KINGSTON UNIVERSITY LONDON

Hydrocarbon Quantification using Neural Networks and Deep Learning based Hyperspectral Unmixing

---

Author:

Asmau Mukhtar AHMED

This thesis being submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

November 22, 2019

Declaration of Authorship

I hereby declare that this thesis and the work presented in it are my own. I confirm that: this work was done wholly while in candidature for a research degree at this University and have not been submitted in whole or in part for consideration for any other qualification in this, or any other University. Where I have consulted the published work of others, this is always clearly attributed. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

Signed:

_____

Date:

_____

Abstract

Hydrocarbon (HC) spills are a global issue, which can seriously impact human life and the environment, therefore early identification and remedial measures taken at an early stage are important. Thus, current research efforts aim at remotely quantifying incipient quantities of HC mixed with soils. The increased spectral and spatial resolution of hyperspectral sensors has opened ground-breaking perspectives in many industries including remote inspection of large areas and the environment. The use of subpixel detection algorithms, and in particular the use of the mixture models, has been identified as a future advance that needs to be incorporated in remote sensing. However, there are some challenging tasks since the spectral signatures of the targets of interest may not be immediately available. Moreover, real time processing and analysis is required to support fast decision-making. Progressing in this direction, this thesis pioneers and researches novel methodologies for HC quantification capable of exceeding the limitations of existing systems in terms of reduced cost and processing time with improved accuracy. Therefore the goal of this research is to develop, implement and test different methods for improving HC detection and quantification using spectral unmixing and machine learning. An efficient hybrid switch method employing neural networks and hyperspectral is proposed and investigated. This robust method switches between state of the art hyperspectral unmixing linear and nonlinear models, respectively. This procedure is well suited for the quantification of small quantities of substances within a pixel with high accuracy as the most appropriate model is employed. Central to the proposed approach is a novel method for extracting parameters to characterise the non-linearity of the data. These parameters are fed into a feedforward neural network which decides in a pixel by pixel fashion which model is more suitable. The quantification process is fully automated by applying further classification

techniques to the acquired hyperspectral images. A deep learning neural network model is designed for the quantification of HC quantities mixed with soils. A three-term backpropagation algorithm with dropout is proposed to avoid overfitting and reduce the computational complexity of the model.

The above methods have been evaluated using classical repository datasets from the literature and a laboratory controlled dataset. For that, an experimental procedure has been designed to produce a labelled dataset. The data was obtained by mixing and homogenizing different soil types with HC substances, respectively and measuring the reflectance with a hyperspectral sensor.

Findings from the research study reveal that the two proposed models have high performance, they are suitable for the detection and quantification of HC mixed with soils, and surpass existing methods. Improvements in sensitivity, accuracy, computational time are achieved. Thus, the proposed approaches can be used to detect HC spills at an early stage in order to mitigate significant pollution from the spill areas.

Acknowledgements

First and foremost, I would love to express my sincere and utmost gratitude to my supervisor, Dr. Olga Duran, for her contribution of time, mentorship, support, patience and ideas over the past four years. Thank you Dr. Olga Duran, for always letting me in if I need to discuss anything and for being very understanding. I will also like to thank Dr. Yahya Zweiri and Dr. Mike Smith for their advices, contributions and meaningful discussions and insightful comments in many ways as my co-supervisors. Their advice helped me grow as a researcher and an individual.

I am highly indebted to my Dad Col. U F Ahmed (rtd) for the countless time he has supported me during my study. I hope this achievement will complete the dream he has for me.

My special appreciation to Prof. Suleiman Elias Bogoro, Executive Secretary Tertiary Education Trust Fund (TETFund) and Kaduna State University for providing me an excellent opportunity to be involved in high standard research and obtaining my PhD.

I would like to acknowledge and extend my heartfelt gratitude to my parents and siblings for all the support and love they showed me which helped in the completion of this work.

I want to express thanks to my colleagues at room 132 for the wonderful time we shared together on the journey to our PhD. Finally, the biggest thanks go to my beloved husband and daughters for their patience and understanding, they have offered everything possible to support me. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

List of Abbreviations

AE　　　　Auto Encoders

AISA　　　Airborne Imaging Spectrometer (for) Application

ALSMA　　Adaptive Linear Spectral Mixture Analysis

ANC　　　Abundance Nonnegative Constraint

ANN　　　Artificial Neural Networks

ASC　　　Abundance Sum-to-one Constraint

AVIRIS　　Airborne Visible Infrared Spectrometer

CCA　　　Convex Cone Analysis

CNN　　　Convolutional Neural Networks

DBN　　　Deep Believe Network

DCA　　　Dependent Component Analysis

DCGAN　　Deep Convolutional Generative Adversarial Networks

DNN　　　Deep Neural Networks

DRIFT　　Diffuse Reflectance Infrared Fourier Transform

FCLS　　　Fully Constraint Least Square

FM　　　　Fan Model

FTIR　　　Fourier Transform Infrared

GBMM　　Generalized Bilinear Mixing Model

GCMS　　Gas Chromatography Mass Spectrometer

GRU　　　Gated Recurrent Units

HC　　　　Hydrocarbon

ICA　　　Independent Component Analysis

ICE　　　Iterative Constrained Endmember

IFOV	Instantaneous Field (of)View
LSTM	Long Short Term Memory
MESMA	Multiple Endmember Spectral Mixture Analysis
MCSUM	Monte Carlo Spectral Unmixing Method
MRF	Markov Random Field
MVSA	Minimum Volume Simplex Analysis
MWR	Microwave Radiometer
NIR	Near Infrared
NM	Nascimento Model
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
PPNMM	Polynomial Post Nonlinear Mixing Model
PPI	Pixel Purity Index
RBF	Radial Basis Function
ReLU	Rectified Linear Units
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
ROI	Region of Interest
RULU	Robust Unconstrained Linear Unmixing
SAD	Spectral Angular Distance
SAR	Synthetic Aperture Radar
SAM	Spectral Angular Mapper
SCG	Scale Conjugate Gradient
SGA	Simplex Growing Algorithm
SGD	Stochastic Gradient Descent
SLAR	Side Looking Airborne Radar
SMP	Subspace Matching Pursuit
SOM	Self Organizing Map

SNR   **S**ignal to **N**oise **R**atio

SU   **S**pectral **U**nmixing

SVM   **S**upport **V**ector **M**achine

SWIR   **S**hort **W**ave **I**nfrared

TIR   **T**hermal **I**nfrared **R**egion

UV   **U**ltraviolet

VCA   **V**ertex **C**omponent **A**nalysis

List of Symbols

a	Abundance reflectance spectra
b	Non-linearity parameter
m	Spectral signature
P	Pixel spectrum
R	Number of endmembers
p	p^{th} pixel
r	Endmember index
n	Additive noise
L	Number of bands
y_p	Measured spectrum
$\odot$	Hadamard product
k	Number of epochs
w	Weights of neuron
b	biases of neuron
e	error for an output neuron
T	Transpose operator
α	Learning rate
β	Momentum factor
γ	Proportional Factor
δ	angular frequency

Dedicated to my dear husband and daughters

Chapter 1

Introduction

Intensive Petro-chemical activities around industrial facilities and oil production sites have led to an increase in the contamination of soils with hydrocarbons causing serious concern to the environment [1]. Between 1988 and 2000, a total of about 2,475 spills were recorded in Toronto and the surrounding regions with a total release of about 800,000 litres of oil [2], also a total of about 5 million barrels of crude oil were recorded from spills in the Gulf of Mexico in 2010 [3]. Other reported spills includes; Exxon Valdez oil spill in Alaska, the Prestige oil spills in Spain and the Erika oil spills in France, etc [4]. Whilst the economic significance of Hydrocarbons (HCs) is attributed to their primary use as fuel and versatile application in downstream industries, they can have severe environmental consequences [5]. Oil exploration, processing and production represents a potential environmental exposure of HCs and because of these activities may result in accidental spillage thereby altering the physical and chemical properties of soils, causing toxicity and limiting soil quality. HCs are therefore environmentally harmful because they can become mutagenic [6]. Since HC contamination raises serious concerns for human health and the environment, there is the need for rapid identification of affected areas in order to allow swift site characterization and prioritization of remediation actions [7].

Measuring spills, particularly in large fields is often difficult because it involves expensive and time consuming process such as inspection of large

areas, surveillance of the affected area and other field measurements.

Remote sensing sensors mounted on airborne platforms has been a promising tool, which can be used to inspect large inaccessible areas in a short time.

Hyperspectral sensors measure signal at hundreds of wavelength ranging from visible to the infrared region of the electromagnetic spectrum; which provides a wealth of information to spectrally distinguish between materials covered in a scene. Due to the rich information content covered by a hyperspectral sensor, it is well suited to the task of oil spill detection and monitoring.

Spectral unmixing (SU) is a source separation problem which has received interest for over a decade [8]. SU is a key process in identifying the spectral signature of different cover types (materials) within a pixel or group of pixels and estimating their spatial distribution over an image. Pixels that contain more than one cover type are called mixed pixels while pixels that have only one cover type are referred to as pure pixels [9]. SU begins with identifying the different cover types referred to as endmembers and a set of corresponding fractions (abundances) that indicate the proportion of each endmember present in the pixel [10]. SU is used to identify the mixing model as observed on the dataset. This explains how the endmembers are mixed, and forms the mixed spectrum as measured by the sensor. Once the mixing model is defined, SU is used to estimates the endmembers and calculate the abundances as measured by the sensor [11]. The two common spectral unmixing methods are the linear and nonlinear methods where the linear mixing methods happens to be the common approach, this method assumes that each photon (incident electromagnetic radiation) comes in contact with only one material before reaching the sensor while the nonlinear mixing methods assumes that interaction occurs at microscopic level, it copes with nonlinear interactions effects that exist in an image or dataset.

As far as we know, no work in hybridizing the linear and nonlinear approaches has been presented in the literature, and in particular the selection of the most appropriate technique in using the two methods. The question on whether a mixed pixel is better explained with a linear or nonlinear process is still an open research problem in spectral analysis.

Research in the field of spectral unmixing of hyperspectral data is not new, but the hybridization of the two methods and the choice of best method to apply for oil spill detection has remained understudied.

Artificial intelligence such as neural networks has been successfully used for abundance estimation with good accuracy [12], however, this method is relatively slow, requires sufficient training data and the problem of overfitting. Therefore the need to explore a more advance technique such as deep learning approach for abundances estimation is considered, which could be fast, avoid overfitting and predict excellent accuracy as demonstrated in [13]. To the best of my knowledge, no studies has been demonstrated to use a deep learning approach for hydrocarbon spill detection and quantification.

The aims of this work is to study the problem of real world hydrocarbon detection and quantification and to propose, implement and test different methods for improving hydrocarbon detection and quantification using data. The key questions for the project are as follows:

- Is information extracted from the data sufficient to evaluate if materials abundances are better (in terms of accuracy and computational time) explained with a linear or nonlinear spectral model?

- Which model is more suitable for the problem of a soil and HC mixture?

- Can a deep learning model be successfully applied to extract abundances of a HC and soil mixture and thus quantify HC levels?

- Can the deep learning model generalise avoiding overfitting with limited datasets?

1.0.1   Hypothesis

To investigate this problem, three hypothesis are considered in this work as follows:

- The choice of best method for spectral unmixing of hyperspectral data can be selected through hybridization between the linear and nonlinear methods.

- The optimum model choice can be extracted from the pixel scene information and varies in a pixel by pixel basis.

- Spectral unmixing methods can be optimized by applying deep learning approach to extract the abundances.

1.0.2   Objectives

To investigate the above hypothesis, the following objectives are to be achieved.

- To propose a hybrid methodology to choose the optimum between linear and nonlinear spectral unmixing methods using artificial neural networks.

- To improve the accuracy of the unmixing techniques by applying deep learning approach to estimate the abundances in a given hyperspectral data.

- To apply the above techniques for oil spill detection and quantification and also identify experimentally which spectral unmixing methods are most suitable to this application.

1.0.3   Contributions

The main contributions of this thesis are:

1. A new hybrid spectral unmixing method for switching between linear and non-linear spectral unmixing methods using artificial neural networks. It is observed that some nonlinear methods perform better in scenes with multiple interactions and a complex mixture of features commonly composed of different cover types, whilst the linear model is appropriate for images that have a single cover type of material in a pixel. The method was developed in such a way that it has the ability to fit the measured spectra and provide an accurate estimation of the abundances.

2. A new deep learning unmixing method is developed. The deep learning algorithm estimates the abundances in a mixed pixel. The architecture of the model utilizes Sigmoid activation function, the network was trained with a three-term back propagation algorithm in order to make the model converge rapidly and dropout was used to avoid overfitting and reduce the complexity of the nets computations. The advantage of using the deep learning approach is the ability of the model to generalize on unseen data .

3. The application of the proposed methods to oil spill detection and the choice of most appropriate method suitable to this application was demonstrated. A real dataset was produced and labeled.

1.0.4   Publications

Journal Papers

1. Ahmed, A.M., Duran, O., Zweiri, Y. and Smith, M., 2019. Quantification of hydrocarbon abundance in soils using deep learning with dropout and hyperspectral data. Remote Sensing, 11(16), p.1938.

2. Ahmed, A. M., Duran, O., Zweiri, Y. and Smith, M., 2017. Hybrid spectral unmixing: Using artificial neural networks for linear/non-linear switching. Remote Sensing, 9(8), p.775.

Conferences

1. Ahmed, A.M., Duran, O., Zweiri, Y. and Smith, M., 2018, November. Quantitative analysis of petroleum hydrocarbon contaminated soils using spectroscopy, spectral unmixing and deep neural networks. In Image and Signal Processing for Remote Sensing XXIV (Vol. 10789, p. 107890N). International Society for Optics and Photonics.

2. Ahmed, A.M., Duran, O., Zweiri, Y. and Smith, M., 2018, September. Application of Hybrid Switch Method to Quantify Oil Spills. In 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-5). IEEE.

3. Ahmed A. M., Duran, O., Zweiri , Y. and Smith, M., 2017, March. Hybridization between Linear and Nonlinear Spectral Unmixing Based on Real and synthetic data: Remote Sensing and Photogrammetry Society (RSPSoc) Annual Conference.

4. Ahmed A. M., Duran, O., Zweiri , Y. and Smith, M., 2016, December. Comparison between Linear and Nonlinear Spectral Unmixing Methods: Geological Remote Sensing Group (GRSG) 27th Annual Conference.

1.0.5   Methodology

In order to test the hypotheses in section 1.0.1, a hybrid switch method in image processing to be able to combine two spectral unmixing methods together (linear and nonlinear methods) is proposed. The goal is a set of recommendations based on the two methods for image processing strategies in spectral unmixing based on the accuracy required to specific application. To address the effect of spectral, spatial and temporal variability, the proposed hybrid switch method was based on the characteristics of the mixture type within the neighboring pixels. Values believed to represent the diversity of the neighboring pixels to the ones in its vicinity were considered, these values are the minimum and maximum Spectral Angular Distance (SAD) [14], covariance [15] and a nonlinearity parameter [16].

To improve on the accuracy of spectral unmixing techniques deep neural network approach to spectrally unmix and estimate the abundance in a given dataset is proposed. A motivation for using the deep learning approach is due to the success of the approach in computer vision, image processing and image classification which have seen a great increase in the last few years. Its main advantage is having multiple hidden layers compared to the conventional neural networks which consist of shallow networks with few hidden layers. Deep learning with dropout has been demonstrated to successfully avoid over-fitting and prove to train faster with good generalization capabilities.

To validate the accuracy of the two proposed methods, they were tested on controlled datasets for hydrocarbon spill detection.

1.0.6   Thesis Outline

This thesis is organized in six chapters summarized as follows:

Chapter 1 introduces the origin and hypothesis of the research, highlights the research questions, defines the objectives, summarizes the contributions and publications and outlines the structure of the thesis.

Chapter 2 explains the background of the research, general oil spill problem, and existing methods to detect oil spills are discussed, existing spectral unmixing problems are explained describing the most commonly used methods including artificial neural networks nonlinear methods. and Deep learning approach.

Chapter 3 describes the methodology of the research. First, the new hybrid spectral unmixing method for switching between linear and non-linear spectral unmixing using artificial neural networks was discussed. A number of parameters that are related to the pixels' neighboring characteristics are used, these values are considered to represent the diversity of the neighborhood of the pixel under consideration to the ones in its vicinity which are the minimum and maximum Spectral Angular Distance (SAD), covariance and a nonlinearity parameter.

Secondly, the use of deep learning approach to extract endmembers and estimate the abundances in a given hyperspectral dataset was discussed. The deep learning model utilizes Sigmoid activation function [17] in all the hidden layers, three-term backpropagation algorithm [18] with dropout [19] was discussed to demonstrate how the accuracy of the network was improved.

Chapter 4 describes experimentation and results for validation of the proposed methods from chapter 3. The two methods were used for the identification and detection of oil spills experimentally using a controlled hyperspectral data. Statistical tests were derived for pixel-by-pixel application to quantify hydrocarbon spills using the proposed methods. The proposed tests are computationally efficient and thus are applicable in practice to quantify oil spills.

Chapter 5 discusses the results of the experimentation from chapter 4.

Chapter 6 provides the conclusion of the thesis, highlights the contributions and discusses possible directions for future research.

Chapter 2

Background Theory

In this chapter, general oil spill problems are discussed, explaining the existing methods used for oil spill detection. The basic idea of spectral unmixing will be introduced, explaining the most commonly used spectral unmixing methods. Deep learning spectral unmixing methods will also be discussed and literature will be discussed demonstrating the usefulness of the approaches in machine learning by showing some examples of recent research areas where they have been successfully applied.

2.1   General Oil Spill Problem

Hydrocarbons (HC) can be described as chemical substances formed exclusively from the combination of carbon and hydrogen. Naturally occurring HC substances occurs in three different states; solid, liquid and gas [20]. These are often a result of the decay of organic substances trapped within sedimentary rocks converted into HCs due to high temperatures and pressures. Crude oil is a natural occurring HC found in a liquid state [20]. Crude oils consist of a complex mixture of various molecular weights; in addition, nitrogen, sulphur and oxygen are found in small quantities [21].

2.1.1   Existing methods to detect oil spills

Whilst the economic significance of HCs is attributed to its primary use as fuel and then versatile application in downstream industries, they can have detrimental environmental consequences [22, 20].

Knowledge about the concentration and nature of a spill is important in order to track their propagation in the environment, assess their risk and damages and propose remediation measures [23, 24]. To effectively protect communities affected by a spill, fast and accurate determination of the area impacted is needed, particularly if monitoring large regions affected by an oil spill or where aging transporting facilities are involved [25] .

Traditional methods employed to track and detect oil spills and the concentration in soils often involve processes which are expensive and time consuming as they mostly require field sampling, chemical analysis and geostatistical interpolation [26, 27]. One such method is the collection of a physical soil sample, then determining the HC level by gravimetric, Fourier- Transform Infrared (FTIR) spectroscopy [7].

Another includes the Gas Chromatography -Mass Spectrometer (GC-MS) method and infrared spectroscopy [28]. Imaging spectroscopy has been recognized as an alternative method for detecting HCs in soils which has proven to be rapid, efficient and cost effective [24]. Spectroscopy measures the diffuse reflected electromagnetic energy from source material to a light source [27]. Spectroscopy analysis includes the Visible - Near-Infrared (Vis-NIR) and Short Wave Infrared (SWIR) spectrum, which has been demonstrated to be a powerful tool for the measurement of HC concentration in soils [26].

More specifically, Near and Shortwave Infrared (NIR – SWIR) spectroscopy has been a popular method for detecting, mapping, quantifying and characterizing HCs in contaminated soils with reasonable accuracy [29, 24]. Moreover, NIR – SWIR spectra provide good information on soils' organic and

inorganic material content [29]. HCs demonstrate good absorption in spectral bands between 1200 mm, 1725 nm and 2310 nm [23, 26, 30]. Therefore, spectral information obtained in the NIR – SWIR range is excellent for both the quantitative and qualitative analysis of detecting HCs in soils [29].

Different methods have been used to analyze reflectance spectroscopy data to detect HCs in soils; Okparanma and Mouazen [23] used regression analysis and spectral pre-processing to generate statistical models to identify different HC products mixed with a mineral substrate. Scafutto *et al* and Webster *et al* [29, 30] in their analyses used Principal Component Analysis (PCA) and Partial Least Square (PLS) regression; they used PCA to differentiate the types and density of the HCs in soils and used PLS regression analysis, which is a multivariate method and includes the correlation between spectral information and corresponding analytical data to rapidly predict the concentration of HCs in soil.

Schwartz *et al* [7] used Spectral Angular Mapper (SAM) to classify oil spills on an image and also used signature matching to distinguish oils from other features, while Webster *et al* [30] reveals that Diffuse Reflectance (mid) Infrared Fourier Transform (DRIFT) spectroscopy; a handheld spectrometer for the prediction of total petroleum hydrocarbon and analysis with PLS regression methods yield positive and accurate results with rapid predictions of HC concentrations in soil. Other researchers show the robustness of visible and infrared spectroscopy for the rapid estimation of HC and polycyclic aromatic hydrocarbons [31, 7].

The main point of concern is the diagnostic absorption band within the visible and the infrared region of the spectrum where both HCs and soils can be detected and analyzed qualitatively and quantitatively.

2.1.2   Satellite sensors for oil spill detection

Remote sensing has proven to be a powerful tool for oil and gas exploration by increasing the understanding of the hydrocarbon system from the surface. It is useful in remote areas where little is known. Different satellite sensors are used for oil spill detection, with systems ranging in spatial resolution from 3 to 30 m depending on the sensor type and the objective of the study with the airborne sensors having 3 m resolution with majority of the satellite sensors within the range 30 m resolution [32]. Airborne surveillance is limited due to the high cost and so tend to be less efficient for wide area surveillance and their limited coverage [33]. Sensors used for HC spill detection range from multispectral (1 - 10 bands) to hyperspectral (hundreds of bands). Hydrocarbon specifically give rise to spectral manifestations around 1200 μm, 1700 μm and 2300 μm within the shortwave infrared (SWIR) portion of the electromagnetic spectrum, Figures 2.1, 2.2 and 2.3 shows spectrum of soil, water and hydrocarbon.
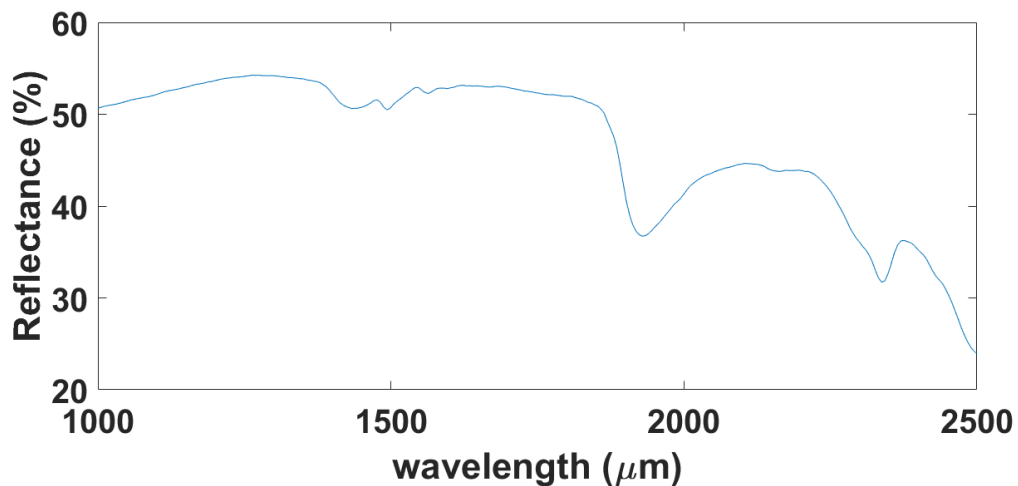
FIGURE 2.1: Spectral reflectance of soil
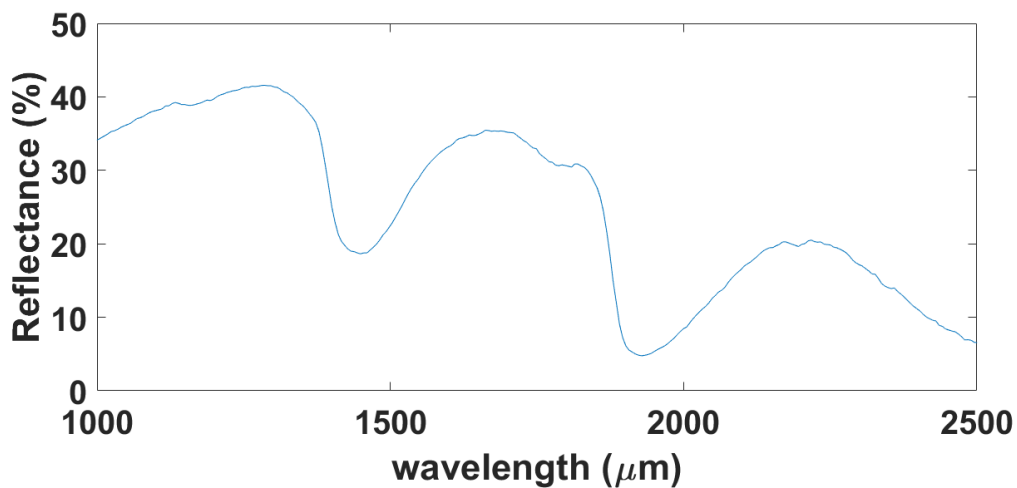
FIGURE 2.2: Spectral reflectance of water
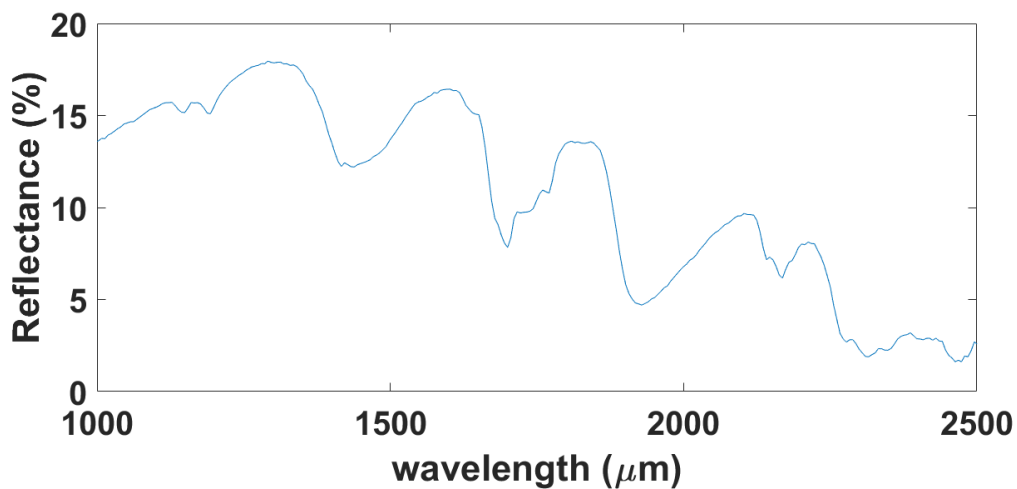
FIGURE 2.3: Spectral reflectance of hydrocarbon

Visible sensors

The Visible systems and aerial photography were the two commonly used airborne sensors used in the early 1970s. Visible sensors are still widely used for HC spill despite their limitations [34]. HC spills are often identified because the spectral reflectance of HC is higher than that of water, however HC also absorbs some radiation in the visible region, and so these sensors are not good for HC detection as it is often difficult to distinguish HC from the background, such as sun glint and wind sheen which create a similar impression to an oil sheen. Also, the presence of seaweed and a darker shoreline may

often be mistaken for HC spills. Another limitation of passive visible sensors is that they cannot operate at night as they depend on sunlight reflectance. The main advantage of the visible sensors is that they are inexpensive and easy to use, therefore they are commonly used to create data in coastal areas [35].

Microwave sensors

Microwave Radiometer (MWR) is a passive sensor commonly used for ocean oil pollution monitoring. These sensors are often preferred to other sensors due to their capability to capture data in all-weather and all night. In addition to detecting oil spills, they also measure oil thickness resulting from interference of radiation within the upper and lower boundaries of the oil film [35]. The MWR sensor works well in adverse weather condition, and also during day and night, however biogenic materials can also produce signals which are similar to oil and thus leads to false positives. Its major limitation is that it requires information about the characteristics of the environment and the properties of oil in order to detect spills [34].

Radio Detection and Ranging (Radar)

Radar is an active sensor operating within the radio wave region. The EMR is reflected by capillary waves on the ocean and thus produces a bright image for ocean water. If oil is present in the ocean, the reflectance is reduced because oil diminishes capillary waves, thus the presence of oil can be detected as a dark part in a bright ocean [36]. However, there might be some interferences or false targets in areas where ice or fresh water flow such as fresh water slicks, wind slicks, shallow seaweed beds, glacial floors and biogenic oils [37]. The main advantage of RADAR is its use in detecting areas with large spills in water as a first assessment tool to possibly locate oil spill.

Its major limitation is that, it can only provide accurate results for spill detection in offshore areas which makes it unsuitable for detection of spills on land surface.

Synthetic Aperture Radar (SAR) and Side - Looking Airborne Radar (SLAR) are the two most commonly used types of radar sensors. SAR is often preferred to SLAR because it has a higher spatial resolution and range [36]. SLAR is less expensive and predominantly used for airborne remote sensing, as a result SAR is commonly used for ocean detection monitoring [35]. SAR is an efficient imaging tool for oil spill detection, although it does not recognize oil type or estimate oil thickness but it is useful for searching large areas and observing oceans at night and during cloudy weather conditions as demonstrated in [33]. In addition to Radar sensor, other space borne sensors capable of monitoring oil spills include RADARSAT-1, Moderate resolution Imaging Spectro radiometer (MODIS) [38]. RADARSAT and Landsat Thematic Mapper (TM) have been demonstrated to detect oil spills in Guanabara bay, Brazil. These two sensors provided suitable temporal coverage while cloud, haze and its 8 days revisit schedule prevented Landsat TM from being used systematically for oil spill monitoring [39].

Ultraviolet sensors (UV)

UV wavelengths (0.32 - 38 m m) can also be used to detect oil spills due to the high reflectivity of thin oil layers, however; they cannot detect oil thicknesses greater than 10 microns, and can only offer information about the relative thickness of oil slicks [40]. Its main limitation is that biogenic material, wind sheen, sun glint as well as sea-weed often causes false positives in the data. In most cases, the UV images are overlayed on infrared images so as to generate oil spill thickness maps [34].

Laser fluorosensors

The laser fluorosensor is a useful instrument for the identification of oil spills since it has the ability to identify oil on different surfaces such as water, soils, weed, ice and snow. The laser fluorosensor also allows the discrimination between different oil types [35]. Oil can be identified when the oil aromatic compound interacts with ultraviolet lights then the light energy is absorbed and subsequently releases extra energy as visible light [34]. Other substances in water such as chlorophyll, fluoresce at distinct wavelengths thus, giving oil a unique spectral signature. It is also able to discriminate between different oil types as they exhibit distinct fluorescence emission signatures.

Infrared sensors

These sensors absorbs and emits part of its radiation as thermal energy within the Thermal Infrared Region (TIR) (8 - 14 m). Oil often has a lower emissivity than water in the TIR and this is typically used for oil spill detection as it has distinctively different spectral signature [41]. Infrared sensors also provide information about the relative thickness of oil slicks, however, these sensors are unable to detect emulsions of oil in water because the emulsion contains 70% water and the thermal properties of emulsion are similar to the background water [35]. An example of such sensor is the Hyperspectral (HS) sensor which collects ten to hundreds of spectral bands for spectral signatures and has high potential for extraction of endmembers and estimates their abundances. HS spectral sensors can have more than 200 wavelengths which makes them capable of exploiting the spectral signatures of different oil types. HS sensors have made it possible to detect oil and gas hydrocarbon information both on land and sea surfaces. HCs have diagnostic absorption bands in the Near and Shortwave Infra-red regions (NIR - SWIR: 700 - 3000 nm) which are derived from the combination of bands or overtones of C – H

stretching modes [6], and mostly occur in the $1720 - 1750$ nm, $2310 - 2350$ nm wavelengths with the strongest absorption features at 2310 - 2350 nm. These often overlap with other mineral absorption features such as calcite but the absorption feature shape differs and does not have a corresponding feature at 1720 - 1750 nm which tends to be weaker at these wavelengths [42]. These features can be used to detect HCs in other to obtain qualitative and quantitative information about the HC [6]. However, the HS sensor used in detecting HC content mixed in soil sample, or the mapping of surface seeps and contaminated soils depends largely on the composition, density and concentration of the HC type and also the soil properties and particle grain size [43]. The spectral information can be used to discriminate between light and heavy crude oils, whilst minute concentrations of crude oil can also be detected with hyperspectral images. The most commonly used HS sensors include the Airborne Visible/Infrared Spectrometer (AVIRIS) and Airborne Imaging Spectrometer for Application (AISA).

The use of a hyperspectral sensor for oil spill detection has been demonstrated in [44] where the United States West Virginia University successfully detected oil and gas microleakage along the Californian coast, where several oil and gas fields are found. This was achieved by delineating the distribution of the hydrocarbon leakage of mineral alteration on the surface using the AVIRIS Airborne Hyperspectral Imager. Another example includes the GEO-SCIENCE company in Australia where the process of surface hydrocarbon leakage and undersea hydrocarbon microleakage reservoirs were detected using the HYMAP airborne hyperspectral sensor [45].

2.2 Spectral Unmixing

Spectral unmixing (SU) is a method of identifying the spectral signatures of materials in an image known as endmembers. SU is applied to estimate the

relative abundance of the endmembers as measured by the sensor [10]. SU involves the analysis of different satellite data such as hyperspectral and multi-spectral data in various fields of application which includes remote sensing, mineralogy, material science and also planetary science [11]. To unmix the spectra of an image, SU often requires the definition of the mixing model underlying the observations as presented on the data. A mixing model describes how the endmembers combine to form the mixed spectrum as measured by the sensor [11]. Having defined the mixing model, spectral unmixing is used to estimate the inverse of the formation process of the data and quantify the endmembers and abundance of interest from the collected spectra [11]. This process is often achieved through a radiative transfer model which explains the process of light scattering by the different cover types in a scene [46].

A hyperspectral image consists of spatial and spectral dimensions of a data corresponding to the different wavelength at which the scene is captured often in the form of reflectance. This can be described as the vector or a spectrum of values corresponding to each pixel in a scene [47]. This spectrum is made up of different materials in a scene such as soil, water, and vegetation etc. These materials are otherwise known as endmembers as demonstrated in Figure 2.4.
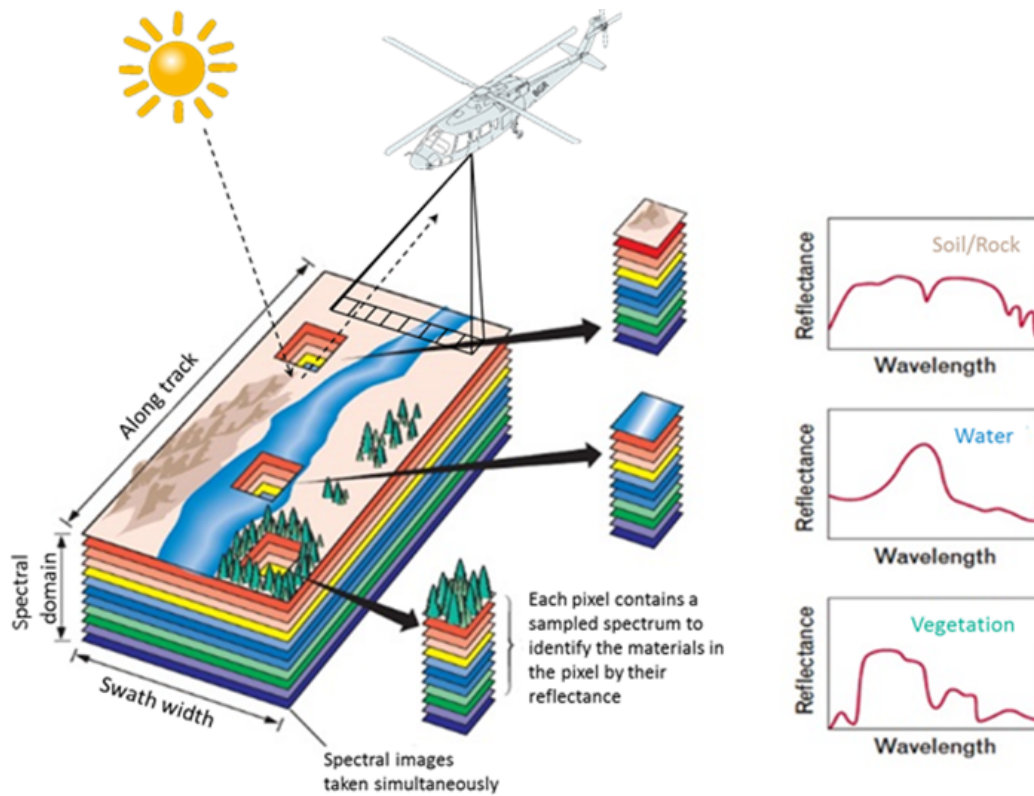
FIGURE 2.4: Example of hyperspectral image concept [48]

In order to decompose the mixed pixels into different materials that are within a given pixel, by identifying the endmembers and their corresponding fractional abundances, SU is applied. This is important in hyperspectral image processing and is commonly referred to unmixing problem [49]. The linear and nonlinear spectral unmixing methods are the most commonly used techniques in decomposing mixed pixels in a hyperspectral data.

2.2.1 Unmixing Strategies

Two steps are involved in spectral unmixing analysis of a dataset, namely, endmember extraction and abundance estimation. Usually most of the methods tackle the two steps, here different methods are described below:

Endmember extraction

Endmember extraction is the first and the most important step in the un-
mixing analysis of an image. It is the process of extracting pure material
signatures of the different features present in an image which forms the basis
of image analysis [50]. This process is mostly based on the results of a sys-
tem involved in a linear equation [51]. In a situation where the endmembers
correspond to real pure substances, the process is challenging, however, in
a highly mixed scenes, endmembers may be difficult to identify commonly
due to low spatial resolution [52].

Different methods for endmember extraction exist [50] these are classified
into two types:

1. The first are the methods that assume the presence of pure pixels of each
   endmember in an image defined by the vertices of the simplex, these
   includes the Vertex Component Analysis (VCA), Robust Unconstrained
   Linear Unmixing (RULU), Endmember Extraction Method (EEM), Sim-
   plex Growing Algorithm (SGA), Minimum Volume Simplex Analysis
   (MVSA), Pixel Purity Index (PPI), N-FINDR, Convex Cone Analysis
   (CCA) etc [50]. These methods identify the purest pixels within the
   dataset.

2. The second are those methods that estimate sets of smaller volumes
   embedded in the data; these algorithms do not need to have pure pix-
   els in an image. These include the Iterative Constrained Endmember
   (ICE), Independent Component Analysis (ICA), Dependent Compo-
   nent Analysis (DCA), Minimum Volume Simplex Analysis (MVSA) etc
   [50].

Abundance estimation methods

After extraction or detection of the endmembers from a spectral library or data set, the next step is inversion; which is a procedure used to estimates the abundance vectors of an image pixel [47]. Fully Constrained Least Square (FCLS) inversion algorithm has been used extensively in the literature and has gained popularity for inversion technique [53]. This algorithm has shown excellent result in different applications. Another algorithm is the Bayesian methods, which has also been used by several authors [11]. Other methods in the literature include Linear Spectral Unmixing, Spectral Angle Mapper etc.

The linear spectral unmixing method is the most common approach to spectral unmixing analysis [11, 46], it assumes that each photon reaching the sensor interacts with only one material in a scene as measured by the spectrum [11]. The linear spectral unmixing methods have recorded a number of excellent results [10], the most commonly used methods includes; Adaptive Spectral Mixture Analysis (ALSMA) [54], Subspace Matching Pursuit (SMP) [55], Orthogonal Matching Pursuit (OMP) [56]. Li *et al* [57] proposed a robust collaborative sparse regression method to unmix a hyperspectral dataset using linear mixture model. Thouvenin *et al* [58] proposed a linear mixing model which directly detects spatial and spectral endmember variability in a hyperspectral dataset. Foody *et al* [59] used linear mixture model and regression based fuzzy membership function to estimate land cover composition whilst [60] used Vertex Component Analysis (VCA) algorithm to unmix a hyperspectral data with relatively low computational complexity compared to other conventional methods.

Another effect that has been considered in spectral analysis is endmember variability, which occurs due to atmospheric and temporal conditions. Machine learning methods have worked well to account for spectral variability

in hyperspectral image analysis.

The combination of both spectral information and spatial context may help to improve the performance and accuracy of unmixing and classification of a hyperspectral data [61]. Techniques such as morphological filters [62], Markov Random Fields (MRF) [63, 64, 61, 65], Support Vector machines (SVM) [66] and Self Organizing Maps (SOM) [67] among others have been proposed to impose spatial information. MRF, in particular, is a powerful tool used to describe the neighborhood dependence between image pixels and has demonstrated to provide good results for image classification. They also provide accurate results for unmixing under the Bayesian inferring framework when spatial information is being incorporated [61]. MRF is a method that combines information obtained from spatial correlation into the posterior probability distribution of the spectral features [63]. SVM has demonstrated to perform well with high classification accuracy when applied to dataset with a limited number of training samples [68].

2.3 Linear Spectral Unmixing Model

The linear unmixing model proposed by [69] is based on the assumption that there are no multiple scattering between different cover types, it is assumed that each photon interact with a single material in a scene [70] as demonstrated in Figure 2.5.
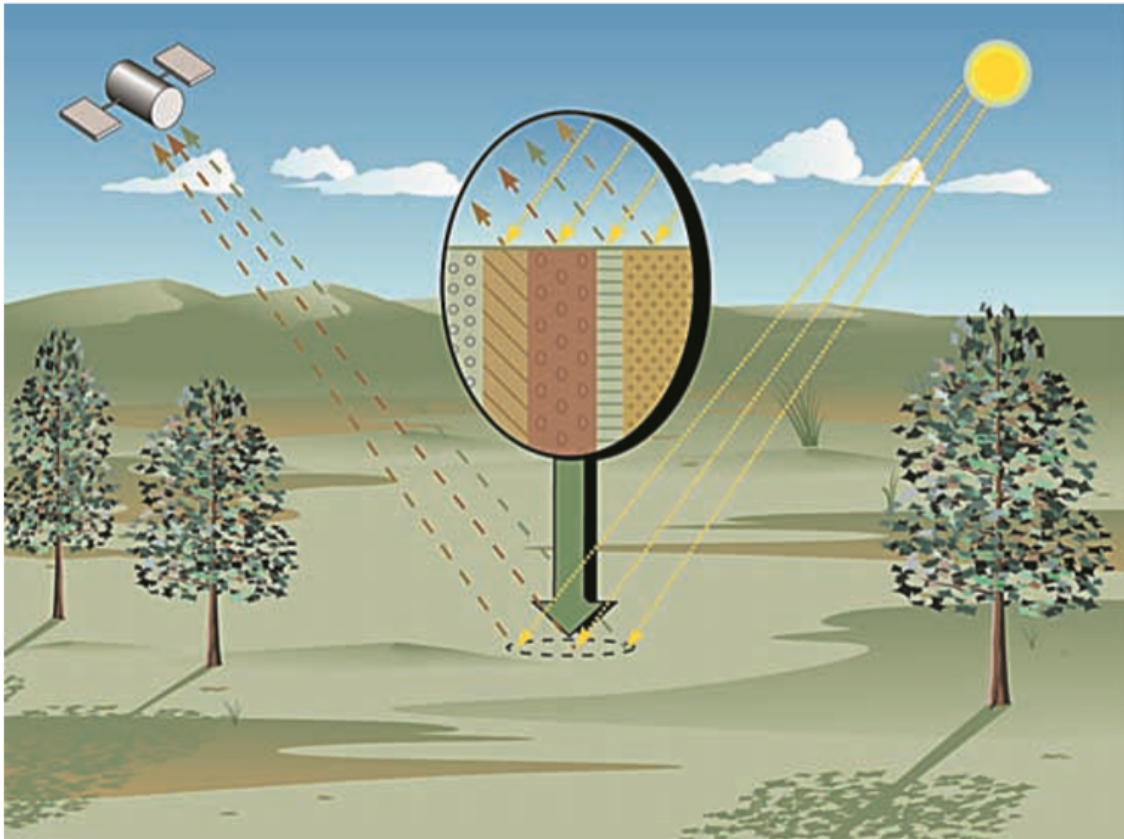
FIGURE 2.5: Linear mixing model [10]

The linear unmixing model is the most commonly used unmixing process, especially in scenes where the endmembers and spectral reflectances are known, the linear mixing model is then applied to estimate the relative abundances of endmembers in each pixel. To obtain the abundances using the linear unmixing, two constraints are imposed on the model; the first requires the sum of abundances for fractions of materials present in an image pixel to be one and the second constraint imposes a condition that the abundance fractions should be non-negative. In practice, the first constraint is easy to implement while the second constraint is difficult to handle because of the irregularities that occurs and can only be solved using numerical methods [53]. However, problem often occur if the model is forced on the two constraints, especially if models information is inaccurate: which could be in the form of number of endmembers and their corresponding signatures. In

the linear mixing model, it is often assumed that a set of endmembers partic-
ipates in the linear mixing problem of all the pixels in an image scene [21]. A
number of research has demonstrated the effect of using too many endmem-
bers in the unmixing process which leads to an error in the abundance esti-
mation and makes the model sensitive to errors such as instrumental noise,
atmospheric contamination and random spectral variation [71].

The linear mixture model has been widely applied by various research
to identify and quantify pure components in remotely sensed images. This
is due to its simple physical interpretation and estimation process [72]. Dif-
ferent image processing techniques can be applied to a hyperspectral image
using the linear mixture model. However, there is a significant difference
between a pure pixel and mixed pixel analysis. Pure pixel is a spatial anal-
ysis technique while the mixed pixel is often processed via spectral analysis
[53]. Mixed pixel classification is used to estimates the abundance fractions
of materials in a particular pixel and classifies them based on their estimated
abundance fractions whereas the pure pixel classification has to do with as-
signing each pixel with a standard class membership.

Linear spectral mixture analysis generally requires a known knowledge
of the material signature present in an image which in most cases is not avail-
able; therefore the selection of an appropriate set of material signatures is the
first and important step for the successful performance of a linear spectral
mixing analysis [51].

When the mixing scale formed in a scene is macroscopic, i.e. where each
photon reaching a sensor interacts with only one material in a scene, then the
measured spectrum $\mathbf{y}_p \in R_L$ in the pth pixel can be accurately described by
the linear mixing model (LMM) [11] which is often defined as

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + \mathbf{n}_p \tag{2.1}$$

Where R is the number of endmembers present in the image, $\mathbf{m}$ is the spectral signature of the rth endmember, $a_{r,p}$ is the abundance of the rth material in the pth pixel and $\mathbf{n}_p$ is an additive noise and modeling error.

The state - of- the- art most commonly used methods for linear spectral unmixing are discussed next.

2.3.1 Pixel Purity Index (PPI)

This algorithm is a well-known approach used for endmember extraction and abundance estimation by different researchers in spectral analysis [73]. This algorithm aims to locate and finds spectrally pure pixels in an image. This is achieved by repeatedly projecting the data in an n-dimensional scatter plot randomly into unit vectors that holds the transformed data [50]. PPI algorithm projects all the spectral vector in an image onto skewers, defined as a large set of random vectors. The points corresponding to extremes for skewer directions are stored [8]. An account is used to keep records of times a particular pixel is found to be an extremes for skewer direction; then the algorithm selects the extreme pixels based on high score ranking as demonstrated in Figure 2.6.
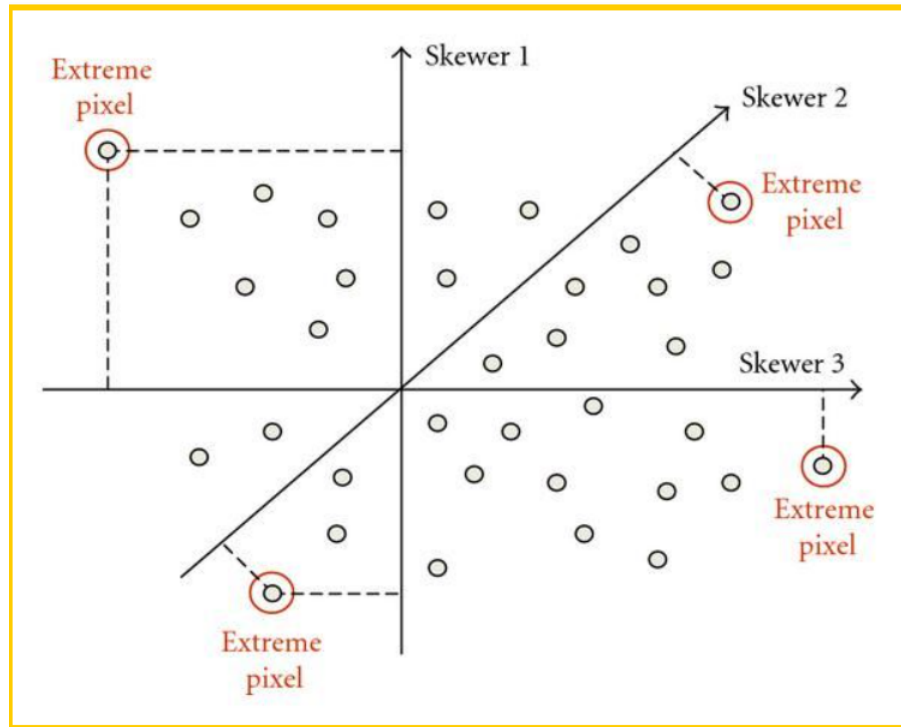
FIGURE 2.6: Schematic representation of the pixel purity index
algorithm [74]

2.3.2   N FINDR

The main aim of this algorithm is to find the simplex of maximum volume
that can be inscribed within a hyperspectral data using nonlinear inversion
technique [75]. Each pixel in the image is evaluated for its likelihood of either
being pure or almost pure pixel; this is obtained by calculating the volume of
each pixel and considering them as endmembers. A trial volume is calculated
for all the pixel in each endmember and replacing them, again the volume
of the endmembers are recalculated, if the calculated volume increases, the
pixels are used to replace the endmember. The process is repeated until all
the pixels are exhausted and no more replacements [50].

2.3.3   Iterative Error Analysis (IEA)

This algorithm performs some constrained unmixing analysis, this algorithm
select the pixels that minimize the remaining error in the unmixed image as

an endmember [50]. This process starts when an initial vector is chosen, then a linear spectral unmixing analysis is performed and the error calculated. The vector that corresponds to the pixel with largest error is located and averaged to obtain the endmember vector. This process is repeated until all the endmembers are selected [76].

2.3.4   Vertex Component Analysis (VCA)

This algorithm, as proposes by [60] is an unsupervised method which originates on the geometry of convex sets and assumes that the endmembers are within the vertices of a simplex. VCA is an algorithm which presumes the presence of at least one pure pixel in a dataset and projects the data iteratively in the direction orthogonal to the subspace traversed by the endmembers already determined. The new endmembers estimated are equivalent to the extreme of the projection, the process is repeated until all endmembers are used [8].

2.3.5   Fully Constrained Least Square Method (FCLS)

The FCLS algorithm as proposed by [77] is derived from an unconstrained least square based orthogonal subspace projection, which has two constraints imposed to the algorithm; the Abundance Non-negative Constraint (ANC) and Abundance Sum-to-one Constraint (ASC). This algorithm considers negative values as 0 whilst the remaining abundance fractions of the materials are normalized to 1. FCLS uses a simplex method to obtain a set of feasible solutions for spectral unmixing of material signatures while discarding the negative abundance values of the other material signatures to unity. This method requires full knowledge of material signatures in an image [77].

2.3.6   Convex Cone Analysis (CCA)

This algorithm considers the fact that some physical quantities such as radiance and reflectance are non-negative. The vectors formed by these physical quantities can be described as linear combinations of non-negative components, often found within a non-negative convex region [50]. The algorithm is used to locate the boundary points of that region; this is achieved by finding the eigen vectors corresponding to the largest eigen values, then the algorithm tries to look for the boundaries of the convex cone. The linear combinations of these eigen vectors produce strictly non-negative vectors [50]. These convex corners identified are used to represent the endmembers as shown in Figure 2.7 where EV1, EV2 and EV3 are the axes as defined by the eigenvectors.
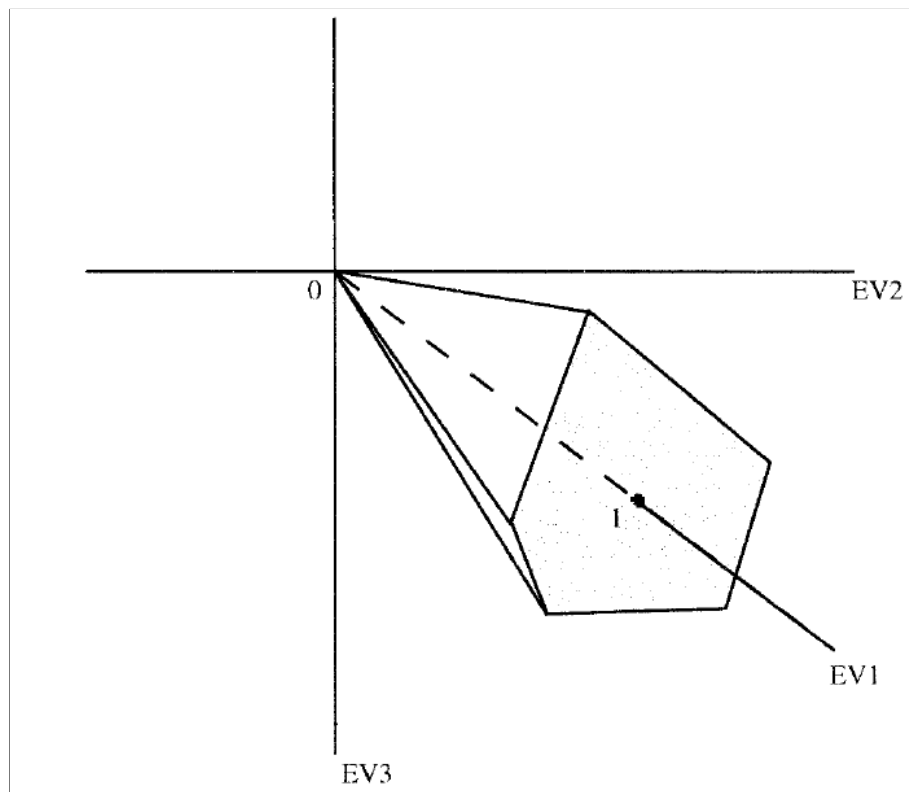
FIGURE 2.7: Schematic representation of the convex cone analysis algorithm [78]

2.3.7 Independent Component Analysis (ICA)

This algorithm is an unsupervised source separation technique used for blind source separation, feature extraction and unsupervised recognition with high accuracy [52]. This algorithm looks for components that are statistically independent rather than uncorrelated. The algorithm requires statistics of orders higher than second order and then a linear model is used to unmix the component of a set of unknown signal sources randomly to produce classification or signal detection [50]. The advantage of this algorithm is its ability to detect endmember spectra from dataset that contains no pure pixels.

2.3.8 Dependent Component Analysis (DECA)

This algorithm unmixes the spectra of a hyperspectral dataset by identifying the endmembers of material present and estimate the abundance fractions of each pixel in the scene [79]. It assumes that each pixel in an image has a linear mixture of the endmember signatures weighted by the corresponding abundance fractions, these abundances are highlighted as Dirichlet densities mixtures, enforcing the non-negativity and sum to one constraints on the abundance fractions. DECA does not require the presence of pure pixels in a scene [50].

2.4 Nonlinear Unmixing Methods

Nonlinear unmixing models cope with nonlinear interaction effects as captured by a sensor which are often present in an image [11]. Due to the simplicity of the linear mixing model, it has been the most popular method used for image analysis. However, it has lately been noted to output inaccurate results, in particular situations [8] because it fails to incorporate back scatter radiation, in this case, other unmixing methods which are assumed to be

more complex are used. Nonlinear unmixing models take into account non-linear interactions; where they capture the effect of non-linearity in an image [11].

The linear spectral unmixing method generally does not perform well and provides poor accuracy when the light from the sensor is affected by multiple interactions between different materials or endmembers otherwise known as intimate interaction. [80, 81]. Here, the nonlinear unmixing method can be advantageously replaced with the linear methods [72, 82] which provide an alternative approach to SU. When interactions occur at a microscopic level, it is said that the materials are intimately mixed. The nonlinear methods account for the intimate mixture of materials as covered by a scene in a dataset [10, 83].

Different nonlinear spectral unmixing models exist; some are motivated by physical arguments such as bilinear models, whilst others exploit a more flexible nonlinear mathematical model to help improve the performance of the unmixing method [11]. Nonlinear models can be grouped into different classes:

2.4.1   Intimate mixtures

Intimate mixture occurs when interactions happen at microscopic level. This can be described as a condition where different materials in a scene are intimately mixed e.g. of this scene is an area that is composed of sand or mineral mixtures as demonstrated in Figure 2.8.
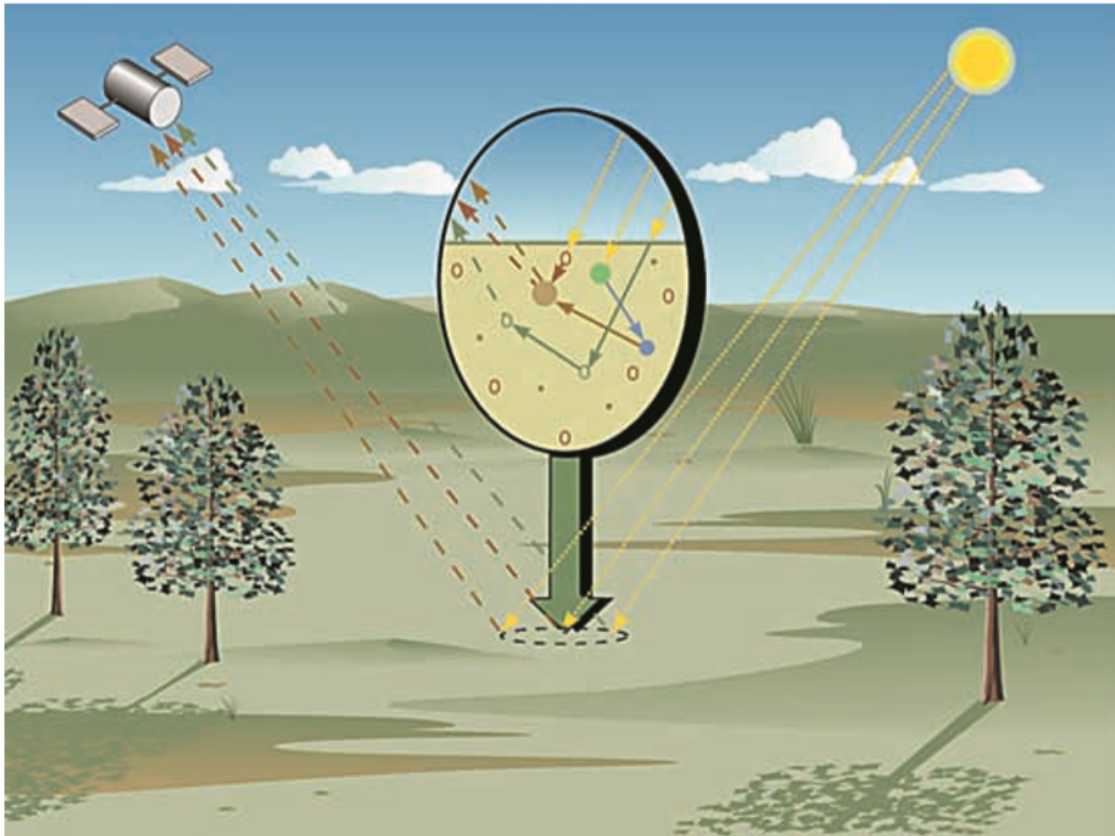
FIGURE 2.8: Nonlinear mixture with multiple reflections [10]

A model described by [46] explains the interactions that occur when light comes in contact with a surface with different materials or particles; they involve meaningful and interpretable quantities that have physical significance, however, these models cannot be solved with a linear model, they require nonlinear models which are more complex and complicates the derivation of the unmixing strategies compared to the linear models [11]. Hapke [46] derive an analytical model to express the measured reflectance as a function of parameters intrinsic to the mixtures, these include mass fraction, density size and single scattering albedo, however, the model's limitation is its dependency on the parameters implicit to the experiment because it need information regarding the geometric position of the sensor against the observed samples, this makes the inversion process more difficult especially when the endmembers are not known [10].

2.4.2 Bilinear models

Bilinear models are nonlinear models, which have gained popularity by researchers due to the model's ability to capture multiple scattering effects in datasets. Multiple scattering occurs when light scatters from a given material and reflects on another material or different surface before being captured by the sensor [47]. This often occurs when the target scene contains different materials such as vegetation cover and interaction takes place between different materials or components. This interaction has been described using different models which often use the power of product of reflectance [47]. These models account for multiple photon interaction by the addition of an interaction term to the linear model, however, these models are often applied in a scenario where the interactions of orders greater than two are neglected. Example of such models are: the Fan Model (FM), Nascimento Model (NM) and Generalized Bilinear Mixing Model (GBMM) [84]. The major difference between these models is the constraints attached to the mixing parameters.

The bilinear mixing model is defined as [11]

$$\mathbf{y}_p = \sum_{r=1}^{R} \alpha_{r,p}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,p}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p \tag{2.2}$$

Where $\mathbf{m}_i \odot \mathbf{m}_j$ is the hadamard (term by term) product of the i and j spectra. The first model describes the linear mixture model and the double sum models the nonlinear effect, the additional coefficient $\beta_{i,j,p}$ adjust the nonlinearity between the components $\mathbf{m}_i$ and $\mathbf{m}_j$ in the pth pixel.

Nascimento bilinear mixing model

The Nascimento bilinear model considers second orders interactions between the ith and the jth endmembers (for $i, j = 1, R$ and $i \neq J$) which is defined as [80]

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p} \mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j,p} \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p \tag{2.3}$$

The parameters $\beta_{i,j,p}$ is the amplitude of the interactions term due to i^{th} and j^{th} component, these parameters have to satisfy the positivity and sum to one constraint [80].

Fan bilinear model

This model, just like the Nascimento model, assumes the interaction term $\mathbf{m}_i \odot \mathbf{m}_j$ are additional spectra, however, this model also assumes that the amplitude of this interaction depends on the component fractions involved in the mixture. The model assumes that $\beta_{i,j,p} = a_{i,p} a_{j,p}$ which is defined as [85]

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p} \mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_{i,p} a_{j,p} \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n} \tag{2.4}$$

Generalized bilinear mixing model

The Generalized bilinear model is based on the assumption that the contribution of the interaction term $\mathbf{m}_i \odot \mathbf{m}_j$ is proportional to the fractions of the components involved with an amplitude $\gamma_{i,j} a_i a_j$ where $\in (0,1)$ which produce the following equation [47]

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p} \mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j,p} a_{i,p} a_{j,p} \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p \tag{2.5}$$

Where $\gamma = [\gamma_{1,2} ..., \gamma_{R-1}, R]$ is a real vector parameter which quantifies the interaction between different spectral components. The parameter introduced makes the model to be more flexible compared to the Fan model and the Nascimento model [86]. This model also adopts the positivity and sum to one constraints.

Nonlinear effect in a dataset have been tackled using different methods. Another way of simplifying a simulation that has no known mixing method and parameters is through the use of a training based approach [72]. Guilfoyle *et al* [87] uses the radial basis function neural network for unmixing analysis, whilst Plaza *et al* [88] discuss the methods for automatic selection and labeling of training samples. In Halimi *et al* [86], a nonlinear unmixing algorithm for the general bilinear mixture model based on Bayesian inferences was used, however the method's limitation is high computational cost and is solely applicable to bilinear models.

2.4.3   Physics based nonlinear mixing model

The physics based model is often applied to unmix macroscopic and microscopic mixtures by introducing a dual model comprising of two terms [11].

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + aR + 1_{,p}R\left(\sum_{r=1}^{R} f_{r,p}\mathbf{w}_r\right) + \mathbf{n}_p \tag{2.6}$$

Where the first term on the right hand side describes a linear mixing model which comes from the macroscopic mixing process, and the second one, considered as an additional endmember with abundance $aR + 1_{,p}$ describes the intimate mixtures by the average single albedo which is expressed in the reflective domain by mapping $R(\cdot)$ [47].

2.4.4   Detecting nonlinear mixtures

Nonlinear mixture models can be used to unmix nonlinear effect in a hyperspectral dataset with good and more accurate results, its major limitation is high computational complexity compared to the linear mixture model [11]. This can be minimized by applying the linear unmixing models to detect linearly mixed pixels, whilst the nonlinear unmixing models can be used to detect nonlinear mixed pixels [11]. Different ways of achieving this is by

applying techniques such as the Polynomial Post Nonlinear Mixing model (PPNMM) and robust model free detectors.

Polynomial Post Nonlinear Mixture (PPNM)

The PPNMM model is based on the assumption that the reflectance occurring in an image are nonlinear functions of pure spectral components, corrupted with additive noise; these nonlinear functions are mostly approximated using polynomial function which in turns leads to a polynomial post nonlinear mixing model [16].

The model involves two terms: the linear and quadratic functions of the abundances. where the R-spectrum $y_p = [y_1, ..., y_R]^T$ of a mixed pixel is described as a nonlinear transformation g of a linear mixture of R spectra m_r and corrupted with additive noise n.

$$\mathbf{y}_p = g_p \left(\sum_{r=1}^{R} a_{r,p} \mathbf{m}_r \right) + \mathbf{n}_p \qquad (2.7)$$

where m_r is the spectrum of the rth material in the scene, a_r its corresponding proportion and g is an appropriate nonlinear function. PPNMM model is also motivated by the Weierstrass approximation theorem which describes every continuous function defined on an interval as said to be uniformly approximated by a polynomial with any desired precision [16].

Robust model-free detector

This model is applied to a situation where the given mixture type in an image does not fit with any unmixing model [11]; this scenario occurs as a result of insufficient information regarding the choice of appropriate unmixing method, this model is used to investigate if a given pixel is linearly or intimately mixed [11].

A comparison of the most commonly used methods are described in Table 2.1. These methods are the state-of-the earth methods extensively used as reference in the literature.

TABLE 2.1: Comparison of the reviewed spectral unmixing methods

Methods	Advantages	Disadvantages	Computational cost
VCA [60]	The algorithm is unsupervised. It assumes that endmembers are the vertices of a simplex and affine transformation of a simplex is also a simplex. Works well with both projected and un-projected data	The algorithm assumes the presence of pure pixels in a dataset which does not always holds.	Relatively low.
N-FINDR [75]	The algorithm works directly in spectral space and its independent of spectral dimension. Convergence property of the algorithm is simplex with the maximum volume found.	Suffers several issues in its practical application e.g search for region which is usually entire space.	Excessive computation.
FCLS [77]	Simultaneously implement the Abundance Non-negative Constraint (ANC) and Abundance Sum-to -one Constraint (ASC) on a linear mixture model. High and promising accuracy	Requires a complete prior knowledge of material signatures present in an image scene	Relatively low.
PPI [73]	The algorithm is supervised and partially interactive	High number of iterations required	High computation.
PPNMM [16]	Used to model different nonlinearities between endmembers and observations.	Considers only scattering effects in a given pixel. Do not consider spatial interaction from materials present in the neighboring pixel.	Low computation.
FAN Bilnear model [85]	Assumes that amplitude of interactions depends on the product of fractional abundance of endmembers present in a pixel. Adopts the sum to one constraint.	Does not generalize the LMM which can be a restrictive property of the model.	Moderate to high
GBM [47]	Accounts for second order interaction. Accounts for the presence of multiple photon by addition of a bilinear term to the LMM	Has the same constraint as the LMM ANC and ASC	High computation
Nascimento Bilinear model [80]	The algorithm ca be considered a linear model with additional virtual endmembers. Abundance vector can be estimated using the linear spectral unmixing algorithm	Has the same constrain as the LMM (ANC and ASC)	High computation

2.4.5   Artificial neural networks based methods

Artificial Neural Networks (ANN) are another nonlinear model developed to initially copy the complex pattern of neuron interconnections in the human brain [89, 90]. ANN often extracts features in a shallow manner, and do not extract deep features. Similar to the human brain, ANNs comprise of 3 layers, namely; the input, hidden and output layers which provide the appropriate output based on the input data. ANNs often require training of the input data in order to influence decision making of the ANN, performed in iterations known as epochs. The training is undertaken in the hidden layer where the activation function and the number of hidden nodes is defined. ANN with a single hidden node and nonlinear activation functions can approximate any function provided that it has sufficient training data. This makes ANNs superior universal approximators with accurate and acceptable results [91].

Presently, a lot of neural networks models have been extensively studied, ANN have been successfully applied for many years with excellent performance in different field of application such as fault detection and diagnosis of mechanical systems [92], pattern recognition [93], and in particular for spectral data [94, 95]. Self Organizing Map (SOM) unsupervised neural network algorithms is one of the most widely used models successfully applied to hyperspectral image classification [67, 96, 97] and data visualization [98]. Alternative approaches include rule base fuzzy logic [99, 92, 100] and Markovian jump systems [101, 102] which could be combined with ANN.

Neural network is often used to classify input data into a set of target classes obtained by training a network to evaluate its performance. The application of neural networks to spectrally unmix a hyperspectral data has been demonstrated due to their ability to recognize complex patterns in high

dimensional images [103]. Neural network based unmixing of hyperspectral imagery has proven to be a powerful tool which produced good and acceptable results [104]. Lyu *et al* [105] have demonstrated the use of neural networks to unmix a hyperspectral data using both linear and nonlinear methods simultaneously. In [106], the use of ANN was reported to detect and count cars in Unmanned Areal Vehicle (UAV) images. Wu *et al* [107] used neural networks for hyperspectral data classification, they used recurrent neural network to model the dependencies between different bands and also to learn more discriminative features for classification. Wu and Prasad [104] used a linear mixture model to unmix hyperspectral data and ANN to predict a fraction of nonlinear mixture in the dataset; ground truth data was used to extract the abundance as estimated by the linear model which was used to train and validate the network. Licciardi *et al* [103] used neural networks for unmixing a hyperspectral data to extract the endmembers and estimate their abundances. Atkinson *et al* [108] used neural networks to unmix a hyperspectral data, the results obtained was compared with a linear unmixing model using a fuzzy c-mean classifier; results showed the capability of the neural networks against the linear unmixing method.

Feed-forward neural networks

Feed-forward neural networks are the simplest and most widely used architecture. A feed-forward neural networks consist of one input layer, several middle layers (hidden) and one output layer. With feed-forward neural networks, neurons have forward connections to neurons in subsequent layers i.e the neurons of each layer only accept output information coming from the neuron of the forward layer [109]. This architecture has demonstrated to produce excellent results in different application such as image processing, pattern recognition and classification etc. Figure 2.9 shows a typical feed forward neural network with a single hidden layer. Antoniades [110] noted that

these networks require specific training before they can be applied to real world problems.
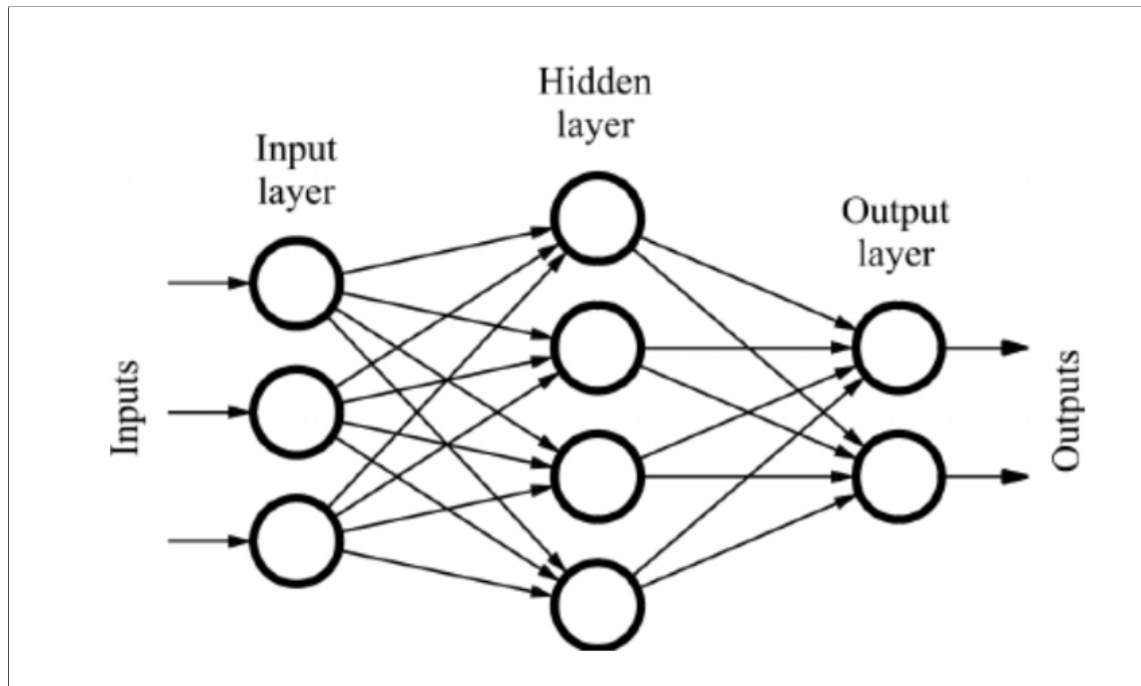
FIGURE 2.9: A feed-forward neural network with a single hidden layer [111]

Supervised learning

Neural networks can be trained using supervised (with human help) or unsupervised (self organized) algorithms. Supervised learning is the most common form of machine learning. Supervised algorithms learn with the help of teaching signals, while unsupervised algorithms take a more independent approach. The supervised algorithm first initializes the weights with adequate values, then takes the input from the training data into the neural networks. The correct output is obtained and the error is calculated from the output, finally the machine then modifies its internal adjustable parameters (weights) to reduce the error [112]. To properly adjust the weight, the learning algorithm computes a gradient vector for each weight where a change in error either increases or decreases, the weight vector is adjusted in the opposite direction to the gradient vector. The Stochastic Gradient Descent (SGD)

is the most commonly used procedure in practice, which consist of the input vector, then compute the outputs error which then calculates the average gradient and adjust the weights accordingly. The procedure is repeated on many small sets of examples from the training data until the average of the objective function stop decreasing. After training the datasets, the system's performance is evaluated on another set of examples unseen by the system during training called the test set. This is often done to test the generalization ability of the machine [107].

2.4.6   Deep neural networks

Deep neural networks often referred to as deep learning was introduced about a decade ago, when results from the approach outperform other machine learning and neural networks techniques with impressive performance. Deep learning can be categorized as a sub-field of machine learning, which often learns high- level abstractions in data by utilizing hierarchical architectures [113]. Deep learning can also be described as the final product of machine learning where the learning rule is considered as the algorithm which generates the model from the training data. Typically, it involves modeling, which hierarchically learn features of input data using ANN and have more than three layers [114]. The main advantage of deep learning is that, these layers of features are not designed by human engineers; they are learned from the input data using some learning procedures [115]. Deep learning model often contains an input layer, two or more hidden layers and an output layer as described in Figure 2.10.
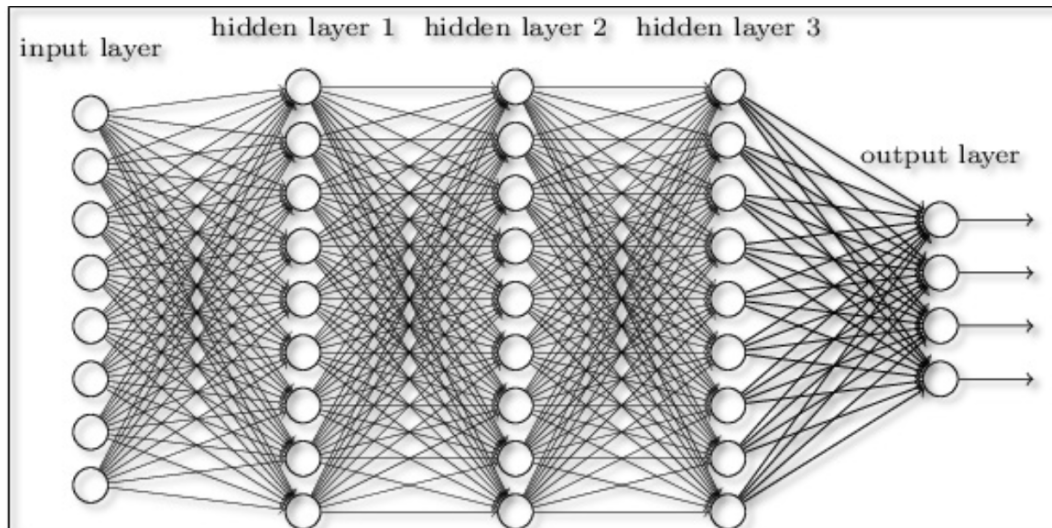
FIGURE 2.10: Deep Neural Network Architecture [115].

Deep learning has been demonstrated to perform well in different applications such as vehicle detection [116, 106], applied deep learning approach to investigate avalanche search and rescue operations using Unmanned Areal Vehicles (UAV) and change detection [117, 105]. There are many types of deep learning architectures whose application have been proven to yield excellent results, the most common are; Deep Believe Network (DBN), Convolutional Neural Network (CNN), Deep Convolutional Generative Adversarial Networks (DCGAN), Recurrent Neural Networks (RNN), Autoencoders etc [118, 119]. Application of deep learning techniques to hyperspectral data is relatively recent, for instance in [120] deep belief networks and a novel texture enhancement algorithm were investigated for their suitability and practical application to hyperspectral image classification. In Yongyang *et al* [121], high resolution remote sensing imagery and deep learning techniques were used to extract buildings in urban districts using guided filters. In Shaohui *et al* [122], a 3D full convolutional neural network model was used for spatial spectral resolution of hyperspectral images by learning end-to-end, with full mapping between low and high spatial resolution hyperspectral images with high accuracy. Transfer learning with a deep convolutional neural network was reported in [123]; in this research, a large amount of unlabeled SAR

scene data was transfered to SAR target recognition tasks and feedback the construction loss to the classification pathway. Others such as [107, 124] used a deep learning approach to classify hyperspectral images.

Here, some of the influencing deep neural network techniques will be introduced.

Convolutional neural networks

A convolutional neural networks (CNN) is a multi-layer neural networks inspired by the organization of neurons in the animal vision contex [125]. CNNs are feedforward networks i.e information flows take place in one direction only, from its inputs to its outputs. The architecture of CNN comes in different variations, however, the most common ones consist of convolutional layers, pooling layers and a fully connected layer(s) as demonstrated in Figure 2.11.
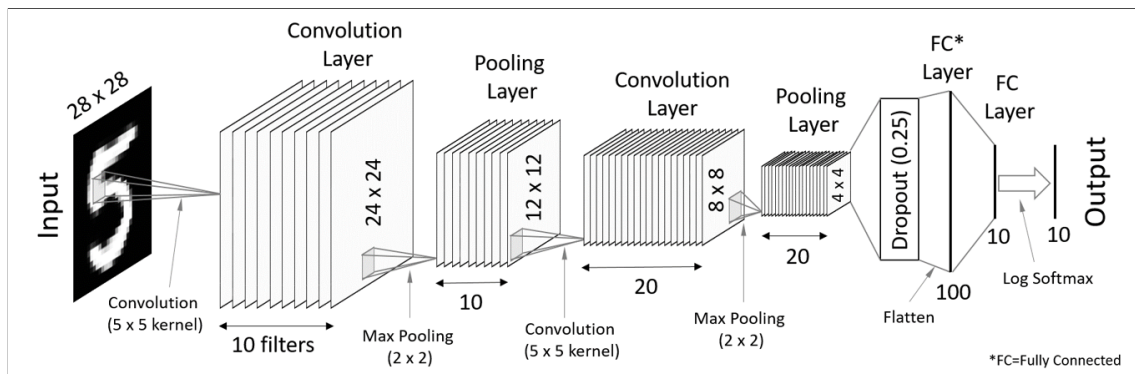
FIGURE 2.11: A typical convolutional neural network structure
[125]

The convolutional layers serve as feature extractors where the feature representation of the input image is learned. The neurons in the convolutional layers are arranged into feature maps with each neuron in the feature map connected to a receptive field which is also connected to a neighbouring neuron in the previous layer through a set of trainable weights [115]. The inputs

convolves with the learned weight and the results are sent through a nonlinear activation function.

Nonlinear activation functions allows the extraction of nonlinear features in the input data, traditionally, the sigmoid and the hyperbolic tangent functions were the most commonly used but recently, the Rectified Linear Units (ReLU) have gained popularity due to their success in various application [126]. The pooling layers reduce the spatial resolution of the feature maps so as to achieve spatial invariance to input distortion and translations. The fully connected layer interpret the feature representation and perform high level reasoning. For classification problems, the softmax operator is commonly used [127], this method was outsmart by Radial Basis Function (RBF) [115] recently, [128] recorded great improvement by replacing a softmax operator with Support Vector Machine (SVM) for classification of a hyperspectral dataset. However, these operators are dependent on the type and size of the input data.

Recurrent Neural Networks

Recurrent Neural Networks (RNN) are a family of deep neural networks designed to recognize patterns in sequences of data such as audiowaves, handwritting, time series data etc. This algorithm takes time and sequence into account and have temporal dimensions [129].

RNNs are often used to model a data where the input and output classes are of variable length. Figure 2.12 shows a typical architecture of RNN with a connection loop which allows the network to be unfolded to handle the variable length issue associated with the dataset.
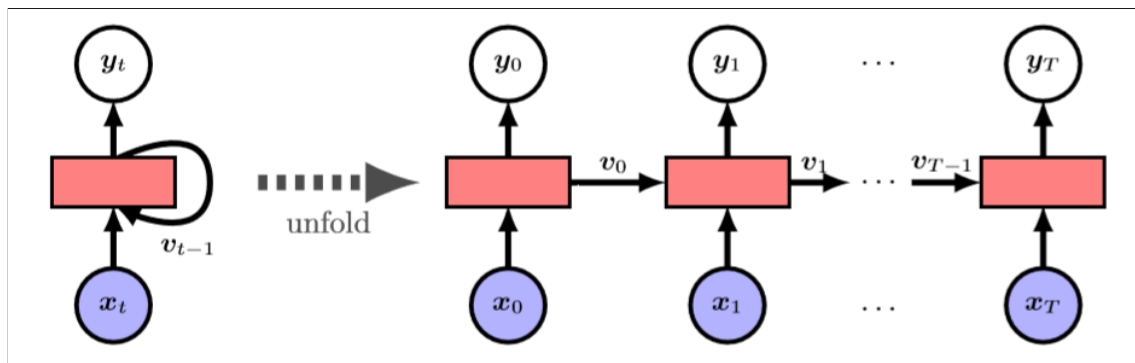
FIGURE 2.12: A Recurrent neural network structure showing
the loop design which allows the network to be unfolded [130]

RNNs can also be applied on images, which can be decomposed into se-
ries of patches and treated as a sequence. Its main drawback is the flexibility
of the data, which means that a fixed- length input layer cannot be utilized
directly, this problem was solved by modelling the RNNs with an internal
memory mechanism which process sequences of variable lengths.

A major challenge in training RNNs is long term dependencies which
causes the gradient to either vanish or explode [131] making the gradient
based optimization method struggle due to the effect of long term dependen-
cies. The Long Short-Term Memory Units (LSTM) [132] and Gated Recurrent
Units (GRU) [133] were developed to handle this problem.

RNNs employing either of this units have shown to perform well in tasks
with long term dependencies as demonstrated in [134, 133]. LSTM has a
memory cell which maintains history information through time, it has gates
specially designed to pass information inward and outward of the cell. The
input of the cell is narrowed by the input gate, while the target cells controls
what history should be retrained. GRU is basically an extension of LSTM but
without an output gate, which in turns writes the contents of its memory cell
into the larger net at each time step [133].

Autoencoders

An Autoencoder (AE) is a type of deep neural networks where training is performed layer by layer in a supervised manner such that each hidden layer maps the input pattern to themselves. Instead of training the network to predict some target values given input. AE is trained to reconstruct its own input, therefore, the output vector have the same dimensionality as the input vector [113]. During the learning process, the AE is optimized by minimizing the reconstruction error thereafter the corresponding code is the learned feature [135]. One such AE which is still extensively studied is the deep autoencoder which was first proposed by [136]. This type of AE is often trained with a variant of back propagation which sometimes becomes ineffective in the presence of errors in the first few layers. This causes the network to learn and reconstruct the average of the training data [113]. This issue can be addressed by pre- training the network with initial weights that approximate the final solutions [135].

2.4.7   Activation functions

The activation functions used in deep neural networks often act like a switch which can either be turned on or off (activated or deactivated). The activation function is controlled by the input signal which is usually modelled as a continuous nonlinear function to induce distributed representations to be tuned. This nonlinear function allows the model to overcome some truvial degradation. The activation functions are also known as transfer functions which are used to determine the output of a neural network and map the resulting values within the range (0 to 1) or (-1 to 1) etc. There are basically two types of activation functions; the linear and nonlinear activation functions. When the linear activation functions are applied, the output of the

function does not usually confide within any range while the nonlinear activation functions make the model to generalize with a variety of data and differentiate between outputs as shown in Figure 2.13.
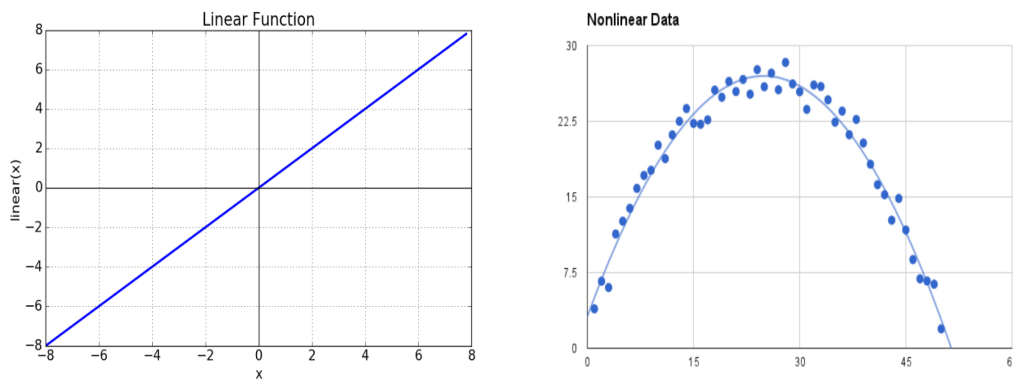
FIGURE 2.13: Linear and non-linear activation function

The linear activation function makes the model to consist of multiple affine transformations which makes the model degrades and become linear, this often happens when a large number of parameters fails to increase the model's capacity. The nonlinear function has the ability to trigger non-trivial model and continuously differentiate direct integration into gradient based optimization algorithm. The most commonly used activation functions used for deep learning applications are sigmoid, hyperbolic tangent, rectified linear unit, softmax, maxout, hermite polynomial and radial basis. The next section briefly introduces some of the commonly used functions.

Sigmoid

The sigmoid activation function also known as logic activation function is a nonlinear function applied to feedforward neural networks [17]. The sigmoid function is commonly used in models where the output predicts a probability, this has been successfully applied in binary classification problem, logistic regression tasks as well as other neural network applications [137]. The sigmoid function has an "S"- shaped curve which ranges between (0 to 1) defined as;

$$f(x) = \left(\frac{1}{(1 + exp^{-x})} \right) \tag{2.8}$$

However, its major limitation include sharp damp gradient during back-propagation from deeper hidden layers to the input layer, gradient saturation, slow convergence and also non zero centred output which as a result makes the gradient update propagates into different directions [17].

Hyperbolic tangent (Tanh)

The hyperbolic tangent or (Tanh) activation function is another type of non-linear function used in deep learning application, it is a smooth, non zero centred function and it ranges between (-1 to 1). The tanh activation function maps negative input values as negative and the zero input values mapped near zero which aids the backpropagation algorithm. The tanh activation function also has an "S"- shape curve similar to the sigmoid function as shown in Figure 2.14 and is defined as;

$$f(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} \right) \tag{2.9}$$

The tanh activation function is mostly preferred to the sigmoid function because it trains neural networks with multiple layers better compared to other activation functions.
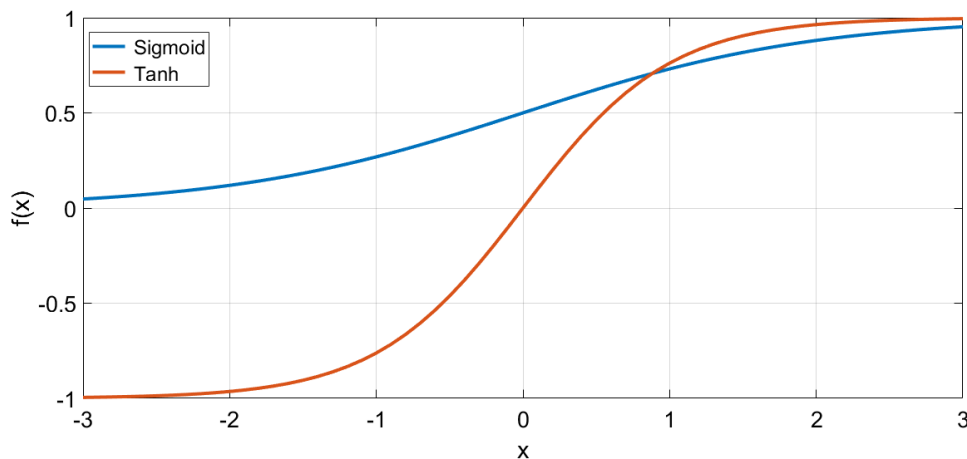
FIGURE 2.14: Sigmoid and Tanh activation functions

The tanh function has been successfully applied in deep neural networks such as recurrent neural networks for speech recognition [138] and language processing [139].

Softmax

The softmax activation function is commonly used for multi class classification task where the target class has the highest probability, it is used to compute probability distribution from a vector of a real number defined as

$$f(x) = \frac{exp\left(x_i\right)}{\sum_j exp\left(x_j\right)} \tag{2.10}$$

The softmax activation function produces an output value ranging between (0 and 1) with the sum of the probability equal to 1. The main difference between the softmax and the sigmoid activation functions is their application. Whilst the sigmoid is applicable to binary classification the softmax is commonly used for multivariate classification [17].

Rectified Linear Units (ReLU)

The Rectified Linear Units (ReLU) activation function also known as the ramp function performs a threshold operation on each input element where

it converts values less than zero to zeros [140], ReLU is defined as;

$$f(x) = max(0, x) \tag{2.11}$$

It has been the most widely used activation function in almost all CNN and deep learning application with outstanding results, it provides good performance with better generalization compared to other activation functions such as sigmoid and tanh activation functions [141]. ReLU ranges from (0 to infinity), The main advantage of ReLU is its ability to train faster mainly due to the fact that it does not compute exponentials and divisions.

Maxout

The maxout activation function proposed by [142] can be categorized as a generalized version of ReLU, here the neurons inherits the properties of ReLU with no saturated neurons in the network's computation. The maxout function is defined as;

$$f(x) = max(w_1^T x + b_1, w_2^T x + b_2) \tag{2.12}$$

where w = weights, b = biases and T = transpose

The maxout activation function has been successfully applied in a study conducted by [143] for phone recognition application. The major draw back of the maxout function is that it doubles the number of parameters used in all neurons there by making it computational expensive.

2.4.8   Learning and parameter optimization

To train a network, a training criterion needs to be defined, this criterion yields a scalar value that measures the performance of a given model, the

mapping of feature vector to the class for the training data. The definition of the criterion depends on the task and the data used.

Cross entropy function is a training criterion used in classification, they are generally known for there good performance except for inevitable cases such as regression [112]. Having defined the training criterion, the next step is to define methods to find the parameters that minimizes the criterion. Deep learning models are often trained with the gradient descent algorithm. This is a simple iterative algorithm for finding the minimum of a function, its main function is to update rules in one iteration. In practice, the Stochastic Gradient Descent (SGD) is mostly used especially with a large training sample. With the SGD, the objective is to evaluate a small training sample set at each step. It calculates error for each training set and immediately adjust the weights [112].

2.4.9 Hyperparameters

A deep learning model requires the modification and fine tuning various hyperparameters in order to provide good results, these largely depend on the application, type of dataset and other hyperparameters. The initial learning rate α is one of the most important hyperparameters; in a standard backpropagation algorithm, α is an important parameter for training the network where too low a learning rate makes the network learn very slowly, and too high learning rate make the weights and cost function diverge slowly.

2.4.10 Backpropagation algorithm

Backpropagation is carried out to train multi-layer architectures to minimize a cost function of the model. It is also used to adjust the free parameters (weights and biases) in order to attain the desired network output. Backpropagation is an expression for the partial derivative of the cost function with

respect to any weight of the network. This expression shows how the cost changes when the weights and biases are changed [112]. Training a network is mostly affected by overfitting, which is poor performance on the test set after the training, which in turns affects the models ability to generalize on unseen data. This is a common and major challenge for deep learning approach. All the parameters are trained with the standard gradient decent algorithm. Misclassifications are obtained from the feedforward process, which are then utilized during backpropagation algorithm to calculate the gradient. These gradients are then used to update the training parameters.

2.4.11   Dropout

Dropout is a powerful technique used for improving the generalization error of large neural networks. It simply refers to randomly dropping out units in a neural network. In other words, it means removing it from the network together with its incoming and outgoing connections as shown in Figure 2.15. By applying dropout to a deep neural networks, it often results in sampling a thinned network from the model. This thinned network consist of all the units that were not dropped out of the network [19].

(a) Standard Neural Net   (b) After applying dropout.
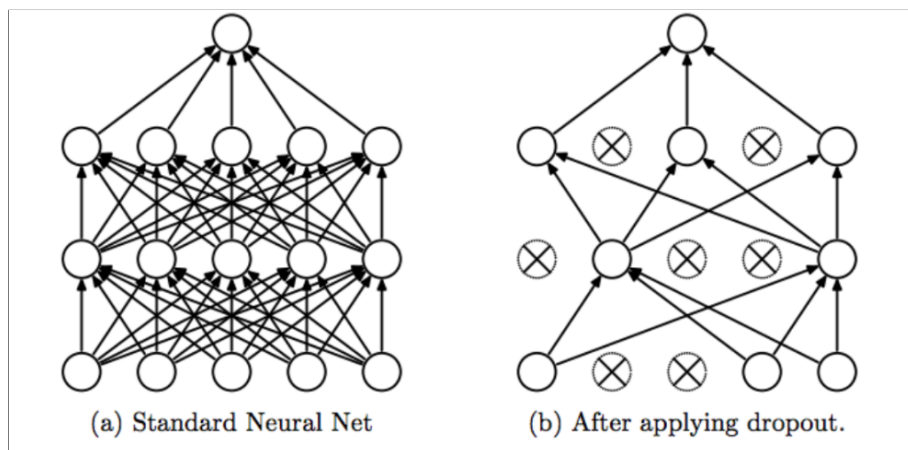
FIGURE 2.15:   A typical network before and after applying dropout [19]

Deep neural network consist of multiple nonlinear hidden layers, this makes them to be very complicated models and as such results in overfitting of the dataset.

The application of deep learning and hyperspectral data for different image processing applications has been demonstrated in the literature with excellent accuracy where the majority of the application is based on hyperspectral image classification [114, 144, 120]. Despite the success recorded in combining the deep learning approach to hyperspectral dataset, the technique has not been used for unmixing and estimation of hydrocarbon.

2.5 Summary

This chapter has presented background theory to the motivation for this research, an overview of spectral unmixing approaches, explaining the different types of spectral unmixing methods which are the linear and nonlinear methods where some of the commonly used algorithms are explained, artificial neural networks nonlinear model was also explained, the architecture of a feed forward network often used in ANNs with a single hidden layer. A supervised learning algorithm for training was also explained. The fundamentals of the deep learning approach are reviewed, the three typical forms of deep neural networks are presented: Convolutional Neural Networks (CNN), Autoencoders (AE) and Recurrent Neural Networks (RNN), explaining some activation functions and training parameters. In order to apply this method for HC spill detection, existing methods were discussed, explaining the different satellite sensors used for HC spills detection. The application of this background knowledge to the development of the hybrid switch method between linear and nonlinear spectral unmixing and deep learning based spectral unmixing methods are covered in the following chapter.

Chapter 3

Methodology

This chapter discusses the proposed methodology of this research. On one hand, a hybrid spectral unmixing method for switching between linear and nonlinear spectral unmixing was proposed. This method deals with solving the problem of weather a mixture is better explained with a linear or nonlinear model. On the other hand, a deep learning based spectral unmixing method which accurately unmixes nonlinear mixtures was developed. The two proposed methods were used for HC spill detection and estimation.

This chapter is organized as follows. Section 3.1 introduces the Spectral Unmixing Method. In 3.2 the assumptions of the linear model are discussed, Section 3.3 describes the vicinity parameters related to the characteristics of neighboring pixels used in the proposed method. Neural network architecture and other parameters employed in the research are discussed in Section 3.4. The hybrid spectral unmixing for switching between linear and nonlinear method is presented in Section 3.5. Section 3.6 introduces the Deep learning based spectral unmixing method explaining the architecture of the model and the three-term backpropagation algorithm with dropout used in training the network. Section 3.7 concludes with a summary of the two proposed methods.

3.1 Introduction

Spectral unmixing analysis of different datasets has risen as a promising method for identifying and extracting the properties of different features such as soils, mineralogy, vegetation, oil spill etc. [10]. Spectral unmixing method requires three factors to be successful: the choice of a good model, the application of an appropriate inversion technique and a set of well calibrated reflectance dataset [145]. Different unmixing models exist in the literature, such as those based on physical assumptions [10], those that consider secondary illumination [80], radiosity based approach [146], manifolds learning methods [147], neural networks [148], nonlinearity detection methods etc [145].

This creates a question of which model is best suited for a given application, or dataset although most of the models described in the literature provide accurate inversion techniques therefore, it is difficult to answer this question [145]. Different unmixing models have been investigated in search of the most accurate unmixing algorithm in literature, for example, [145] compared four different models in order to choose the most suitable method and evaluate their ability to estimate canopy biophysical parameters. Models are often chosen based on the application and the data itself. However, there is no methodology to choose the most appropriate model for a given dataset.

This section presents a hybrid spectral unmixing method for selecting between linear and nonlinear models using hyperspectral data. As outlined in chapter 2, spectral unmixing is a source separation problem which deals with mixture analysis techniques to remote sensing data. The collected spectral signature of materials in a scene is commonly a mixture of different materials found within the spatial extent of the field of view of a sensor [10].

The linear and nonlinear spectral unmixing approaches are often used

to unmix the spectra of the mixed pixel and estimates the fractional abundances. Figure 3.1 shows a flow chart detailing the procedure of spectral unmixing.
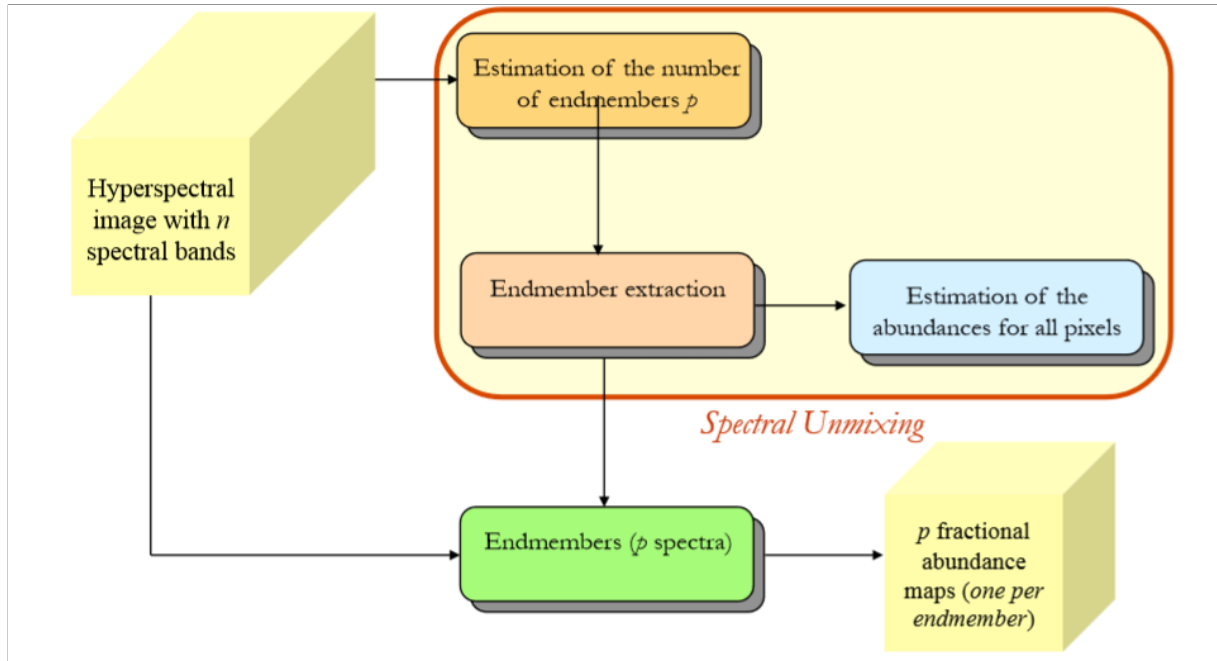
FIGURE 3.1: Schematic presentation of spectral unmixing [149].

The linear models assume that the observed spectra collected by a sensor can be expressed as a linear combination of endmembers, weighted by their corresponding abundances [150]. The main practical advantage of the linear model is due to its simplicity in implementation and flexibility of the model while the nonlinear models describe mixed spectra by assuming that part of the source radiation has multiple scattering between at least two materials before being collected by the sensor [10]. The choice between linear and nonlinear model largely depends on the nature of problem and the scale of the endmembers.

Moreover, the extraction of endmember is an important part of spectral unmixing problem, as discussed in the literature different endmember extraction techniques exist which allows validation of extracted endmembers with respect to the reference signatures using different distance metrics such

as Vertex Component Analysis (VCA), Pixel Purity Index (PPI), Minimum Volume Simplex Analysis (MVSA) [151]. These algorithms are the most commonly used methods due to their ease of computation and clear conceptual meaning with good accuracy [150], therefore VCA was chosen for endmember extraction in this study.

Spectral unmixing models are applied to estimate the fractional abundances from the extracted endmembers. The choice of a suitable unmixing model is very important and comes with implications if a wrong model is chosen. For example, if a linear model is used on a data where the abundances are nonlinear, the calculated fractional abundances will have a significant amount of error which could be as much as 30% [152]. This is mainly because of the multiple scattering of different materials collected by the sensor in a nonlinear mixture which makes them unfit with a linear model [10].

Once the endmembers and their corresponding spectral signatures are known, there is need to estimate the accuracy of the spectral mixture analysis, which is mostly based on the fit between the model and the observed mixed spectral signals such as the error metric [153], residual term [154] or Root Mean Square Error (RMSE) [155] and in some cases by checking the discrepancy between the estimated and real endmember fractions such as fractional abundances error [154] and coefficient of determination [156].

However, there is no methodology to assess the validity of a model to a given dataset, and to predict which model is more suitable to a given dataset. It may be assumed, however, that if the main assumptions for the linear model are satisfied within a pixel, most likely neighboring pixels will have a similar behavior. For instance, if multiple scattering occurs, the variability of neighboring pixels spectra will be higher. Similarly, if endmembers spectra is not constant for instance due to shadows, the resulting spectra of neighboring pixels might change dramatically. Nevertheless, it is known that scenes often contain cases of both linear and nonlinear mixing on a pixel-by-pixel

basis so, although it may be assumed that pixels within a close vicinity may exhibit similar behavior, not all pixels within an image will follow the same assumptions [157].

Here, an approach is proposed to close that gap and assess model suitability based only on the data. Moreover, the method does not assume that all data in an image might be better suited to a particular model type since different parameters such as multiple reflectance, absorption, endmember variability, etc. might be different within an image and as such results to some dependencies within immediate pixels.

3.2 Assumptions on the linear model

In this section, different assumptions on the linear model and in particular how this effect affects the resulting spectra of pixels are discussed. It shows that this problem has been neglected, especially the effects of multiple scattering and the resulting nonlinear mixing [158].

The linear spectral unmixing relies on some assumptions [159], namely:

- There is no significant multiple scattering between the different surface components.

- There are no intimate interactions between the different surface components.

- The measured radiance at a pixel is a weighted average of the radiances of the components present at the pixel.

- Endmembers are known and are invariant within the image.

3.2.1 Multiple scattering

The linear unmixing models are very popular and mostly used because of their simplicity in application with good and acceptable results in different application, especially in areas with large flat surfaces with well defined end-members [160]. However, there are some situations which makes the linear mixing model not to function well where the assumptions do not hold such as scenes with large geometrical structures, e.g. building and trees; in such scenario, shadow and mutual illumination plays a large role.

In mineral mixtures where incoming radiation interact with different mineral grains, the linear mixture model assumption of single interactions does not hold due to an intimate mixture of material components as shown in Figure 3.2.
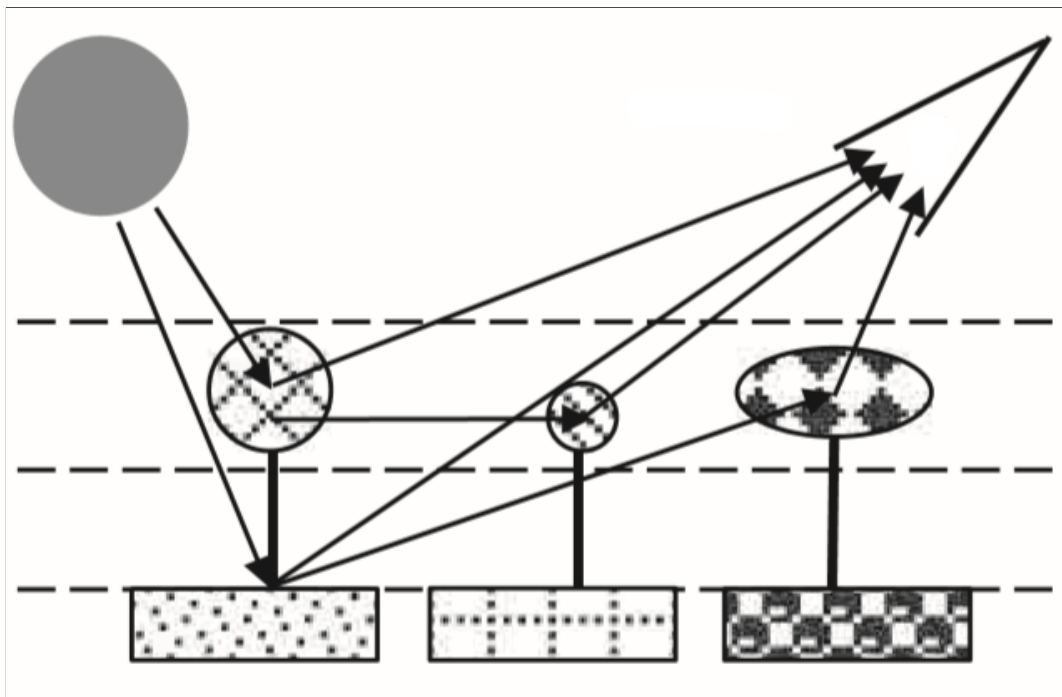
FIGURE 3.2: Multiple reflections of different materials [161]

When mixing occurs as a result of multiple reflection, it is not always possible to determine what material are present in the pixels directly from the measured spectral vector [161].

3.2.2 Intimate Mixing

Intimate mixture of materials occurs when each component in a scene is randomly distributed in a homogenous way, the incident radiation experience multiple reflections on different materials and the aggregated spectrum of reflected radiation may no longer uphold the linear proportions. Therefore, the linear mixture model is inappropriate to describe this interaction with a scenario containing multiple interaction.

Linear mixture model can be inappropriate for some hyperspectral images, for examples in cases where multiple reflections occur as a result of mixed spectra of certain endmember distributions randomly distributed through out the field of view of the sensor [162]. In that case, the collected mixed spectra can be described by assuming part of the source radiation is multiple scattered before being collected by the sensor [163] as demonstrated in Figure 3.3.
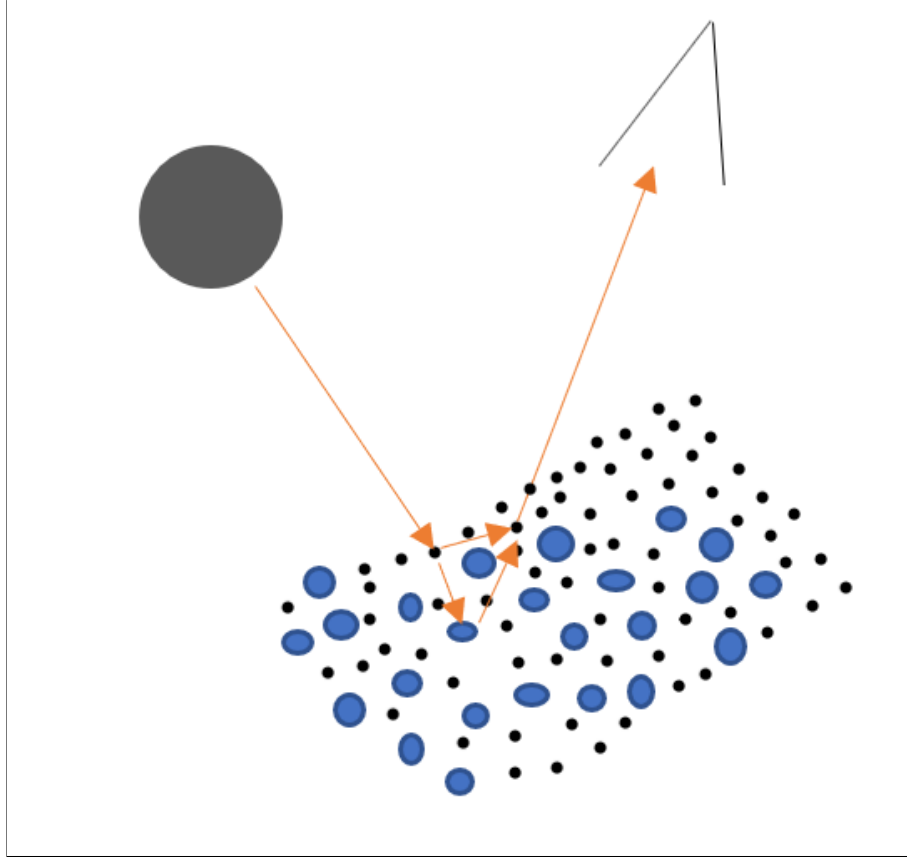
FIGURE 3.3: Schematic diagram of intimate mixtures

Intimate mixing happens, especially when dealing with granular materials such as soils. Soils are often significant constituents of spectral remote sensing scenes, thus it may be assumed that intimate mixing is very common. In such cases, a linear model is not sufficient [157]. In particular, with mixtures of soils and oils or water and oils intimate mixtures.

To cope with both intimate mixtures and multiple scattering problems, machine learning technologies has been proposed to train artificial neural networks for nonlinearity or bilinear nonlinear models using the following equation [161].

$$\mathbf{y}_p = \sum_i a_i \mathbf{m}_i + \sum_{i \neq j} a_i a_j \mathbf{m}_i \odot \mathbf{m}_j \tag{3.1}$$

Where the sum on the left hand term for the single scattering which is

similar to the linear mixing model; and the sum on the right hand side accounts for the double scattering where the vector $\mathbf{m}_i \odot \mathbf{m}_j \, y_p \in R^{nb}$ accounts for the pair wise interactions.

3.2.3 Endmember variability

One important aspect that has been understudied and subsequently raised many concerns in the literature is the effects of spectral, spatial and temporal variability, which arises due to variable illumination, atmospheric, environmental and temporal conditions in the scenes. These effects have been considered to account for the main errors in spectral unmixing. This variability introduces a modification in the shape and the scale of the endmembers spectrum in each pixel as shown in Figure 3.4. This figure shows the endmember spectra of soil and HC mixture of a normalized data.
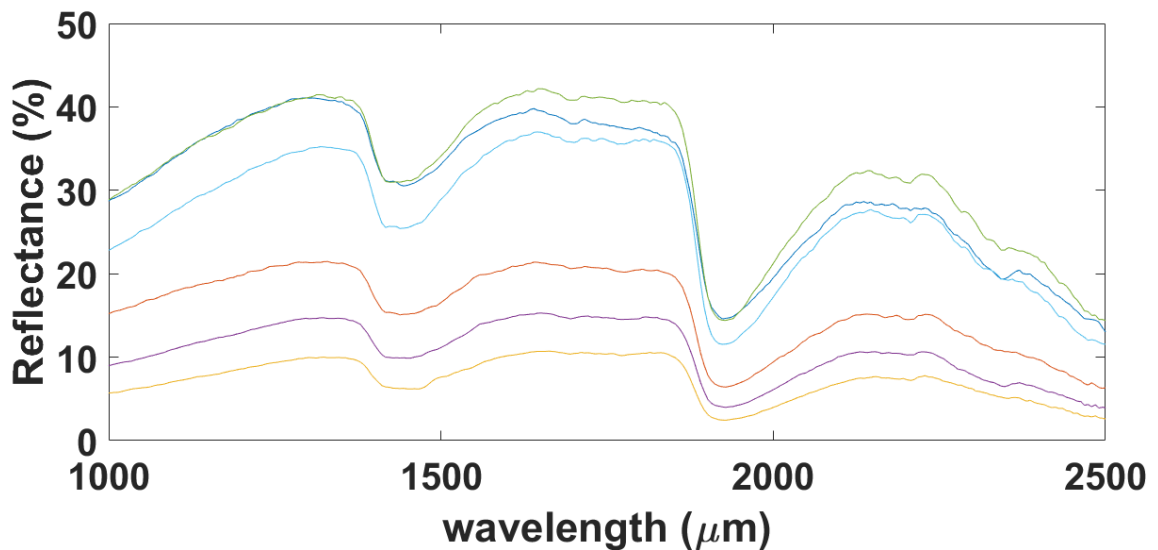
FIGURE 3.4: Effect of endmember variability

Library spectra has been used in spectral mixture analysis to address this issue. It shows variation in endmember spectral signatures, often caused by spatial and temporal variability in the scene components. While differential illumination conditions are not accounted for which results in significant

fraction estimation error [153]. Different methods have been proposed to address endmember variability in spectral mixture analysis, such as Multiple Endmember Spectral Mixture Analysis (MESMA) [155], Monte Carlo Spectral Unmixing Model (MCSUM) [164], and endmember bundle using Principle Component Analysis (PCA) [165].

However the limitation of these techniques is that endmember signatures are not treated as constants rather they are represented by probability density functions. Also the computational complexity of the methods is another drawback, especially when applied to hyperspectral data [166].

Thus, in HC spill quantification problems, the main nonlinear factors relate to multiple scattering due to the granulous terrain and interaction of organic/ non organic substances, and multiple reflections due to the nature of images where trees and different canopy might be present.

The proposed method took that into consideration and proposes the hybrid switch methods based on the characteristics of the mixture type within the neighboring pixels. Values that were believed to represent the diversity of the neighboring pixels were considered which are discussed in the next section.

3.3   Vicinity Parameters

The vicinity parameters proposed here are used to evaluate the linearity assumptions in a pixel by pixel basis. These parameters depend on characteristics of a Region of Interest (ROI) around the pixel being unmixed. The linear model is expected to provide acceptable results when the following two assumptions are satisfied: (1) mixing at macroscopic level which simply means that the probability of interacting with a given endmember is proportional to its areal abundance in the Instantaneous Field of View (IFOV) and (2) photons interact with single material before reaching the sensor. This suggest

that each incoming light ray interacts with a single material in the IFOV before reaching the sensor. Generally this is not known a priori and changes might occur over different parts on a given scene due to spatial resolution and mixing phenomena in the scene. In fact, it is more likely that the position and extent depends on the spatial, spectral and temporal complexity and also the composition of endmembers present in a given scene [153]. Therefore, it would be an interesting idea to design new model capable of simultaneously exploiting the spatial correlation and nonlinearities between abundances in a dataset to produce good and acceptable results. To address this, the hybrid switch method was proposed where a number of parameters related to the pixels' neighboring characteristics are used. These parameters are: the minimum and maximum Spectral Angular Distance (SAD), covariance and a nonlinearity parameter. In order to compute these parameters, a ROI W was defined around the examined pixels of size $n \times n$ as demonstrated in Figure 3.5.
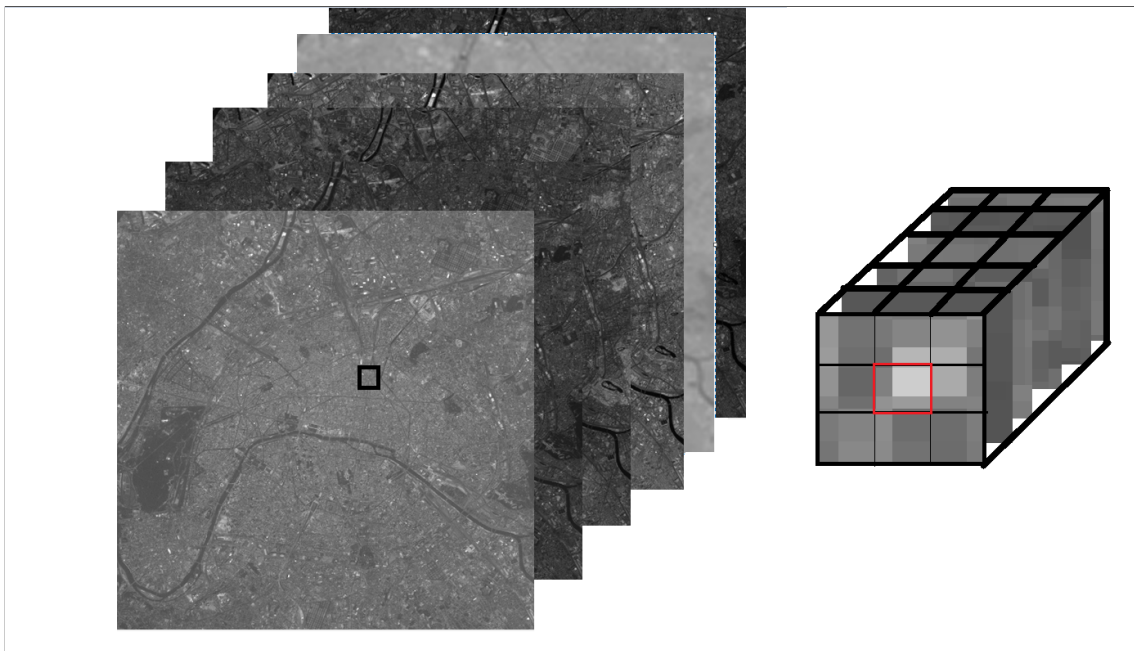
FIGURE 3.5: Hyperspectral data showing Region of Interest (ROI) and defined window size.

It can be hypothesise that neighboring pixels in a linear scene will have

more spectral and spatial coherent spectrum compared to those in a nonlinear scene as demonstrated in [167].

3.3.1 Spectral Angular Distance (SAD)

Spectral Angular Distance (SAD) describes the angular distance between two vectors. This method provides a way of measuring separability of different endmembers quantitatively and objectively and can be estimated by computing the cosine of the angles between the actual and the estimated endmembers [14].

The SAD between two spectra: $\mathbf{U} = (U_i, ..., U_L)^T$ and $\mathbf{V} = (V_i, ..., V_L)^T$ is defined as:

$$SAD(\mathbf{U}, \mathbf{V}) = cos^{-1}\left(\frac{\sum_{i=1}^{L} U_i V_i}{\|\mathbf{U}\| \quad \|\mathbf{V}\|}\right) \tag{3.2}$$

where L is the number of bands and $\|\mathbf{U}\| and \quad \|\mathbf{V}\|$ are the modulo of the two spectra vectors.

3.3.2 Covariance Matrix

The Covariance matrix is a method used to fuse multiple spectra that are correlated where the variance of each spectra are represented by the diagonal values of the covariance matrix and the non-diagonal values represent the correlation [15]. Also, the noise corrupting individual samples are often filtered out using an average filter during the computational process of the covariance. The covariance matrix is defined as:

$$C = \frac{1}{L-1} \sum_{i=1}^{L} (\mathbf{X}_i - \mu_i)(\mathbf{X}_i - \mu_i)^T \tag{3.3}$$

Where $\boldsymbol{\mu}_i$ is the mean vector of pixels in the ROI in band i and $\mathbf{X}_i$ is the vector containing all pixel values in band i and L is the number of bands

It is noted that the covariance may be calculated in the spatial or spectral domain to describe the variability of the spectral signals with respect to those in the ROI. In the proposed method, it is calculated in the spatial domain. On the other hand, it may also be used to describe the variability of the consequent spectral bands.

3.3.3 Nonlinearity parameter

From the Polynomial Post Nonlinear Mixing Model (PPNMM) described in section 2.4.4, nonlinearity can be calculated using the PPNMM model and is characterized by a parameter b for each pixel in a scene. This parameter quantifies the levels of nonlinearity in a pixel which is defined as:

$$b = \sum_{i=1}^{R} \sum_{j=i+1}^{R} \mathbf{a}_i \mathbf{a}_j \; \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n} \qquad (3.4)$$

Where $\odot$ is the Hadamard (term by term) product operation, $\mathbf{a}_i$ and $\mathbf{a}_j$ are the abundance reflectance spectra of endmembers i and j and R is the number of endmembers.

3.4 Neural Networks

The methods proposed in this work uses Neural Networks (NN) and deep learning approach. The following sections describe the neural networks employed in the proposed methods together with the parameters that are required.

3.4.1   Nonlinearity for feed-forward networks

Feed-forward neural networks, also known as multi-layer perceptrons are a basic neural network architecture which is widely used in a variety of machine learning tasks [130]. Inspired by the biological neural systems which constitute the animal brains, a feed-forward neural networks imitates signal transmissions from a collection of artificial neurons often referred to as units. These units are organized in a chain of layers referred to as hidden layers. As discussed in section 2.4.5 feed-forward neural networks consist of three layers; input, hidden and output layers.

The weighted sum of the inputs produces the activation signal which is then passed to the activation function to obtain one output from the neuron as shown in Figure 3.6.
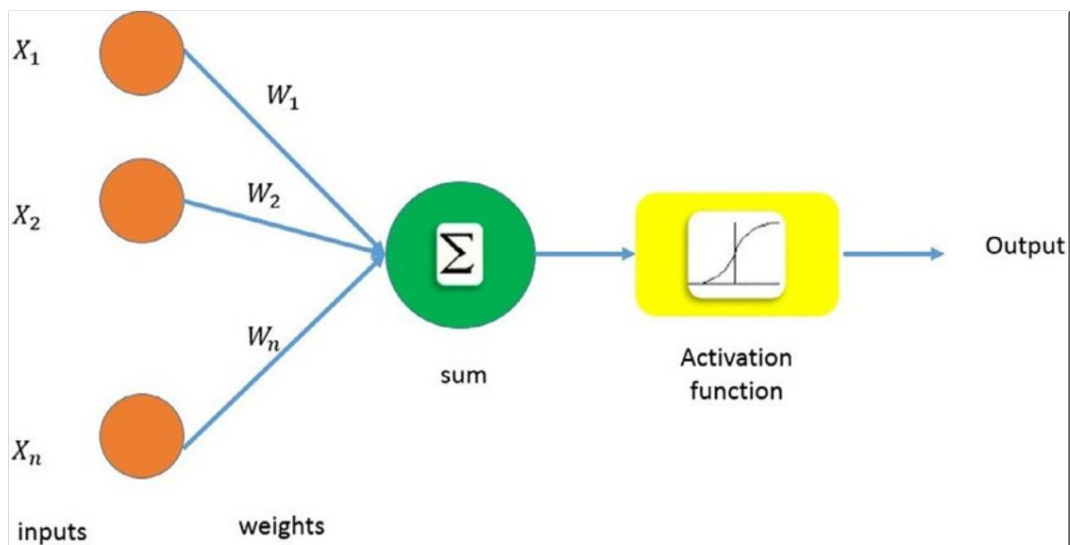
FIGURE 3.6: Schematic diagram of a feed-forward neural network [168].

3.4.2   Scale conjugate gradient backpropagation algorithm

Backpropagation algorithm is a supervised learning method used to train an artificial neural networks. Training of the networks is often carried out by

iterative updating the weights based on error signal, then the error is back propagated to the lower layers as demonstrated in Figure 3.7.
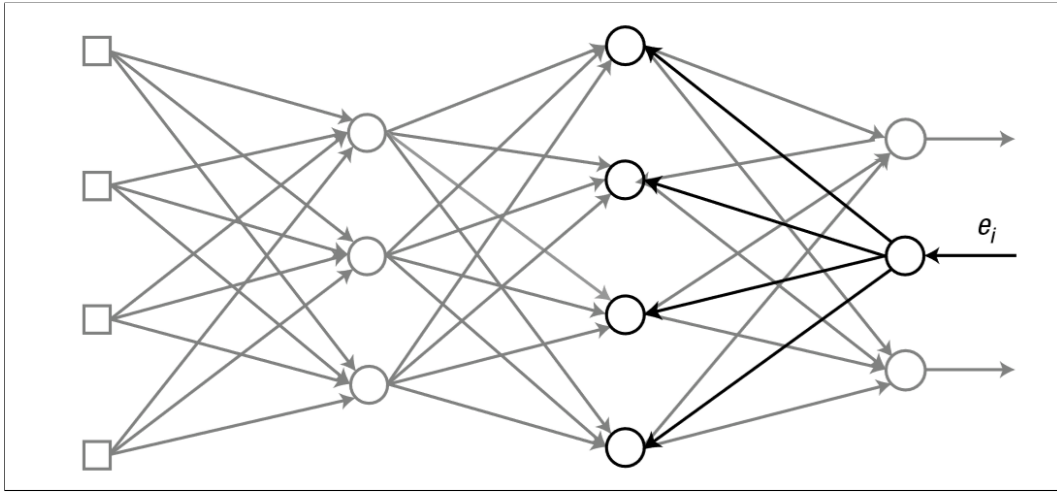
FIGURE 3.7: Schematic diagram of backpropagation algorithm [112].

Backpropagation algorithm is a descent algorithm which attempts to minimize the error rate at each iteration for a particular training pattern. Its also the famous training algorithm for multi-layer perceptions [18]. Despite the general success of the backpropagation algorithm, it still has some limitations such as the existence of temporary local minima which results from saturation behaviour of the activation function. It also has slow convergence rate which makes it relatively slow to train a networks with more than one hidden layer. The backpropagation algorithm involves two parameters in updating the weights during training which are; the learning rate (α) and momentum factor (β) which is given as:

$$\Delta W(k) = \alpha(-\nabla E(W(k))) + \beta \Delta W(k-1) \tag{3.5}$$

Where $\nabla E(W(k))$ is gradient of E at $W = W(k)$, with $K = 1, 2, 3 .... N$ being the iteration number. $\Delta W(k-1)$ is a previous weight change, α and β are the learning rate and momentum factor respectively.

The best choice of these parameters depends on the problem which often requires a trial an error process before a good choice is found [169]. A larger learning rate results in the network learning with bigger steps and converge faster, in other words, if the learning rate value is too big, oscillation and overshooting occurs while too small a learning rate makes the algorithm to converge slowly [170].

Scale Conjugate Gradient (SCG) is a variant of the conjugate gradient method. It is one of the most popular second order gradient supervised procedure [171]. This method shows super-linear convergence on most problems. SCG uses a scale size scaling mechanism to avoid time consuming line search per learning iteration, which makes the algorithm faster than other second order algorithms [172]. This algorithm simply avoids the line search per learning iteration by using a Levenberg – Marquardt approach to decrease the step size. The Levenberg-Marquardt introduces scalar factor lamda (λ) to the algorithm which regulates the indefiniteness by adjusting the λ with a constant factor for each iteration.

3.4.3 Hyperparameters

An artificial neural network requires different hyperparameters in order to improve results, these largely depend on the dataset used in training the network. The SCG backpropagation algorithm involves three parameters in updating the weights during training which are; the learning rate (α), momentum factor (β) and lamda (λ). These parameters are automatically selected to fit the network during training.

Learning rate

The initial learning rate α is one of the most important hyperparameters. The learning rate determines how much the parameters are changed in one update; too low a learning rate makes the network learn slowly, and too large a learning rate possibly leads to oscillation preventing the error to fall below certain value. Several approaches have been proposed to adaptively change the learning rate to improve the performance of the network during training. These includes; the NewBob method [173] this method adaptively determines the learning rate according to temporary system performance during training. Decay method [174] this method reduce the learning rate gradually after each training update. Another alternative method is associated with individual learning rate with each parameter adjusted according to heuristic rules [175].

Momentum

The momentum factor β is a strategy to make the learning procedure more stable and accelerate convergence in shallow regions of the error function which in practice does not always happens [176]. The momentum method recursively accumulates a decaying average of past gradient and then add it to the current update. The advantage of momentum is that the inertia on the update direction can be maintained to reduce the risk of oscillation which often results in a smoother decrease of the training criterion [130].

3.4.4 Three-term backpropagation algorithm

The backpropagation algorithm has been modified by different researchers aimed at improving efficiency and convergence rate of the algorithm. One such method is the three-term backpropagation algorithm proposed by [18]. This algorithm proposes an extra term called the Proportional Factor (PF)

to the standard backpropagation algorithm which speeds up the weight adjustment process by increasing the convergence rate and decreasing learning stalls whilst maintaining the simplicity and efficiency of the standard backpropagation algorithm [177]. This approach outperforms the standard backpropagation algorithm in terms of convergence rate. The three-term backpropagation algorithm is defined as:

$$\Delta W(k) = \alpha(-\nabla E(W(k))) + \beta \Delta W(k-1) + \gamma E(W(k)) \qquad (3.6)$$

Where γ is the propotional term and $E(W)(k))$ represents the difference between the output and the target at each iteration.

Some backpropagation algorithm modifications require complex and costly calculations which often results in offsetting the faster rate of convergence obtained with the modified algorithm. Unlike other modification methods, the three-term backpropagation algorithm maintains the simplicity of the standard backpropgation algorithm as evidence from equation 3.6. This has been demonstrated in [178] where the convergence speed, simplicity and efficiency of the algorithm were tested.

Proportional factor

The extra term introduced in the three-term backpropagation algorithm called the proportional factor (γ) speeds up the weight adjustment process by increasing the convergence rate and decreasing learning stalls whilst maintaining the simplicity and efficiency of the algorithm compared to the standard backpropagation algorithm.

3.4.5 Cross validation

To evaluate the performance of a model, it needs to be evaluated as a seperate dataset from previously seen dataset by the model. It is a common practice to split a full dataset into training and testing sets which is a way of evaluating the generalizability of a model. The most common used cross validation method is the k-fold cross validation. In this method, the dataset is seperated into k-folds where the training and the testing process is repeated k times, and one of the folds used as test set and the remaining folds used for training the model. The accuracy of the model is assesed based on the mean performance across the k-folds, which can be considered as a better estimate compared to a single training - testing split of dataset [179].

The training dataset can further be divided into training and validation sets where the validation dataset is used to assess the performance of the model before testing on unseen data by the model. This is important because models such as deep neural networks have non-trainable hyperparameters therefore, an independent validation dataset is needed to evaluate their choices [130]. Another reason for using the validation dataset to evaluate a deep neural network model is because they are trained iteratively therefore a stopping criterion is needed to avoid overfitting which can be obtained from the performance of the validation dataset after each training iteration. If the performance keeps improving on the training data but deteriorates on the validation set, it is a sign of overfitting and training should stop [180].

3.5   Hybrid Spectral Unmixing Method for Switching Between Linear and Nonlinear Models

The proposed hybrid method switches between different spectral unmixing methods using artificial neural networks to assess if a particular pixel is better explained with a given model. In order to do that, a supervised neural network was used which required ground truth datasets. Although the idea is that, the neural network should be able to generalize and assess models validity on unseen pixels provided that sufficient dataset is used to train the network.

The following unmixing models: VCA, FCLS, PPNMM and GBM which have been discussed in details in Sections 2.3 and 2.4 were used here. These methods were chosen because they are state-of-the-art methods and have been studied extensively and used as benchmark in the literature. However, the same methodology could be applied to other models. The advantage of the hybrid switch method is its ability to decide the best model to use for a particular application in a pixel-by-pixel basis. It also addresses the question of whether a mixed pixel is better explained with a linear or nonlinear process.

The two models, one linear and one nonlinear were selected to train the neural network. The following pairs were used: VCA – PPNMM, VCA – GBM, FCLS – PPNMM, and FCLS – GBM, respectively. Moreover, VCA was always used to estimate the endmembers contained in the dataset and the corresponding spectra. In order to provide a comparison, the four individual methods as well as the hybrid methods were used to unmix spectra of mixtures and estimate the fractional abundances.

The key idea here is that the neural network will be able to decide whether a pixel should be unmix with a particular given model, when it is fed with the pixel spectra or a vector containing some of the vicinity parameters described

in section 3.3.

The accuracy of the individual methods to the proposed hybrid methods for switching between linear and nonlinear spectral unmixing based on the diversity of the neighboring pixels will be compared in chapter 4. The algorithms were coded according to [60, 77, 47, 16].

3.5.1 Architecture of the neural network

An Artificial Neural Network (ANN) was used to conduct and predict the best method when switching between linear and nonlinear spectral unmixing. Neural pattern recognition was used in this study, this was chosen because it solves a pattern recognition problem and classify the inputs into a set of target categories based on the class with highest values. This method has demonstrated to output good results with high level of accuracy in image processing.

The ANN is a two layered feedforward network, with sigmoid activation function in its hidden neurons and softmax in its output neurons. ANN has been proven to classify vectors arbitrarily well, having defined enough neurons in its hidden layers and its trained with scale conjugate gradient backpropagation algorithm.

The architecture of the ANN used to conduct the switching is described; the networks have 3 layers namely: input, hidden and output layers. The input layer has 12 nodes when using a 3×3 window corresponding to the vicinity parameters as described in 3.1.1 (min. SAD, max. SAD, $c_1, c_2, c_3.....c_9$, b); the hidden layer has 10 nodes while the output layer has 1 node corresponding to either linear or nonlinear model. The output layer provides the decision between linear and non-linear unmixing models with a 0.5 threshold.

The data for the ANNs were divided into 3 categories, namely: training, validation and testing sets. For these, ground truth data was required. The data was unmixed with all individual methods and the best in terms of better estimate of abundances was labeled for each pixel. Subsequently, the data was split into 3 categories.

1. The training set is used to fit the parameters of the classifier.

2. Validation set is used to minimize over-fitting (i.e verifying the accuracy of the training data) over some untrained data by the networks.

3. Testing sets are used to test the final solution in order to confirm the actual predictive power of the network [181].

3.5.2 Training

The networks were trained with scaled conjugate gradient backpropagation algorithm. This algorithm adopts the chain rule derivative [115]. In this algorithm, the gradient of the objective with respect to the input module is computed backwards from the output module [115]. This algorithm was considered because of its performance when updating the weight and bias values when using the scaled conjugate gradient; training the network stops when number of conditions are met which includes: reaching the maximum number of epochs, maximum time exceeded, performance is minimized to the goal and the validation performance increasing more than it records [181]. Algorithm 1 explains the training procedure. The Matlab Neural Pattern Recognition app was used for the hybridization between the linear and non-linear models.

Algorithm 1: Learning method using SCG backpropagation algorithm used in training the ANN

Data: linear model, nonlinear model, α , ω, δ, λ, e

NN weights ω, are randomly initialized ω_{11} **... ... ...** ω_{ij}

for *Number of epochs (k)* **do**

    initialize the learning rates α , β, ω, λ and evaluate the conjugate direction

    for *image pixel = 1, 2 ..... n* **do**

        for *Number of hidden layers = 1, 2 ..... f* **do**

            /* Calculate the errors and the delta, δ, of the output nodes

            $e = d - y$

            $\delta = \phi'(v)\, e$

            /* propagate the network output δ backwards, and calculate the delta, δ

            $e^p = W^T\, \delta$

            $\delta^p = \phi'(v^p)\, e^p$

            /* adjust the weights according to δ

            $\Delta \omega_{ij}(k) = \alpha\, \delta_i\, x_j + \beta\, \Delta \omega_{ij}(k-1) + \lambda\, e_i\, x_j$

            $\omega_{ij} \leftarrow \omega_{ij} + \Delta \omega_{ij}$

        end

        /* Repeat the algorithm for each pixel

    end

end

/* Repeat error calculation δ updating until NN is properly trained

---

3.6   Deep Learning Based Spectral Unmixing Method

This section further investigates spectral unmixing method using deep learning approach. Deep learning has been identified as a popular feature transformation technique due to its ability to implement layer by layer supervised pre-training of neural networks to discover new features formed from the combination of the original features. The deep learning approach proposed here has better generalization ability compared to standard neural network.

The goal of this section is to develop deep learning approach with dropout to spectrally unmix and estimate the abundance in a given dataset without using endmember estimation or other unmixing methods. A motivation for using the deep learning approach is due to the success of the approach in computer vision, image processing and image classification which have seen a great increase in the last few years. A review of the literature shows that deep learning approaches have been used to some extent in hyperspectral unmixing problems [182].

3.6.1   Deep Learning

Deep learning is a kind of machine learning technique which employs a deep neural network (DNN), different activation functions and dropout. It is simply referred to as a network of sufficient complexity in order to interpret raw data without human derived explanatory variables [115, 183]. A DNN contains two or more hidden layers as described in Section 2.4.6, where the learning rule becomes the algorithm that generates the model from the training data, its main drawbacks are over-fitting and vanishing gradient [112]. The most common solutions for over-fitting and vanishing gradient involves in co-operating dropout and best suited activation function respectively. The dropout trains only some of the randomly selected nodes rather than the entire network.

Deep learning is now considered state-of-the-art algorithm which has won several contests in pattern recognition and machine learning, having achieved unprecedented accuracy on benchmark datasets [184]. The structure and operation of a deep learning model is organized into three layers, namely; input layer, hidden layer(s) and output layer as shown in Figure 3.8. The input layer contains input neurons, which send information to the hidden layers. The hidden layers send data to the output layer where every neuron has weight inputs, activation function and one output. These layers are fine-tuned using different algorithms to update the weights and biases of the network. One commonly used method is the backpropagation algorithm.
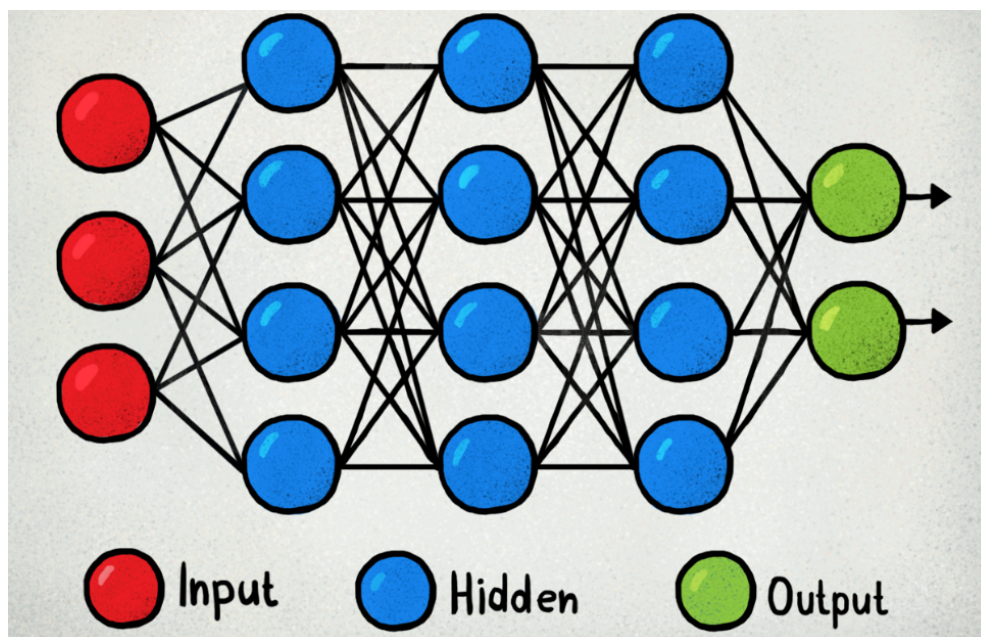
FIGURE 3.8: Structure of a deep learning model [130].

In this section, a deep learning model to estimate the amount of Hydrocarbon (HC) mixed with different soil samples was developed using a three-term backpropagation algorithm (which is a modified version of the standard backpropagation algorithm) to rapidly converge the network and dropout was used to avoid overfitting and reduce the complexity of the nets computations.

Different methods utilizing both linear and nonlinear models have been demonstrated in the literature for the analysis of different hydrocarbon types. Scafutto *et al* and Webster *et al* [29, 30] used Principal Component Analysis (PCA) and PLS regression were used. The authors used PCA to differentiate the types and density of HCs in soils while they used PLS to predict the concentration of oils and fuels in soil samples. Schwartz *et al* [7] used Spectral Angular Mapper (SAM) to classify oil spills on an image and also used signature matching to distinguish oils from other features. However, most of these methods adopt a linear model and smoothing threshold function for feature extraction. Other approaches such as a Kernel-based transformation [185] and manifold learning algorithm [5] were based on nonlinear models.

In [186], it was proven experimentally that HCs abundance in soils was estimated with higher accuracy when nonlinear unmixing models were applied. Nevertheless, spectral unmixing and specifically the abundance estimation of HCs such as gasoline can be challenging [187] and may require more advanced techniques such as deep learning. An alternative method for solving nonlinear problems deep learning network, which can be fast and accurate since the model does not rely on assumptions to simplify the model.

3.6.2 Dropout

Dropout allows neurons to randomly drop out of the network during training while other neurons steps in and handle the representation required to make predictions for the missing neurons [19]. As discussed in section 2.4.11, it refers to randomly dropping out units in a neural network as demonstrated in Figure 3.9. The dropout works in such a way that the network becomes less sensitive to the original weights of neurons. This results to having a network that is able to generalize and not likely to overfit the training data.
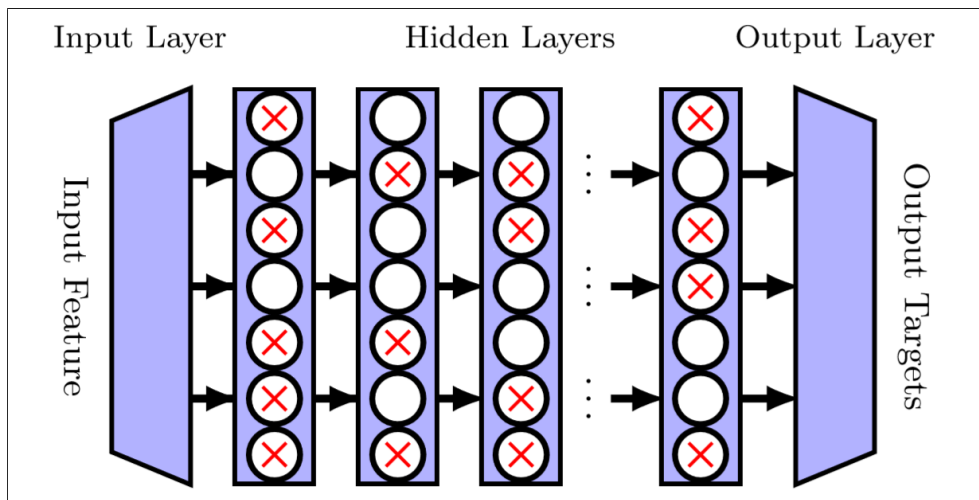
FIGURE 3.9: A thinned network produced by dropout where the crossed are dropped [130].

Since deep neural network consists of multiple nonlinear hidden layers, this makes them very expressive models that can learn complicated relationships between the input and output nodes which often results to overfitting.

The dropout function randomly sets a hidden unit to zero during training with a certain probability range and modifies the final output of the nodes [188]. Each training set can be viewed as providing gradient for a different randomly sampled architecture so that the final network resembles efficiently with good generalization capability. The function takes the output vector and dropout ratio which then returns the new vector that is multiplied to the output vector. Dropout is given as; $yd = Dropout(y, ratio)$ where y is the output vector and *ratio* is the ratio of the dropout of the output vector. Dropout only trains randomly selected nodes at certain percentages rather than the entire network where the selected node outputs are set to zero to deactivate the nodes [112]. Deep neural networks with dropout have proven to perform well in many different applications and greatly reduces computational time [189].

3.6.3   Deep Learning Architecture

A deep learning model requires the modification of various hyperparameters in order to have acceptable results, these largely depend on the type of dataset and other hyperparameters.

The best choice of these parameters depends on the problem which often requires a trial an error process before a good choice is found [169]. Having run the simulation a number of times based on trial an error, the optimum values of the parameters were found which were then used to train the network model.

All training trails for different HCs were conducted with the learning rate α set to 0.01, β set to 0.5, γ set to 0.1 and a dropout ratio of 0.20 which allowed convergence of the objective function. The results illustrate the effectiveness of the developed deep learning approach based on the three-term backpropagation algorithm with dropout.

An exploratory approach was used in considering the size of the hidden layer of the neural network, different sizes up to 10 layers and 100 nodes were considered for the hidden layer. After the exploratory stage, the effective range was narrowed down to between 3 - 10 layers and between 20 - 50 nodes. All possible values within the effective hyperparameter range were evaluated in order to obtain the best fit for the model.

Finally, given similar performance with hyperparameters that results in lower model complexity, the deep learning model with 4 hidden layers and 30 nodes happens to be the best fit for the other hyperparameters.

Regarding sensitivity of the deep learning model to the choice of hyperparameters, a decrease in performance was observed with increase or decrease of the parameter values such as number of hidden layers, number of nodes, values of alpha (α), beta (β) and gamma (γ). It was, however noted to be difficult to isolate the effect of individual hyperparameter, as the value of one

affects the optimal value of the other. For example, it was noted that there is a close relationship between the number of nodes in each hidden layer and the regularization weight, specifically, a network with large number of nodes may demand higher regularization weight [179].

The deep learning model was designed using the number of bands (288) as input to the network and 4 hidden layers each containing 30 nodes 1 output corresponding to the abundance of hydrocarbon. The network was trained using the ground truth abundances for the different mixtures. The network used Sigmoid activation function and a range of dropout ($10\% - 50\%$).

In the proposed deep learning approach, dropout on hidden layers and on the visible layer are developed. Dropout on hidden layers is applied to hidden neurons in the hidden layers and between the last hidden layer and the output layer of the body of the deep networks' model. Dropout on the visible layer is applied between the input and the first hidden layer.

The algorithm was run iteratively with 20 epochs. The deep learning model was implemented using MatLab 2018b. The experiments were carried out on an LG desktop with Intel (R) core $(TM)^2$ Duo CPU 3.00 GHZ processor 8.00 GB RAM. The learning process terminates when the iterations are over a fixed number or the total squared error is less than a small threshold.

3.6.4 Training with three-term backpropagation algorithm with dropout

The three-term backpropagation algorithm with dropout was used to train the network in order to minimize the measured error by adjusting the weights.

To train the network, initial weights and bias must be defined as well as the activation function. The values of the three-terms (α, β and γ) determined using trial an error method were used.

Training a network is mostly affected by overfitting, which is poor performance on the test set after the training, this affects the model's ability to generalize on unseen data which is a major challenge for the deep learning approach. However, combining this approach with dropout can be a good way of overcoming this challenge. Using this method, all the parameters are trained with the standard gradient decent algorithm. misclassifications obtained from the feedforward process are utilized during backpropagation algorithm to calculate the gradient. These gradients are then used to update the training parameters. The three-term backpropagation algorithm is used to calculate the gradients error with the ground truth data and the parameters of the model are adjusted to optimize the learning model.

Deep learning model with dropouts can be trained using the stochastic gradient descent which can be similar to a standard neural networks, the only difference here is the random dropping of units in the network's hidden layers. Different methods have been used to improve the standard gradient descent algorithm such as momentum, annealed learning rates and weight decay [19].

The three-term backpropagation algorithm with dropout was used to train the deep learning model. The effectiveness of using this algorithm to train the model is demonstrated, where the dropout probability for each hidden variable is computed so as to achieve better results of fine-tuning as demonstrated in algorithm 2.

To evaluate the generalizability of our proposed model, the results were obtained using 20- fold cross validation. Each dataset was split into training and testing thus obtaining a one - to - one correspondence of cross validation folds.

With this experimental design, the overall performance of the model was assessed and the accuracy of the model on all the experiments.

Algorithm 2: Learning method using the three-term backpropaga-
tion with dropout used in training the DNN model

Data: $\alpha, \beta, \gamma, \omega, \delta, e$

DNN weights ω, are randomly initialized ω_{11} **... ... ...** ω_{ij}

for *Number of epochs (k)* **do**

    initialize the learning rates α , β, γ, ω

    for *Number of HC estimates = 1, 2 ..... n* **do**

        for *Number of hidden layers = 1, 2 ..... f* **do**

            /* Calculate the errors and the delta, δ, of the output nodes

            $e = d - y$

            $\delta = \phi'(v)\,e$

            - compute the nodes' output y_{ij}

            $y_{ij} = y_{ij} * \text{Dropout}(y_{ij}, \text{ratio})$

            - propagate the network output y_z backwards, and

            calculate the delta, δ

            $e^p = W^T\,\delta$

            $\delta^p = \phi'(v^p)\,e^p$

            /* update and adjust the weights according to δ

            $\Delta\omega_{ij}(k) = \alpha\,\delta_i\,x_j + \beta\,\Delta\omega_{ij}(k-1) + \gamma\,e_i\,x_j$

            $\omega_{ij} \leftarrow \omega_{ij} + \Delta\omega_{ij}$

        end

    end

end

3.7   Summary

In this chapter, the hybrid switch method for switching between linear and nonlinear spectral unmixing of hyperspectral data based on artificial neural networks and the deep learning based spectral unmixing method were discussed, detailing the architectures of the two proposed methods, the training algorithms as well as validation methods of the proposed models. Validation of the method through experimentation with simulated and real hyperspectral data sets will be discussed in the next chapter. An experiment using a controlled hyperspectral data for HC spill detection and estimation will be validated using the two proposed methods in the next chapter.

Chapter 4

Experimentation and results

4.1 Introduction

In this chapter, experimentation and results from the two proposed methods as described in chapter 3 are presented. The results aim to show validation of the proposed methods and the their performances against conventional neural networks. Another aim is to demonstrates the effectiveness of the proposed methods for HC spill detection using controlled datasets to validate the methods. This chapter is organized as follows. Section 4.2 describes all the datasets used for the experiments, including the materials used and the detailed experimental protocol observed during the image capturing of the controlled datasets. Section 4.3 explain the results obtained from the hybrid switch methods using different datasets. Section 4.4 describes the hydrocarbon abundance estimation using the hybrid switch methods and the deep learning approach for HC spill detection. This chapter concludes with a summary in Section 4.5.

4.2 Data Description

In order to validate the hybrid switch method, different experiments were conducted with different datasets, including both simulated and real datasets which are described in this section.

4.2.1   Simulated Data

Simulated images of size 36×36 pixels and 224 channels were generated with abundances computed according to a Dirichlet distribution (a method to model random probability mass functions for finite sets [80]) defined as :

$$y(\gamma) = (\Gamma(\beta_1 + \beta_2)\Gamma(\beta_1)\Gamma(\beta_2))\gamma^{\beta_1 - 1}(\gamma - 1)^{\beta_2 - 1} \qquad (4.1)$$

Where $\Gamma(\cdot)$ denotes the Gamma function and parameter γ is Beta $(\beta_1 \ \beta_2)$ distributed.

The data was generated using the ENVI spectral library that contains 21 endmembers corresponding to mineral reflectances with 224 bands in the 500 - 2500 m range [60]. Additionally, a nonlinearity co-efficient was added ranging between $[0, 1]$, these parameters were tuned accordingly with different numbers of endmembers ranging between 3 and 9. The images were corrupted with Random Gaussian noise with Signal to Noise Ratio (SNR) 10 dB, 30 dB and 50 dB respectively. Figure 4.1 shows the spectral reflectance of endmembers of the simulated data.
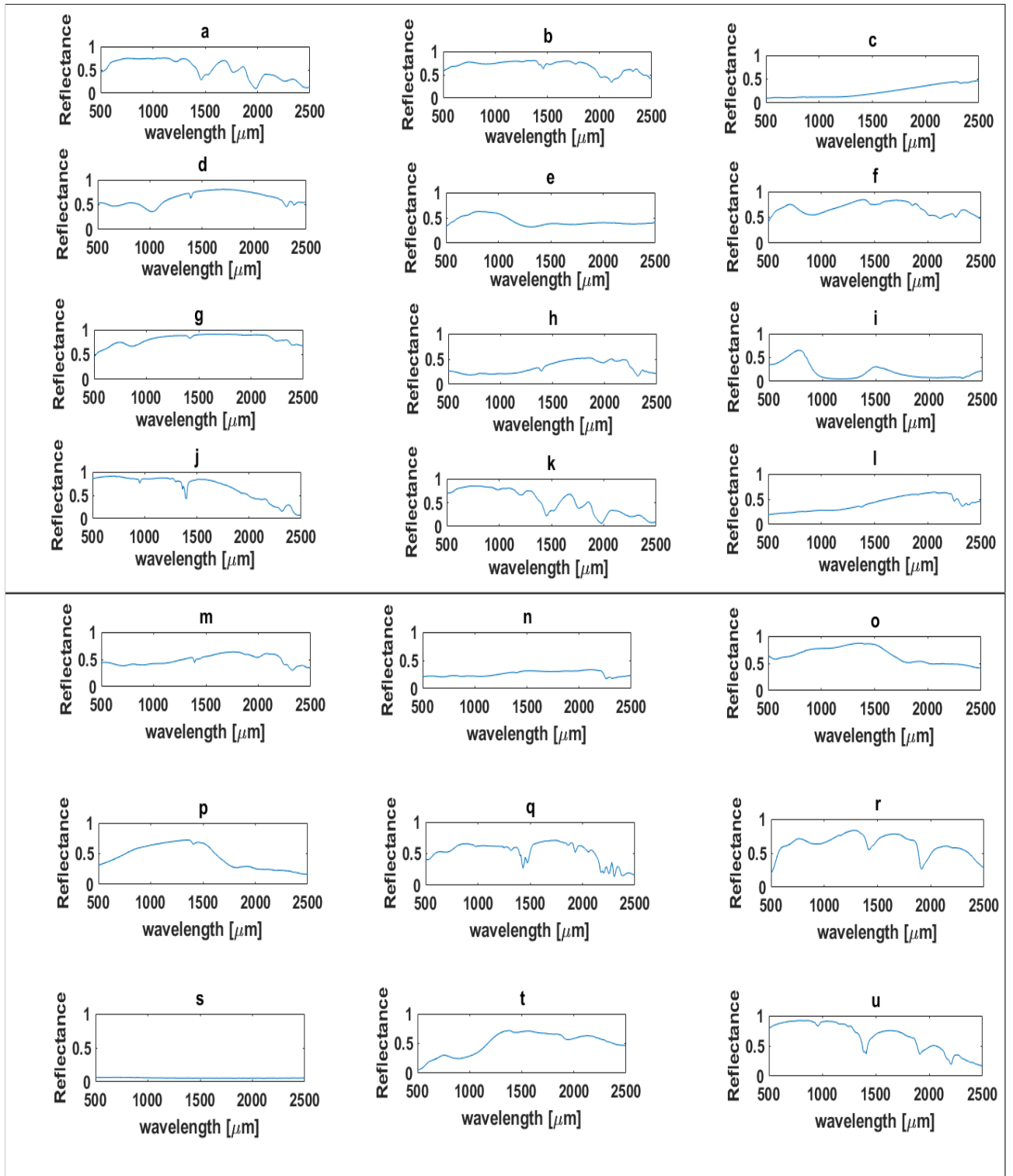
FIGURE 4.1: Spectral reflectance of endmembers of the simulated data plotted, Normalised reflectance[%] against wavelength (a) Brucite; (b) Clinochlore; (c) Axinite; (d) Erionite; (e) Ammonioalunite; (f) Clintonite; (g) Almandine; (h) Carnallite; (i) Actinolite; (j) Andradite; (k) Antigorite; (l) Elbaite; (m) Ammonio-jarosite; (n) Diaspore; (o) Halloysite; (p) Biotite; (q) Galena; (r) Carnallite; (s) Chlorite; (t) Goethite; (u) Corundum.

4.2.2 Samson data

The Samson data used in this research is a real hyperspectral data owned by Oregon State University provided by WeoGeo [190]. The data is captured with a push broom visible to near infrared sensor. The pixel responses are captured by 156 bands in the spectral range of 401 nm – 889 nm with resolution up to 3.13 nm. The data has 952 scan lines with 952 pixels in each line [167]. For this experiment a subset of the image covering 95×95 pixels was used, which comprises of three endmembers i.e soil, tree and water. Figure 4.2 shows the spectral reflectance of endmembers of the Samson dataset.
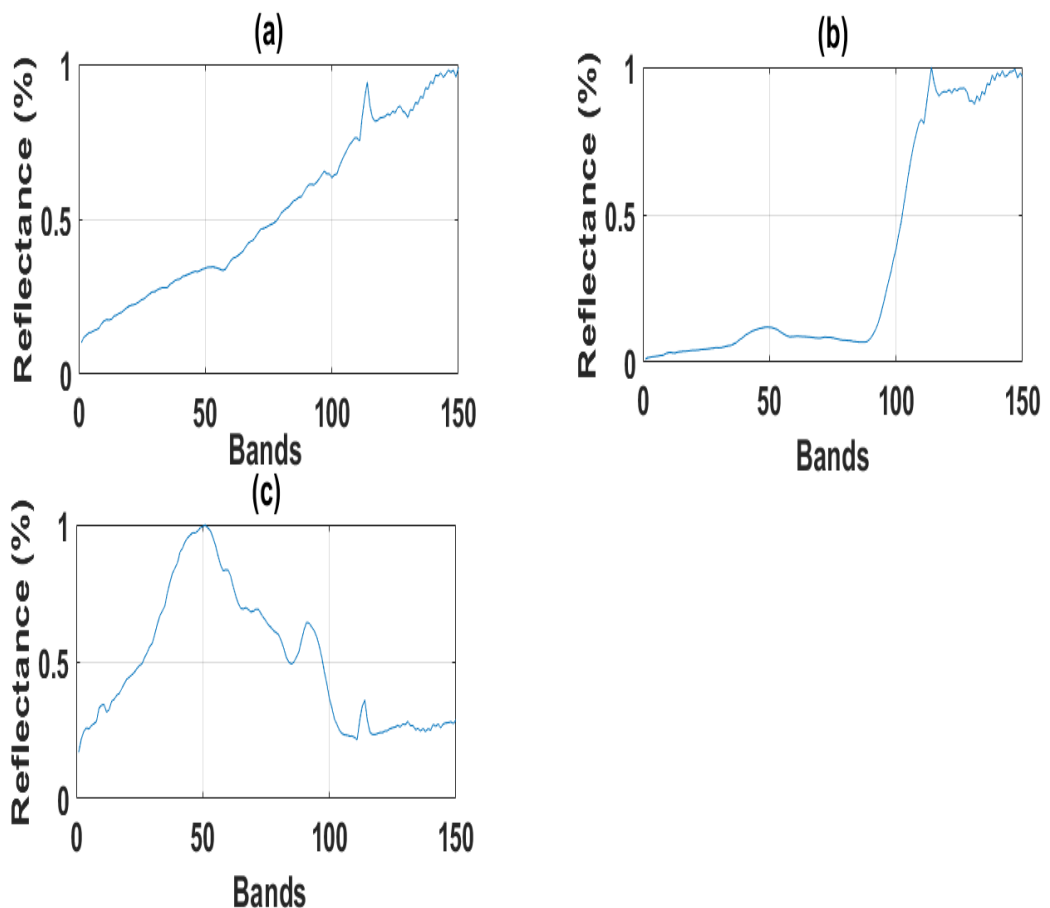
FIGURE 4.2: Spectral reflectance of endmembers of the Samson data (a) rock (b) tree (c) water [190].

4.2.3 Jasper Ridge

The Jasper Ridge dataset is also a real hyperspectral data cube recorded by AVIRIS over the standard scene of the Jasper Ridge. This image covers a biological reserve in California. The dataset consist of 512×614 pixels recorded in 224 channels ranging from 380 nm to 2500 nm. The data has a spectral resolution of 9.46 nm. A subset of the image of size 100×100 pixels was used from the original image and 198 bands were selected after removing those bands with atmospheric effects and dense water vapor. There are four main endmembers in this image: road, soil, water and tree.

Both Samson and Jasper Ridge datasets and their corresponding abundance ground truth are available at [190]. These datasets have been used as benchmark to test different unmixing algorithms and classification. Figure 4.3 show the spectral reflectance of endmembers of the Jasper Ridge dataset.
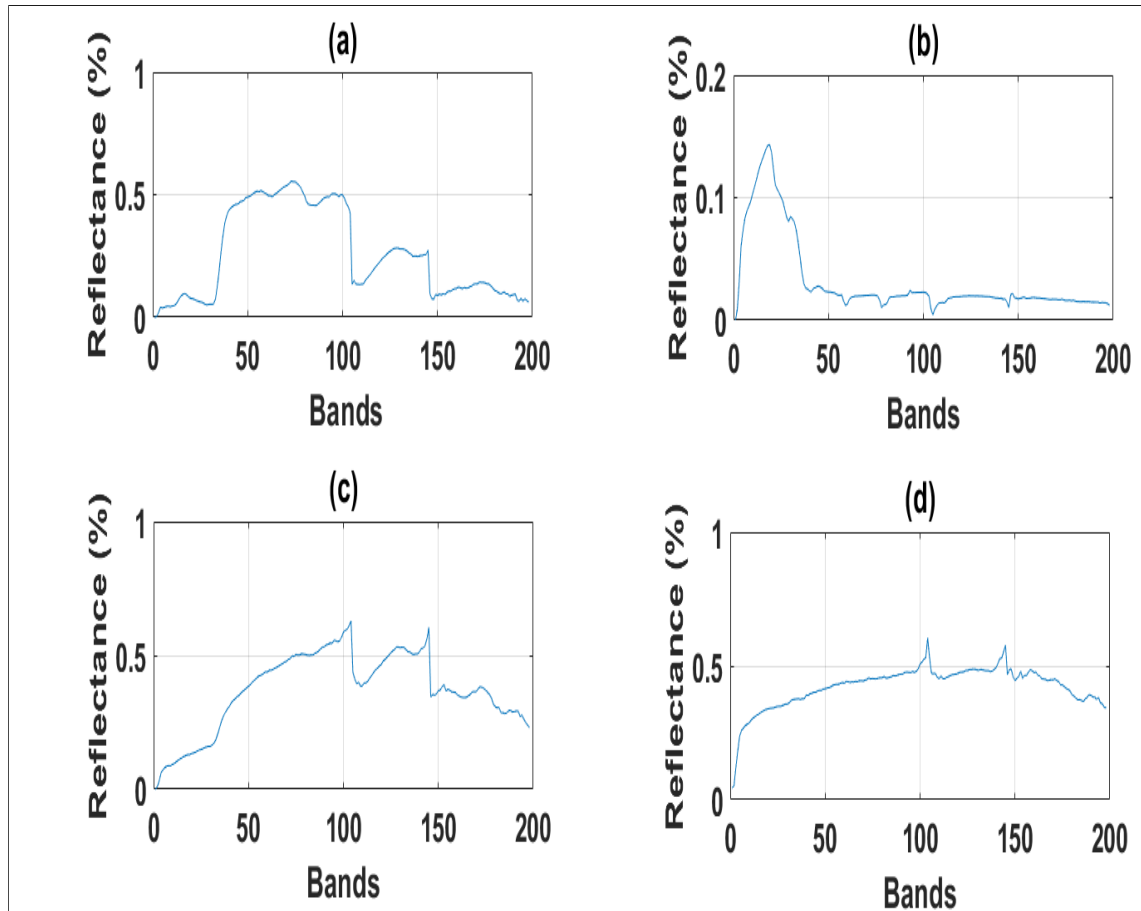
FIGURE 4.3: Spectral reflectance of endmembers of the Jasper
Ridge data (a) tree (b) water (c) soil (d) road [190].

4.2.4   Controlled dataset

A controlled laboratory experiment was designed where we analysed several soils contaminated with various types and concentration of HC in order to create predictive model to correctly estimate the amount of HC in each soil mixture which could be applied worldwide for rapid remediation to both oil exploration and environmental monitoring.

4.2.5   Materials for the controlled dataset

The hydrocarbon types used for this experiment include: Diesel, Bio-diesel, Ethanol and Petroleum. These are the most commonly used HCs in the literature. Soil types include typical mixtures of: Clay ($< 0.002mm$ in diameter),

Silt $(0.002 - 0.05$ mm in diameter) and Sand $(0.05 - 1$ mm in diameter). In particular, we used mixtures with different grain size ranging from medium to coarse as follows: Clay, Clay Loam , Sand Clay Loam and Sand Loam. Percentages of textural classes of different soil types are shown in Figure 4.4 [191].
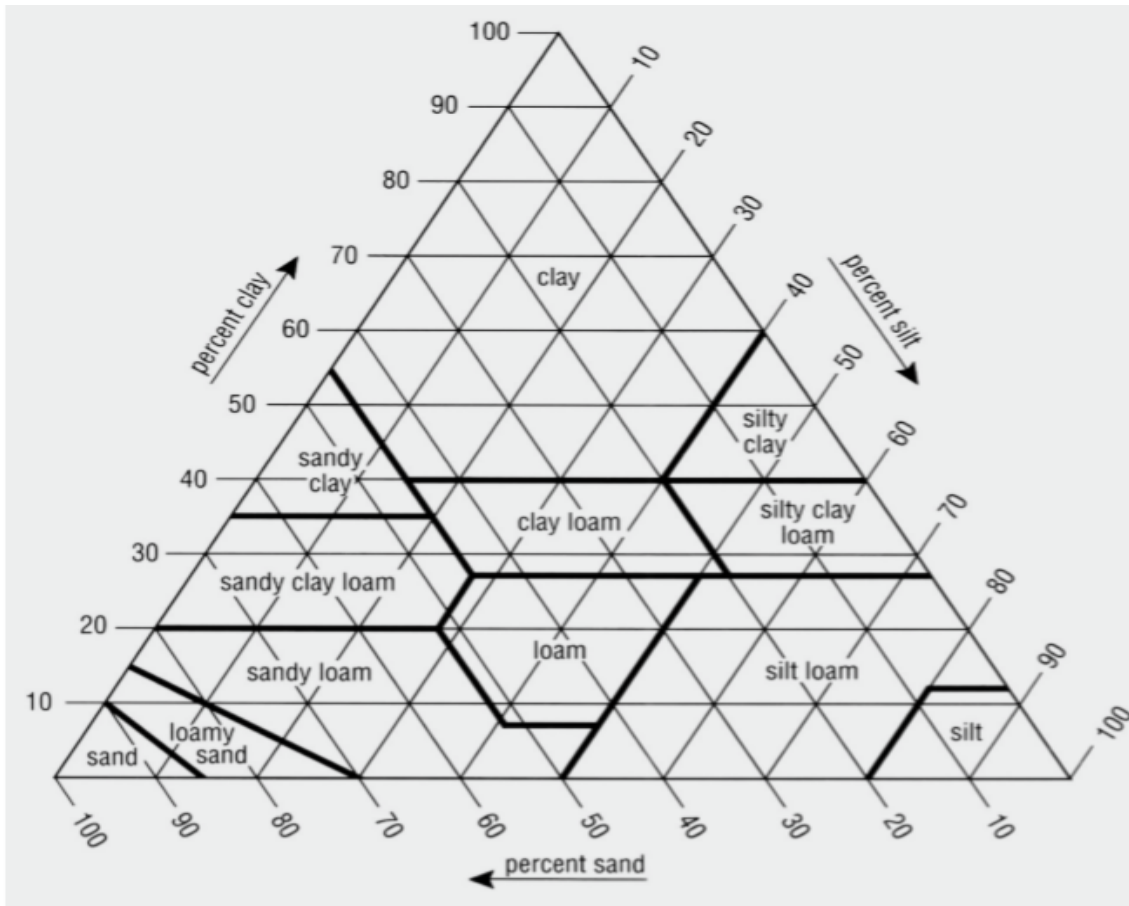
FIGURE 4.4: Percentages of different soil samples in a basic textural classes [191] .

4.2.6 Sample Preparation

In order to achieve uniform datasets on all the different materials, the measurement and scanning was conducted under constant illumination.

The preparation of the samples consisted of the following steps:

- Each soil type was air-dried for 12 hours so all samples contained similar levels of moisture.

- 50 g of a soil sample type was added to a petri dish (12 cm in diameter)

- The petri dish was placed on the camera's field of view along side a white paper (to calibrate) in order to obtain an image with the best possible contrast.

- The sample was scanned with a Hyspex SWIR 384 m camera under constant illumination.

- In the same sample, 5 ml of the HC was added to the soil using a syringe.

- A disposable plastic spoon was used to homogenize the mixture and to flatten its surface.

- The sample was scanned with a Hyspex SWIR 384 m camera under constant illumination.

- In the same sample, a further 5 ml of HC was added to the mixture

- The disposable spoon was used to homogenize the mixture and another scan was taken.

- The procedure was repeated with increments of 5 ml of HCs until the mixture was saturated and formed a shallow local pool as shown in Figure 4.5.

FIGURE 4.5: Sample preparation of the experiment combining sandy - loam with diesel. Photos show the HC contaminant being increasingly added to the same soil sample until saturated. From left; addition of 5*ml* , followed by 10*ml*, 15*ml*, 20*ml* and 25*ml* of the HC

The procedure was repeated on all the soil samples contaminated with all the different hydrocarbon types. A total of 15 combinations were produced as shown in Table 4.1 with four mixtures each for clay - loamy, sandy - clay - loam and sandy - loam soil types, while clay had three mixtures. The complete data set used here consisted of 96 spectral images.

TABLE 4.1: Samples created for each combination made in
the experiment and their corresponding absolute HC and soil
quantities, respectively

Sample combination	HC (ml)	Soil (gr)	Sample combination	HC (ml)	Soil (gr)
Clay - Diesel 0	0	50	Clay - Bio- diesel 0	0	50
Clay - Diesel 1	5	50	Clay - Bio- diesel 1	5	50
Clay - Diesel 2	10	50	Clay - Bio- diesel 2	10	50
Clay - Diesel 3	15	50	Clay - Bio- diesel 3	15	50
Clay - Diesel 4	20	50	Clay - Bio- diesel 4	20	50
Clay - Diesel 5	25	50	Clay - Bio- diesel 5	25	50
Clay - Ethanol 0	0	50	Clay Loam - Ethanol 0	0	50
Clay - Ethanol 1	5	50	Clay Loam - Ethanol 1	5	50
Clay - Ethanol 2	10	50	Clay Loam - Ethanol 2	10	50
Clay - Ethanol 3	15	50	Clay Loam - Ethanol 3	15	50
Clay - Ethanol 4	20	50	Clay Loam - Ethanol 4	20	50
Clay - Ethanol 5	25	50	Clay Loam - Ethanol 5	25	50
Clay Loam - Diesel 0	0	50	Clay Loam - Bio- diesel 0	0	50
Clay Loam - Diesel 1	5	50	Clay Loam - Bio- diesel 1	5	50
Clay Loam - Diesel 2	10	50	Clay Loam - Bio- diesel 2	10	50
Clay Loam - Diesel 3	15	50	Clay Loam - Bio- diesel 3	15	50
Clay Loam - Diesel 4	20	50	Clay Loam - Bio- diesel 4	20	50
Clay Loam - Petrol 0	0	50	Sand Loam - Petrol 0	0	50
Clay Loam - Petrol 1	5	50	Sand Loam - Petrol 1	5	50
Clay Loam - Petrol 2	10	50	Sand Loam - Petrol 2	10	50
Clay Loam - Petrol 3	15	50	Sand Loam - Petrol 3	15	50
Clay Loam - Petrol 4	20	50	Sand Loam - Petrol 4	20	50
Clay Loam - Petrol 5	25	50	Sand Loam - Petrol 5	25	50
Clay Loam - Petrol 6	30	50	Sand Loam - Petrol 6	30	50
Clay Loam - Petrol 7	35	50	Sand Loam - Petrol 7	35	50
Clay Loam - Petrol 8	40	50	Sand Loam - Petrol 8	40	50
Clay Loam - Petrol 9	45	50	Sand Loam - Petrol 9	45	50
Sand Clay Loam - Diesel 0	0	50	Sand Clay Loam - Bio- diesel 0	0	50
Sand Clay Loam - Diesel 1	5	50	Sand Clay Loam - Bio- diesel 1	5	50
Sand Clay Loam - Diesel 2	10	50	Sand Clay Loam - Bio- diesel 2	10	50
Sand Clay Loam - Diesel 3	15	50	Sand Clay Loam - Bio- diesel 3	15	50
Sand Clay Loam - Diesel 4	20	50	Sand Clay Loam - Bio- diesel 4	20	50
Sand Clay Loam - Diesel 5	25	50	Sand Clay Loam - Bio- diesel 5	25	50
Sand Clay Loam - Ethanol 0	0	50	Sand Clay Loam - Petrol 0	0	50
Sand Clay Loam - Ethanol 1	5	50	Sand Clay Loam - Petrol 1	5	50
Sand Clay Loam - Ethanol 2	10	50	Sand Clay Loam - Petrol 2	10	50
Sand Clay Loam - Ethanol 3	15	50	Sand Clay Loam - Petrol 3	15	50
Sand Clay Loam - Ethanol 4	20	50	Sand Clay Loam - Petrol 4	20	50
Sand Clay Loam - Ethanol 5	25	50	Sand Clay Loam - Petrol 5	25	50
Sand Clay Loam - Ethanol 6	30	50	Sand Clay Loam - Petrol 6	30	50
			Sand Clay Loam - Petrol 7	35	50
Sand Loam - Diesel 0	0	50	Sand Loam - Bio- diesel 0	0	50
Sand Loam - Diesel 1	5	50	Sand Loam - Bio- diesel 1	5	50
Sand Loam - Diesel 2	10	50	Sand Loam - Bio- diesel 2	10	50
Sand Loam - Diesel 3	15	50	Sand Loam - Bio- diesel 3	15	50
Sand Loam - Diesel 4	20	50	Sand Loam - Bio- diesel 4	20	50
Sand Loam - Ethanol 0	0	50			
Sand Loam - Ethanol 1	5	50			
Sand Loam - Ethanol 2	10	50			
Sand Loam - Ethanol 3	15	50			
Sand Loam - Ethanol 4	20	50			

4.2.7 **Hyperspectral Imaging**

The spectral data was obtained using a Norsk Electro Optikk Hyspex SWIR 384 m line-scan hyperspectral camera and is equipped with a Mercury Cadmium Telluride (MCT) detector array. For this experiment, a user friendly table-top laboratory set up with a translation stage, SWIR light source and close up lenses were used during the scanning stage to scan the sample and build a hyperspectral data cube (Figure 4.6). The camera simultaneously captured a full SWIR spectrum, with a spectral sample interval of 5.45 nm between 930 and 2500 nm, each along a line of 384 pixels for 288 bands with a radiometric resolution of 16 bit [192]. The 384 columns of the detector array formed one line of the hyperspectral image in the x-axis. The hyperspectral image was obtained line by line using the so-called "pushbroom" scanning mode, where the platform holding the sample was translated onto the y-axis at constant speed (Figure 4.7). The scanning speed was automatically controlled by the data acquisition unit based on the selected lens option. The images produced had a spatial resolution of 0.22 mm/pixel. Radiometric calibration was performed using the vendor's software package. A more detailed specification of the system is given in Table 4.2.

TABLE 4.2: Hyspex 384m main specifications

Spectral range (nm)	930 - 2500
Spatial Pixels (pixels)	384
Spectral Channels	288
Spectral Sampling (nm)	5.45
FOV (degrees)	16○
Pixel FOV across/along (mrad)	0.73/0.73
Bit resolution (bit)	16
Noise floor (e$^-$)	150
Dynamic range	7500
Peak SNR (at full resolution)	> 1100
Max speed (at full resolution)(fps)	400
Power consumption (W)	30
Dimensions (l -w- h) (cm)	38 - 12 - 17.5
Weight (kg)	5.7
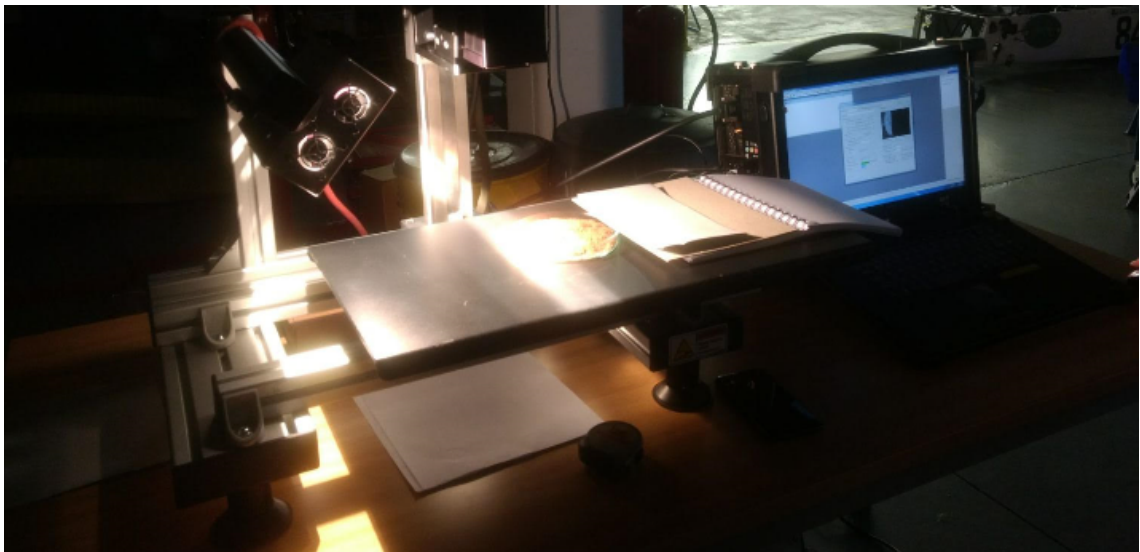
FIGURE 4.6: Scanning process of the dataset
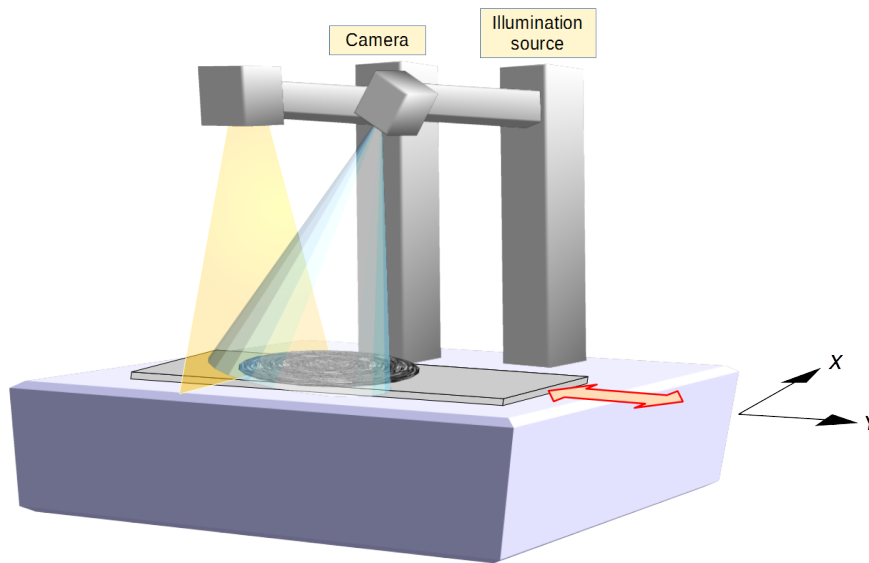
FIGURE 4.7: HySpex 384m line scan acquisition process. The camera (nadir) acquires hyperspectral lines of pixels. The hyperspectral image is obtained by translation of the object under constant illumination.

4.3 Results from the Hybrid Switch Method

4.3.1 Simulated data

This experiment was conducted using the simulated dataset described in 4.2.1, which allows a full control of the data. Here, the algorithms: VCA, FCLS, PPNMM and GBM were used to unmix spectra of different mineral mixtures in the simulated dataset. The accuracy of the individual methods to the proposed hybrid methods for switching between linear and nonlinear spectral unmixing methos based on the diversity of the neighboring pixels was compared. The hybrid methods for switching were between VCA – PPNMM, VCA – GBM, FCLS – PPNMM, and FCLS – GBM, respectively. The experiment was repeated with different numbers of endmembers ranging 3,

5, 7 and 9 and different Signal to Noise Ratios (SNR) of 10 dB, 30 dB and 50 dB respectively on the simulated dataset. Monte Carlo simulations based on 100 iterations were ran on the generated images for each experiment.

To conduct the switching, the samples were randomly split into 3 sets namely: training sets, validation sets and test sets. To train a network the amount of training data required depends on the nature of the problem and the structure/ architecture of the network, however, it is noted that the greater the amount of training data, the better the accuracy of the network. As a rule of thumb, training a network requires roughly about 10 times as many examples as there are degrees of freedom in a particular model [193].

During training, 70% of the datasets were randomly selected to train the network, 15% were used for validating the model so as to learn the hyperparameters of the neural networks and 15% of the remaining samples were used to test the accuracy of the networks.

In this experiment, a 3×3 window was used around the pixel of interest. A vector containing 12 values i.e SAD min, SAD max, covariance matrix (9 values) and nonlinearity were computed for each pixel as input to train the network. Each input data consisted of 12 nodes with the number of hidden nodes set to 10 and the output layer having 1 node which output (0 or 1) corresponding to either a linear or nonlinear approach with a threshold of 0.5 set for the switching. The overall accuracy of the experiment and the abundance estimation error of the methods were computed and summarized in Table 4.3. Result shows that the VCA – PPNMM hybrid method predicted better overall accuracy of 98.8% as estimated by the confusion matrix with neural networks in switching between linear and nonlinear spectral unmixing, followed by FCLS – PPNMM with an overall accuracy of 95.6%, VCA – GBM and FCLS – GBM both have an overall accuracy of 92% and 92.4% respectively.

TABLE 4.3: Abundance estimation error (3×3 window) of the individual and hybrid methods between linear and nonlinear spectral unmixing with different signal to noise ratios and endmembers. The best results are shown in bold

.

SNR (dB) = 50	P = 3	P = 5	P = 7	P = 9
INDIVIDUAL METHODS				
PPNMM	0.0206	0.0307	0.0371	0.0486
GBM	0.0207	0.0303	0.0346	0.0449
VCA	0.0521	0.0696	0.0777	0.0778
FCLS	0.0714	0.0916	0.0922	0.0924
HYBRID METHODS				
VCA – PPNMM	**0.0117**	0.0201	**0.0143**	0.0373
VCA – GBM	0.0189	0.0201	0.0158	0.0353
FCLS – PPNMM	0.0177	**0.0179**	0.0177	**0.0340**
FCLS – GBM	0.0193	0.0196	0.0199	0.0174
SNR (dB) = 30	**P = 3**	**P = 5**	**P = 7**	**P = 9**
INDIVIDUAL METHODS				
PPNMM	0.0696	0.0951	0.0914	0.0886
GBM	0.0965	0.1193	0.1405	0.1285
VCA	0.0597	0.0662	0.0886	0.0945
FCLS	0.0684	0.0747	0.0894	0.0911
HYBRID METHODS				
VCA – PPNMM	**0.0390**	**0.0317**	**0.0421**	**0.0556**
VCA – GBM	0.0591	0.0412	0.0579	0.0662
FCLS – PPNMM	0.0396	0.0320	0.0539	0.0645
FCLS – GBM	0.0866	0.0926	0.0990	0.1081
SNR (dB) = 10	**P = 3**	**P = 5**	**P = 7**	**P = 9**
INDIVIDUAL METHODS				
PPNMM	0.0907	0.1510	0.1640	0.1733
GBM	0.1106	0.1222	0.1334	0.1740
VCA	0.1289	0.1514	0.1257	0.1988
FCLS	0.1169	0.1702	0.1791	0.1763
HYBRID METHODS				
VCA – PPNMM	0.0401	**0.0421**	**0.0736**	**0.0775**
VCA – GBM	0.0704	0.0911	0.0813	0.0915
FCLS – PPNMM	**0.0440**	0.0508	0.0813	0.0814
FCLS – GBM	0.0917	0.0959	0.1099	0.1112

The abundance estimation error of the generated data with different Signal to Noise Ratio (SNR = 50 and SNR = 10) are shown in Figures 4.8 and 4.9. Here the results displays the abundance estimation error of the individual as well as the hybrid methods.

The first row shows the ground truth abundances displayed in grayscale where a white pixel means abundance equal to one for that particular class and a black pixel means no abundance found for that class. The other rows shows the abundance estimation error for each class and each method, these are also displayed in grayscale where bright pixel indicates high error.
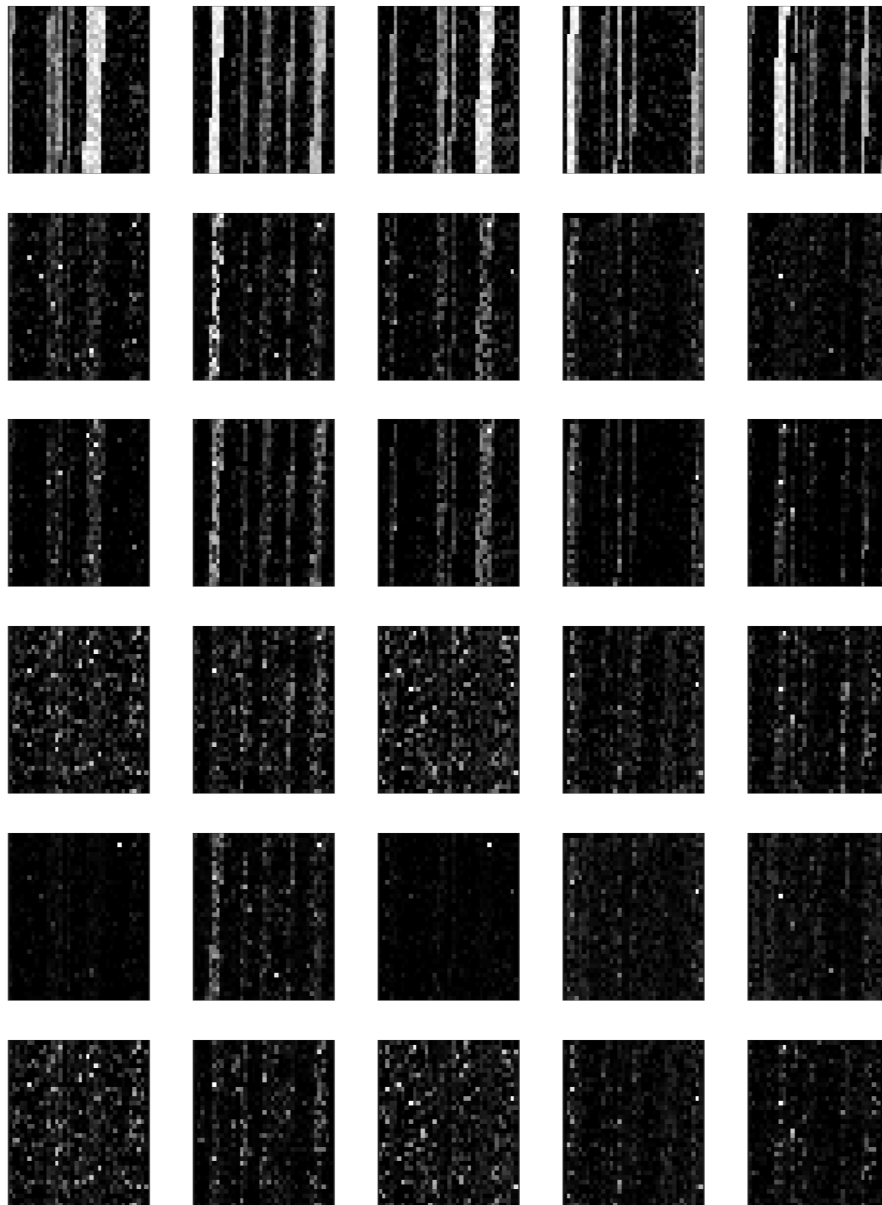
FIGURE 4.8: Abundance estimation errors with simulated data with 5 endmembers (SNR = 50 dB) The first row shows the ground truth abundances for the 5 classes from the top, then the error in abundances as estimated by the hybrid, VCA, FCLS, PPNMM and GBM methods, respectively.
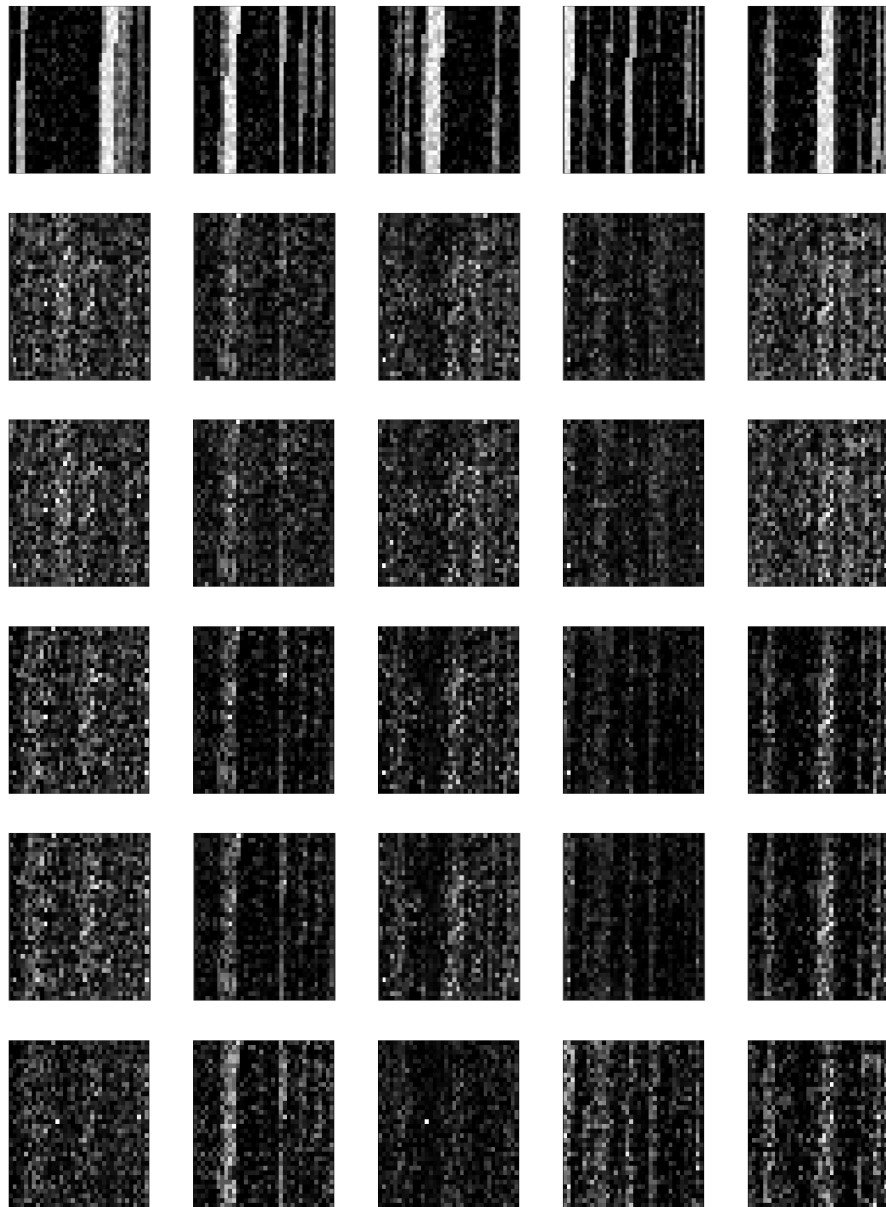
FIGURE 4.9: Abundance estimation errors with simulated data with 5 endmembers (SNR = 10 dB) The first row shows the ground truth abundances for the 5 classes from the top, then the hybrid, VCA, FCLS, GBM and PPNMM methods, respectively.

4.3.2 Vicinity parameters

Another experiment was conducted with the simulated dataset around a 3×3 window. In this experiment, the vicinity parameters used in creating

the input training data were excluded one at a time. This was done to evaluate the importance of the parameters in the vector created. The experiment was also repeated for SNR values $10dB$, $30dB$, and $50dB$ with different endmembers of 3, 5, 7 and 9 respectively. Here it is expected that higher error values will be recorded when each of the parameters are removed from the vector in comparison with the results in Table 4.3 where all the parameters are involved in the experiment. Table 4.4 shows that all parameters play an important role in the vector for the hybrid switch methods. SAD max has demonstrated to be the most important parameter in the vector created with the highest error value in the experiment where it was excluded as compared to the other parameters.

TABLE 4.4: .

Abundance estimation error (3×3 window) of the individual and hybrid methods between linear and nonlinear spectral unmixing with different signal to noise ratios and 3 endmembers where each of the parameters is removed one at a time The best results are shown in bold

WITHOUT SAD MIN.			
	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.1503	0.0537	0.0179
GBM	0.1220	0.1274	0.0168
VCA	0.1090	0.1000	0.0952
FCLS	0.1670	0.1370	0.0997
HYBRID METHODS			
VCA – PPNMM	**0.0433**	**0.0392**	0.0150
VCA – GBM	0.0854	0.0784	0.0163
FCLS – PPNMM	0.0434	0.0402	**0.0143**
FCLS – GBM	0.1180	0.1080	0.0161
WITHOUT SAD MAX.			
	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.0969	0.0876	0.0878
GBM	0.1002	0.0920	0.0741
VCA	0.1216	0.0791	0.0451
FCLS	0.2726	0.1073	0.0560
HYBRID METHODS			
VCA – PPNMM	0.0584	**0.0467**	**0.0251**
VCA – GBM	0.0885	0.0731	0.0525
FCLS – PPNMM	**0.0521**	**0.0467**	**0.0251**
FCLS – GBM	0.1689	0.1000	0.0772
WITHOUT COVARIANCE DISTANCE			
	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.0940	0.0518	0.0173
GBM	0.1243	0.0921	0.0166
VCA	1.0488	0.0824	0.0590
FCLS	1.1673	0.0966	0.0680
HYBRID METHODS			
VCA – PPNMM	**0.0506**	0.0340	**0.0145**
VCA – GBM	0.0902	0.0588	0.0163
FCLS – PPNMM	**0.0506**	**0.0336**	**0.0145**
FCLS – GBM	0.1231	0.0916	0.0161
WITHOUT NONLINEARITY PARAMETER			
	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.1447	0.0532	0.0183
GBM	0.1100	0.1039	0.0185
VCA	0.1251	0.0982	0.0865
FCLS	0.1852	0.1167	0.0927
HYBRID METHODS			
VCA – PPNMM	**0.0448**	0.0432	0.0173
VCA – GBM	0.0856	0.0789	0.0178
FCLS – PPNMM	**0.0448**	**0.0431**	**0.0171**
FCLS – GBM	0.1236	0.1096	0.0137

In order to assess the accuracy of the hybrid switch methods, the network was also trained with the raw data (i.e 224 bands) as input instead of the vicinity parameters. Here we have 224 while the rest of the parameters remain the same. Results of 100 Monte Carlo simulations are summarized in Table 4.5. From the results obtained, it was observed that the trend is in the same order of magnitude as obtained when the vicinity parameters were used to train the network as demonstrated in Table 4.3.

TABLE 4.5: .

Abundance estimation error with the individual and hybrid methods of the raw hyperspectral data between linear and nonlinear spectral unmixing with different signal to noise ratios and different endmembers. The best results are shown in bold

SNR (dB) = 50	P = 3	P = 5	P = 7	P = 9
INDIVIDUAL METHODS				
PPNMM	0.0253	0.0276	0.0378	0.0418
GBM	0.0253	0.0276	0.0347	0.0383
VCA	0.0775	0.0612	0.0717	0.0719
FCLS	0.0891	0.0663	0.0877	0.0612
HYBRID METHODS				
VCA – PPNMM	**0.0125**	**0.0127**	**0.0230**	**0.0285**
VCA – GBM	0.0457	0.0164	0.0269	0.0317
FCLS – PPNMM	0.0217	0.0214	0.0236	0.0316
FCLS – GBM	0.0513	0.0627	0.0850	0.0981
SNR (dB) = 30	P =3	P = 5	P = 7	P = 9
INDIVIDUAL METHODS				
PPNMM	0.1520	0.1759	0.1464	0.1353
GBM	0.1568	0.1442	0.1473	0.1337
VCA	0.1007	0.1195	0.0313	0.2767
FCLS	0.1072	0.1713	0.1344	0.1819
HYBRID METHODS				
VCA – PPNMM	**0.0231**	**0.0223**	**0.0219**	**0.0268**
VCA – GBM	0.0317	0.0360	0.0364	0.0370
FCLS – PPNMM	0.0308	0.0358	0.0458	0.0654
FCLS – GBM	0.0437	0.0787	0.0901	0.0956
SNR (dB) = 10	P = 3	P = 5	P = 7	P = 9
INDIVIDUAL METHODS				
PPNMM	0.1809	0.1816	0.1856	0.1883
GBM	0.1517	0.1506	0.1440	0.1481
VCA	0.1196	0.0612	0.0717	0.0717
FCLS	0.1072	0.0663	0.0877	0.0612
HYBRID METHODS				
VCA – PPNMM	**0.0548**	**0.0564**	**0.0570**	**0.0584**
VCA – GBM	0.0751	0.0940	0.0962	0.0961
FCLS – PPNMM	0.0714	0.0739	0.0740	0.0763
FCLS – GBM	0.0974	0.0981	0.0990	0.1170

From the experiments conducted between the 3×3 window and the raw data, it can be seen that similarities exist in the results where the Signal to Noise Ratios are set to 10 dB and 50 dB. However, a better result was observed with the 3×3 window and a Signal to Noise Ratio of 30 dB. Therefore, it shows that the ANN does not require the whole raw data to train the network and the reduced chosen parameters provide good and acceptable results.

4.3.3   Jasper Ridge

This experiment was conducted with the Jasper Ridge dataset to evaluate the accuracy of the methods involved. In this experiment, the raw data, and the vicinity parameters computed within a 3×3 and a 4×4 windows respectively were used to train the neural network. In this experiment, the training samples for each experiment were selected randomly, 70% of the raw data was used for training corresponding to ($7,000$ samples), 15% each were considered for validation and testing ($1,500$ samples each for validation and testing) the neural networks. In another experiment, the number of training samples was reduced, with 30% used for training corresponding to ($3,000$ samples), 35% each was used for validation and testing which corresponds to ($3,500$ samples) each for validation and testing the neural networks. Finally, the experiment was repeated with $1,000$ and 300 training samples respectively. Results of the experiment is summarized in Table 4.6

TABLE 4.6: Average abundance estimation error of the hybrid methods with different numbers of training samples (7000 to 300) and different window size vectors on the Jasper Ridge data as compared with the abundance estimation error of the individual methods which are: PPNMM = 0.2115, GBM = 0.2441, VCA = 0.6513, and FCLS = 0.1832. The best results are shown in bold.

Raw data	7,000	3,000	1,000	300
VCA – PPNMM	0.1417	0.1478	**0.1405**	0.1994
VCA – GBM	0.2079	0.2049	0.3087	0.3897
FCLS – PPNMM	0.1402	0.1402	0.1590	0.1663
FCLS – GBM	**0.1399**	**0.1397**	0.1483	**0.1495**
3×3 **WINDOW**				
VCA – PPNMM	**0.1697**	**0.1607**	**0.1781**	**0.1763**
VCA – GBM	0.2595	0.2454	0.2932	0.3350
FCLS – PPNMM	0.1765	0.1624	0.1783	0.1790
FCLS – GBM	0.2448	0.2448	0.3442	0.3642
4×4 **WINDOW**				
VCA – PPNMM	0.1712	**0.1640**	0.1736	**0.1704**
VCA – GBM	0.2488	0.2250	0.3117	0.3460
FCLS – PPNMM	**0.1632**	0.1659	**0.1705**	0.1722
FCLS – GBM	0.2647	0.2459	0.2488	0.2732

Figure 4.10 shows the ground truth abundances and the abundances as estimated by a linear (VCA), nonlinear (PPMM) and the corresponding hybrid methods on the Jasper Ridge data.
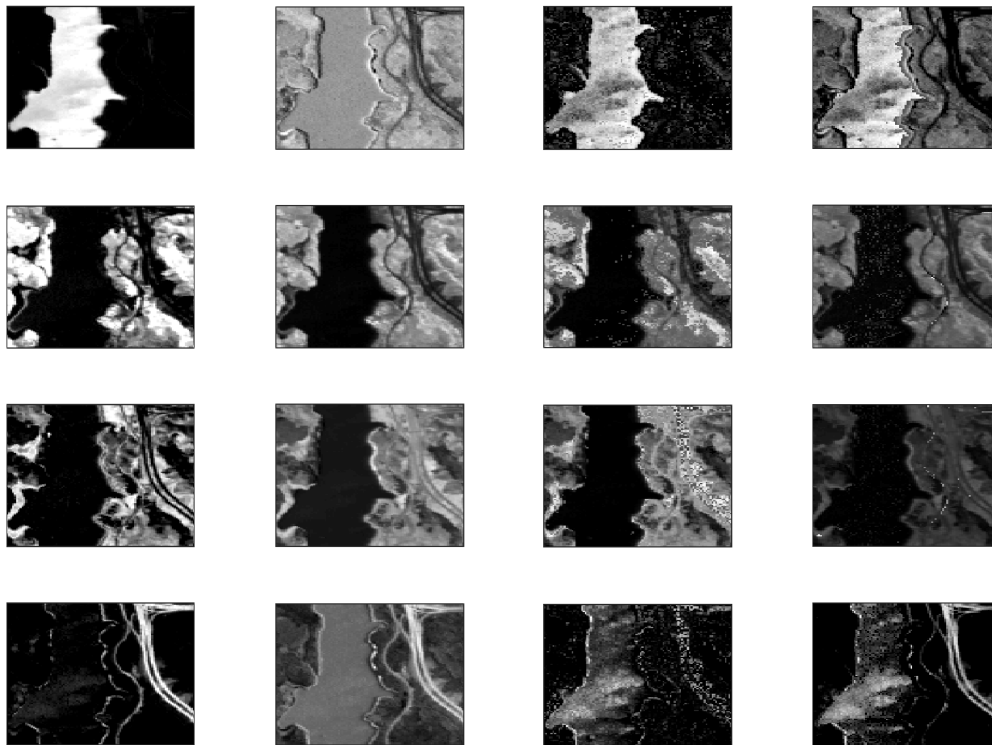
FIGURE 4.10: Abundance estimate of endmembers of the Jasper Ridge data showing from left; the ground truth, linear (VCA), nonlinear (PPMM) and the hybrid methods. From top water, tree, soil and road.

4.3.4 Samson data

Another experiment was conducted with the Samson data, the samples were randomly split into training, validation and testing sets. 70% for training, 15% for validation and 15% for testing respectively which corresponds to $6,317$ samples for training, $1,353$ samples each for testing and validation, Another experiment was conducted where the number of training samples

were reduced; 30% for training; 35% each for validation and testing equivalent to 2,707 samples for training, 3,158 samples each for validation and testing the network respectively. Finally, the experiment was repeated with 1,000 and 300 training samples, respectively. Results of the experiment is summarized in Table 4.7.

TABLE 4.7: Average abundance estimation error of the hybrid methods with different numbers of training samples (6317 to 300) and different window size vectors on the Samson data as compared with the abundance estimation error of the individual methods which are: PPNMM = 0.1455 , GBM = 0.1588, VCA = 0.1254, and FCLS = 0.1577. The best results are shown in bold.

Raw data	6,317	3,158	1,000	300
VCA – PPNMM	**0.0839**	**0.0841**	**0.0871**	0.0979
VCA – GBM	0.0841	0.0846	0.0879	**0.0939**
FCLS – PPNMM	0.1229	0.1230	0.1258	0.1308
FCLS – GBM	0.1614	0.1615	0.1674	0.1696
3×3 **WINDOW**				
VCA – PPNMM	**0.0888**	**0.0885**	**0.0902**	**0.0973**
VCA – GBM	0.0975	0.1040	0.1079	0.1112
FCLS – PPNMM	0.1148	0.1151	0.1197	0.1292
FCLS – GBM	0.1615	0.1617	0.1657	0.1710
4×4 **WINDOW**				
VCA – PPNMM	**0.0904**	**0.0905**	**0.0949**	**0.0994**
VCA – GBM	0.0945	0.0945	0.1061	0.1106
FCLS – PPNMM	0.1154	0.1197	0.1216	0.1245
FCLS – GBM	0.1616	0.1616	0.1636	0.1658

Figure 4.11 shows the corresponding ground truth abundances, and the abundances as estimated by both the linear (VCA), nonlinear (PPMM) as well as the hybrid methods on the Samson data.
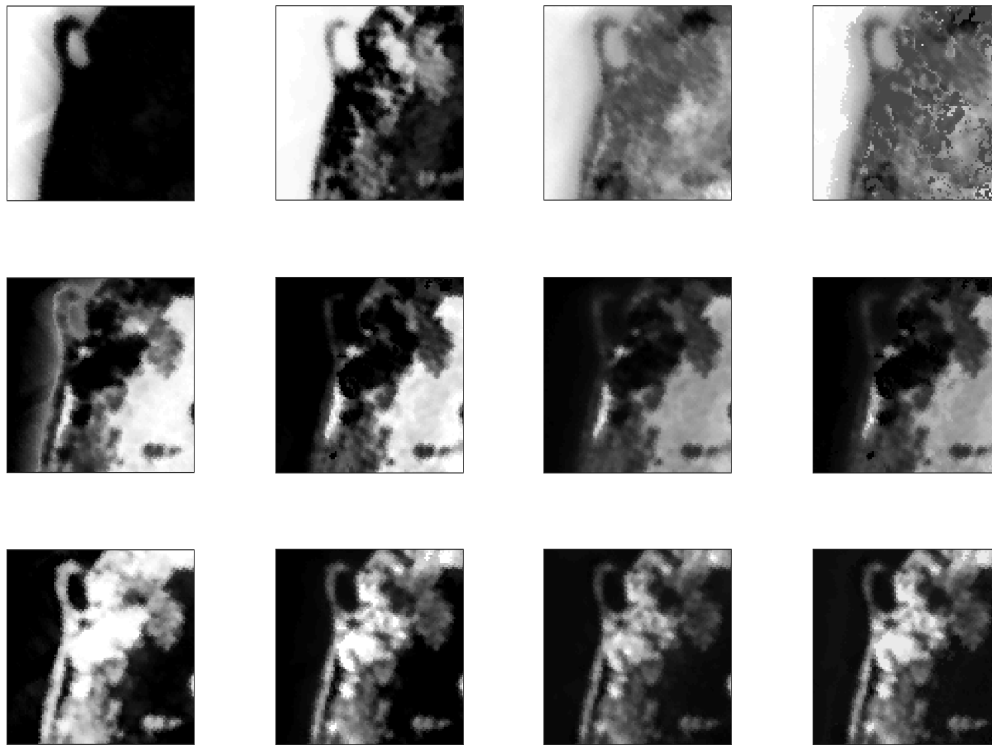
FIGURE 4.11: Abundance estimate of endmembers of the Samson data showing the groundtruth, linear (VCA), nonlinear (PPMM) and the hybrid methods. From top water, rock and tree.

The performance of the proposed hybrid switch method was evaluated using the overall accuracy of the network, the abundance estimation error, the training, validation and testing error of the network. From the results obtained, experiments with the raw dataset, 3×3 and 4×4 windows produce similar overall accuracy in all the experiments. With regards to the four hybrid switch methods, the VCA – PPNMM method outperforms the other hybrid methods with a higher overall accuracy of 96% as compared to FCLS – PPNMM method which has an overall accuracy of 94.5% while

VCA – GBM and FCLS – GBM methods both returned overall accuracies of 92.8%. VCA – PPNMM also has the lowest abundance estimation error and produced the lowest abundance error in terms of training, validation and testing of the neural networks. However, experiments from the proposed hybrid switch method, a similar result was obtained with the 3×3 and 4×4 windows used to conduct the experiment when fewer samples were used to train the networks. Therefore, it shows that the proposed hybrid method also does not requires all the raw data for training the networks and can be used effectively to switch between linear and nonlinear spectral unmixing of hyperspectral data.

The computational cost of the individual and the hybrid methods were evaluated, the individual methods are 40% more time consuming compared to the hybrid method which makes them computationally expensive in terms of simulation. Tables 4.8 and 4.9 summarizes the result of the experiments demonstrating the accuracy of the neural network based on training, testing and validation of the networks.

TABLE 4.8: Abundance estimation error on Jasper Ridge data showing training, validation and testing accuracy on the individual and hybrid methods with different training samples and different window size vectors. The best results are shown in bold

	7,000 samples				3,000 samples			
Raw data	**VCA – PPNMM**	**VCA – GBM**	**FCLS –PPNMM**	**FCLS — GBM**	**VCA – PPNMM**	**VCA – GBM**	**FCLS –PPNMM**	**FCLS — GBM**
TRAIN	**0.0905**	0.1025	0.1184	0.1084	0.0953	**0.0859**	0.1085	0.1200
VALIDATION	**0.0777**	0.0780	0.1008	0.0980	**0.0809**	0.0866	0.1006	0.0995
TEST	**0.0751**	0.0797	0.1012	0.1000	**0.0811**	0.0832	0.1013	0.0906
3×3 **Window**								
TRAIN	**0.0967**	0.1054	0.0981	0.1268	**0.0473**	0.0533	0.0991	0.1229
VALIDATION	0.0524	**0.0505**	0.1274	0.1138	**0.0465**	0.0549	0.0923	0.1125
TEST	**0.0486**	0.0614	0.1276	0.1147	**0.0454**	0.0506	0.0914	0.1135
4×4 **Window**								
TRAIN	**0.0906**	0.1523	0.0941	0.1171	**0.0393**	0.1531	0.0997	0.1146
VALIDATION	0.1696	**0.0704**	0.0911	0.0954	**0.0351**	0.0530	0.0918	0.1117
TEST	0.1608	**0.0382**	0.0938	0.0944	**0.0354**	0.0445	0.0920	0.1121

TABLE 4.9: Abundance estimation error on Samson data showing training, validation and testing accuracy on the individual and hybrid methods with different training samples and different window size vectors. The best results are shown in bold.

	6,317 Samples				3,158 Samples			
Raw data	**VCA – PPNMM**	**VCA – GBM**	**FCLS –PPNMM**	**FCLS — GBM**	**VCA – PPNMM**	**VCA – GBM**	**FCLS –PPNMM**	**FCLS — GBM**
TRAIN	**0.0255**	0.0741	0.1058	0.1585	**0.0280**	0.0732	0.1182	0.1167
VALIDATION	0.0466	0.0101	0.0792	**0.0098**	**0.0553**	0.1026	0.0733	0.1311
TEST	0.0494	0.0105	0.0762	**0.0098**	**0.0569**	0.1053	0.0719	0.1311
3×3 **Window**								
TRAIN	**0.0726**	0.0842	0.1046	0.1581	**0.0722**	0.0897	0.1021	0.1161
VALIDATION	0.0530	0.0100	0.0748	**0.0098**	**0.0588**	0.0692	0.0703	0.1309
TEST	0.0533	0.0100	0.0740	**0.0098**	**0.0594**	0.0696	0.0706	0.1309
4×4 **Window**								
TRAIN	**0.0836**	0.0842	0.1046	0.1581	**0.0822**	0.0843	0.1007	0.1160
VALIDATION	0.0536	0.0100	0.0748	**0.0098**	**0.0569**	0.0654	0.0710	0.1308
TEST	0.0545	0.0100	0.0740	**0.0098**	**0.0569**	0.0640	0.0712	0.1308

4.4 Hydrocarbon Abundance Estimation

In this section, the hybrid spectral unmixig method and the deep learning model were validated to estimate the amount of Hydrocarbon (HC) mixed with different soil samples. Hyperspectral Imaging utilizes information obtained from surface reflectance in the electromagnetic spectrum using different sensors [32], and the resultant information is used for spectral and spatial analysis. Spectral analysis involves evaluation of energy reflected and absorbed at different wavelengths.

4.4.1 Hydrocarbon abundance estimation using the hybrid switch method

The hybrid spectral unmixing method for switching between linear and nonlinear models described in chapter 3 was used to quantify HC spills using the controlled dataset. The aim here is to robustly choose the most suitable method between the linear and nonlinear spectral unmixing methods to

quantifies the amount of HC content in each sample. The experiments were conducted with 70% and 30% of training samples in order to check if the size of the training data affects the accuracy of the network. The remaining samples were split for testing and validation. The results of the experiments were based on the abundance estimation error of the individual and hybrid methods as summarized in Tables 4.10 - 4.13.

TABLE 4.10: Average abundance estimation error of the hybrid switch methods with different numbers of training samples (70% and 30%) as compared with the abundance estimation error of the individual methods with clay mixture.

(Clay - Biodiesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0217	0.0217
GBM	0.0082	0.0082
VCA	0.0955	0.0955
FCLS	0.0083	0.0083
Hybrid methods		
VCA – PPNMM	0.0073	0.0079
VCA – GBM	0.0104	0.0103
FCLS – PPNMM	0.0097	0.0097
FCLS – GBM	0.0082	0.0082
(Clay - Diesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0045	0.0045
GBM	0.3775	0.3775
VCA	0.0845	0.0845
FCLS	0.3776	0.3776
Hybrid methods		
VCA – PPNMM	0.0041	0.0043
VCA – GBM	0.0634	0.0640
FCLS – PPNMM	0.0048	0.0053
FCLS – GBM	0.3770	0.3776
(Clay - Ethanol)	70% SAMPLES	30% SAMPLES
PPNMM	0.0036	0.0036
GBM	0.7046	0.7046
VCA	0.7499	0.7499
FCLS	0.7048	0.7048
Hybrid methods		
VCA – PPNMM	0.0025	0.0030
VCA – GBM	0.2516	0.2513
FCLS – PPNMM	0.0047	0.0047
FCLS – GBM	0.0056	0.0060

TABLE 4.11: Average abundance estimation error of the hybrid switch methods with different numbers of training samples (70% and 30%) as compared with the abundance estimation error of the individual methods with clayloam mixture.

(Clay Loam- Biodiesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0084	0.0084
GBM	0.0824	0.0824
VCA	0.0766	0.0766
FCLS	0.0842	0.0842
Hybrid methods		
VCA – PPNMM	0.0068	0.0065
VCA – GBM	0.0501	0.0438
FCLS – PPNMM	0.0084	0.0086
FCLS – GBM	0.0824	0.0824
(Clay Loam - Diesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0025	0.0025
GBM	0.0824	0.0824
VCA	0.0245	0.0245
FCLS	0.0828	0.0828
Hybrid methods		
VCA – PPNMM	0.0020	0.0022
VCA – GBM	0.0267	0.0268
FCLS – PPNMM	0.0027	0.0027
FCLS – GBM	0.0824	0.0830
(Clay Loam - Ethanol)	70% SAMPLES	30% SAMPLES
PPNMM	0.0546	0.0546
GBM	0.0825	0.0825
VCA	0.0085	0.0085
FCLS	0.0852	0.0852
Hybrid methods		
VCA – PPNMM	0.0080	0.0082
VCA – GBM	0.0085	0.0085
FCLS – PPNMM	0.0540	0.0541
FCLS – GBM	0.0825	0.0823
(Clay Loam - Petrol)	70% SAMPLES	30% SAMPLES
PPNMM	0.0084	0.0084
GBM	0.8242	0.8242
VCA	0.1560	0.1560
FCLS	0.8240	0.8240
Hybrid methods		
VCA – PPNMM	0.0050	0.0051
VCA – GBM	0.1524	0.1521
FCLS – PPNMM	0.0080	0.0084
FCLS – GBM	0.0824	0.0826

TABLE 4.12: Average abundance estimation error of the hybrid switch methods with different numbers of training samples (70% and 30%) as compared with the abundance estimation error of the individual methods with sandyloam mixture.

(Sandy Loam - Biodiesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0025	0.0025
GBM	0.0083	0.0083
VCA	0.0205	0.0205
FCLS	0.0080	0.0080
Hybrid methods		
VCA – PPNMM	0.0016	0.0012
VCA – GBM	0.0057	0.0055
FCLS – PPNMM	0.0077	0.0077
FCLS – GBM	0.0082	0.0082
(Sandy Loam - Diesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0036	0.0036
GBM	0.8244	0.8244
VCA	0.0097	0.0097
FCLS	0.0821	0.0821
Hybrid methods		
VCA – PPNMM	0.0016	0.0017
VCA – GBM	0.0073	0.0073
FCLS – PPNMM	0.0031	0.0031
FCLS – GBM	0.0824	0.0824
(Sandy Loam - Ethanol)	70% SAMPLES	30% SAMPLES
PPNMM	0.3223	0.3223
GBM	0.4018	0.4018
VCA	0.1119	0.1119
FCLS	0.4067	0.4067
Hybrid methods		
VCA – PPNMM	0.0921	0.0922
VCA – GBM	0.0895	0.0897
FCLS – PPNMM	0.3185	0.3188
FCLS – GBM	0.4001	0.4001
(Sandy Loam - Petrol)	70% SAMPLES	30% SAMPLES
PPNMM	0.1886	0.1886
GBM	0.3646	0.3646
VCA	0.1965	0.1965
FCLS	0.3640	0.3640
Hybrid methods		
VCA – PPNMM	0.0560	0.0547
VCA – GBM	0.0722	0.0777
FCLS – PPNMM	0.1870	0.1870
FCLS – GBM	0.0363	0.0363

TABLE 4.13: Average abundance estimation error of the hybrid switch methods with different numbers of training samples (70% and 30%) as compared with the abundance estimation error of the individual methods with sandyclayloam mixture.

(Sandy Clay Loam - Biodiesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0118	0.0118
GBM	0.0089	0.0089
VCA	0.1773	0.1773
FCLS	0.0087	0.0087
Hybrid methods		
VCA – PPNMM	0.0117	0.0117
VCA – GBM	0.0080	0.0083
FCLS – PPNMM	0.0079	0.0079
FCLS – GBM	0.0080	0.0080
(Sandy Clay Loam - Diesel)	70% SAMPLES	30% SAMPLES
PPNMM	0.0054	0.0054
GBM	0.0822	0.0822
VCA	0.0178	0.0178
FCLS	0.0823	0.0823
Hybrid methods		
VCA – PPNMM	0.0038	0.0030
VCA – GBM	0.0116	0.0116
FCLS – PPNMM	0.0054	0.0054
FCLS – GBM	0.0822	0.0820
(Sandy Clay Loam - Ethanol)	70% SAMPLES	30% SAMPLES
PPNMM	0.3216	0.3216
GBM	0.4978	0.4978
VCA	0.2326	0.2326
FCLS	0.4973	0.4973
Hybrid methods		
VCA – PPNMM	0.1234	0.1235
VCA – GBM	0.1449	0.1449
FCLS – PPNMM	0.3219	0.3220
FCLS – GBM	0.4972	0.4973
(Sandy Clay Loam - Petrol)	70% SAMPLES	30% SAMPLES
PPNMM	0.2749	0.2749
GBM	0.4503	0.4503
VCA	0.1357	0.1357
FCLS	0.4490	0.4490
Hybrid methods		
VCA – PPNMM	0.0708	0.0708
VCA – GBM	0.0803	0.0803
FCLS – PPNMM	0.2300	0.2300
FCLS – GBM	0.4470	0.4474

4.4.2 Hydrocarbon abundance estimation using DNN

The deep learning based spectral unmixing approach was applied to estimate the amount of HC in each mixture.

The reflectance spectra of different soil samples with 15% hydrocarbon concentration mixture are shown in Figure 4.12 showing specific absorption at around 1700 μm and 2300 μm, respectively
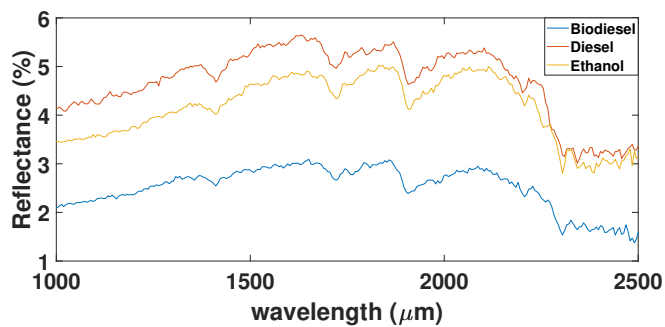
(A) Clay with different hydrocarbon mixtures
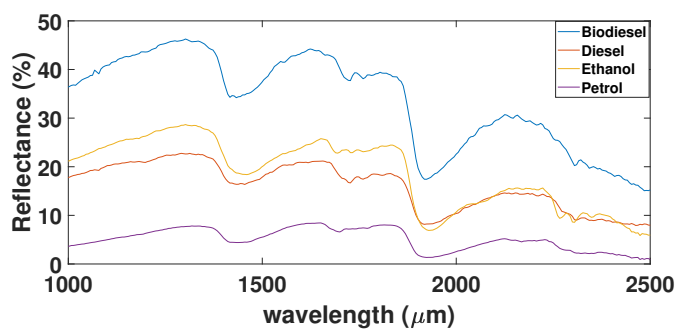
(B) Clay loam with different hydrocarbon mixtures
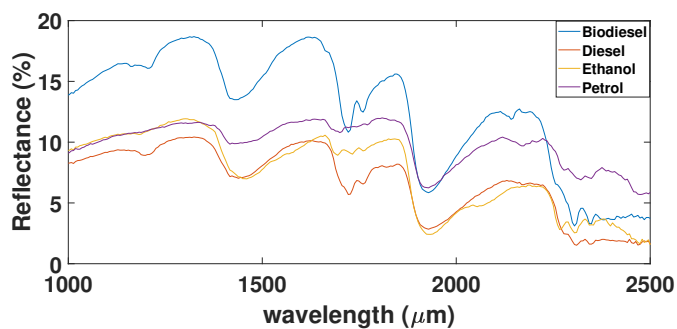
(C) Sandy loam with different hydrocarbon mixtures
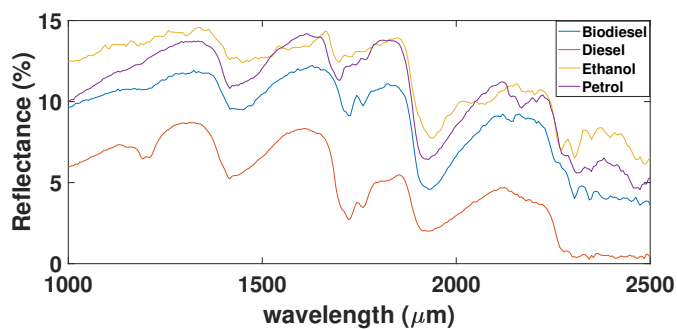
(D) Sandy clay loam with different hydrocarbon mixtures

FIGURE 4.12: Spectral reflectance of different soils and 15% hydrocarbon concentration mixtures.

The experiments were carried out with the purpose of obtaining optimum hyperparameters so as to achieve maximum accuracy in estimating the amount of HCs in each soil type based on the control experiment. The ground truth or known abundances from the sample preparation were used as class labels (targets) to train the network for the abundance estimation. These abundances were estimated based on the ground truth of the different HC mixtures corresponding to the different percentages of HC contents added in every mixture as detailed in the experimental protocol section 4.2.4 in each dataset as detailed in Table 4.14.

TABLE 4.14: Ground truth abundances for the different mixtures

Corresponding mixtures (ml)	Petrol	Diesel	Biodiesel	Ethanol
0.05	0.068	0.080	0.080	0.073
0.10	0.128	0.148	0.149	0.136
0.15	0.181	0.206	0.208	0.191
0.20	0.227	0.258	0.260	0.240
0.25	0.269	0.303	0.305	0.283
0.30	0.340	0.342	0.345	0.321

The ground truth abundances for the different HC type were calculated based on the density type of each hydrocarbon multiplied by the amount of the HC in the mixture (ml) divided by the amount of soil (50g) plus the density of the HC multiplied by the amount of HC in the mixture (ml). The densities of the different Hcs are; $0.73g/cm^3$ for petroleum, $0.79g/cm^3$ for ethanol, $0.88g/cm^3$ for biodiesel and $0.85g/cm^3$ for diesel.

Cross validation was used to assess the overall accuracy of the HC quantification. Cross validation estimates the overall accuracy of correct classes

by the number of instances in the datasets. Part of the data (training samples) were used to train the network and the remaining (validation samples) were used to validate the accuracy of the network. The cross validation algorithm also avoids overfitting because the training sample is independent of the validation sample [194]. The size of the data sets depended on the soils' absorption level during the experiment (i.e. when a local shallow pool was formed). Only image pixels corresponding to data from inside the petri dish were considered. Moreover, for each scanned image, 1000 pixels were randomly selected. Thus the data sets ranged between $5000 pixels \times 288$ bands (where 5 mixture types were available) to $10000 pixels \times 288$ bands (for samples with 10 possible mixtures). The size of the data sets and number of mixtures used for the experiments as shown in (Table 4.1) are summarized in Table 4.15. Subsets of the hyperspectral data were fed into the network as follows: 80% of the data were randomly selected for training the network and 20% were used for cross validation.

TABLE 4.15: Size of datasets and target class

Dataset	Size	Number of mixtures
Clay biodiesel	6000×288	6
Clay diesel	6000×288	6
Clay ethanol	6000×288	6
Clay loam biodiesel	5000×288	5
Clay loam diesel	5000×288	5
Clay loam ethanol	6000×288	6
Clay loam petrol	10000×288	10
Sandy loam biodiesel	5000×288	5
Sandy loam diesel	5000×288	5
Sandy loam ethanol	5000×288	5
Sandy loam petrol	10000×288	10
Sandy clay loam biodiesel	6000×288	6
Sandy clay loam diesel	6000×288	6
Sandy clay loam ethanol	7000×288	7
Sandy clay loam petrol	8000×288	8

To evaluate the effectiveness of the three- term backpropagation algorithm with dropout in the proposed deep learning model, the experimental process was repeated with different dropout ratio on the hidden layers of 10%, 20%, 30%, 40% and 50% respectively. The results demonstrate both training and validation accuracy of the network. Tables 4.16 - 4.19 illustrate the mean square error of the proposed method with the different dropout ratios. It is noted that in all cases the error is 10 times lower for dropout ratio of 40% than for 50%. Then again, when the error drops significantly for dropout ratio 20%. However, when it is further reduced to 10%, the error increases. The 20% dropout is adopted subsequently in the rest of the experiments.

TABLE 4.16: Mean Square Error (MSE) of the deep learning model for clayloam datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio - diesel					
MSE	3.5×10^{-4}	0.22×10^{-4}	0.69×10^{-4}	6.5×10^{-4}	7.2×10^{-4}
Diesel					
MSE	0.31×10^{-4}	0.25×10^{-4}	0.52×10^{-4}	7.3×10^{-4}	5.9×10^{-4}
Ethanol					
MSE	3.1×10^{-4}	0.21×10^{-4}	5.6×10^{-4}	7.4×10^{-4}	7.6×10^{-4}
Petrol					
MSE	2.9×10^{-4}	0.22×10^{-4}	0.73×10^{-4}	6.7×10^{-4}	7.2×10^{-4}

TABLE 4.17: Mean square error of the deep learning model for clay datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio - diesel					
MSE	2.9×10^{-4}	0.25×10^{-4}	2.4×10^{-4}	4.4×10^{-4}	8.3×10^{-4}
Diesel					
MSE	1.7×10^{-4}	0.35×10^{-4}	1.7×10^{-4}	3.2×10^{-4}	7.6×10^{-4}
Ethanol					
MSE	0.20×10^{-4}	0.18×10^{-4}	0.25×10^{-4}	3.3×10^{-4}	9.9×10^{-4}

TABLE 4.18: Mean square error of the deep learning model for sandyclayloam datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio - diesel					
MSE	2.6×10^{-4}	0.21×10^{-4}	0.63×10^{-4}	2.1×10^{-4}	6.6×10^{-4}
Diesel					
MSE	2.1×10^{-4}	0.02×10^{-4}	4.1×10^{-4}	7.2×10^{-4}	7.3×10^{-4}
Ethanol					
MSE	3.1×10^{-4}	0.26×10^{-4}	4.2×10^{-4}	6.3×10^{-4}	7.4×10^{-4}
Petrol					
MSE	2.7×10^{-4}	0.27×10^{-4}	0.51×10^{-4}	3.6×10^{-4}	7.7×10^{-4}

TABLE 4.19: Mean square error of the deep learning model for sandyloam datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio - diesel					
MSE	3.2×10^{-4}	0.24×10^{-4}	0.47×10^{-4}	7.6×10^{-4}	6.4×10^{-4}
Diesel					
MSE	2.0×10^{-4}	0.23×10^{-4}	0.46×10^{-4}	4.1×10^{-4}	7.0×10^{-4}
Ethanol					
MSE	2.0×10^{-4}	0.2×10^{-4}	9.5×10^{-4}	6.5×10^{-4}	9.4×10^{-4}
Petrol					
MSE	0.35×10^{-4}	0.22×10^{-4}	0.42×10^{-4}	4.7×10^{-4}	6.4×10^{-4}

To demonstrate the accuracy of the proposed method, results of the experiment showing the root mean square error and the networks ability to converge rapidly with low number of epochs are shown in Figures 4.13 - 4.16 for individual soil types contaminated with different HCs. The plots in Figures 4.17 - 4.20 shows the training and corresponding output target of the network,
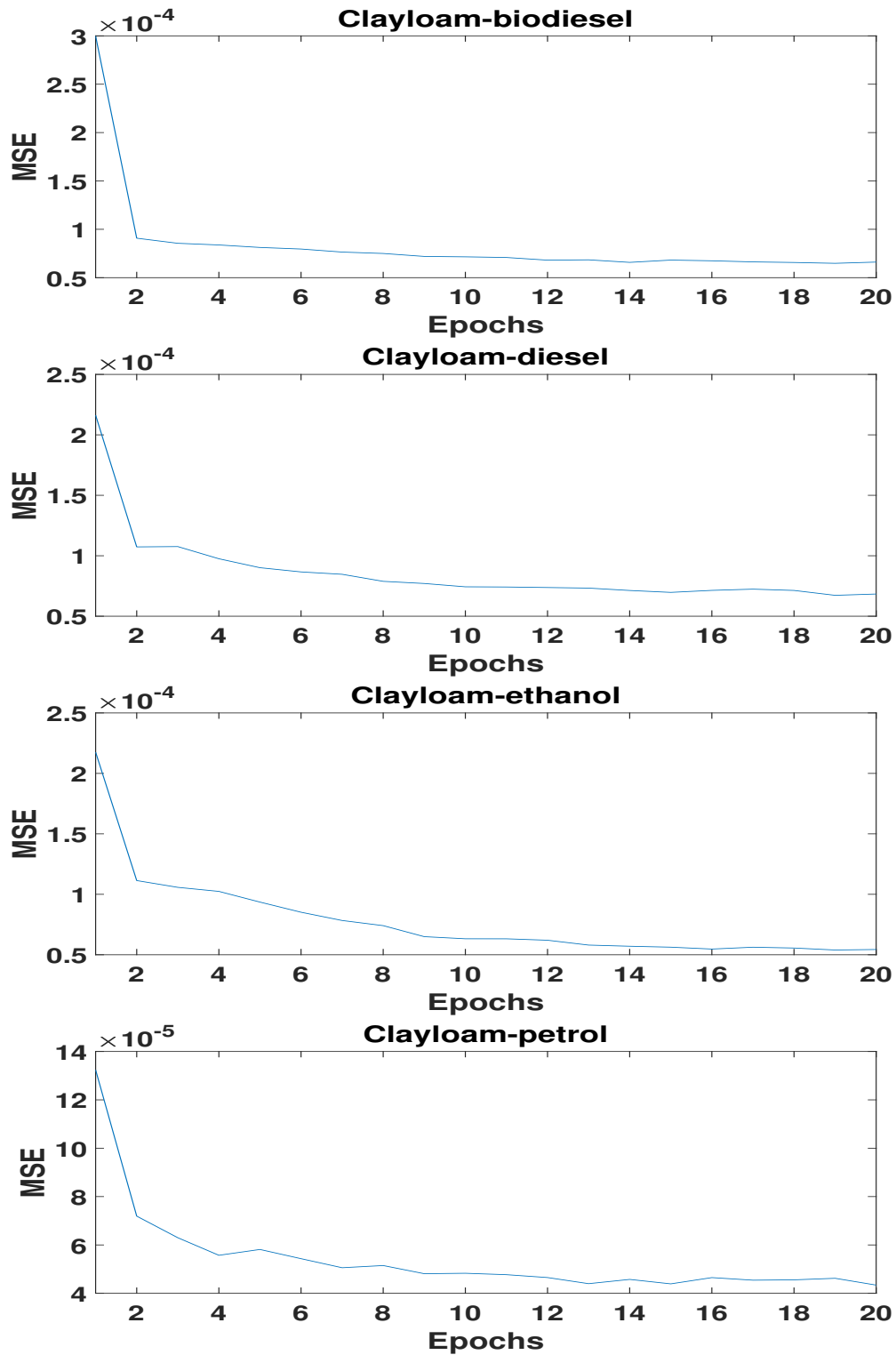
FIGURE 4.13: Mean square error of Clayloam mixture with different HC concentration
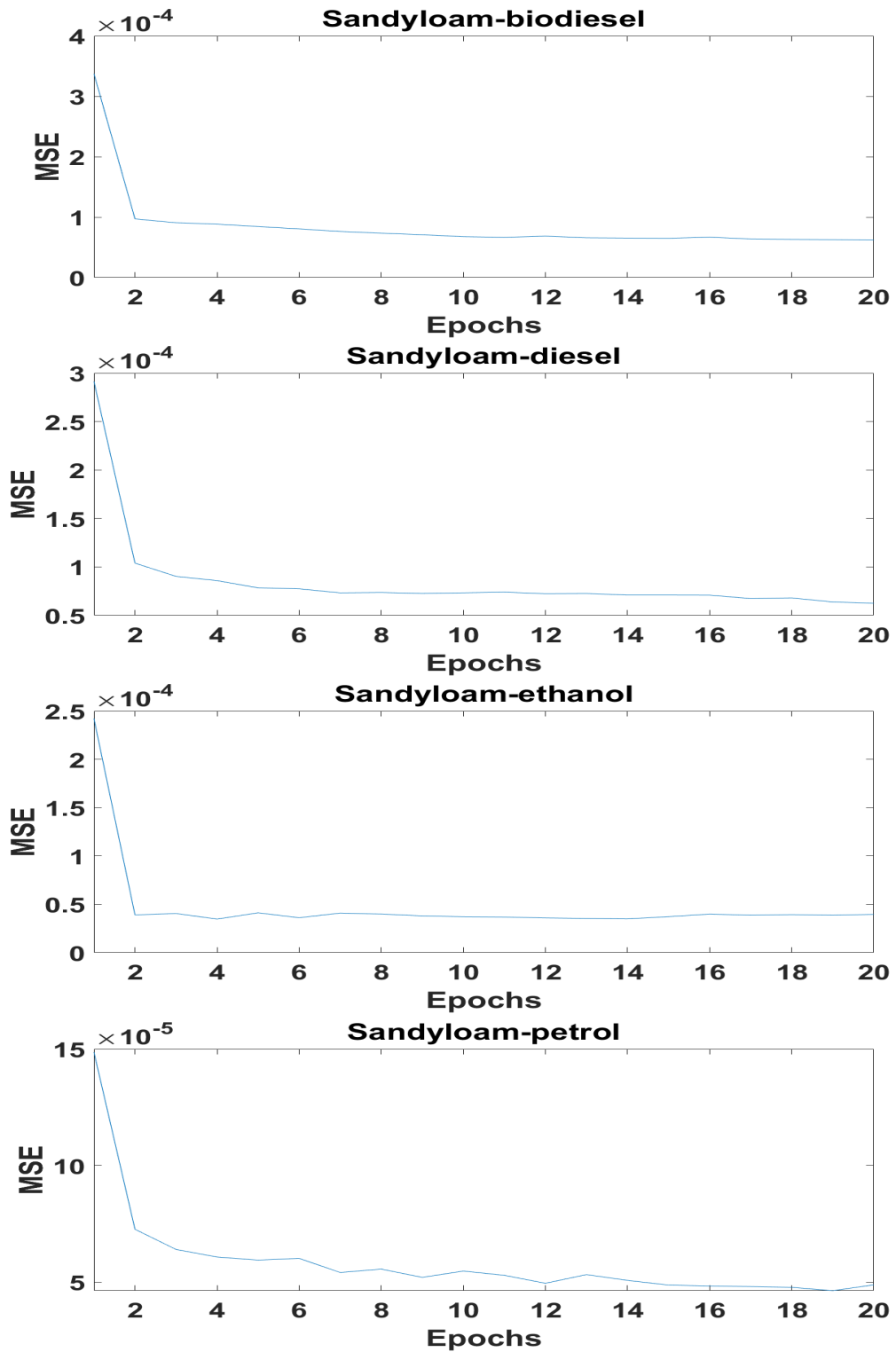
FIGURE 4.14: Mean square error of Sandyloam mixture with different HC concentration
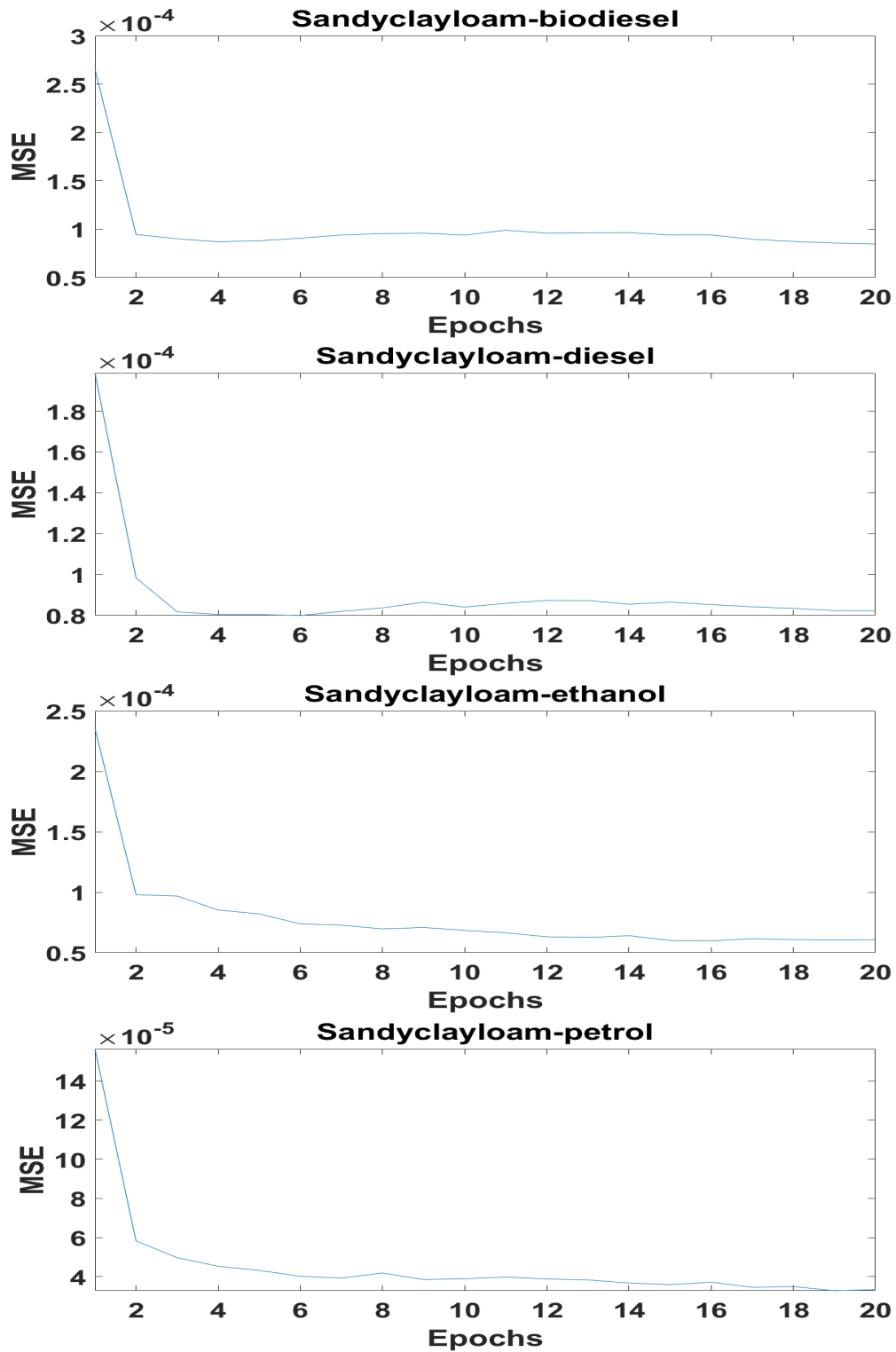
FIGURE 4.15: Mean square error of Sandyclayloam mixture with different HC concentration
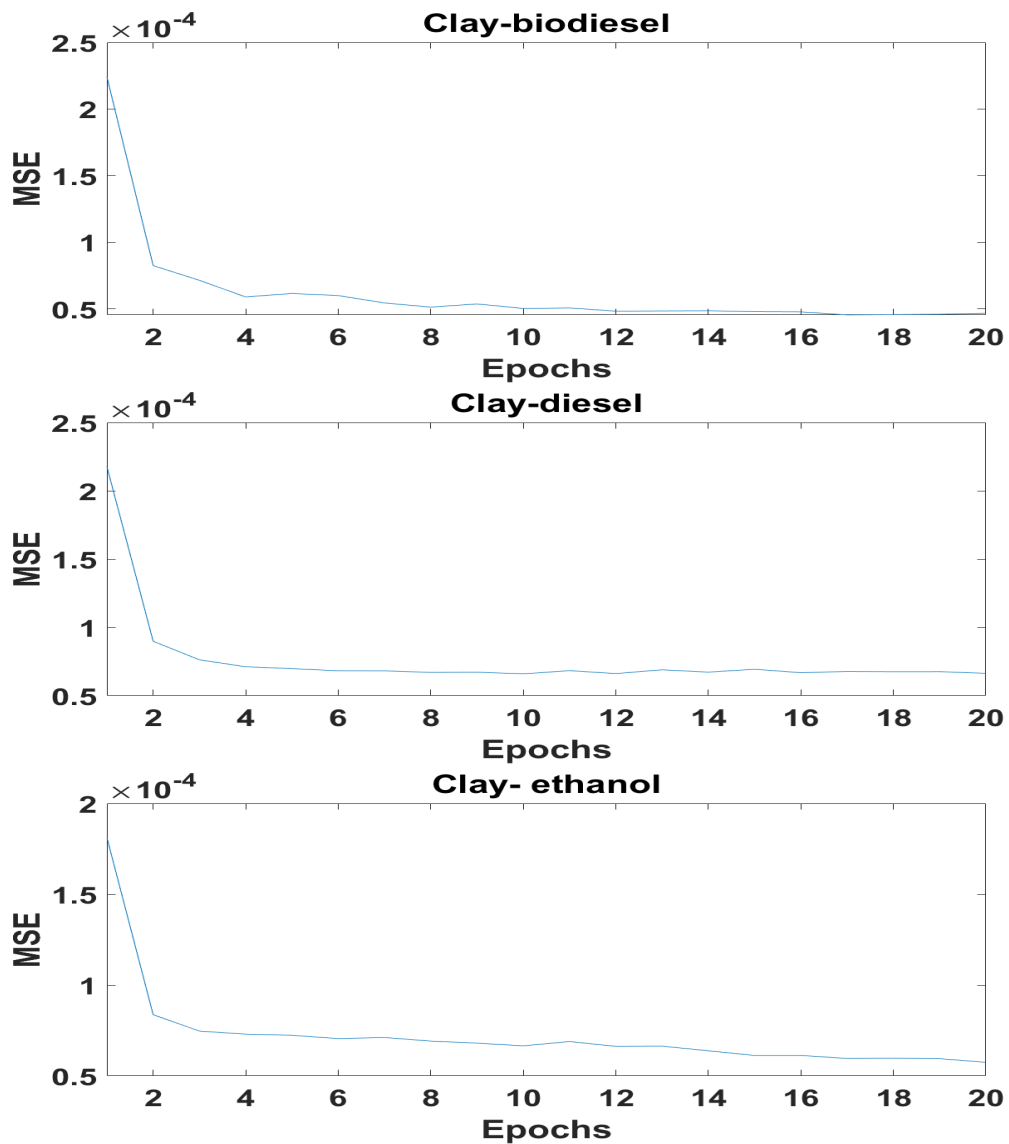
FIGURE 4.16: Mean square error of Clay mixture with different
HC concentration
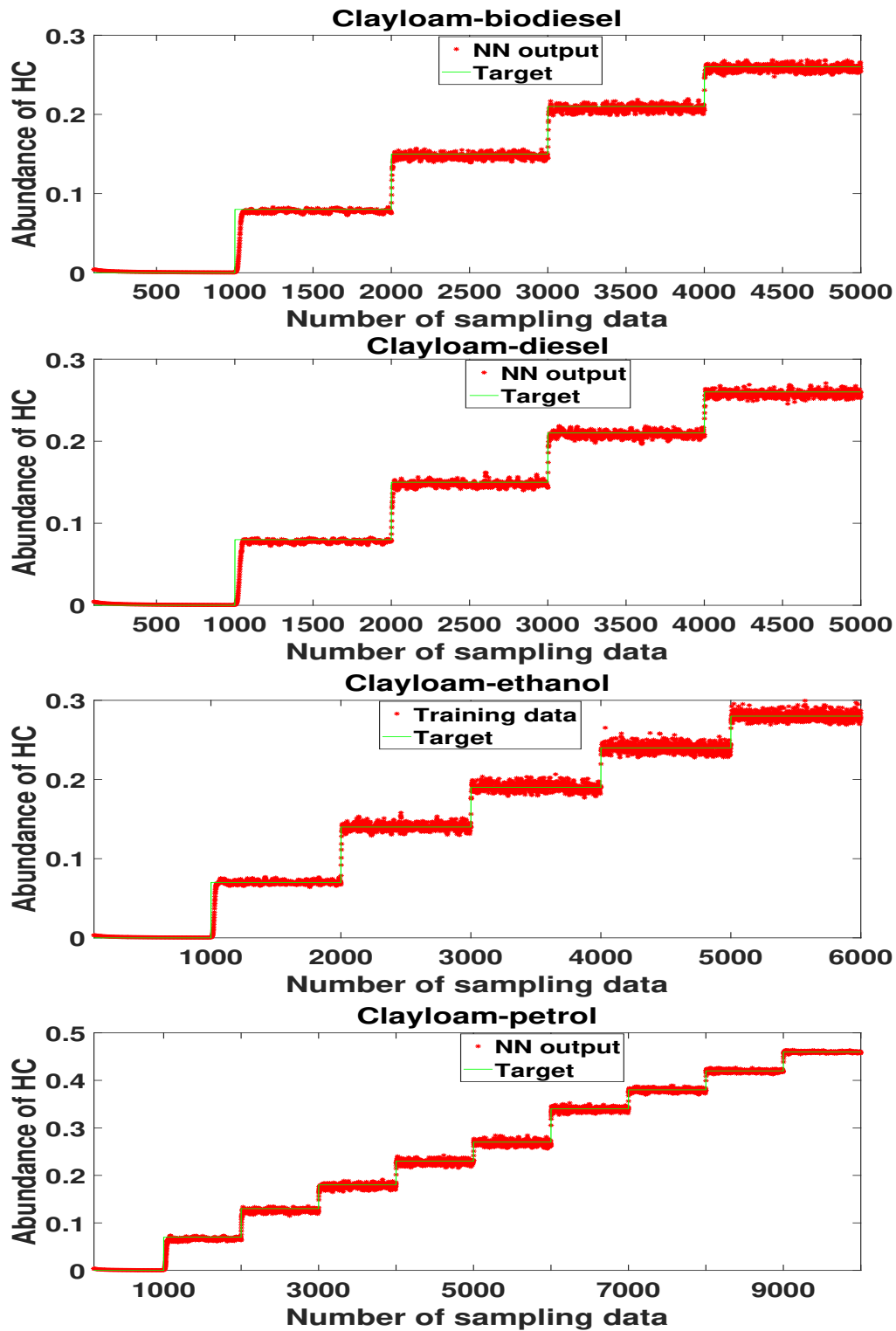
FIGURE 4.17: Training and output targets of Clayloam mixture
with different HC concentration
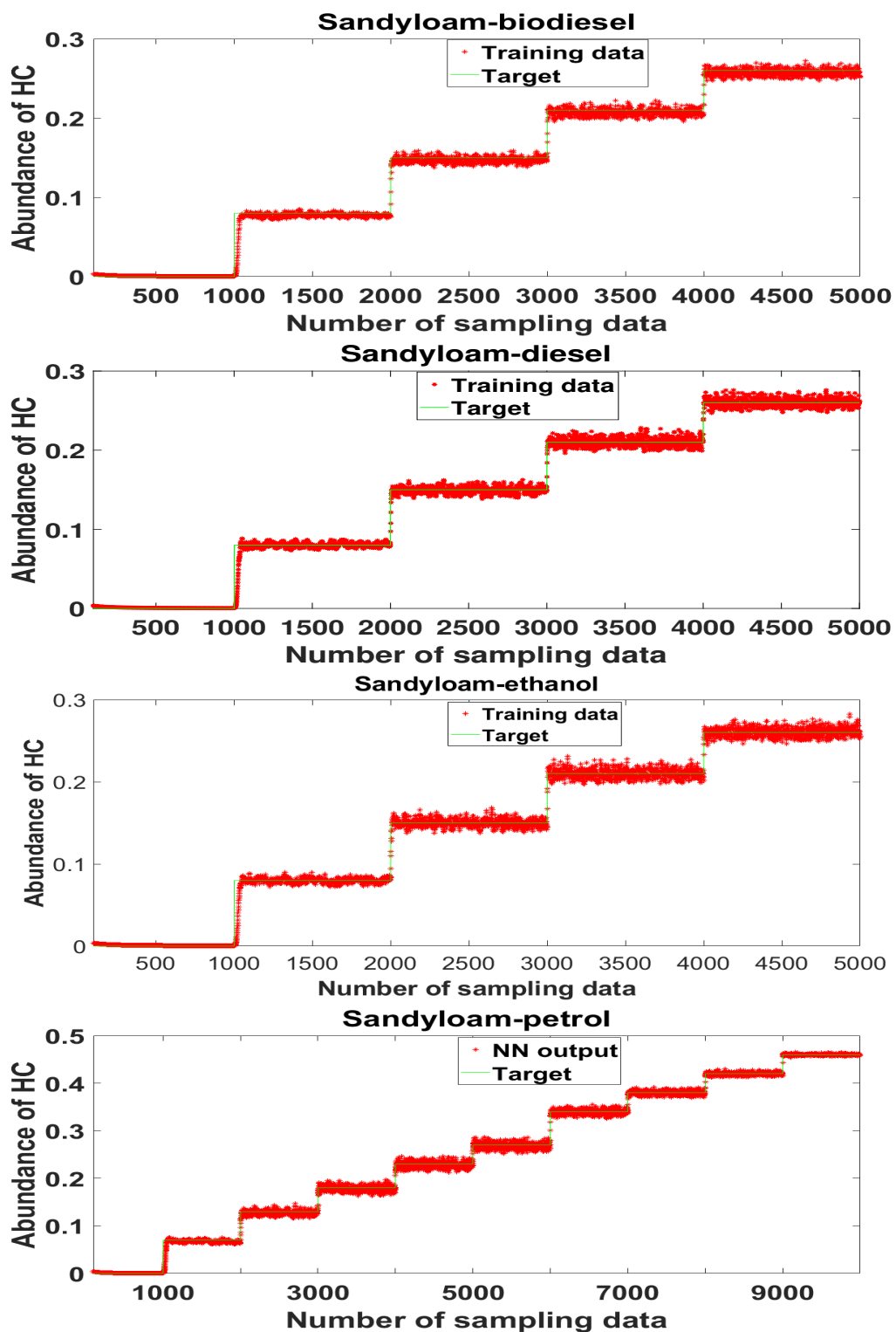
FIGURE 4.18: Training and output targets of Sandyloam mix-
ture with different HC concentration
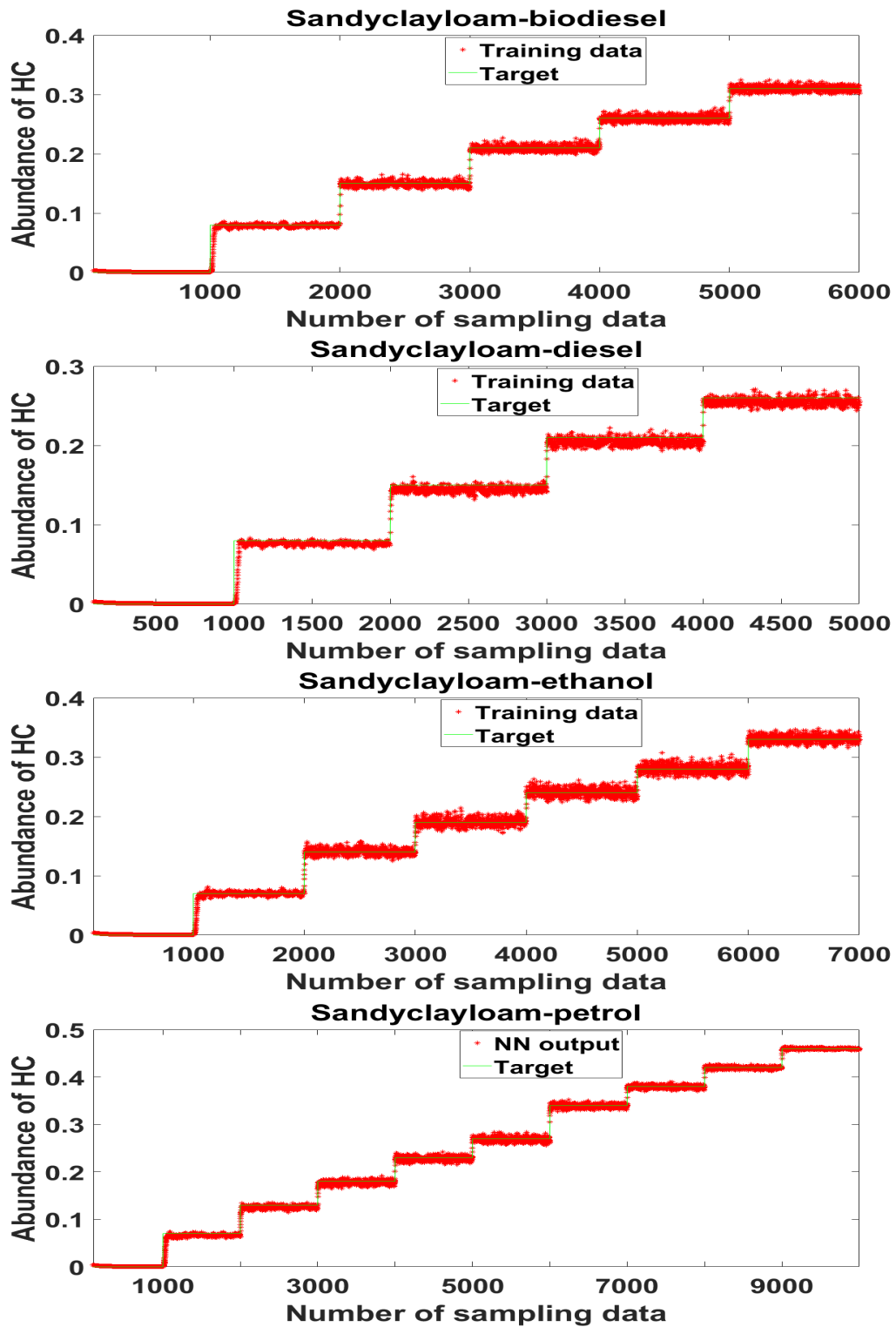
FIGURE 4.19: Training and output targets of Sandyclayloam mixture with different HC concentration
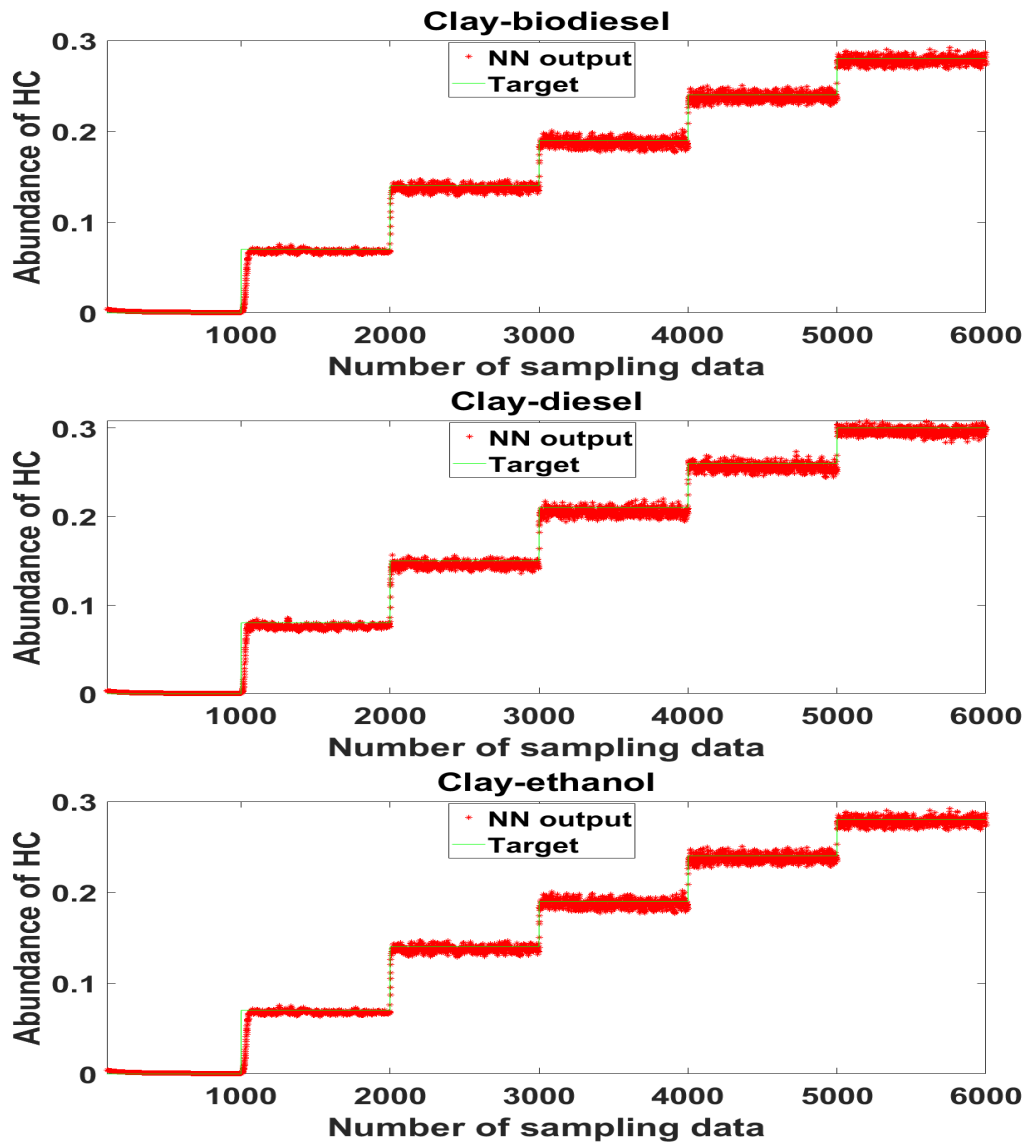
FIGURE 4.20: Training and output targets of Clay mixture with
different HC concentration

From the results obtained, it is noted that the proposed method trained with three-term backpropagation algorithm with dropout was able to generalize on unseen data with high prediction accuracy. We observed a similar trend on all the datasets used for the experiment which indicates a reduction in the error rate and high convergence rate.

To demonstrate the ability of the proposed deep learning model to generalize on unseen data, Table 4.20 displays the results obtained from the test sets with and without dropout, respectively.

TABLE 4.20: Mean square error of the deep learning model on unseen data with and without dropout, respectively.

Dataset	Test Set with Dropout	Test Set without Dropout
Clay biodiesel	7.11×10^{-3}	9.1×10^{-3}
Clay diesel	1.16×10^{-3}	6.9×10^{-3}
Clay ethanol	8.26×10^{-4}	6.2×10^{-3}
Clay loam biodiesel	7.62×10^{-4}	1.4×10^{-3}
Clay loam diesel	2.20×10^{-3}	3.3×10^{-3}
Clay loam ethanol	8.80×10^{-4}	7.1×10^{-3}
Clay loam petrol	2.50×10^{-3}	8.1×10^{-3}
Sandy loam biodiesel	8.00×10^{-4}	1.2×10^{-3}
Sandy loam diesel	1.20×10^{-3}	9.8×10^{-3}
Sandy loam ethanol	4.01×10^{-3}	8.4×10^{-3}
Sandy loam petrol	6.71×10^{-3}	9.3×10^{-3}
Sandy clay loam biodiesel	8.25×10^{-4}	3.5×10^{-3}
Sandy clay loam diesel	8.84×10^{-4}	9.8×10^{-3}
Sandy clay loam ethanol	2.10×10^{-3}	7.5×10^{-3}
Sandy clay loam petrol	5.13×10^{-3}	6.1×10^{-3}

Soil Continuity Experiment

In this research, four different mixtures of soil were created and HC were added in discrete steps. However, in real life situations, both HC and soil levels of given samples are continuous rather than discrete. Therefore, in order to simulate a more realistic scenario, several strategies were explored. The first strategy was to create a generic model with all soils combined as opposed to separate models for each soil as in the previous experiments. It is noted that the soils were prepared and mixed manually and contained grains of different size (e.g., clay and sand mixture). By feeding the DL network with all types of soils, differences in the soil composition would appear from pixel to pixel. DLs were created including all four different soil mixtures (Clay, Clay–loam, Sandy loam, and Sandy clay–loam) rather than individually. Using the same architecture of the deep learning model, 80% of the resultant data was used to train the model, 10% was used as test sample and the remaining 10% was used for cross-validation. Table 4.21 summarizes the results obtained for biodiesel. Average mean square error for the individual models are shown in brackets for comparison purposes.

In order to simulate a more realistic scenario, and following a similar approach presented in the work by the authors of [195], noise was added to the data to simulate continuous spectra values instead of discrete and also to evaluate the noise rejection of the models. Here, the datasets were corrupted with Random Gaussian noise with signal-to-noise ratio (SNR) ranging from 10 to 40 dB.

TABLE 4.21: Soil continuity experiments. Mean square error of the bio-diesel deep learning model using generic models and individual models with added noise. Training and testing results are shown.

Dataset	Training Data	Test Data
Biodiesel with generic model	7.2238×10^{-4} (1.525×10^{-4})	6.8×10^{-3} (2.374×10^{-3})
Biodiesel with added noise		
SNR (dB)	**Training data**	**Test data**
40	8.2117×10^{-4}	8.9821×10^{-4}
30	8.2594×10^{-4}	8.9333×10^{-4}
20	9.393×10^{-4}	0.001
10	9.671×10^{-4}	0.001
Biodiesel with added noise on testing data		
SNR (dB)	Training data	Test data
40	6.9997×10^{-4}	9.3811×10^{-4}
30	6.9997×10^{-4}	9.0321×10^{-4}
20	6.9997×10^{-4}	9.0657×10^{-4}
10	6.9997×10^{-4}	0.0012

4.4.3 Comparison to conventional backpropagation algorithm

To demonstrate the effectiveness of the proposed method, a deep learning model trained with conventional backpropagation algorithm was similarly used to quantify the HC abundances; first without dropout and then with 20% dropout to train the networks. For fair comparison, the same network structure was used and these includes; number of layers, number of nodes for each layer, range of initial values and learning rate. The hybrid spectral unmixing method for switching between linear and nonlinear methods as discussed in Chapter 3 was similarly used to estimate the amount of HC in the datasets.

The abundance estimate for each mixture type was estimated as shown in

Tables 4.22 and 4.23 respectively. It shows that our proposed method outperforms the deep learning model trained with conventional backpropagation algorithm with the closest estimation from the reference mixture types, followed by the hybrid methods. It can be concluded that the two methods were able to estimate the abundance of hydrocarbon spills with minimum error compared to the deep neural networks trained with conventional backpropagation algorithm.

TABLE 4.22: Estimated hydrocarbon abundance predicted by the proposed deep learning method, compared with the hybrid switch method, deep neural network trained with conventional backpropagation algorithm with and without dropout (DO) for Clay and Clay Loam (CL) mixtures. A comparative summary showing the average estimation error in percentage is included for each mixture type.

Mixtures	Reference (%)	Proposed method (%)	Hybrid switch method (%)	Conventionally trained NN with (DO 0.20) (%)	Conventionally trained NN (%)
Clay- biodiesel	0	0.004	0.03	0.57	0.75
Clay- biodiesel	8	8.5	9.2	9.7	9.9
Clay- biodiesel	14.9	15.0	16.7	17.3	17.8
Clay- biodiesel	20.8	21.3	22.0	16.5	15.9
Clay- biodiesel	26.0	26.2	27.7	29.5	28.9
Clay- biodiesel	30.5	31.3	33.0	25.7	23.8
Average error (%)		2	10	17	20
Clay- diesel	0	0.004	0.07	1.02	1.55
Clay- diesel	8	8.3	7.2	5.8	4.9
Clay- diesel	14.8	15.3	12.1	18.4	18.9
Clay- diesel	20.6	21.1	17.3	24.0	24.7
Clay- diesel	25.8	25.6	22.3	29.4	29.9
Clay- diesel	30.3	30.6	34.2	35.4	35.6
Average error (%)		2	14	20	24
Clay- ethanol	0	0.004	0.09	1.01	2.00
Clay- ethanol	7.3	7.7	8.7	5.3	4.7
Clay- ethanol	13.6	14.0	15.2	10.1	9.6
Clay- ethanol	19.1	19.5	18.4	22.1	22.7
Clay- ethanol	24	24.8	22.5	27.3	27.7
Clay- ethanol	28.3	28.4	27.1	32.3	32.9
Average error (%)		3	9	19	23
CL- biodiesel	0	0.004	0.10	0.79	1.99
CL- biodiesel	8	8.3	7.1	9.8	10.3
CL- biodiesel	14.9	15.2	12.6	17.3	17.7
CL- biodiesel	20.8	21.3	18.2	24.1	25.4
CL- biodiesel	26.0	26.2	25.1	29.2	29.8
Average error (%)		2	11	17	21
CL- diesel	0	0.004	0.001	0.99	1.75
CL- diesel	8	8.2	10.0	11.2	11.9
CL- diesel	14.8	14.2	12.8	18.3	18.9
CL- diesel	20.6	21.3	22.6	25.7	26.8
CL- diesel	25.8	26.4	27.2	29.5	30.1
Average error (%)		3	13	26	31
CL- ethanol	0	0.004	0.07	1.57	1.92
CL- ethanol	7.3	7.3	6.3	9.3	9.9
CL- ethanol	13.6	14.2	11.0	16.1	16.9
CL- ethanol	19.1	19.3	17.4	22.0	22.7
CL- ethanol	24	24.5	25.7	27.7	27.9
CL- ethanol	28.3	28.5	29.8	31.2	31.8
Average error (%)		2	11	17	21
CL- petrol	0	0.003	0.007	0.56	1.95
CL- petrol	6.8	6.8	7.6	8.3	8.7
CL- petrol	12.8	12.6	13.3	10.1	9.3
CL- petrol	18.1	18.5	17.1	16.2	15.8
CL- petrol	22.7	23.4	24.0	19.3	18.1
CL- petrol	26.9	27.0	27.8	24.0	23.3 CL- petrol
34	34.6	35.9	29.7	29.2	
CL- petrol	38.1	38.4	39.2	36.4	36.1
CL- petrol	42	42.5	44.1	38.0	37.4
CL- petrol	46.2	46.6	45.0	43.3	42.7
Average error (%)		1	5	13	16

TABLE 4.23: Estimated hydrocarbon abundance predicted by the proposed deep learning method, compared with the hybrid switch method, deep neural network trained with conventional backpropagation algorithm with and without dropout (DO) for Sandy Loam (SL) and Sandy Clay Loam (SCL) mixtures. A comparative summary showing the average estimation error in percentage is included for each mixture type.

Mixtures	Reference (%)	Proposed method (%)	Hybrid switch method (%)	Conventionally trained NN with (DO 0.20) (%)	Conventionally trained NN (%)
SL- biodiesel	0	0.004	0.03	1.76	2.57
SL- biodiesel	8	8.2	8.9	10.3	10.9
SL- biodiesel	14.9	15.2	13.1	12.3	11.7
SL- biodiesel	20.8	21.1	22.3	23.1	23.9
SL- biodiesel	26	26.4	25.1	24.0	23.4
SL- biodiesel	30.5	31.5	32.3	35.4	36.7
Average error (%)		2	8	16	21
SL- diesel	0	0.004	0.007	1.46	2.95
SL- diesel	8	8.2	8.8	11.9	12.2
SL- diesel	14.8	15.3	15.9	17.4	17.9
SL- diesel	20.6	20.6	21.1	22.9	23.2
SL- diesel	25.8	26.3	27.3	28.3	29.2
Average error (%)		2	6	22	25
SL- ethanol	0	0.004	0.01	1.56	2.77
SL- ethanol	7.3	7.2	7.7	9.5	10.4
SL- ethanol	13.6	13.8	15.0	20.0	20.5
SL- ethanol	19.1	19.3	21.1	23.3	23.7
SL- ethanol	24	24.4	25.8	27.4	27.8
Average error (%)		1	8	28	33
SL- petrol	0	0.003	0.009	0.95	1.83
SL- petrol	6.8	6.6	5.1	4.4	4.0
SL- petrol	12.8	12.5	10.7	9.6	9.1
SL- petrol	18.1	18.3	17.1	15.1	14.6
SL- petrol	22.7	23.1	24.0	19.2	18.6
SL- petrol	26.9	27.4	28.1	23.1	22.9
SL- petrol	34	34.4	32.2	29.9	28.3
SL- petrol	38.1	38.4	36.5	35.4	34.9
SL- petrol	42	42.2	41.2	39.5	39.0
Average error (%)		2	9	16	19
SCL- biodiesel	0	0.004	0.009	1.02	1.99
SCL- biodiesel	8	8.4	8.9	10.6	11.1
SCL- biodiesel	14.9	15.1	15.7	11.9	10.7
SCL- biodiesel	20.8	21.5	23.0	24.2	25.9
SCL- biodiesel	26	26.4	28.1	29.1	30.7
SCL- biodiesel	30.5	31.2	33.0	34.3	36.1
Average error (%)		3	9	20	27
SCL- diesel	0	0.004	0.002	0.97	2.02
SCL- diesel	8	8.4	7.2	10.2	10.9
SCL- diesel	14.4	15.1	15.9	16.8	17.1
SCL- diesel	20.6	21.2	21.9	18.2	17.3
SCL- diesel	25.8	26.4	27.6	20.0	19.6
Average error (%)		4	8	20	24
SCL- ethanol	0	0.004	0.09	2.12	3.01
SCL- ethanol	7.3	7.5	8.1	9.6	10.3
SCL- ethanol	13.6	14.5	14.9	15.8	16.2
SCL- ethanol	19.1	19.5	18.3	21.1	21.9
SCL- ethanol	24	24.3	25.0	26.6	27.8
SCL- ethanol	28.3	28.7	27.1	31.4	32.6
Average error (%)		3	7	16	21
SCL- petrol	0	0.003	0.002	0.99	1.76
SCL- petrol	6.8	6.6	5.2	4.6	3.9
SCL- petrol	12.8	12.5	13.2	9.4	9.0
SCL- petrol	18.1	18.3	17.1	15.7	14.2
SCL- petrol	22.7	23.1	20.6	19.1	18.2
SCL- petrol	26.9	27.4	24.4	23.4	22.7
SCL- petrol	34	34.4	32.1	29.9	29.1
SCL- petrol	38.1	38.4	36.8	35.5	35.0
Average error (%)		2	9	17	22

4.5 Summary

The experimentation and results of the two proposed methods, both for validation and HC spill detection has been detailed in this chapter. Experiments using synthetic data, real data as well as the controlled datasets demonstrates the effectiveness and accuracy of the proposed methods for HC spill detection.

From the hybrid switch methods, the use of the vicinity parameters to train the network has improved the accuracy of the model. Likewise with the deep learning model, the choice of suitable hyperparameters and good training algorithm has demonstrated to influence the resultant model for accurately estimating the amount of HC in each mixture type.

Chapter 5

Discussion

5.1   Introduction

In this chapter, results generated in the previous chapter will be discussed, advantages and limitations of the proposed methods from chapter 3 will be addressed.

5.2   Hybrid Spectral Unmixing Method for Switching Between Linear and Nonlinear Spectral Unmixing

Nonlinearity usually occurs when photons interact with different cover types in a scene with materials before reaching the sensor. We assumed here that the linear mixing could be associated with mixtures for which the pixel components appear in spatially segregated patterns. More specifically the linear models are better explained in areas that are spatially correlated.

From the experiments in the previous chapter, a controlled simulated data was used for the first experiment. Each image consisted of a series of regions. Each region had the same type of ground cover with added noise. Figure 4.8 showed the results for the simulated dataset with 5 classes and SNR =50 dB. Although the average error is of the same order of magnitude for both linear

and nonlinear approaches, the distribution of error differs. It is noted that the linear models FCLS and VCA detected the low abundances of classes contained in each pixel (shown in black on the ground truth figures). The errors are related to quantification rather than detecting the wrong class. This might be due to the algorithm performing poorly with high spectral variability within the classes. The nonlinear method, especially the PPNMM, outputs an error which is more uniform and not so related to the spatial pattern of the data or spectral variability as displayed in Figures 4.8 and 4.9. The proposed approach assumptions are further validated with the real data sets. In particular, the Jasper Ridge data set which has 3 classes; water, soil and road. Figure 4.10 shows the abundance estimation for the individual and hybrid switch methods. It is noted that VCA has been reported to under perform in this dataset [65]. However, the road class is identified compared to the nonlinear methods that failed to detect this class. On the other hand, the linear methods failed to correctly classify the water class which is a more spectrally variable medium. Thus, it seems that noise and endmember spectral variability makes the nonlinear models outperform the linear ones while spatially structured areas are well defined with the linear model. The vicinity parameters used in the proposed method address both the spatial and spectral diversity of the datasets used in the experiments. The test in Table 4.4 showed that all parameters played an important role in the decision making process. Moreover, Figures 4.8 and 4.9 also support that the chosen features are suitable and that the switching is appropriate in achieving good and improved results. To demonstrate the ability of the method to output accurate results with different training samples sizes, the neural network was trained using 70%, 30%, 10% and 0.3% samples. Experimentation with the real data, 3×3 window and 4×4 window vectors, proved the effectiveness of the hybrid switch methods - the results show that the size of datasets used for training the network and the vector size does not affect the accuracy of

the hybrid methods in switching between linear and nonlinear spectral unmixing, which means that the network can be trained with less sample data without the loss of prediction accuracy.

5.2.1 Advantages and Limitations

The proposed hybrid switch method decides on the best method and provides a switch between unmixing methods for given spectral images on a pixel by pixel basis. It cannot only provide more accurate results, as shown in the experimental section but also reduce computational costs by selecting the most appropriate approach and performing computations for just the selected approach. This research study has proven the capabilities of the proposed methodology based on certain parameters. However, the supervised ANN relies on having ground truth data for training which is not always available. It will be interesting to test the hybrid switch method using unsupervised approaches such as self-organizing maps which have been successfully used in spectral data for classification and anomaly detection tasks [196].

Although spatial and spectral features within windows were used for learning and thus to make the decision, the switching was made at the individual pixel level.

The raw data was also used to train the network for deciding between the linear and nonlinear models. The network was trained with fewer samples to validate the effectiveness of the model in the decision making for the most suitable method.

The results indicated that the hybrid methods for switching between linear and nonlinear spectral unmixing are more effective than the individual methods, meanwhile, it can also be said that ANN pattern recognition has

good capability in recognizing patterns. The proposed hybrid switch method is effective even with few samples used to train the network.

5.3   Hydrocarbon Abundance Estimation Using The Hybrid Switch Method

Results of the experiment using the hybrid switch spectral unmixing method for HC spill detection in Section 4.4.1 suggest that the hybrid switch methods and most importantly the nonlinear models are well suited to the task of HC spill detection.

From the four hybrid methods, the VCA – PPNMM method outperforms the other methods with the lowest average abundance estimation error of about 0.0016 when mixed with sandyloam - biodiesel (Table 4.12), 0.0016 when mixed with sandyclayloam -diesel (Table 4.13) and 0.0025 when mixed with clay- ethanol (Table 4.10).

The overall accuracy of the hybrid switch methods for HC spill also demonstrates that the VCA – PPNMM has an overall accuracy of 98.9 %, followed by FCLS – PPNMM with an overall accuracy of 96%. VCA – GBM has an overall accuracy of 95% while FCLS – GBM has an accuracy of 93%. The result is similar to what was obtained in Section 4.3.With regards to the grain size of the soil samples, it is expected that the nonlinear models will detect well the endmembers with low abundance estimation error when the mixture contains medium grain size (clay and clayloam) and the linear models will detect well on mixtures with coarse grain size (sandyclayloam and sandyloam). Results in Tables 4.10 and 4.11 shows that the nonlinear models detected well the abundances of the mixture with medium grain size. Although not all the nonlinear models detected well, for example; in Table 4.10 clay contaminated with biodiesel mixture, GBM performs well considering the grain size of the

mixture but the FCLS method perform better. From Tables 4.12 and 4.13, the linear models perform well in detecting the abundance of the mixtures with coarse grain size.

The size of the training samples did not affect the accuracy of the results as shown from the results when the experiment was conducted with 70% and 30% of samples to train the network. The PPNMM method produced the best results with lower abundance estimation error on all the mixture types in comparison to other individual methods. It could be said that the nonlinear models are best suited for this application if a single method is to be used, although some pixels are better fitted with a linear model. The abundance estimation error of the network produced low error, thus it can be concluded that the network is suitable for the prediction of HC in soils.

5.4 Deep Learning Based Spectral Unmixing Method for Hydrocarbon Spill Quantification

From the experimentation and results chapter, controlled datasets were used to predict and quantify the amount of HC on different soil types using hyperspectral data and a deep learning model. The deep learning approach was trained using a three-term backpropagation algorithm with dropout. The model was trained using a known output so as to produce acceptable results. The deep learning model designed for the experiment utilizes a sigmoid activation function and dropout of 20% in all the hidden layers of the architecture in order to avoid overfitting. Another advantage of utilizing dropout is its ability to generalize on unseen data.

Given the high disagreement across epochs between human experts [197], 1 - 2% in improvement in performance may not be considered significant.

The proposed deep learning model has two characteristics that renders it better than current state-of-the-art methods, first there is a significant decrease in error for the prediction estimation of hydrocarbon in all the experiments conducted. The proposed deep learning method estimated a small concentration of spills of about 5% (Tables 4.22 and 4.23) and is the closest in match to the reference compared to the state-of-the-art methods which mainly concentrate on quantifying large spills. For instance, Scafutto *et al* [24] reports the estimation of 30% of HC contamination in soils.

Recently, [187] presented regression models based on HC absorption bands in order to estimate the pollution level of different HCs. Although they were able to observe changes in the spectral response, in some cases, for 2% of contaminant and successfully applied their models to identify soils contaminated with just 3% of heavy oil and 14% of diesel they were not able to detect gasoline in soils which is a major limitation to their model. The results show the capabilities of the spectral data in the VNIR–SWIR interval to detect low HC concentrations.

However, this is not the same as achieving high performance in estimation using all the different hydrocarbon types with the closest match in accuracy of estimation to the reference target. The proposed deep learning model is preferable to a method that achieves high performance in one hydrocarbon type and low accuracy in others or even fail to recognise a type of hydrocarbon.

Secondly, the proposed deep learning model was able to achieve acceptable results using a maximum of 20 epochs. This could be attributed to the use of the three-term backpropagation algorithm to train the model. The main aim of the three-term backpropagation algorithm is to train faster, reduce the number of training epochs and maintain the system's stability during training. The proposed method was able to achieve the objective of the three-term backpropagation algorithm which could also estimate the amount

of HCs in each dataset with high accuracy using a low number of epochs. The network was able to achieve an average of 2.5×10^{-4} mean square error on an average of 18 epochs as shown in Tables $4.16 - 4.19$. Figures 4.13 - 4.16 show the mean square error of the proposed deep learning model on all the different mixtures. Although variation was observed in the output target in each mixture type used for the experiment, this could be attributed to the density types of the hydrocarbons used and also the soils' properties such as grain size and texture which can lead to variation in the absorption level and thus the difference in the detection of the different hydrocarbon types.

The soil continuity experiment was conducted so as to simulate a more realistic scenario, this was conducted to validate the network's ability to estimate the amount of HC regardless of the soil type and allowing for different soil types. The results were in the same range as for the different HC as shown in Table 4.21. It is noted that the training mean square error was in the same range as the individual model, although the number of epochs required increased to 178. The generalization mean square error on the training data is higher than in the individual models. However, it is noticed that this data is more complex as the individual models as it contains four different soil types. In the work of the authors of [195] , an individual model per soil type was recommended as the generic models degraded the responses. It is believed that with further tuning of the hyperparameters, improved results could have been achieved. However, this is out of the scope of this research.

The model shows similar performance even with low SNR, showing good noise rejection and accurate prediction at different ranges of SNR. For SNR lower than 20 dB, the generalization error deteriorates slightly. Initially, noise was added to the test data, but not to the training, to simulate training with discrete samples and testing with continuous samples. The performance deteriorates for SNR = 10 dB, while for higher SNR, it is very stable, showing good adaptation of the model to continuous spectra.

Krizhevsky *et al* [198] revealed that deep learning requires large amounts of training data to achieve optimal performance and good generalization with minimum error. It was recognised that better results could have been achieved with a large dataset. However despite the size of the dataset used, the training process was faster with the proposed deep learning method, attributed to the use of the three-term backpropagation algorithm with dropout for training the model. It is anticipated that including a wider range of data such as field HC leaks, will require larger datasets and longer training to account for the variability of the data. Nevertheless, the proposed methodology has proven to decrease significantly both the learning time and sample data required to achieve accurate generalization.

Learning takes an average of 28.16 seconds with our proposed deep learning model compared to the conventionally trained neural networks which takes an average of 300.68 seconds to learn. The data processing-to-end-product time of our proposed method is relatively more time-consuming compared to the traditional spectral unmixing method, which is approximately 35.45 seconds compared to the hybrid spectral unmixing method which has an average of 24.64 seconds. This could be attributed to the number of parameters and training in the deep learning method compared to the spectral unmixing method.

Srivastava *et al* [19] reported that large neural networks trained in the standard way tend to overfit on small datasets. To see if dropout can improve this condition, the experiment was conducted on all the datasets and the dropout ratio varied as demonstrated in Tables 4.16 - 4.19. This shows that the error rate is relatively low when the dropout ratio is between 10% and 30%, with a slight increase when the ratio is set above 50%. Therefore, it can be concluded that a dropout ratio between 10% and 30% provides an acceptable prediction estimation and could be used in any dataset for estimation analysis.

To evaluate the network's ability to generalize on unseen data, the training data was plotted against the corresponding output targets of the different hydrocarbon types used in each mixture. The results reveal that the deep learning model is a promising tool in detecting the amount of HC used in the experiment as shown in Figures 4.17 - 4.20.

From the results obtained, it is noted that the proposed deep learning model trained using the three-term backpropagation algorithm with dropout was able to generalize on unseen data with high prediction accuracy. It showed that the method was able to estimate the amount of hydrocarbon spill with the closest match compared to the hybrid spectral unmixing method. A similar trend was observed on all the datasets used for the experiment which indicates a reduction in the error rate and a high convergence rate.

The effectiveness of the proposed deep learning model was proven with a comparison of the deep learning method against the hybrid spectral unmixing method and a deep learning model trained with the conventional backpropagation algorithm. Results obtained (Tables 4.22 and 4.23) shows that the proposed deep learning model outperforms the deep learning model trained with conventional backpropagation algorithm even when dropout was used in training the network. The proposed methodology shows the closest match to the reference target in all the experiments conducted with all the different mixture types.

For instance, it is noted from the first mixture in Table 4.22 (biodiesel mixed with clay), it is noted that for the pure clay sample (reference 0% HC), all methods provide a close estimate. The second row presents the results for the same mixture with 8% of biodiesel, 92% of clay. The deep learning method estimated 8.5% of biodiesel, the hybrid switch method 9.2% while the deep learning model trained with conventional backpropagation algorithm with and without dropout estimated 9.7% and 9.9% respectively. Similar results were obtained for all the different samples across the two tables

(Tables 4.22 and 4.23).

To the best of our knowledge, the proposed deep learning model has the best performance in the literature using spectral data and deep learning methods to detect and estimate the percentage of HCs in soil. There are examples in the literature that use different methods for HC estimation [29, 6] but none could estimate with a high prediction estimation for small concentrations. The proposed method achieved high performance with all the different hydrocarbon types used in the experiments.

5.5   Summary

To test the optimization of the proposed methods, firstly, the models were validated on synthetic data to ensure a reasonable comparison on real datasets. The efficiency of the hybrid switch spectral unmixing method between linear and nonlinear models and the deep learning model for HC spill estimation was tested on controlled datasets as discussed in this chapter. The results show the efficiency of the two methods in estimating the amount of HC as mixed in the datasets.

Chapter 6

Conclusions and Future work

6.1 Conclusions

A review of the hypothesis proposed in chapter 1 of this study is provided here with a summary of the findings in relation to each hypothesis.

1. The choice of best method for spectral unmixing of hyperspectral data can be selected through hybridization between the linear and nonlinear models.

2. The optimum model choice can be extracted from the pixel scene information and varies in a pixel by pixel basis.

3. Spectral unmixing methods can be optimized by applying deep learning approach to extract the abundances.

6.1.1 Hypothesis 1 and 2

Hypothesis 1 and 2 were investigated as discussed in chapter 3 and validated in chapter 4. In Chapter3, a hybrid method for switching between linear and nonlinear spectral unmixing models using artificial neural networks was presented. In this method, the neural networks were trained with parameters within a window of the pixel under consideration. These parameters were computed to represent the diversity of the neighboring pixels and are

based on the Spectral Angular Distance, Covariance and a nonlinearity parameter. The endmembers were extracted using Vertex Component Analysis (VCA) while the abundances were estimated using the method identified by the neural networks. The methods used for the switching were the state-of-the-art methods which have been used extensively as reference in the literature. These methods are Vertex Component Analysis (VCA), Fully Constraint Least Square Method (FCLS), Polynomial Post Nonlinear Mixing Model (PPNMM) and Generalized Bilinear Model (GBM). Results show that the hybrid switch method performs better than each of the individual techniques with high overall accuracy and low computational cost, whilst the abundance estimation error is significantly lower than that obtained using the individual methods. The hybrid switch method decides which of the individual methods is most suitable for a particular application. In this study, the PPNMM nonlinear model proves to be best for oil spill quantification with an overall accuracy of about 98% compared to the other individual methods.

In Chapter 4, the hybrid switch method was used for estimating the quantity of HC spills in a hyperspectral data. Here the aim was to robustly choose the most suitable method among linear and nonlinear spectral unmixing approaches to quantify different HC substances in different soils. Hypersectral data sets have been acquired using mixtures of different HCs and soils. Then an artificial neural network was used for switching between linear and nonlinear methods to assess the most suitable method in quantifying the amount of HC. Results show that the hybrid methods are more suitable than the individual technique with high overall accuracy and lower abundance estimation error compared to those obtained with the individual methods. Again the VCA – PPNMM hybrid switch method proved to be the best with 98.9% accuracy in all the experiments conducted. With regards to the individual methods, the PPNMM nonlinear model also proves to be a good method with lowest abundance estimation error in comparison to other individual

methods. Therefore, it can be concluded that the nonlinear models are suited for the task of estimating hydrocarbon spills on different soil types.

6.1.2 Hypothesis 3

Deep learning is found in the literature review in section 2.4.6 to be a complex model that involves selection and fine-tuning of numerous hyperparameters in order to achieve efficient results. In order to optimize the deep learning model for spectral unmixing of a hyperspectral data, different parameters were selected so as to achieve optimum results.

Deep learning models provide excellent results with the ability to extract stronger features, but in turn lead to vanishing gradient, overfitting and computational load. Deep learning model utilizes backpropagation algorithm to train a network which is often done by iterative updating of weights employing the negative gradient of the mean square error function. In order to reduce this problem, the proposed deep learning model was trained using a three-term backpropagation algorithm with dropout. The aim is to improve the accuracy of the model, avoid overfitting and the ability to converge faster.

The conventional backpropagation algorithm builds co-adaptation which works well for training data, but the network does not generalize to unseen data. Dropout neutralizes these co-adaptations by improving the network performance, thus enabling it to generalize. The choice of dropout ratio used in any neural networks depends on the type of dataset and application.

The three-term backpropagation algorithm improves the network's ability to train faster and overcome local minima when compared to conventional backpropagation algorithm. The effectiveness of the deep learning model was verified when tested on dataset containing different mixture types of HCs to estimate the amount and concentration of each HC type as mixed in each dataset. The result of the experiments consistently shows that the

proposed method provides high prediction accuracy with low error. There-fore, it can be concluded that the three-term backpropagation algorithm with dropout significantly improves the model's operation.

Recent research in deep learning suggests that a large dataset is required for training remote sensing data which is its major setback, however, this was not a challenge in the proposed model because of the three-term back-propagation algorithm used in training the network makes the model to train faster using an average of 18 epochs to converge compared to a deep learn-ing model trained using a conventional backpropagation algorithm where the parameter updates are very noisy in architectures with dropout.

In summary, the deep learning model can be improved by adopting the three-term back propagation algorithm with dropout to correctly estimate the amount of HC spills using a hyperspectral dataset. A model is presented in this work that incorporates such algorithm with a number of hyperparam-eters to help fine-tune the model in order to provide acceptable results for HC spill quantification.

6.2   Future work

The results of the work presented here provide a number of opportunities for future research based on existing outcomes. Some of the directions are listed as follows:

In Chapter 3, a hybrid method for switching between linear and nonlinear spectral unmixing models for given spectral images using supervised ANN was proposed. The capabilities of the proposed method based on certain parameters have been proven and validated in chapter 4. The main limitation of the supervised ANN is the requirement of ground truth data for training which is always not available.

Future work will expand to unsupervised approaches such as self-organizing maps which have been successfully used in spectral data for classification and anomaly detection tasks [196, 93].

Although the proposed hybrid switch method was made at individual pixel level. Another area of interest for future research will base the decision on a group of pixels, or areas using for instance Markovian Jump method for switching between linear and nonlinear spectral unmixing.

In Chapter 3, a deep learning based unmixing method was proposed to correctly estimate hydrocarbon spills on different soil samples measured using imaging spectroscopy. The deep learning model was trained using a three-term backpropagation algorithm with dropout with the aim to improve the accuracy of the model, avoid overfitting and converge faster. Compared with a deep neural network trained with conventional backpropagation algorithm, the proposed model could achieve higher accuracy using all the experimental datasets. It will be interesting to further this research to classify and identify each type of HC based on an image based convolutional neural network and a real world datasets.

Another aspect to consider, would be the use of airborne or space borne data. The models would consequently need to be trained with a wider range of soils, HC, and minerals in order to perform accurate detection. Other aspects such as spatial resolution would need to be assessed.

A general problem when it comes to HS image processing is the lack of data samples with labelled ground truth within the scope of this research, a dataset has been acquired and data provided has been labelled. However, the samples used homogenized mixtures and assumes that spills were at the soil surface.

Future research should also look at different penetration of spills. Moreover, it not just consider the spatial and spectral domains, but also the temporal as HC spills will certainly produce dynamic spectral images as the spills

propagate through the soil and eventually partially dissipate into the air.

Appendix A

Appendix

Quantification of Hydrocarbon Abundance in Soils Using Deep Learning with Dropout and Hyperspectral Data

Asmau M. Ahmed [1], Olga Duran [1,*], Yahya Zweiri [1,2] and **Mike Smith [3]**

[1] Faculty of Science, Engineering and Computing, Kingston University London, London SW15 3DW, UK
[2] Khalifa University Center for Autonomous Robotic Systems (KUCARS), Khalifa University of Science and Technology, P.O. Box 127788 Abu Dhabi, UAE
[3] School of Geography, Earth and Environmental Sciences, University of Plymouth, Drake Circus Plymouth Devon, Plymouth PL4 8AA, UK
[*] Correspondence: o.duran@kingston.ac.uk; Tel.: +44-020-8417-4740

check for updates

Abstract: Terrestrial hydrocarbon spills have the potential to cause significant soil degradation across large areas. Identification and remedial measures taken at an early stage are therefore important. Reflectance spectroscopy is a rapid remote sensing method that has proven capable of characterizing hydrocarbon-contaminated soils. In this paper, we develop a deep learning approach to estimate the amount of Hydrocarbon (HC) mixed with different soil samples using a three-term backpropagation algorithm with dropout. The dropout was used to avoid overfitting and reduce computational complexity. A Hyspex SWIR 384 m camera measured the reflectance of the samples obtained by mixing and homogenizing four different soil types with four different HC substances, respectively. The datasets were fed into the proposed deep learning neural network to quantify the amount of HCs in each dataset. Individual validation of all the dataset shows excellent prediction estimation of the HC content with an average mean square error of ~2.2×10^{-4}. The results with remote sensed data captured by an airborne system validate the approach. This demonstrates that a deep learning approach coupled with hyperspectral imaging techniques can be used for rapid identification and estimation of HCs in soils, which could be useful in estimating the quantity of HC spills at an early stage.

1. Introduction

Hydrocarbons refer to chemical substances formed exclusively from carbon and hydrogen. Naturally occurring hydrocarbon (HC) substances, depending on the length of the carbon chain, occur in different forms; solid, liquid, and gas [1]. Liquid HCs found in nature consist of a complex mixture of various molecular weights; in addition nitrogen, sulfur, and oxygen exist in small quantities [2].

While the economic significance of HCs is attributed to its primary use as fuel and then versatile application in downstream industries, they can have detrimental environmental consequences [1,3]. Oil exploration, production, and processing represent potential environmental exposure to HCs resulting in accidental terrestrial spillage thereby altering the physical and chemical properties of soils. HCs may therefore be environmentally harmful, causing toxicity, and limiting soil quality [4].

Knowledge about the concentration and nature of a spill is important in order to track their propagation in the environment, assess their risk and propose remediation strategies [5,6]. To effectively protect communities affected by a spill, fast and accurate determination of the area impacted is

needed, particularly if monitoring large regions affected by an oil spill or where aged oil transporting facilities are involved [7]. Traditional methods employed to track and detect oil spills and the concentration of HCs in soils often involve processes which are expensive and time consuming as they require field sampling, chemical analysis, and geostatistical interpolation [8,9]. Imaging spectroscopy has been recognized as a reliable alternative method for detecting HCs in soils and is rapid and cost-effective [6,10].

Imaging spectroscopy (hyperspectral imaging) can be described as the combination of digital imaging and spectroscopy. A hyperspectral camera captures the light intensity for a large number of spectral bands, providing much more information about a scene when compared to a standard camera which only covers the visible wide bandwidth portion of the electromagnetic spectrum [6]. Due to the rich information content in hyperspectral imagery, it is well suited to a range of applications such as crop/vegetation classification, disaster monitoring, oil spill detection, etc. There are several uses of imaging spectroscopy for oil spills, such as the enforcement of ship discharge laws, surveillance and general slick detection, mapping of spills, and direction of spills [11], due to its high spectral and spatial capabilities [12].

More specifically, Near- and Shortwave Infrared (NIR-SWIR) spectroscopes have been popular methods for detecting, mapping, quantifying, and characterizing HCs in contaminated soils with reasonable accuracy [6,13,14]. Moreover, NIR-SWIR spectra provide good information on soils organic and inorganic material content [13]. HCs demonstrate good absorption in spectral bands 1200 nm, 1725 nm, and 2310 nm [5,8,15]. Therefore, spectral information obtained in the NIR-SWIR range is excellent for both the quantitative and qualitative analysis of HCs in soils [13]. Recent works have also successfully demonstrated the use of Longwave Infrared (LWIR) for Petroleum HC detection [16].

Different methods have been used to analyze reflectance spectroscopy data to detect HCs in soils; the authors of [5] used regression analysis and spectral preprocessing to generate statistical models to identify different HC products mixed with a mineral substrate. The authors of [15] used Diffuse Reflectance Infrared Fourier-Transform (DRIFT) spectroscopy which is a hand held spectrometer for the prediction of total petroleum hydrocarbons in contaminated soils. It uses Partial Least Square (PLS) regression analysis, which is a multivariate method and includes correlation between spectral information and corresponding analytical data to rapidly predict the concentration of HCs in soil. Other researchers show the robustness of visible and infrared spectroscopy for the rapid estimation of HCs [17,18].

However, state-of-the art methods for estimating HC concentration in soils mainly concentrate on the quantification of large spills [19]. For instance, the authors of [6] report the estimation of 30% of HC contamination in soils. Recently, the authors of [20] presented regression models based on HC absorption bands in order to estimate the pollution level of different HC. They were able to observe changes in the spectral response, in some cases, for 2% of contaminant and successfully applied their models to identify soils contaminated with just 3% of heavy oil and 14% of diesel. However, the spectrum of soils contaminated with gasoline showed only subtle changes for pollution levels higher than 8%. Thus, they concluded that it would be difficult to detect soils contaminated with gasoline by assessing the VNIR–SWIR interval.

One of the characteristics of hyperspectral remotely sensed data is that the recorded reflectance is the result of multiple interactions of the electromagnetic radiation with the constituents of the soil creating mixed pixels. Numerous studies address the mixing problem and propose analysis techniques [19]. Spectral Unmixing (SU) is the process of identifying spectral signatures of materials, often referred to as endmembers, and then estimate their relative abundance to the measured spectra within a pixel [21]. Endmembers play an important role in exploring spectral information of a hyperspectral image [22,23], as usually the extraction of endmembers, which is the process of obtaining pure signatures of different features present in an image, is the first step in the unmixing algorithms [24–26]. SU often requires the definition of the mixing model underlying the observations as presented on the data. A mixing model describes how the endmembers combine to form the mixed

spectrum as measured by the sensor [27]. Given the mixing model, SU then estimates the inverse of the formation process to infer the quantity of interest, specifically the endmembers, and abundance from the collected spectra [28–30]. This could be achieved through a radiative transfer model that accurately describes light-scattering by the materials in the observed scene by a sensor [27,31]. The two main approaches to spectral unmixing are linear and nonlinear models [21,22,25,26,28].

Different methods utilizing both linear and nonlinear models have been demonstrated in the literature for the analysis of different hydrocarbon types. In a work by the authors of [13,15], Principal Component Analysis (PCA) and PLS regression are used. The authors used PCA to differentiate the types and density of HCs in soils while they used PLS to predict the concentration of oils and fuels in soil samples. The authors of [18] used Spectral Angular Mapper (SAM) to classify oil spills on an image and also used signature matching to distinguish oils from other features. However, most of these methods adopt a linear model and smoothing threshold function for feature extraction. Other approaches such as a Kernel-based transformation [32] and manifold learning algorithm [33] are based on nonlinear models.

In the work by the authors of [34], we proved experimentally that HCs abundance in soils was estimated with higher accuracy when non linear unmixing models were applied. Nevertheless, spectral unmixing and specifically the abundance estimation of HCs such as gasoline, can be challenging [20], and may require more advanced techniques such as deep learning. Deep learning network can be considered a powerful technique to solve nonlinear problems, which can be fast, accurate and does not rely on any assumptions to estimate the abundances in a given dataset. However, to the best of our knowledge, there is no study that uses spectral data and deep learning methods to detect and estimate the percentage of HCs in soils. While the value and application of these two techniques have been presented in independent research activities, the techniques have not yet been combined. Therefore, in this paper, a deep learning approach is developed to estimate the amount of HC contamination in soil samples using SWIR imaging spectroscopy. The remainder of the paper is organized as follows. Section 2 describes the data acquisition process including the materials used, sample preparation and the hyperspectral sensor used. Section 3 discusses the methodology, including the parameters used in training the network, the architecture of the deep learning approach, as well as the validation method. Results are presented in Section 4 and discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Data Acquisition

2.1. Materials

The hyperspectral imaging sensor used for this experiment covers the Shortwave Infrared (SWIR) range (930–2500 nm), which has been found suitable for the detection of HCs [5,6,13,35,36], mineral identification and mapping [36], rock mapping [37], and mapping of mafic and ultramafic units in the Cape Smith Belt [38].

The soils and HC types selected here have been used extensively in the literature for assessments of HC contamination in different soil types [8,13,17,18,39]. Different HC types, namely Diesel, Bio-diesel, Ethanol, and Petroleum were used. These are the most commonly used HCs in the literature. Soil types include typical mixtures of clay (<0.002 mm in diameter), silt (0.002–0.05 mm in diameter), and sand (0.05–1 mm in diameter). In particular, we used mixtures with different grain size ranging from medium to coarse as follows; Clay, Clay Loam , Sand Clay Loam, and Sand Loam [40].

2.2. Sample Preparation

The preparation of the samples consisted of the following steps.

- Each soil type was air-dried, and therefore all samples contained similar levels of moisture.
- Fifty grams of a soil sample type was added to a petri dish (12 cm in diameter)
- The sample was scanned with a Hyspex SWIR 384 m camera under constant illumination.

- In the same sample, initially, 2 mL of the HC were added to the soil using a syringe (to clay and clayloam), which was subsequently changed to 5 mL of the HC to the other soil types.
- A disposable plastic spoon was used to homogenize the mixture and to flatten its surface in order to have even surfaces, except for some soil samples containing clay which tends to be sticky and difficult to flatten due to the characteristics of the soil type, e.g., Figure 1b.
- The sample was scanned with a Hyspex SWIR 384 m camera under constant illumination.
- In the same sample, a further 5 mL of HC was added to the mixture.
- The disposable spoon was used to homogenize the mixture and another scan was taken.
- The procedure was repeated with increments of 5 mL of HCs until the mixture was saturated and formed a shallow local pool (see Figure 1).

Figure 1. Sample preparation of the experiment combining sandy-clay-loam with diesel. Photos show the HC contaminant being increasingly added to the same soil sample until saturated. From left; addition of 5 mL , followed by 10 mL, 15 mL, 20 mL, and 25 mL of the HC.

The procedure was repeated on all the soil samples contaminated with all the different hydrocarbon types.

A calibration panel was used as white reference, the acquired images were calibrated from radiance to reflectance using HYSPEX REF software which normalizes the images to an area of known reflectance. A total of 15 combinations (see Table 1) were produced with four mixtures each for clay–loamy, sandy–clay–loam, and sandy–loam soil types, while clay had three mixtures. The complete data set used here consisted of 96 spectral images.

Table 1. Samples created for each combination made in the experiment and their corresponding absolute HC and soil quantities, respectively.

Sample Combination	HC (mL)	Soil (gr)	Sample Combination	HC (mL)	Soil (gr)
Clay - Diesel 0	0	50	Clay - Bio- diesel 0	0	50
Clay - Diesel 1	2	50	Clay - Bio- diesel 1	2	50
Clay - Diesel 2	4	50	Clay - Bio- diesel 2	4	50
Clay - Diesel 3	5	50	Clay - Bio- diesel 3	5	50
Clay - Diesel 4	10	50	Clay - Bio- diesel 4	10	50
Clay - Diesel 5	15	50	Clay - Bio- diesel 5	15	50
Clay - Diesel 6	20	50	Clay - Bio- diesel 6	20	50
Clay - Diesel 7	25	50	Clay - Bio- diesel 7	25	50
Clay - Ethanol 0	0	50	Clay Loam - Ethanol 0	0	50
Clay - Ethanol 1	2	50	Clay Loam - Ethanol 1	2	50
Clay - Ethanol 2	4	50	Clay Loam - Ethanol 2	4	50
Clay - Ethanol 3	5	50	Clay Loam - Ethanol 3	5	50
Clay - Ethanol 4	10	50	Clay Loam - Ethanol 4	10	50
Clay - Ethanol 5	15	50	Clay Loam - Ethanol 5	15	50
Clay - Ethanol 6	20	50	Clay Loam - Ethanol 6	20	50
Clay - Ethanol 7	25	50	Clay Loam - Ethanol 7	25	50

Table 1. *Cont.*

Sample Combination	HC (mL)	Soil (gr)	Sample Combination	HC (mL)	Soil (gr)
Clay Loam - Diesel 0	0	50	Clay Loam - Bio- diesel 0	0	50
Clay Loam - Diesel 1	2	50	Clay Loam - Bio- diesel 1	2	50
Clay Loam - Diesel 2	4	50	Clay Loam - Bio- diesel 2	4	50
Clay Loam - Diesel 3	5	50	Clay Loam - Bio- diesel 3	5	50
Clay Loam - Diesel 4	10	50	Clay Loam - Bio- diesel 4	10	50
Clay Loam - Diesel 5	15	50	Clay Loam - Bio- diesel 5	15	50
Clay Loam - Diesel 6	20	50	Clay Loam - Bio- diesel 6	20	50
Clay Loam - Petrol 0	0	50			
Clay Loam - Petrol 1	2	50			
Clay Loam - Petrol 2	4	50	Sand Loam - Petrol 0	0	50
Clay Loam - Petrol 3	5	50	Sand Loam - Petrol 1	5	50
Clay Loam - Petrol 4	10	50	Sand Loam - Petrol 2	10	50
Clay Loam - Petrol 5	15	50	Sand Loam - Petrol 3	15	50
Clay Loam - Petrol 6	20	50	Sand Loam - Petrol 4	20	50
Clay Loam - Petrol 7	25	50	Sand Loam - Petrol 5	25	50
Clay Loam - Petrol 8	30	50	Sand Loam - Petrol 6	30	50
Clay Loam - Petrol 9	35	50	Sand Loam - Petrol 7	35	50
Clay Loam - Petrol 10	40	50	Sand Loam - Petrol 8	40	50
Clay Loam - Petrol 11	45	50	Sand Loam - Petrol 9	45	50
Sand Clay Loam - Diesel 0	0	50	Sand Clay Loam - Bio- diesel 0	0	50
Sand Clay Loam - Diesel 1	5	50	Sand Clay Loam - Bio- diesel 1	5	50
Sand Clay Loam - Diesel 2	10	50	Sand Clay Loam - Bio- diesel 2	10	50
Sand Clay Loam - Diesel 3	15	50	Sand Clay Loam - Bio- diesel 3	15	50
Sand Clay Loam - Diesel 4	20	50	Sand Clay Loam - Bio- diesel 4	20	50
Sand Clay Loam - Diesel 5	25	50	Sand Clay Loam - Bio- diesel 5	25	50
Sand Clay Loam - Ethanol 0	0	50	Sand Clay Loam - Petrol 0	0	50
Sand Clay Loam - Ethanol 1	5	50	Sand Clay Loam - Petrol 1	5	50
Sand Clay Loam - Ethanol 2	10	50	Sand Clay Loam - Petrol 2	10	50
Sand Clay Loam - Ethanol 3	15	50	Sand Clay Loam - Petrol 3	15	50
Sand Clay Loam - Ethanol 4	20	50	Sand Clay Loam - Petrol 4	20	50
Sand Clay Loam - Ethanol 5	25	50	Sand Clay Loam - Petrol 5	25	50
Sand Clay Loam - Ethanol 6	30	50	Sand Clay Loam - Petrol 6	30	50
			Sand Clay Loam - Petrol 7	35	50
Sand Loam - Diesel 0	0	50	Sand Loam - Bio- diesel 0	0	50
Sand Loam - Diesel 1	5	50	Sand Loam - Bio- diesel 1	5	50
Sand Loam - Diesel 2	10	50	Sand Loam - Bio- diesel 2	10	50
Sand Loam - Diesel 3	15	50	Sand Loam - Bio- diesel 3	15	50
Sand Loam - Diesel 4	20	50	Sand Loam - Bio- diesel 4	20	50
Sand Loam - Ethanol 0	0	50			
Sand Loam - Ethanol 1	5	50			
Sand Loam - Ethanol 2	10	50			
Sand Loam - Ethanol 3	15	50			
Sand Loam - Ethanol 4	20	50			

2.3. Hyperspectral Imaging

The spectral data was obtained using a Hyspex SWIR 384 m line-scan hyperspectral camera and is equipped with a Mercury Cadmium Telluride (MCT) detector array. For this experiment, a user friendly table-top laboratory set-up with translation stage, SWIR light source, and close-up lenses were used during the scanning stage to scan the sample and build a hyperspectral data cube (see Figure 2). The camera simultaneously captured a full SWIR spectrum, with a spectral sample interval of 5.45 nm between 930 and 2500 nm, each along a line of 384 pixels for 288 bands with a radiometric resolution of 16 bit [41]. The 384 columns of the detector array formed one line of the hyperspectral image in the x-axis. The hyperspectral image was obtained line by line using the so-called "pushbroom" scanning mode, where the platform holding the sample was translated onto the y-axis at constant speed (see Figure 3). The scanning speed was automatically controlled by the data acquisition unit based on the selected lens option. The images produced had a spatial resolution of

0.22 mm/pixel. Radiometric calibration was performed using the vendor's software package. A more detailed specification of the system is given in Table 2.
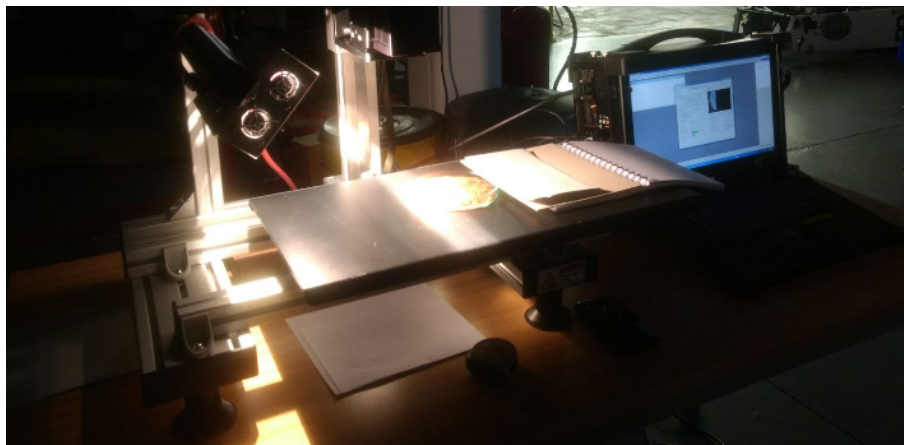
Figure 2. Scanning process of the dataset.

Table 2. Hyspex 384 m main specifications.

Specification	HySpex SWIR-384 m
Spectral Range (nm)	930–2500
Spatial Pixels (pixels)	384
Spectral Channels	288
Spectral Sampling (nm)	5.45
FOV (degrees)	16°
Pixel FOV across/along (mrad)	0.73/0.73
Bit resolution (raw data)/Digitization	16
Noise floor (e^-)	150
Dynamic range	7500
Peak SNR (at full resolution)	>1100
Max speed (at full resolution)(fps)	400
Full Width Half Maximum	~1 pixel
Power consumption (W)	30
Dimensions (l-w-h) (cm)	38-12-17.5
Weight (kg)	5.7
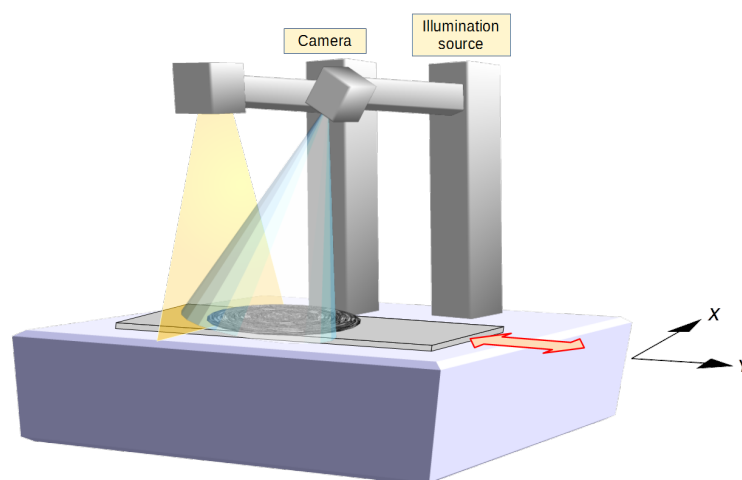
Figure 3. HySpex 384 m line scan acquisition process. The camera (nadir) acquires hyperspectral lines of pixels. The hyperspectral image is obtained by translation of the object under constant illumination.

The resultant reflectance spectra were used to estimate the percentages of the HCs using the abundances calculated based on the different mixture types as shown in Table 3.

Table 3. Size of datasets and target class.

Dataset	Size	Number of Mixtures
Clay biodiesel	8000×288	8
Clay diesel	8000×288	8
Clay ethanol	8000×288	8
Clay loam biodiesel	7000×288	7
Clay loam diesel	7000×288	7
Clay loam ethanol	8000×288	8
Clay loam petrol	$12,000 \times 288$	12
Sandy loam biodiesel	5000×288	5
Sandy loam diesel	5000×288	5
Sandy loam ethanol	5000×288	5
Sandy loam petrol	$10,000 \times 288$	10
Sandy clay loam biodiesel	6000×288	6
Sandy clay loam diesel	6000×288	6
Sandy clay loam ethanol	7000×288	7
Sandy clay loam petrol	8000×288	8

3. Methodology

3.1. Workflow

Spectral information was obtained from the controlled dataset and used with ground truth abundances to evaluate the performance of the proposed deep learning model for estimating the abundance of HCs in each dataset. The workflow of the study is as follows.

- Obtaining the dataset via a controlled experiment by mixing and homogenizing different Hydrocarbon (HC) types with soil samples and scanning them with a Hyspex Shortwave Infrared (SWIR) 384 m camera.
- Applying the Deep Learning (DL) model trained using a three-term backpropagation algorithm with dropout for the abundance estimation of the HCs.
- Structuring the DL model with different dropout ratios to determine the most efficient DL setting.
- Testing and validating the performance of the proposed method for abundance estimation of the different HCs by using the same network structure and hyperparameters.
- Comparing the accuracy and performance of the DL model with a hybrid spectral unmixing method [21] and DL models trained using a standard backpropagation algorithm with and without dropout (to prove the generalization ability of dropout), respectively.

The description and experimental results of this workflow are organized in the following sections. Further explanation and discussion regarding abundance estimation of the HCs by the DL model as well as the other methods can be found in the data acquisition, results, and discussion sections.

3.2. Deep Learning

Deep learning has been shown to outperform other machine learning and neural networks techniques. Deep learning can be categorized as a subfield of machine learning, which learns high level abstractions in data by utilizing hierarchical architectures [42]. Deep learning can also be described as the final product of machine learning where the learning rule becomes the algorithm that generates the model from the training data. It typically involves modeling, which hierarchically learn features of input data using Artificial Neural Networks (ANN) and usually has more than three layers [43]. The main advantage of deep learning is that these layers of features are not designed by an operator; they are learned from the input data using learning procedures. A deep neural network can simply be

referred to as a network of sufficient complexity in order to interpret raw data without human derived explanatory variables [44,45]. Deep learning models provide excellent results with the ability to extract stronger features, but in turn lead to vanishing gradient, overfitting, and computational load [46]. These problems can be addressed and improved by employing dropout, three-term backpropagation and a Rectified Linear Unit (ReLU) activation function which is known to transmit error better when compared to other functions.

There are many types of deep learning architectures whose application have been proven to yield excellent results, the most common are Deep Believe Network (DBN), Convolutional Neural Network (CNN), Deep Convolutional Generative Adversarial Networks (DCGAN), Recurrent Neural Networks (RNN), etc. [47,48]. The application of deep learning techniques to hyperspectral data is relatively recent, for instance, in the work by the authors of [49], deep belief networks, and a novel texture enhancement algorithm were investigated for their suitability and practical application to hyperspectral image classification. The authors of [50] utilized high-resolution remote sensing imagery and deep learning techniques to extract buildings in urban districts using guided filters. In the work of the authors of [51], a 3D full convolutional neural network model was used for spatial-spectral resolution of hyperspectral images by learning end-to-end, with full mapping between low and high spatial resolution hyperspectral images at high accuracy. Transfer learning with a deep convolutional neural network was reported in the work by the authors of [52]; in this research, a large amount of unlabeled SAR scene data was transferred to SAR target recognition tasks with feedback of the construction loss to the classification pathway. Others, such as the authors of [53,54], used a deep learning approach to classify hyperspectral images. Most of the aforementioned methods used the standard backpropagation algorithm to train the network which has been characterized as having low convergence rates especially when used to train a network with more than one hidden layer. Thus, in this paper, the main aim of using the three-term backpropagation algorithm with dropout to train the network is to increase the convergence rate and the ability to generalize to unseen data with good prediction accuracy compared to existing methods.

3.3. Dropout

Dropout allows neurons to randomly drop out of the network during training, while other neurons can step in and handle the representation required to make predictions for the missing neurons [55]. This simply means removing neurons from the network along with all its incoming and out going connections. By applying dropout to a deep neural network, a thinned network often results. This thinned network consists of all the units that survive dropout [56] as shown in Figure 4. The dropout effect is that the network becomes less sensitive to the specific weights of neurons. This in turn results in a network that is capable of better generalization and is less likely to overfit the training data.

(a) Standard Neural Net　　　　　(b) After applying dropout.
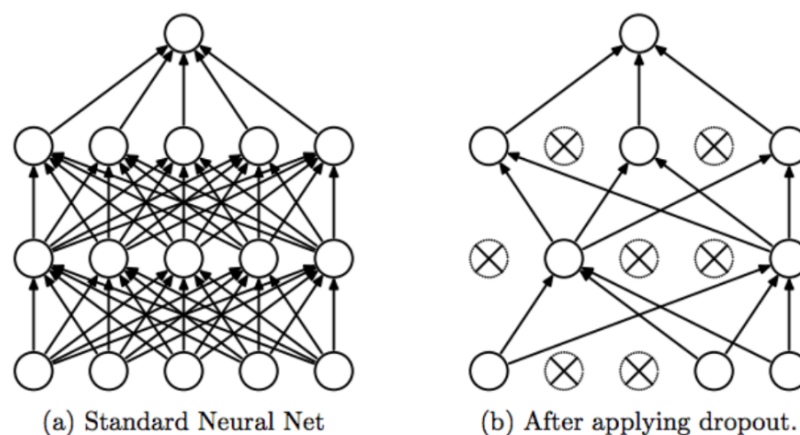
Figure 4. (**a**) A typical network before and (**b**) after applying dropout (adapted from the work by the authors of [55]).

In this paper, dropout on hidden layers and on the visible layer are developed. Dropout on hidden layers is applied to hidden neurons in the hidden layers and between the last hidden layer and the output layer of the body of the deep networks' model. Dropout on the visible layer is applied between the input and the first hidden layer. Since deep neural networks consist of multiple nonlinear hidden layers, this makes them expressive models that can learn complex relationships between the input and output nodes which often results in overfitting.

3.4. Backpropagation

Backpropagation is carried out to train multilayer architectures to minimize the cost function of the model. It is also used to adjust the free parameters weights (ω) and biases in order to attain the desired network output. Traditionally, the learning rate and momentum factors are used to control the weight adjustments and damping oscillations. This is a popular training algorithm in many applications, however the main limitation is its slow convergence especially when used to train a deep neural network with multiple hidden layers. Therefore, the three-term backpropagation algorithm with dropout tend to improve the accuracy of the trained model.

3.5. Three-Term Backpropagation

The backpropagation algorithm has been modified by different researchers to improve the efficiency and convergence rate of the algorithm. One such method is the three-term backpropagation algorithm proposed by the authors of [57], shown in Algorithm 1. This algorithm uses an extra term called the Proportional Factor (PF) to the standard backpropagation algorithm. This PF speeds up the weight adjustment process by increasing the convergence rate and decreasing learning stalls while maintaining the simplicity and efficiency of the standard backpropagation algorithm [58].

Algorithm 1: Learning method using the three-term backpropagation with dropout used in training the DNN model.

Data: α, β, γ, ω, δ, e

DNN weights, ω, are randomly initialized ω_{11} **... ... ...** ω_{ij}

initialize the learning rate, α ; momentum factor, β; and proportional factor, γ

for *Number of epochs (k)* **do**

　　for *Number of data samples = 1, 2 ..... n* **do**

　　　　for *Number of hidden layers = 1, 2 ..... f* **do**

　　　　　　/* Calculate the errors and the delta, δ, of the output nodes

　　　　　　$e = d - y$

　　　　　　$\delta = \phi'(v)\, e$

　　　　　　/* compute the nodes' output y_{ij}

　　　　　　$y_{ij} = y_{ij} *$ Dropout $(y_{ij},$ ratio)

　　　　　　/* propagate the network output y_z backwards, and calculate the delta, δ

　　　　　　$e^p = W^T\, \delta$

　　　　　　$\delta^p = \phi'(v^p)\, e^p$

　　　　　　/* update and adjust the weights according to δ

　　　　　　$\Delta\, \omega_{ij}(k) = \alpha\, \delta_i\, x_j + \beta\, \Delta\, \omega_{ij}(k-1) + \gamma\, e_i\, x_j$

　　　　　　$\omega_{ij} \leftarrow \omega_{ij} + \Delta\, \omega_{ij}$

　　　　end

　　end

end

A stability analysis of the three-term backpropagation was studied in the work by the authors of [58] to test the convergence rate and stability of the algorithm. This training algorithm has proven to be effective in training a network with good prediction accuracy and a high convergence rate [58,59].

A deep learning model with dropout can be trained using the stochastic gradient descent which can be similar to a standard neural network, the only difference here is the random dropping of units in the network's hidden layers. Different methods have been used to improve the standard gradient descent algorithm such as momentum, annealed learning rates, as well as L2 weight decay [55]. Here, the effectiveness of the dropout trained method using the three-term backpropagation algorithm is demonstrated. The three-term backpropagation algorithm speeds up weight space adjustment compared to a conventional backpropagation algorithm. The dropout has proven to be successful for computer vision tasks as it helps to avoid overfitting and improve generalization [60,61].

3.6. Hyperparameters

A deep learning model requires the modification of various hyperparameters in order to improve the results, and these largely depend on the dataset and other hyperparameters. The backpropagation algorithm involves two parameters in updating the weights during training which are: the learning rate (α) and momentum factor (β).

The initial learning rate α is one of the most important hyperparameters; too small a learning rate makes the network learn slowly, and too large a learning rate possibly leads to oscillation preventing the error falling below a certain value.

The momentum factor β is believed to make the learning procedure more stable and accelerate convergence in shallow regions of the error function, which in practice does not always happen [62].

The extra term introduced by the three-term back propagation algorithm, called the proportional factor (γ), speeds up the weight adjustment process by increasing the convergence rate and decreasing learning stalls of the algorithm.

The best choice of these parameters depends on the problem which often requires a trial and error process before a suitable choice is found [63]. Having run the experiments a number of times based on trial an error, the optimum values of the parameters were achieved which trained the network and output good results.

3.7. Architecture of the Deep Learning Model

The deep learning model was designed using the 288 bands as input to the network. Each pixel is taken as an independent input to the network. In this research study, we do not consider the spatial information. The network has four hidden layers each containing 30 nodes and one output corresponding to the abundance of hydrocarbon. The network was trained using the ground truth abundances for the different mixtures, as detailed in Table 4.

The data was randomly divided into 3 categories, namely: training, validation, and test sets. The training set is used to fit the parameters of the deep learning model, the test set (unseen data) is used to investigate the predictive power of the model while the validation set is used to avoid overfitting using the cross-validation algorithm.

The cross-validation algorithm avoids overfitting because the training sample is independent of the validation sample [64]. The size of the data sets depended on the soils' absorption level during the experiment (i.e., when a local shallow pool was formed). Only image pixels corresponding to data from inside the Petri dish were considered. Moreover, for each scanned image, 1000 pixels were randomly selected. Thus the data sets ranged between 5000 pixels × 288 bands (where five mixture types were available) to 10,000 pixels × 288 bands (for samples with ten possible mixtures). The size of the data sets and number of mixtures used for the experiments (see Table 1) are summarized in Table 3. Subsets of the hyperspectral data were fed into the network as follows: 80% of the data were randomly selected for training the network, 10% were used to test the network and 10% were used for cross-validation.

We compared two neural network architectures one with and another without dropout, respectively. This is to prove the dropout's efficiency to improve the generalization capabilities of the neural network.

The network used a sigmoid activation function which was applied to the hidden and output nodes.

The deep learning abundance estimation experiments were conducted to obtain optimum hyperparameters in order to achieve maximum accuracy in estimating the amount of HCs in each soil mixture type. The ground truth, or known abundances from the sample preparation, were used as class labels (targets) to train the network for the abundance estimation. These ground truth abundances were estimated based on the HC type in each data set as detailed in Table 4 and depend on the density of each HC.

Table 4. Ground truth abundances (expressed in wt%) for the different mixtures corresponding to 2 mL, 4 mL, 5 mL, 10 mL, 15 mL, 20 mL, 25 mL, 30 mL, 35 mL, 40 mL, and 45 mL of HC , respectively.

Corresponding mixtures (mL)	Petrol	Diesel	Biodiesel	Ethanol
2.0	0.02	0.023	0.034	0.029
4.0	0.055	0.063	0.065	0.059
5.0	0.068	0.08	0.08	0.073
10.0	0.128	0.148	0.149	0.136
15.0	0.181	0.206	0.208	0.191
20.0	0.227	0.258	0.260	0.240
25.0	0.269	0.303	0.305	0.283
30.0	0.340	0.342	0.345	0.321
35.0	38.1	–	–	–
40.0	42.0	–	–	–
45.0	46.2	–	–	–

All the experiments were conducted with the learning rate of α set to 0.01, β set to 0.5, and γ set to 0.1, which allowed convergence of the objective function at a high rate. The algorithm was run iteratively with 20 epochs.

Moreover, in order to find the optimum level of dropout, the models were trained using the three-term backpropagation algorithm with different ranges of dropout (10–50%).

4. Results

In this section, we present the results obtained from the deep learning model demonstrating the abundance estimation of the different HCs. Results are presented to demonstrate the effectiveness of dropout in the model in terms of generalization capabilities. We also show the accuracy of the proposed method compared to the hybrid spectral unmixing method and DL models trained with conventional backpropagation with and without dropout, respectively. Results with laboratory and remote sensed data are presented. The algorithms were implemented using MatLab 2018b. The experiments were carried out on an LG desktop with Intel (R) core $(TM)^2$ Duo CPU 3.00 GHZ processor 8.00 GB RAM.

4.1. Experiment with Laboratory Data

The reflectance spectra of different soil samples with 15% hydrocarbon concentration mixture are shown in Figure 5 showing specific absorption at around 1700 μm and 2300 μm, respectively

(**a**) Clay with different hydrocarbon mixtures

(**b**) Clay loam with different hydrocarbon mixtures

(**c**) Sandy loam with different hydrocarbon mixtures

(**d**) Sandy clay loam with different hydrocarbon mixtures
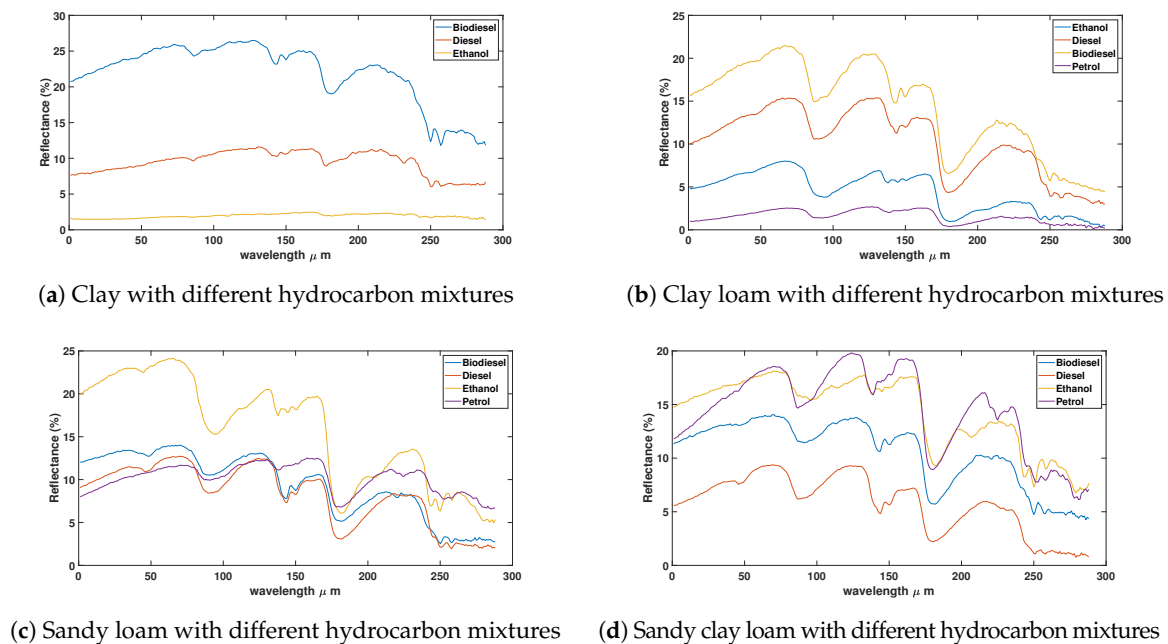
Figure 5. Spectral reflectance of different soils and 15% hydrocarbon concentration mixtures.

The ground truth abundances in Table 4 were used to estimate the amount of hydrocarbon used in the experiment. The abundances were calculated based on the density of the different hydrocarbon types. The aim is to quantify the percentage or amount of HC in each pixel using deep learning and hybrid spectral unmixing method (using abundance estimation). This was calculated based on the saturation level of the different hydrocarbons as shown in Table 1.

To demonstrate the ability of the proposed deep learning model to generalize on unseen data, Table 5 displays the results obtained from the test sets with and without dropout, respectively.

Table 5. Mean square error of the deep learning model on unseen data with and without dropout, respectively.

Dataset	Test Set with Dropout	Test Set without Dropout
Clay biodiesel	7.11×10^{-3}	9.1×10^{-3}
Clay diesel	1.16×10^{-3}	6.9×10^{-3}
Clay ethanol	8.26×10^{-4}	6.2×10^{-3}
Clay loam biodiesel	7.62×10^{-4}	1.4×10^{-3}
Clay loam diesel	2.20×10^{-3}	3.3×10^{-3}
Clay loam ethanol	8.80×10^{-4}	7.1×10^{-3}
Clay loam petrol	2.50×10^{-3}	8.1×10^{-3}
Sandy loam biodiesel	8.00×10^{-4}	1.2×10^{-3}
Sandy loam diesel	1.20×10^{-3}	9.8×10^{-3}
Sandy loam ethanol	4.01×10^{-3}	8.4×10^{-3}
Sandy loam petrol	6.71×10^{-3}	9.3×10^{-3}
Sandy clay loam biodiesel	8.25×10^{-4}	3.5×10^{-3}
Sandy clay loam diesel	8.84×10^{-4}	9.8×10^{-3}
Sandy clay loam ethanol	2.10×10^{-3}	7.5×10^{-3}
Sandy clay loam petrol	5.13×10^{-3}	6.1×10^{-3}

The experimental process was repeated with different dropout ratios on the hidden layers of 10%, 20%, 30%, 40%, and 50% respectively. Results demonstrate both the training and validation accuracy of the network. Tables 6–9 illustrate the mean square error of the proposed method with the different dropout ratios. It is noted that in all cases the error is 10 times lower for dropout ratio of 40% than for 50%.

Then again when the error drops significantly for dropout ratio 20%. However, when it is further reduced to 10%, the error increases. The 20% dropout is adopted subsequently in the rest of the experiments.

Table 6. Mean Square Error (MSE) of the deep learning model for Clay Loam datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC Types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio-diesel MSE	3.5×10^{-4}	0.22×10^{-4}	0.69×10^{-4}	6.5×10^{-4}	72×10^{-4}
Diesel MSE	0.31×10^{-4}	0.25×10^{-4}	0.52×10^{-4}	7.3×10^{-4}	59×10^{-4}
Ethanol MSE	3.1×10^{-4}	0.21×10^{-4}	5.6×10^{-4}	74×10^{-4}	76×10^{-4}
Petrol MSE	2.9×10^{-4}	0.22×10^{-4}	0.73×10^{-4}	6.7×10^{-4}	72×10^{-4}

Table 7. Mean square error of the deep learning model for Clay datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC Types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio-diesel MSE	2.7×10^{-4}	2.3×10^{-4}	1.9×10^{-4}	5.1×10^{-4}	77×10^{-4}
Diesel MSE	1.7×10^{-4}	0.35×10^{-4}	1.7×10^{-4}	3.2×10^{-4}	76×10^{-4}
Ethanol MSE	1.6×10^{-4}	2.2×10^{-4}	2.8×10^{-4}	4.3×10^{-4}	83×10^{-4}

Table 8. Mean square error of the deep learning model for Sandy Clay Loam datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC Types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio-diesel MSE	2.7×10^{-4}	0.33×10^{-4}	0.61×10^{-4}	3.4×10^{-4}	86×10^{-4}
Diesel MSE	2.7×10^{-4}	2.0×10^{-4}	5.3×10^{-4}	6.6×10^{-4}	94×10^{-4}
Ethanol MSE	3.4×10^{-4}	2.2×10^{-4}	4.3×10^{-4}	7.2×10^{-4}	88×10^{-4}
Petrol MSE	2.2×10^{-4}	1.3×10^{-4}	3.6×10^{-4}	4.7×10^{-4}	83×10^{-4}

Table 9. Mean square error of the deep learning model for Sandy Loam datasets with different hydrocarbon types and different Dropout (DO) ratios.

HC Types	DO 10%	DO 20%	DO 30%	DO 40%	DO 50%
Bio-diesel MSE	3.6×10^{-4}	3.4×10^{-4}	1.6×10^{-4}	7.3×10^{-4}	72×10^{-4}
Diesel MSE	2.7×10^{-4}	0.2×10^{-4}	0.66×10^{-4}	3.1×10^{-4}	68×10^{-4}
Ethanol MSE	2.5×10^{-4}	0.2×10^{-4}	7.3×10^{-4}	5.8×10^{-4}	86×10^{-4}
Petrol MSE	0.68×10^{-4}	1.5×10^{-4}	2.7×10^{-4}	6.8×10^{-4}	81×10^{-4}

Results of the experiments are shown in Figures 6 and 7 for individual soil types contaminated with different HCs to confirm the accuracy of the method. Figure 6 shows the mean square error during training, and demonstrates the network's ability to converge rapidly with low numbers of epochs. The plots in Figure 7 show the model's estimated output and target output for 4 different combinations. It is observed that the DL model quantifies correctly all the different HC abundances with low error.

(**a**) Sandy cay loam–biodiesel

(**b**) Clay loam–petrol

(**c**) Sandy loam–diesel

(**d**) Clay-biodiesel
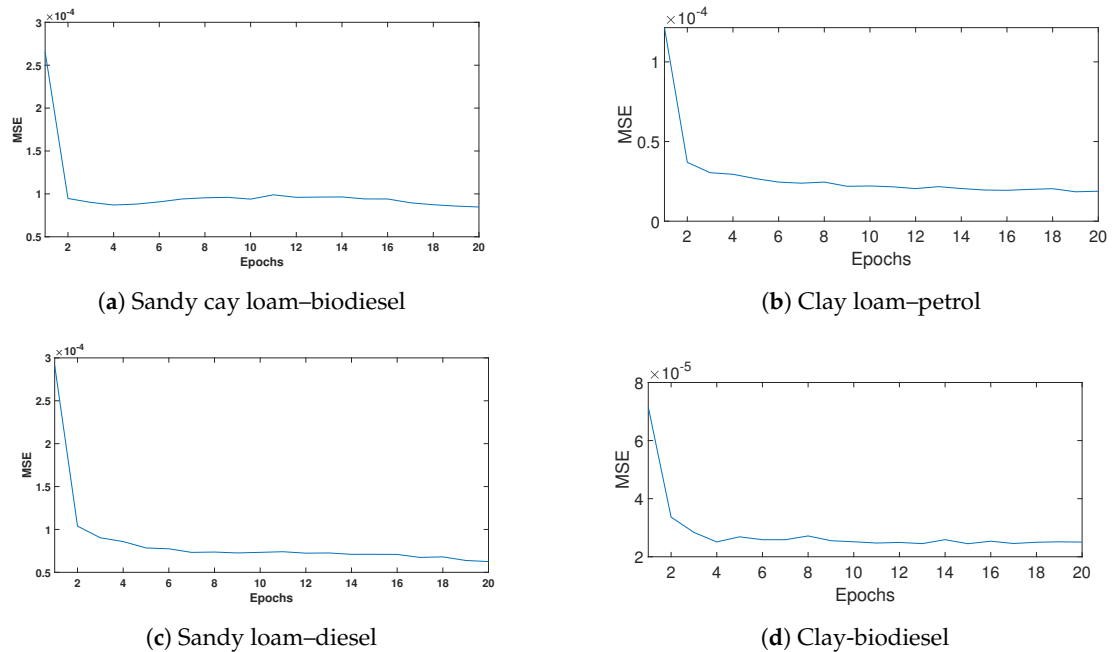
Figure 6. Mean square error of different soils contaminated with different HC contents.

(**a**) Sandy clay loam–biodiesel

(**b**) Clay loam–petrol

(**c**) Sandy loam–diesel

(**d**) Clay–bio-diesel

Figure 7. Neural network estimated output and target output of different soils contaminated with different HC contents.

From the results obtained, it is noted that the proposed method was able to generalize on unseen testing and validation data with high prediction accuracy. We observed a similar trend on all the datasets used for the experiment which indicates a reduction in the error rate and a high convergence rate.

To demonstrate the effectiveness of the proposed method, a deep learning model trained with a conventional backpropagation algorithm was similarly used to quantify the HC abundances; first without dropout, and then with 20 % dropout to train the networks. For fair comparison, the same network structure was used and these include: number of layers, number of nodes for each layer, range of initial values, and learning rate. Another comparison was conducted with the hybrid spectral unmixing method for switching between linear and nonlinear methods [21]. Hybrid spectral unmixing uses a neural network to determine the most appropriate method among a set of linear and non linear unmixing method for each pixel in the scene. Specifically, we used Vertex Component Analysis (VCA) [65], Fully Constrained Least Square Method (FCLS) [66], Generalized Bilinear Mixing Model (GBM) [67], and the Polynomial Post Nonlinear Mixing Model (PPNMM) [68]. This means that the hybrid switch method selects the best of these four methods for each pixel.

From the results obtained, it may be observed that our proposed method outperforms the hybrid switch method and conventionally trained networks with the closest estimate from the ground truth values as demonstrated in Tables 10 and 11.

4.2. Soil Continuity Experiments

In this research, four different mixtures of soil were created and HC were added in discrete steps. However, in real life situations, both HC and soil levels of given samples are cont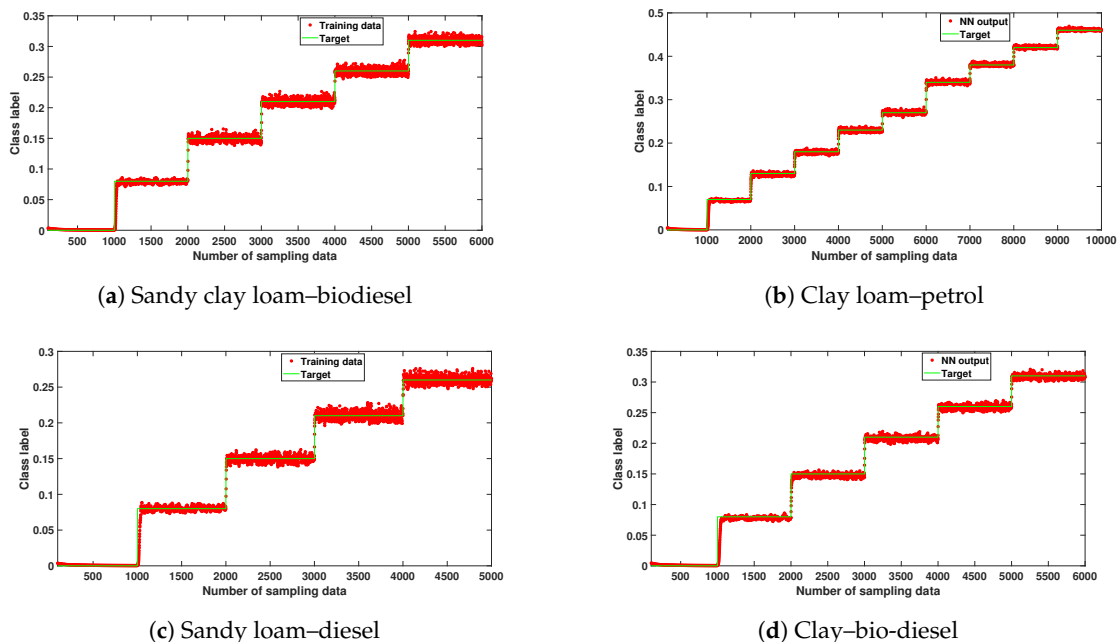inuous rather than discrete. Therefore, in order to simulate a more realistic scenario, several strategies were explored. The first strategy was to create a generic model with all soils combined as opposed to separate models for each soil as in the previous experiments. It is noted that the soils were prepared and mixed manually and contained grains of different size (e.g., clay and sand mixture). By feeding the DL network with all types of soils, differences in the soil composition would appear from pixel to pixel. DLs were created including all four different soil mixtures (Clay, Clay–loam, Sandy loam, and Sandy clay–loam) rather than individually. Using the same architecture of the deep learning model, 80% of the resultant data was used to train the model, 10% was used as test sample and the remaining 10% was used for cross-validation. This was conducted to validate the network's ability to estimate the amount of HC regardless of the soil type and allowing for different soil types. The results were in the same range as for the different HC. Table 11 summarizes the results obtained for biodiesel. It is noted the training MSE was in the same range as the individual model although the number of epochs required increased to 178. Average MSE for the individual models are shown in brackets for comparison purposes. The generalization MSE on the training data is higher than in the individual models. However, it is noticed that this data is more complex as the individual models as it contains four different soil types. In the work of the authors of [16], an individual model per soil type was recommended as the generic models degraded the responses. We believe that with further tuning of the hyperparameters, improved results could have been achieved. However, this is out of the scope of this paper.

In order to simulate a more realistic scenario, and following a similar approach presented in the work by the authors of [16], noise was added to the data to simulate continuous spectra values instead of discrete and also to evaluate the noise rejection of the models. Here, the datasets were corrupted with Random Gaussian noise with signal-to-noise ratio (SNR) ranging from 10 to 40 dB. The model shows similar performance even with low SNR, showing good noise rejection and accurate prediction at different ranges of SNR. For SNR lower than 20 dB, the generalization error deteriorates slightly. Initially, we added noise to the test data but not to the training, to simulate training with discrete samples and testing with continuous samples. The performance deteriorates for SNR = 10 dB, while for higher SNR, it is very stable showing good adaptation of the model to continuous spectra.

Table 10. Estimated hydrocarbon abundance predicted by the proposed method, compared with the hybrid switch method, conventional neural network, and conventional neural network with dropout (DO) for Clay and Clay Loam (CL) mixtures. A comparative summary showing the average estimation error in percentage is included for each mixture type.

Mixtures	Reference	Proposed Method	Hybrid Switch Method	Conventionally Trained NN with DO (0.2)	Conventionally Trained NN
Clay–biodiesel	0	0.002	0.03	0.61	0.69
Clay–biodiesel	3	3.4	3.7	3.9	4.5
Clay–biodiesel	6	6.5	6.9	7.3	7.9
Clay–biodiesel	8	8.4	9.4	9.8	9.9
Clay–biodiesel	14.9	15.3	16.9	18.0	18.9
Clay–biodiesel	20.8	21.4	22.6	17.3	17.9
Clay–biodiesel	26.0	26.8	28.3	29.9	30.3
Clay–biodiesel	30.5	31.9	32.2	35.1	38.6
Average error (%)		2	10	17	20
Clay–diesel	0	0.003	0.04	1.06	1.79
Clay–diesel	3	3.3	3.7	4.6	4.9
Clay–diesel	6	6.5	6.3	6.9	7.21
Clay–diesel	8	8.1	7.6	4.8	3.7
Clay–diesel	14.8	15.3	13.0	19.1	19.8
Clay–diesel	20.6	21.4	18.5	23.8	25.5
Clay–diesel	25.8	25.6	23.6	29.1	29.4
Clay–diesel	30.3	30.6	32.2	36.1	37.7
Average error (%)		2	14	20	24
Clay–ethanol	0	0.004	0.05	1.41	2.04
Clay–ethanol	2	2.7	2.3	3.4	3.9
Clay–ethanol	5	5.5	5.3	6.01	6.63
Clay–ethanol	7.3	7.6	8.2	4.9	4.1
Clay–ethanol	13.6	14.1	15.6	10.4	9.9
Clay–ethanol	19.1	19.6	18.1	22.7	22.9
Clay–ethanol	24	24.9	22.9	27.7	28.3
Clay–ethanol	28.3	28.9	27.7	32.8	33.7
Average error (%)		3	9	19	23
CL–biodiesel	0	0.002	0.16	0.90	1.63
CL–biodiesel	3	3.2	3.6	4.5	4.9
CL–biodiesel	6	6.4	6.6	7.1	7.8
CL–biodiesel	8	8.5	7.7	9.9	10.8
CL–biodiesel	14.9	15.5	13.6	17.9	18.3
CL–biodiesel	20.8	21.6	18.7	24.6	24.9
CL–biodiesel	26.0	26.4	24.1	29.6	29.9
Average error (%)		2	11	17	21

Table 10. *Cont.*

Mixtures	Reference	Proposed Method	Hybrid Switch Method	Conventionally Trained NN with DO (0.2)	Conventionally Trained NN
CL-diesel	0	0.003	0.001	0.76	1.88
CL-diesel	3	3.3	3.1	3.7	4.3
CL-diesel	6	5.8	6.6	7.2	7.8
CL-diesel	8	8.4	10.4	11.6	11.9
CL-diesel	14.8	14.3	12.8	18.7	19.0
CL-diesel	20.6	21.1	22.1	24.4	25.3
CL-diesel	25.8	26.8	27.7	28.2	30.7
Average error (%)		3	13	26	31
CL-ethanol	0	0.002	0.05	1.22	1.79
CL-ethanol	2	2.4	2.7	3.6	3.9
CL-ethanol	5	5.2	5.4	5.9	6.4
CL-ethanol	7.3	7.6	6.7	9.6	9.9
CL-ethanol	13.6	14.4	11.6	16.7	17.0
CL-ethanol	19.1	19.7	18.5	22.6	22.9
CL-ethanol	24	24.6	25.9	26.7	27.5
CL-ethanol	28.3	28.7	29.9	30.7	31.2
Average error (%)		2	11	17	21
CL-petrol	0	0.002	0.006	0.36	1.40
CL-petrol	2	2.4	2.2	2.9	3.3
CL-petrol	5	5.1	5.5	6.3	6.9
CL-petrol	6.8	5.9	7.9	8.4	8.9
CL-petrol	12.8	12.9	13.0	11.0	9.9
CL-petrol	18.1	18.4	17.4	16.8	15.6
CL-petrol	22.7	23.1	24.4	19.8	18.5
CL-petrol	26.9	27.4	27.8	24.6	23.9
CL-petrol	34	34.9	35.6	29.7	29.8
CL-petrol	38.1	38.5	39.6	36.6	36.9
CL-petrol	42	42.4	44.4	38.3	36.1
CL-petrol	46.2	46.9	45.2	43.5	43.7
Average error (%)		1	5	13	16

Table 11. Soil continuity experiments. Mean square error of the bio-diesel deep learning model using generic models and individual models with added noise. Training and testing results are shown.

Dataset	Traning Data	Test Data
Biodiesel with generic model	7.2238×10^{-4} (1.525×10^{-4})	6.8×10^{-3} (2.374×10^{-3})
Biodiesel with added noise		
SNR (dB)	Traning data	Test data
40	8.2117×10^{-4}	8.9821×10^{-4}
30	8.2594×10^{-4}	8.9333×10^{-4}
20	9.393×10^{-4}	0.001
10	9.671×10^{-4}	0.001
Biodiesel with added noise on testing data		
SNR (dB)	Training data	Test data
40	6.9997×10^{-4}	9.3811×10^{-4}
30	6.9997×10^{-4}	9.0321×10^{-4}
20	6.9997×10^{-4}	9.0657×10^{-4}
10	6.9997×10^{-4}	0.0012

4.3. Experiment with Remote Sensed Data

A remote sensed data captured by an airborne system, adjusted to work under stationary condition in the field, was used to validate our proposed method. This dataset contains soils contaminated with different levels of hydrocarbon (between 0 to 10 *wt%* in steps of 1 *wt%*) that were acquired at three different locations (Hamra, Kokhav, and Evrona) with a Hyper-Cam LW instrument. Each pixel responses are captured by 88 spectral bands in the spectral range of 8 to 12 µm with spectral resolution of 0.25 cm^{-1}. The experimental protocol, data capturing, and preprocessing of these datasets are fully described in the work by the authors of [16].

Each one of the 3 datasets was independently trained with a DL network. For each network, 75% of the samples were randomly selected for training and 25% were used for testing. These parameters were selected similar to the work by the authors of [16] in order to provide a fair comparison to the results they presented. A dropout ratio of 20% was used and all other hyperparameters were left as in our previous configuration. Results are presented as MSE for each dataset in Table 12. It was noted that our results surpassed in term of prediction accuracy the ones presented in the literature for these datasets. Moreover, the results show good generalization capabilities.

Table 12. Summary of the evaluation on the 3 different soils using the deep learning model trained with and without dropout.

Soil Type	MSE on Training Set	MSE on the Test Set
Hamra	0.48×10^{-4}	1×10^{-4}
Evona	0.78×10^{-4}	3.5×10^{-4}
Kokhav	2×10^{-4}	3×10^{-4}

Results on all 3 datasets shows that our proposed DL method achieved acceptable results with consistent MSE values as shown in Figure 8c for both training and generalization.
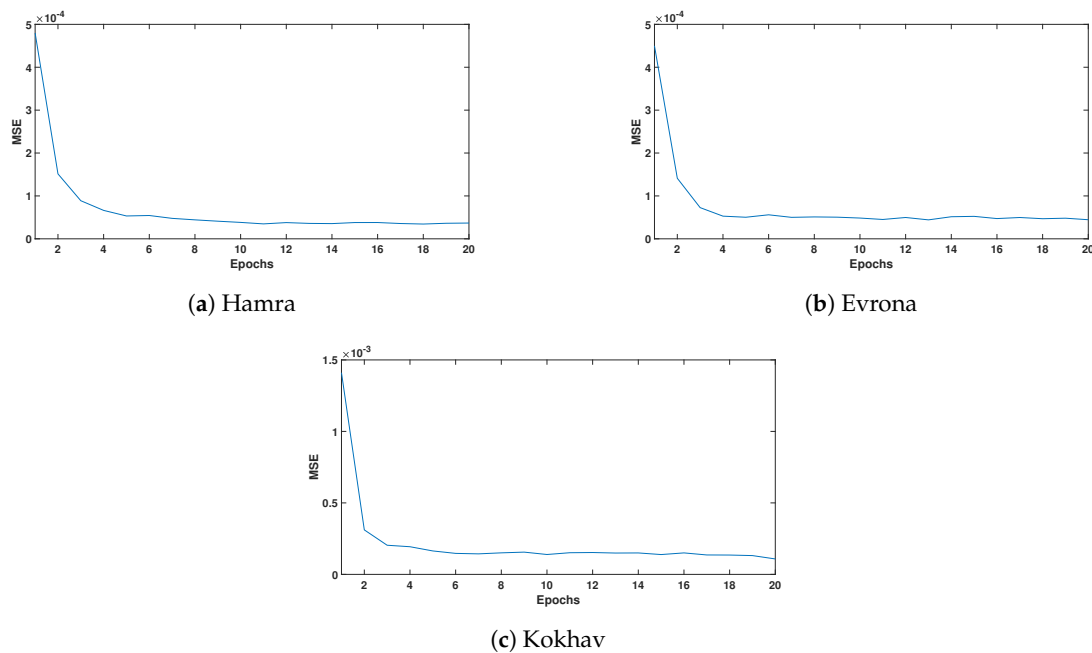
(**a**) Hamra

(**b**) Evrona

(**c**) Kokhav

Figure 8. Mean square error of the 3 different soils contaminated with different HC contents.

5. Discussion

In this study, controlled hyperspectral datasets were used to assess the capabilities of the deep learning model to predict and quantify the amount of HC spills on different soil types. The deep learning approach was trained using a three-term backpropagation algorithm with dropout technique. The deep learning model designed for this experiment utilizes a sigmoid activation function and dropout of 20% in all the hidden layers of the architecture in order to avoid overfitting. Another advantage of utilizing dropout is its ability to generalize.

The main aim of the three-term backpropagation algorithm was to reduce the number of training epochs and maintain the system's stability during training. Our proposed method was able to estimate the amount of HCs in each dataset with high accuracy using a low number of epochs. The network was able to achieve an average of 2.2×10^{-4} mean square error on an average of 18 epochs as shown in Tables 5–8 and Figure 6.

Dropout plays an important role in the architecture of the proposed deep learning model by improving the performance of the model and avoided overfitting on the training data. This can be proven from Table 5, where the results show the ability of the model to generalize on unseen data with good accuracy.

From the results obtained, it may be observed that hydrocarbon can be estimated even at low levels as shown in Tables 10 and 11.

Tables 10 and 11 summarize the abundance estimation of the quantity of HCs in the different mixture types using our proposed method, the hybrid spectral unmixing method, and the conventionally trained NN with and without dropout. For instance, if we observe the first mixture from Table 10 (biodiesel mixed with clay), it is noted that for pure clay sample (reference 0% HC), all methods provide very close estimate. The second raw presents the results for the same mixture with 8% of biodiesel, 92% of clay. Our proposed method estimated 8.4% of biodiesel with the hybrid switch method at 9.4%, while the neural network trained with a standard backpropagation algorithm with and without dropout estimated at 9.8% and 9.9%, respectively. Similar results were obtained for all the different samples as shown in Tables 10 and 11.

In addition, the soil's properties such as grain size and texture can lead to a variation in the absorption level, and thus the difference in detection of the different hydrocarbon types. The hybrid switch method was also able to estimate the amount of hydrocarbon spills with reasonable

accuracy unlike the conventionally trained neural network which has low accuracy compared to the proposed method.

In the work by the authors of [69], it was revealed that deep learning requires large amounts of training data to achieve optimal performance and good generalization with minimum error. However, despite the size of the datasets used, the training process is faster with our proposed method, attributed to the use of the three-term backpropagation algorithm for training and the use of cross-validation with dropout. The authors of [55] reported that large neural networks trained in the standard way tend to overfit on small datasets. To see if dropout can improve this condition, we ran the experiment on all the datasets and varied the dropout ratio as shown in Tables 6–9. From the results obtained, the error rate is relatively low when the dropout ratio is between 10% and 30%, with a slight increase when the ratio is set above 50%. Therefore, it can be concluded that dropout ratio between 10% and 30% provides an acceptable prediction estimation and could be used in any dataset for estimation analysis.

The proposed deep learning method was further validated on field datasets. In particular, the Hamra soils produced a better results with lower MSE of 0.48×10^{-4} compared to the Kokhav and Evrona soils as shown in Figure 8c. This shows a similar trend to what was obtained in the work by the authors of [16]. Although the deep learning model did not estimate well abundances between 1 and 3 $wt\%$, this could be attributed to the fact that the data was obtained within the Longwave Infrared Region (LWIR), which has a different spectral range with the laboratory controlled data used in this research. Nevertheless, our proposed deep learning model perform better with all 3 datasets in terms of MSE accuracy.

Recent research in deep learning suggests that a large dataset is required for training remote sensing data, which is a major drawback. This was not the case for our proposed model because the three-term backpropagation algorithm allows it to train faster using a minimum number of epochs to converge. The training process was relatively fast compared to standard networks where parameter updates are noisy in architectures with dropout. Learning takes an average of 28.16 seconds with our proposed deep learning model compared to the conventionally trained neural networks which takes an average of 300.68 seconds to learn. The data processing-to-end-product time of our proposed method is relatively more time-consuming compared to traditional spectral unmixing method, which is approximately 35.45 seconds compared to the hybrid spectral unmixing method which has an average of 24.64 seconds. This could be attributed to the number of parameters and training in the deep learning method compared to the spectral unmixing method.

However, we anticipate that including a wider range of data such as field HC leaks, will require larger datasets and longer training to account for the variability of the data. Nevertheless, the proposed methodology has proven to decrease significantly both the learning time and sample data required to achieve accurate generalization.

6. Conclusions

In this paper, we developed a deep learning approach to accurately estimate hydrocarbon spills on different soil samples measured using imaging spectroscopy. The deep learning model was trained using a three-term backpropagation algorithm with dropout. The aim was to improve the accuracy of the model, avoid overfitting, and converge faster.

Standard backpropagation algorithms build co-adaptation, which work well on training data but the network does not generalize to unseen data. Dropout neutralizes these co-adaptations by improving the network's performance, thus enabling it to generalize. The choice of dropout ratio used in any neural network depends on the type of dataset and application.

The three-term backpropagation algorithm improves the network's ability to train faster and overcome local minima when compared to conventional backpropagation algorithms. The effectiveness of the deep learning model was verified when tested on the datasets containing different soil samples mixed with different hydrocarbon types to estimate the amount of hydrocarbon spills in each dataset. The datasets were acquired using a Hyspex 384 SWIR camera under laboratory conditions.

Many studies have shown the ability to detect HC using spectroscopy in the SWIR region. The results of the experiments consistently show that the proposed method provides high prediction accuracy with low error even for amounts of HC as low as 6.8%. Therefore, it can be concluded that the three-term backpropagation algorithm with dropout significantly improves the model's operation.

The deep learning model was further applied on three datasets acquired with an airborne LWIR camera in field conditions which proved the effectiveness of the proposed method and its applicability in real world scenarios.

Satellite and airborne hyperspectral data with ground truth are expensive, thus making it difficult to obtain; but with the emergence of new lightweight sensors mounted on Unmanned Aerial Vehicles (UAV), the potential application of this research is very large. It is noted that data acquired in field conditions could be affected by several limitations such as variable illumination, atmospheric conditions, and sensor sampling distance which could affect the accuracy in using such dataset. However, in the work by the authors of [70], the correlation between datasets obtained under laboratory and outdoor conditions was demonstrated. Thus, a neural network could be trained with laboratory data and validated using remote UAV or airborne data. The information provided in this research study can be used as a guide to understand the potential and limitations of a hyperspectral sensor for HC abundance estimation.

Future work will develop networks that are able to classify and identify different types of HC incorporating also spatial information using convolutional neural networks.

References

1.　United Nations Environment Programme (UNEP). *Environmental Assessment of Ogoniland*; United Nations Environment Programme: Nairobi, Kenya, 2011; pp. 1–262.

2.　Raksuntorn, N.; Du, Q. A new linear mixture model for hyperspectral image analysis. In Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008; Volume 3, pp. III-258–III-261.

3.　Eweje, G. Environmental costs and responsibilities resulting from oil exploitation in developing countries: The case of the Niger Delta of Nigeria. *J. Bus. Ethics* **2006**, *69*, 27–56. [CrossRef]

4.　Latimer, J.S.; Zheng, J. The sources, transport, and fate of PAHs in the marine environment. *PAHs Ecotoxicol. Perspect.* **2003**, 9. [CrossRef]

5.　Okparanma, R.N.; Mouazen, A.M. Determination of total petroleum hydrocarbon (TPH) and polycyclic aromatic hydrocarbon (PAH) in soils: A review of spectroscopic and nonspectroscopic techniques. *Appl. Spectrosc. Rev.* **2013**, *48*, 458–486. [CrossRef]

6.　Scafutto, R.D.M.; de Souza Filho, C.R.; de Oliveira, W.J. Hyperspectral remote sensing detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for onshore exploration and monitoring. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 146–157. [CrossRef]

7.　Salem, F.; Kafatos, M. Hyperspectral image analysis for oil spill mitigation. Paper presented at the 22nd Asian Conference on Remote Sensing, Taipei, Taiwan, 3–7 October 2011; Volume 5, p. 9.

8.　Okparanma, R.N.; Mouazen, A.M. Visible and near-infrared spectroscopy analysis of a polycyclic aromatic hydrocarbon in soils. *Sci. World J.* **2013**, *2013*, 160360. [CrossRef]

9. Chakraborty, S.; Weindorf, D.C.; Li, B.; Ali, M.N.; Majumdar, K.; Ray, D. Analysis of petroleum contaminated soils by spectral modeling and pure response profile recovery of n-hexane. *Environ. Pollut.* **2014**, *190*, 10–18. [CrossRef]

10. Lammoglia, T.; de Souza Filho, C.R. Spectroscopic characterization of oils yielded from Brazilian offshore basins: Potential applications of remote sensing. *Remote Sens. Environ.* **2011**, *115*, 2525–2535. [CrossRef]

11. Fingas, M.; Brown, C. Oil spill remote sensing. In *Earth System Monitoring*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 337–388.

12. Clasen, A.; Somers, B.; Pipkins, K.; Tits, L.; Segl, K.; Brell, M.; Kleinschmit, B.; Spengler, D.; Lausch, A.; Förster, M. Spectral unmixing of forest crown components at close range, airborne and simulated Sentinel-2 and EnMAP spectral imaging scale. *Remote Sens.* **2015**, *7*, 15361–15387. [CrossRef]

13. Scafutto, R.D.M.; de Souza Filho, C.R. Quantitative characterization of crude oils and fuels in mineral substrates using reflectance spectroscopy: Implications for remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *50*, 221–242. [CrossRef]

14. Pelta, R.; Ben-Dor, E. Assessing the detection limit of petroleum hydrocarbon in soils using hyperspectral remote-sensing. *Remote Sens. Environ.* **2019**, *224*, 145–153. [CrossRef]

15. Webster, G.T.; Soriano-Disla, J.M.; Kirk, J.; Janik, L.J.; Forrester, S.T.; McLaughlin, M.J.; Stewart, R.J. Rapid prediction of total petroleum hydrocarbons in soil using a hand-held mid-infrared field instrument. *Talanta* **2016**, *160*, 410–416. [CrossRef]

16. Pelta, R.; Ben-Dor, E. An Exploratory Study on the Effect of Petroleum Hydrocarbon on Soils Using Hyperspectral Longwave Infrared Imagery. *Remote Sens.* **2019**, *11*, 569. [CrossRef]

17. Okparanma, R.N.; Coulon, F.; Mayr, T.; Mouazen, A.M. Mapping polycyclic aromatic hydrocarbon and total toxicity equivalent soil concentrations by visible and near-infrared spectroscopy. *Environ. Pollut.* **2014**, *192*, 162–170. [CrossRef]

18. Schwartz, G.; Ben-Dor, E.; Eshel, G. Quantitative analysis of total petroleum hydrocarbons in soils: Comparison between reflectance spectroscopy and solvent extraction by 3 certified laboratories. *Appl. Environ. Soil Sci.* **2012**, *2012*, 751956. [CrossRef]

19. Gholizadeh, A.; Saberioon, M.; Ben-Dor, E.; Boruuvka, L. Monitoring of selected soil contaminants using proximal and remote sensing techniques: Background, state-of-the-art and future perspectives. *Crit. Rev. Environ. Sci. Technol.* **2018**, *48*, 243–278. [CrossRef]

20. Pabón, R.E.C.; de Souza Filho, C.R. Spectroscopic characterization of red latosols contaminated by petroleum-hydrocarbon and empirical model to estimate pollutant content and type. *Remote Sens. Environ.* **2016**, *175*, 323–336. [CrossRef]

21. Ahmed, A.M.; Duran, O.; Zweiri, Y.; Smith, M. Hybrid spectral unmixing: Using artificial neural networks for linear/nonlinear switching. *Remote Sens.* **2017**, *9*, 775. [CrossRef]

22. Xu, M.; Zhang, L.; Du, B.; Zhang, L.; Fan, Y.; Song, D. A mutation operator accelerated quantum-behaved particle swarm optimization algorithm for hyperspectral endmember extraction. *Remote Sens.* **2017**, *9*, 197. [CrossRef]

23. Uezato, T.; Murphy, R.J.; Melkumyan, A.; Chlingaryan, A. A novel spectral unmixing method incorporating spectral variability within endmember classes. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2812–2831. [CrossRef]

24. Weeks, A.R. *Fundamentals of Electronic Image Processing*; SPIE Optical Engineering Press: Bellingham, WA, USA, 1996.

25. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57. [CrossRef]

26. Drumetz, L.; Tochon, G.; Chanussot, J.; Jutten, C. Estimating the number of endmembers to use in spectral unmixing of hyperspectral data with collaborative sparsity. In Proceedings of the Submitted to the 13th International Conference on Latent Variable Analysis and Signal Separation (LVA-ICA), Grenoble, France, 21–23 February 2017; pp. 1–10.

27. Hapke, B. Bidirectional reflectance spectroscopy: 1. Theory. *J. Geophys. Res. Solid Earth* **1981**, *86*, 3039–3054. [CrossRef]

28. Dobigeon, N.; Tourneret, J.Y.; Richard, C.; Bermudez, J.; McLaughlin, S.; Hero, A.O. Nonlinear unmixing of hyperspectral images: Models and algorithms. *Signal Process. Mag. IEEE* **2014**, *31*, 82–94. [CrossRef]

29. Halimi, A.; Altmann, Y.; Buller, G.S.; McLaughlin, S.; Oxford, W.; Clarke, D.; Piper, J. Robust unmixing algorithms for hyperspectral imagery. In Proceedings of the Sensor Signal Processing for Defence (SSPD), Edinburgh, UK, 22–23 September 2016; pp. 1–5.

30. Liu, J.; Luo, B.D.S.C.J. Exploration of planetary hyperspectral images with unsupervised spectral unmixing: A case study of planet Mars. *Remote Sens.* **2018**, *10*, 737. [CrossRef]

31. Zhang, X.; Li, C.; Zhang, J.; Chen, Q.; Feng, J.; Jiao, L.; Zhou, H. Hyperspectral Unmixing via Low-Rank Representation with Space Consistency Constraint and Spectral Library Pruning. *Remote Sens.* **2018**, *10*, 339. [CrossRef]

32. Kang, S.; Xiurui, G.; Hairong, T. A new target detection method using nonlinear PCA for hyperspectral imagery. *Bull. Surv. Mapp.* **2015**, *1*, 105–108.

33. Feng, D.C.; Chen, F.; Wen-Li, X. Detecting local manifold structure for unsupervised feature selection. *Acta Autom. Sin.* **2014**, *40*, 2253–2261. [CrossRef]

34. Ahmed, A.; Duran, O.; Zweiri, Y.; Smith, M. Application of hybrid switch method to quantify oil spills. In Proceedings of the Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 23–26 September 2018.

35. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2014–2039. [CrossRef]

36. Feng, J.; Rogge, D.; Rivard, B. Comparison of lithological mapping results from airborne hyperspectral VNIR-SWIR, LWIR and combined data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 340–353. [CrossRef]

37. Harris, J.; Rogge, D.; Hitchcock, R.; Ijewliw, O.; Wright, D. Mapping lithology in Canada's Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering. *Can. J. Earth Sci.* **2005**, *42*, 2173–2193. [CrossRef]

38. Rogge, D.; Rivard, B.; Segl, K.; Grant, B.; Feng, J. Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. *Remote Sens. Environ.* **2014**, *152*, 302–317. [CrossRef]

39. Khamehchiyan, M.; Charkhabi, A.H.; Tajik, M. Effects of crude oil contamination on geotechnical properties of clayey and sandy soils. *Eng. Geol.* **2007**, *89*, 220–229. [CrossRef]

40. *Soil Survey Division Staff. Soil Survey Manual*; Number 18; Government Printing Office: Washington, DC, USA, 1993.

41. Mathieu, M.; Roy, R.; Launeau, P.; Cathelineau, M.; Quirt, D. Alteration mapping on drill cores using a HySpex SWIR-320m hyperspectral camera: Application to the exploration of an unconformity-related uranium deposit (Saskatchewan, Canada). *J. Geochem. Explor.* **2017**, *172*, 71–88. [CrossRef]

42. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]

43. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [CrossRef]

44. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

45. Ayrey, E.; Hayes, D.J. The Use of Three-Dimensional Convolutional Neural Networks to Interpret LiDAR for Forest Inventory. *Remote Sens.* **2018**, *10*, 649. [CrossRef]

46. Kim, P. Deep Learning. In *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*; Springer: New York, NY, USA, 2017; pp. 103–120.

47. Gallego, A.J.; Pertusa, A.; Gil, P. Automatic Ship Classification from Optical Aerial Images with Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 511. [CrossRef]

48. Jin, X.; Jie, L.; Wang, S.; Qi, H.J.; Li, S.W. Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field. *Remote Sens.* **2018**, *10*, 395. [CrossRef]

49. Li, J.; Xi, B.; Li, Y.; Du, Q.; Wang, K. Hyperspectral Classification Based on Texture Feature Enhancement and Deep Belief Networks. *Remote Sens.* **2018**, *10*, 396. [CrossRef]

50. Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sens.* **2018**, *10*, 144. [CrossRef]

51. Mei, S.; Yuan, X.; Ji, J.; Zhang, Y.; Wan, S.; Du, Q. Hyperspectral image spatial super-resolution via 3D full convolutional neural network. *Remote Sens.* **2017**, *9*, 1139. [CrossRef]

52. Huang, Z.; Pan, Z.; Lei, B. Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. *Remote Sens.* **2017**, *9*, 907. [CrossRef]

53. Wu, H.; Prasad, S. Convolutional recurrent neural networks for hyperspectral data classification. *Remote. Sens.* **2017**, *9*, 298. [CrossRef]

54. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* **2017**, *9*, 498. [CrossRef]

55. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

56. Brownlee, J. *Deep Learning with Python: Develop Deep Learning Models on Theano and TensorFlow Using Keras*; Machine Learning Mastery: Vermont Victoria, Australia, 2016.

57. Zweiri, Y.H.; Whidborne, J.F.; Althoefer, K.; Seneviratne, L.D. A new three-term backpropagation algorithm with convergence analysis. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02), Washington, DC, USA, 11–15 May 2002; Volume 4, pp. 3882–3887.

58. Zweiri, Y.H.; Seneviratne, L.D.; Althoefer, K. Stability analysis of a three-term backpropagation algorithm. *Neural Netw.* **2005**, *18*, 1341–1347. [CrossRef]

59. Zweiri, Y.H.; Whidborne, J.F.; Seneviratne, L.D. Optimization and stability of a three-term backpropagation algorithm. In Proceedings of the International Conference of Neural Networks (CI'2000), Como, Italy, 23–27 July 2000.

60. Dahl, G.E.; Sainath, T.N.; Hinton, G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 8609–8613.

61. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference on Machine Learning (ICML 2013), Atlanta, GA, USA, 16–21 June 2013; Volume 30, p. 3.

62. Riedmiller, M.; Braun, H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, USA, 28 March–1 April 1993; pp. 586–591.

63. Abdulkadir, S.J.; Shamsuddin, S.M.; Sallehuddin, R. Three term back propagation network for moisture prediction. In Proceedings of the International Conference on Clean and Green Energy, Hong Kong, China, 5 January 2012; pp. 103–107.

64. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [CrossRef]

65. Nascimento, J.M.; Dias, J.M.B. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]

66. Heinz, D.C. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [CrossRef]

67. Altmann, Y. Nonlinear spectral unmixing of hyperspectral images. Ph.D. Thesis, l'Institut National Polytechnique de Toulouse (INP Toulouse), Toulouse, France, 2013.

68. Altmann, Y.; Halimi, A.; Dobigeon, N.; Tourneret, J.Y. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Trans. Image Process.* **2012**, *21*, 3017–3025. [CrossRef]

69. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, CA, USA, 3–6 December 2012; pp. 1097–1105.

70. Crucil, G.; Castaldi, F.; Aldana-Jague, E.; van Wesemael, B.; Macdonald, A.; Van Oost, K. Assessing the Performance of UAS-Compatible Multispectral and Hyperspectral Sensors for Soil Organic Carbon Prediction. *Sustainability* **2019**, *11*, 1889. [CrossRef]

Hybrid Spectral Unmixing: Using Artificial Neural Networks for Linear/ Non-Linear Switching

Asmau M Ahmed [1], Olga Duran [1], Yahya Zweiri [1,2,]* 🆔 **and Mike Smith [3]**

[1] Faculty of Science, Engineering and Computing, Kingston University London, London SW15 3DW, UK; k1270385@kingston.ac.uk (A.M.A.); o.duran@kingston.ac.uk (O.D.)
[2] Robotics Institute, Khalifa University of Science and Technology, P.O. Box 127788, Abu Dhabi 127788, UAE
[3] Royal Geographical Society (with IBG), 1 Kensington Gore, London SW7 2AR, UK; mike@hsm.org.uk
[*] Correspondence: y.zweiri@kingston.ac.uk; Tel.: +44-020-8417-4846

Abstract: Spectral unmixing is a key process in identifying spectral signature of materials and quantifying their spatial distribution over an image. The linear model is expected to provide acceptable results when two assumptions are satisfied: (1) The mixing process should occur at macroscopic level and (2) Photons must interact with single material before reaching the sensor. However, these assumptions do not always hold and more complex nonlinear models are required. This study proposes a new hybrid method for switching between linear and nonlinear spectral unmixing of hyperspectral data based on artificial neural networks. The neural networks was trained with parameters within a window of the pixel under consideration. These parameters are computed to represent the diversity of the neighboring pixels and are based on the Spectral Angular Distance, Covariance and a non linearity parameter. The endmembers were extracted using Vertex Component Analysis while the abundances were estimated using the method identified by the neural networks (Vertex Component Analysis, Fully Constraint Least Square Method, Polynomial Post Nonlinear Mixing Model or Generalized Bilinear Model). Results show that the hybrid method performs better than each of the individual techniques with high overall accuracy, while the abundance estimation error is significantly lower than that obtained using the individual methods. Experiments on both synthetic dataset and real hyperspectral images demonstrated that the proposed hybrid switch method is efficient for solving spectral unmixing of hyperspectral images as compared to individual algorithms.

Keywords: hyperspectral image; spectral unmixing; endmembers; artificial neural networks; hybrid switch method

1. Introduction

Spectral Unmixing (SU) is the process of identifying spectral signatures of materials often referred to as endmembers and also estimates their relative abundance to the measured spectra. Spectral unmixing is used in a wide range of applications including crop/vegetation classification, disaster monitoring, surveillance, planetary exploration, food industry, fire and chemical spread detection and wild animal tracking [1]. Endmembers play an important role in exploring spectral information of a hyperspectral image [2,3] the extraction of endmembers is the first and most crucial step in any image analysis which is the process of obtaining pure signatures of different features present in an image [1,4,5]. SU often requires the definition of the mixing model underlying the observations as presented on the data. A mixing model describes how the endmembers are combined to form the mixed spectrum as measured by the sensor [6]. Given the mixing model, SU then estimates the inverse of the formation process to infer the quantity of interest, specifically the endmembers, and abundance

from the collected spectra [7,8]. This could be achieved through a radiative transfer model which accurately describes the light scattering by the materials in the observed scene by a sensor [6].

The most common approach to spectral unmixing is the linear spectral unmixing [6,7], which assumes that each photon reaching the sensor interacts with only one material as measured by the spectrum [7]. Promising and excellent results have been recorded with linear spectral unmixing methods as proposed by Keshava and Mustard [1], with some of the commonly used linear mixture models being; Adaptive Spectral Mixture Analysis (ALSMA) [9], Subspace Matching Pursuit (SMP) [10], Orthogonal Matching Pursuit (OMP) [11]. Li et al. [12] proposed a robust collaborative sparse regression method to spectrally unmix hyperspectral data based on a robust linear mixture model. Thouvenin et al. [13] proposed a linear mixing model which explicitly accounts for spatial and spectral endmembers variability. Foody and Cox [14] used a linear mixture model and regression based fuzzy membership function to estimate land cover composition while in [15] the use of the VCA algorithm is demonstrated to unmix hyperspectral data with relatively lower computational complexity compared to other conventional methods. Non linear mixing models cope with nonlinear interactions capturing effects that are mostly present in an image [7]. Li et al. [12] proposed a robust collaborative sparse regression method using a robust linear mixture model which takes into account nonlinearity in the image and treat them as mere outliers. The linear spectral unmixing method generally provides poor accuracy when the light suffers multiple interactions between distinct endmembers or intimate interaction before reaching the sensor [16,17]. In this case, the linear mixture model can be advantageously replaced with nonlinear methods [18,19] which provides an alternative approach to SU. When interactions occur at a microscopic level, it is said that the materials are intimately mixed. A model proposed by Hapke [6] describes the interactions suffered by light when it comes into contact with a surface composed of particles; they involve meaningful and interpretable quantities that have physical significance, however, these models require a nonlinear formulation which is complex and complicates the derivation of the unmixing strategies [7]. These methods account for the intimate mixture of materials, as covered by a scene, in a dataset [1,8]. Different nonlinear mixing models exist, some motivated by physical arguments such as bilinear models, while others exploit a more flexible nonlinear mathematical model to improve the performance of the unmixing method [7]. Nonlinear models can be grouped into several classes such as: intimate mixture models [1], bilinear models [20], physics based nonlinear mixing models [20], polynomial post nonlinear mixing models [21]. Nascimento and Dias [22] solve the nonlinear unmixing problem with an intimate mixture model. This method first converts the observed reflectance into albedo using a look-up table, then a linear algorithm estimates the end members albedo and mass fraction for each sample. Chen et al. [18] formulated a new kernel-based paradigm that relies on the assumption that the mixing mechanism can be described by a linear mixture of end member spectra, with additive nonlinear fluctuations defined in a reproducing Kernel Hilbert Space. Hapke [6] derive an analytical model used to express the measured reflectance as a function of parameters intrinsic to the mixtures, these include mass fraction, density size and single scattering albedo. The main limitation is that these models depend solely on parameters inherent to the experiment because they require the full information of the geometric position of the sensor with respect to the observed samples therefore making the inversion process more challenging to implement especially when the spectral signatures of the endmembers are unknown [1].

Another effect that has been considered to great extend is the endmember variability during spectral unmixing due to atmospheric and temporal conditions. Machine learning methods have worked well to account for spectral variability. The combination of spectral information and spatial context may improve the accuracy of the results for hyperspectral unmixing and classification [23]. Techniques such as morphological filters [24], Markov Random Fields (MRF) [23,25,26] Zhang [27] , Support Vector machines (SVM) [28] and Self Organizing Maps (SOM) [29] among others have been proposed to impose spatial information. MRF, in particular, is a very powerful tool used to describe neighborhood dependence between image pixels and have proven to provide accurate results for

hyperspectral image classification. MRF are effective under the Bayesian inferring framework to incorporate spatial information which proves to provide accurate results in classification and unmixing of hyperspectral data [23]. Markov Random Fields is a method that integrates spatial correlation information into the posterior probability distribution of the spectral features [25]. SVM have shown excellent performance with high classification accuracies when applied to datasets with limited number of training samples [30]. Artificial Neural Networks (ANN) are mathematical models that were initially developed to mimic the complex pattern of neuron interconnections in the human brain [31,32]. Presently, a lot of feed-forward neural networks models have been extensively studied in fault detection and diagnosis of mechanical systems. Moreover, ANN have been successfully applied for many years with excellent performance in pattern recognition [33], and in particular for spectral data [34,35]. SOM is one of the most widely used unsupervised neural network algorithms successfully applied for hyperspectral image classification [29,36,37] and data visualization [38]. Alternative approaches include rule base fuzzy logic [39–41] and Markovian jump systems [42,43] which could be combined with ANN for switching decision making.

Deep learning involves modeling, which hierarchically learn features of input data using Artificial Neural Networks (ANN) and typically have more than three layers [44]. Deep learning has been extensively used in the literature for a range of different applications such as vehicle detection [45,46], investigated avalanche search and rescue operations with Unmanned Areal Vehicles (UAV), change detection [47,48]. In this scheme, high level features are learned from low level ones where the features derived can be formulated for pattern recognition classification [49]. Neural network pattern recognition is often used to classify input data into a set of target categories by training a network to evaluate its performance using a confusion matrix. The application of neural networks has been demonstrated in the field of remote sensing and hyperspectral unmixing due to their ability to recognize complex patterns in high dimensional images [50]. Neural network based unmixing of hyperspectral imagery has produced excellent results [51]. Lyu et al. [48] have demonstrated neural networks to be a good tool for unmixing using both linear and nonlinear methods simultaneously [52]. In [46], the use of artificial neural networks was reported to detect and count cars in Unmanned Areal Vehicle (UAV) images. Wu and Prasad [53] used neural networks for hyperspectral data classification, where a recurrent neural network was used to model the dependencies between different spectral bands and learn more discriminative features for hyperspectral data classification. Li et al. [35] reported the use of a 3D convolution neural network to extract spectral - spatial combined features from a hyperspectral image. Kumar et al. [51] used a linear mixture model to unmix hyperspectral data and then neural networks to predict a fraction of the data that accounts for the nonlinear mixture; they used ground truth data and the abundance estimated by the linear method to train the network for effective validation. Giorgio and Frate [50] used neural networks to unmix hyperspectral data to estimate endmembers and their abundance. Atkinson and Lewis [54] applied neural networks to decompose hyperspectral data and compared their results with a linear unmixing model and a fuzzy c-mean classifier; results showed that the neural networks outperformed the conventional linear unmixing method.

Little work in combining the linear and nonlinear approaches has been presented in the literature, and in particular the selection of the most appropriate technique in using the two methods. In this paper, we note that some nonlinear methods are a better method in scenes with multiple interactions and a complex mixture of features commonly composed of multi-layered materials. The linear model is appropriate for images that have a single cover type of material in a pixel. The objective of this paper is to propose a new hybrid methodology for switching between linear and nonlinear spectral unmixing methods using artificial neural networks based on deep learning strategies. The paper is organized as follows. Section 2 describes our methodology. Experimental results are presented in Section 3, results were discussed in section 4 and Conclusions are drawn in Section 5.

2. Methodology

2.1. Research Design

In this study, two linear and two nonlinear spectral unmixing methods were adopted to unmix hyperspectral data. The question as to whether a mixed pixel is better explained with a linear or nonlinear process is still an unresolved problem in spectral analysis. Researchers have identified temporal, spectral and spatial variability that maybe due for instance to variable illumination, environmental, atmospheric, and temporal conditions in the scenes as the main error in spectral unmixing [55]. Thus endmember variability problem has been deeply studied, neglecting the effects of multiple scattering and the resulting nonlinear mixing [56]. Non linearities may occur when the photons interact with different material before reaching the sensor. In that sense, studies suggest that linear mixing is associated to mixtures for which the pixel components appear in spatially segregated patterns, (checker board scene) [55]. The structure of the canopy and the spatial distribution of the plants area are also known to play an important role in nonlinearity [56]. This paper proposes a novel approach to decide whether a mixed pixel is better explained with a linear or non linear model. Here we use a ANN to learn and to decide the non-linearity of a pixel based on some simple spatial and spectral features. The methods chosen were the state of the art methods that have been used extensively as reference in literature. They are: the Vertex Component Analysis (VCA) [15] and Fully Constrained Least Square Method (FCLS) [57] for the linear models, and the Polynomial Post Nonlinear Mixing Model (PPNMM) [21] and Generalized Bilinear Model (GBM) [58] for the nonlinear models. Hybridization between the methods was experimented with Artificial Neural Networks (ANN) to conduct a switch between the linear and nonlinear models.

2.1.1. Vertex Component Analysis (VCA)

This algorithm is based on the geometry of convex sets and exploits the fact that endmembers occupy the vertices of a simplex [59]. The VCA algorithm assumes the presence of spectrally pure pixels in a dataset and iteratively projects the data onto the direction orthogonal to the subspace spanned by the end members that are already determined Weeks [4]. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are exhausted Bioucas et al. [60].

2.1.2. Fully Constrained Least Square Method (FCLS)

The FCLS algorithm is derived from an unconstrained least square based orthogonal subspace projection Heinz [57]; in this method, negative values are considered 0 and the abundance fractions of the remaining material signatures are normalized to 1. FCLS utilizes a simplex method to produce a set of feasible solutions for spectral unmixing of material signatures while discarding the negative abundance values of the remaining material signatures to unity [57].

2.1.3. Polynomial Post Nonlinear Mixture (PPNM)

This model assumes that the reflectance of an image are nonlinear functions of pure spectral components contaminated by additive noise; the nonlinear functions are often approximated using polynomial function leading to a polynomial post nonlinear mixing model Altmann et al. [21].

The model involves linear and quadratic functions of the abundances. In this case, the R-spectrum $Y = [y_1, ..., y_R]^T$ of a mixed pixel is defined as a nonlinear transformation g of a linear mixture of L spectra $\mathbf{m}_l$ contaminated by additive noise $\mathbf{n}$.

$$Y = g_p \left(\sum_{l=1}^{L} a_l \mathbf{m}_l \right) + \mathbf{n} \tag{1}$$

where $\mathbf{m}_l$ is the spectrum of the lth material in the scene, a_l its corresponding proportion, L is the number of endmembers contained in the image and g is an appropriate nonlinear function. Another motivation for the PPNMM is the Weierstrass approximation theorem which states that every continuous function defined on an interval can be uniformly approximated by a polynomial with any desired precision [21].

2.1.4. Generalized Bilinear Mixing Model

The GBM model introduces a second term that accounts for multiple photon interactions [20]. This model proposes that the spectrum of a mixed pixel, Y can be derived as follows:

$$Y = \sum_{l=1}^{L} a_l \mathbf{m}_l + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} \gamma_{i,j} a_i a_j \ \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n} \qquad (2)$$

where $\mathbf{m}_i \odot \mathbf{m}_j$ is the Hadamard (term by term) product of the i and j spectra, $\mathbf{m}_i$ is the spectrum of the endmember i, a_i is the corresponding abundance and $\mathbf{n}$ is an additive noise. The first model term describes the linear mixture model and the double sum models the nonlinear effect. $\gamma = [\gamma_{1,2}..., \gamma_{L-1, \ L}]$ is a real parameter vector, $\gamma_{i,j} \in (0,1)$, that quantifies the interaction between different spectral components. The parameter introduced in this model is used to obtain a more flexible model Halimi et al. [61]. This model also adopts the positivity and sum to one constraints.

2.2. Vicinity Parameters

The objective of this study is to switch between linear and nonlinear methods depending on the mixture type of the neighboring pixels. The linear model is expected to provide acceptable results when two assumptions are satisfied [1] i.e., the mixing process should occur at macroscopic level and the photons must interact with single material before reaching the sensor (checker board scene). Generally this can not be known a priori and might change in different parts on a given scene. Most profound sources of error in spectral mixture analysis, however, lies in the lack of ability to account for sufficient temporal and spatial spectral variability [55]. Endmember variability problem is often caused by spatial and temporal changes thereby neglecting the effects of multiple scattering and the resulting nonlinear mixing [62]. In fact, it is more likely that the position, extent and number of stable spectral zones depends on the spatial, spectral and temporal complexity and composition of the endmembers present in the scene [55], therefore, it would be very interesting to design new models and nonlinear unmixing procedures that are capable of simultaneously exploiting the spatial correlation between abundances and nonlinearities to produce best results. Here, we propose a methodology to automatically switch between linear and nonlinear spectral unmixing to provide more accurate results based on deep learning neural network strategies. A number of parameters that are related to the pixels' neighboring characteristics are used. We assume that neighboring pixels in a checkerboard type of scene have more spectral spatial coherent spectrum than those in a nonlinear scene. The following values represent the diversity of the neighborhood for the pixel under consideration to the ones in its vicinity. These values are the minimum and maximum Spectral Angular Distance (SAD), covariance and a nonlinearity parameter. In order to compute these parameters, we defined a window W around the examined pixels of size $n \times n$.

2.2.1. Spectral Angular Distance (SAD)

Spectral Angular Distance (SAD) describes the angular distance between two vectors, this is estimated by computing the cosine of the angles between the actual and the estimated endmembers [63].

The SAD between two spectra: $\mathbf{U} = (U_i, ..., U_R)^T$ and $\mathbf{V} = (V_i, ..., V_R)^T$ is defined as

$$SAD(\mathbf{U}, \mathbf{V}) = cos^{-1}\left(\frac{\sum_{i=1}^{R} U_i V_i}{\|\mathbf{U}\| \quad \|\mathbf{V}\|} \right) \tag{3}$$

where R is the number of bands and $\|\mathbf{U}\| \quad \|\mathbf{V}\|$ are the modules of the vectors. Here, we compute the SAD between all pixels within the window W and use the minimum and maximum values respectively.

2.2.2. Covariance Matrix

The Covariance matrix proposes a way of fusing multiple spectra that are correlated. The variance of each spectra are represented by the diagonal values of the covariance matrix while the non-diagonal values represent the correlation [64]. The covariance matrix is defined by the following equation:

$$C = \frac{1}{R-1} \sum_{i=1}^{R} (\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_i - \boldsymbol{\mu}_i)^T \tag{4}$$

where $\boldsymbol{\mu}_i$ is the mean vector of all pixels in band i and $\mathbf{X}_i$ is the vector containing all pixel values in band i within window W.

2.2.3. Nonlinearity Parameter

The non-linearity parameter, b as computed to a window, is a parameter which quantifies the levels of nonlinearity in a pixel, given as:

$$b = \sum_{i=1}^{L} \sum_{j=i+1}^{L} \mathbf{a}_i \mathbf{a}_j \ \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n} \tag{5}$$

where $\odot$ is the Hadamard (term by term) product operation, $\mathbf{a}_i$ and $\mathbf{a}_j$ are the abundance reflectance spectra of endmembers i and j and L is the number of endmembers.

2.3. Learning

An Artificial Neural Network was used to predict the best method when switching between linear and nonlinear spectral unmixing.

The data for the Artificial Neural Networks were divided into 3 categories, namely: training, validation and testing sets.

1. The training set is used to fit the parameters of the classifier.
2. Validation set is used to minimize over-fitting (i.e., verifying the accuracy of the training data) over some untrained data by the networks, while
3. testing sets are used to test the final solution in order to confirm the actual predictive power of the network [65].

The networks were trained with scale conjugate gradient back propagation because it has proven to be efficient and produce accurate results [66–68] . The back propagation procedure simply adopts the chain rule derivative [69], this is achieved where the gradient of the objective with respect to the input module, is computed backwards from the output module [69,70]. This was considered due to its performance in updating the weight and bias values according to the scaled conjugate gradient; the training stops when certain conditions are met such as the maximum number of epochs is reached, maximum amount of time is exceeded, performance is minimized to the goal and the validation performance has increased more than the maximum it recorded [65]. We expect that the linear model

will perform better if neighborhood pixels are similar, on the other hand, when the pixel have multiple interactions, we expect higher diversity in the pixels.

The neural networks have 3 layers namely: input, hidden and output layers. The input layer has 12 nodes when using a 3×3 window corresponding to the vicinity parameters as described in Section 2.2 (min. SAD, max. SAD, c_1, c_2, c_3.....c_9, b); the hidden layer has 10 nodes while the output layer has 1 node. The output layer provides the decision between linear and non-linear unmixing models.

3. Experimental Setup and Results

3.1. Data Description

3.1.1. Simulated Data

A simulated dataset of images of size 36×36 pixels and 224 channels was generated with abundances computed according to a Dirichlet distribution with 21 endmembers. The spectral signatures of the endmembers are mineral reflectance with 224 bands from the ENVI spectral library [15]. Additionally, a nonlinearity co-efficient was added ranging between $[0, 1]$ these parameters were tuned accordingly with different numbers of endmembers ranging from 3 to 9. The images were corrupted with Random Gaussian noise with Signal to Noise Ratio (SNR) 10 dB, 30 dB and 50 dB respectively. Figure 1 show the spectral reflectance of endmembers of the simulated data.
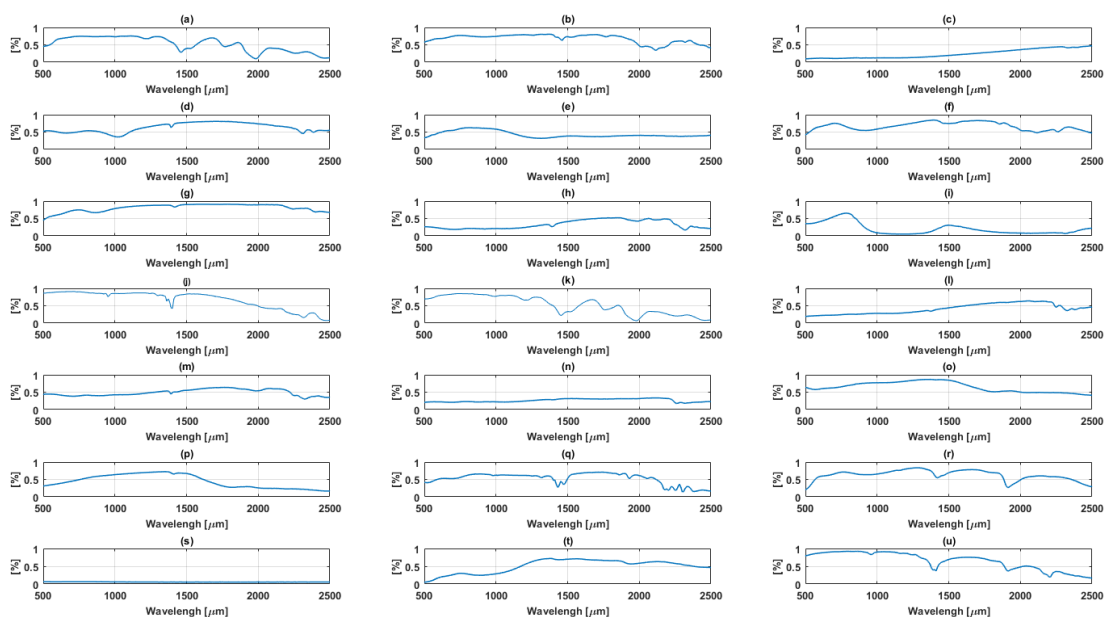
Figure 1. Spectral reflectance of endmembers of the simulated data plotted, reflectance [%] against wavelength (**a**) Brucite; (**b**) Clinochlore; (**c**) Axinite; (**d**) Erionite; (**e**) Ammonioalunite; (**f**) Clintonite; (**g**) Almandine; (**h**) Carnallite; (**i**) Actinolite; (**j**) Andradite; (**k**) Antigorite; (**l**) Elbaite; (**m**) Ammonio-jarosite; (**n**) Diaspore; (**o**) Halloysite; (**p**) Biotite; (**q**) Galena; (**r**) Carnallite; (**s**) Chlorite; (**t**) Goethite; (**u**) Corundum.

3.1.2. Real data

Samson Data

Samson data is a hyperspectral data owned by Oregon State University provided by WeoGeo [71], which is a push broom visible to near infrared sensor. The pixel responses are captured by 156 bands in the spectral range of 401 nm–889 nm with resolution up to 3.13 nm. The data has 952 scan lines with 952 pixels in each line. For this experiment a subset of the image covering 95×95 pixels was used, which is

comprised of three endmembers i.e., soil, tree and water. Figure 2 shows the spectral reflectance of endmembers of the Samson data.
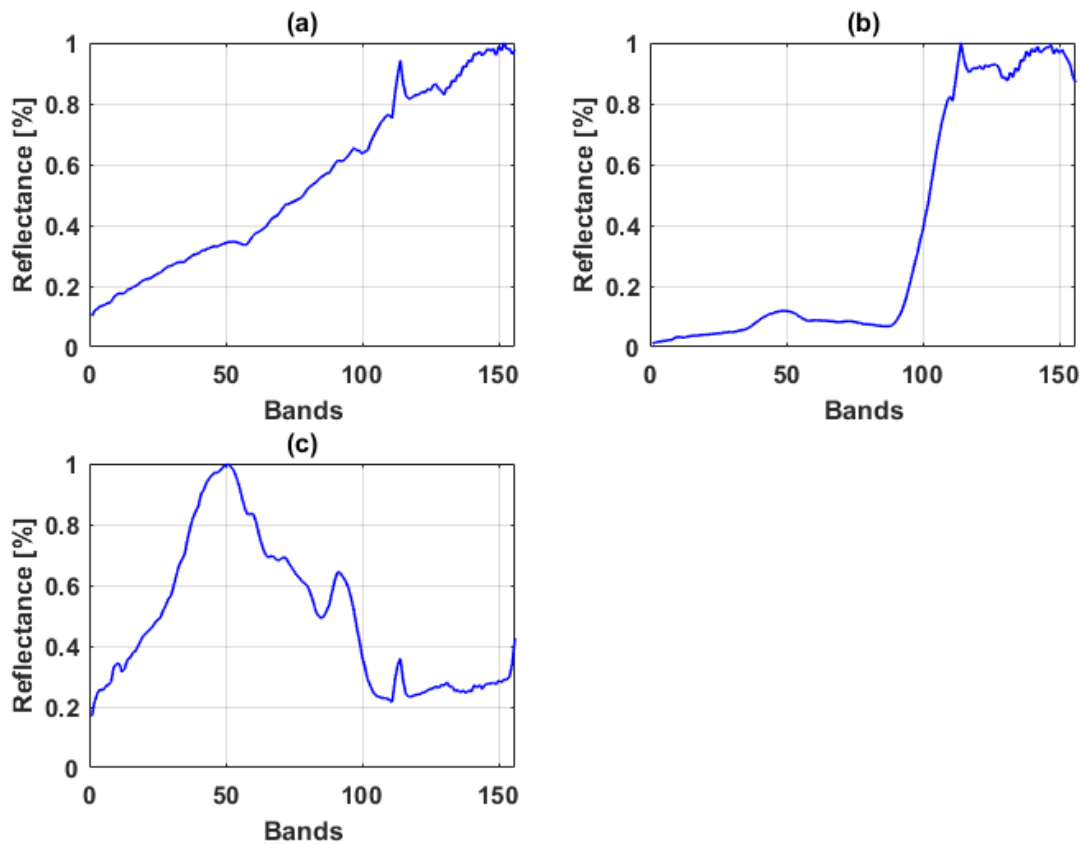
Figure 2. Spectral reflectance of endmembers of the Samson data (**a**) rock; (**b**) tree; (**c**) water.

Jasper Ridge

Jasper Ridge is a hyperspectral data cube recorded by AVIRIS over the standard scene of the Jasper Ridge, a biological reserve in California. The dataset consist of 512×614 pixels recorded in 224 channels ranging from 380 nm to 2500 nm. The data has a spectral resolution of 9.46 nm. In this experiment, a subset of 100×100 pixels was used from the original image and 198 bands were selected after removing those bands with atmospheric effects and dense water vapor. There are four main endmembers in this image: road, soil, water and tree [71]. Both datasets and corresponding abundance ground truth are available at [71], and are used as a benchmark to test classification and unmixing algorithms. Figure 3 show the spectral reflectance of endmembers of the Jasper Ridge data.

Figure 3. Spectral reflectance of endmembers of the Jasper Ridge data (**a**) tree; (**b**) water; (**c**) soil; (**d**) road.

3.2. Experiments with Synthetic Data

This experiment was carried out using the synthetic dataset described in Section 3.1.1, which allows a priori control of the data. Here, VCA, FCLS, PPNMM and GBM methods were used to unmix spectra of mineral mixtures. We compared the accuracy of the individual methods to the proposed hybrid methods for switching between linear and nonlinear spectral 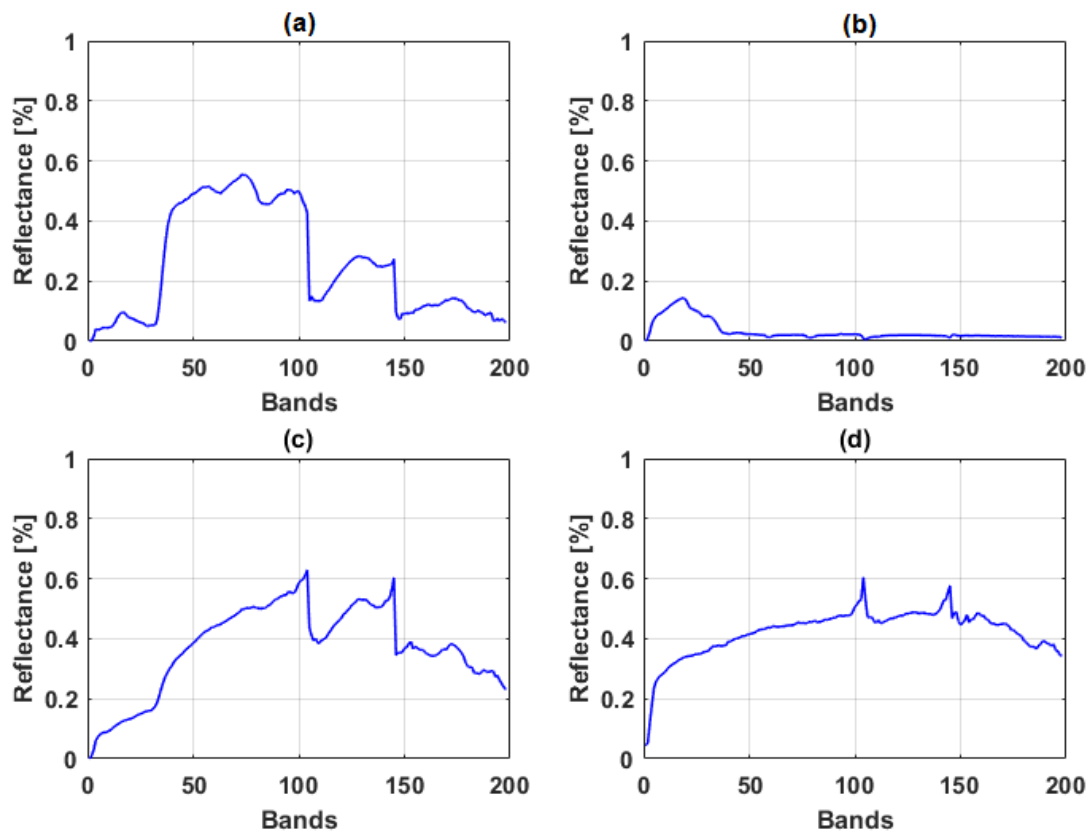unmixing based on the diversity of the neighboring pixels. The algorithms were coded according to [15,21,57,58]. The hybrid methods for switching were between VCA–PPNMM, VCA–GBM, FCLS–PPNMM, and FCLS–GBM, respectively. VCA was used to estimate the endmembers as contained in the dataset, while the four methods as well as the hybrid methods were used to estimate the fractional abundances. The experiment was conducted with different numbers of endmembers ranging 3, 5, 7 and 9 and different Signal to Noise Ratios of 10 dB, 30 dB and 50 dB respectively on the simulated dataset. We ran Monte Carlo simulations based on 100 generated images for each experiment.

The switching was predicted using Artificial Neural Networks (ANN). Here, we randomly split the samples into training, validation and test sets. During training, 70% of the datasets were selected to train the network, 15% were used as validation set to learn the hyperparameters of the neural networks and 15% of the remaining samples were used to test the accuracy of the networks.

In the first experiment, a 3×3 window was used around the pixel of interest. A vector containing 12 values i.e., SAD min, SAD max, covariance matrix (9 values) and nonlinearity was computed for each pixel as input to train the ANN. Each input data consisted of 12 nodes with the number of hidden nodes set to 10 and the output layer having 1 node which output (0 or 1) corresponding to either a linear or nonlinear approach where a threshold was set at 0.5 for the switching. The Artificial Neural Network was used to choose between a linear or nonlinear approach for each pixel. The overall accuracy and the abundance estimation error of the methods were computed and summarized in Table 1. Results

shows that the VCA – PPNMM hybrid method predicted better overall accuracy of 98.8% as estimated by the confusion matrix with neural networks in switching between linear and nonlinear spectral unmixing, followed by FCLS–PPNMM with an overall accuracy of 95.6%, VCA–GBM and FCLS–GBM both have an overall accuracy of 92% and 92.4% respectively. Examples showing the generated data and the error in abudance estimation are shown in Figures 4 and 5. Here have chosen to display a linear method (VCA) and a non linear (GBM) for comaprison purposes and two different signal to noise ratios (SNR=10 and 50, respectively).

Table 1. Abundance estimation error (3×3 window)of the individual and hybrid methods between linear and nonlinear spectral unmixing with different signal to noise ratios and endmembers. The best results are shown in bold.

SNR (dB) $= 50$	$P = 3$	$P = 5$	$P = 7$	$P = 9$
INDIVIDUAL METHODS				
PPNMM	0.0206	0.0307	0.0371	0.0486
GBM	0.0207	0.0303	0.0346	0.0449
VCA	0.0521	0.0696	0.0777	0.0778
FCLS	0.0714	0.0916	0.0922	0.0924
HYBRID METHODS				
VCA – PPNMM	**0.0117**	0.0201	**0.0143**	0.0373
VCA – GBM	0.0189	0.0201	0.0158	0.0353
FCLS – PPNMM	0.0177	**0.0179**	0.0177	**0.0340**
FCLS – GBM	0.0193	0.0196	0.0199	0.0174
SNR (dB) $= 30$	$P = 3$	$P = 5$	$P = 7$	$P = 9$
INDIVIDUAL METHODS				
PPNMM	0.0696	0.0951	0.0914	0.0886
GBM	0.0965	0.1193	0.1405	0.1285
VCA	0.0597	0.0662	0.0886	0.0945
FCLS	0.0684	0.0747	0.0894	0.0911
HYBRID METHODS				
VCA – PPNMM	**0.0390**	**0.0317**	**0.0421**	**0.0556**
VCA – GBM	0.0591	0.0412	0.0579	0.0662
FCLS – PPNMM	0.0396	0.0320	0.0539	0.0645
FCLS – GBM	0.0866	0.0926	0.0990	0.1081
SNR (dB) $= 10$	$P = 3$	$P = 5$	$P = 7$	$P = 9$
INDIVIDUAL METHODS				
PPNMM	0.0907	0.1510	0.1640	0.1733
GBM	0.1106	0.1222	0.1334	0.1740
VCA	0.1289	0.1514	0.1257	0.1988
FCLS	0.1169	0.1702	0.1791	0.1763
HYBRID METHODS				
VCA – PPNMM	0.0401	**0.0421**	**0.0736**	**0.0775**
VCA – GBM	0.0704	0.0911	0.0813	0.0915
FCLS – PPNMM	**0.0440**	0.0508	0.0813	0.0814
FCLS – GBM	0.0917	0.0959	0.1099	0.1112

Figure 4. Abundance estimation errors with simulated data with 5 endmembers (SNR = 10 dB). The first row shows the ground truth abundances for the 5 classes. From the top, then the error in abundances as estimated by the hybrid, VCA and GBM methods, respectively.
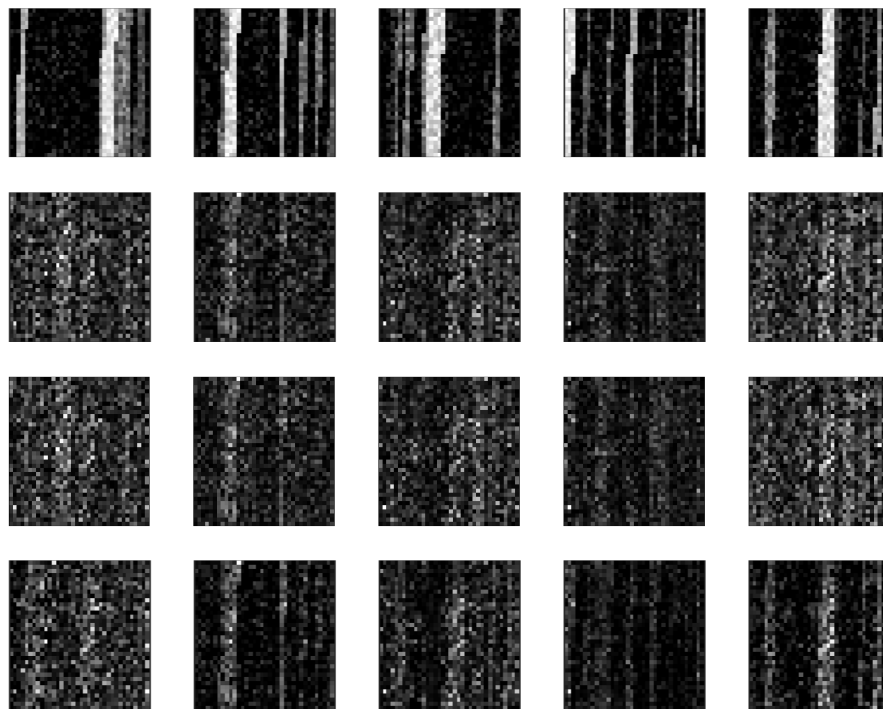

Figure 5. Abundance estimation errors with simulated data with 5 endmembers (SNR = 50 dB). The first row shows the ground truth abundances for the 5 classes. From the top, then the error in abundances as estimated by the hybrid, VCA and GBM methods, respectively.
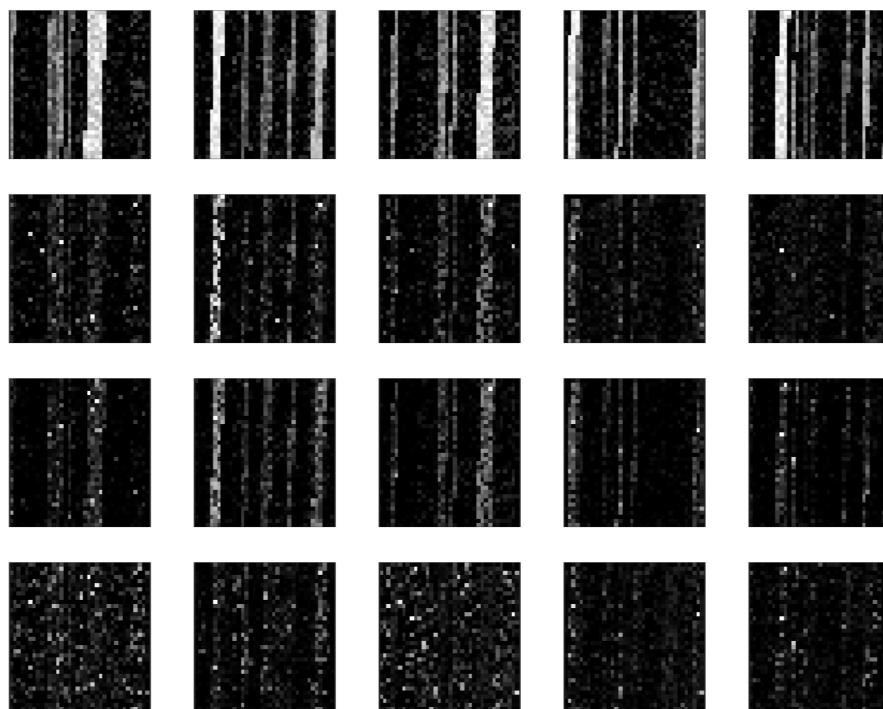
A second experiment was conducted with the 3×3 window where each of the parameters used in creating the input training data was excluded one at a time. This was performed in order to assess the importance of the parameters in the vector created. The experiment was also repeated for SNR values 10 dB, 30 dB, and 50 dB with different endmembers of 3, 5, 7 and 9 respectively. Here we expect to have higher error values when each of the parameters are removed from the vector in comparison with the results in Table 1 where all the parameters are involved in the experiment. Results (Table 2) show that all parameters play an important role in the vector and hybrid switch methods. SAD max proves to be the most important parameter with the highest error value in the experiment where it was excluded for all SNR values as compared to the other parameters. For comparison purpose, the size of the window was increased to 4×4 and the experiment repeated to evaluate the accuracy of the parameters. Results show an increase in the error value when each of the parameters is excluded from the ANN input data.

Table 2. Abundance estimation error (3×3 window) of the individual and hybrid methods between linear and nonlinear spectral unmixing with different signal to noise ratios and 3 endmembers where each of the parameters is removed one at a time. The best results are shown in bold.

WITHOUT SAD MIN.	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.1503	0.0537	0.0179
GBM	0.1220	0.1274	0.0168
VCA	0.1090	0.1000	0.0952
FCLS	0.1670	0.1370	0.0997
HYBRID METHODS			
VCA – PPNMM	**0.0433**	**0.0392**	0.0150
VCA – GBM	0.0854	0.0784	0.0163
FCLS – PPNMM	0.0434	0.0402	**0.0143**
FCLS – GBM	0.1180	0.1080	0.0161
WITHOUT SAD MAX.	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.0969	0.0876	0.0878
GBM	0.1002	0.0920	0.0741
VCA	0.1216	0.0791	0.0451
FCLS	0.2726	0.1073	0.0560
HYBRID METHODS			
VCA – PPNMM	0.0584	**0.0467**	**0.0251**
VCA – GBM	0.0885	0.0731	0.0525
FCLS – PPNMM	**0.0521**	**0.0467**	**0.0251**
FCLS – GBM	0.1689	0.1000	0.0772
WITHOUT COVARIANCE DISTANCE	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.0940	0.0518	0.0173
GBM	0.1243	0.0921	0.0166
VCA	1.0488	0.0824	0.0590
FCLS	1.1673	0.0966	0.0680
HYBRID METHODS			
VCA – PPNMM	**0.0506**	0.0340	**0.0145**
VCA – GBM	0.0902	0.0588	0.0163
FCLS – PPNMM	**0.0506**	**0.0336**	**0.0145**
FCLS – GBM	0.1231	0.0916	0.0161

Table 2. *Cont.*

WITHOUT NONLINEARITY PARAMETER	SNR (dB) = 10	SNR (dB) = 30	SNR (dB) = 50
INDIVIDUAL METHODS			
PPNMM	0.1447	0.0532	0.0183
GBM	0.1100	0.1039	0.0185
VCA	0.1251	0.0982	0.0865
FCLS	0.1852	0.1167	0.0927
HYBRID METHODS			
VCA – PPNMM	**0.0448**	0.0432	0.0173
VCA – GBM	0.0856	0.0789	0.0178
FCLS – PPNMM	**0.0448**	**0.0431**	**0.0171**
FCLS – GBM	0.1236	0.1096	0.0137

In order to assess the accuracy of the methods, we also trained a network with the raw data (i.e., 224 bands) as input instead of the vicinity parameters. Here we have 224 while the rest of the parameters remain the same. Table 3 summarizes the results of 100 Monte Carlo simulations. It is noted that the results are of the same order of magnitude as obtained in Table 1. Figure 3 displays the abundance estimated error by the 4 methods with SNR = 10 dB.

Table 3. Abundance estimation error with the individual and hybrid methods of the raw hyperspectral data between linear and nonlinear spectral unmixing with different signal to noise ratios and different endmembers. The best results are shown in bold.

SNR (dB) = 50	$P = 3$	$P = 5$	$P = 7$	$P = 9$
INDIVIDUAL METHODS				
PPNMM	0.0253	0.0276	0.0378	0.0418
GBM	0.0253	0.0276	0.0347	0.0383
VCA	0.0775	0.0612	0.0717	0.0719
FCLS	0.0891	0.0663	0.0877	0.0612
HYBRID METHODS				
VCA – PPNMM	**0.0125**	**0.0127**	**0.0230**	**0.0285**
VCA – GBM	0.0457	0.0164	0.0269	0.0317
FCLS – PPNMM	0.0217	0.0214	0.0236	0.0316
FCLS – GBM	0.0513	0.0627	0.0850	0.0981
SNR (dB) = 30	$P = 3$	$P = 5$	$P = 7$	$P = 9$
INDIVIDUAL METHODS				
PPNMM	0.1520	0.1759	0.1464	0.1353
GBM	0.1568	0.1442	0.1473	0.1337
VCA	0.1007	0.1195	0.0313	0.2767
FCLS	0.1072	0.1713	0.1344	0.1819
HYBRID METHODS				
VCA – PPNMM	**0.0231**	**0.0223**	**0.0219**	**0.0268**
VCA – GBM	0.0317	0.0360	0.0364	0.0370
FCLS – PPNMM	0.0308	0.0358	0.0458	0.0654
FCLS – GBM	0.0437	0.0787	0.0901	0.0956

Table 3. *Cont.*

SNR (dB) = 10	$P = 3$	$P = 5$	$P = 7$	$P = 9$
INDIVIDUAL METHODS				
PPNMM	0.1809	0.1816	0.1856	0.1883
GBM	0.1517	0.1506	0.1440	0.1481
VCA	0.1196	0.0612	0.0717	0.0717
FCLS	0.1072	0.0663	0.0877	0.0612
HYBRID METHODS				
VCA – PPNMM	**0.0548**	**0.0564**	**0.0570**	**0.0584**
VCA – GBM	0.0751	0.0940	0.0962	0.0961
FCLS – PPNMM	0.0714	0.0739	0.0740	0.0763
FCLS – GBM	0.0974	0.0981	0.0990	0.1170

From the experiments conducted between the 3 × 3 window and the raw data, it can be seen that the results are similar between the Signal to Noise Ratios 10 dB and 50 dB. However, the results were better with the 3 × 3 window with a Signal to Noise Ratio of 30 dB. Therefore, it can be concluded that the ANN does not require the whole raw data and the reduced choosen parameters provide good results. Figure 5 shows the abundance results with simulated data (SNR = 50 dB). The first row shows the ground truth abundances in grayscale where a white pixel means abundance equal to one for that class and a black pixel means no abundance for that class. The other rows display the error in abundance estimation for each class and each method also coded in grayscale where the brighter the pixel, the higher the error is.

3.3. Experiment with Real Data

To evaluate the accuracy of the methods involved, the raw data, as well as the vicinity parameters computed in a 3 × 3 window, and 4 × 4 window respectively, were used to train the neural network. In the first experiment, the Jasper Ridge data was used. The training samples for each experiment were selected randomly, 70% of the samples were used for training (7000 samples), 15% each were considered for validation and testing (1500 samples for validation and 1500 samples for testing) of the neural networks. In a second experiment, the number of training samples were reduced, with 30% used for training (3000 samples), 35% each used for validation and testing (3500 samples for validation and testing) of the neural networks respectively. Finally, the experiment was repeated with 1000 and 300 training samples, respectively. Figure 6 shows the groundtruth abundances and the abundances as estimated by a linear (VCA), nonlinear (PPMM) and the corresponding hybrid methods on the Jasper Ridge data.

The next experiment was with Samson data, where the training, validation and testing samples were randomly selected at 70%, 15% and 15% respectively resulting in 6317 samples for training, 1353 samples each for testing and validation, then number of training samples were reduced to 30%; 35% for validation and 35% for testing which is equivalent to 2707 samples for training, 3158 samples each for validation and testing the neural networks respectively. Finally, the experiment was repeated with 1000 and 300 training samples, respectively. Figure 7 shows the groundtruth abundances, and the abundances as estimated by the a linear (VCA), nonlinear (PPMM) and hybrid methods on the Samson data.

The experiment was repeated on a 3 × 3 window, and 4 × 4 window. This was to compare and evaluate the accuracy of the hybrid methods with regards to the size of the data used to train the networks. The results on both datasets show that our proposed methods achieved the best results in all scenarios. Results of the experiments based on the abundance estimation error, are summarized in Tables 4 and 5.
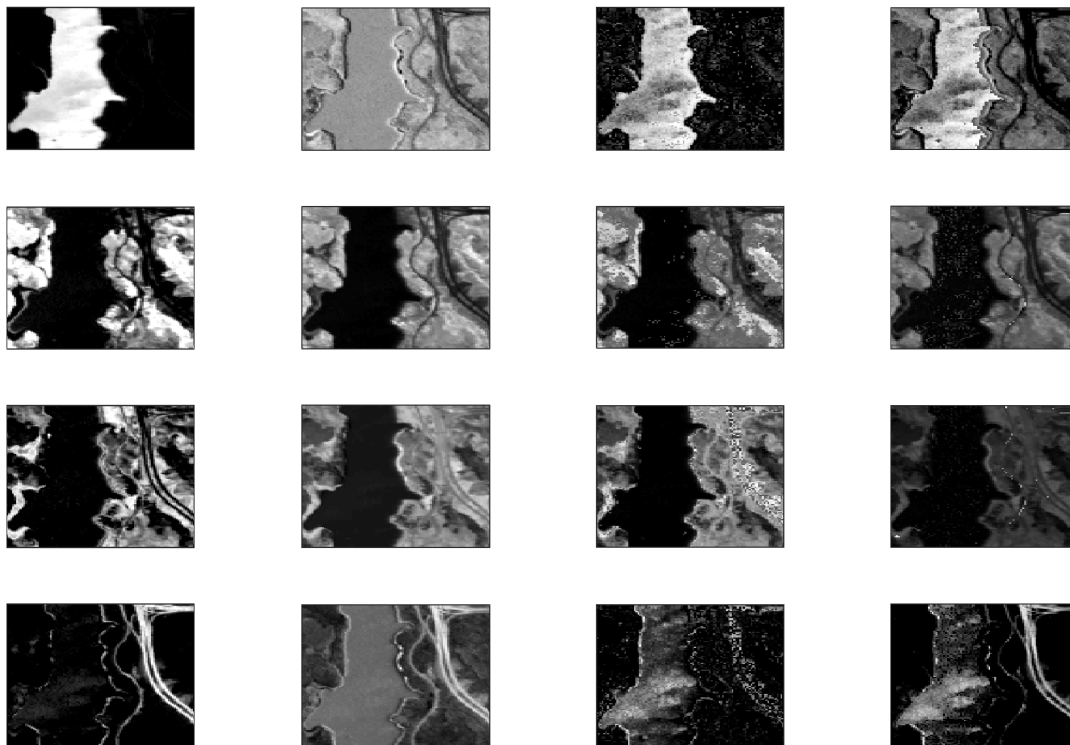
Figure 6. Abundance estimate of endmembers of the Jasper Ridge data showing from left; the groundtruth, linear (VCA), nonlinear (PPMM) and the hybrid methods. From top **water**; **tree**; **soil**; and **road**.

The overall accuracy of the network, the abundance estimation error, the training, validation and testing abundance error of the networks were used to evaluate the performance of the methods investigated in this paper. From the results obtained, experiments with the raw dataset, 3×3 and 4×4 windows produce similar overall accuracy in all the experiments. It indicates that the hybrid methods for switching between linear and nonlinear spectral unmixing are more effective than the individual methods, meanwhile, it can also be said that ANN pattern recognition has good capability in recognizing patterns which is very effective even with fewer samples used to train the network. From the four hybrid switch methods, the VCA – PPNMM method outperforms the other methods with a higher overall accuracy of 96% as compared to the other methods, FCLS – PPNMM has an overall accuracy of 94.5% while VCA – GBM and FCLS – GBM both have overall accuracies of 92.8%. VCA – PPNMM also has the lowest abundance estimation error and produced the lowest abundance error in terms of training, validation and testing of the neural networks. However, it can be observed that the proposed hybrid switch methods obtained similar results when using the 3×3 and 4×4 window to conduct the experiment when fewer samples were used to train the networks. Therefore, it shows that the proposed hybrid method does not requires all the raw data for training the networks and can be used effectively to switch between linear and nonlinear spectral unmixing of hyperspectral data. In terms of computational time, the individual methods are 40% more time consuming compared to the hybrid method thereby making them computationally expensive in terms of simulation. Tables 6 and 7 summarizes the result of the experiments showing the accuracy of the neural network based on training, testing and validation of the networks.
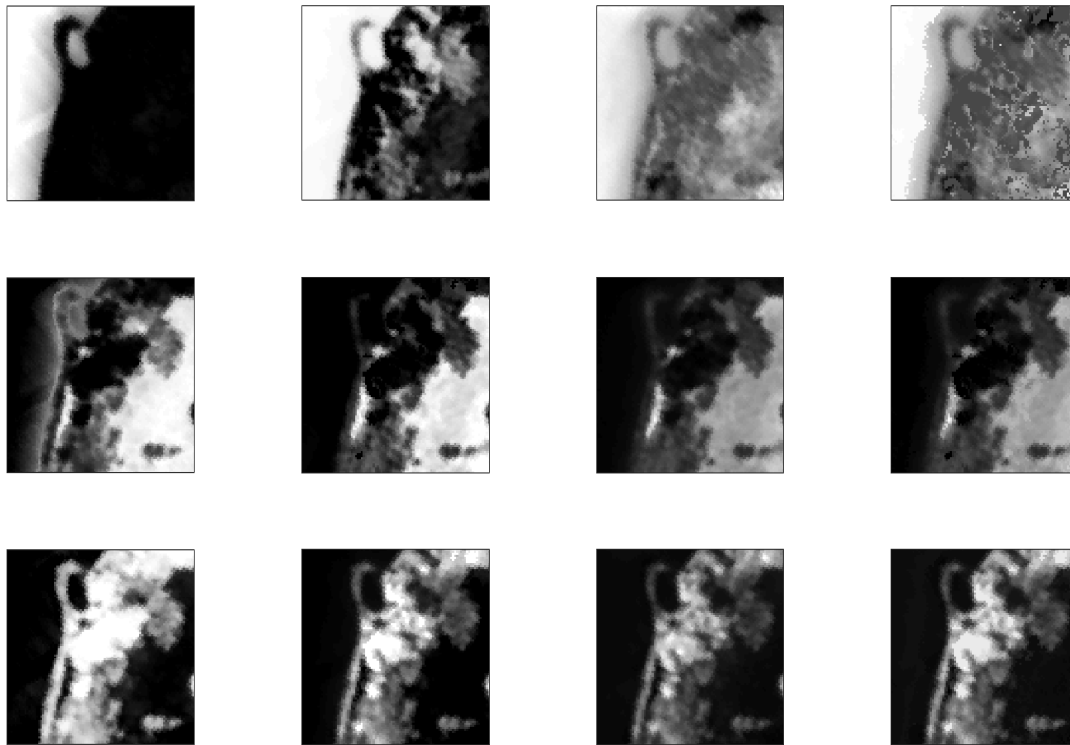
Figure 7. Abundance estimate of endmembers of the Samson data showing the groundtruth, linear (VCA), nonlinear (PPMM) and the hybrid methods. From top **water**; **rock**; **tree**.

Table 4. Average abundance estimation error of the hybrid methods with different numbers of training samples (7000 to 300) and different window size vectors on the Jasper Ridge data as compared with the abundance estimation error of the individual methods which are: PPNMM = 0.2115, GBM = 0.2441, VCA = 0.6513, and FCLS = 0.1832. The best results are shown in bold.

Raw Data	7000	3000	1000	300
VCA – PPNMM	0.1417	0.1478	**0.1405**	0.1994
VCA – GBM	0.2079	0.2049	0.3087	0.3897
FCLS – PPNMM	0.1402	0.1402	0.1590	0.1663
FCLS – GBM	**0.1399**	**0.1397**	0.1483	**0.1495**
3×3 **WINDOW**				
VCA – PPNMM	**0.1697**	**0.1607**	**0.1781**	**0.1763**
VCA – GBM	0.2595	0.2454	0.2932	0.3350
FCLS – PPNMM	**0.1765**	0.1624	0.1783	0.1790
FCLS – GBM	0.2448	0.2448	0.3442	0.3642
4×4 **WINDOW**				
VCA – PPNMM	0.1712	**0.1640**	0.1736	**0.1704**
VCA – GBM	0.2488	0.2250	0.3117	0.3460
FCLS – PPNMM	**0.1632**	0.1659	**0.1705**	0.1722
FCLS – GBM	0.2647	0.2459	0.2488	0.2732

Table 5. Average abundance estimation error of the hybrid methods with different numbers of training samples (6317 to 300) and different window size vectors on the Samson data as compared with the abundance estimation error of the individual methods which are: PPNMM = 0.1455, GBM = 0.1588, VCA = 0.1254, and FCLS = 0.1577. The best results are shown in bold.

Raw Data	6317	3158	1000	300
VCA – PPNMM	**0.0839**	**0.0841**	**0.0871**	0.0979
VCA – GBM	0.0841	0.0846	0.0879	**0.0939**
FCLS – PPNMM	0.1229	0.1230	0.1258	0.1308
FCLS – GBM	0.1614	0.1615	0.1674	0.1696
3 × 3 WINDOW				
VCA – PPNMM	**0.0888**	**0.0885**	**0.0902**	**0.0973**
VCA – GBM	0.0975	0.1040	0.1079	0.1112
FCLS – PPNMM	0.1148	0.1151	0.1197	0.1292
FCLS – GBM	0.1615	0.1617	0.1657	0.1710
4 × 4 WINDOW				
VCA – PPNMM	**0.0904**	**0.0905**	**0.0949**	**0.0994**
VCA – GBM	0.0945	0.0945	0.1061	0.1106
FCLS – PPNMM	0.1154	0.1197	0.1216	0.1245
FCLS – GBM	0.1616	0.1616	0.1636	0.1658

Table 6. Abundance estimation error on Jasper Ridge data showing training, validation and testing accuracy on the individual and hybrid methods with different training samples and different window size vectors. The best results are shown in bold.

Raw Data	7000 Samples				3000 Samples			
	VCA–PPNMM	VCA–GBM	FCLS–PPNMM	FCLS—GBM	VCA–PPNMM	VCA–GBM	FCLS–PPNMM	FCLS—GBM
TRAIN	**0.0905**	0.1025	0.1184	0.1084	0.0953	**0.0859**	0.1085	0.1200
VALIDATION	**0.0777**	0.0780	0.1008	0.0980	**0.0809**	0.0866	0.1006	0.0995
TEST	**0.0751**	0.0797	0.1012	0.1000	**0.0811**	0.0832	0.1013	0.0906
3 × 3 Window								
TRAIN	**0.0967**	0.1054	0.0981	0.1268	**0.0473**	0.0533	0.0991	0.1229
VALIDATION	0.0524	**0.0505**	0.1274	0.1138	**0.0465**	0.0549	0.0923	0.1125
TEST	**0.0486**	0.0614	0.1276	0.1147	**0.0454**	0.0506	0.0914	0.1135
4 × 4 Window								
TRAIN	**0.0906**	0.1523	0.0941	0.1171	**0.0393**	0.1531	0.0997	0.1146
VALIDATION	0.1696	**0.0704**	0.0911	0.0954	**0.0351**	0.0530	0.0918	0.1117
TEST	0.1608	**0.0382**	0.0938	0.0944	**0.0354**	0.0445	0.0920	0.1121

Table 7. Abundance estimation error on Samson data showing training, validation and testing accuracy on the individual and hybrid methods with different training samples and different window size vectors. The best results are shown in bold.

Raw Data	6317 Samples				3158 Samples			
	VCA–PPNMM	VCA–GBM	FCLS–PPNMM	FCLS—GBM	VCA–PPNMM	VCA–GBM	FCLS–PPNMM	FCLS—GBM
TRAIN	**0.0255**	0.0741	0.1058	0.1585	**0.0280**	0.0732	0.1182	0.1167
VALIDATION	0.0466	0.0101	0.0792	**0.0098**	0.0553	0.1026	0.0733	0.1311
TEST	0.0494	0.0105	0.0762	**0.0098**	0.0569	0.1053	0.0719	0.1311
3 × 3 Window								
TRAIN	**0.0726**	0.0842	0.1046	0.1581	**0.0722**	0.0897	0.1021	0.1161
VALIDATION	0.0530	0.0100	0.0748	**0.0098**	0.0588	0.0692	0.0703	0.1309
TEST	0.0533	0.0100	0.0740	**0.0098**	0.0594	0.0696	0.0706	0.1309
4 × 4 Window								
TRAIN	**0.0836**	0.0842	0.1046	0.1581	**0.0822**	0.0843	0.1007	0.1160
VALIDATION	0.0536	0.0100	0.0748	**0.0098**	0.0569	0.0654	0.0710	0.1308
TEST	0.0545	0.0100	0.0740	**0.0098**	0.0569	0.0640	0.0712	0.1308

4. Discussion

4.1. Results

Nonlinearity occurs when the photons interact with different materials before reaching the sensor. We assumed here that the linear mixing could be associated to mixtures for which the pixel components appear in spatially segregated patterns. More specifically areas that are spatially correlated are more likely to be explained with the linear model. In this paper, we first used controlled simulated data. Each image consisted of a series of regions. Each region had the same type of ground cover with added noise. Figure 5 shows the results for the simulated data with 5 classes and SNR = 50 dB. Although the average error is in the same order of magnitude for both linear and nonlinear approaches, the distribution of the error differs. It is noted that the linear model (VCA) detects well the low abundances of classes and pixels that do not contain a particular class (shown in black in the ground truth figures). The errors are related to quantification rather than to detecting the wrong class. This might be due to the algorithm performing worse with high spectral variability within the classes. The non-linear method (GBM), on the other hand, returns an error which is more uniform and not so related to the spatial pattern of the data or spectral variability as displayed in Figures 4 and 5. The proposed approach assumptions are further proven with the real data sets. In particular, the Jasper Ridge data set includes different classes; water, soil and road. Figure 6 shows the abundance estimation for the different methods. It is noted that VCA has been reported to underperform in this data set [27]. However, the road class is very well identified against the non-linear methods that failed to detect this class. On the other hand, the linear methods failed to classify correctly the water class which is more spectrally variable medium. Thus, it seems that noise and endmember spectral variability makes the non-linear models outperforming the linear ones while the spatially structured areas are well defined with the linear model. The vicinity parameters used in this paper address both the spatial and spectral diversity of the data. The test shown in Table 2 showed that all parameters played an important role in the decision making process. Moreover, Figures 6 and 7 also support that the chosen features are suitable and that the switching is appropriate achieving improved results.

4.2. Advantages and Limitations

The proposed method provides a switch between unmixing methods for given spectral images. It can not only provide more accurate results, as shown in the experimental section but also reduce computational costs by selecting the most appropriate approach. This research study has proven the capabilities of the proposed methodology based on certain parameters. However, the supervised ANN relies on having ground truth data for training which is not always available. Future work will expand to unsupervised approaches such as self-organizing maps which have been successfully used in spectral data for classification and anomaly detection tasks [33,72]. Although we used spatial and spectral features within windows for learning and thus to make the decision, the switching was made at individual pixel level. Thus future work will base the decision on group of pixels or areas using for instance Markov random fields.

5. Conclusions

In this paper, a new hybrid switch method for switching between linear and nonlinear spectral unmixing of hyperspectral data based on deep learning neural networks is proposed. The endmembers were extracted using VCA while the abundances were estimated using individual and hybrid methods. The ANN was trained with a set of parameters extracted from the diversity of the neighboring pixels of the images computed within a 3×3 and 4×4 window. These parameters are spectral angular distance, covariance and nonlinearity parameters. Experiments were conducted with different Signal to Noise Ratio (SNR) ranging between 10 dB, 30 dB, 50 dB and different numbers of endmembers: 3, 5, 7 and 9. We have noted that the hybrid methods are more suitable than the individual technique with high overall accuracy and the abundance estimation error is significantly lower than that obtained with the

individual methods in particular, VCA – PPNMM proved to be the best with about 98% accuracy in all the experiments conducted. The experiment with the Jasper Ridge and Samson data confirmed the effectiveness of the approach. The method was applied to two real datasets with ANN trained using 70%, 30%, 10% and 0.3% samples. Experimentation with the real data, 3×3 window and 4×4 window vectors, proved the effectiveness of the hybrid switch methods, the results show that the size of datasets used for training the network and the vector size does not affect the accuracy of the hybrid methods in switching between linear and nonlinear spectral unmixing, which means that the network can be trained with less sample data without the loss of prediction accuracy. An area to consider for future research is the application of Markovian Jump method for switching between linear/nonlinear spectral unmixing.

Author Contributions: All authors have made great contributions to the work. Asmau Ahmed and Olga Duran conceived and designed the experiments, Asmau Ahmed, Olga Duran, Yahya Zweiri and Mike Smith analyzed the data and revised the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57.
2. Xu, M.; Zhang, L.; Du, B.; Zhang, L.; Fan, Y.; Song, D. A Mutation Operator Accelerated Quantum-Behaved Particle Swarm Optimization Algorithm for Hyperspectral Endmember Extraction. *Remote Sens.* **2017**, *9*, 197.
3. Uezato, T.; Murphy, R.J.; Melkumyan, A.; Chlingaryan, A. A novel spectral unmixing method incorporating spectral variability within endmember classes. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2812–2831.
4. Weeks, A.R. *Fundamentals of Electronic Image Processing*; SPIE Optical Engineering Press: Bellingham, WA, USA, 1996.
5. Drumetz, L.; Tochon, G.; Chanussot, J.; Jutten, C. Estimating the number of endmembers to use in spectral unmixing of hyperspectral data with collaborative sparsity. In Proceedings of the the 13th International Conference on Latent Variable Analysis and Signal Separation (LVA-ICA), Grenoble, France, 21–23 February 2017; pp. 1–10.
6. Hapke, B. Bidirectional reflectance spectroscopy: 1. Theory. *J. Geophys. Res.* **1981**, *86*, 3039–3054.
7. Dobigeon, N.; Tourneret, J.Y.; Richard, C.; Bermudez, J.; McLaughlin, S.; Hero, A.O. Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal Process. Mag.* **2014**, *31*, 82–94.
8. Halimi, A.; Altmann, Y.; Buller, G.S.; McLaughlin, S.; Oxford, W.; Clarke, D.; Piper, J. Robust unmixing algorithms for hyperspectral imagery. In Proceedings of the Sensor Signal Processing for Defence (SSPD), Edinburgh, UK, 22–23 September 2016; pp. 1–5.
9. Chang, C.I. Adaptive Linear Spectral Mixture Analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1240–1253.
10. Shi, Z.; Tang, W.; Duren, Z.; Jiang, Z. Subspace matching pursuit for sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3256–3274.
11. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2014–2039.
12. Li, C.; Ma, Y.; Mei, X.; Liu, C.; Ma, J. Hyperspectral unmixing with robust collaborative sparse regression. *Remote Sens.* **2016**, *8*, 588.
13. Thouvenin, P.A.; Dobigeon, N.; Tourneret, J.Y. Hyperspectral unmixing with spectral variability using a perturbed linear mixing model. *IEEE Trans. Signal Process.* **2016**, *64*, 525–538.
14. Foody, G.M.; Cox, D. Subpixel Land Cover Composition Estimation Using Linear Mixture Model and Fuzzy Membership Functions. *Int. J. Remote Sens.* **1994**, *15*, 619–631.
15. Nascimento, J.M.; Dias, J.M.B. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910.

16. Nascimento, J.M.; Bioucas-Dias, J.M. Nonlinear mixture model for hyperspectral unmixing. In Proceedings of the SPIE Europe Remote Sensing. International Society for Optics and Photonics, Berlin, Germany, 31 August 2009; p. 74770I.

17. Li, C.; Ma, Y.; Huang, J.; Mei, X.; Liu, C.; Ma, J. GBM-based unmixing of hyperspectral data using bound projected optimal gradient method. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 952–956.

18. Chen, J.; Richard, C.; Honeine, P. Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *IEEE Trans. Signal Process.* **2013**, *61*, 480–492.

19. Marinoni, A.; Gamba, P. Accurate detection of anthropogenic settlements in hyperspectral images by higher order nonlinear unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1792–1801.

20. Altmann, Y. Nonlinear Spectral Unmixing of Hyperspectral Images. Ph.D. Thesis, l'Institut National Polytechnique de Toulouse (INP Toulouse), Toulouse, France, 2013.

21. Altmann, Y.; Halimi, A.; Dobigeon, N.; Tourneret, J.Y. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Trans. Image Process.* **2012**, *21*, 3017–3025.

22. Nascimento, J.M.; Bioucas-Dias, J.M. Unmixing hyperspectral intimate mixtures. In Proceedings of the Remote Sensing. International Society for Optics and Photonics, Toulouse, France, 20 September 2010; p. 78300C.

23. Sun, L.; Wu, Z.; Liu, J.; Xiao, L.; Wei, Z. Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1490–1503.

24. Fauvel, M.; Benediktsson, J.A.; Chanussot, J.; Sveinsson, J.R. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3804–3814.

25. Yu, H.; Gao, L.; Li, J.; Li, S.S.; Zhang, B.; Benediktsson, J.A. Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive markov random fields. *Remote Sens.* **2016**, *8*, 355.

26. Ni, L.; Gao, L.; Li, S.; Li, J.; Zhang, B. Edge-constrained Markov random field classification by integrating hyperspectral image with LiDAR data over urban areas. *J. Appl. Remote Sens.* **2014**, *8*, 085089.

27. Zhang, B.; Li, S.; Jia, X.; Gao, L.; Peng, M. Adaptive Markov random field approach for classification of hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 973–977.

28. Jimenez, L.O.; Landgrebe, D.A. Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Trans. Syst. Man Cybern. Part C* **1998**, *28*, 39–54.

29. Martínez, P.; Gualtieri, J.; Aguilar, P.; Pérez, R.; Linaje, M.; Preciado, J.; Plaza, A. Hyperspectral Image Classification Using a Self-organizing Map. Available online: http://www.umbc.edu/rssipl/people/aplaza/Papers/Conferences/2001.AVIRIS.SOM.pdf (accessed on 1 June 2017).

30. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675.

31. Raith, S.; Vogel, E.P.; Anees, N.; Keul, C.; Güth, J.F.; Edelhoff, D.; Fischer, H. Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data. *Comput. Biol. Med.* **2017**, *80*, 65–76.

32. Han, T.; Jiang, D.; Zhao, Q.; Wang, L.; Yin, K. Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Trans. Inst. Meas. Control* **2017**, doi:10.1177/0142331217708242.

33. Pal, S.K.; Mitra, S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Trans. Neural Netw.* **1992**, *3*, 683–697.

34. Duran, O.; Althoefer, K.; Seneviratne, L.D. Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network. *IEEE Trans. Autom. Sci. Eng.* **2007**, *4*, 118–126.

35. Li, Y.; Zhang, H.; Shen, Q. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 67.

36. Duran, O.; Petrou, M. A time-efficient clustering method for pure class selection. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'05), Seoul, Korea, 29 July 2005; Volume 1, doi:10.1109/IGARSS.2005.1526223.

37. Pérez-Hoyos, A.; Martínez, B.; García-Haro, F.J.; Moreno, Á.; Gilabert, M.A. Identification of ecosystem functional types from coarse resolution imagery using a self-organizing map approach: A case study for Spain. *Remote Sens.* **2014**, *6*, 11391–11419.

38. Penn, B.S. Using self-organizing maps to visualize high-dimensional data. *Comput. Geosci.* **2005**, *31*, 531–544.

39. Wei, Y.; Qiu, J.; Shi, P.; Lam, H.K. A new design of H-infinity piecewise filtering for discrete-time nonlinear time-varying delay systems via TS fuzzy affine models. *IEEE Trans. Syst. Man Cybern.* **2017**, doi:10.1109/TSMC.2016.2598785.

40. Wei, Y.; Qiu, J.; Lam, H.K.; Wu, L. Approaches to TS fuzzy-affine-model-based reliable output feedback control for nonlinear Itô stochastic systems. *IEEE Trans. Fuzzy Syst.* **2016**, doi:10.1109/tfuzz.2016.2566810.

41. Lam, H.K.; Li, H.; Liu, H. Stability analysis and control synthesis for fuzzy-observer-based controller of nonlinear systems: A fuzzy-model-based control approach. *IET Control Theory Appl.* **2013**, *7*, 663–672.

42. Wei, Y.; Qiu, J.; Karimi, H.R.; Wang, M. Model reduction for continuous-time Markovian jump systems with incomplete statistics of mode information. *Int. J. Syst. Sci.* **2014**, *45*, 1496–1507.

43. Wang, H.; Shi, P.; Agarwal, R.K. Network-based event-triggered filtering for Markovian jump systems. *Int. J. Control* **2016**, *89*, 1096–1110.

44. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107.

45. Audebert, N.; Saux, B.L.; Lefèvre, S. Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. *Remote Sens.* **2017**, *9*, 368.

46. Ammour, N.; Alhichri, H.; Bazi, Y.; Benjdira, B.; Alajlan, N.; Zuair, M. Deep Learning Approach for Car Detection in UAV Imagery. *Remote Sens.* **2017**, *9*, 312.

47. Bejiga, M.B.; Zeggada, A.; Nouffidj, A.; Melgani, F. A convolutional neural network approach for assisting avalanche search and rescue operations with uav imagery. *Remote Sens.* **2017**, *9*, 100.

48. Lyu, H.; Lu, H.; Mou, L. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection. *Remote Sens.* **2016**, *8*, 506.

49. Makantasis, K.; Karantzalos, K.; Doulamis, A.; Doulamis, N. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2015), Milan, Italy, 26–31 July 2015; pp. 4959–4962.

50. Giorgio, L.; Frate, F.D. *A Neural Network Approach for Pixel Unmixing in Hyperspectral Data;* Earth Observation Laboratory- Tor Vergata University Via del Politecnico: Rome, Italy.

51. Kumar, U.; Raja, K.S.; Mukhopadhyay, C.; Ramachandra, T.V. A Neural Network Based Hybrid Mixture Model to Extract Information from Nonlinear Mixed Pixels. *Information* **2012**, *3*, 420–441.

52. Giorgio, A.L.; Frate, F.D. Pixel Unmixing in Hyperspectral Data by Means of Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4163–4172.

53. Wu, H.; Prasad, S. Convolutional Recurrent Neural Networks forHyperspectral Data Classification. *Remote Sens.* **2017**, *9*, 298.

54. Atkinson, P.M.; Cutler, M.E.J.; Lewis, H. Mapping Sub-Pixel Proportional Land Cover With AVHRR Imagery. *Int. J. Remote Sens.* **1997**, *18*, 917–935.

55. Somers, B.; Asner, G.P.; Tits, L.; Coppin, P. Endmember variability in spectral mixture analysis: A review. *Remote Sens. Environ.* **2011**, *115*, 1603–1616.

56. Somers, B.; Cools, K.; Delalieux, S.; Stuckens, J.; Van der Zande, D.; Verstraeten, W.W.; Coppin, P. Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards. *Remote Sens. Environ.* **2009**, *113*, 1183–1193.

57. Heinz, D.C.; Chang, C.-I. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545.

58. Altmann, Y.; Dobigeon, N.; Tourneret, J.Y. Bilinear models for nonlinear unmixing of hyperspectral images. In Proceedings of the 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 6–9 June 2011; pp. 1–4.

59. Nascimento, J.M.; Dias, J.M.B. Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 175–187.

60. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379.

61. Halimi, A.; Altmann, Y.; Dobigeon, N.; Tourneret, J.Y. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4153–4162.

62. Altmann, Y.; Pereyra, M.; McLaughlin, S. Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis. *IEEE Trans. Comput. Imaging* **2015**, *1*, 174–185.

63. Sohn, Y.; McCoy, R.M. Mapping desert shrub rangeland using spectral unmixing and modeling spectral mixtures with TM data. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 707–716.

64. Tuzel, O.; Porikli, F.; Meer, P. Region covariance: A fast descriptor for detection and classification. In *Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006;* Springer: Berlin/Heidelberg, Germany, 2006; pp. 589–600.

65. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007.

66. Charalambous, C. Conjugate gradient algorithm for efficient training of artificial neural networks. *IEEE Proc. G* **1992**, *139*, 301–310.

67. Saini, L.M.; Soni, M.K. Artificial neural network-based peak load forecasting using conjugate gradient methods. *IEEE Trans. Power Syst.* **2002**, *17*, 907–912.

68. Nabipour, M.; Keshavarz, P. Modeling surface tension of pure refrigerants using feed-forward back-propagation neural networks. *Int. J. Refrig.* **2017**, *75*, 217–227.

69. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.

70. Hu, X.; Yuan, Y. Deep-Learning-Based Classification for DTM Extraction from ALS Point Cloud. *Remote Sens.* **2016**, *8*, 730.

71. Zhu, F.; Wang, Y.; Fan, B.; Xiang, S.; Meng, G.; Pan, C. Spectral unmixing via data-guided sparsity. *IEEE Trans. Image Process.* **2014**, *23*, 5412–5427.

72. Duran, O.; Petrou, M. A time-efficient method for anomaly detection in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3894–3904.

PROCEEDINGS OF SPIE

Quantitative analysis of petroleum hydrocarbon contaminated soils using spectroscopy, spectral unmixing and deep neural networks

Asmau M. Ahmed, Olga Duran, Yahya Zweiri, Mike Smith

Quantitative Analysis of Petroleum Hydrocarbon Contaminated Soils using Spectroscopy, Spectral Unmixing and Deep Neural Networks

Asmau M. Ahmed[a], Olga Duran[b], Yahya Zweiri[c], and Mike Smith[d]

[b] Faculty of Science, Engineering and Computing, Kingston University London, London SW15 3DW, UK.
[c] Khalifa University of Science and Technology, P.O. Box 127788, Abu Dhabi, United Arab Emirates
[d] School of Geography, Earth and Environmental Sciences. University of Plymouth, Drake Circus Plymouth Devon. PL4 8AA. UK.

ABSTRACT

Oil spill can be described as a global issue, which causes serious concern to human life and the environment, therefore early identification and remedial measures taken at an early stage is very important. Spectral Unmixing is the process of identifying the constituent spectra of a mixed pixel referred to as endmembers and computing the corresponding proportions or abundances within each pixel in a given image. Many spectral unmixing methods have being proposed in the literature based on linear or nonlinear models. Deep neural networks allow computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstractions. Deep neural networks have shown excellent performance on various task on image processing with better accuracy compared to shallow learning networks and are increasingly gaining popularity with Hyperspectral imaging. Here we propose to use deep neural network to quantify different Hydrocarbon (HCs) substances in sandy clay loam soil type. Hyperspectral data sets have been acquired using mixtures of different HCs with the soil type. Vertex Component Analysis (VCA) algorithm was used to identify the endmembers and deep neural network was used to predict the quantity of each endmember. Experimental result shows the effectiveness of the proposed method with high accuracy.

Keywords: Spectroscopy, Hydrocarbons, Spectral unmixing, Hyperspectral imaging, Deep neural networks

1. INTRODUCTION

Intensive Petro-chemical activities around industrial facilities and oil production sites have led to an increase in the contamination of soils with hydrocarbons causing serious concern to the environment[1] . Hydrocarbon (HC) contamination in soils is a worldly concern phenomenon, it posses significant environmental problem, which raises serious concern for the environment and human health, therefore, there is need for rapid identification of the affected areas in order to allow swifter site characterization and prioritizing remediation actions.[2] HCs are pollutants of major concern due to their significant toxicity to humans. Soils contaminated with HCs produce unpredictable effect on the soils as a result compromise the soils health.[3]

Some of the existing methods used to track and detect oil spills in soils are often expensive and time consuming.[4] Some of the methods as described in literature includes; analytical methods such as Gas Chromatography -mass Spectro- Meter (GC-MS) method and infrared spectroscopy.[5] Imaging spectroscopy has been recognized as an alternative method for detecting HCs in soils which has proven to be rapid and cost effective.[6] Spectroscopy measures the diffuse reflected electromagnetic energy from source object to a light source.[7] Spectroscopy analysis

Further author information: (Send correspondence to Yahya Zweiri)
Yahya Zweiri.: E-mail: y.zweiri@kingston.ac.uk, Telephone: +44-020-8417-4846
Olga Duran: E-mail:o.duran@kingston.ac.uk, Telephone: +(0)2084 174 740

includes Visible - Near-Infrared (Vis-NIR) and ShortWave Infrared (SWIR) spectrum, which has been demonstrated to be a powerful tool for the measurement of HC concentration in soils.[4]

Different methods have been described to effectively analyze reflectance spectroscopy data to detect HCs in soils; Okparanma et al[8] used regression analysis and spectral pre- processing to generate statistical models to identify different HC products mixed with mineral substrates. Other methods includes; Partial Least Squares Regression (PLSR),[3] Stepwise Multiple Linear Regression (SMLR),[9] Standard Normal Variate Pre-processing followed by Detrending (SNV-DT)[7] etc.

One important problem in hyperspectral image processing is to decompose the mixed pixels into the materials that contribute to the pixel endmembers and the corresponding fractions of the spectral signatures in the mixed pixel which is often referred to as unmixing problem. Spectral Unmixing (SU) is the process of identifying and decomposing mixed pixel into a collection constituent spectral and estimates their corresponding abundance to the measured spectra.[10] Endmembers play an important role in exploring spectral information of a hyperspectral image which is the first step in hyperspectral unmixing applications.[11] The two main approaches to hyperspectral unmixing are the linear and nonlinear methods to which their application rely on some certain conditions and the type of mixing citedobigeon et al 2016.

Artificial Neural Networks (ANN) has proven to rapidly grow in remote sensing application mainly due to its ability to learn complex patterns and has proven to be an effective approach for unmixing problem and image classification.[12] ANN often extract features in a shallow manner which do not extract deep features, on the other hand, Deep Neural Networks (DNN) hierarchically learn features from the input data, this typically is a multi- layer neural network with two or more hidden layers thereby extracting high level abstract features of different characteristics such as scaling and translation invariance.[13] DNN has demonstrated to be of excellent performance on various image processing tasks with better accuracy compared to shallow neural network. DNN are a class of machine learning that learn features hierarchically using high level features from low level ones.[14] In practice, DNN, Convolutional Neural Networks (CNN) and Restricted Boltzmann Machines (RBM) are more often used to implement deep learning procedures.[15] Deep learning techniques have shown promising results in various application such as image classification,[16] object detection[17] and pixel unmixing.[18]

More specifically, DNN has proven to be of excellent performance on various image processing tasks with better accuracy compared to shallow neural networks. In a similar way, we propose to use spectroscopy analysis and deep neural networks to quantitatively analyse a hyperspectral data to detect different hydrocarbon contents contaminated with sandy clay loam soil type.

2. MATERIALS AND METHODS

2.1 Materials

Different Hydrocarbon (HC) types were used in this experiment, which includes Diesel, Bio - diesel, petrol and Ethanol. Soil type used in this experiment is sandy clay loam. This soil type has high concentration of sand in them which gives them the gritty feel. Sandy clay loam soils are capable of quickly draining excess water but do not hold a significant amount which make them saturates easily.

2.2 Sample Preparation

In the experiment, each hydrocarbon type was mixed with sandy clay loam soil untill saturated. The sample preparation began with the addition of 50 g of the sandy clay loam soil to a petri dish of size 12 cm in diameter, followed by the addition of 5 ml of the HC to the soil using a syringe. The mixture was homogenized with a disposable plastic spoon after which the sample was measured with Hyspex SWIR 320 m camera. In the same sample, a further addition of 5 ml was added to the mixture; again the mixture was homogenized using a disposable plastic spoon and another measurement taken. This procedure was repeated with a step of 5 ml until a total of 25 ml was added or until the soil sample mixture was saturated and forms a shallow local pool as shown in figure 1. For the experiment with petrol, saturation occurs when a total of 35 ml was added to the soil sample.

The procedure was repeated with all four HCs. Experiments using bio-diesel and ethanol, 10 samples were produced with 5 mixtures for each HC and soil combination, experiment with diesel produced 4 samples and

Figure 1. Sandy clay loam soil contaminated with 5 ml and 20 ml diesel.

experiment with petrol produced 7 mixtures as summarized in Table 1. The hyperspectral data was measured and scanned with Hyspex SWIR 320 m camera under constant illumination.

Table 1. Samples combination and their saturation levels.

Soil type	Hydrocarbon type	Number of samples	Saturation levels (ml)
Sand clay loam	Diesel	4	20
Sand clay loam	Bio- diesel	5	25
Sand clay loam	Ethanol	5	25
Sand clay loam	Petrol	7	35

2.3 Hyperspectral Imaging

The hyperspectral image was obtained using Hyspex SWIR 320 m camera as shown in figure 2. The camera is equipped with a Mercury Cadmium- Telluride (MCT or HgCaTe) detector array which simultaneously acquires a full Short Wave Infa-Red (SWIR) spectrum, it has a spectral sample interval of 5 nm between 1300 and 2500 nm, each along a line of 384 pixels.[19] The images acquired are 700×384 pixels. The obtained datasets were then analyzed to extract the endmembers and estimate the percentages of the HCs using deep learning approach.
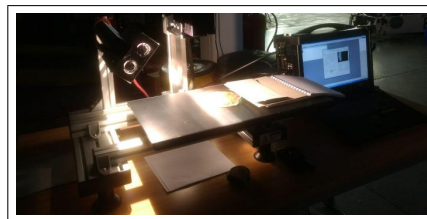

Figure 2. Scanning process of the dataset.

2.4 Endmember Extraction

Vertex Component Analysis (VCA) algorithm was used to identify the spectra of the hyperspectral data and extract the abundance materials. This algorithm is based on the geometry of convex sets and exploits the fact that endmembers occupy the vertices of a simplex.[20] This algorithm assumes the presence of spectrally pure pixels in a dataset. The algorithm iteratively projects the data onto the direction orthogonal to the subspace spanned by the endmembers already determined.[21] The new endmember signature as estimated by the method corresponds to the extreme of the projection. The algorithm iterates until all endmembers are exhausted.[22]

2.5 Deep Neural Networks

A DNN is a network which comprises of multiple nonlinear processing layers, operating using simple elements, it has an input layer, several hidden layers as well as an output layer. The hidden layers are interconnected via nodes or neurons where each hidden layer utilizes the output of the previous layer as input.[23] DNN has proven to be competitive and even outperform shallow neural networks in different applications,[24] likewise, the application of DNN to hyperspectral data has been explored in,[25] where application is mostly based on spectral and spatial information of the hyperspectral data because DNN learns the representation of a spectral signature for each class and uniquely distinguish it from the other classes.
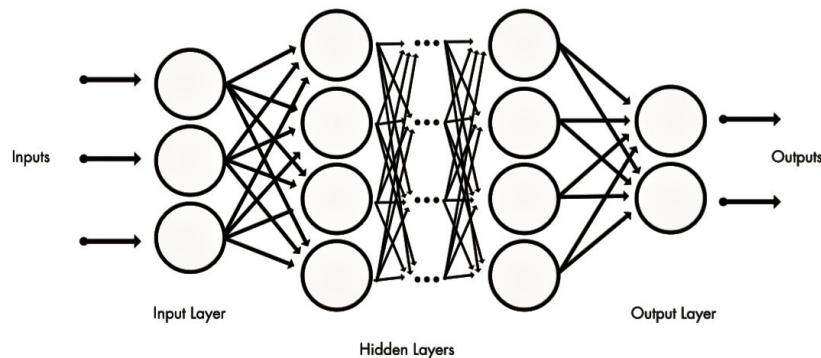

Figure 3. Deep neural network system architecture.

Here, we built a simple DNN architecture to quantify the amount of HCs mixed with a soil type using hyperspectral data. The DNN model comprises of multiple layers with different nodes, followed by one fully connected layer, the neurons in the fully connected layer have full connections to all activation in the previous layer and are responsible for the classification of the data. The output layer has a very similar structure with the fully connected layer, this layer has the same number of neurons as the number of classes we're trying to classify the input data. Each neuron in the outputs layer yields a probability which corresponds to each of the predicted classes which is often presumed to have a maximum value,[14] DNN requires tweaking of various hyperparameters to achieve optimum results, these largely depends on the type of dataset used for the model and other hyperparameters such as the initial learning rate alpha, beta and the algorithm used to train the network.[13]

3. EXPERIMENTAL SETUP AND RESULTS

In this study, imaging spectroscopy and deep neural networks were used to estimate the amount of HC content in sand - clay -loam soil. Imaging spectroscopy is a method which utilizes information obtained from surface reflectance in the electromagnetic spectrum using different sensors and the resultant information is used for spectral and spatial analysis.[8] Spectral analysis is a process that involves the evaluation of energy reflected and absorbed at different wavelengths[26] .The hyperspectral imaging sensor used for this experiment covers the Shortwave Infrared (SWIR); this range of the electromagnetic spectrum has shown positive response in detecting different HC content,[3] for mineral and lithological mapping[27] and the distribution and mapping of mafic and ultramafic units in Cape Smith Belt.[28] The soil sample and the HC types chosen for this experiment have been used extensively as reference in the literature for assessments of HC contamination with different soil types.[3,4]

The Deep Neural Networks (DNN) experiment was conducted using MatLab 2017a. The experiments were carried out on an LG desktop with Intel (R) core $(TM)^2$ Duo CPU 3.00 GHZ processor. The DNN was trained with scale conjugate back propagation which adopts the chain rule derivate with learning rate α set to 0.9, A subset of the raw hyperspectral vector was fed into the network for all the obtained datasets which corresponds to 4000×288 for samples mixed with ethanol, bio-ethanol and diesel (where each 1000 point corresponds to 5 ml, 10 ml, 15 ml and 20 ml of the HC content as mixed in the experiment) while for sample mixed with

petrol, a subset of 7000×288 was used. 288 points were used as input for all the datasets, the input data were then propagated through 3 successive hidden layers each containing 30 nodes and 1 output corresponding to the abundance of the HC. The labeled samples were split into 50 % training and 50 % generalization sets. The algorithm was run iteratively with 1000 epochs, the training stops when the maximum number of epoch is reached or when the performance is minimized to the goal and the validation performance increases more than the maximum recorded. A sigmoid activation function was computed to predict the probabilities of the different categories. Finally, the extracted high level features were then flattened to fixed dimensional vectors, which were fully connected to the output layer for prediction estimation of the quantity of the HCs in each dataset.

VCA, which is a popular endmember extraxtion method, was used as a baseline to identify and compare the endmembers as mixed in the dataset for the accuracy of the deep neural network model. VCA has proven to show excellent results in literature as a method to find pure materials in mixed pixels.[20] Results of the experiments were based on the prediction estimation error of the test data and the generalized dataset as summarized in (Table 2).

Table 2. Abundance estimation error of the dataset showing training and generalization accuracy of the deep neural network model.

Data	Training	Generalization
Sand clay loam -Petrol	0.0523	0.0535
Sand clay loam - Ethanol	0.0400	0.0422
Sand clay loam - Diesel	0.0441	0.0457
Sand clay loam - Biodiesel	0.0482	0.0486

The effectiveness of the proposed deep neural network model has been demonstrated to quantify the amount of HCs in the soil sample with relatively low estimation error thus high accuracy. From the four datasets, the soil sample contaminated with ethanol predicted the lowest average abundance estimation error of about 0.0400 for the training data and 0.0422 on the generalization dataset as shown in table 2. The prediction estimation on all the datasets has been demonstrated in figures 4, 5, 6 and 7 by plotting the output of the DNN against the real abundances.
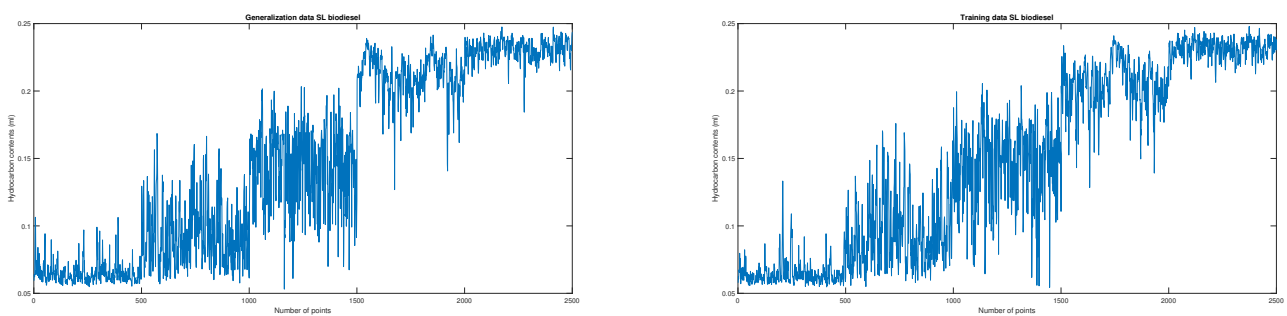

Figure 4. Training and generalization error plots for sandy clay loam soil contaminated with biodiesel.
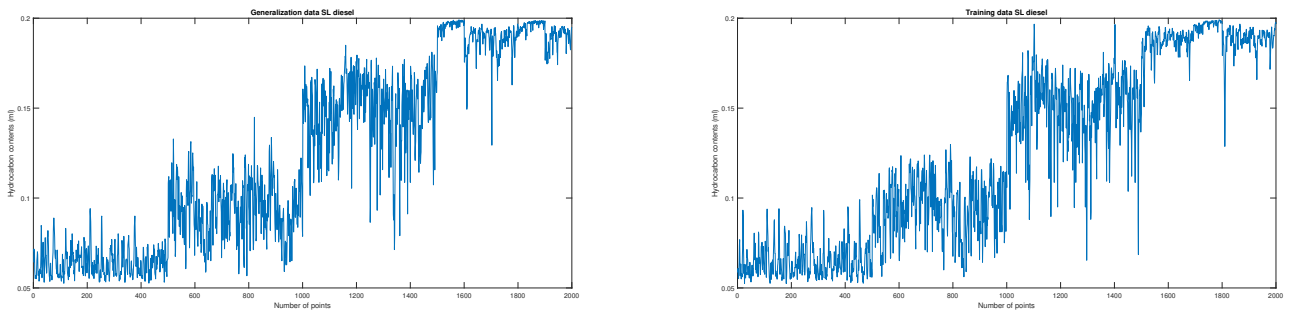
Figure 5.   Training and generalization error plots for sandy clay loam soil contaminated with diesel.
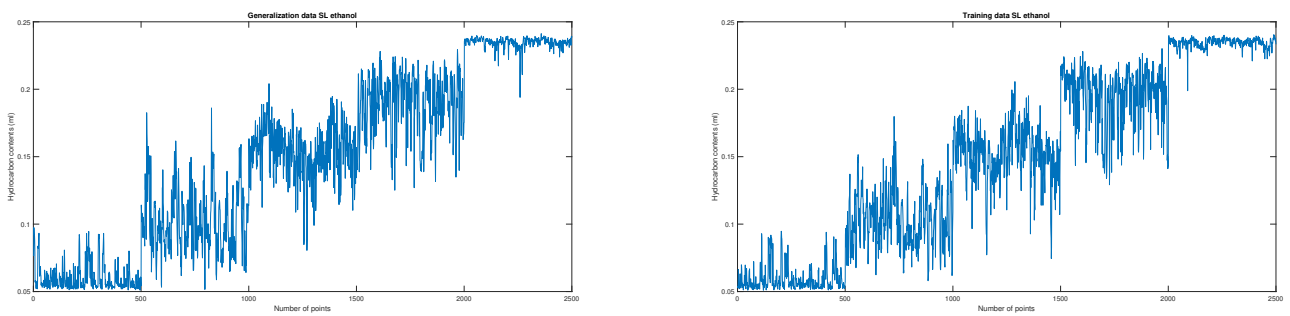

Figure 6.   Training and generalization error plots for sandy clay loam soil contaminated with ethanol.
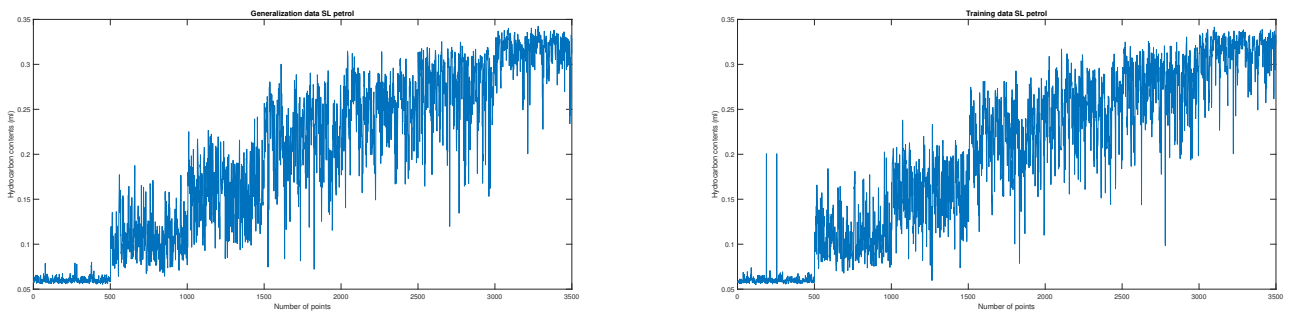

Figure 7.   Training and generalization error plots for sandy clay loam soil contaminated with petrol.

4. CONCLUSIONS

In this paper, we proposed a deep neural network model to correctly estimate hydrocarbon spills on sandy clay loam soil sample measured using spectroscopy analysis, the proposed method could achieve higher accuracy using the experimental datasets. The deep neural network was trained with the scale conjugate back propagation algorithm, the network had 3 hidden layers each with 30 nodes. Experimental validation shows that the deep neural network approach presents superior performance on all the datasets used with low average abundance estimation error. The quite promising quantitative evaluation indicate the high potentials of the developed approach. Finally, in the future, we will aim to apply the deep neural network and deep learning approach to quantify the amount of HC using different soil types.

4.1 Acknowledgments

REFERENCES

[1] Douglas, R. K., Nawar, S., Alamar, M. C., Mouazen, A., and Coulon, F., "Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-nir spectroscopy and regression techniques," *Science of the Total Environment* **616**, 147–155 (2018).

[2] Medina, R., Gara, P. M. D., Fernández-González, A. J., Rosso, J. A., and Del Panno, M. T., "Remediation of a soil chronically contaminated with hydrocarbons through persulfate oxidation and bioremediation," *Science of the Total Environment* **618**, 518–530 (2018).

[3] Scafutto, R. D. M. and de Souza Filho, C. R., "Quantitative characterization of crude oils and fuels in mineral substrates using reflectance spectroscopy: Implications for remote sensing," *International Journal of Applied Earth Observation and Geoinformation* **50**, 221–242 (2016).

[4] Okparanma, R. N. and Mouazen, A. M., "Visible and near-infrared spectroscopy analysis of a polycyclic aromatic hydrocarbon in soils," *The Scientific World Journal* **2013** (2013).

[5] Fernández-Varela, R., Andrade, J., Muniategui, S., Prada, D., and Ramírez-Villalobos, F., "The comparison of two heavy fuel oils in composition and weathering pattern, based on ir, gc-fid and gc–ms analyses: Application to the prestige wreackage," *Water research* **43**(4), 1015–1026 (2009).

[6] Scafutto, R. D. M., de Souza Filho, C. R., and de Oliveira, W. J., "Hyperspectral remote sensing detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for onshore exploration and monitoring," *ISPRS Journal of Photogrammetry and Remote Sensing* **128**, 146–157 (2017).

[7] Chakraborty, S., Weindorf, D. C., Li, B., Ali, M. N., Majumdar, K., and Ray, D., "Analysis of petroleum contaminated soils by spectral modeling and pure response profile recovery of n-hexane," *Environmental Pollution* **190**, 10–18 (2014).

[8] Okparanma, R. N. and Mouazen, A. M., "Determination of total petroleum hydrocarbon (tph) and polycyclic aromatic hydrocarbon (pah) in soils: a review of spectroscopic and nonspectroscopic techniques," *Applied Spectroscopy Reviews* **48**(6), 458–486 (2013).

[9] van der Meijde, M., Knox, N. M., Cundill, S. L., Noomen, M. F., Van der Werff, H., and Hecker, C., "Detection of hydrocarbons in clay soils: A laboratory experiment using spectroscopy in the mid-and thermal infrared," *International journal of applied earth observation and geoinformation* **23**, 384–388 (2013).

[10] Ahmed, A. M., Duran, O., Zweiri, Y., and Smith, M., "Hybrid spectral unmixing: using artificial neural networks for linear/non-linear switching," *Remote Sensing* **9**(8), 775 (2017).

[11] Xu, M., Zhang, L., Du, B., Zhang, L., Fan, Y., and Song, D., "A mutation operator accelerated quantumbehaved particle swarm optimization algorithm for hyperspectral endmember extraction," *Remote Sensing* **9**(3), 197 (2017).

[12] Giorgio, L. and Frate, F. D., "A neural network approach for pixel unmixing in hyperspectral data," tech. rep., Earth Observation Laboratory- Tor Vergata University Via del Politecnico, 1–00133 Rome, Italy.

[13] Hu, X. and Yuan, Y., "Deep-learning-based classification for dtm extraction from als point cloud," *Remote Sensing* **8**(9), 730 (2016).

[14] Makantasis, K., Karantzalos, K., Doulamis, A., and Doulamis, N., "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in [*Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*], 4959–4962, IEEE (2015).

[15] Petersson, H., Gustafsson, D., and Bergstrom, D., "Hyperspectral image analysis using deep learninga review," in [*Image Processing Theory Tools and Applications (IPTA), 2016 6th International Conference on*], 1–6, IEEE (2016).

[16] Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y., "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected topics in applied earth observations and remote sensing* **7**(6), 2094–2107 (2014).

[17] Mnih, V. and Hinton, G. E., "Learning to label aerial images from noisy data," in [*Proceedings of the 29th International conference on machine learning (ICML-12)*], 567–574 (2012).

[18] Licciardi, G. A. and Del Frate, F., "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE transactions on Geoscience and remote sensing* **49**(11), 4163–4172 (2011).

[19] Mathieu, M., Roy, R., Launeau, P., Cathelineau, M., and Quirt, D., "Alteration mapping on drill cores using a hyspex swir-320m hyperspectral camera: Application to the exploration of an unconformity-related uranium deposit (saskatchewan, canada)," *Journal of Geochemical Exploration* **172**, 71–88 (2017).

[20] Nascimento, J. M. and Dias, J. M. B., "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on* **43**(4), 898–910 (2005).

[21] Weeks, A. R., [*Fundamentals of electronic image processing*], SPIE Optical Engineering Press Bellingham (1996).

[22] Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **5**(2), 354–379 (2012).

[23] Morchhale, S., *Deep convolutional neural networks for classification of fused hyperspectral and LiDAR data*, PhD thesis, Wake Forest University (2016).

[24] LeCun, Y; Bengio, Y. and Hinton, G., "Deep learning," *Nature* **521**(7553), 436–444 (2015).

[25] Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., and Zuair, M., "Deep learning approach for car detection in UAV imagery," *Remote Sensing* **9**(4), 312 (2017).

[26] Lord, B., "Remote sensing techniques for onshore oil and gas exploration," *The Leading Edge* **36**(1), 24–32 (2017).

[27] Feng, J., Rogge, D., and Rivard, B., "Comparison of lithological mapping results from airborne hyperspectral vnir-swir, lwir and combined data," *International Journal of Applied Earth Observation and Geoinformation* **64**, 340–353 (2018).

[28] Rogge, D., Rivard, B., Segl, K., Grant, B., and Feng, J., "Mapping of nicu–pge ore hosting ultramafic rocks using airborne and simulated enmap hyperspectral imagery, nunavik, canada," *Remote sensing of environment* **152**, 302–317 (2014).

Application of Hybrid Switch Method to Quantify Oil Spills

Asmau M. Ahmed[1], Olga Duran, [1], Yahya Zweiri, [2] and Mike Smith, [3]

Abstract— Oil spill occurs across large landscape in a variety of soils which causes serious concern to the environment, therefore, it is important to identify and quantify spills at an early stage. Linear umixing methods are primary used due to their simplicity and computational cost with respect to non linear to monitor oil spills. However, the effectiveness of the different methods to quantify oil spills has not been assessed yet. Here we propose to robustly choose the most suitable method among linear and non linear spectral unmixing approaches to quantify different Hydrocarbon (HCs) substances in different soils. Hypersectral data sets have been acquired using mixtures of different HCs and soils. Then Artificial Neural Networks was used to switch between linear and non-linear methods to assess the most suitable method in quantifying the amount of spills. Results are presented for Vertex Component Analysis (VCA) and Fully Constrained Least Square Method (FCLS) for the linear models, and the Polynomial Post Nonlinear Mixing Model (PPNMM) and Generalised Bilinear Model (GBM).

I. INTRODUCTION

Hydrocarbons refers to chemical substances formed exclusively from carbon and hydrogen. Hydrocarbons (HC) are naturally occurring substances and depending on the length of the carbon chain, HCs often occur in different forms; solid, liquid and gas [1]. These often occur as a result of decay of organic substances trapped within sedimentary rocks. High temperatures and pressures convert the trapped malta into hydrocarbons. Liquid hydrocarbons found in nature are also referred to as crude oil [1]. Crude oil consist of a complex mixture of hydrocarbons of various molecular weights, in addition nitrogen, sulphur and oxygen occur in small quantities [2].

It is important to have knowledge about the concentration and nature of oil spill in order to track it's propagation in the environment, assess its risk and propose remediation strategies [3]. Fast and accurate determination of the affected area is needed, in order to effectively protect communities and species affected by the spill, particularly if monitoring large areas affected by an oil spill or aged oil transporting facilities [4] .

A hyperspectral camera captures the light intensity (radiance) for a large number of spectra. More specifically, Near and Shortwave Infrared (NIR SWIR) spectroscopy is a popular method for detecting, quantifying and mapping HCs in contaminated soils with reasonable accuracy [4]. Moreover, NIR SWIR spectra have proven effective at identifying and predicting soil organic and inorganic material content [5]. HCs demonstrate good absorption in spectral bands between 1200 mm, 1725 nm and 2310 nm [3]. Therefore, spectral information obtained in the NIR SWIR spectrum output excellent results for both quantitative and qualitative analysis of detecting HCs in soils [4].

Spectral Unmixing (SU) is described as the process of identifying spectral signatures of materials referred to as endmembers and estimating their relative abundance to the measured spectra [6]. Endmembers play an important role in exploring spectral information of a hyperspectral image [7]. SU often requires the definition of the mixing model underlying the observations as presented on the data. A mixing model describes how the endmembers combine to form the mixed spectrum as measured by the sensor [8]. Given the mixing model, SU then estimates the inverse of the formation process to infer the quantity of interest, specifically the endmembers, and abundance from the collected spectra [9]. The two main approaches to spectral unmixing are: linear and nonlinear. Analysis with these methods generally outputs excellent results as reported in the literature [6].

Artificial Neural Networks (ANN) are mathematical models initially developed to mimic the complex pattern of neuron interconnections as presumed in the human brain [10]. Presently, a lot of feed-forward neural network models have been extensively studied in the literature that show excellent results [11]. The deeper network layers provide excellent results with the ability to extract stronger features but in turn leads to a vanishing gradient [12]. The objective of this paper is to use hybrid switching method between linear and nonlinear spectral unmixing using artificial neural networks to identify and quantify different Hydrocarbon (HC) substances in real controlled hyperspectral datasets.

II. MATERIALS AND METHODS

A. Materials

Different Hydrocarbon (HC) types used in spectral mixing experiment includes: Diesel, Bio - diesel and Ethanol. Soil type used in this experiment is Clay. Clay is made of fine material particles which largely accounts for its properties, the particles are smooth and in a collidal state. Clay particles retains and absorb water, it also exhibit some properties such as flocculation, deflocculation , plasticity and stickiness [13].

B. Sample Preparation

In the experiment, the preparation of each sample began with the addition of 50g of clay soil to a petri dish 12cm in diameter, followed by the addition of 5ml of the HC to the soil using a syringe. A disposable plastic spoon was

[1] Faculty of Science, Engineering and Computing, Kingston University London, London SW15 3DW, UK. `o.duran@kingston.ac.uk`
[2] Khalifa University of Science and Technology, P.O. Box 127788, Abu Dhabi, United Arab Emirates
[3] School of Geography, Earth and Environmental Sciences. University of Plymouth, Drake Circus Plymouth Devon. PL4 8AA. UK.

used to homogenize the mixture after which the sample was measured with a Hyspex SWIR 320m camera. In the same sample, a further addition of 5 ml was added to the mixture; again the disposable spoon was used to homogenize the mixture and another measurement taken. This procedure was repeated with increments of 5 ml until a total of 25 ml was added or until the soil sample mixture was saturated and formed a shallow local pool.

Fig. 1. Clay Soil mixed with 5 ml and 25 ml diesel.

This procedure was repeated with all three HCs. A total of 15 samples (Table 1) were produced, five mixtures for each HC, mixed and homogenized at the addition of every 5 ml, measured and scanned with Hyspex SWIR 320 m camera under constant illumination.

Fig. 2. Scanning process of the dataset.

TABLE I

SAMPLES GENERATED FOR EACH COMBINATION MADE IN THE
EXPERIMENT AND THEIR SATURATION LEVELS

Soil type	Hydrocarbon type	Number of samples	HC saturation (ml)
Clay	Diesel	5	25
Clay	Bio- diesel	5	25
Clay	Ethanol	5	25

C. Hyperspectral Imaging

The data was obtained using Hyspex SWIR 320 m camera. The camera is equipped with a Mercury Cadmium- Telluride (MCT or HgCaTe) detector array which simultaneously acquires a full Short Wave Infa-Red (SWIR) spectrum, with a spectral sample interval of 5 nm between 1300 and 2500 nm, each along a line of 384 pixels [14]. The images acquired are 700×384 pixels. The obtained datsets were then analyzed to extract the endmembers and estimate the percentages of the HCs using the hybrid switch method.

D. Unmixing Methods

The hybrid switch method as described in [6] was used in this study, the methods are currently state of the art methods and have been used extensively in the literature. They
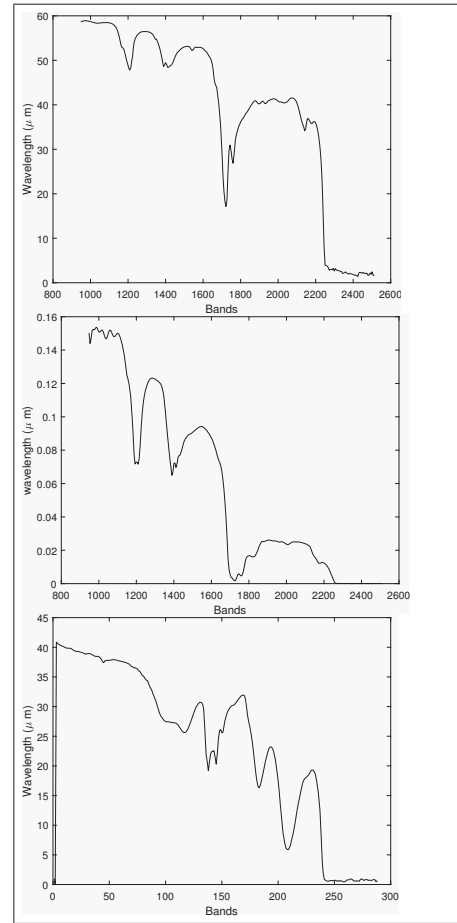

Fig. 3. Spectral Reflectance of the Petroleum Hydrocarbons from the top; **(a)** Biodiesel, **(b)** Diesel and **(c)** Ethanol

are: the Vertex Component Analysis (VCA) [15] and Fully Constrained Least Square Method (FCLS) [16] for the linear models, and the Polynomial Post Nonlinear Mixing Model (PPNMM) [17] and Generalized Bilinear Model (GBM) [18] for the nonlinear models. The linear mixing model is applicable when the photon interacts with a single cover type (checker board scene) while non linearity occurs when the photon interacts with different materials before reaching the sensor. The question of using either linear or nonlinear methods for spectral unmixing is still an unresolved problem. We propose hybridization between the linear and nonlinear spectral unmixing methods with Artificial Neural Networks (ANN) to conduct the switch between the two models. The models used are;

1) Vertex Component Analysis (VCA): This algorithm is based on the geometry of convex sets and exploits the fact that endmembers occupy the vertices of a simplex [19]. The VCA algorithm assumes the presence of spectrally pure pixels in a dataset and iteratively projects the data onto the direction orthogonal to the subspace spanned by the end members that are already determined [20]. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are

exhausted [21].

2) Fully Constrained Least Square Method (FCLS): The FCLS algorithm is derived from an unconstrained least square based orthogonal subspace projection [16]; in this method, negative values are considered 0 and the abundance fractions of the remaining material signatures are normalized to 1. FCLS utilizes a simplex method to produce a set of feasible solutions for spectral unmixing of material signatures while discarding the negative abundance values of the remaining material signatures to unity [16].

3) Polynomial Post Nonlinear Mixture (PPNM): This model assumes that the reflectance of an image are nonlinear functions of pure spectral components contaminated by additive noise; the nonlinear functions are often approximated using polynomial function leading to a polynomial post nonlinear mixing model [17].

The model involves linear and quadratic functions of the abundances. In this case, the R-spectrum $Y = [y_1, ..., y_R]^T$ of a mixed pixel is defined as a nonlinear transformation g of a linear mixture of L spectra $\mathbf{m}_l$ contaminated by additive noise $\mathbf{n}$.

$$Y = g_p \left(\sum_{l=1}^{L} a_l \mathbf{m}_l \right) + \mathbf{n} \tag{1}$$

where $\mathbf{m}_l$ is the spectrum of the lth material in the scene, a_l its corresponding proportion, L is the number of endmembers contained in the image and g is an appropriate nonlinear function. Another motivation for the PPNMM is the Weierstrass approximation theorem which states that every continuous function defined on an interval can be uniformly approximated by a polynomial with any desired precision [17].

4) Generalized Bilinear Mixing Model: The GBM model introduces a second term that accounts for multiple photon interactions [22]. This model proposes that the spectrum of a mixed pixel, Y can be derived as follows:

$$Y = \sum_{l=1}^{L} a_l \mathbf{m}_l + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} \gamma_{i,j} a_i a_j \ \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n} \tag{2}$$

Where $\mathbf{m}_i \odot \mathbf{m}_j$ is the Hadamard (term by term) product of the i and j spectra, $\mathbf{m}_i$ is the spectrum of the endmember i, a_i is the corresponding abundance and $\mathbf{n}$ is an additive noise. The first model term describes the linear mixture model and the double sum models the nonlinear effect. $\gamma = [\gamma_{1,2}...,\gamma_{L-1, \ L}]$ is a real parameter vector, $\gamma_{i,j} \in (0,1)$, that quantifies the interaction between different spectral components. The parameter introduced in this model is used to obtain a more flexible model [23]. This model also adopts the positivity and sum to one constraints.

E. Hybrid Switch

The hybrid switching was experimented with Artificial Neural Networks (ANN), the networks were trained with scale conjugate back propagation which adopts the chain rule derivate. The training stops when the maximum number of

epoch is reached or when the performance is minimized to the goal and the validation performance increases more than the maximum recorded. The neural networks have 3 layers (input, hidden and output). A dataset of size 5000×288 was used as input, the hidden layer has 10 nodes and the output layer has 1 node which corresponds to the decision between the linear and nonlinear unmixing models.

III. EXPERIMENTAL SETUP AND RESULTS

In the experiment, each category of the mixture type was analyzed separately, where clay contaminated with each HC was used as input vector. A subset of 5000×288 per data set was used for the experiment. We randomly split the labeled samples into training, testing and validation sets. We implemented the ANN algorithm using MatLab 2018a. The experiments were carried out on LG desktop with intel (R) core $(TM)^2$ Duo CPU 3.00 GHZ processor.

Here, we want to know which of the unmixing methods best quantifies the amount of HC content in each sample. The experiments were conducted with 70% and 30% of training samples in order to check if the size of the training data affects the accuracy of the network. The remaining samples were split for testing and validation. Results of the experiments were based on the abundance estimation error of the individual and hybrid methods as summarized in Table 2.

TABLE II

AVERAGE ABUNDANCE ESTIMATION ERROR OF THE HYBRID METHODS WITH DIFFERENT NUMBERS OF TRAINING SAMPLES (70% AND 30%) AS COMPARED WITH THE ABUNDANCE ESTIMATION ERROR OF THE INDIVIDUAL METHODS.

(CLAY - BIODIESEL)	70% SAMPLES	30% SAMPLES
PPNMM	0.0217	0.0217
GBM	0.0082	0.0082
VCA	0.0955	0.0955
FCLS	0.0083	0.0083
Hybrid methods VCA – PPNMM	0.0073	0.0079
VCA – GBM	0.0104	0.0103
FCLS – PPNMM	0.0097	0.0097
FCLS – GBM	0.0082	0.0082
(CLAY - DIESEL)	70% SAMPLES	30% SAMPLES
PPNMM	0.0045	0.0045
GBM	0.3775	0.3775
VCA	0.0845	0.0845
FCLS	0.3776	0.3776
Hybrid methods VCA – PPNMM	0.0041	0.0043
VCA – GBM	0.0634	0.0640
FCLS – PPNMM	0.0048	0.0053
FCLS – GBM	0.3770	0.3776
(CLAY - ETHANOL)	70% SAMPLES	30% SAMPLES
PPNMM	0.0036	0.0036
GBM	0.7046	0.7046
VCA	0.7499	0.7499
FCLS	0.7048	0.7048
Hybrid methods VCA – PPNMM	0.0025	0.0030
VCA – GBM	0.2516	0.2513
FCLS – PPNMM	0.0047	0.0047
FCLS – GBM	0.0056	0.0060

The overall accuracy of the network was based on the training, validation and testing error of the trained network. From the results obtained (Table 3), it shows that the hybrid

methods are more effective in quantifying the amount of hydrocarbons in each sample with minimum abundance estimation error as compared to the individual methods.

TABLE III

ABUNDANCE ESTIMATION ERROR OF THE DATASETS SHOWING TRAINING, VALIDATION AND TESTING ACCURACY OF THE HYBRID METHODS WITH DIFFERENT TRAINING SAMPLES

70% TRAINING SAMPLES				
(CLAY - BIODIESEL)	VCA – PPNMM	VCA – GBM	FCLS – PPNMM	FCLS – GBM
TRAIN	0.0083	0.0137	0.0207	0.0246
VALIDATION	0.0082	0.0196	0.0694	0.0644
TEST	0.0082	0.0173	0.0713	0.0632
(CLAY - DIESEL)	VCA – PPNMM	VCA – GBM	FCLS – PPNMM	FCLS – GBM
TRAIN	0.0071	0.0635	0.0117	0.2183
VALIDATION	0.0093	0.1020	0.0128	0.0694
TEST	0.0078	0.1071	0.0138	0.0694
(CLAY - ETHANOL)	VCA – PPNMM	VCA – GBM	FCLS – PPNMM	FCLS – GBM
TRAIN	0.0039	0.3018	0.0177	0.0069
VALIDATION	0.0057	0.5385	0.0309	0.0060
TEST	0.0049	0.5439	0.0249	0.0077
30% TRAINING SAMPLES				
(CLAY - BIODIESEL)	VCA – PPNMM	VCA – GBM	FCLS – PPNMM	FCLS – GBM
TRAIN	0.0083	0.0169	0.0471	0.0103
VALIDATION	0.0083	0.0161	0.0401	0.0375
TEST	0.0083	0.0155	0.0403	0.0369
(CLAY - DIESEL)	VCA – PPNMM	VCA – GBM	FCLS – PPNMM	FCLS – GBM
TRAIN	0.0098	0.0723	0.0174	0.0514
VALIDATION	0.0097	0.0631	0.0184	0.0466
TEST	0.0117	0.0610	0.0165	0.0466
(CLAY - ETHANOL)	VCA – PPNMM	VCA – GBM	FCLS – PPNMM	FCLS – GBM
TRAIN	0.0044	0.4110	0.0455	0.0093
VALIDATION	0.0056	0.3704	0.0278	0.0097
TEST	0.0064	0.3628	0.0464	0.0151

From the four hybrid methods, the VCA – PPNMM method outperforms the other methods with the lowest average abundance estimation error of about 0.0073 when mixed with biodiesel, 0.0041 when mixed with diesel and 0.0025 when mixed with ethanol as shown in table 2. The overall accuracy of the hybrid switch methods was calculated, it shows that the PPNMM has an overall accuracy of 98%, followed by FCLS – PPNMM with an overall accuracy of 96 %, VCA – GBM has an overall accuracy of 95 % while FCLS – GBM has an accuracy of 93 %. The result is similar to what was obtained in [6]. From the individual methods, the average abundance estimation error as displayed on the table was for 100% of samples thus the values are same irrespective of the number of samples used, this is because the hybrid switching cannot be done on the individual methods alone. The PPNMM method produced the best results with lower abundance estimation error on all the mixture types in comparison to other individual methods. It could be said that the nonlinear models are best suited for this application. The abundance estimation error of the network produced very low error which can be concluded that the network is good for the prediction of HCs in soils. The network has similar pattern in terms of training, validation and testing abundance error with similar or almost the same values.

IV. CONCLUSIONS

In this paper, the hybrid switch method for switching between linear and nonlinear spectral unmixing of hyperspectral data based on artificial neural networks was used to estimate the quantity of petroleum hydrocarbons mixed with different soil samples. The endmembers were extracted using VCA while the abundances were estimated using individual and hybrid methods. The method was validated with ANN to propose the best method in the estimation of the abundance. Results show that the hybrid methods are more suitable than the individual technique with high overall accuracy and lower abundance estimation error compared to those obtained with the individual methods in particular, the VCA – PPNMM hybrid method proved to be the best with 98% accuracy in all the experiments conducted. With regards to the individual methods, the PPNMM nonlinear model proves to be a good method with lowest abundance estimation error in comparison to other individual methods. therefore, it can be concluded that the nonlinear models are suited for the task of estimating hydrocarbon spills on different soil types. Although, the abundances estimated by VCA were not the correct ones, we replaced the estimates with known endmembers from our experiment and obtained good results with the method. The size of datasets used for training the network produced similar results with different number of training samples used, therefore, it can be concluded that the size of the training data does not affect the accuracy of the hybrid methods in switching between linear and nonlinear spectral unmixing,

ACKNOWLEDGMENT

REFERENCES

[1] UNEP. *Environmental assessment of Ogoniland. 1–262 United National Environmental Programme (UNEP): Nairobi.*

[2] Gabriel Eweje. Environmental costs and responsibilities resulting from oil exploitation in developing countries: The case of the niger delta of nigeria. *Journal of Business Ethics*, 69(1):27–56, 2006.

[3] Reuben Nwomandah Okparanma and Abdul Mounem Mouazen. Determination of total petroleum hydrocarbon (tph) and polycyclic aromatic hydrocarbon (pah) in soils: a review of spectroscopic and non-spectroscopic techniques. *Applied Spectroscopy Reviews*, 48(6):458–486, 2013.

[4] Rebecca Del'Papa Moreira Scafutto, Carlos Roberto de Souza Filho, and Wilson José de Oliveira. Hyperspectral remote sensing detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for onshore exploration and monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128:146–157, 2017.

[5] Rebecca DelPapa Moreira Scafutto and Carlos Roberto de Souza Filho. Quantitative characterization of crude oils and fuels in mineral substrates using reflectance spectroscopy: Implications for remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 50:221–242, 2016.

[6] Asmau M Ahmed, Olga Duran, Yahya Zweiri, and Mike Smith. Hybrid spectral unmixing: using artificial neural networks for linear/non-linear switching. *Remote Sensing*, 9(8):775, 2017.

[7] Mingming Xu, Liangpei Zhang, Bo Du, Lefei Zhang, Yanguo Fan, and Dongmei Song. A mutation operator accelerated quantum-behaved particle swarm optimization algorithm for hyperspectral endmember extraction. *Remote Sensing*, 9(3):197, 2017.

[8] Bruce Hapke. Bidirectional reflectance spectroscopy: 1. theory. *Journal of Geophysical Research: Solid Earth*, 86(B4):3039–3054, 1981.

[9] Abderrahim Halimi, Yoann Altmann, Gerald S Buller, Steve McLaughlin, William Oxford, Damien Clarke, and Jonathan Piper. Robust unmixing algorithms for hyperspectral imagery. In *Sensor Signal Processing for Defence (SSPD), 2016*, pages 1–5. IEEE, 2016.

[10] Stefan Raith, Eric Per Vogel, Naeema Anees, Christine Keul, Jan-Frederik Güth, Daniel Edelhoff, and Horst Fischer. Artificial neural networks as a powerful numerical tool to classify specific features of a tooth based on 3d scan data. *Computers in biology and medicine*, 80:65–76, 2017.

[11] Te Han, Dongxiang Jiang, Qi Zhao, Lei Wang, and Kai Yin. Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of Measurement and Control*, page 0142331217708242, 2017.

[12] Phil Kim. Matlab deep learning: With machine learning, neural networks and artificial intelligence. 2017.

[13] JB Page. Role of physical properties of clays in soil science. *Clays and Clay Minerals*, 1:167–176, 1952.

[14] Magali Mathieu, Régis Roy, Patrick Launeau, Michel Cathelineau, and David Quirt. Alteration mapping on drill cores using a hyspex swir-320m hyperspectral camera: Application to the exploration of an unconformity-related uranium deposit (saskatchewan, canada). *Journal of Geochemical Exploration*, 172:71–88, 2017.

[15] José MP Nascimento and José M Bioucas Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(4):898–910, 2005.

[16] Daniel C Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3):529–545, 2001.

[17] Yoann Altmann, Abderrahim Halimi, Nicolas Dobigeon, and Jean-Yves Tourneret. Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery. *Image Processing, IEEE Transactions on*, 21(6):3017–3025, 2012.

[18] Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret. Bilinear models for nonlinear unmixing of hyperspectral images. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pages 1–4. IEEE, 2011.

[19] José MP Nascimento and José M Bioucas Dias. Does independent component analysis play a role in unmixing hyperspectral data? *Geoscience and Remote Sensing, IEEE Transactions on*, 43(1):175–187, 2005.

[20] Arthur R Weeks. *Fundamentals of electronic image processing*. SPIE Optical Engineering Press Bellingham, 1996.

[21] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379, 2012.

[22] Yoann ALTMANN. *Nonlinear spectral unmixing of hyperspectral images*. PhD thesis, lInstitut National Polytechnique de Toulouse (INP Toulouse), October 2013.

[23] Abderrahim Halimi, Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(11):4153–4162, 2011.

Reference

[1] Nawar Said Alamar M Carmen Mouazen AM Douglas Reward K and Frederic Coulon. "Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques". In: *Science of the Total Environment* 616 (2018), pp. 147–155.

[2] J Li. "Spill Management for the Toronto AOC: The City of Toronto Study". In: *Report and factsheet prepared for the Great Lakes Sustainability Fund, Burlington, Ontario, Canada* (2002).

[3] Campbell Robertson and Clifford Krauss. "Gulf spill is the largest of its kind, scientists say". In: *The New York Times* 2.8 (2010).

[4] Abdul Aziz Al-Majed, Abdulrauf Rasheed Adebayo, and M Enamul Hossain. "A sustainable approach to controlling oil spills". In: *Journal of environmental management* 113 (2012), pp. 213–227.

[5] Ding-Cheng Feng, Feng Chen, and XU Wen-Li. "Detecting local manifold structure for unsupervised feature selection". In: *Acta Automatica Sinica* 40.10 (2014), pp. 2253–2261.

[6] Rebecca Del'Papa Moreira Scafutto, Carlos Roberto de Souza Filho, and Benoit Rivard. "Characterization of mineral substrates impregnated with crude oils using proximal infrared hyperspectral imaging". In: *Remote Sensing of Environment* 179 (2016), pp. 116–130.

[7] Guy Schwartz, Eyal Ben-Dor, and Gil Eshel. "Quantitative analysis of total petroleum hydrocarbons in soils: comparison between reflectance

spectroscopy and solvent extraction by 3 certified laboratories". In: *Applied and Environmental Soil Science* 2012 (2012).

[8]   Plaza Antonio. Dobigeon Nicolas. Parente Mario. Du Qian. Gader Paul Bioucas-Dias José M. and Chanussot Jocelyn. "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches". In: *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 5.2 (2012), pp. 354–379.

[9]   Antonio Plaza et al. "A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles". In: *Pattern Recognition* 37.6 (2004), pp. 1097–1116.

[10] N Keshava and J. F. Mustard. "Spectral unmixing". In: *IEEE Signal Process. Mag* vol. 19.no.1 (2002), pp. 44–57.

[11] Nicolas Dobigeon et al. "Nonlinear unmixing of hyperspectral images: Models and algorithms". In: *Signal Processing Magazine, IEEE* 31.1 (2014), pp. 82–94.

[12] Giorgio A Licciardi and Fabio Del Frate. "Pixel unmixing in hyperspectral data by means of neural networks". In: *IEEE transactions on Geoscience and remote sensing* 49.11 (2011), pp. 4163–4172.

[13] Baronio Annalisa Signoroni Alberto. Savardi Mattia. and Sergio Benini. "Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review". In: *Journal of Imaging* 5.5 (2019), p. 52.

[14] Youngsinn Sohn and Roger M McCoy. "Mapping desert shrub rangeland using spectral unmixing and modeling spectral mixtures with TM data". In: *Photogrammetric Engineering and Remote Sensing* 63.6 (1997), pp. 707–716.

[15] Oncel Tuzel, Fatih Porikli, and Peter Meer. "Region covariance: A fast descriptor for detection and classification". In: *European conference on computer vision*. Springer. 2006, pp. 589–600.

[16] Dobigeon Nicolas Altmann Yoann Halimi Abderrahim and Jean-Yves Tourneret. "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery". In: *Image Processing, IEEE Transactions on* 21.6 (2012), pp. 3017–3025.

[17] Gachagan Anthony Nwankpa Chigozie Ijomah Winifred and Stephen Marshall. "Activation Functions: Comparison of trends in Practice and Research for Deep Learning". In: *arXiv preprint arXiv:1811.03378* (2018).

[18] Whidborne James F Althoefer Kaspar Zweiri Yahya H and Lakmal D Seneviratne. "A new three-term backpropagation algorithm with convergence analysis". In: *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*. Vol. 4. IEEE. 2002, pp. 3882–3887.

[19] Krizhevsky Alex Sutskever Ilya Srivastava Nitish Hinton Geoffrey and Ruslan Salakhutdinov. "Dropout: A simple way to prevent neural networks from overfitting". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 1929–1958.

[20] UNEP. *Environmental assessment of Ogoniland. 1–262 United National Environmental Programme (UNEP): Nairobi*. 2011, pp. 1–262.

[21] Nareenart Raksuntorn and Qian Du. "A new linear mixture model for hyperspectral image analysis". In: *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*. Vol. 3. IEEE. 2008, pp. III–258.

[22] Gabriel Eweje. "Environmental costs and responsibilities resulting from oil exploitation in developing countries: The case of the Niger Delta of Nigeria". In: *Journal of Business Ethics* 69.1 (2006), pp. 27–56.

[23]   Reuben Nwomandah Okparanma and Abdul Mounem Mouazen. "Determination of total petroleum hydrocarbon (TPH) and polycyclic aromatic hydrocarbon (PAH) in soils: a review of spectroscopic and nonspectroscopic techniques". In: *Applied Spectroscopy Reviews* 48.6 (2013), pp. 458–486.

[24]   Rebecca Del'Papa Moreira Scafutto, Carlos Roberto de Souza Filho, and Wilson José de Oliveira. "Hyperspectral remote sensing detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for onshore exploration and monitoring". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 128 (2017), pp. 146–157.

[25]   Foudan Salem and Menas Kafatos. "Hyperspectral image analysis for oil spill mitigation". In: *Paper presented at the 22nd Asian Conference on Remote Sensing*. Vol. 5. 2001, p. 9.

[26]   Reuben N Okparanma and Abdul M Mouazen. "Visible and near-infrared spectroscopy analysis of a polycyclic aromatic hydrocarbon in soils". In: *The Scientific World Journal* 2013 (2013).

[27]   Li Bin Ali Md Nasim Majumdar K Chakraborty Somsubhra Weindorf David C and DP Ray. "Analysis of petroleum contaminated soils by spectral modeling and pure response profile recovery of n-hexane". In: *Environmental Pollution* 190 (2014), pp. 10–18.

[28]   JM Muniategui S Prada D Fernández-Varela R Andrade and F Ramírez-Villalobos. "The comparison of two heavy fuel oils in composition and weathering pattern, based on IR, GC-FID and GC–MS analyses: Application to the prestige wreackage". In: *Water research* 43.4 (2009), pp. 1015–1026.

[29]   Rebecca Del'Papa Moreira Scafutto and Carlos Roberto de Souza Filho. "Quantitative characterization of crude oils and fuels in mineral substrates using reflectance spectroscopy: Implications for remote sensing". In: *International Journal of Applied Earth Observation and Geoinformation* 50 (2016), pp. 221–242.

[30]   Soriano-Disla José M Kirk Joel Janik Leslie J Forrester Sean T McLaughlin Mike J Webster Grant T and Richard J Stewart. "Rapid prediction of total petroleum hydrocarbons in soil using a hand-held mid-infrared field instrument". In: *Talanta* 160 (2016), pp. 410–416.

[31]   Mayr Thomas Okparanma Reuben N Coulon Frederic and Abdul M Mouazen. "Mapping polycyclic aromatic hydrocarbon and total toxicity equivalent soil concentrations by visible and near-infrared spectroscopy". In: *Environmental Pollution* 192 (2014), pp. 162–170.

[32]   Ben Lord. "Remote sensing techniques for onshore oil and gas exploration". In: *The Leading Edge* 36.1 (2017), pp. 24–32.

[33]   Camilla Brekke and Anne HS Solberg. "Oil spill detection by satellite remote sensing". In: *Remote sensing of environment* 95.1 (2005), pp. 1–13.

[34]   Maya Jha, Jason Levy, and Yang Gao. "Advances in remote sensing for oil spill disaster management: state-of-the-art sensors technology for oil spill surveillance". In: *Sensors* 8.1 (2008), pp. 236–255.

[35]   Mervin F Fingas and Carl E Brown. "Review of oil spill remote sensing". In: *Spill Science & Technology Bulletin* 4.4 (1997), pp. 199–208.

[36]   Carl E Brown, Mervin F Fingas, and Robert Hawkins. "Synthetic aperture radar sensors: Viable for marine oil spill response?" In: *ARCTIC AND MARINE OILSPILL PROGRAM TECHNICAL SEMINAR*. Vol. 1. Environment Canada; 1999. 2003, pp. 299–310.

[37]   Merv Fingas and Carl Brown. "Review of oil spill remote sensing". In: *Marine pollution bulletin* 83.1 (2014), pp. 9–23.

[38]   Frank E Taylor Charles Myhre Douglas Murch Brock Odriozola Ana L Hu Chuanmin Müller-Karger and Gonzalo Godoy. "MODIS detects oil spills in Lake Maracaibo, Venezuela". In: *Eos, Transactions American Geophysical Union* 84.33 (2003), pp. 313–319.

[39]   Cristina Bentz and F Pellon de Miranda. "Application of remote sensing data for oil spill monitoring in the Guanabara Bay, Rio de Janeiro, Brazil". In: *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International*. Vol. 1. IEEE. 2001, pp. 333–335.

[40]   K Grüner, R Reuter, and H Smid. "A new sensor system for airborne measurements of maritime pollution and of hydrographic parameters". In: *GeoJournal* 24.1 (1991), pp. 103–117.

[41]   John W Salisbury, Dana M D'Aria, and Floyd F Sabins Jr. "Thermal infrared remote sensing of crude oil slicks". In: *Remote Sensing of Environment* 45.2 (1993), pp. 225–231.

[42]   Edward A Cloutis. "Spectral reflectance properties of hydrocarbons: remote-sensing implications". In: *Science* 245.4914 (1989), pp. 165–168.

[43]   P Fine, ER Graber, and B Yaron. "Soil interactions with petroleum hydrocarbons: abiotic processes". In: *Soil Technology* 10.2 (1997), pp. 133–153.

[44]   Heather Freeman. "Evaluation of the use of hyperspectral imagery for identification of microseeps near Santa Barbara, California". In: *Project Report forM aster of Science in Geolo2 gy W est V irginia U niversity* 9 (2003).

[45]   Alan Williams and Geoff Lawrence. "The role of satellite seep detection in exploring the South Atlantic's ultradeep water". In: *Surface exploration case histories: Applications of geochemistry, magnetics, and remote sensing. AAPG Studies in Geology* 48 (2002), pp. 327–344.

[46]   Bruce Hapke. "Bidirectional reflectance spectroscopy: 1. Theory". In: *Journal of Geophysical Research: Solid Earth* 86.B4 (1981), pp. 3039–3054.

[47]   Yoann Altmann. "Nonlinear spectral unmixing of hyperspectral images". PhD thesis. l'Institut National Polytechnique de Toulouse (INP Toulouse), 2013.

[48]   Wenzhi Liao. "Feature extraction and classification for hyperspectral remote sensing images". PhD thesis. Ghent University, 2012.

[49]   Adams JB Willis SC Fischer AF Gillespie A R Smith MO and DE Sabol. "Interpretation of residual images: spectral mixture analysis of AVIRIS images, Owens Valley, California". In: *Proceedings of the 2nd airborne visible/infrared imaging spectrometer (AVIRIS) workshop*. NASA Jet Propulsion Laboratory Pasadena, CA. 1990, pp. 243–270.

[50]   Scientific Technical Report (SRT). *Hyperspectral algorithm: Report of the frame of Enmap preparation activities*. Tech. rep. Potsdam: Deutsches Geo Forschungs Zentrum GFZ, 268 S.Scientific, 2008.

[51]   Olga Duran and Maria Petrou. "Subpixel temporal spectral imaging". In: *Pattern Recognition Letters* 48 (2014), pp. 15–23.

[52]   Zhao Xiao-Li Althouse Mark LG Chang Chein-I and Jeng Jong Pan. "Least squares subspace projection approach to mixed pixel classification for hyperspectral images". In: *Geoscience and Remote Sensing, IEEE Transactions on* 36.3 (1998), pp. 898–912.

[53] Daniel C Heinz and Chein-I Chang. "Unsupervised fully constrained squares linear spectral mixture analysis method for multispectral imagery". In: *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*. Vol. 4. IEEE. 2000, pp. 1681–1683.

[54] Chein-I Chang. "Adaptive Linear Spectral Mixture Analysis". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.3 (2017), pp. 1240–1253.

[55] Wei Duren-Zhana Shi Zhenwei Tang and Zhiguo Jiang. "Subspace matching pursuit for sparse unmixing of hyperspectral data". In: *IEEE Transactions on Geoscience and Remote Sensing* 52.6 (2014), pp. 3256–3274.

[56] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. "Sparse unmixing of hyperspectral data". In: *IEEE Transactions on Geoscience and Remote Sensing* 49.6 (2011), pp. 2014–2039.

[57] Ma Yong Mei-Xiaoguang Li Chang, Chengyin Liu, and Jiayi Ma. "Hyperspectral unmixing with robust collaborative sparse regression". In: *Remote Sensing* 8.7 (2016), p. 588.

[58] Pierre-Antoine Thouvenin, Nicolas Dobigeon, and Jean-Yves Tourneret. "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model". In: *IEEE Transactions on Signal Processing* 64.2 (2016), pp. 525–538.

[59] G M Foody and D.P Cox. "Subpixel Land Cover Composition Estimation Using Linear Mixture Model and Fuzzy Membership Functions". In: *International Journal of Remote Sensing* 15 (), pp. 619–631.

[60] José MP Nascimento and José M Bioucas Dias. "Vertex component analysis: A fast algorithm to unmix hyperspectral data". In: *Geoscience and Remote Sensing, IEEE Transactions on* 43.4 (2005), pp. 898–910.

[61]   Liu Jianjun Xiao Liang Sun Le Wu Zebin and Zhihui Wei. "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.3 (2015), pp. 1490–1503.

[62]   Chanussot Jocelyn Fauvel Mathieu Benediktsson Jón Atli and Johannes R Sveinsson. "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles". In: *IEEE Transactions on Geoscience and Remote Sensing* 46.11 (2008), pp. 3804–3814.

[63]   Haoyang Yu et al. "Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive markov random fields". In: *Remote Sensing* 8.4 (2016), p. 355.

[64]   Li Shanshan Li Jun Ni Li Gao Lianru and Bing Zhang. "Edge-constrained Markov random field classification by integrating hyperspectral image with LiDAR data over urban areas". In: *Journal of Applied Remote Sensing* 8.1 (2014), pp. 085089–085089.

[65]   Jia Xiuping Gao Lianru Zhang Bing Li Shanshan and Man Peng. "Adaptive Markov random field approach for classification of hyperspectral imagery". In: *IEEE Geoscience and Remote Sensing Letters* 8.5 (2011), pp. 973–977.

[66]   Luis O Jimenez and David A Landgrebe. "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 28.1 (1998), pp. 39–54.

[67]   Aguilar PL Pérez R Linaje M Preciado JC Martínez P Gualtieri JA and A Plaza. "Hyperspectral image classification using a self-organizing map". In: *Summaries of the X JPL Airborne Earth Science Workshop*. 2001.

[68]   Mathieu Fauvel et al. "Advances in spectral-spatial classification of hyperspectral images". In: *Proceedings of the IEEE* 101.3 (2013), pp. 652–675.

[69]   Harold M Horwitz et al. "Estimating the proportions of objects within a single resolution element of a multispectral scanner." In: (1971).

[70]   O Duran and M Petrous. "Mixed pixel classification in remote sensing—literature survey". In: *School of Electronics and Physical Sciences, University of Surrey: Guildford, UK* (2004).

[71]   Donald E Sabol, John B Adams, and Milton O Smith. "Quantitative subpixel spectral detection of targets in multispectral images". In: *Journal of Geophysical Research: Planets* 97.E2 (1992), pp. 2659–2672.

[72]   Jie Chen, Cédric Richard, and Paul Honeine. "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model". In: *Signal Processing, IEEE Transactions on* 61.2 (2013), pp. 480–492.

[73]   Joseph W Boardman, Fred A Kruse, and Robert O Green. "Mapping target signatures via partial unmixing of AVIRIS data". In: (1995).

[74]   Carlos González et al. "FPGA implementation of the pixel purity index algorithm for remotely sensed hyperspectral image analysis". In: *EURASIP Journal on Advances in Signal Processing* 2010.1 (2010), p. 969806.

[75]   Michael E Winter. "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data". In: *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics. 1999, pp. 266–275.

[76]   Antonio Plaza et al. "A quantitative and comparative analysis of end-member extraction algorithms from hyperspectral data". In: *IEEE transactions on geoscience and remote sensing* 42.3 (2004), pp. 650–663.

[77] Daniel C Heinz et al. "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery". In: *IEEE transactions on geoscience and remote sensing* 39.3 (2001), pp. 529–545.

[78] Agustin Ifarraguerri and C-I Chang. "Multispectral and hyperspectral image analysis with convex cones". In: *IEEE transactions on geoscience and remote sensing* 37.2 (1999), pp. 756–770.

[79] José MP Nascimento and José M Bioucas-Dias. "Hyperspectral unmixing algorithm via dependent component analysis". In: *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*. IEEE. 2007, pp. 4033–4036.

[80] José MP Nascimento and José M Bioucas-Dias. "Nonlinear mixture model for hyperspectral unmixing". In: *SPIE Europe Remote Sensing*. International Society for Optics and Photonics. 2009, pp. 74770I–74770I.

[81] Chang Li et al. "GBM-based unmixing of hyperspectral data using bound projected optimal gradient method". In: *IEEE Geoscience and Remote Sensing Letters* 13.7 (2016), pp. 952–956.

[82] Andrea Marinoni and Paolo Gamba. "Accurate detection of anthropogenic settlements in hyperspectral images by higher order nonlinear unmixing". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.5 (2016), pp. 1792–1801.

[83] Abderrahim Halimi et al. "Robust unmixing algorithms for hyperspectral imagery". In: *Sensor Signal Processing for Defence (SSPD), 2016*. IEEE. 2016, pp. 1–5.

[84] Wenyi Fan et al. "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data". In: *International Journal of Remote Sensing* 30.11 (2009), pp. 2951–2962.

[85]   Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret. "Bilinear models for nonlinear unmixing of hyperspectral images". In: *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*. IEEE. 2011, pp. 1–4.

[86]   Abderrahim Halimi et al. "Nonlinear unmixing of hyperspectral images using a generalized bilinear model". In: *Geoscience and Remote Sensing, IEEE Transactions on* 49.11 (2011), pp. 4153–4162.

[87]   Althouse Mark L Guilfoyle Kerri J and Chein-I Chang. "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks". In: *Geoscience and Remote Sensing, IEEE Transactions on* 39.10 (2001), pp. 2314–2318.

[88]   Perez Rosa Plaza Javier Plaza Antonio and Pablo Martinez. "On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images". In: *Pattern Recognition* 42.11 (2009), pp. 3032–3045.

[89]   Stefan Raith et al. "Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data". In: *Computers in biology and medicine* 80 (2017), pp. 65–76.

[90]   Te Han et al. "Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery". In: *Transactions of the Institute of Measurement and Control* (2017), p. 0142331217708242.

[91]   Kurt Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural networks* 4.2 (1991), pp. 251–257.

[92]   Yanling Wei et al. "Approaches to TS fuzzy-affine-model-based reliable output feedback control for nonlinear Itô stochastic systems". In: *IEEE Transactions on fuzzy systems* (2016).

[93] Sankar K Pal and Sushmita Mitra. "Multilayer perceptron, fuzzy sets, and classification". In: *IEEE Transactions on neural networks* 3.5 (1992), pp. 683–697.

[94] Olga Duran, Kaspar Althoefer, and Lakmal D Seneviratne. "Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network". In: *Automation Science and Engineering, IEEE Transactions on* 4.1 (2007), pp. 118–126.

[95] Ying Li, Haokui Zhang, and Qiang Shen. "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network". In: *Remote Sensing* 9.1 (2017), p. 67.

[96] Olga Duran and Maria Petrou. "A time-efficient clustering method for pure class selection". In: *Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International*. Vol. 1. IEEE. 2005, 4–pp.

[97] Ana Pérez-Hoyos et al. "Identification of ecosystem functional types from coarse resolution imagery using a self-organizing map approach: A case study for Spain". In: *Remote Sensing* 6.11 (2014), pp. 11391–11419.

[98] Brian S Penn. "Using self-organizing maps to visualize high-dimensional data". In: *Computers & Geosciences* 31.5 (2005), pp. 531–544.

[99] Yanling Wei et al. "A new design of H-infinity piecewise filtering for discrete-time nonlinear time-varying delay systems via TS fuzzy affine models". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2017).

[100] Hak Keung Lam, Hongyi Li, and Honghai Liu. "Stability analysis and control synthesis for fuzzy-observer-based controller of nonlinear systems: a fuzzy-model-based control approach". In: *IET Control Theory & Applications* 7.5 (2013), pp. 663–672.

[101] Yanling Wei et al. "Model reduction for continuous-time Markovian jump systems with incomplete statistics of mode information". In: *International Journal of Systems Science* 45.7 (2014), pp. 1496–1507.

[102] Huijiao Wang, Peng Shi, and Ramesh K Agarwal. "Network-based event-triggered filtering for Markovian jump systems". In: *International Journal of Control* 89.6 (2016), pp. 1096–1110.

[103] L Giorgio and Fabio Del Frate. *A Neural Network Approach For Pixel Unmixing In Hyperspectral Data*. Tech. rep. Earth Observation Laboratory-Tor Vergata University Via del Politecnico, 1–00133 Rome, Italy.

[104] S J Kumar U. Chiranjit and M T.V Ramachandra. "A Neural Network Based Hybrid Mixture Model to Extract Information from Nonlinear Mixed Pixels". In: *International Journal of Remote Sensing* 3 (), pp. 420 –441.

[105] Haobo Lyu, Hui Lu, and Lichao Mou. "Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection". In: *Remote Sensing* 8.6 (2016), p. 506.

[106] Nassim Ammour et al. "Deep learning approach for car detection in UAV imagery". In: *Remote Sensing* 9.4 (2017), p. 312.

[107] Hao Wu and Saurabh Prasad. "Convolutional Recurrent Neural Networks forHyperspectral Data Classification". In: *Remote Sensing* 9.3 (2017), p. 298.

[108] Cutler M.E.J Atkinson P M and H Lewis. "Mapping Sub-Pixel Proportional Land Cover With AVHRR Imagery". In: *International Journal of Remote Sensing* 18 (), pp. 917–935.

[109] Xingui He and Shaohua Xu. "Artificial neural networks". In: *Process Neural Networks: Theory and Applications* (2010), pp. 20–42.

[110] Antoniades Andreas. "Interpreting Biomedical Data via Deep Neural Networks". PhD thesis. PhD thesis, 2017.

[111] Mohamad H Hassoun. *Fundamentals of artificial neural networks*. MIT press, 1995.

[112] Phil Kim. "MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence". In: (2017).

[113] Yanming Guo et al. "Deep learning for visual understanding: A review". In: *Neurocomputing* 187 (2016), pp. 27–48.

[114] Yushi Chen et al. "Deep learning-based classification of hyperspectral data". In: *IEEE Journal of Selected topics in applied earth observations and remote sensing* 7.6 (2014), pp. 2094–2107.

[115] Yoshua LeCun Y; Bengio and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[116] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. "Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images". In: *Remote Sensing* 9.4 (2017), p. 368.

[117] Mesay Belete Bejiga et al. "A convolutional neural network approach for assisting avalanche search and rescue operations with uav imagery". In: *Remote Sensing* 9.2 (2017), p. 100.

[118] Antonio-Javier Gallego, Antonio Pertusa, and Pablo Gil. "Automatic Ship Classification from Optical Aerial Images with Convolutional Neural Networks". In: *Remote Sensing* 10.4 (2018), p. 511.

[119] Xiu Jin et al. "Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field". In: *Remote Sensing* 10.3 (2018), p. 395.

[120]   Jiaojiao Li et al. "Hyperspectral Classification Based on Texture Feature Enhancement and Deep Belief Networks". In: *Remote Sensing* 10.3 (2018), p. 396.

[121]   Yongyang Xu et al. "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters". In: *Remote Sensing* 10.1 (2018), p. 144.

[122]   Shaohui Mei et al. "Hyperspectral Image Spatial Super-Resolution via 3D Full Convolutional Neural Network". In: *Remote Sensing* 9.11 (2017), p. 1139.

[123]   Zhongling Huang, Zongxu Pan, and Bin Lei. "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data". In: *Remote Sensing* 9.9 (2017), p. 907.

[124]   Gang Fu et al. "Classification for high resolution remote sensing imagery using a fully convolutional network". In: *Remote Sensing* 9.5 (2017), p. 498.

[125]   Norbert Kruger et al. "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?" In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1847–1871.

[126]   Waseem Rawat and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review". In: *Neural computation* 29.9 (2017), pp. 2352–2449.

[127]   Bing Xu et al. "Empirical evaluation of rectified activations in convolutional network". In: *arXiv preprint arXiv:1505.00853* (2015).

[128]   Yichuan Tang. "Deep learning using linear support vector machines". In: *arXiv preprint arXiv:1306.0239* (2013).

[129] Ronald J Williams and David Zipser. "A learning algorithm for continually running fully recurrent neural networks". In: *Neural computation* 1.2 (1989), pp. 270–280.

[130] Chunyang Wu. "Structured Deep Neural Networks for Speech Recognition". PhD thesis. University of Cambridge, 2018.

[131] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult". In: *IEEE transactions on neural networks* 5.2 (1994), pp. 157–166.

[132] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[133] Junyoung Chung et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling". In: *arXiv preprint arXiv:1412.3555* (2014).

[134] Alex Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. "Hybrid speech recognition with deep bidirectional LSTM". In: *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*. IEEE. 2013, pp. 273–278.

[135] Qianqian Hao, Hua Zhang, and Jinkou Ding. "The hidden layer design for staked denoising autoencoder". In: *Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015 12th International Computer Conference on*. IEEE. 2015, pp. 150–153.

[136] Geoffrey E Hinton and Ruslan R Salakhutdinov. "Reducing the dimensionality of data with neural networks". In: *science* 313.5786 (2006), pp. 504–507.

[137] Joseph Turian, James Bergstra, and Yoshua Bengio. "Quadratic features and deep architectures for chunking". In: *Proceedings of Human*

Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers. Association for Computational Linguistics. 2009, pp. 245–248.

[138] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. "Rectifier non-linearities improve neural network acoustic models". In: *Proc. icml*. Vol. 30. 1. 2013, p. 3.

[139] Yann N Dauphin et al. "Language modeling with gated convolutional networks". In: *arXiv preprint arXiv:1612.08083* (2016).

[140] Vinod Nair and Geoffrey E Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010, pp. 807–814.

[141] Matthew D Zeiler et al. "On rectified linear units for speech processing". In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE. 2013, pp. 3517–3521.

[142] Ian J Goodfellow et al. "Maxout networks". In: *arXiv preprint arXiv:1302.4389* (2013).

[143] László Tóth. "Phone recognition with hierarchical convolutional deep maxout networks". In: *EURASIP Journal on Audio, Speech, and Music Processing* 2015.1 (2015), p. 25.

[144] Yushi Chen, Xing Zhao, and Xiuping Jia. "Spectral–spatial classification of hyperspectral data based on deep belief network". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.6 (2015), pp. 2381–2392.

[145] S Jacquemoud et al. "Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode". In: *Remote Sensing of Environment* 74.3 (2000), pp. 471–481.

[146] Rob Heylen, Mario Parente, and Paul Gader. "A review of nonlinear hyperspectral unmixing methods". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.6 (2014), pp. 1844–1868.

[147] Dalton Lunga et al. "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning". In: *IEEE Signal Processing Magazine* 31.1 (2014), pp. 55–66.

[148] Giles M Foody et al. "Non-linear mixture modelling without end-members using an artificial neural network". In: *International Journal of Remote Sensing* 18.4 (1997), pp. 937–953.

[149] Miguel A Veganzones et al. "Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation". In: *IEEE Transactions on Image Processing* 23.8 (2014), pp. 3574–3589.

[150] Antonio Plaza et al. "Foreword to the special issue on spectral unmixing of remotely sensed data". In: *IEEE transactions on geoscience and remote sensing* 49.11 (2011), pp. 4103–4110.

[151] Jin Chen et al. "Generalization of subpixel analysis for hyperspectral data with flexibility in spectral similarity measures". In: *IEEE transactions on geoscience and remote sensing* 47.7 (2009), pp. 2165–2171.

[152] SG Herzog and JF Mustard. "Reflectance spectra of five-component mineral mixtures: implications for mixture modeling". In: *Lunar and Planetary Science Conference*. Vol. 27. 1996.

[153] Ben Somers et al. "Endmember variability in spectral mixture analysis: A review". In: *Remote Sensing of Environment* 115.7 (2011), pp. 1603–1616.

[154] Derek M Rogge et al. "Iterative spectral unmixing for optimizing per-pixel endmember sets". In: *IEEE Transactions on Geoscience and Remote Sensing* 44.12 (2006), pp. 3725–3736.

[155] Dar A Roberts et al. "Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models". In: *Remote sensing of environment* 65.3 (1998), pp. 267–279.

[156] Andrew J Elmore et al. "Quantifying vegetation change in semiarid environments: precision and accuracy of spectral mixture analysis and the normalized difference vegetation index". In: *Remote sensing of environment* 73.1 (2000), pp. 87–102.

[157] Robert S Rand, Ronald G Resmini, and David W Allen. "Characterizing intimate mixtures of materials in hyperspectral imagery with albedo-based and kernel-based approaches". In: *Imaging Spectrometry XX*. Vol. 9611. International Society for Optics and Photonics. 2015, p. 961112.

[158] Ben Somers et al. "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards". In: *Remote Sensing of Environment* 113.6 (2009), pp. 1183–1193.

[159] JJ Settle and NA Drake. "Linear mixing and the estimation of ground cover proportions". In: *International Journal of Remote Sensing* 14.6 (1993), pp. 1159–1177.

[160] F Poulet et al. "Comparison between the Shkuratov and Hapke scattering theories for solid planetary surfaces: Application to the surface composition of two Centaurs". In: *Icarus* 160.2 (2002), pp. 313–324.

[161] José M Bioucas-Dias et al. "Hyperspectral remote sensing data analysis and future challenges". In: *IEEE Geoscience and remote sensing magazine* 1.2 (2013), pp. 6–36.

[162] Christoph C Borel and Siegfried AW Gerstl. "Nonlinear spectral mixing models for vegetative and soil surfaces". In: *Remote sensing of environment* 47.3 (1994), pp. 403–416.

[163] Antonio Plaza and Javier Plaza. "Parallel implementation of linear and nonlinear spectral unmixing of remotely sensed hyperspectral images". In: *High-Performance Computing in Remote Sensing*. Vol. 8183. International Society for Optics and Photonics. 2011, p. 81830D.

[164] Gregory P Asner and David B Lobell. "A biogeophysical approach for automated SWIR unmixing of soils and vegetation". In: *Remote sensing of environment* 74.1 (2000), pp. 99–112.

[165] Ann Bateson and Brian Curtiss. "A method for manual endmember selection and spectral unmixing". In: *Remote Sensing of Environment* 55.3 (1996), pp. 229–243.

[166] Ben Somers et al. "Spectral mixture analysis to monitor defoliation in mixed-aged Eucalyptus globulus Labill plantations in southern Australia using Landsat 5-TM and EO-1 Hyperion data". In: *International Journal of Applied Earth Observation and Geoinformation* 12.4 (2010), pp. 270–277.

[167] Asmau M Ahmed et al. "Hybrid spectral unmixing: using artificial neural networks for linear/non-linear switching". In: *Remote Sensing* 9.8 (2017), p. 775.

[168] Majid Jamshidian et al. "Estimation of minimum horizontal stress, geomechanical modeling and hybrid neural network based on conventional well logging data–a case study". In: *Geosystem Engineering* 20.2 (2017), pp. 88–103.

[169] Said Jadid Abdulkadir, Siti Mariyam Shamsuddin, and Roselina Salle-huddin. "Three term back propagation network for moisture prediction". In: *International Conference on Clean and Green Energy*. 2012, pp. 103–7.

[170] Chien-Cheng Yu and Bin-Da Liu. "A backpropagation algorithm with adaptive learning rate and momentum coefficient". In: *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*. Vol. 2. IEEE. 2002, pp. 1218–1223.

[171] Prashant Borkar, MV Sarode, and LG Malik. "Employing speeded scaled conjugate gradient algorithm for multiple contiguous feature vector frames: an approach for traffic density state estimation". In: *Procedia Computer Science* 78 (2016), pp. 740–747.

[172] Martin Fodslette Møller. "A scaled conjugate gradient algorithm for fast supervised learning". In: *Neural networks* 6.4 (1993), pp. 525–533.

[173] Steve Renals et al. *Connectionist speech recognition: Status and prospects*. Tech. rep. Citeseer, 1991.

[174] Léon Bottou. "Large-scale machine learning with stochastic gradient descent". In: *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[175] John Duchi, Elad Hazan, and Yoram Singer. "Adaptive subgradient methods for online learning and stochastic optimization". In: *Journal of Machine Learning Research* 12.Jul (2011), pp. 2121–2159.

[176] Martin Riedmiller and Heinrich Braun. "A direct adaptive method for faster backpropagation learning: The RPROP algorithm". In: *Neural Networks, 1993., IEEE International Conference on*. IEEE. 1993, pp. 586–591.

[177] Yahya H Zweiri, Lakmal D Seneviratne, and Kaspar Althoefer. "Stability analysis of a three-term backpropagation algorithm". In: *Neural Networks* 18.10 (2005), pp. 1341–1347.

[178] Yahya H Zweiri, James F Whidborne, and Lakmal D Seneviratne. "A three-term backpropagation algorithm". In: *Neurocomputing* 50 (2003), pp. 305–318.

[179] Orestis Tsinalis. "Deep Learning for Automated Sleep Monitoring". PhD thesis. Imperial College London, 2016.

[180] Ron Kohavi et al. "A study of cross-validation and bootstrap for accuracy estimation and model selection". In: *Ijcai*. Vol. 14. 2. Montreal, Canada. 1995, pp. 1137–1145.

[181] Brian D Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.

[182] Burkni Palsson et al. "Hyperspectral Unmixing Using a Neural Network Autoencoder". In: *IEEE Access* 6 (2018), pp. 25646–25656.

[183] Elias Ayrey and Daniel J Hayes. "The Use of Three-Dimensional Convolutional Neural Networks to Interpret LiDAR for Forest Inventory". In: *Remote Sensing* 10.4 (2018), p. 649.

[184] Yoshua Bengio, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives". In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1798–1828.

[185] Sun Kang, Geng Xiurui, Tang Hairong, et al. "A New Target Detection Method Using Nonlinear PCA for Hyperspectral Imagery [J]". In: *Bulletin of Surveying and Mapping* 1 (2015), pp. 105–108.

[186] Asmau Ahmed et al. "Application of hybrid switch method to quantify oil spills". In: *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE. June 2018.

[187] Rosa Elvira Correa Pabón and Carlos Roberto de Souza Filho. "Spectroscopic characterization of red latosols contaminated by petroleum-hydrocarbon and empirical model to estimate pollutant content and type". In: *Remote Sensing of Environment* 175 (2016), pp. 323–336.

[188] Jimmy Ba and Brendan Frey. "Adaptive dropout for training deep neural networks". In: *Advances in Neural Information Processing Systems*. 2013, pp. 3084–3092.

[189] Zhe Li, Boqing Gong, and Tianbao Yang. "Improved dropout for shallow and deep learning". In: *Advances in Neural Information Processing Systems*. 2016, pp. 2523–2531.

[190] Feiyun Zhu et al. "Spectral unmixing via data-guided sparsity". In: *IEEE Transactions on Image Processing* 23.12 (2014), pp. 5412–5427.

[191] *Soil Survey Division Staff. Soil survey manual*. 18. Government Printing Office, Washington, D.C., 1993.

[192] Magali Mathieu et al. "Alteration mapping on drill cores using a HySpex SWIR-320m hyperspectral camera: Application to the exploration of an unconformity-related uranium deposit (Saskatchewan, Canada)". In: *Journal of Geochemical Exploration* 172 (2017), pp. 71–88.

[193] Sebastian Thrun and Peter Norvig. "Introduction to artificial intelligence". In: *Retrieved March* 19 (2012), p. 2012.

[194] Sylvain Arlot, Alain Celisse, et al. "A survey of cross-validation procedures for model selection". In: *Statistics surveys* 4 (2010), pp. 40–79.

[195] Ran Pelta and Eyal Ben-Dor. "An Exploratory Study on the Effect of Petroleum Hydrocarbon on Soils Using Hyperspectral Longwave Infrared Imagery". In: *Remote Sensing* 11.5 (2019), p. 569.

[196] Olga Duran and Maria Petrou. "A time-efficient method for anomaly detection in hyperspectral images". In: *Geoscience and Remote Sensing, IEEE Transactions on* 45.12 (2007), pp. 3894–3904.

[197] Carl Stepnowsky et al. "Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters". In: *Sleep medicine* 14.11 (2013), pp. 1199–1207.

[198] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.