

EFFICIENT CLASSIFICATION AND SOIL QUANTIFICATION USING HYPERSPECTRAL DATA

**THESIS SUBMITTED
FOR THE AWARD OF DEGREE OF
DOCTOR OF PHILOSOPHY**

BY

SIMRANJIT SINGH

(REGISTRATION NO. : 901503037)

**UNDER THE GUIDANCE OF
DR. SINGARA SINGH KASANA**

ASSOCIATE PROFESSOR



**Computer Science and Engineering Department
Thapar Institute of Engineering and Technology, Patiala**

INDIA-147004

September 2020

Certificate

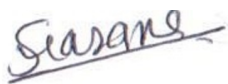
I hereby declare that the work being presented in this thesis "Efficient Classification and Soil Quantification using Hyperspectral Data", in fulfilment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** submitted in Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my work carried out under the supervision of Dr. Singara Singh Kasana, Associate Professor, Computer Science and Engineering Department and refers other researcher works which are duly listed in the reference section. The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.



(Simranjit Singh)

Registration No. 901503037

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge and belief.



(Dr. Singara Singh Kasana)

Associate Professor

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

Supervisor

Acknowledgements

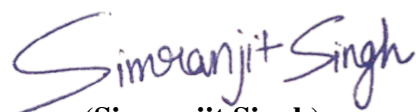
I owe the grace of the Almighty, whose spiritual illumination brings me much courage, encouragement, motivation, confidence, and power to carry out this job.

I want to convey my sincere gratitude to my supervisor, **Dr. Singara Singh Kasana** Associate Professor, Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala for his advice and analysis, providing me with insightful feedback and help in this research work. For having time to listen and offer his insightful guidance, I am deeply indebted to him.

I would like to thank the Doctoral Committee which includes **Dr. Maninder Singh, Dr. Rajiv Kumar, Dr. Shreelekha Pandey**, and **Dr. Sunil Singla** to track the development and have useful recommendations for this work to be strengthened. I would like to thank all the professors and staff members of the Computer Science and Engineering Department and my colleagues **Dr. Jashanpreet Singh, Er. Sahil Sharma** and **Er. Jatinder Singh** for their partnership and encouragement.

Words are insufficient with regard to my father **Shri Balwinder Singh** and my mother **Smt. Sukhjnder Kaur** for their constant support and patience. I convey my sincere appreciation to my brother **Ajit Singh** whose affection and caring have been an indispensable source of motivation to me.

I would like to convey my appreciation to all those who have made their commitment, directly or indirectly, on this voyage of research work to render my Ph.D. a successful event.



(Simranjit Singh)

Abstract

Hyperspectral data accommodates a large amount of information generally in the range of 400 - 2500 nm in the electromagnetic regions. It contains the reflectance values that are captured from the measuring device. This information is high dimensional in nature, which can be utilised in several real term applications.

Classification of the hyperspectral image is a prevalent domain. Most of the existing classification techniques are not able to excerpt the deep features adjacent to the hyperspectral image as it comprises of a significant number of bands. So, we proposed a deep learning-based approach to extract the deep features efficiently. This approach is a hybrid of Locality Preserving Projection, Stacked Autoencoders, and Logistic Regression. Locality Preserving Projection is a dimensionality reduction procedure that is employed to decrease the hyperspectral input. The reduced input contains the preserved local information of the input hyperspectral image. Afterward, the reduced hyperspectral image is passed to the Stacked Autoencoders for taking out the essential approximations of the input that is regarded as deep features which are then passed to the Logistic Regression for efficient construction of the prediction model. Standard hyperspectral image dataset is utilised for validating the developed model results, and it is compared with 21 machine learning models to prove the efficacy of the proposed model.

Most of the existing frameworks for classification are based on spectral-spatial information. The hyperspectral image is termed as high dimensional image because it comprises of many values corresponding to the bands and adding more spatial information to already a large dataset leads to more complex dimension set. So, we proposed a pre-processing framework by using spectral values set of the hyperspectral image. The spectral features are reduced with the two most popular dimensionality reduction algorithms named- Locality Preserving Projection and Principal Component Analysis. The feature set obtained from these algorithms are integrated to design a more efficient feature set called 'hybrid features' which contains the local and global characteristics corresponding to the original hyperspectral image. The hybrid feature set has fewer dimensions than the original dataset due to which it takes less time for supervised training of the learning models. In addition to this, linear interpolation is incorporated in the framework to approximate the bands which are lost during the atmospheric correction process of the hyperspectral image which supports in procuring the better results in classification rate as well as training time in comparison with the spectral-spatial and spectral

frameworks.

The progress achieved in deep learning changed the conventional methods of analysing the information in every domain. It is rarely used in the field of soil science. A Principal Component Analysis and Long Short-Term Memory Networks based framework is proposed to perform the effective quantity estimation of soil characteristics from the spectral library. Firstly, the dimensions of the hyperspectral dataset are reduced with the help of Principal Component Analysis after pre-processing of the dataset. Then the output of Principal Component Analysis is passed to Long Short-Term Memory Networks for efficient learning. Long Short-Term Memory Networks perform better on sequential data as it can remember long term and short-term dependencies in the input. In the end, it is compared with other commonly employed models to show the effectiveness of the proposed model.

We further improved the performance of the Principal Component Analysis and Long Short-Term Memory Networks framework by using the hybrid framework. Previously, we are only using the Principal Component Analysis features which denote the global approximation of the original dataset. So, local features obtained from the Locality Preserving Projection are combined with the Principal Components to form the hybrid features which are passed to Long Short-Term Memory Networks. It generated excellent results in comparison to the Principal Component Analysis and Long Short-Term Memory Networks framework due to efficient learning *via* hybrid features.

Abbreviations

<i>AE</i>	Autoencoders
<i>CEC</i>	Cation Exchange Capacity
<i>DL</i>	Deep Learning
<i>DN-SAE</i>	Deep Network Stacked Autoencoders
<i>DT</i>	Decision Trees
<i>ED</i>	Euclidean Distance
<i>HF</i>	Hybrid Framework
<i>HSI</i>	Hyperspectral Image
<i>JMD</i>	Jeffries-Matusita Distance
<i>KSC</i>	Kennedy Space Centre
<i>LPP</i>	Locality Preserving Projections
<i>LSTM</i>	Long Short-Term Memory Network
<i>LUCAS</i>	Land Use and Coverage Area frame Survey
<i>ML</i>	Machine Learning
<i>PLSR</i>	Partial Least Square Regression
<i>MLE</i>	Maximum Likelihood Estimation
<i>PCA</i>	Principal Component Analysis
<i>PF</i>	Proposed Framework
<i>RMSE</i>	Root Mean Square Error
<i>SAE</i>	Stacked Autoencoders
<i>SAM</i>	Spectral Angle Mapping
<i>SCA</i>	Spectral Correlation Angle
<i>SVM</i>	Support Vector Machines
<i>SVR</i>	Support Vector Regression

Notations and Symbols

H_i^k	HSI with total i pixels, each of k dimensions.
T_b	Total number of bands
p_i	Pixel Spectrum
r_i	Reference Spectrum
r	Pearson Coefficient
$P(T)$	True Positive
$E(S)$	Error Sum of Squares
y_k	Actual values
$\hat{y}_k$	Predicted values
ρ_C	Lin's Concordance Coefficient
$N(T)$	True Negative
ρ	Spearman Correlation Coefficient
$T(S)$	Total Sum of Squares
μ	Mean
$P(C)$	Total Positive Cases
σ	Variance
λ	Eigenvalue
V	Eigenvector
$N(C)$	Total Negative Cases
R^2	Coefficient of Determination

Contents

Certificate	i
Acknowledgements	ii
Abstract	iii
Abbreviations	v
Notations and Symbols	vi
List of Figures	x
List of Tables	xii
List of Publications	xiii
1 Introduction	1
1.1 Spectral Images and Data	1
1.2 Classification of Hyperspectral Images	3
1.3 Quantitative Estimation of Soil	6
1.4 Problem Statement	7
1.5 Motivation of the Proposed Work	8
1.6 Thesis Contribution	8
1.7 Novelty of the Research Work	9
1.8 Thesis Organisation	9
2 Literature Survey	11
2.1 Background of the Research Work	11
2.1.1 Classification Techniques	11
2.1.2 Quantification Techniques	14
2.2 Literature Review on Hyperspectral Classification	16
2.3 Literature Review on Soil Properties Quantification	24

2.4	Research Gaps	36
2.5	Objectives	36
2.6	Methodology	36
3	Efficient Classification of the Hyperspectral Images using Deep Learning	39
3.1	Background of the Proposed Work	39
3.1.1	AutoEncoders	39
3.1.2	Stacked Autoencoders	41
3.1.3	Local Preserving Projections	42
3.2	Proposed Framework	43
3.3	Experimental Results and Analysis	44
3.3.1	Description of the Datasets	44
3.3.2	Experimental Results	46
3.4	Conclusion of the Chapter	51
4	A Pre-processing Framework for Spectral Classification of Hyperspectral Images	53
4.1	Background of the proposed work	53
4.1.1	Interpolation	53
4.1.2	Principal Component Analysis and Locality Preserving Projections	54
4.2	Proposed Preprocessing Framework	57
4.3	Experimental Results and Analysis	58
4.3.1	Description of Datasets	59
4.3.2	Experimental Results	60
4.4	Conclusion of the Chapter	72
5	Quantification of Soil Properties from the Hyperspectral Data using Long Short-Term Memory Networks	73
5.1	Background of the Proposed Work	73
5.1.1	Recurrent Neural Networks	73
5.1.2	Long Short-Term Memory Networks	74
5.2	Materials and Methods	77
5.2.1	LUCAS Dataset	77
5.2.2	Field Sampling	78
5.2.3	Soil Testing	79
5.2.4	Spectral Acquisition and Pre-treatments	79
5.3	Proposed Framework	80
5.4	Experimental Results and Analysis	81
5.5	Conclusion of the Chapter	87

6 Quantification of Soil Properties Using Hybrid features and Deep Learning 89

6.1 Background of the proposed work 89

6.1.1 Principal Component Analysis and Locality Preserving Projections 89

6.1.2 Long Short-Term Memory Networks 90

6.2 Proposed Framework 90

6.3 Results and Analysis 91

6.4 Conclusion of the Chapter 97

7 Conclusions and Future Scope 99

7.1 Conclusions 99

7.2 Future Scope 100

Bibliography 101

List of Figures

1.1	Spectral Reflectance Curve of Vegetation	3
1.2	Electromagnetic Spectrum	4
1.3	Hyperspectral Image	4
2.1	Structure of Decision Tree	13
2.2	Process of Ensembling	14
2.3	Methodology of Proposed Framework	37
3.1	Single Layer Autoencoder	40
3.2	4-layered Stacked AE including input and Softmax classifier layer	41
3.3	Discriminating power of PCA (a) and LPP (b)	42
3.4	Principal Component Analysis	43
3.5	Proposed framework for classification	44
3.6	Indian pines (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13	45
3.7	Salinas dataset (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13	45
3.8	Salinas classification map a) Original HSI b) Proposed framework	47
3.9	Indian Pines classification map a) Original HSI b) Proposed framework	47
4.1	Interpolation	54
4.2	Procedure of extracting Principal Components	56
4.3	Proposed Preprocessing Framework for Classification	57
4.4	KSC dataset (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13	59
4.5	Pavia University dataset (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13	60
4.6	Effect of changing the hybrid features size on overall accuracy and training time (A) OA (Indian Pines) (B) TT (Indian Pines) (C) OA (Salinas) (B) TT (Salinas)	66
4.7	Effect of changing the hybrid features set size on overall accuracy and training time (A) OA (KSC) (B) TT (KSC) (C) OA (Pavia University) (B) TT (Pavia University)	68
4.8	Effect of changing the feature set size on overall accuracy (A) Indian Pines (B) Salinas (C) KSC (D) Pavia University	69

4.9	Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)	70
4.10	KSC Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)	71
4.11	Salinas Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)	71
4.12	Pavia University Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)	72
5.1	Recurrent Neural Network Unrolled	74
5.2	Independent operations performed on (a) Input gate (b) Forget gate (c) New cell state (d) Output gate (e) Hidden layer	75
5.3	LSTM with two layers	76
5.4	Sampling points of European Union	77
5.5	Proposed PCA-LSTM Framework	80
5.6	Mean absolute Error (loss function) in training and validation (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand	83
5.7	Residuals (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand	85
5.8	Predicted vs Ground Truth graphs of PPLF (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand	87
6.1	Proposed Hybrid Framework	91
6.2	Mean absolute Error (loss function) in training and validation (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand	92
6.3	Predicted vs Ground Truth graphs calculated from HF (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand	95
6.4	Residuals (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand	96

List of Tables

3.1	Class wise prediction and Overall accuracies of 22 learning Models (Indian Pines) . .	49
3.2	Class wise prediction and Overall accuracies of 22 learning Models (Salinas)	50
4.1	Correlation between LPP and PCA	56
4.2	Description of train and test set samples for HSIs	61
4.3	Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (Indian Pines)	62
4.4	Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (Salinas)	62
4.5	Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (KSC)	63
4.6	Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (Pavia University)	63
4.7	Class wise prediction results after applying the Proposed Framework (Indian Pines) .	64
4.8	Class wise prediction results after applying the Proposed Framework (Salinas)	64
4.9	Class wise prediction results after applying the Proposed Framework (KSC)	65
4.10	Class wise prediction results after applying the Proposed Framework (Pavia)	65
4.11	Comparison of PF with previously proposed DNN	70
5.1	% of samples acquired by countries and land uses	78
5.2	Statistics of different soil properties from samples ($n=19036$)	79
5.3	Parameters/ Hyperparameter set in LSTM	82
5.4	PPLF comparison with other models based on six evaluation metrics.	84
5.5	Earlier studies comparison with HF on Test set and R^2	86
6.1	Parameters/ Hyperparameter set in LSTM	93
6.2	Hybrid Framework comparison with other models based on six evaluation metrics. .	94
6.3	Earlier studies comparison with HF on Test set and R^2	97

List of Publications

- I. **Simranjit Singh** and Singara Singh Kasana (2018). Efficient classification of the hyperspectral images using deep learning. *Multimedia Tools and Applications*, 77(20), pp. 27061-27074, IF=2.313. **SCI Indexed**
- II. **Simranjit Singh** and Singara Singh Kasana (2019). Estimation of soil properties from the EU spectral library using long short-term memory networks. *Geoderma Regional*, e00233, pp. 1-12, IF=2.667. **SCI Indexed**
- III. **Simranjit Singh** and Singara Singh Kasana (2019). A Pre-processing Framework for Spectral Classification of Hyperspectral images, *Multimedia Tools and Applications*, DOI: <https://doi.org/10.1007/s11042-020-09180-2>, IF=2.313. **SCI Indexed**
- IV. **Simranjit Singh** and Singara Singh Kasana (2019). Spectral-Spatial Hyperspectral Image Classification using Deep Learning. *IEEE International Conference on Artificial Intelligence*, pp. 411-417, Dubai.
- V. **Simranjit Singh** and Singara Singh Kasana (2019). Quantitative Assessment of soil properties from hyperspectral data for optimising sustainable agriculture using Deep learning, *16th International Conference on Environmental Science and Technology*, September 4-7, Greece.
- VI. **Simranjit Singh** and Singara Singh Kasana (2019). Estimation of Soil Organic Carbon for sustainable agriculture using Deep learning, *16th International Conference on Environmental Science and Technology*, September 4-7, Greece.
- VII. **Simranjit Singh** and Singara Singh Kasana. Classification of Hyperspectral Image Classification using LSTM, *40th Asian Conference on Remote Sensing*, October 14-18. South Korea.

Chapter 1

Introduction

Soil is the most critical resource that occurs naturally and provides numerous forms of metals, micronutrients that are important for plant growth. It has various advantages as it serves as a carbon sink to minimise global warming and also functions as a buffer for radioactive metals and other toxins. It is gene reserve and a natural habitat for many organisms that provide the foundation for many construction activities. It also has cultural significance. It is in the environment since the creation of the earth. It also contains valuable information on how humans have migrated in the past.

Soil's physical, chemical, structural, and biological properties change both spatially and temporally (Blum 1993; Jenny 2012) in the same field, which inhibits the capacity to carry out each of the above said tasks. Due to the diverse nature of the soil, it is required to have the information on its properties to keep it healthy. This information can also help us to understand which part of the field needs more attention. There exist various methods to capture this information like spectroscopic analysis (Morra et al. 1991), Nuclear magnetic resonance (Harris 1986), etc. Out of these, the spectroscopic analysis is the most preferred method as it captures the radiation intensity as a wavelength function in the electromagnetic spectrum. There are various devices used to measure this information like spectrometers, spectrographs, spectrophotometers, etc. Nowadays, these spectrometers are in-built in the satellites to capture quick information regarding the fields.

There is a significant development in the field of remote sensing during the past decade. The high-resolution sensors can capture images in 400 to 2500 nm range of the electromagnetic spectrum *via* onboard spectrometers. The images which contain 100 or more bands are called hyperspectral images. These images contain reflectance spectra values which are employed to classify the pixels into a certain category based on their spectral signature.

1.1 Spectral Images and Data

In remote sensing, spectral images captured by satellite sensor are of multispectral or hyperspectral nature. In acquiring high-resolution data, many improvements are being made from sensors such as

— Bhuvan, Hymap, Hyperion, Hot Spot Recognition Sensor System (HSRS), *etc.* Rourke (O'Rourke and Holden 2011) reported that with the help of onboard satellite spectrometers, valuable information could be obtained rapidly from the collected data, which reduces time. These satellites collect details from 400 – 2500 nm, and includes sound detection (Rossel et al. 2006), through the entire electromagnetic spectrum.

In the 1960s the multispectral sensors were designed to deliver more reliable 8-10 bands to the high-resolution image that collect information from the electromagnetic range. It was able to differentiate between surface soils effectively (Kristof 1971). But the image set captured from these sensors, is not available for research purposes. Following the launch of LANDSAT-1, captured information was made available to the research community. Many scientists use this information for designing different applications such as material identification, quantification of soil, vegetation mapping *etc.*

Hyperspectral imaging is a growing area in the domain of remote sensing. Hyperspectral Images (HSI) captured by satellite sensors contain 100 or more bands and have an extremely fine resolution. There is a limited number of sensors which can capture HSIs, namely Bhuvan, HyMap, Earth Observing-1 Hyperion, Hymap *etc.* The captured images from these satellites contain reflectance values in the range of 400 to 2500 nm captured by the onboard spectrometers. The output of these sensors is an image which consists of n dimensions, as shown in Figure 1.3.

Hyperspectral data contains high dimensional spectral values. This type of data is typically captured in the laboratory with the help of spectrometers that captures the reflectance values 400-2500nm. It generally contains instrumental artifacts in the range of 400 to 500nm. The captured reflectance values present in this region is removed before processing the data. There are several machines present to capture this information like Analytical Spectral Devices (ASD), FOSS spectrometer *etc.*

The sun's radiation can be reflected, refracted or absorbed depending upon the kind of surface. The object's reflective properties depend upon its physical and chemical specifications, its surface hardness and its geometric conditions (*e.g.*-sun's incident angle). The values that are captured represent the reflectance captured by sensor that often varies with the electromagnetic strength wavelengths. Every material has a unique spectral signature, which can be recognised by the analysis of the captured reflectance values from the satellites. Spectral reflectance of vegetation is shown in Figure 1.1.

For several years, reflectance spectra have been utilised to measure soil properties. In the near-infrared area, Condit (Condit 1970, 1972) attempted to describe the soil's characteristics by using the reflectance spectrum. It is also found that the quantity of the nutrient present in the soil depends upon the captured reflectance value (Baumgardner et al. 1986). Afterwards, With the advent of advanced equipment in laboratories, hyperspectral data became available. The measured reflectance is partitioned into mainly three regions Visible (VIS), Infrared (IR) and Ultraviolet (UV) as shown in Figure 1.2 :

- I. Visible region lies in the range of 400-700 nm with a frequency of $790 - 430 THz$. This region contains all the primary colors that can be seen by the human eye.

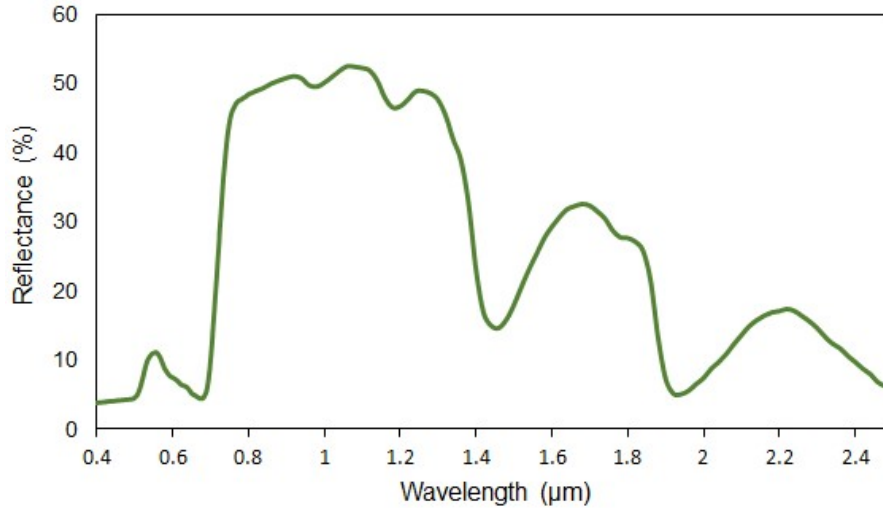


Figure 1.1: Spectral Reflectance Curve of Vegetation

- II. Infrared region lies in the range of 700 nm – 1 mm having a frequency in the range of 430 THz – 300 GHz . This region contains the heat effects. It is most commonly divided into three parts:
- (a) Near-Infrared- It lies in the range of 0.75 - 1.4 μm with a frequency of 214-400 THz . It can be used to see cooler red stars.
 - (b) Mid-Infrared- It lies in the range of 1 - 4 μm with a frequency of 100-214 THz . It is used to see protoplanetary disks, dust warmed by starlight.
 - (c) Far-Infrared- It lies in the range of 15-1000 μm with a frequency of 0.3 - 20 THz . In this region, central regions of galaxies and cold dust emissions can be seen.
- III. Ultraviolet region lies in the 10 – 400 nm having a frequency in the range of 30 PHz – 790 THz . These regions rays are most harmful to the living organisms and are mostly stopped by the ozone layer.

1.2 Classification of Hyperspectral Images

HSI classification is a popular domain among the remote sensing community. Many efforts are focussed on this issue. Classification of a particular pixel in a particular category based on spectra provides a better understanding of spectral data captured by satellite sensors. HSI contains 100 or more bands and has continuous spectral information available. The captured HSIs have spatial and spectral features. Due to high spectral bands, the applications of these images are extensive. The acquisition of high-resolution images has been made feasible with the development of spectral sensors that can capture the images having a significant number of bands, due to which it is employed in many domains such as border surveillance, quantification of soil, material detection, food processing

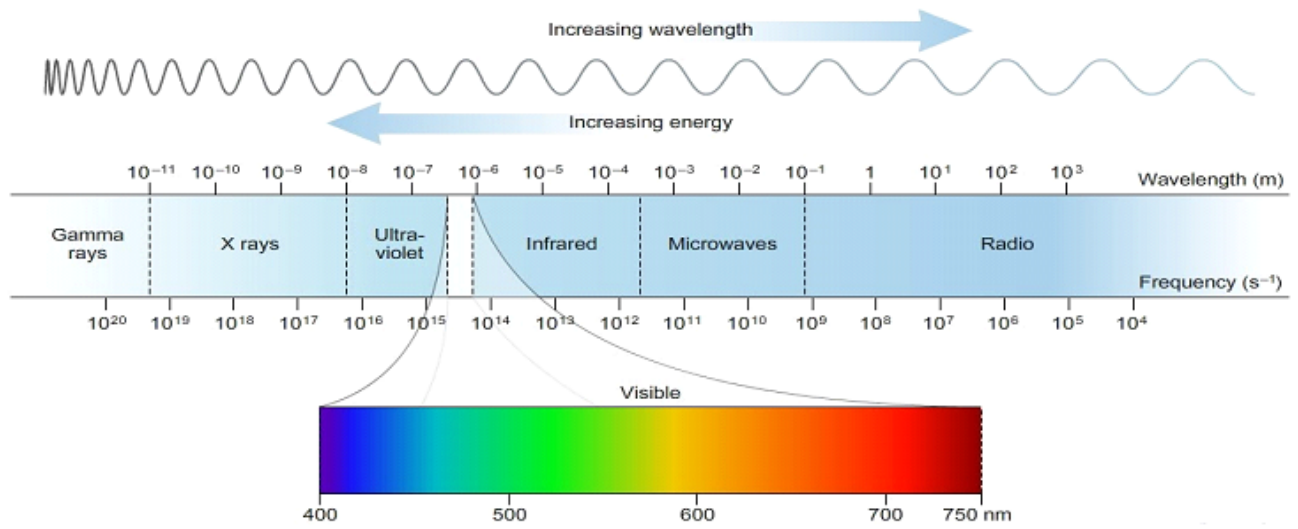


Figure 1.2: Electromagnetic Spectrum

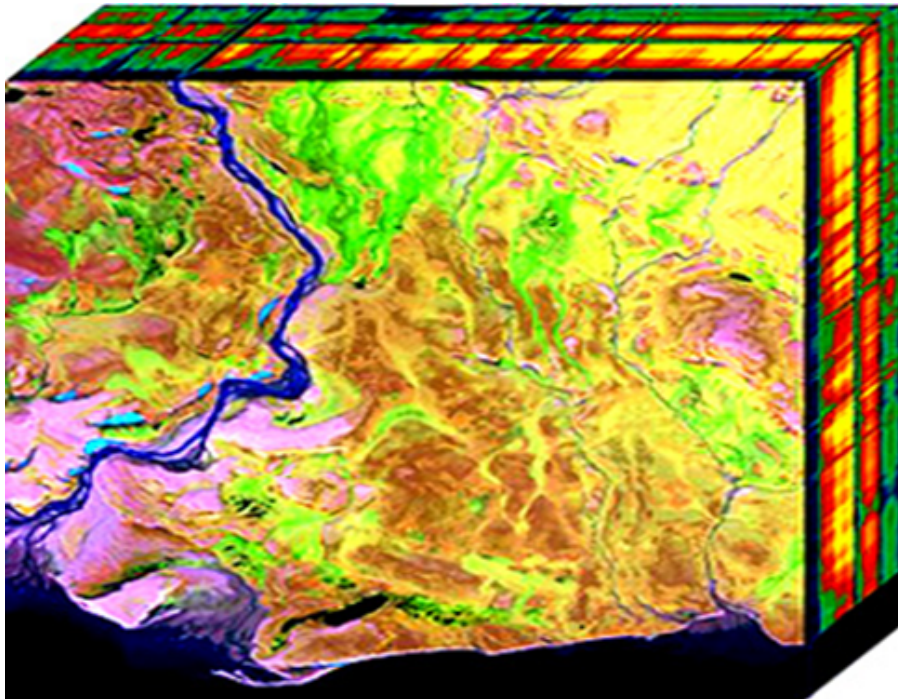


Figure 1.3: Hyperspectral Image

etc. As the HSI contains the high dimensional spectral signature, it becomes very complicated to efficiently analyse the hyperspectral data.

If the data of the HSIs are carefully handled, it can generate higher classification accuracies. There are several different classification algorithms for this and the most popular among these is the spectral angle mapping method. Others include the maximum likelihood, euclidean distance, spectral information divergence and correlation angle, *etc.* All of these techniques mentioned are based on a distance measure. When the distance between two pixels is less than the threshold, they belong to the same class. These measures can also be employed for spectral matching by taking the pixels from images and matching them with the present spectral libraries for identifying certain materials. Machine learning algorithms such as SVM, naive Bayes *etc* are more efficient in classifying the hyperspectral pixels into certain classes. SVM is considered to be the most robust machine learning model (Borges 1998). In most cases, it is able to outperform the traditional artificial neural networks (Zanaty 2012). But SVM is not able to perform well on high dimensional input due to the Hughes phenomenon. Others works also applied clustering techniques (Pompilio et al. 2014) similar to network-based clusters (Chand et al. 2006).

In HSI classification, spectral-spatial techniques generate high classification accuracy. The improvements in spectral-spatial architecture are discussed by Fauvel et al.(Fauvel et al. 2013). Their work also presented another method, that is hybrid of morphology operators and SVM. Similarly, the use of a composite kernel to enhance the classification accuracy of hyperspectral images is performed by Camps (Camps-Valls et al. 2014). The designed framework exploited mercer's kernels properties for developing the family of composite kernels. It also incorporated the spectral-spatial information, and it also has greater computation efficiency. Spectral-spatial features are able to generate a high classification rate as compared to spectral features. These features, when passed to machine learning classifiers like the random forest, naive Bayes, SVM *etc* is able to improve the classification performance of the learners. Also, linear discriminant analysis is used widely in combination with other learning models to form a hybrid model (Kumar and Rani 2013; Rani et al. 2013). After supervised learning, the trained model is provided with testing data for evaluating the performance of the prediction. But due to HSI's high dimensions, most of the learning models are not able to perform well. To solve this, many works stated that utilising deep learning frameworks for the classification is beneficial as these frameworks can handle high dimensional input. Stacked Autoencoder, a deep learning architecture is employed by Chen (Chen et al. 2015) to extract the deep features, which when are passed to classifier generates high overall accuracy. Similarly, another deep learning technique named Long-Short Term Memory Network (LSTM) is also employed by Zhou (Zhou et al. 2019), for classifying the spectral-spatial features, which produced effective results.

Most of the present classifiers utilise the spectral-spatial features of the HSIs (Wang et al. 2017b). The spectral features can be extracted from the HSI directly, and the spatial features are extracted from the local neighbor information. These spatial features constitute immense information as each pixel

comprises 50-400 bands. Suppose an average HSI captured image has 115 bands. For extracting a local neighborhood information a 3×3 mask is placed on the image. Now, this mask coincides with 9 neighbors having 115 bands each, which when extracted becomes 1035 bands, which contributes to more information. So, to solve this, different feature extraction (Chen et al. 2015) and selection (Sanasam et al. 2010) algorithms have been employed for selecting important features and maintaining spatial information. Commonly used algorithms include ISOMAP, Principal Component Analysis(PCA), Locally Linear Embedding, Locality Preserving Projection(LPP) *etc.*

1.3 Quantitative Estimation of Soil

Quantification of soil properties using spectral data is a growing field. The spectral region is partitioned into three parts- visible, Infrared, and Ultraviolet. There are many popular quantification techniques like Multiple linear regression, Machine learning models, etc. that can take the visible and near-infrared region from the hyperspectral data to quantify the soil properties. The spectra set is partitioned into training and testing. Afterward, these techniques take the spectra values as input for training and build a prediction model. Then, the prediction model predicts the quantity values from the testing set.

Multiple Linear Regression Analysis (MLR) was employed by Ben-Dor and Banin (1995) to find the association of soil characteristics with particular bands in NIR. The stepwise linear regression (SWR) was tested by Kadupitiya *et al.* (2010) for the construction of an estimation model for soil properties from NIR and VIS regions of the electromagnetic spectrum. Similarly, the trend of employing regression was continued by the Dalal *et al.* (1986) and Jenny *et al.* (2012) utilising Partial Least Square Regression (PLSR) for quantity estimation of soil properties. Then Shepard and Walsh (2002) employed another regression-based method called multivariate adaptive regression splines to precisely predict soil characteristics by analysing the spectral libraries. Afterward, hyperspectral satellite Hyperion is utilised by Harris *et al.* (2016) for detecting as well as estimating the salts exist in the desert plateau located in Chile by employing spectral mixture analysis on the acquired data from the satellites. The widely employed methods for quantity estimation are Principal Component Regression (PCR), MLR, SWR, PLSR. PLSR, in most cases, is regarded as stable, since its orthogonal components can describe the maximum covariance between input and output, and it can accommodate lost and noisy records. The association between spectrum and soil characteristics can easily be observed by employing PLSR as compared to other current approaches. The utilisation of Artificial Neural Network (ANN) is performed by Daniel *et al.* (2003) for estimating many key soil attributes like phosphorus, organic matter and potassium from the reflectance values obtained in NIR region. Afterward, many other works also employed ANN for estimating soil attributes like organic carbon, Zinc, pH, Magnesium, etc. as ANN provides a reliable way for getting a quick preliminary quantity estimate of various soil minerals (Puno et al. 2017; Tiwari et al. 2015).

The use of deep learning is widespread in every domain for analysing the data. But it is hardly employed in soil spectroscopy analysis with the exception of some studies. Deep learning-based Convolutional Neural Networks (CNN) and Multi-CNN architectures are employed by Padarian *et al.* (2019) for the quantitative estimation of soil characteristics from LUCAS high-dimensional dataset. In addition to this, they smoothed the data using spectrogram before giving it to CNN for analysis. Their models are able to produce more accurate results as compared to other regression models. Similarly, CNN was also employed by Padarian *et al.* (2019) again in another study. Instead of the conventional Digital Soil Mapping (DSM) model, CNN was able to generate efficient soil maps. They claimed that Conventional DSM models are based on point have been calibrated with a point observation that fits covariates, and these are not capable of understanding the contextual data. That is why CNN was employed to produce contextual data for building well-defined maps.

Similarly, Liu *et al.* (2018) also investigated transfer learning for the estimating of clay material exist in soil. Their designed model achieved a R^2 of 0.60 and 0.70 on the dataset of Hymap and LUCAS, respectively. In the same way, the LUCAS and AFSIS dataset were used by Veres *et al.* (2015) to approximate soil property with the help of the conditional Boltzmann machine along with CNN. They claimed that analysis based on regression is not capable of finding the associations exist in the samples. They employed different deep learning frameworks for quantity estimation of the various soil properties and achieved good results.

Mou *et al.* (2017) termed the HSI as sequential data for classifying the different pixels of HSI by employing LSTM and Gated Recurrent Unit (GRU). They also described a novel activation function named ‘Pretanh’ that is used by LSTM and GRU for classifying the HSI. The results achieved a high overall classification accuracy, indicating that HSI is sequential in nature. Afterwards, other studies also tested the LSTM (Hang et al. 2019; Ndikumana et al. 2018) and achieved a high classification rate when LSTM is being employed.

1.4 Problem Statement

Classification and soil quantification using hyperspectral data is a popular research domain in hyperspectral analytics. In hyperspectral classification, most current methods are based on similarity measure and traditional machine learning techniques. These techniques are very time consuming and yield less precise performance. Likewise, existing soil quantification models are based on traditional regression techniques. Deep learning advancements have improved traditional ways of processing data in diverse fields. In this study, novel frameworks based on deep learning models like deep autoencoders and Long Short-Term Memory Networks have been proposed. The proposed models are capable of delivering high performance compared to traditional regression and machine learning models.

1.5 Motivation of the Proposed Work

- I. Most of the existing matching techniques used for classification of hyperspectral images use similarity measures which are not very accurate, and most of them cannot have the capability to extract deep features from the high dimensional hyperspectral images.
- II. The acquired HSIs from the onboard satellite sensors cannot be directly utilised as these have atmospheric errors. During the correction of these errors, a lot of bands are lost. Missing HSI bands decrease the classifier's discriminatory strength.
- III. By utilising the HSIs spatial features to form spectral-spatial based frameworks, the total feature set increases, due to which supervised training needs more time.
- IV. The equipment employed for capturing the hyperspectral data contains artifacts in the range of 400–500 nm, due to which the prediction models are not trained well.
- V. Present techniques for quantification do not take benefit of the sequential nature of spectral data.

1.6 Thesis Contribution

- I. To extract the efficient deep features from the hyperspectral images, stacked autoencoder is employed.
- II. For the approximation of the deleted bands of *HSI*, linear interpolation is employed. It approximates these extracted bands as *HSI* is a sequential image (Mou et al. 2017).
- III. The evaluation of all HSI bands is tedious as most of the present bands are redundant. The use of PCA and LPP can remove these redundancies and preserve the overall structure of the input HSI. The resultant output obtained from PCA and LPP are merged together to form hybrid features, which constitute of local and global approximation of the HSI and have lower features compared to the original HSI. By utilising these features, the prediction model is able to generate high classification rates with lower training time.
- IV. Pre-processing is conducted after artifact elimination, and PCA is employed to minimise the excess hyperspectral data dimensions, which enhances the prediction model performance.
- V. LSTMs are utilised to design an efficient prediction model for exploiting the sequential nature of hyperspectral data.

1.7 Novelty of the Research Work

The proposed research work presents an effective hyperspectral classification and soil quantification technique based on deep learning frameworks.

Hyperspectral data is composed of high dimensional features, which constitute redundant information. As a result, most of the existing techniques are not able to classify the data efficiently. We employed popular dimensionality reduction algorithms named Principal Component Analysis and Locality Preserving Projection to reduce the dimensions effectively.

Hyperspectral image data collected by space-borne satellites contain atmospheric defects. As these errors are corrected, a number of useful bands are lost. As hyperspectral data is sequential, we employed linear interpolation to estimate the bands that were missing during the correction process.

The use of Stacked Autoencoders provides deep features of hyperspectral data. These features, when combined with dimensionality reduction algorithms, provide a better classification of the hyperspectral pixel. Pixels typically include a variety of soil surfaces, trees, plants *etc* that can provide sufficient information on the identification of change, deriving forest maps etc.

Hyperspectral data is sequential in nature, and the existing techniques for soil quantification are not leveraging this feature of hyperspectral data. To handle this issue, Long Short-Term Memory Networks, a deep learning model is used to quantify soil properties, which effectively analyse soil spectroscopy data.

1.8 Thesis Organisation

Our thesis is organised as follows:

Chapter 1 presents the general introduction to different types of spectral data. Chapter 2 presents a background of the research work and literature review regarding all the existing works on various classification and quantification techniques. It also shows the motivations, objective, and methodology. Chapter 3 presents an efficient classification technique using deep learning for the classification of hyperspectral images, that is compared with 22 machine learning models to show its effectiveness. Chapter 4 presents a pre-processing framework using PCA and LPP for the classification of hyperspectral images. This chapter also shows the accuracy with and without the application of our proposed framework. Chapter 5 presents a deep learning approach called PPLF for quantification of the soil minerals from the hyperspectral library. Chapter 6 presents an improved hybrid framework called HF for quantification the minerals from the hyperspectral data. Chapter 7 presents the conclusions drawn based on the present study and recommendations for future work.

Chapter 2

Literature Survey

Background of the research work is discussed, and the Present hyperspectral classification and soil quantification techniques have been examined in this chapter. The limitations of these approaches have been described. On the basis of these limitations, the aims and methodology of the work have been described in this chapter, which contributes to the formulation and solution of the problems addressed in the following chapters.

2.1 Background of the Research Work

There are various existing classification and soil quantification techniques, discussed in following subsections.

2.1.1 Classification Techniques

The classification techniques which are based on the matching of one-pixel spectra with another. It can be defined as:

I. Spectral Angle Mapping

Spectral Angle Mapping (*SAM*) is a similarity index measure which is based on cosine angle similarity. It is one of the most frequently used methods in HSIs for classification (Robila 2005). It is defined as:

$$SAM(p, r) = \cos^{-1} \frac{\sum_{i=1}^{T_b} p_i \cdot r_i}{\sqrt{\sum_{i=1}^{T_b} p_i^2} \sqrt{\sum_{i=1}^{T_b} r_i^2}} \quad (2.1)$$

where T_b denotes total bands, r_i denotes reference spectrum and p_i is the pixel spectrum and $\cos^{-1}$ is the inverse of the cosine angle.

II. Euclidean Distance

Euclidean Distance (*ED*) is the most simple similarity measure which is based on the straight line distance between two spectra. For n - dimensions it is given as:

$$ED(p, r) = \sqrt{\sum_{i=1}^{T_b} (p_i - r_i)^2} \quad (2.2)$$

III. Spectral Correlation Angle

Spectral Correlation Angle (*SCA*) is the measure of cosine angle of the pearson coefficients between the pixel and reference spectral. It is defined as:

$$SCA(p, r) = \frac{T_b \sum_1^{T_b} p_i r_j - \sum_1^{T_b} p_i \sum_1^{T_b} r_j}{\sqrt{[T_b \sum_1^{T_b} (p_i)^2 - (\sum_1^{T_b} p_i)^2][T_b \sum_1^{T_b} (r_j)^2 - (\sum_1^{T_b} r_j)^2]}} \quad (2.3)$$

IV. Maximum Likelihood Estimation

Maximum likelihood estimation MLE calculates the values of parameters of a model. The parameters are set in a way that it maximizes the likelihood that the process identified by the model create the observed data. Suppose $H_1, H_2, \dots, H_n$ have the joint density denoted as $f_\theta(h_1, h_2, \dots, h_n)$ with the given observed values $H_1 = h_1, H_2 = h_2, \dots, H_n = h_n$ the likelihood of θ is given by:

$$f_{likli(\theta)} = fr(h_1, h_2, \dots, h_n | \theta) \quad (2.4)$$

where $f_{likli(\theta)}$ is considered as a function of θ and fr is the frequency if the distribution is discrete. Then the MLE of θ is obtained after maximizing the $f_{likli(\theta)}$

V. Jeffries-Matusita Distance

Jeffries-Matusita Distance (*JMD*) is a very popular measure in remote sensing. It is a modification of Bhattacharyya distance because using it can generalise any distributions, for which Bhattacharyya distance can be calculated. It is defined as:

$$JMD(p, r) = \sqrt{2(1 - \exp(-bh(p, r)))} \quad (2.5)$$

where $\exp$ is the exponential function and bh is the Bhattacharyya distance which calculates the dissimilarity between two distribution and can be defined as:

$$b(p, r) = -\ln \int \sqrt{p(x) \cdot r(x)} dx \quad (2.6)$$

where $\ln$ is the natural logarithm.

There are machine learning algorithms that are used for classification. The data is partitioned into

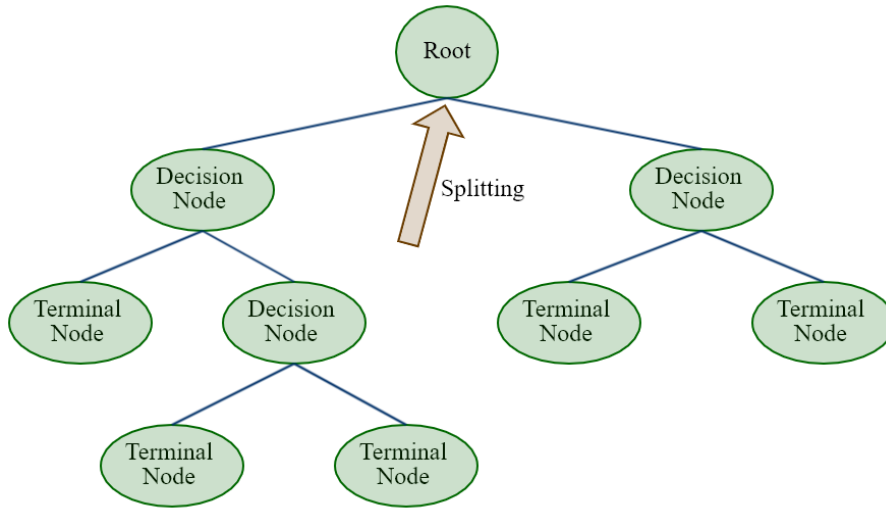


Figure 2.1: Structure of Decision Tree

training and testing. In the training phase, these algorithms perform supervised learning. Afterward, a trained model is built to which testing data is passed for prediction. These methods include:

I. Support Vector Machines

Support Vector Machines (SVM) is the most widely employed procedure in hyperspectral classification (Qi et al. 2016; Zhong et al. 2014) and in many other domains (Anand and Suganthan 2009; Kovačević et al. 2010). It comes under supervised machine learning methods. It separates the hyperplanes with the help of kernel function in case of the classification problem. It produces efficient results classification as well as regression problems. But as the dimension size increases, it performs poorly due to the Hughes phenomenon.

II. Decision Trees

Decision Trees (DT) is a tree-like structure which contains a series of decision with possible outcomes. It is considered to be one of the best and most frequently used learning models. It is also referred to as the Classification and Regression Tree (CART). It can be used to make linear as well as non-linear models. The structure of simple DT is shown in Figure 2.1 where each internal node is the decision node, and the terminal node is the response or resultant class values.

III. K-nearest neighbor

K-nearest neighbor (KNN) is a non-parametric and supervised lazy learning method employed for classification as well as regression problems. It is based on k closest training inputs. In a learning example, the weights of KNN are changed in such a way that the closest neighbors contribute more than the distant ones. Its application is simple, and it takes less processing time. In this, the Euclidean distance is the most preferred distance metric for finding the similarity measure.

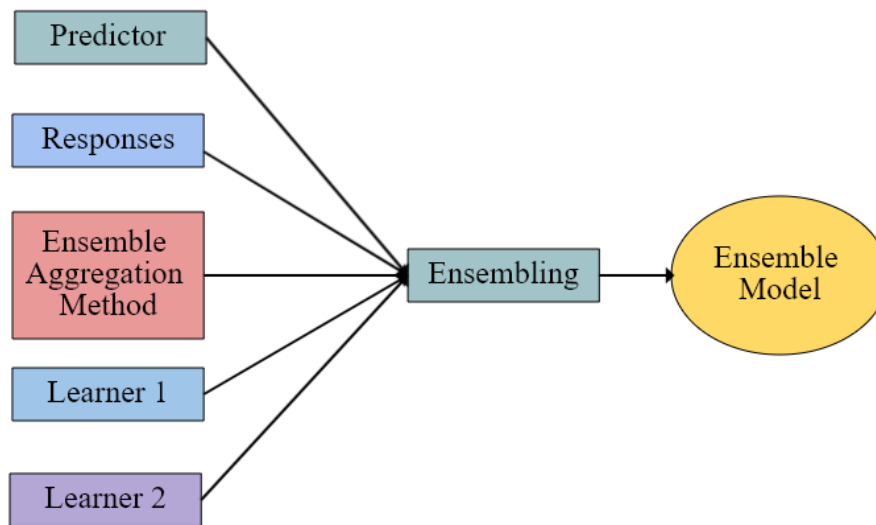


Figure 2.2: Process of Ensembling

IV. Ensemble

Ensemble learning is the process by which weak learners are combined to form a strong one, as shown in Figure 2.2. It improves the performance of a prediction. The learner obtained in the end has greater performance than all its weak learners. There are many types of ensembles out of which commonly used are:

- (a) Bootstrap aggregation- It is proposed by (Breiman 1996) to ameliorate the output of ML algorithms. It avoids the overfitting of the learning models by lowering the variance of the input. It is also known as bagging,
- (b) Boosting- It is proposed by Kearns (1988). It is a supervised machine learning algorithm which combines weak classifiers to improve the performance by lowering the bias and variance of the input.
- (c) Bayes optimal classifier- It is an ensemble of various weak learners. It is based on voting where the input data is independent of conditions, which makes it computationally efficient. Every hypothesis has a vote which is proportional to the likelihood that the train set is sampled if that hypothesis is correct.

2.1.2 Quantification Techniques

There are many quantification techniques which are mainly based on regression. These techniques take the spectra as input and output the quantity of minerals. Commonly used methods are:

I. Multiple Linear Regression

Multiple Linear Regression (MLR) is the commonly employed quantification technique used in

soil sciences. It has one continuous dependent variable and two or more independent variables. The spectral values are considered as independent (predictor) variable, and the corresponding is taken as the dependent (response) variable. MLR evaluates the association between these dependent and independent variables. MLR can be represented as:

$$y_i = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots \beta_j \times x_j + \epsilon \quad (2.7)$$

Where y_i is the dependent variable, β_0 is the y_i -intercept, β_j is the slope coefficient, ϵ is the model's error term.

II. Principal Component Regression

PCR is based on the principal components. Firstly, PCA is applied to extract out the non-redundant and important features with maximum variance. Afterward, extracted principal components are passed to the standard linear regression model for regression. It also overcomes the multicollinearity problem.

III. Stepwise Regression

Stepwise Regression (SWR) selects or removes the values from a pool of candidate predictors to build an efficient prediction model. The choice of predictor candidates is automatic and can be chosen by three approaches:

- (a) **Forward selection**- In this, the model begins with no variable and start adding the predictors in such a way that the addition statistically improves the fit. This process is repeated until a significant fit is obtained.
- (b) **Backward elimination**- It this, the model includes all the predictor variables and start eliminating the variables n such a way that the elimination statistically improves the fit. This process is repeated until a significant fit is obtained.
- (c) **Bidirectional elimination**- It is a combination of both forward and backward selection.

IV. Partial Least Square Regression

PLSR is a type of regression, which is related to PCA. It is not employed to calculate the high variance hyperplanes between the response and the independent variable, but instead, the linear regression with expected and observed values are calculated by PLSR. PLS is given as:

$$A = M \times L^T + G \quad (2.8)$$

$$B = N \times R^T + H \quad (2.9)$$

where A represents predictor matrix having dimensions $a \times b$, B represents response matrix having dimensions $a \times c$; M and N denotes a matrix having dimensions $a \times d$ that represents

projections of A and B respectively; L and R denotes the matrices having dimensions $b \times d$ and $c \times d$ orthogonal loading and G and H denotes errors, assumed to be independent and identically distributed random normal variables.

V. Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) is a prevalent method for quantification of soil properties (Shepherd and Walsh 2002). Friedman designed it in 1992 for solving the regression problems. It is a non-parametric technique which forms the relationship between the set coefficients and basis functions that are drawn from the input data. It is based on ‘divide and conquer’ strategy which firstly divides the input into various regions, with each region has its regression equation which makes it convenient for the problems which have high dimensions like the hyperspectral data.

2.2 Literature Review on Hyperspectral Classification

Burnase et al. (2002) stated that HSI acquisition systems are proliferating with increasing dimensions. They further stressed that a proper technique is needed to extract useful information from these dimensions. So, they used DWT to reduce the high dimensional features of HSIs and explored different families of wavelet for this purpose. They concluded that the biorthogonal wavelet transform requires less processing time as compared to frequently used HAAR wavelet transform.

Plaza et al. (2004) proposed an approach for the HSI analysis to perform unsupervised mixed pixel classification. It is focused on the morphology of mathematics which is extended to multi-channel imagery. Extended morphological profiles for hyperspectral analysis provided satisfactory results. The experiments are performed on the HSI captured by AVIRIS and ROSIS sensors.

Extended morphological transformations by simultaneously considering the spatial as well as spectral features contained in the HSIs, helped inaccurate classification as reported by Plaza et al. (2005). They found that if the spectral-spatial characteristics are factored into the equation coupled with the morphology, the classification of agricultural and urban land cover from the images will yield good classification results.

The use of a composite kernel to enhance the classification accuracy of HSIs is performed by Camps-Valls et al. (2006). The proposed framework exploited mercer’s kernels properties for developing the family of composite kernels. It also incorporated the spectral-spatial information, and also has greater computation efficiency.

The method employed by Tarabalka et al. (2009) merged the result of classification performed by SVM with the segmentation map that is achieved by partitional clustering methods. The clustering techniques used are ISODATA and Gaussian mixture resolving methods. The combination of spatial information obtained by these techniques with the classifier produces homogeneous maps which help

in efficient spectral-spatial classification of HSIs. The images considered by them are - Indian Pines and Pavia University.

Phillips et al. (2009) tried a new way to increase the classification accuracy of the HSI by taking care of excessive noise in the HSIs, which generally occurs during the acquisition process. They used to adaptive filter based on signal to noise ratio (SNR). The algorithm developed uses the Hermite curves by which area under the SNR curve can be approximated. The SNR curve is a function of bands of HSI. Afterward, every band is passed to the median filter. This adaptive noise filtering procedure improves the classification accuracy by 10% when using the same discriminant method.

As HSI is an N-dimensional image so to classify it Bourennane et al. 2010 proposed a tensorial method, by which denoising and dimensionality reduction can be performed jointly on a spectral-spatial HSI. Wiener filter with some modification is used for denoising of the multiway data. The resultant quadtree obtained after filtering is decomposed to tensor to account the local information in the Wiener filtering. This approach is called the Multidimensional Wiener filtering, which performed well for the denoising followed by dimensionality reduction, which is integrated to SVM for boosting the overall classification rate.

The combination of Gaussian Processes (GPs) and Conditional random fields (CRFs) is performed by Yao et al. (2010) for HSI classification. GP is used for utilising spectral correlations, and CRF employs the spatial information. Afterward, the integration of these represents the spectral-spatial features set that is utilised for classification.

Li et al. (2012) used the active learning algorithms with loopy belief propagation to propose a spectral-spatial hyperspectral classification framework. The highlight of this work is that they incorporated the marginal probability distribution to learn *via* spectral and spatial information of HSIs. Classification is performed with the help of logistic regression classifier that is contained in the probabilistic discriminative random field. The results are generated on AVIRIS and ROSIS images with a maximum overall accuracy of 94.76%.

The use of multiple classifiers for classifying the HSIs is performed by Wang et al. (2012). They combined the different classifiers named minimum distance, maximum likelihood, and SVM using C5.0 decision tree. In addition to this, they used wavelet-PCA transformation to reduce the dimensions of HSIs. Experiments are performed on Indian pines, and maximum class accuracy of 99.23% is achieved by their method.

The use of fractional-order Darwinian particle swarm optimisation to build an efficient classification model is performed by Ghamisi et al. (2013). They build this model on top of the SVM classifier, which improved the classification accuracy of SVM. Their method takes less CPU time and provides better fitness value. The practices use the multiple solutions generated by swarm optimisation, which increases the accuracy and improves CPU performance.

As HSIs are formed of hundreds of bands which covers a large set of information. So, to efficiently process these images, Sujithra et al. (2013) used discrete wavelet transform (DWT) and Walsh

Hadamard transform to exploit the spectral-spatial components of HSI. Firstly, the HSI is divided into sub-sample images with the application of DWT, and then Walsh Hadamard transform is applied to each divided block to achieve the best compression ratio. Their method produced the best compression ratio as compared to some well-known methods.

Tsai et al. (2013) employed a novel method to extract the features from HSIs. These features are then passed to the standard classifier for classification. They extended the traditional texture measure of grey-level co-occurrence to a tensor of the hyperspectral cube to efficiently extract the discriminant features. The kernel size for both spectral and spatial domain is set using the semi-variance and spectral separability measure.

A robust framework is proposed by Ghamisi et al. (2014), which is based on spectral-spatial data. To extract the spatial information, they used the extended multivariate attribute profile and passed the hybrid spectral-spatial features to the random forest for classification. In addition to this, feature extraction is employed to the HSI to decrease the excess dimensions. Experiments are performed on Pavia University and centre dataset. Better classification results are achieved.

Edge-preserving filtering, in conjunction with the spectral-spatial classification framework, is used by Kang et al. (2013) to produce impressive classification results. The procedure is divided into three parts. Firstly, the HSI is classified using a standard classifier. Afterward, the generated classification response is shown as multiple probability map. On each map, edge filtering is performed. Then, the filtered maps point to resultant classes. Their proposed method generated the maximum classification accuracy of 99.43% in case of Pavia dataset.

Vasques et al. (2014) stated that the utilisation of visible/near-infrared diffuse reflectance spectroscopy (VNIRDRS) in soil survey and taxonomic is still underexplored. They constructed a pseudo-multi-depth spectral curve which was used to form the classification model. Then Principle Component Analysis and Multinomial Regression techniques were used to classify 291 soil samples, based on levels of order, suborder, and suborder + textural classification. Order level obtained 67% agreement in comparison to suborder and suborder + textural classification, which obtained only 48% and 24% agreement respectively.

Multiple-kernel learning-based spectral mixture analysis framework is proposed by Liu et al. (2014). It takes a set of non-linear kernels for characterising hyperspectral data for improving the discrimination capacity in classification. Kernel combination process is learned automatically, which provides the feature to choose kernels easily. Due to kernel combinations, it generates promising results in the classification of real-time HSIs.

Zhou et al. (2014) stated that Extreme Learning Machine is drawing attention due to its learning parameters. So, they proposed the spectral-spatial composite kernel combined with ELM which generated good classification results for HSI. To show the effectiveness of their method, they compared it with SVM and SVM with a composite kernel.

Markov random fields and rotating forests are combined to form a robust framework for classifi-

cation by Xia et al. (2014). Class probabilities are calculated by rotation forests, which extracts the essential features with assigned learners. Then the obtained features are divided into many disjoint sets. On each set, feature extraction is applied, and the new feature set is obtained. The ensemble of classifiers is trained on this new feature set, which generated accurate results.

Zhong et al. (2014) used support vector conditional random fields classifier. The designed classifier is based on the Mahalanobis distance for similarity measure. SVM is used for spectral features to estimate the precise class probabilities. Mahalanobis distance is used for spatial information to perform spatial smoothing. This spectral-spatial framework is used to classify the HSIs obtained from sensors like HYDICE, IKONOS, and QuickBird which produced excellent results with a maximum classification accuracy of up to 99.50%

Zhou et al. (2014) observed that dimensionality reduction is an important preprocessing for HSIs for classification. They proposed a spectral-spatial local discriminant embedding for dimensionality reduction. In this method, the HSIs are first smoothed by mean filtering, and then the spectral information is preserved by local preserving matrix. The spatial information is preserved by the local pixel neighborhood matrix. In the end, the discriminative projection is learned, which provided $95.9 \pm 0.40\%$ accuracy in case of Indian pines.

Regularised sparse representation is proposed by Yuan et al. (2014). They stated that due to independent sparse reconstruction procedure, the similarity in the sparse vectors of the same sample is lost. To rectify it, they used a centralised quadratic constraint to regularise the model, which generated quite accurate results.

A posteriori framework for spectral-spatial classification is proposed by Sun et al. (2014). The spectral data fidelity and a spatially adaptive Markov random field are incorporated their proposed framework. The data fidelity is based on the sparse multinomial adaptive total variation regularisation. Thus, the model takes the edge in fully utilising the spatial and contextual information of the HSI, which improves the classification accuracy.

Singular spectrum analysis is a non-parametric spectral analysis method which is hybrid of many classical techniques including time series analysis and signal processing. Zabalza et al. (2015) extended the singular spectrum analysis to 2-D. With this, all the bands are decomposed into various trends, oscillations, and noise. With the help of trends and selecting oscillations as features, the reconstructed signal became more robust, which are then passed to SVM for classification.

Multiple kernels are used by Fang et al. (2015) to classify the HSIs using the spectral-spatial data. Firstly, they used an over-segmentation algorithm to cluster the HSI into superpixels. Then, they used three kernels to utilise spectral-spatial information properly. In the end, these filters are combined, and combined data is given to SVM classifiers, which outperformed well-known methods.

Deep learning is used by Liu et al. (2015) to design a spectral-spatial framework for hyperspectral classification. The spectral features are passed to stacked autoencoders. These autoencoders non-linear features are projected to high dimensional space with compact representation. The spatial

features are refined by utilising superpixel. Then both are combined to form the hybrid spectral-spatial features which are classified by logistic regression.

Deep belief network is used by Chen et al. (2015) for efficient classification of HSI. The spectral features are given to deep belief network for feature extraction. Afterward, PCA is applied on spectral features, and spatial features are extracted by placing a $\times n$ mask for principal components. Then both are combined to form joint spectral-spatial features which are classified by logistic regression. Experiments are performed on Pavia University and Indian pines dataset.

Slavkovikj et al. (2015) proposed a CNN based hyperspectral classification framework. They used CNN for feature learning of high dimensional images. The proposed framework can learn structured features, which helps to improve the discriminatory power of the classifier. Experiments are performed on Indian pines dataset.

A deep learning framework with feature extraction to classify the HSIs is proposed by Zhao et al. (2016). They used the local discriminant embedding algorithm to extract the spectral features. The convolutional neural network extracts the spatial features. Afterward, the extracted spectral and spatial are stacked together to extract the fused feature, which is used for classification by classifiers. Their proposed method produced the highest accuracy.

Li et al. (2016) observed that the existing sparse representation methods ignore the residuals associated with the sparse representation. They proposed a hyperspectral classification method which can handle these residuals and called it robust sparse representation-based classification (RSRC). They further extended this RSRC to robust joint RSRC where they represented pixels around the neighborhood with residuals combined with a linear combination of training pixels. This method generated good results and outperformed some popular classifiers.

Chen et al. (2016) designed the applications of convolutional neural network (CNN) for extracting features from HSIs. They developed 3-D CNN by using many pooling and convolution layers. It extracted the deep features from HSI with $L2$ regularisation and dropout to avoid the overfitting of the trained CNN prediction model. The designed network is tested on Indian pines, Pavia University and Kennedy space center dataset, which generated quite impressive results with a maximum overall accuracy of 94.11% in case of Kennedy space centre.

Qi et al. (2016) focussed on deep learning and SVM for classification of HSI. Firstly, they used an Ex-Adaboost learning technique and then proposed a deep learning model called DeepSVM. Ex-Adaboost selects the SVMs with least error rates and also saves the weights associated with them. In the end, the selected SVMs are stacked into layers to extract the deep features of the HSI. This method combines the ensemble learning with deep learning to achieve some impressive results.

A semi-supervised multi-category framework for HSI classification is proposed by Li et al. (2016). This framework uses 1-D multi-embedding, which is based on interpolation. Using this, they embedded data into various 1D manifolds to perform binary classification. Afterward, it is extended to multiple categories with changing weight strategy to improve the classification rate of their method.

The experiments are performed on MNIST dataset to show the effectiveness of the work.

A spectral-spatial framework using stacked autoencoders, CNN, spatial pyramid pooling, and logistic regression is proposed by Yue et al. (2016). Stacked autoencoders are used to get the deep high-level features from spectral data and spatial features are extracted using modified deep CNN where spatial pyramid pooling is used for the first time for pooling the spatial map of CNN layers into a fixed-length feature. Then the spectral-spatial features are combined for classification. Their framework generated maximum class accuracy of $99.91 \pm 0.08\%$ in case of Pavia university dataset.

Active deep learning architecture is employed by Liu et al. (2016) for effective HSI classification. Their model performs incremental supervised training by assigning weights to the connections. The designed model tries to maximise the selection conventions defined by uncertainty and representative. Their model performed well as compared to other popular models.

HSI classification with the help of curvelet transform is performed by Qiao et al. (2016). Singular spectrum analysis for interpreting the HSI spectrum is also employed. After the processing, the curvelet is utilised for extracting the effective features. The curvelet processed feature does not contain any kind of noisy bands with preserved geometric structure. These features are passed for SVM for the construction of effective prediction model. The designed model is able to improve the overall classification of 10% in case of Indian Pines dataset.

Zhou et al. (2016) presented a new feature extraction technique, which is based on deep learning. Their technique is able to utilise the spectral-spatial features in the deep network. They combined the group data in the training process of the network with the help of weight-decay. Their technique is able to produce good results and outperformed other commonly employed models, including SVM with PCA and ICA features, Deep belief network. It achieved the maximum overall classification rate of approximately 99.45% in the case of Salinas dataset, proving that the model is able to effectively reduced the hyperspectral feature set and still generate accurate results.

Cao et al. (2017) fully exploited the spatial information of the HSI for efficient classification. They observed that recent works extract the spatial information mostly in the pre-classification step or in the post-classification step. So, they proposed a 3-D DWT which fully utilises the spatial information in both stages. In addition to this, they used Markov random field to model the local correlation of pixels followed by a Bayesian perspective. Experiments conducted showed that the proposed method achieves a significant performance gain.

Mou et al. (2017) used the recurrent neural network for the first for classification of HSIs. They demonstrated that hyperspectral data could be considered as sequential data for RNN. Two different RNN's – LSTM and Gated Recurrent Unit (GRU) are designed with new activation function termed as 'PRetanh'. Experimental results showed an overall classification accuracy of 88.85% in case of Houston dataset inferring that the HSI data is sequential in nature.

Liu et al. (2017) stated that deep learning had made significant achievements in the field of data mining and analysis. They designed a spectral-spatial framework which combines the CNN, deep

belief network, and PCA. They also solved the problem of imbalanced samples by introducing virtual samples. The experiments are performed on HSIs captured by HYDICE sensors. Their work showed promising results with 96.68% overall classification accuracy.

Four deep learning spectral-spatial frameworks are proposed by Yang et al. (2018) namely 2-D CNN, 3-D CNN, recurrent 2-D CNN, recurrent 3-D CNN for HSI classification. Rigorous experiments have been conducted for a comparative analysis of these frameworks. All these frameworks performed better especially the recurrent version of 2-D and 3-D CNN in comparison with other state-of-art classifiers.

Deng et al. (2018) stated that in recent years, hyperspectral classification using CNN showed tremendous growth with exceptional results as compared with conventional machine learning models. They tried a new form of neural networks called capsule networks for classification to further improve upon CNN. They used a modified two layers CapsNet, which is used to classify the Salinas and Pavia University HSIs. Effectiveness of the method is shown by comparing it with CNN, random forest, and SVM. The method performed more accurately with the average accuracy of 95.9%.

A spectral-spatial framework for HSI classification is proposed by Abdelfattah et al. (2018). Spatial features of HSIs are extracted using PCA-Randomized Singular Value transform (PCA-RSVD). Also, the 3-D DWT is applied for spatial information followed by Markov random fields, which is used to find the local spatial correlations. The classification map is generated by SVM and Logistic regression, which is taken as a maximum posterior classification problem by Bayesian perspective, which generated quite accurate results.

Deep learning is successfully implemented in HSI classification by Ma et al. (2019). They used LSTM, a deep learning framework. LSTMs are a form of recurrent neural networks that work best on sequential problems. So, they considered the HSI, a sequential problem. In addition to this, pixel matching and block matching techniques are used to select the sequence candidates from the images which are then passed to LSTM for classification. The overall accuracy of 96.2% is achieved in the case of Pavia University.

The hyperspectral signature is considered as a sequence by Hang et al. (2019) for passing signatures as inputs to RNNs. They proposed a cascaded RNN model with the help of GRUs to explore the redundant and complementary information in HSIs. Two RNN layers are used, the first one is used to remove the redundant bands, and the second layer is used to extract relevant information from non-adjacent bands. In addition to this, they extended this model by obtaining spatial information from CNN. The experimental results showed that deep learning model provides a better classification accuracy.

A spectrum-based deep metric learning model and conditional random field algorithm are used by Liang et al. (2019) for HSI classification. The distance metric used in the deep model is Euclidean space, and the loss function is a centre loss. The conditional random field is based on Gaussian edge potentials and is applied to take care of the geometric and Euclidean distances between two

pixel's features. Experiments are performed on Pavia University and Salinas datasets with a maximum overall accuracy of 99.10%.

Kang et al. (2019) used CNN for classification of HSIs. Before passing it to the CNN, PCA based edge-preserving features and extended morphological profiles are applied to the HSIs to generate the probability maps for classification. Afterward, a decision function is constructed for obtaining the pseudo-training samples. In the end, it is combined to form a deep feature fusion network (DFFN) to which HSIs are passed for classification. DFFN produced good classification accuracies in case of Indian pines and Pavia university dataset.

Deep learning architecture named Generative adversarial networks (GANs) are employed by Feng et al. (2019) for effectively classifying the HSI. They stated that the sample gathering for HSI is cumbersome, and the gathered sample does not have much discrimination. Their method employed two generators, which are capable of generating efficient spectral-spatial samples for effective training. Their models produced a high classification rate with 99.5% overall accuracy by utilising only 3% samples from Salinas dataset.

A new technique where 3-D CNN is merged with ConvLSTM for HSI classification is designed by Seydgar et al. (2019). Their model utilised the capability of CNN and ConvLSTM. Firstly, with the help of 3-D CNN, important features are extracted from the input HSI. The output features have more spectral features as compared to spatial features. Afterward, these features are passed to ConvLSTM for effective recurrent analysis. Their model is able to achieve high classification rate with 1 to 2% improvement as compared to existing models.

Gated recurrent unit (GRU) is utilised by Pan et al. (2019) for HSI classification. Their model used GRU for identifying the associations in the spectral bands, and the spatial characteristic of the HSI is fused as the hidden state in GRU. The model is able to generate high overall accuracy of 99.14% in case of Pavia center dataset.

Multi-task GANs are employed by Hang et al. (2020) for effective classification of HSI. They stated that deep learning architectures required a high amount of data for effective training. Their model utilises the unlabelled samples. The designed model is able to reconstruct the HSI cube and identify the class of the cube. The adversarial learning employed in the technique provides a great framework for effective training of the model as the generator produces the near-real samples to improve the discriminator. Their model produces quite accurate results with 92.32 % overall classification rate in case of Houston dataset. CNN is utilised by Ye et al. (2020) for effective classification of HSI. Their model assigned weights to the spectral bands prior to performing the classification. They partitioned the HSI pixels into various regions such as visible, infrared and passed it to the model for training. Afterward, the samples that are unclassified and incorrectly classified are recognised from the prediction output. Then the model calculates the individual weight for assignment to the bands which is passed to the CNN classifier for classification, that generated highly accurate results for classification.

Dense residual-network is employed by Cai et al. (2020) for effective HSI classification. Their model is based on the spectral-spatial feature extraction of HSI. For extracting spatial characteristics, their model employed enhanced residual framework with RELU as the activation function. The model showed a maximum increase of 0.49% in case of Indian pines as compared to other state-of-art models.

2.3 Literature Review on Soil Properties Quantification

Regular monitoring of soil is needed to keep it healthy. It can be done remotely with the help of satellites or by laboratory analysis. Various works provided a solution to quickly estimate the soil properties by using various techniques which are discussed below:

Bowers et al. (1971) experimented several different thermally dependent soil process such as evaporation by manipulating the reflectance values. They presented that the data recorded from spectrophotometers about organic matter, particle size, and moisture content have a direct relation with the reflectance and absorption of solar radiant energy by soils.

The study of spectral reflectance of the sampled soil in the infrared and visible region is performed by Krishnan et al. (1980). They collected soil samples from different parts of Illinois and recorded soil reflectance values with a spectrophotometer. With these values, Krishnan developed a stepwise multiple linear regression model for the prediction of the organic matter content of the soil. The results showed the optimal wavelength that is ideal for predicting organic matter content of the soil.

Near-infrared diffuse reflectance spectrophotometry was examined by Dalal et al. (1986) to determine soil Moisture, Carbon, and Nitrogen from air-dried soils. Multiple regression is used to build the prediction model, which shows good accuracy. The authors concluded that the near-infrared spectrophotometry provides a rapid solution to approximate the soil characteristics.

Agbu et al. (1990) conducted a study to establish a relation between the high-resolution satellite data from SPOT (Satellite Pour l'Observation de la Terre) and soil characteristics which helps in map unit delineation and classification. Morphology, chemical, and physical properties are predicted by using the spectral data from satellite and vice-versa with the help of multiple regression models. Results showed that that surface and some subsurface properties are relevant to the classification of soil and can be used to predict satellite spectral data.

The particular bands were selected by Coleman et al. (1991) to quantify the various soil characteristics. The bands were selected from the electromagnetic spectrum. They found high correlations between the soil variables and spectrum values. The critical bands were selected by using the regression equations. The chosen bands were quite accurate in predicting soil properties.

Coleman et al. (1993) investigated the possibility of remote sensing in determining different soil properties from the radiance values of Landsat's Thematic Mapper. The overall accuracy reported by Coleman is 97.2% for differentiating the different types of soil. They concluded that Spatial resolution

of Thematic mapper is too coarse for prediction of other soil properties like iron oxide and organic matter.

Airborne Multispectral sensor and satellite sensor (SPOT and Landsat TM) are used by Barnes and Baker (2000) for mapping soil texture in Maricopa, Arizona. Soil textural maps were compared between the textural analysis resulted from the top 30 cm of the soil profile at 120m of grid spacing. They concluded that the spectral classification using field basis generated a good accuracy in mapping the soil textural classes.

In 2000, Chang et al. studied the use of passive microwave remote sensing to estimate physical properties of soil. They built the prediction model using an ANN to predict the soil texture from temperature and moisture data. They used the data from the Washita '92 experiment and concluded that it is feasible to deduce the texture of soil from temperature and soil moisture data.

Thomasson et al. (2001) shifted the focus to soil reflectance values for the determination of different soil properties in Mississippi. The collected samples values were recorded from 250-2500nm. They concluded that there exist significant relationship exists between the spectra and properties of soil, but it not consistent when the field differs. The spectral regions of 400-800 and 950-1500nm have high discriminatory power.

Chang et al. (2001) evaluated the ability of near-infrared reflectance spectroscopy in the prediction of soil characteristics. The PCR model is built using the first-order derivatives of optical density. The characteristics predicted were soil carbon, nitrogen, calcium, and cation-exchange capacity. The model showed good accuracy proving it can be used for rapid results.

The use of radial basis function in estimating the soil properties in the near-infrared region by Fidencio et al. (2002) proved to be more accurate than multi-layer perceptron (MLP), and PLSR with root mean square error of prediction of 0.25%. Radial basis function is used with regularised forward selection to decrease the complexity of the prediction model.

Hutchinson et al. (2003) used a series of Radar images to determine the soil moisture in Kansas accurately. Different techniques used are field data collection, radar image interpretation, and modelling using geographical information system (GIS). They also worked on determining the scattering behavior of the vegetation and concluded that the radar images more accurately predict soil moisture in case of evenly dense vegetation area.

To predict the soil nutrients, Daniel et al. (2003) built an artificial neural network to estimate the soil Phosphorous, organic matter, and potassium from reflectance values measured in situ photometers. The neural network is trained and tested on 41 soil reflectance values from Lop Buri, Thailand. The authors concluded that the neural networks have great potential for accurate prediction by solving the problem of high cross-channel correlations that exists in classical statistical methods.

Islam et al. (2003) analysed ultraviolet, visible and infrared range in the spectrums to estimate pH, electrical conductivity, magnesium, potassium, sodium, clay, sand, and silt contents. Principle component regression is used to build the prediction model between the reflectance and ground-truth

values of the minerals. They concluded that the UV-VIS-IR ranges have the potential to provide a quick estimation of soil properties.

Udelhoven et al. (2003) developed a prediction model based on PLSR for quantitative prediction of calcium, manganese, iron, and potassium in the visible and near-infrared areas. They found that in a landscape context, the prediction of soil organic carbon can be improved, but they also concluded that these types of prediction models could also be used accurately in homogenous geological areas. They also suggested that the brightness of the samples in the visible and infrared region plays a critical role in the prediction of carbon. Their method can be used for quick estimation of organic carbon and total metals when laboratory costs are a vital factor.

Odhambo et al. (2004) designed a fuzzy neural network for the classification of soil from radar satellite images. The network clusters the different soil profiles from the radar images. The newly formed network outperforms all the conventional methods and is at par with visual inspection and interpretation of radargrams.

Bajwa and Tian (2005) used the airborne HSIs for determining fertility characteristics of soil in two fields of Illinois, Missouri. They determined pH, organic matter, calcium, magnesium, potassium, and electrical conductivity. They built the PLSR model in the very near-infrared region on the derivative spectra and reported high accuracy of the fertility factors.

High-resolution multispectral data by Ikonos satellite is used by Sullivan et al. (2005) to determine the variability of topsoil properties. Sullivan experiments were 4-fold for determining the soil variability. Firstly, they used kriging, followed by the application of co-kriging with soil characteristics and multispectral data. Thirdly, multiple regression is used based on the relation between the multispectral data and soil properties. Then, they used to cluster the data by using fuzzy c-means. They concluded that the results produced by fuzzy c-means are most accurate for predicting soil attributes.

Bricklemyer et al. (2005) observed that the minerals and characteristics of the soil vary in the same field at shorter distances. They examined that the quantity of soil organic carbon varies in two wheat fields of Montana. They found no variance in carbon within sites. But they found that the soil carbon is more in the no-till field as compared to tillage. They used variability within and between microsites to optimise the microsites within fields and the number of cores within microsites for accurate soil carbon reading.

In the same way, Chen et al. (2005) mapped the concentration of soil organic carbon from the remotely sensed images of bare surfaces of three fields. They developed the regression model, to find the relation between the reflectance values of the images and the corresponding soil organic carbon. The model developed was different for each field due to the high spatial variability of carbon. R^2 values of 0.89, 0.93, and 0.95 of three fields showed a substantial relation between the reflectance values and the carbon content. Then classification and clustering are applied for mapping the carbon from 21 collected soil samples, which showed R^2 of 0.81 and 0.87.

Selige et al. (2006) studied the variability of the soil contents using the airborne sensor for the

mapping of the soil in the high resolution. They found that the soil organic matter and soil texture are correlated with each other *via* the spectral properties. The quantitative prediction of the soil organic matter is predicted with the help of multivariate analysis using the PLSR Model or with the Multiple Linear Regression model. The validation test reveals 89% and 91% between organic carbon and nitrogen.

Whiting et al. (2006) advocated that HSIs are better than multispectral images in terms of the mapping of the crops for Precision agriculture. They developed a spectral analysis method for the plant growth with the help of information like soil attributes, pre-harvest defoliant, etc. from HSIs. They founded the most influential bands and narrow-band index method for mapping the variability in crops and soils.

Dematte et al. (2007) used the Enhanced Thematic Mapper/ Landsat-7 data to quantify several soil properties in Sao Paulo and Mato Grosso, Brazil. They collected 110 samples to validate the predicted results. They used multiple linear regression to accurately predict clay, sand, organic matter, potassium, calcium, magnesium, phosphorous, etc. They proved that the estimation of soil properties from satellite radiometry provides a quick and accurate result without the harming environment.

Rossel et al. (2007) used diffuse reflectance spectroscopy to rapidly obtain spectral data which is used in building the PLS with bootstrap aggregation. Its main task was to reduce the variance of predictions by combining different models. This approach proves much robust than PLSR and is less prone to overfitting.

Ge et al. (2007) tried to make Near-Infrared Spectroscopy (NIRS) more accurate by including the wavelet approach in the pre-processing of near-infrared spectroscopic data captured by proximal sensing. They evaluated this technique with two datasets—synthetic and original. The resultant wavelet regression model showed less error and more R^2 of 0.99.

The advantages and disadvantages of imaging spectroscopy for the characterisation of soil properties are discussed by Ben-dor et al. (2009). The discussion extended to include other problems like soil degradation, classification, and contamination. The problems occurred during the conversion to reflectance data and noise in the imaging dataset are also underlined. Various new ideas were presented in their work going forward in the field of soil science and remote sensing.

Improved analysis of diffuse reflectance spectra is done by Viscarra et al. (2009) by using wavelets to smooth signals and reduce the significant representations of other signals that result in less computation power and low storage space. Daubechies's wavelet with two vanishing moments is used. Multivariate calibration was performed by PLSR and multiple linear regression using the reduced features resulting after wavelet transformation.

Kadupitiya et al. (2010) performed the quantitative assessment of the soil chemical properties by using the visible and near-infrared data from hyperspectral data. In this, the soil is collected from Jalandhar district of Punjab. Lab analysis is performed on that soil for the ground reality, which is compared with the result obtained from the Hyperion sensor. Fast Line-of-sight Atmospheric Analysis

of Hypercubes (FLAASH) is also applied to the Hyperion data for atmospheric correction. Prediction Evaluations indices confirmed that hyperspectral derived models were predicted well except Zn, Cu, and pH.

Shahid et al. (2010) observed that soil salinity is a major global issue due to which the soil has more salt as usual, which results in less crop production. Different salinity assessment, mapping, and monitoring techniques, and the similarity between them are also discussed.

The comparison of three popular prediction models—Principle Component Regression, PLSR, and Artificial Neural Network is performed by Mouazen et al. (2010) for accuracy of measurement of soil properties (carbon, sodium, potassium, magnesium and phosphorous). This comparison is performed on 168 soil samples collected from Belgium and Northern France. They concluded that ANN is the most accurate among the other models in predicting soil properties.

Dematte et al. (2010) developed narrow-band models of hyperspectral data to quantify organic matter, cations, aluminium saturation, etc. and they also compared the chemical methods and laboratory analysis of reflectance spectra to find the correlation between them. The reflectance data were obtained from the Infrared intelligent spectroradiometer. The 948 samples were collected from regions of Paraguacu Paulista and Rio Brilhante along with their Global Positioning (GPS) Coordinates. Multiple regression is used to develop the prediction model. They concluded that the laboratory spectra have a high correlation between the traditional chemical methods with R^2 of 0.84, 0.85 in case of clay and sand, respectively.

Kovacevic et al. (2010) used the Support Vector Machine (SVM) to classify and predict soil attributes. The application of SVM increased the accuracy of the prediction, and the effectiveness is shown by comparing the existing regression methods with SVM. They also reported that in the case of classification, the SVM performance is at par with the regression models.

The review of the use of pedo-transfer functions for prediction of the soil characteristics in the tropics is done by Minasny et al. (2011). The authors presented the analysis of ISRIC and USDA soil databases, concluding that various methods designed for the soils in the temperate region can be successfully used in the tropical region if the calibration is done carefully. They also found that pedo-transfer functions could be used when there is no soil data available in other locations.

Doolette et al. (2011) used a mixture of sodium hydroxide ethylenediaminetetraacetic acid, which extracted 89% of the total phosphorus present in the soil samples. They presented an improved method for assessing the composition of soil phosphorous using solution ^{31}P NMR (nuclear magnetic resonance) and spectral deconvolution. The technique not only provided a more detailed characterisation of organic P forms in Australian soils than has hitherto been undertaken but also offered a more accurate estimation of the quantity of these forms by minimising the potential for overestimation of the P concentrations of simple orthophosphate monoesters.

The use of remote sensing and geographical information system is done by Sarajlic et al. (2012) to map land cover classes and minerals characteristics in Tushka depression, Egypt. The resultant

map showed an increase in vegetation, potential agriculture land, and cover area over the past five years.

Desaules et al. (2012) suggested that to identify the natural and man-made trace contents in soils is not very efficient. They compared five different models and concluded that the soil contamination assessment methods have a methodological specificity and bias. It was due to the less availability of site-specific reference values.

Al Maliki et al. (2012) reviewed HSIs for detecting the lead contamination, which is useful in health risk assessment. They used the Very Near Infra-Red (VNIR) in the spectral range 400-2500 nm to quantify the metal concentration. This review is based on detecting the heavy metals that are harmful to humans *via* remote sensing hyperspectral imaging.

Toth et al. (2013) introduced the Land Use/Land Cover Area Frame Survey dataset, a project which contains all physical and chemical properties of the European Union. Soil basic properties like sodium, potassium, phosphorous, carbon, particle size and cation exchange capacity with location-based multispectral signatures are contained in the survey. They suggested that this dataset has the potential to increase the base soil knowledge in the European Union.

Tian et al. (2013) developed prediction models for the estimation of soil organic matter collected from the seven eco-friendly zones in the middle and eastern China. Three formats are used in building the models. The first model is based on the original captured hyperspectral data. The second model is based on the first derivative spectra corrected using Savitzky-Golay technique and third based on Norris Savitzky-Golay smoothing filters. Principle component analysis is used to reduce the high dimensions of the hyperspectral data. The reduced data is given as input to neural networks with backpropagation for the construction of a robust prediction model. This model generated a quite high accuracy with R^2 of 0.98.

Garfagnoli et al. (2013) used the high-resolution Hyper SIM-GA sensor for quantitative assessment of clay minerals in Italy. Simultaneous ground samples were also collected for verification. The spectral region chosen was from 2000-2450 nm. They concluded that 2200 band produces a reliable result about the clay content, and there exists a considerable correlation between the original and predicted values.

Liao et al. (2013) estimated the soil texture using remote sensing. They used the Landsat Enhanced Thematic Mapper (ETM) in the regions of Pingdu, China. The study evaluated the three basic techniques to determine the variability in the soil texture. Firstly, Multiple stepwise regression is used to exploit the relation between the sand, clay contents with imaging data from the satellite. Then kriging of soil, silt and clay and co-kriging of the satellite data is performed. In the end, they concluded that co-kriging of satellite data improves the estimation of soil texture as compared to the kriging and Multiple Stepwise Regression.

The prediction of soil organic carbon is made by Stevens et al. (2013) at the European scale in the regions of Visible and Near-Infrared. They used a soil survey of the EU of 20,000 samples collected

over 23 countries. The multivariate model tested by Stevens has shown the least error of around 4-15g C/kg⁻¹.

The use of spectroscopy in the visible and mid-near-infrared is done by Gholizadeh et al. (2013) for the prediction of soil organic matter. They used different pre-treatments models to increase the accuracy of the model. Also, the review of varying calibration scales is given. They concluded that the mid-infrared when used in combination of Savitzky-Golay smoothing and derivative, generates good accuracy in case of least square Support Vector Machine to determine soil organic matter.

Luce et al. (2014) presented modified PLSR to construct the prediction model for the rapid determination of soil organic matter quality from the visible and near-infrared region. The Principal Components are used to eliminate the outliers that were present in the collected spectra, which increased the prediction accuracy.

Li et al. (2014) evaluated soil nutrients using three different models, namely- Support Vector Machines (SVM), Regression Model, and Artificial Neural Network (ANN). They reported that ANN performed the best with 92.86 % accuracy in predicting Organic Matter, Nitrogen, Phosphorous, and Potassium.

The Environmental Mapping and Analysis Program (EnMAP) hyperspectral sensor is used by Chabrillat et al. (2014) to evaluate various algorithms for quantification of soil characteristics. Spectral data and multivariate analysis are used to determine soil carbon, clay, and iron oxide. The authors reported that which algorithm is best suited for each sensor resolution.

The estimation of Soil Organic Carbon (SOC) using the multivariate prediction model and artificial neural network (ANN) is done by Tiwari et al. (2015). EO-1 generated HSIs in the range of 400-2500 nm. ANN is applied to these images obtained from the satellite, and SOC is predicted, which was then compared with the laboratory scale data (350-2500 nm). It is concluded that the ANN prediction model shows high accuracy for detecting the SOC in soil.

The web camera is used by Adhiwibawa et al. (2015) as a multispectral sensor to determine the quantity of chlorophyll in Soybeans leaves in Indonesia. This study was based on the premise that cost can be reduced to get the quantity as laboratory analysis of chlorophyll takes a lot of time and have a high price of testing. They concluded that normalised difference vegetation index correlates with Soil Plant Development Analysis with R^2 of 0.70 and prediction error was less than 18%.

Das et al. (2015) presented the current status and the challenges faced in India for soil assessment in precision agriculture. The non-availability or limited availability of spectral libraries is a big challenge for research in India. The country has 121 million agriculture field, and most of the soil testing is done by chemical analysis by 1049 labs, which are less as compared to the area of the agriculture field. It has been discussed that Diffuse Reflectance Spectroscopy (DRS) is an attractive technique as the proximal and remote mode is useful in estimating the soil characteristics as compared to the hyperspectral remote sensing.

Veres et al. (2015) leveraged the deep learning architectures particularly Convolutional Neural

Network and Boltzmann Machines for constructing the prediction model for estimating soil attributes. The constructed model was tested on two datasets- LUCAS and Africa Soil Information Service (AFSIS). A considerable amount of correlation between the spectral values and the quantity of measured properties has been reported.

Xu et al. (2016) retrieved soil moisture and salt concentration using hyperspectral data. Laboratory and field analysis are performed to examine three variables— soil moisture, salt, and texture. Moisture content model is based on hyperspectral data around 1460, 1900 and 2100 nm, which has shown a high accuracy of around 93%.

The use of airborne hyperspectral Imagery and the corresponding field spectra is done by Vaudour et al. (2016) to build a prediction model that can estimate soil organic carbon efficiently in the western region of Paris, France. PLSR is used to calibrate the prediction from 101 soil samples. It generated quite a high accuracy in terms of RMSE and Residual Prediction Deviation (RPD) values.

Morellos et al. (2016) tried the prediction capability of machine learning models to predict soil carbon, nitrogen, and moisture in the visible and near-infrared regions. They constructed the Least square SVM and cubist models and compared these models with conventional regression models. Morellos reported that the machine learning model named Least Square SVM have less error of 0.457-0.062% and cubist provided the best result for predicting nitrogen.

Sarathjith et al. (2016) considered the spectral aspect of the wavelength values to improve the performance of diffuse reflectance spectroscopy. They stressed the importance of selecting the correct spectral values to form a robust spectral algorithm. They used PLSR, correlation coefficients, weight-specific correlation, *etc.* for optimal selection of the spectral variables, which increased the performance of spectroscopy.

Harris et al. (2016) used NASA's Hyperion sensor to identify and quantify the salts which are abundant in the Atacama Desert. The use of Hyperion sensor provided a high-resolution HSI of 242 bands. This study mainly focused on the near-infrared and very near-infrared regions for identification and quantification of various salts present in the soil using spectral mixture analysis. To cross verify the collected spectra, they used the ASD. Field spec and quantitative XRD of the samples collected. They concluded that spectral mixture analysis accurately predicts the sulfate minerals present in the soil from HSIs.

LUCAS dataset is used by Ballabio et al. (2016) for mapping topsoil physical properties. They predicted the soil texture and other properties using hybrid kriging. They also fitted regression models with remotely sensed data from MODIS satellite and generated a R^2 between 0.47 and 0.50 for texture and errors between 4% to 10%, respectively.

One hundred and two samples are collected by Nawar et al. (2016) from northern Sinai, Egypt to build the prediction models for the estimation of soil clay content and organic matter. They constructed three regression techniques for prediction, namely PLSR, Support Vector Regression, and Multivariate Adaptive Regression Splines. They concluded that the Multivariate Adaptive Regression

Spline is the best performing model with $R^2 = 0.85$, RMSE = 0.19% and Ratio of Performance to Deviation = 2.63.

Coopersmith et al. (2016) identified that soil moisture is an essential parameter for analysing the energy flux at a land boundary. They used moisture sensors from US reference network as input to machine learning techniques. Validation was performed on in situ readings captured earlier. The developed model generated good results with RMSE of only $0.0215 \text{ m}^3/\text{m}^3$. After that, the same model was compared on AMSR-E reading and reported RMSE of $0.03 \text{ m}^3/\text{m}^3$. In the end, the developed model was used on the ECONet in North Carolina also provided impressive results.

Soil moisture is an essential parameter of soil. Khedri et al. (2017) estimated the soil moisture with the help of PULSOR data captured from the satellite. It was observed that it is not feasible to collect moisture data over vast areas. The data available from satellite has a high spatial resolution by which machine learning models can be trained. They used Support Vector Regression to build the prediction model for the estimation of moisture. The building model was accurate with a $R^2 = 0.97$.

Vohland et al. (2017) determined the potential of both imaging spectroscopy and laboratory analysis for predicting organic carbon, Nitrogen, Biomass-C, Extractable-C. The images were captured from Hymap sensor, which provided the hyperspectral data for easy determination of these properties. The use of a specific portion spectrum is done to efficiently predict the chemical properties with prediction deviation of 2.35 for Biomass-C and 2.20 for organic carbon in case of sensor data.

A study on four soil properties clay, free iron oxides, calcium carbonate (CaCO_3), and pH is conducted by Adeline et al. (2017). The uniqueness in this study that the PLSR model is built for the laboratory data in the visible and infrared region, which is solely based on spectral degradation.

The comparison of machine learning and multiple linear regression is done by Forkuor et al. (2017) for the digital soil mapping. Remotely sensed images were used from satellite RapidEye and Landsat for determination of sand, organic carbon, silt, cation exchange capacity. The learning techniques random forest regression, SVM, stochastic gradient boosting, multiple linear regression was compared. They concluded that overall, all machine learning techniques performed better compared with multiple linear regression.

Chivasa et al. (2017) presented a review of estimating maize production using satellite imagery. Maize account 50% of the primary energy resource in Africa. They discussed in great detail about the various high-resolution sensors that can reduce the cost and time in yield prediction and also reported that classification techniques for obtaining the ground truth from the fields.

Arrouays et al. (2017) reviewed significant improvements in the area of soil mapping. They also highlighted the areas where some work is needed. They underlined a significant point that soil mapping should be done at the regional level to ensure the integrity of data, and more investment is required for this area.

Wang et al. (2017) investigated the visible and near-infrared region to identify the rare earth materials present in the soil. A total of 130 samples were collected near rare earth elements mines.

Then a PLSR model was used for calibration of the spectrum. They observed that absorption features of rare elements aren't distinct in the spectra. But still, the built model provided a good accuracy with R^2 greater than 0.5 and RPD greater than 1.4.

Liu et al. (2017) stated that PLSR is the most employed technique in the estimation of various soil properties, which can be also be utilised for decreasing the excess dimensions of high dimensional spectra. Their technique is a hybrid of ML model gradient-boosted decision tree (GBDT) and PLS, that is able to accurately quantify the soil's OC, clay and nitrogen. The dimension reduced output obtained from PLS is passed to GBDT for effective training. They also compared their hybrid approach with individual PLS and GBDT to show that by merging these two techniques, their model is able to achieve high performance as compared to using them individually.

To predict soil organic carbon from the visible-near infrared region, Dotto et al. (2018) built multivariate models for efficient prediction. Pre-processing approaches are used before the prediction to test their influence on the models. These approaches are divided into two parts, namely—scatter correction and spectral derivatives. They observed that the best model for the soil organic prediction is built on the base of weighted PLS method with R^2 of 0.82 and RMSE of 0.48%.

Copper distribution in the European Union is mapped by Ballabio et al. (2018) from LUCAS dataset. The authors utilised the linear model for investigating the elements that affect the quantity of copper in regions. They designed two models based on regression. The regression utilised for finding copper distribution proved that the quantity of copper is associated with the OC, pH and texture of the soil, due to which its quantity is more in the vineyard fields. The second model based on Gaussian process regression is employed for creating maps for copper distribution. The model is able to account for 66% of the deviance of copper concentration.

Behrens et al. (2018) compared various multi-scale terrain feature construction techniques and their effectiveness for digital soil mapping is shown with the help of deep learning. They introduced 'mixed scaling' to overcome the traditional problems faced in mapping and also extended the Gaussian pyramid with the help of intermediate scales. Their method produced 4-7% more accuracy compared to conventional random forests.

Meier et al. (2018) used eight machine learning models to map soil in mountain areas of Brazil. The morphometric maps are generated with the combination of satellite data of Landsat-8 and digital elevation model. Essential features are selected by Recursive Feature Elimination for prediction by machine learning models. The model results were satisfactory with all models showed a kappa value of 0.42-0.48 agreeing with the conventional soil maps.

The artificial neural network grows popular due to its ability to learn by training. Bittar et al. (2018) used ANN to estimate the physical and chemical properties of soil using 120 soil samples. The results are compared with the mean errors of kriging estimates, which showed that ANN performs better due to the inbuilt learning provided by backpropagation.

The use of RGB cameras for soil composition analysis in the visible region is performed by

Aitkenhead et al. (2018). They calibrated the prediction models solely on the base of the visible region and tested the organic matter, carbon, nitrogen, pH, clay, etc. in combination with spatial variables including topography, climate, soil, vegetation, parent material. The model is prepared with the help of training by neural networks and provided excellent accuracy.

Liu et al. (2018) stated that prediction models constructed from soil library have less or no correlation with spectral data collected from field survey. They suggested transfer learning can be used to bridge this gap. They used transfer learning (1D-CNN) for estimating the soil clay content using HSIs, and the build model has shown R^2 of 0.75 and low error.

The deep learning-based framework was proposed by Chen et al. (2018) for quantitative analysis of nutrition in the soil. The framework is trained on the Fourier Transform Near Infrared (FTIR) values. The features were selected using CARS algorithm. Competitive Adaptive Reweighted Sampling (CARS) algorithm outputs 91 optimal features, which generated superior results and effectiveness of the proposed method is shown by comparing with PCR and PLSR.

Padarian et al. (2019) used CNN for digital soil mapping. The CNN takes the images as input and analyses the spatial contextual information from the neighbourhood of a pixel to improve the prediction accuracy over traditional digital soil mapping models. The model was able to predict the organic carbon accurately and reduced the error by almost 30% in comparison with other conventional techniques. They proved that deep learning is a new and efficient way of analysing data and can generate high accuracy result in case of soil mapping.

Transfer learning for quantity estimation of soil's properties is employed Padarian et al. (2019). They tried to localise the generalised models by utilising transfer learning. Their model first trained on the big dataset like LUCAS spectral library. Then the trained model is extracted, and hyperparameter tuning is performed on local data. As LUCAS contains samples collected from over 23 countries, any country data can be used for parameter tuning. This approach proved to effective as the author proved in 91% cases it is able to lower the RMSE for quantification of clay, OC, pH and CEC.

The soil's OC is modelled by Gomes et al. (2019) in Brazil. They employed machine learning (ML) models in the study. Their model was validated on 37693 samples. The authors tested 4 ML models named- Support Vector Machines, Random Forests (RF), Generalized Linear Model Boosting and Cubist. They concluded that RF produces the best results with R^2 of 0.33 on the validation set. Important features are selected using recursive feature elimination.

Memory-based learning model is employed by Tziolas et al. (2019) for enhanced estimation of soil properties. The model is validated on the standardised spectral library, which has 1760 samples. Their designed model was accurately able to predict the quantity of clay, soil, OC etc. The developed learning-based model is able to generate the 6.47% lower error as compared to existing state-of-art prediction models.

Ward et al. (2019) stated the reflectance spectroscopy is an optimistic method for quantifying the soil's OC effectively. They tested various techniques for estimating SOC, including PLSR, k-means

clustering. The authors concluded local PLSR is capable of accurately quantifying SOC when remote sensing images are given as input to it. In addition to this, they also employed PLSR to field captured LUCAS data by reconfiguring it to the hyperspectral sensor EnMap specifications. PLSR is able to achieve a R^2 of 0.66.

Lee et al. (2019) estimated the soil moisture with the help of deep learning from satellite data in South Korea. They reported that deep learning could perform better in comparison with the traditional models due to the fact that it can form a meaningful learning connection on a large data set. The proposed method shows the correlation coefficient of 0.89 and RMSE of 3.82%.

The use of infrared spectroscopy allows quick estimation of soil properties. Based on this premise, Padarian et al. (2019) build a deep learning model consisting Convolutional Neural Network to estimate the underlying soil properties—Organic carbon, Cation Exchange, clay, sand, pH and nitrogen. They compared CNN based models with the most popular conventional model PLSR and Cubist Regression Tree and reported that multi-CNN decreased the error by 87%.

Gaussian process regression is employed by Ballabio et al. (2019) for quantification of chemical properties of soil from LUCAS spectral library. Their model was able to effectively estimate soil properties like potassium, nitrogen etc. In addition to this, the regression model utilised was able to assess the uncertainty of the model, and it also has a provision of appending the previous knowledge.

Wadoux (2019) utilised the CNN for estimating the properties of soil from the LUCAS spectral library. He stated that CNN is able to interpret spatial contextual information more effectively as compared to traditional estimation models. He concluded that the CNN utilised the available contextual bands more efficiently within a radius of 4.5 km. In addition to this, the author also concluded that the model is able to effectively quantify the uncertainty in prediction.

Yang et al. (2020) merged two popular deep learning architecture named CNN and RNN for estimating the critical properties of soil from LUCAS spectral library. Their work extracted the local features by utilising CNN. The extracted features are then passed to RNN for learning various dependencies present. Their model generated good results with the highest R^2 achieved in the case of pH.

Safanelli et al. (2020) utilised the available data of Landsat satellite for estimating the soil's properties. They produced two topsoil composites from GEOS3 system. Afterward, machine learning algorithms are employed for calibration of the synthetic soil images. The model developed for estimating Cation exchange capacity, organic carbon, sand, clay etc. But after calibration, the authors concluded that the images from Landsat could only be utilised for estimating calcium carbonate and clay.

Soil's OC and nitrogen quantities are estimated by Zhou et al. (2020) via Sentinel-1 and 2 satellite radar data. The acquired data is passed to ML techniques for designing a prediction model. They tested both satellite images in order to find the best fit. The ML models employed are RF, SVM, CART and regression trees. The models are calibrated on 179 samples. In the end, the predicted

results are evaluated on the basis on RMSE, R^2 and MAE and the maximum R^2 of 0.38 is achieved in the case of SVM.

2.4 Research Gaps

Based on literature the following gaps are identified:

- I. Most of the existing matching techniques used for the identification of soil minerals use similarity measure, which is not very accurate and also has a high computation time.
- II. Existing prediction models used for quantification of the soil are based on multivariate linear regression approach which does not provide much accurate results as compared to models applicable to other domains because the only predefined functional structure is captured by multivariate linear regression models, that might not be enough to correctly estimate the nonlinear design of the variables concerned. Also, these models require complex calculations.
- III. Present techniques for quantification do not take benefit of the sequential nature of spectral data.

2.5 Objectives

On the basis of the gaps following objectives are proposed

- I. To analyse the existing spectral classification and soil quantification techniques.
- II. To design efficient matching technique using obtained spectra of the hyperspectral image.
- III. To develop efficient quantification technique for the minerals, present in the soil.
- IV. Comparison of the proposed techniques with the existing techniques.

2.6 Methodology

In this section, the methodology used in the proposed work has been discussed.

Phase 1-Data Preprocessing

Firstly, the standard hyperspectral dataset is acquired for conducting experiments. The acquired datasets are Salinas Dataset, LUCAS Dataset, Pavia University Dataset, Kennedy Space Center Dataset, Indian Pines Dataset. General pre-processing is then implemented to hyperspectral data to boost the accuracy of the proposed techniques. The methodology is shown in Figure 2.3.

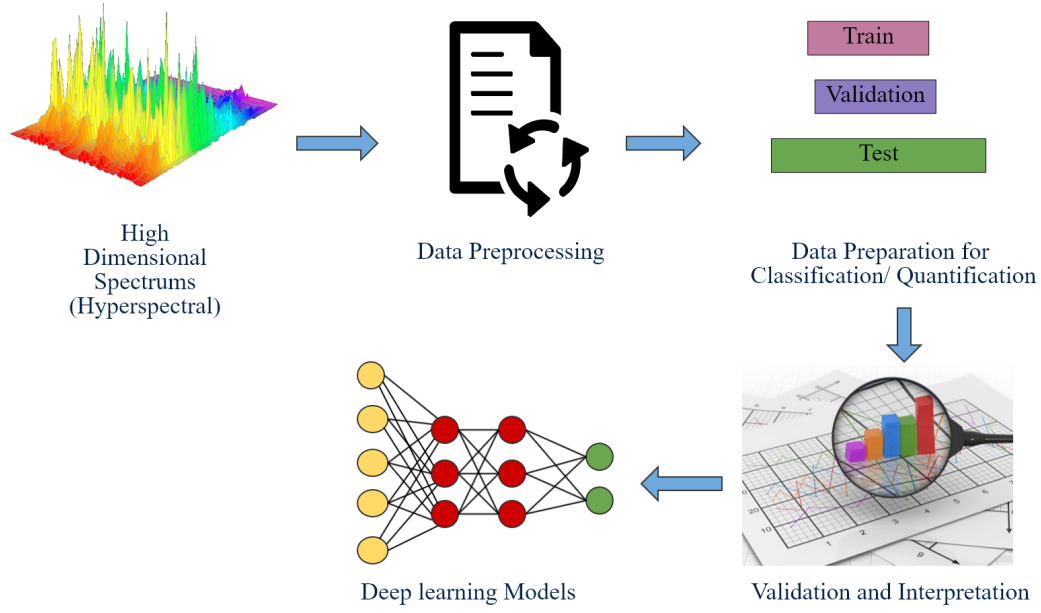


Figure 2.3: Methodology of Proposed Framework

Phase 2-Data preparation for classification/ quantification

In this phase, the prepared hyperspectral data is partitioned into training, validation, and testing set. With the help of training data, the prediction models are prepared.

Phase 3-Preparation of prediction model based on deep learning

In this phase, the training and validation set is given as input to the deep learning model for supervised training. Generally, the validation set eliminates the over-fitting; it arrives in the middle to verify whether or not the training process is progressing well. If the validation set does not work well, it means that the training was not done effectively, which is then strengthened by hyperparameter optimisation.

Phase 4- Validation and interpretation of the results

In this phase, all the obtained values will be matched against the ground truth to evaluate the robustness of the proposed prediction models. The following quality parameters are used in the classification and quantification of hyperspectral data:

$$Q = \frac{P(T) + N(T)}{P(C) + N(C)} \quad (2.10)$$

where ACC is the accuracy, P(T) is the true positive, N(T) is the true negative, P(C) is the number of positive cases, and N(C) is the number of negative cases.

Quantification quality parameters used are RMSE, Coefficient of Determination (R^2), Mean Absolute Error (MAE), Spearman correlation coefficient (ρ), Pearson correlation coefficient (r) and Lin's concordance coefficient (ρ_C) which are defined below:

$$RMSE = \sqrt{\frac{1}{t} \sum_{r=1}^t (I_r - \hat{P}_r)^2} \quad (2.11)$$

$$MAE = \frac{1}{t} \sum_{r=1}^t |I_r - \hat{P}_r| \quad (2.12)$$

where t denotes total observations, I_r represents original values and $\hat{P}_r$ represents the predicted value

$$R^2 = 1 - \frac{E(S)}{T(S)} \quad (2.13)$$

where $E(S)$ is the sum of squares (errors) and $T(S)$ is the sum of squares (total)

$$\rho = 1 - \frac{6 \sum k s_k^2}{t(t^2 - 1)} \quad (2.14)$$

where s_k is the difference in the ranks of the k^{th} element

$$r = \frac{A \sum PQ - (\sum P \sum Q)}{\sqrt{[A \sum p^2 - (\sum p)^2][A \sum q^2 - (\sum q)^2]}} \quad (2.15)$$

Where A pair scores, $\sum PQ$ is the product of paired scores, $\sum p$ is the sum of p scores, $\sum p^2$ is the sum of squared p scores

$$\rho_C = \frac{2\rho v_p v_q}{v^2 + v^2 + (m_p - m_q)^2} \quad (2.16)$$

where v is the variance and m is the mean.

Chapter 3

Efficient Classification of the Hyperspectral Images using Deep Learning

This chapter presents an efficient classification technique of HSIs using deep learning approach. Existing HSI classification techniques do not efficiently extract deep HSI features. So, we proposed a novel deep learning framework, which is capable of extracting deep features from the input HSI. The framework is a combination of LPP, Deep Neural Network (DNN), and logistic regression. LPP is employed for decreasing the excess dimensions of HSI. The LPP processed reduced dimensions are then given to Stacked Autoencoders, a deep learning architecture for extracting deep features. Then the deep features are passed to the softmax layer for classification. Salinas and Indian pines HSIs are employed for validating the effectiveness of the proposed framework. In contrast with current machine learning models, high classification accuracy is obtained with the proposed system. .

3.1 Background of the Proposed Work

3.1.1 AutoEncoders

Autoencoder (AE) as shown in Figure 3.1 is the category of deep learning models that effectively learns the encodings present in the input via unsupervised learning. There are different types of autoencoders:

Sparse AE- This type of AE generally used to learn features that help in efficient classification tasks. It includes more hidden units than the input dimensions typically. Only a small number of these hidden units are active at once.

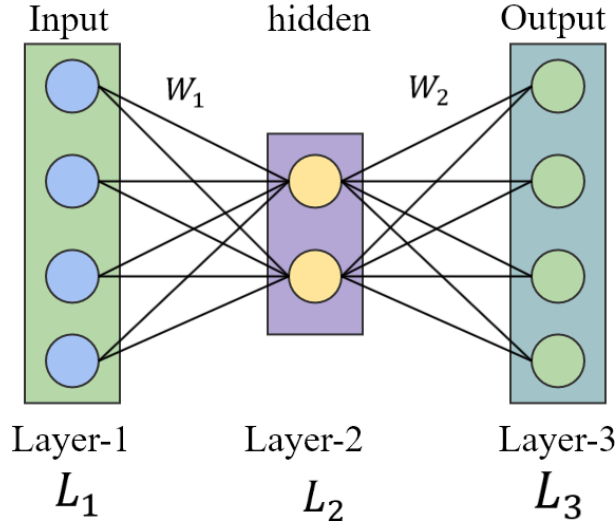


Figure 3.1: Single Layer Autoencoder

Denoising AE- It is a stochastic variant of the AE. It focuses more on the reconstruction of the input. It takes corrupted data as input and recovers the original input. It is mostly used for eliminating the noise in images.

Variational AE- This type of AEs is powerful generative models like Generative Adversarial Networks. It generates a random new data that is similar to the input training data which we need to train a robust model.

Contractive AE- It is used to encode the unlabelled training data. It is done with the help of a regulariser or a penalty term. The major aim of contractive AE is to decrease the sensitivity of the learned feature towards the input with the help of the Frobenius norm of the jacobian matrix used in the encoder sequence.

Autoencoders (AEs) are simple three-layered neural networks. These present an unsupervised method for efficient encoding of the data. Like PCA, AEs can perform efficient dimensionality reduction. Mostly, AEs are used for compression, de-noising, machine translation, *etc.* It can conserve the information in the form of deep features in the hidden layers. The main task of the AEs is an efficient representation of the lower dimensional data from high dimensional input.

$$I^k \longrightarrow H^m \quad (3.1)$$

If I^k is the input to the AEs with k dimensions, then the output should be H^m where $m < k$. The efficiency of the AE depends upon the reconstruction layer. If the reconstructed features are approximately the same to the input features, the AEs are considered to be optimal.

Firstly, the encoding of the input features is done as:

$$H^m = W_1 \times I^k + b_1$$

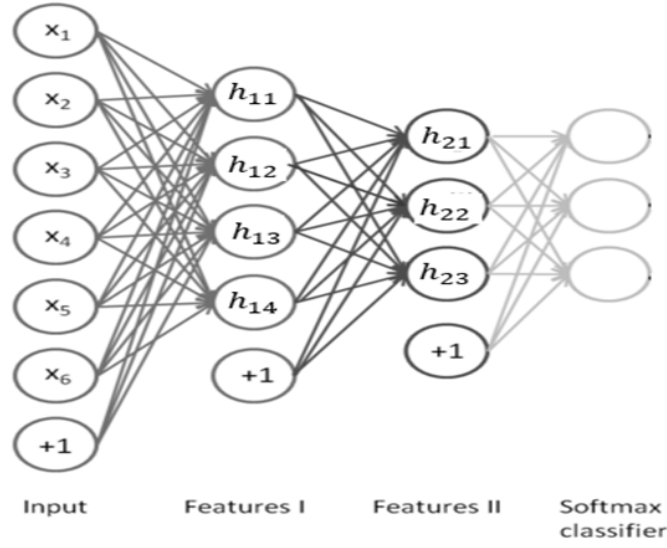


Figure 3.2: 4-layered Stacked AE including input and Softmax classifier layer

Where H represents hidden layer to which the k features of input are mapped to m features of hidden layer and $m < k$.

Afterwards, the redevelopment of the input is performed with the aid of reduced features from the hidden layer:

$$\hat{I} = W_2 \times H^m + b_2$$

Where b_2 is the bias

In the end if the I is same as $\hat{I}$ than the backpropagation is stopped and H features are considered to be the efficient features. If not then objective function is applied in the end to reduce the difference among the original and the redeveloped features. It is defined as:

$$O(I, \hat{I}) = \left\| \hat{I} - I \right\|^2$$

This equation is representing distance loss, which can be minimised using the stochastic gradient descent.

3.1.2 Stacked Autoencoders

Stacked AE is obtained by combining the different autoencoders, as shown in Figure 3.2. A hidden layer h_{1k} for input x_k contains the reduced features which are obtained by unsupervised training. The obtained reduced feature set is then passed to the second hidden layer (h_{2k}), which further reduces the input features. This AE is called the sparse AE as the feature set is reducing on every hidden layer. These reduced features are termed as a deep feature set, which is then passed to the SoftMax regression layer for supervised training. This placement of the layers is termed as stacked autoencoders.

3.1.3 Local Preserving Projections

LPP is an effective algorithm that is proficient in extracting the important features from the input dataset. It is employed to decrease the surplus dimensions of HSI. The application of LPP preserves the local neighborhood information efficiently. We employed LPP to map the high dimensional feature space of the HSI to lower-dimensional space. It is common practice in the analysis of HSI to decrease the feature set so that the classifiers discriminatory strength can be increased. PCA is another algorithm that is widely employed in HSI analysis for decreasing the feature set (Chen et al. 2015). LPP was introduced by He and Niyogi in 2003. They stated that in many cases, LPP is able to perform better compared to PCA .

LPP can be applied to any new point unlike ISOMAP and LLE, that are only defined on the training points. It generates graphs that consist of neighborhood details. Its sole purpose is to conserve the local spatial information by mapping the high dimensions to the reduced representation space. In the first step, the adjacency graph G is constructed by placing an edge between the nodes if the nodes are considered as ‘near’. The closeness of the nodes can be calculated by Euclidean norm or by k -nearest neighbors. Afterward, the weights are chosen for the edges by using a heat kernel which is defined as:

$$W_{ab} = e^{\|x_a - x_b\|^2/q} \quad (3.2)$$

Where $q \in \mathbb{R}$.

In the end, the eigen maps are computed:

$$XLX^T a = \lambda XD X^T a \quad (3.3)$$

The lower-dimensional output is obtained in the form of vectors that is represented by a . These column vectors are arranged according to the increasing eigenvalues. The selected vectors represent the lower manifold space preserving local spatial features.

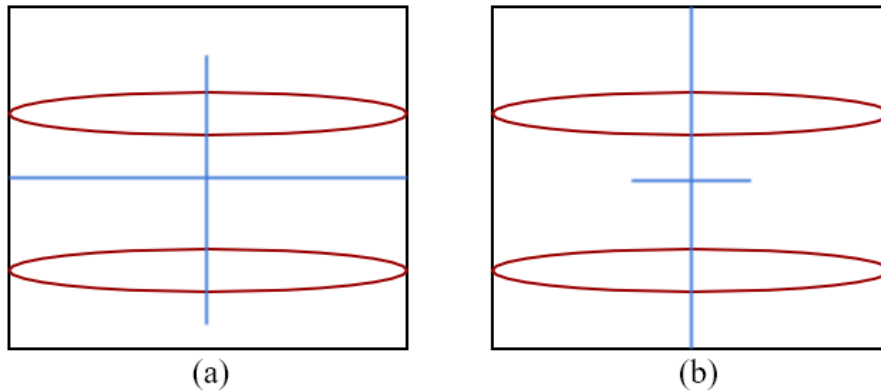


Figure 3.3: Discriminating power of PCA (a) and LPP (b)

On the other hand, PCA is more effective in maintaining the global approximation of the given input. Its aim is to maximise the variance in the resultant output. The Principal components are computed by utilising covariance matrix, eigenvalues and vectors. The computed eigenvector, in the end, represents the low feature space. It is efficient in maintaining global features, but the local features are not effectively captured by PCA as compared to LPP. In both PCA and LPP, the selected eigenvectors represent the low dimensional output. The PCA's selected eigenvectors are arranged according to the eigenvalues. The eigenvalues are placed in decreasing order in the case of PCA, which is completely opposite to LPP, as LPP's eigenvalues are placed in increasing order. Due to this, LPP is able to decrease the dimensions effectively by preserving the local features. Figure 3.4 shows the two selected principal components plotted on x and y-axis and Figure 3.3 shows that the discriminating strength of LPP is more than PCA, due to this LPP performs better than PCA in many cases.

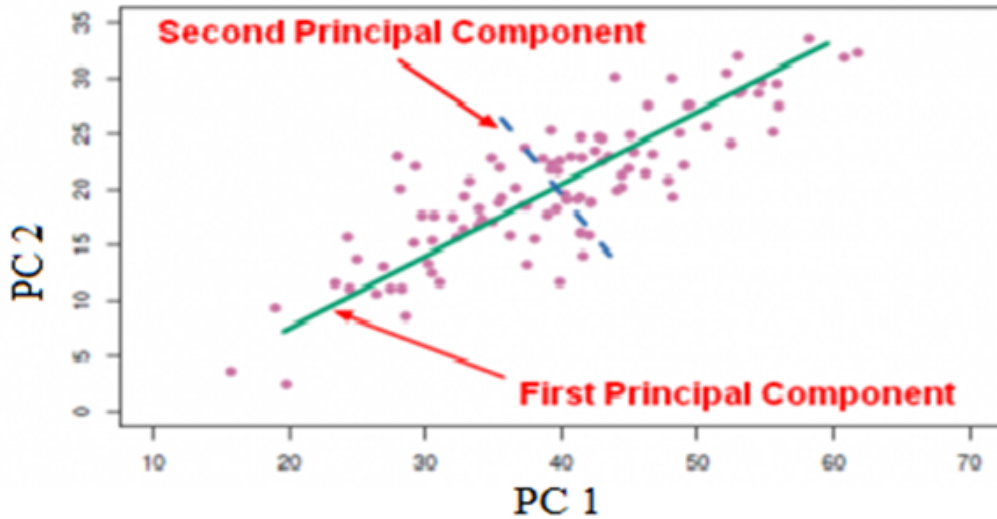


Figure 3.4: Principal Component Analysis

3.2 Proposed Framework

A novel HSI classification framework, as shown in Figure 3.5 is proposed in this section. It has the three components, namely -LPP, SAE and logistic regression. The LPP is employed to remove redundancies and extract useful information from the input HSI. LPP is utilised because of the large information represented by HSI in the form of the continuous bands. The output of LPP is the low dimensional feature set that has the preserved local information of the original HSI. Then from the reduced feature set, 50% data is selected for training. Afterward, from the training data, pixel vectors are extracted, which are passed to the SAEs for the extraction of deep features from the input set. These obtained deep features are passed to the SoftMax layer for supervised learning. The trained model is provided with the complete dataset, and the predicted classes are noted, which is compared

with ground truth for calculation of the overall accuracy. The datasets employed for validation are Indian Pines and Salinas.

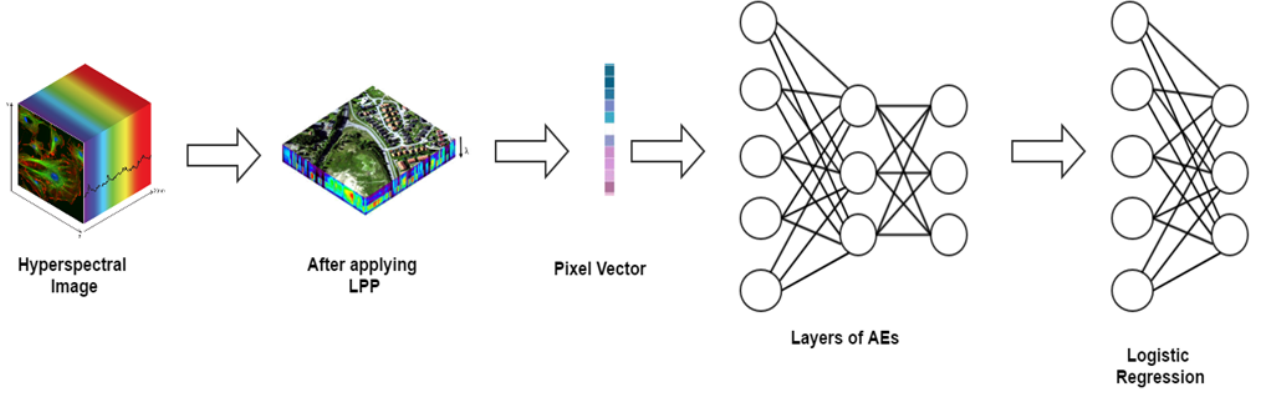


Figure 3.5: Proposed framework for classification

3.3 Experimental Results and Analysis

Real HSIs are considered for performing experiments whose description is given below.

3.3.1 Description of the Datasets

3.3.1.1 Indian Pines Dataset

Indian Pines is acquired *via* AVIRIS sensor which captures high-resolution image with 224 bands. AVIRIS has a group of 4 spectrometers, able to handle 677 pixels in 224 bands in size. It acquires an HSI of 20m or $2e+7$ microns in pixel dimension. This HSI is acquired from NASA ER-2, that has 20 km of altitude and 20 m of spatial resolution, meaning 20×20 m surface area is represented per pixel. The image size is 145×145 pixels of 224 bands, meaning HSI has a complete resolution of $145 \times 145 \times 224$. It includes all areas of agriculture, woodland and other natural vegetation. Soya is sowed in farmland with coverage of roughly less than 5% cover.

Indian Pines dataset constitutes of total 16 classes as shown in Figure 3.6. The 224 bands of the AVIRIS sensor are reduced for more accurate performance. Due to the atmospheric errors bands [220], [104-108], [150-163] have been eliminated. It contains variety of classes that are suitable for DNN training. It consists of 21025 total pixels including the background pixel.

3.3.1.2 Salinas Dataset

Salinas dataset is acquired *via* AVIRIS. This HSI is acquired from the Valley, situated in California, as shown in Figure 3.7. The resolution is high at 3.7m, and the complete resolution of the salinas is $512 \times 217 \times 224$. Due to the atmospheric errors bands [108-112], [154-167], 224 have been eliminated.

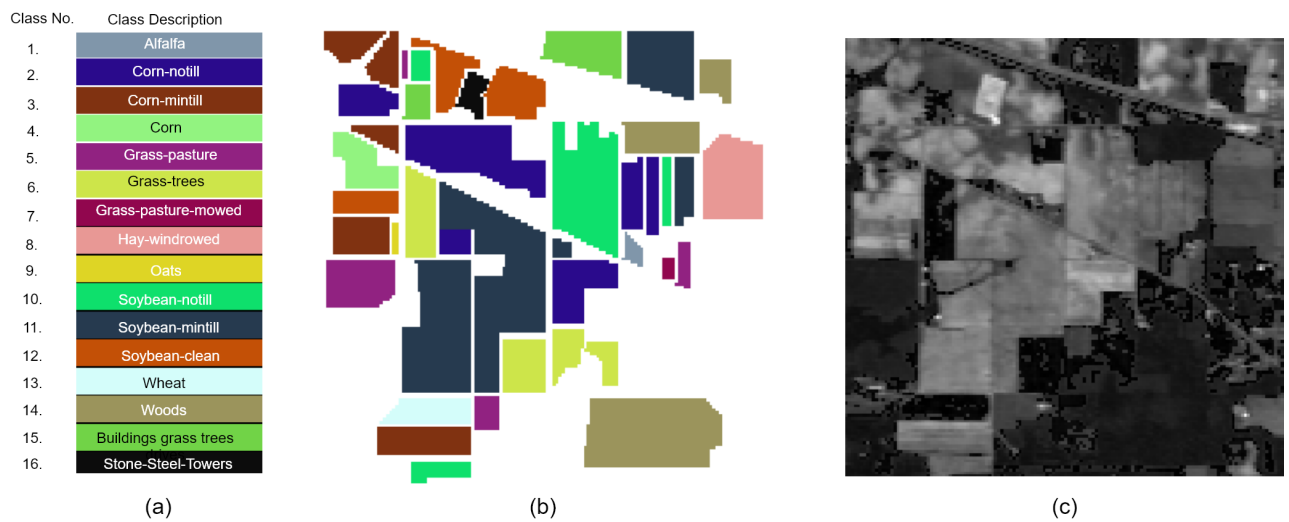


Figure 3.6: Indian pines (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13

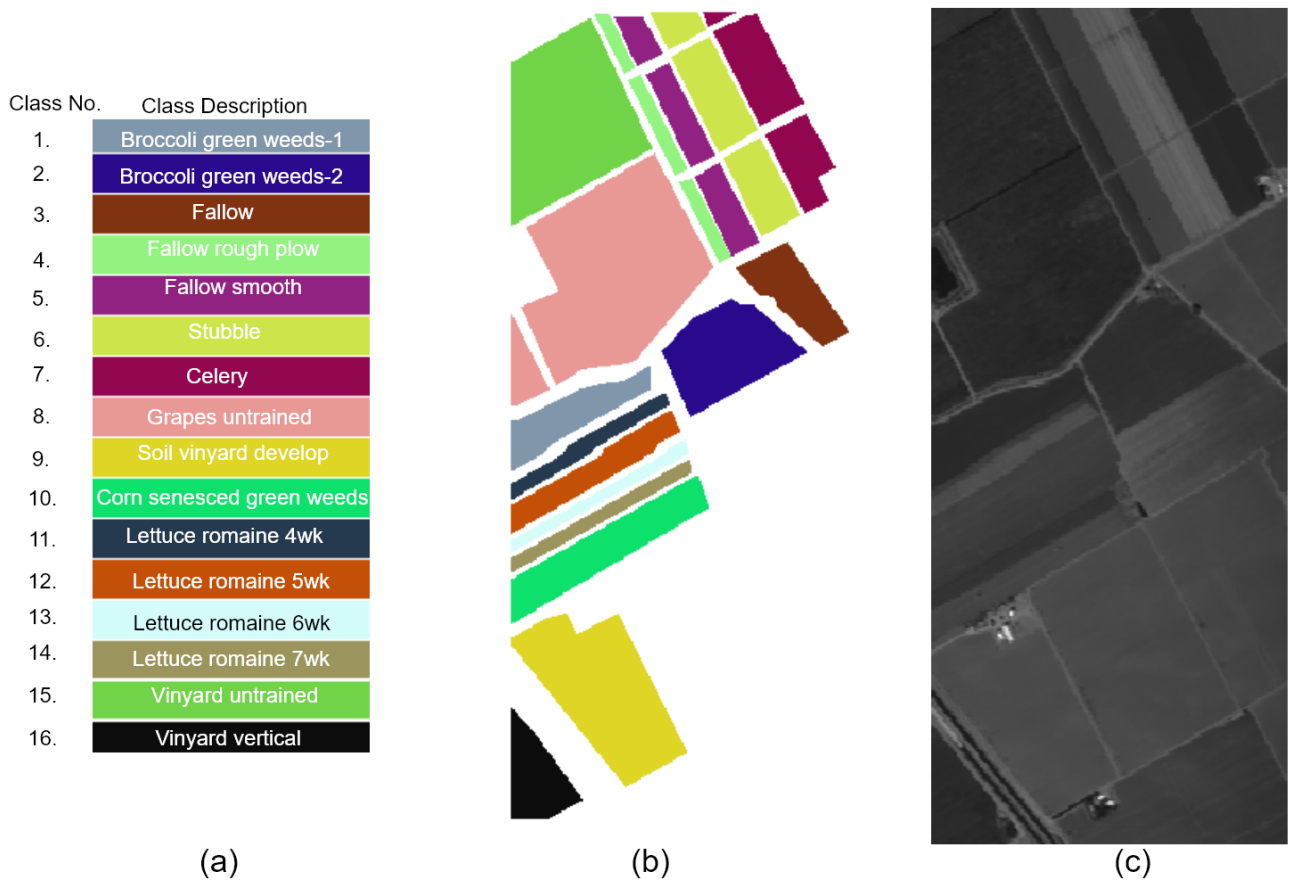


Figure 3.7: Salinas dataset (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13

The band count decreased to 204 after elimination of noisy and useless bands. It constitutes of a total of 16 classes which contains a variety of trees, bare soils and different fields of a vineyard.

3.3.2 Experimental Results

Experiments are performed on the real HSIs. Due to high dimensional features present in HSIs, these are passed to LPP for reducing the redundant dimensions. LPP decrease the band counts to 100 from 200, 100 from 204 in case of Indian Pines and Salinas respectively. Afterward, from the reduced HSIs, pixels are extracted in which every pixel constitutes of 100 values. Then the data is partitioned into training and testing. 50% training data is given as input to SAEs. SAEs extract the deep features with the help of two autoencoder layers. These features are then passed to the SoftMax layer (logistic regression) for supervised training. Afterward, a trained classifier is extracted, which is able to classify the pixels into 16 different classes. Finally, the remaining 50% testing data is passed to the trained classifier, which predicts the corresponding classes. The predicted readings are compared against the actual readings for evaluating the discriminating strength of the trained model.

The LPP is first supplied with datasets that decrease the surplus dimensions. The neighbourhood pixel size and the weight width are set as 10×10 and 1 respectively, and the weight parameter employed is 'adjacency'. LPP outputs 100 dimensions when both the datasets are passed to it with the output values is of double precision type. Afterward, the decreased features set obtained from LPP is transformed into a pixel vector. This vector is given as input to sparse autoencoders with a weight regularisation set at 0.001, a sparsity regularisation to 4 and the learning rate at 0.005 with 1000 epochs. The autoencoders are then stacked together into stacked autoencoders, which extracts the deep features using unsupervised learning. Afterward, 50% of these deep features are passed to SoftMax layer for supervised training with 5000 epochs.

The resultant images are shown in Figures 3.8, 3.9. The overall classification accuracy is contrasted to other twenty-one machine learning models, as shown in Tables 3.1, 3.2. For Indian Pines and Salinas, the proposed framework produces 84.2 and 87.2% overall classification accuracy, respectively, and it also outperforms all machine learning models. In Indian Pines dataset, Cubic SVM is nearest to the DNN with 83.1% classification accuracy. In the same way, Ensemble (Bagged Tree) is closest to DNN with 86.5% accuracy in case of Salinas. DNN performed better as it can able to extract deep features from the datasets. The comparison is also made with Ensembling methods and widely used KNN techniques. Ensembling combines weak classifier together to increase classification accuracy. In KNN, the distance metric used is 'euclidean'. To have a fair evaluation, all the 22 models are given the output obtained from LPP. Simple Tree performed worst with only 35.7% and 31.7% accuracy in case of Indian pines and Salinas, respectively. Decision Tree cannot be able to form meaningful, decisive points due to the high dimensional dataset. The maximum class accuracy achieved by the DNN is 99.5% in class 9 of Indian Pines, 99.5% in class 14 of Salinas. The proposed DNN improved the maximum classification accuracy of 236.41% and 275.08% in cases of

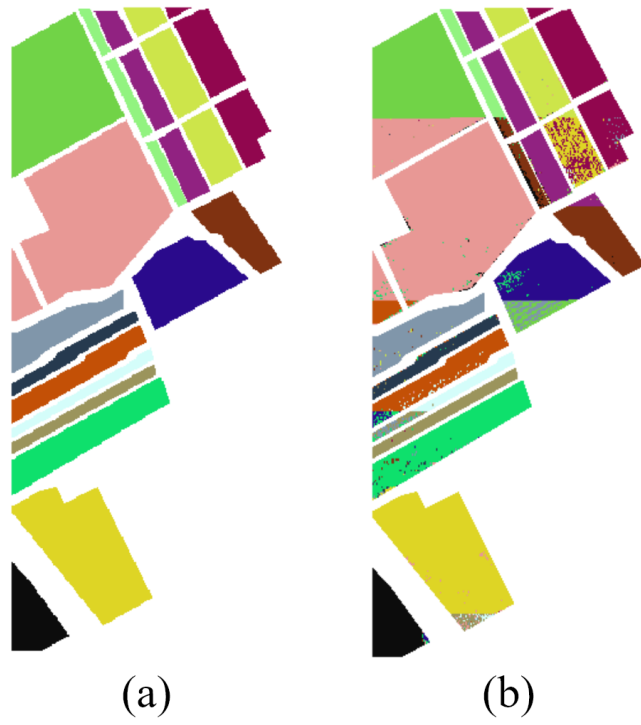


Figure 3.8: Salinas classification map a) Original HSI b) Proposed framework

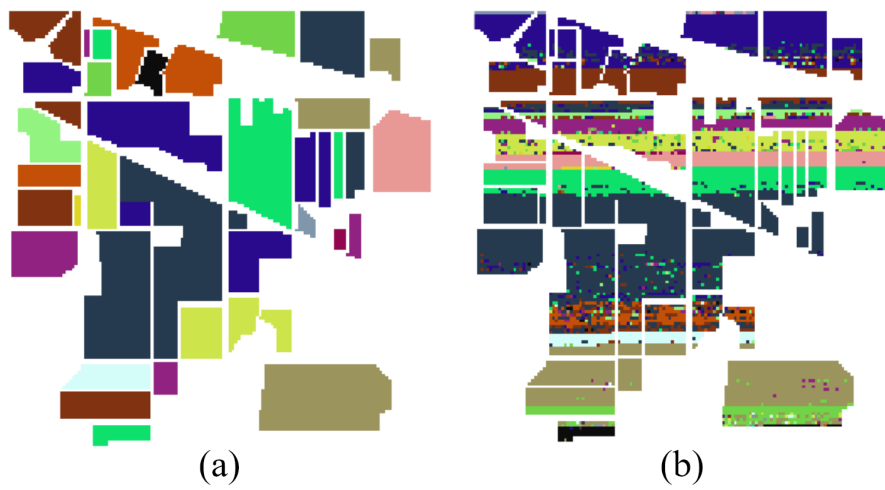


Figure 3.9: Indian Pines classification map a) Original HSI b) Proposed framework

Indian Pines and Salinas, respectively. Earlier works reveal that the use of deep features using autoencoders (Li et al. 2016c; Özdemir et al. 2014) improves the performance of the learning models. In proposed work, spectral characteristics are employed in combination with Stacked Autoencoders are used. Li et al. (2016c) used Stacked Autoencoders with 95% training samples and produced the 78.99% classification rate. The proposed approach used just 50% training data and still produces maximum classification accuracy of 87.2% and lower error rates.

Table 3.1: Class wise prediction and Overall accuracies of 22 learning Models (Indian Pines)

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	OA(%)
Train	23	743	500	111	271	80	13	221	10	527	1500	59	105	633	183	37	
Test	46	1428	830	237	483	730	28	478	20	972	2455	593	205	1265	386	93	
DNN	33	1200	591	155	451	586	21	476	16	830	2279	222	191	1238	276	88	84.4
SVM(CU)	29	1145	581	139	448	502	19	478	14	793	2345	162	191	1262	327	85	83.1
SVM(Q)	28	1140	581	146	460	507	19	477	14	755	2339	185	192	1262	316	87	83
En(BA)	28	1058	547	119	439	500	18	475	12	766	2380	59	193	1263	245	89	79.9
En(SK)	29	1013	561	131	422	563	23	438	16	719	2278	88	190	1239	242	85	78.4
SVM(MG)	22	1121	562	128	435	412	14	471	7	708	2306	65	172	1239	290	72	78.3
SVM(L)	18	1066	563	119	454	501	18	478	4	649	2179	129	192	1250	272	82	77.8
KNN(F)	30	1003	556	136	417	547	24	428	15	744	2129	102	190	1212	249	84	76.7
KNN(W)	25	1020	549	131	419	533	19	450	11	702	2165	61	192	1228	235	81	76.3
Dis(L)	39	991	549	119	422	566	28	425	15	538	1931	271	195	1195	185	83	73.7
SVM(CG)	0	1071	504	38	392	347	0	476	0	358	2208	0	183	1255	212	55	69.3
En(SD)	31	997	489	67	372	560	27	400	15	406	1978	63	198	1203	179	83	69
KNN(COS)	9	933	425	69	421	517	17	452	0	556	1810	0	191	1188	138	75	66.4
KNN(M)	3	934	398	60	382	521	11	442	3	513	1867	1	189	1176	121	50	65.1
Tree(C)	30	863	374	80	395	455	10	431	0	437	1928	0	196	1214	214	0	64.7
KNN(CU)	3	932	384	45	378	468	4	402	0	488	1800	0	188	1158	116	46	62.5
En(BT)	0	953	385	34	400	403	0	451	0	208	1688	0	202	1247	113	0	59.4
Tree(M)	0	926	338	0	362	498	0	444	0	0	1738	0	190	1198	104	0	56.6
En(RUS)	39	585	450	82	340	443	23	354	19	703	834	302	203	1150	127	88	56
SVM(FG)	0	743	512	39	272	0	0	221	0	527	2455	0	26	633	183	0	54.7
KNN(CO)	0	930	342	1	295	48	0	158	0	223	1966	0	174	1104	24	0	51.4
Tree(S)	0	1241	0	0	278	0	0	470	0	0	490	0	0	1181	0	0	35.7

S-simple, Q-Quadratic, SK-Subspace KNN, MG-Medium Gaussian, L-Linear, F-Fine, CG-Coarse Gaussian, C-Complex, En-Ensemble, Dis-Discriminant, RUS-RUS Boosted, FG-Fine Gaussian, BA-Bagged Trees, SD-Subspace Discriminant, W-Weighted, CO-Coarse, BT-Boosted Trees, M-Medium, COS-Cosine, CU-Cubic.

Table 3.2: Class wise prediction and Overall accuracies of 22 learning Models (Salinas)

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	OA(%)
Training	749	3000	1189	695	1814	999	999	5362	2915	1596	109	197	354	678	3287	973	
Testing	2009	3726	1976	1394	2678	3959	3579	11271	6203	3278	1068	1927	916	1070	7268	1807	
DNN	1977	2857	1817	765	2641	2454	3510	10808	5900	3138	1020	1713	629	1065	5126	1775	87.2
En(BA)	2004	2940	1819	765	2647	2460	3457	10887	5406	3217	935	1711	630	1066	5123	1766	86.5
SVM(L)	1973	2811	1819	765	2633	2407	3566	10919	4598	3200	784	1720	628	999	5081	1743	84.3
KNN(COS)	2005	2734	1819	765	2650	2447	3571	10965	4039	3236	908	1756	631	1030	5124	1742	83.9
SVM(CU)	2005	2720	1819	765	2653	2448	3570	10962	3667	3234	873	1744	631	1012	5124	1667	82.9
Tree(C)	2002	2910	1819	765	2645	2451	3266	10868	3708	3170	1002	1643	626	1049	5103	1782	82.8
SVM(MG)	1955	2185	1819	758	2643	2395	2701	11242	5236	3195	574	1215	631	995	5112	1760	82.1
SVM(CG)	1920	2666	1815	762	2638	2396	3551	10937	4724	3117	1	365	628	899	5122	1579	79.6
En(BT)	2009	2903	1819	764	2667	2993	3562	10593	3683	3055	0	0	645	967	5058	1781	78.5
KNN(F)	1680	2598	1688	727	1993	1341	2306	7017	1662	2202	154	119	584	906	4651	1336	57.2
Dis(L)	1379	2637	1819	702	2048	1985	3565	8733	1535	2142	734	862	623	591	4668	1542	65.7
En(RUS)	2005	2889	1817	902	2629	2629	2719	4509	2076	2606	984	1626	663	583	5094	1625	65.3
Tree(M)	2001	2902	1819	1361	2576	2445	3562	4562	3231	2234	0	0	593	996	4962	1450	64.1
EN(SK)	1769	1933	1676	763	2031	1194	2788	7888	2925	1946	137	274	609	959	5341	1005	61.4
En(SD)	1375	2873	1819	677	2044	2054	3489	7728	1426	1825	178	63	597	296	4760	1468	60.4
KNN(W)	1790	2273	1515	759	1961	1096	2645	6454	2928	1826	233	356	625	820	5444	989	58.6
SVM(Q)	1680	2598	1688	727	1993	1341	2306	7017	1662	2202	154	119	587	906	4651	1336	57.2
SVM(FG)	762	1008	1113	583	1838	1008	1002	11271	2916	1600	26	96	531	344	3295	980	52.4
KNN(M)	1769	2312	1170	732	1665	404	2647	5246	431	1323	120	214	607	637	5374	204	45.9
CoarseKNN	1438	2482	676	724	1468	61	2603	5100	272	719	0	13	628	427	5472	35	40.9
KNN(CU)	1491	2033	634	712	1539	264	2323	4137	393	1158	65	123	584	584	5324	224	39.9
Tree(S)	0	2949	0	0	0	0	0	4663	3231	0	0	0	0	0	5126	1215	31.7

3.4 Conclusion of the Chapter

In this chapter, a novel DNN framework is presented. Most of the previous research use PCA for dimension reduction but in this framework, the LPP is used for the first time which reduced the dimensions of an HSI to almost half without losing much information and the previous work on hyperspectral classification (Chen et al. 2015; Yue et al. 2016) employs the HSIs which comprises of significant hyperspectral signatures. In this chapter, the dataset like Indian Pines, which have fewer samples, is also tested to show the robustness of the proposed model. With the utilisation of LPP, the feature set is reduced due to which the DNN processes the data quickly, and the processing time is decreased drastically. The results generated by LPP with the accuracy of around 84.4 and 87.2% which leads to 236.41 and 275.08% improvement in overall accuracy in case of Indian pines and Salinas dataset respectively.

Chapter 4

A Pre-processing Framework for Spectral Classification of Hyperspectral Images

The previous chapter presented an efficient framework for classification, which only uses the extracted features from LPP. This chapter focusses on the pre-processing of the HSIs so that maximum classification results can be obtained. The HSI classification is primarily focused on spectral-spatial features in current methods. The satellite acquired HSIs may have several useless bands because of H_2O absorption. These bands are removed in correction process due to which numerous useful bands are also removed from the image. This chapter presents a pre-processing framework that utilises linear interpolation to deal with the lost bands. In addition to this, hybrid features are designed by utilising PCA and LPP. These hybrid features preserve both the local and global information of the input HSI. Afterward, these features are given as input to the various machine and deep learning models for obtaining classification accuracy. In this chapter, we have extended the testing datasets to four HSIs namely Indian Pines, Salinas, Pavia University, and Kennedy Space Centre to show the effectiveness of the proposed framework.

4.1 Background of the proposed work

4.1.1 Interpolation

Captured raw hyperspectral images from the satellite cannot be directly utilised. It needs corrective measures in the atmosphere, geometry and radiometry readings and many important bands are lost in these operations. The missing *HSI* bands restrict the classifier's discriminatory capability. So, linear interpolation is employed for an approximation of the removed *HSI* bands to overcome this issue. It

The contents of the work presented in this chapter have been published in *Multimedia Tools and Applications*, DOI: <https://doi.org/10.1007/s11042-020-09180-2> (SCI Indexed)

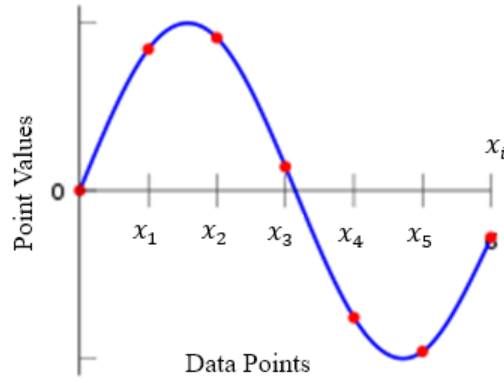


Figure 4.1: Interpolation

offers the approximation of these eliminated bands, as HSI is a sequential image (Mou et al. 2017). Interpolation is the method to generate new data set from the old one. The data set contains the data points to which the corresponding value is mapped. Interpolation can embed new values provided that all the desired data points are between the minimum and the maximum of the available data points. It helps in an approximation of the removed bands of HSIs (Ye et al. 2016).

Suppose the data set contains q_i points that are mapped to w_i values corresponding to the points using the concept of one to one mapping. Insert the desired points s_m such that the values of s lies between $[min(q_i), max(q_i)]$. After the interpolation, the new data set will contain the s more points and values than the previous one. Interpolation can be linear, cubic, spline *etc.* In this chapter, the linear interpolation is used. Interpolation is shown in Figure 4.1.

Before Interpolation- $(q_1, w_1), (q_2, w_2), (q_3, w_3), (q_4, w_4) \dots (q_i, w_i)$

After Interpolation- $(q_1, w_1), (q_2, w_2), (q_3, w_3), (q_4, w_4) \dots (q_{i+s}, w_{i+s})$

4.1.2 Principal Component Analysis and Locality Preserving Projections

HSIs have high dimensional features due to which the spectrum produced by the information becomes very complex to interpret. So, to process the hyperspectral signatures, LPP (He and Niyogi 2004). and PCA can be employed to reduce the redundant dimensions present in the HSIs. Chen (Chen et al. 2015) also utilised PCA for eliminating the excess dimensions and preserving the spatial information. LPP has a capability of retaining the local characteristics with but also its description is not restricted to the specified training points. These points are everywhere, and it is also identified by approaching the Laplacian operator's eigenfunction.

PCA is among the most prominent dimensionality reduction technique used in the creation of prediction models. The usage of it will minimise redundant details from the spectral data and maximise the variance in the resultant output. So, the sequential data may still stay sequential after dimensionality reduction. PCA's primary role is to maximise the variance in the resultant output. With the help of Singular Value Decomposition (SVD) and eigenvalues, vectors principal components of the input

can be calculated.

PCA is employed to map the higher dimensional space $H = H_1, H_2, H_3 \dots H_a$ to lower dimensional space $L = L_1, L_2, L_3 \dots L_b$ where $b < a$. It can be shown as:

$$H^a \rightarrow L^b \quad (4.1)$$

Where H denotes the spectral data with high dimensions denoted by a .

In the first step, the spectral data with higher dimensions (S) is converted to normalised (S) by subtracting it from its mean without reducing the dimensions. Afterward, from the normalised S, covariance matrix is calculated, that is a square symmetric matrix. It measures the changes between two random variables. It can be represented as:

$$C = \begin{bmatrix} Var(a_1) & Covar(a_1, a_2) & \cdots & Covar(a_1, a_k) \\ Covar(a_2, a_1) & Var(a_2) & \cdots & Covar(a_2, a_k) \\ \vdots & \vdots & \ddots & \vdots \\ Covar(a_k, a_1) & Covar(a_k, a_2) & \cdots & Var(a_k) \end{bmatrix}$$

where a_i represents variable and the value of i lies between 1 to k .

Then, the appropriate eigenvalues λ and eigenvectors V are then calculated from the covariance matrix. The eigenvalues and vectors represent the direction of the mapped space, magnitude of the direction respectively. Every eigen vector has the corresponding eigen value and the vectors are placed based on the decreasing values. The first eigen vector represent the maximum variance followed by second and so on. Now, from this placement the first p principal components (eigenvectors) are chosen that represents the maximum variance, which is mapped to L of PCA, as shown in Figure 4.2. L is procured by the following equation:

$$L = V^T \times normalised(S) \quad (4.2)$$

Where T denotes Transpose.

LPP is a linear approach like PCA. LPP output is obtained in the form of non-linear manifold embedding in lower-dimensional space. In the first step, it builds the adjacency graph. In graph G, the edge E is placed vertices p and q if they are considered close. The closeness is calculated by utilising Euclidean norm in the higher dimensional space or by employing the k -nearest neighbor algorithm. Afterward, the weights are allocated to the edges based on the heat kernel. Finally, the eigenvalues and vectors are calculated, which provides the resultant desired HSI output with low dimensions. More details are discussed in chapter 3.

Due to the high number of dimensions, the evaluation of all spectral bands is ineffective as redun-

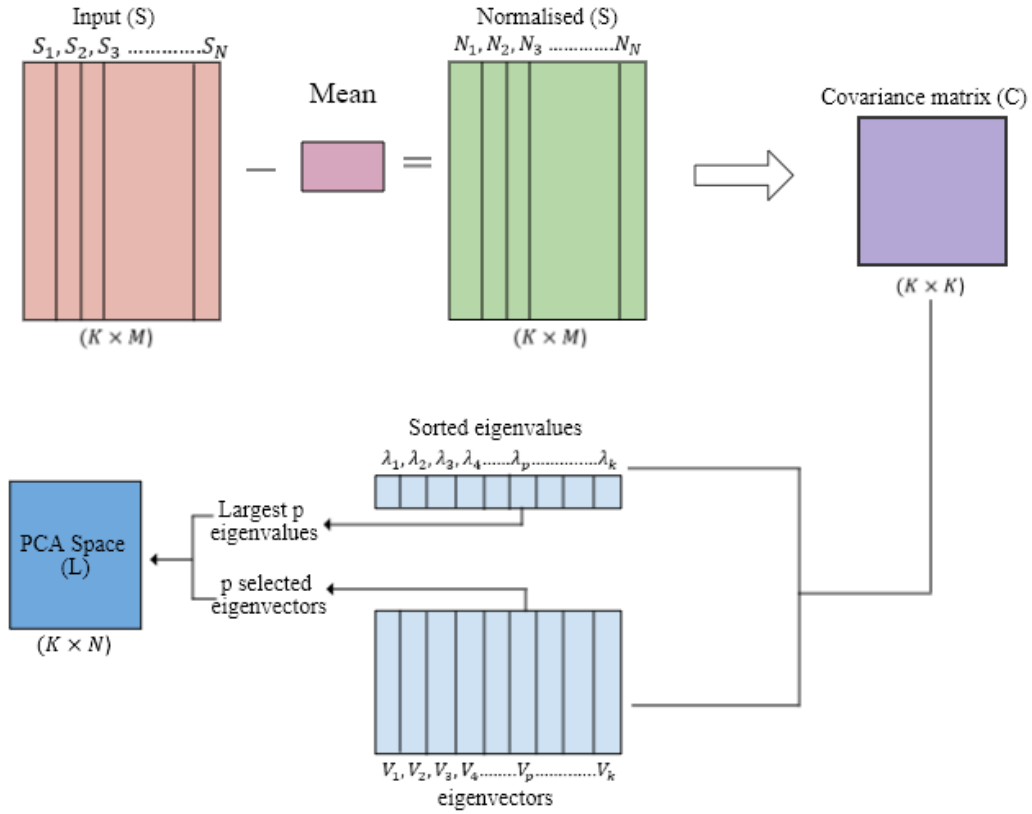


Figure 4.2: Procedure of extracting Principal Components

dancies and associations exist amongst them. This redundancy is eliminated by a PCA and LPP. In reducing the dimensions, both algorithms are effective. Both of these algorithms can extract linear information, and LPP has the added capability to represent data in non-linear forms such as Laplacian Eigenmaps. Both minimise the spectral characteristics without sacrificing a lot of spatial information. The output generated by both of these algorithms is separate from one another. The independent outcomes of PCA and LPP provide outstanding accuracy in classification. However, a mixture of both can contribute to more accuracy in classification than individual usage. Spatial information is effectively maintained by LPP and PCA. Their resultant outputs are blended together to design hybrid features. The integrated features boost the overall accuracy of different classifiers.

Table 4.1: Correlation between LPP and PCA

Dataset	Pearson coefficient between LPP and PCA
PaviaU	0.0037
Indian Pines	-0.0013
KSC	-0.0051
Salinas	0.0064

Assume the PCA reduced features are $P1$, which includes lower dimensions having maximum vari-

ance. $P2$ is obtained from LPP in the same manner. Both $P1$ and $P2$ are placed together to create a new matrix. Table 4.1 displays the correlation coefficient (r) between them. It suggests that there is no correlation between the features obtained via PCA and LPP. Thus, these are merged to maximise the classifier's discrimination capability. In the end, the resultant combined features having both the local and global approximation are given to various learners for evaluating the overall classification accuracy.

4.2 Proposed Preprocessing Framework

The pre-processing framework for HSI classification is proposed in this section, as shown in Figure 4.3. Firstly, the acquired HSI is transformed into a 2-D array. The acquired HSI contains radiometric, geometric and atmospheric errors due to which some bands become noisy and useless. After applying the suitable corrections, many useful bands are also lost. So, to procure the omitted information, we employed linear interpolation as HSI is sequential in nature. It provides the approximation of the missing bands, which is close to the missing values. Afterward, the processed data is passed to effective dimensionality reduction algorithms named PCA and LPP. The resultant output from both of these is combined to form one single array, which constitutes 'hybrid' features. These hybrid features have the local and global approximation of the input HSI. Afterward, it is partitioned into a training and testing set. Then, the pixel vectors from the training set is passed to the machine and deep learning classifiers for supervised training. Then, the testing data is passed to these trained classifiers for prediction. In the end, the predicted output is compared against the actual output for evaluating the model.

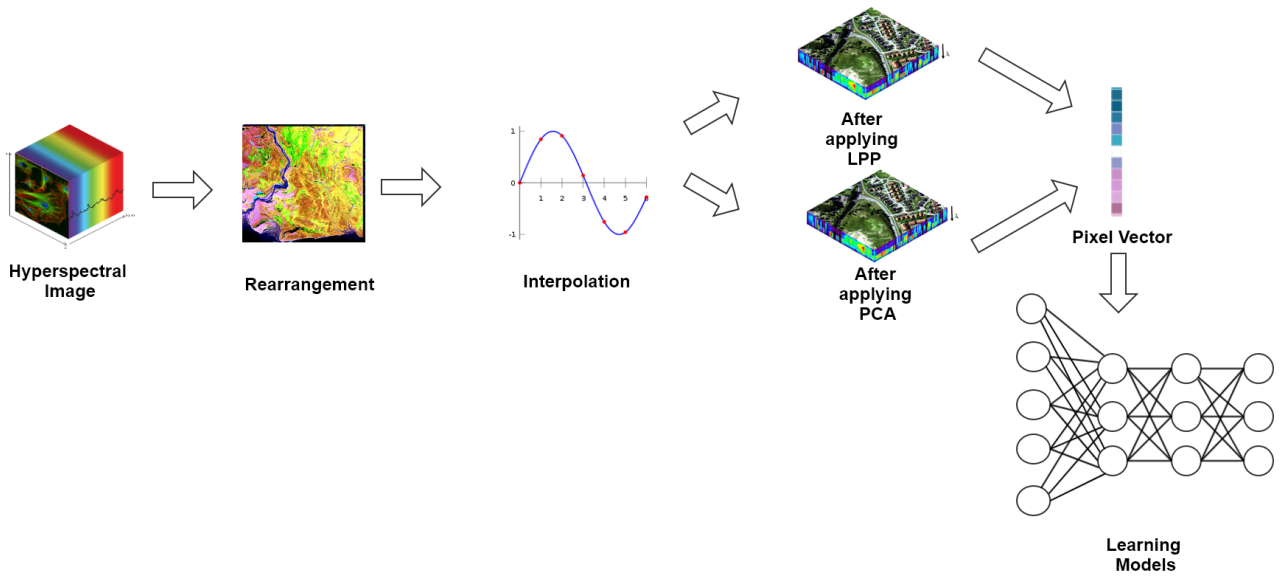


Figure 4.3: Proposed Preprocessing Framework for Classification

Algorithm Classification with the Proposed framework

Input-An HSI image H_n^m which has total n pixels, each of m dimensions. Each pixel belongs to class $C=1, 2, 3, \dots C_i$.

Output-Predicted Classes $C=1, 2, 3, \dots C_i$.

- 1: Convert H_i^k to array of $i \times k$ by reshaping.
 - 2: **For** iteration =1 to x_q [x_q are the desired points]
 - 3: F =Interpolation at each row
 - 4: **endFor**
 - 5: H =Evaluated PCA on F (no of neighbours= j , no. of components= n)
 - 6: G =Perform LPP on F (weight width=1.0, no. of neighbours= j , no. of components= m)
 - 7: **For** iteration=1 : no. of rows(G)
 - 8: $J(\text{iteration},:)$ =Add $H(\text{iteration},:)$ to the tail of $G(\text{iteration}, :)$.
 - 9: **endFor**
 - 10: Prepare training data from J and store it in I .
 - 11: Give pixel vector from I as input to various machine learning models
 - 12: Now associate the respective class $C=1, 2, 3, \dots C_i$ with every pixel vector for supervised training.
 - 13: Trained classifiers TC_i is obtained.
 - 14: Input H to TC_i which predicts the respective class of each pixel vector.
-

4.3 Experimental Results and Analysis

Real HSIs are considered for performing experiments on a device with core i5-5200U processor and 12 GB of RAM in Matlab version 2018a. In these experiments, the data sets involved are KSC, Salinas and Indian Pines, which are acquired by the AVIRIS. Pavia University dataset acquisition is performed by a separate sensor named Reflective Optics System Imaging Spectrometer (ROSIS). AVIRIS captures the reflectance values ranging from 430 to 860nm in the electromagnetic spectrum, which is spread across 224 bands. It has a group of 4 spectrometers, able to handle 677 pixels in 224 bands in size. It captures an image of 20m or 2e+7 microns in pixel dimension. It can yield interesting details about several minerals in the present on various surfaces. ROSIS sensor captures the reflectance value ranging from 430-860nm in the electromagnetic spectrum, which is spread across 115 bands. It is employed specifically for water structure identification. The proposed framework results are compared against spectral and spatial-spectral frameworks as shown in Tables 4.3, 4.4, 4.5, and 4.6. The spectral characteristics present are the atmospherically corrected features that are procured from the HSI. The spectral features are selected from the dataset directly. For obtaining the spatial feature, PCA is employed first to reduce the dimensions. Afterward, a $f \times f$ filter is placed on the reduced dimension set for obtaining the neighbourhood information of pixels. In the end, the spatial features are combined with spectral features to build the spectral-spatial framework.

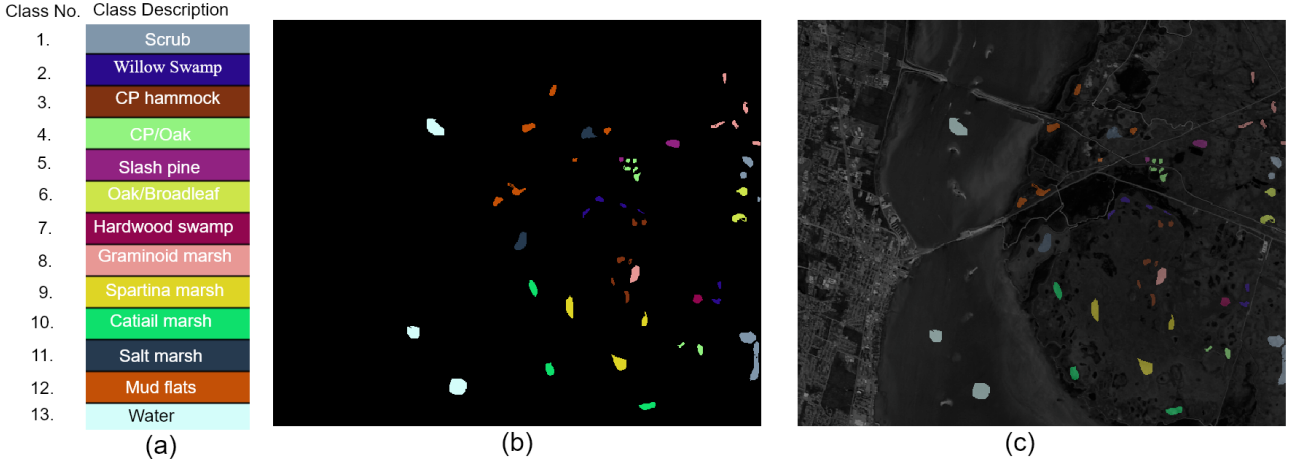


Figure 4.4: KSC dataset (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13

4.3.1 Description of Datasets

4.3.1.1 Indian Pines Dataset

The Indian Pines is based in North-western Indiana. The scene is acquired *via* AVIRIS sensor and complete resolution of the scene is $145 \times 145 \times 224$. It constitutes 16 different classes which cover the agricultural area, woodland and other natural vegetation etc. as shown in Figure 3.6. Due to atmospheric errors, many bands are eliminated after correction. So, the complete resolution of the scene after correction is $145 \times 145 \times 220$. More details of this dataset are discussed in chapter 3.

4.3.1.2 Salinas Dataset

Salinas scene is based in Salinas Valley. The scene is acquired *via* AVIRIS sensor and complete resolution of the scene is $512 \times 217 \times 224$. It constitutes 16 different classes which cover different vegetables and other natural vegetation etc. as shown in Figure 3.7. Due to atmospheric errors, many bands are eliminated after correction. So, the complete resolution of the scene after correction is $512 \times 217 \times 204$. More details of this dataset are discussed in chapter 3.

4.3.1.3 Kennedy Space Center Dataset

The Kennedy Space Centre (KSC) scene is based in Florida. The scene is acquired *via* AVIRIS sensor and complete resolution of the scene is $512 \times 614 \times 224$. It constitutes 13 different classes which cover different marsh, water and other natural vegetation etc. as shown in Figure 4.4. Due to atmospheric errors, many bands are eliminated after correction. So, the complete resolution of the scene after correction is $512 \times 614 \times 176$.

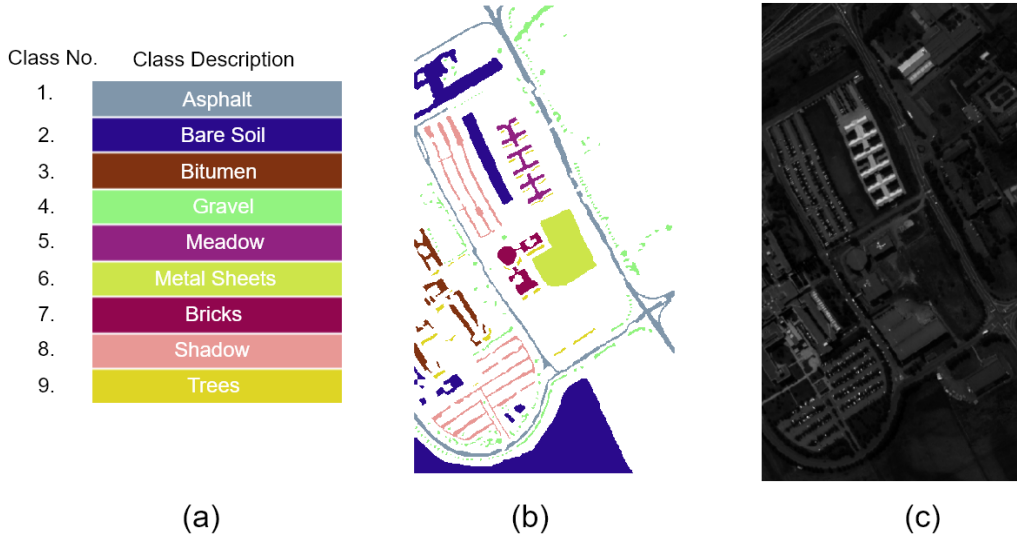


Figure 4.5: Pavia University dataset (a) Ground Truth Classes (b) Original HSI (c) Spectral band 13

4.3.1.4 Pavia University Dataset

Pavia University scene is based in Italy. The scene is acquired *via* ROSIS sensor and complete resolution of the scene is $610 \times 610 \times 115$. It constitutes 9 different classes which cover asphalt, soil, other natural vegetation etc. as shown in Figure 4.5. Due to atmospheric errors, many bands are eliminated after correction. So, the complete resolution of the scene after correction is $610 \times 340 \times 103$.

4.3.2 Experimental Results

The n -dimensional HSIs are first transformed into 2-D array by reorganising the pixels. An HSI is denoted as H_n^m which has total n points, each having m features/dimensions. A single-pixel have m dimensions spread in the z -axis, that is regarded as 1 HSI point. Now, every HSI point is taken from the z -axis and placed it in the x -axis in the form of rows. At the end of the row, the corresponding class value is added. So, for n HSI points, n rows with m columns are produced in 2-Dimensional. Then, linear interpolation is employed for adding the approximation of eliminated bands, that are removed during the correction process. Afterwards, PCA and LPP are employed to decrease the hyperspectral feature set to x and y respectively. Then both the reduced feature set obtained from PCA and LPP is combined ($x + y$) to form ‘hybrid’ feature set. These features contain the global and local features of the input dataset.

Hybrid features developed using PCA and LPP are the key parameters. Individually, these can considerably boost the differentiation ability of learning models, and the combination of these can enhance the model’s selective capability even further due to the presence of local and global information in hybrid features. These features are configured so that the size of these are lower than the input HSI features to reduce the processing time. Other feature set from spectral and spectral-spatial framework have more features than these hybrid features, due to which the proposed framework needs

less time for training. The designed hybrid features are chosen from LPP and PCA. PCA processed feature set preserves the spatial information and provides a global approximation having a maximum variance of the input. In most cases, 20 to 70 principal components are chosen with variance ranging from 99.72% to 100%. Similarly, LPP processed feature set preserve the locality information of the pixel and provides local approximation based on the eigenvalues. The vectors are selected in a way, so the feature set has minimum of aggregated eigenvalues. The output hybrid spectrum incorporates neighborhood and global approximation, thus increasing the discriminating capability of employed ML and DL models.

Table 4.2: Description of train and test set samples for HSIs

	Pavia University		Salinas		Indian Pines		KSC	
Class No.	Train set	Test set	Train set	Test set	Train set	Test set	Train set	Test set
1	3000	6631	749	2009	23	46	350	761
2	9000	18649	3000	3726	743	1428	100	243
3	1000	2099	1189	1976	500	830	125	256
4	1500	3064	695	1394	111	237	120	252
5	600	1345	1814	2678	271	483	80	161
6	2600	5029	999	3959	80	730	120	229
7	800	1330	999	3579	13	28	55	105
8	1800	3682	5362	11271	221	478	230	431
9	500	947	2915	6203	10	20	250	520
10	–	–	1596	3278	527	972	202	404
11	–	–	109	1068	1500	2455	200	419
12	–	–	197	1927	59	593	250	503
13	–	–	354	916	105	205	500	927
14	–	–	678	1070	633	1265	–	–
15	–	–	3287	7268	183	386	–	–
16	–	–	973	1807	37	93	–	–

Two autoencoders are employed in a deep network, which are then stacked into Stacked Autoencoders (SAEs), that extract deep features efficiently. These are then passed for classification to the SoftMax layer. The parameters of the learning models used in the experiments are set according to the standard. As HSI classes are non-uniform, so for every class, 50% samples are utilised for training and the testing is performed on 100% samples as shown in Table 4.2. Training data for supervised learning were given to learning models. These models have been trained for each dataset. After supervised learning, trained models are obtained. These trained models are given testing data for predicting the corresponding classes. The predicted output of these trained models is compared against the actual class value for evaluating the robustness of the trained models. Evaluation is based on overall accuracy and total time needed for training the prediction model. The obtained results from the proposed framework are contrasted with existing spectral-spatial and spectral frameworks.

Table 4.3: Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (Indian Pines)

Indian Pines							
Sr.No	Classifiers	OA-PF (%)	TT (sec)	OA-S (%)	TT (sec)	OA-S.S (%)	TT (sec)
1	DN-SAEs	88.9	621.8	80.7	2872.8	80.4	2390
2	Dis(L)	87.8	3.1	75.6	3.6	76.9	3.7
3	SVM(Q)	87	18.3	81.2	28	82.2	37.4
4	SVM(CU)	86.9	30.6	83.2	47.9	82.9	38.5
5	SVM(L)	85.5	15.8	75.8	22.3	77.7	20.6
6	SVM(MG)	82.7	41.8	76.9	60.2	78.6	68.4
7	EN (SD)	80.3	63.3	75	100.8	76.3	110.8
8	SVM(CG)	77.9	51.7	59.8	79.7	62.4	73.7
9	KNN(Cosine)	74.7	50.5	57	77.2	70.2	84.3
10	EN (Boost)	62.9	94.6	60.8	124.3	62.7	134.4

OA-PF–Overall Accuracy with Proposed Framework, **TT**–Training Time **OA-S**–Overall Accuracy with Spectral features, **OA-S.S**–overall Accuracy with Spectral-Spatial Features, **DN-SAEs**-Deep Network using Stacked Autoencoders, **CU**-Cubic, **Q**-Quadratic, **MG**-Medium Gaussian, **L**-Linear, **Dis**-Discriminant, **CG**-Coarse Gaussian, **EN**-Ensemble, **SD**-Subspace Discriminant, **BT**-Boosted Trees.

Table 4.4: Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (Salinas)

Salinas							
Sr.No	Classifiers	OA-PF (%)	TT (sec)	OA-S (%)	TT (sec)	OA-S.S (%)	TT (sec)
1	SVM(CU)	95.8	110.9	58	140.5	59	206.3
2	SVM(Q)	95.3	93.8	54.5	95.1	56.2	137.2
3	SVM(L)	93.7	68.7	53.9	72.5	55.6	71.5
4	SVM(MG)	93.5	188	49.6	188.5	49.4	289.6
5	DN-SAEs	93.3	3095.4	52.5	6436.3	57.8	6552
6	SVM(CG)	93.3	116.2	47	234.4	47.6	362.4
7	Dis(L)	92.4	8.2	52.8	15.1	53.8	14.9
8	EN (SD)	90.6	58.9	52.5	411.9	53.7	1003
9	KNN(Cosine)	89.9	204.3	46.2	423.3	52	817.8
10	EN (Boost)	86.5	188	41.3	312.9	43.8	905.7

Table 4.5: Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (KSC)

KSC							
Sr.No	Classifiers	OA-PF (%)	TT (sec)	OA-S (%)	TT (sec)	OA-S.S (%)	TT (sec)
1	Dis(L)	94.8	1.9	89.5	2.9	88.6	2.5
2	SVM(L)	92.7	7.6	87.2	8.9	87.3	7.9
3	SVM(Q)	92	7.8	87.6	9.1	87.8	10.3
4	SVM(CU)	92	13	87.8	13.9	87.7	14.9
5	EN (SD)	91.3	22.2	89.2	29.2	89.2	33.4
6	DN-SAEs	91	247.3	82.6	1193.3	88.4	1398
7	KNN(Cosine)	89.9	14.7	80	18	78	21.3
8	EN (Boost)	89.8	30.1	79.9	30.8	80.8	38.7
9	SVM(CG)	89.6	16.8	82.1	20	82.1	22.2
10	SVM(MG)	83.1	13.1	78	16	78.7	17.1

Table 4.6: Performance comparison of PF with other learning models based on overall accuracy with corresponding training time (Pavia University)

Pavia University							
Sr.No	Classifiers	OA-PF (%)	TT (sec)	OA-S (%)	TT (sec)	OA-S.S (%)	TT (sec)
1	SVM(MG)	94.8	93.6	86.9	154.27	89.4	194.4
2	SVM(CU)	94.7	78.6	90.1	225.9	91.4	133.2
3	SVM(Q)	94.1	44	91.4	106.4	93.2	73
4	SVM(CG)	92.1	85.8	83.3	222.56	85.4	283.3
5	DN-SAEs	91.9	914.3	86.5	1046.6	89.8	1492
6	SVM(L)	91.5	31	88.6	65.7	90.7	70
7	Dis(L)	88.4	3.4	84.9	5.8	86.1	4.6
8	EN (Boost)	87.6	163.2	73.6	307.73	75.8	200.8
9	EN (SD)	87.5	146.8	84.5	354.34	87.3	212.8
10	KNN(Cosine)	83.5	129.7	77.9	393.9	79.6	224.3

Table 4.7: Class wise prediction results after applying the Proposed Framework (Indian Pines)

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	OA(%)
DN-SAEs	45	1332	579	169	474	681	22	472	18	884	2410	170	180	1257	334	80	88.9
Dis(L)	44	1348	669	153	450	643	28	466	20	784	2346	266	197	1221	290	82	87.9
SVM(Q)	37	1275	653	160	477	600	22	478	17	848	2396	118	188	1258	308	72	86.9
SVM(CU)	38	1264	653	160	475	590	21	478	18	865	2395	114	189	1258	319	71	86.9
SVM(L)	33	1299	634	149	471	623	20	476	6	775	2381	90	193	1252	291	70	85.5
SVM(MG)	27	1255	644	132	460	461	16	446	6	793	2408	85	173	1244	261	61	82.7
EN (SD)	43	1286	610	87	420	573	27	450	15	543	2344	177	191	1200	185	80	80.3
SVM(CG)	5	1284	605	38	436	461	0	478	0	542	2429	0	184	1256	204	67	77.9
KNN(Cosine)	18	1124	488	58	382	513	16	446	0	640	2293	62	181	1227	126	82	74.7
EN (Boost)	16	928	320	8	364	359	0	475	0	183	2221	0	170	1244	91	0	62.2

Table 4.8: Class wise prediction results after applying the Proposed Framework (Salinas)

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	OA(%)
SVM(CU)	2007	3726	1976	1374	2669	3955	3578	11188	6203	3237	976	1910	913	1070	5289	1784	95.8
SVM(Q)	2007	3726	1976	1374	2669	3958	3578	11081	6202	3238	966	1908	909	1070	5132	1786	95.3
SVM(L)	2007	3726	1976	1373	2670	3958	3578	10841	6203	3235	879	1889	904	1057	4626	1788	93.7
SVM(MG)	2007	3726	1976	1373	2670	3958	3578	10841	6203	3235	879	1889	904	1057	4626	1788	93.7
DN-SAEs	2005	3726	1928	1385	2653	3954	3571	10755	6179	3222	994	1698	896	1054	4689	1777	93.3
SVM(CG)	2005	3726	1976	1351	2665	3951	3573	10882	6203	3225	851	1839	892	1052	4522	1794	93.3
Dis(L)	2008	3725	1889	1374	2646	3943	3568	10513	6111	3175	982	1887	905	1012	4489	1807	92.4
EN (SD)	2006	3725	1781	1348	2644	3945	3568	10099	6203	3114	995	1724	899	1000	4219	1792	90.6
KNN(Cosine)	2007	3720	1974	1375	2620	3929	3575	10270	6163	3126	1008	1188	846	1026	4037	1792	89.9
EN (Boost)	2003	3710	1896	1389	2453	3917	3568	10427	6116	2737	644	0	827	1037	4303	1768	86.5

Table 4.9: Class wise prediction results after applying the Proposed Framework (KSC)

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	OA (%)
Dis(L)	750	242	245	225	128	219	104	273	520	398	414	497	926	94.8
SVM(L)	752	181	243	217	126	202	103	274	520	391	406	491	925	92.7
SVM(Q)	755	175	238	208	135	195	99	274	516	392	408	481	918	92
SVM(CU)	754	169	240	200	131	193	99	279	518	392	408	483	926	92
EN (SD)	749	211	202	181	108	208	103	257	516	395	411	491	926	91.3
DN-SAEs	750	152	247	191	130	224	103	289	518	389	407	462	880	91
KNN(Cosine)	744	181	242	165	102	191	100	279	490	384	418	483	908	89.9
EN (Boost)	726	215	231	139	106	200	103	260	514	391	415	473	907	89.8
SVM(CG)	739	193	224	178	101	196	88	260	519	388	394	479	908	89.6
SVM(MG)	752	193	242	202	119	190	82	268	438	393	415	489	545	83.1

Table 4.10: Class wise prediction results after applying the Proposed Framework (Pavia)

Class	1	2	3	4	5	6	7	8	9	OA (%)
SVM(MG)	6311	18525	1119	2843	1343	4834	1206	3481	883	94.8
SVM(CU)	5969	18514	1241	2893	1339	4866	1258	3475	947	94.7
SVM(Q)	5815	18529	1148	2909	1342	4876	1227	3476	946	94.1
SVM(CG)	6267	18343	899	2796	1326	4638	778	3424	942	92.1
DN-SAEs	5661	18532	1245	2890	1341	4893	1273	3481	0	91.9
SVM(L)	5452	18409	993	2829	1340	4815	851	3492	946	91.5
Dis(L)	5358	18549	958	1933	1337	4667	756	3309	944	88.4
EN (Boost)	5740	16930	941	2633	1327	4499	1124	3325	943	87.6
EN (SD)	5964	18058	918	2776	1337	3871	518	3027	944	87.5
KNN(Cosine)	4816	18553	1105	2450	1343	3939	811	1749	942	83.5

The proposed preprocessing framework results are compared with spectral and spectral-spatial frameworks as shown in Tables 4.3, 4.4, 4.5 and 4.6. By employing the proposed framework, hybrid features are extracted, and these features are passed to various other ML and DL based classifiers for evaluating the overall accuracy and time required for training. It can be seen from these tables that the PF has achieved the highest accuracy and needed the least amount of time for training. It is also observed that spectral-spatial features deliver higher overall accuracy relative to spectral features. In the case of Salinas dataset, the proposed framework shows the maximum improvement of 209.4% in SVM (CG) learning model. Salinas contains a high number of samples in the dataset due to which spectral and spectral-spatial framework performs poorly on it. When the samples are higher, the learning models are not able to train efficiently. The hybrid features of the proposed framework allow all learning models to learn rapidly and efficiently. Linear discrimination requires the least training time in all datasets. The highest time for training is taken by Deep Network-Stacked Autoencoders. The predicted classes of all the learning models is shown in Figures 4.9, 4.10, 4.11 and 4.12. These figures show the classification map of the testing sets. The resultant predicted map is very identical to the actual classification map. SVM cubic performed best with the maximum overall classification rate of 95.8% in the case of Salinas.

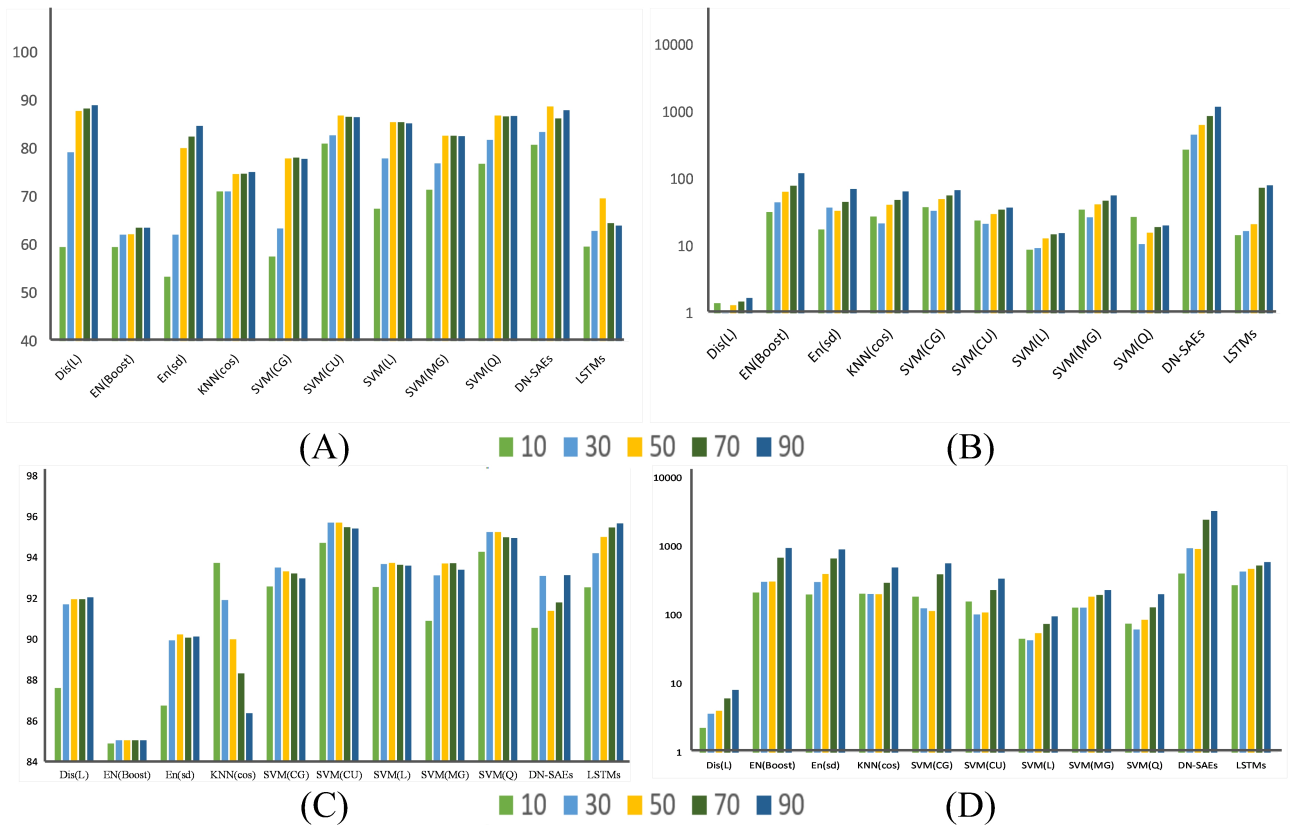


Figure 4.6: Effect of changing the hybrid features size on overall accuracy and training time (A) OA (Indian Pines) (B) TT (Indian Pines) (C) OA (Salinas) (B) TT (Salinas)

Effect of changing the hybrid feature set size on overall accuracy and training time is also tested.

The results are shown in Figure 4.6 and 4.7, which reveals that in most cases, the training time and overall accuracy is increased when the hybrid feature set size is increased. The graphs shown uses the log scale for showing the training time as the time taken by DN-SAE is very high. The mentioned 10 on graphs indicate that individual ten features are integrated that are taken from LPP and PCA, respectively. We have employed 100 hybrid features in most cases, which are considered as 50 on graphs, that shows the optimal accuracy with less processing time. In addition to this, it can also be inferred from the graphs, that when the size of hybrid features is extended, classifiers such as DN-SAEs and quadratic SVM takes lower training time as compared to smaller features. These models are able to make a more useful connection in shorter training time.

The ablation analysis is also being performed, and the results are shown in Figure 4.8. The pre-processing framework is a combination of three main parts named linear Interpolation, PCA and LPP. We evaluated five separate variations of these parts to show that the proposed framework's combination performs best. The combinations are LPP with PCA (woinLPPPCA), LPP (woinLPP), interpolation with LPP (winLPP), interpolation with PCA (winPCA), PCA (woinPCA). The graphs prove that the hybrid feature size selected for PF performs well with higher overall accuracy and lower training time as contrasted with other combinations.

The designed framework is tested on big datasets such as Salinas and Pavia and on a dataset having fewer samples like Indian pines and KSC for assessing the PF's scalability. Many previous studies only employed Indian Pines and Pavia for experimentation (Mou et al. 2017; Yue et al. 2016). This chapter also employed KSC and Salinas, which proves the validity of the proposed work. Class wise results of the PF are also shown in Tables 4.7, 4.8, 4.9 and 4.10.

The proposed framework also outperforms the previous works based on hyperspectral classification. Ozdemir (Özdemir et al. 2014) used Stacked Autoencoders with 90% training data and achieved 89.69% accuracy on the Pavia dataset. The proposed framework is trained on just 50% and is capable of achieving a maximum accuracy of 94.8% on the same Pavia dataset. Similarly, Lin (Lin et al. 2013) used autoencoders on spectral and spectral-spatial features of Pavia and achieved a minimum error rate of 14.851 and 14.355, respectively with 4 AE layers. The proposed framework has a minimum error rate of 6.2% by using only the spectral features. Singh (Singh and Kasana 2019) employed stacked autoencoders to classify the Salinas and Pavia datasets. Interpolation was used to cover the missing bands, that occurred during the acquisition of the images. Afterwards, the random forest is used to extract the essential features that are paired with extracted spatial features. Their model was able to achieve a classification accuracy of 89.9 and 94%. The proposed framework is capable of providing 95.8 and 94.8% accuracy on Salinas and Pavia dataset, indicating that the spectral features extracted with the help of LPP and PCA are very effective.

The proposed framework is focused on both LPP and PCA. Thus, the convergence of the proposed algorithm relies on PCA and LPP's convergence. For PCA, when the data is independent, identically

distributed having covariance matrix MC_a , then the standard sample covariance matrix is defined as:

$$MC = \frac{1}{n} i^T \cdot i \quad (4.3)$$

Where $i = i_1, i_2 \dots i_n$ denotes the total points. MC will converge on MC_a as $n \rightarrow \infty$. When the value of r grows, MC will converge (Shalizi 2009). We utilised PCA and LPP to extract global and local features respectively. These features are integrated to build hybrid features, which has local and global approximation of the input HSI. Let $\bar{i} = \frac{1}{n} \sum i_j$. To extract the principal components the following objective function is utilised:

$$q_{opt} = \underset{q}{\operatorname{argmax}} \sum_{i=1}^n (w_j - \bar{w})^2 \quad (4.4)$$

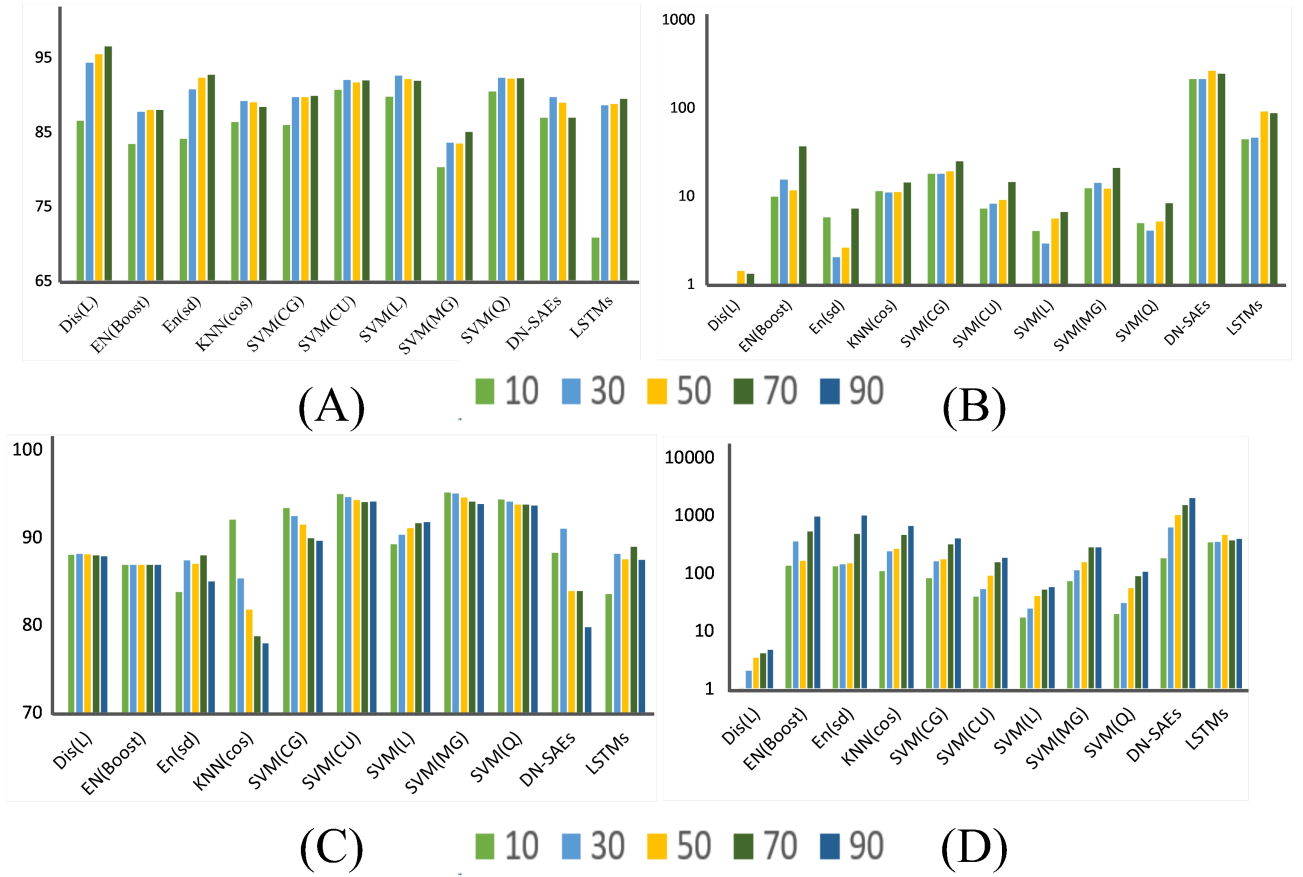


Figure 4.7: Effect of changing the hybrid features set size on overall accuracy and training time (A) OA (KSC) (B) TT (KSC) (C) OA (Pavia University) (D) TT (Pavia University)

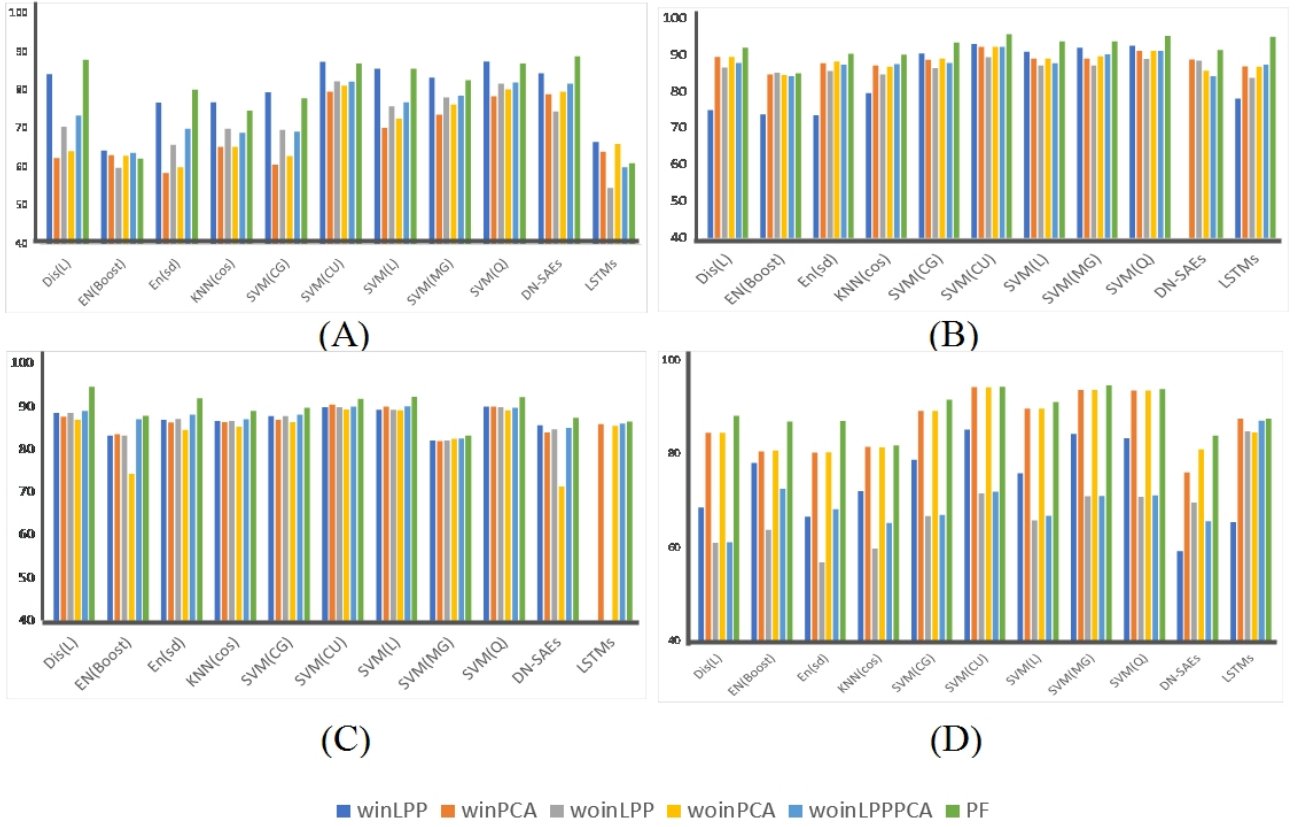


Figure 4.8: Effect of changing the feature set size on overall accuracy (A) Indian Pines (B) Salinas (C) KSC (D) Pavia University

$$= \operatorname{argmax}_q \sum_{i=1}^n q^T (i - \bar{i})(i - \bar{i})^T . q \quad (4.5)$$

$$= \operatorname{argmax}_q q^T . MC . q \quad (4.6)$$

where q is the transformation vector, $w_j = q^T . i_j$ and $MC = \frac{1}{n} \sum_{i=1}^n (i - \bar{i})(i - \bar{i})^T$. It ensures the resultant output have the maximum variance.

The LPP algorithm preserves local spatial characteristics by employing the following objective function(He and Niyogi 2004), which is defined as:

$$q_{opt} = \operatorname{argmin}_q \sum_{jk} (q^T i_j - q^T i_k)^2 Z_{jk} \quad (4.7)$$

$$= \operatorname{argmin}_q q^T A M A^T q \quad (4.8)$$

where $M = U - Z$ denotes Laplacian graph and $U_{jj} = \sum_k Z_{jk}$. Both of the dimensionality reduction algorithms rely on the eigenvalues. PCA eigenvalues are arranged in decreasing order, and the

eigenvalues calculated from $AMAT^T$ is arranged in increasing order. $AMAT^T$ convergence will lead to MC_τ . When the value of τ approach to ∞ , LPP's covariance matrix transforms to PCA's standard covariance matrix (He et al. 2005). The small value of τ enables LPP's covariance matrix to preserve the neighborhood information efficiently with fast convergence. As HSIs constitutes of high dimensions n corresponds to high value, because of this the PCA also converge. So, the proposed methodology tends to converge quicker.

We also contrasted the proposed framework in case of DN-SAEs with the previously proposed DNN on the basis of OA. Results of this comparison have been shown in Table 4.11. Both are given the identical training samples for supervised training and the same parameters are set in both SAEs.

Table 4.11: Comparison of PF with previously proposed DNN

	DN-SAEs With PF (OA %)	DNN (OA %)
Indian Pines	88.9	84.4
Salinas	93.3	87.2

PF also outperforms the previously proposed DNN. DNN is based on LPP and PF employed the hybrid features procured from LPP and PCA. The combination of both PCA and LPP leads to local and global spatial information. But in DNN, only local information is extracted *via* LPP. Due to the presence of hybrid features in PF, the DN-SAEs are able to extract more meaningful deep features as compared to DNN, which leads to better classification accuracy.

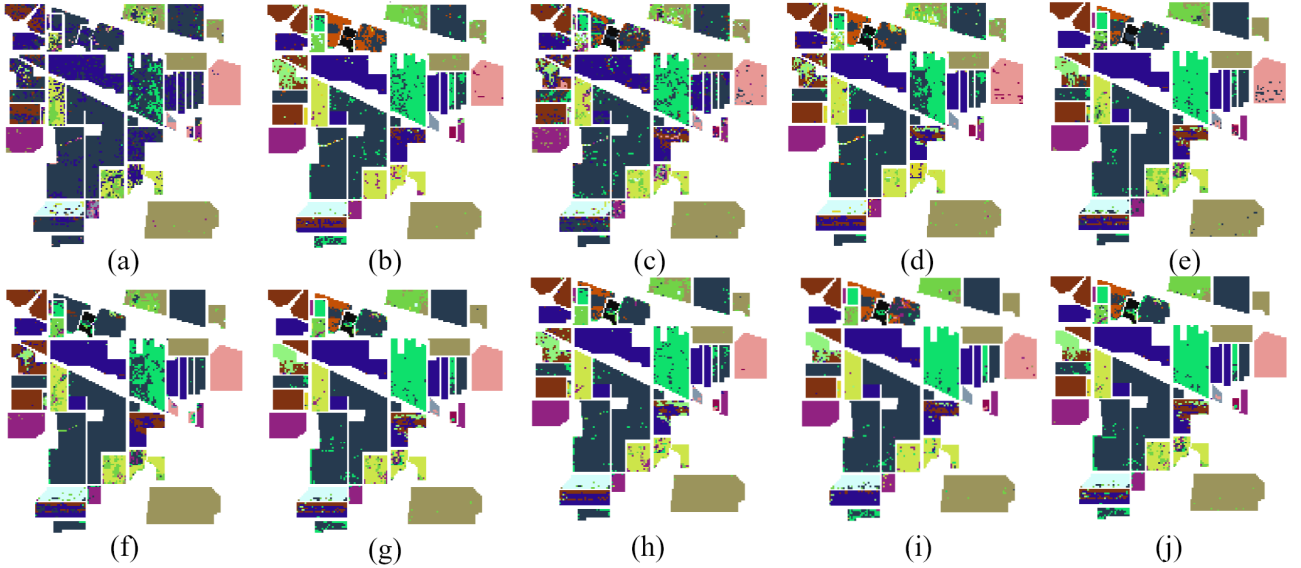


Figure 4.9: Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)

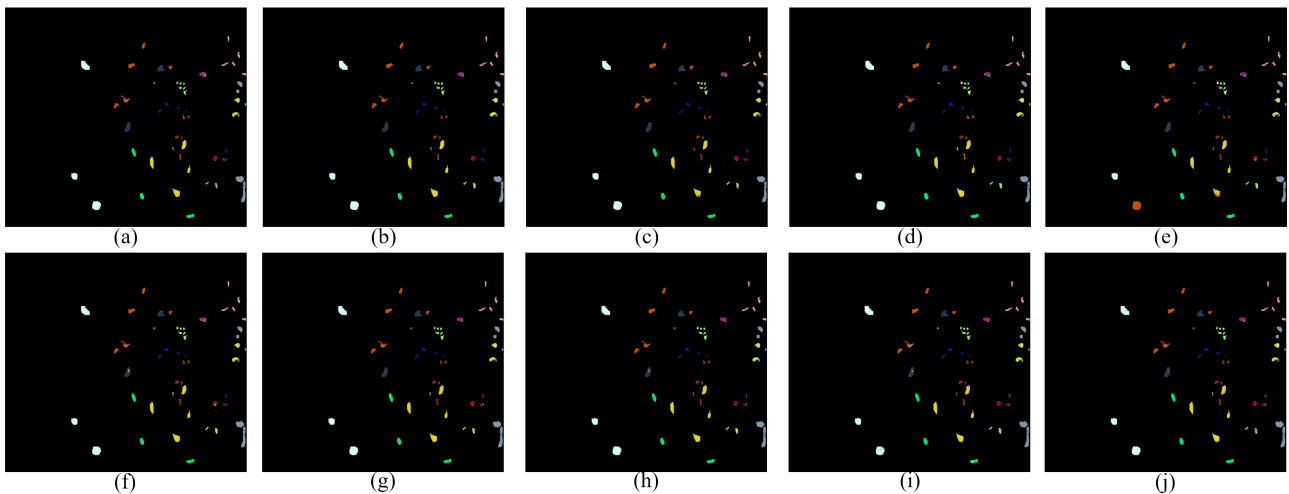


Figure 4.10: KSC Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)

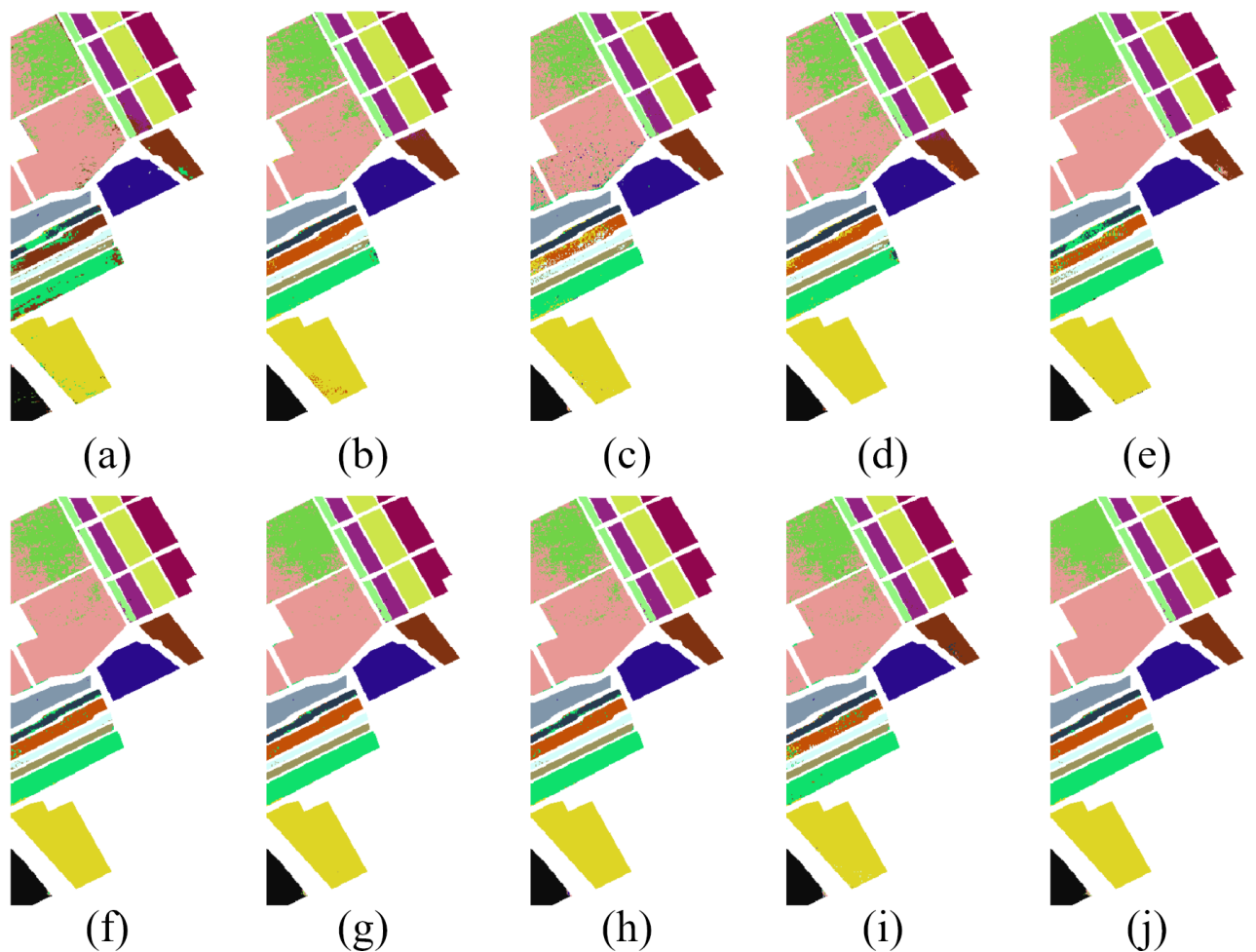


Figure 4.11: Salinas Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)

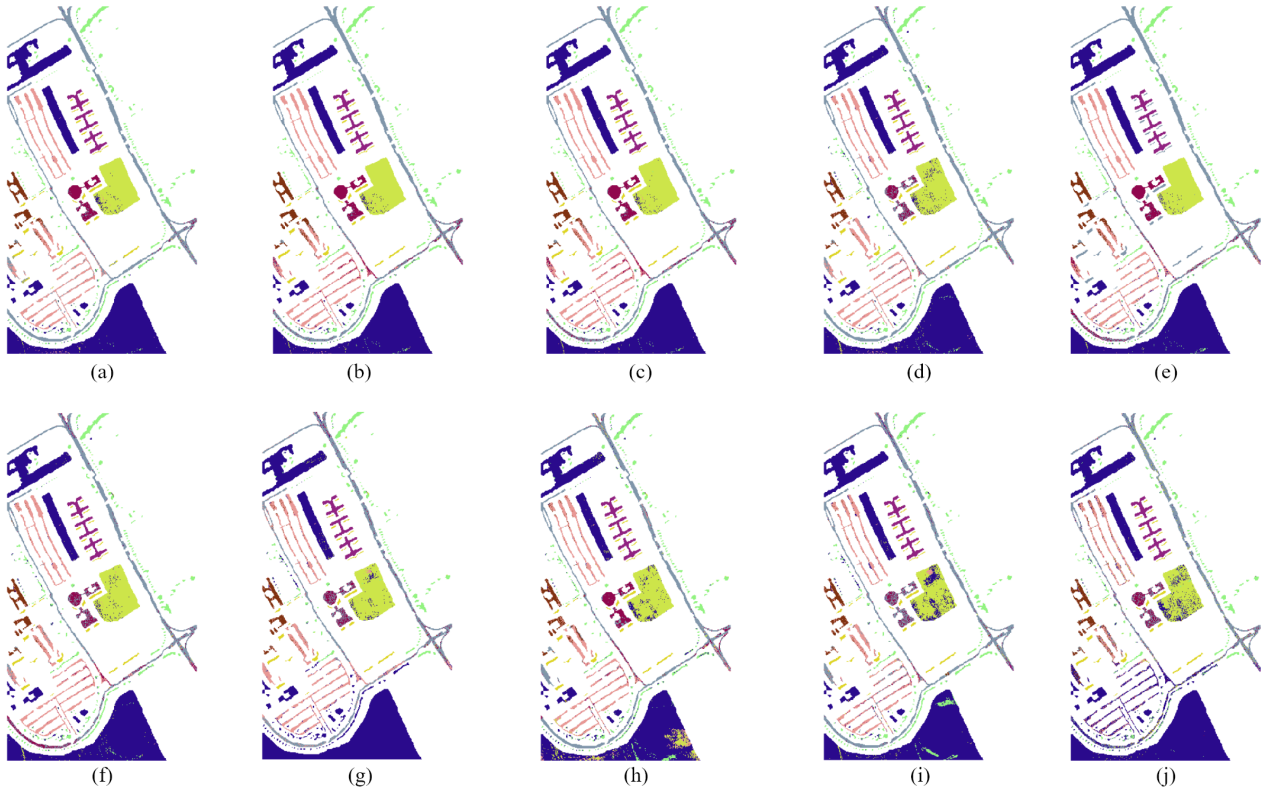


Figure 4.12: Pavia University Indian pines classification map (a) EN (BT) (b) Dis(L) (c) KNN(Cosine) (d) En (SD) (e) SVM(MG) (f) SVM(CG) (g) SVM(Q) (h) SVM(L) (i) DN-SAEs (j) SVM(CU)

4.4 Conclusion of the Chapter

In this chapter, effective classification of HSIs is performed by employing the proposed framework. During the atmospheric correction, a significant number of bands are eliminated. So, to approximate the eliminated bands, linear interpolation is employed in this framework to approximate them due to the fact the HSI is a sequential image. It also utilised the low dimensional feature set called hybrid features from PCA and LPP. This feature set is able to maintain the global with local approximations of the HSI. The proposed framework takes 4.3, 5.3, 2 and 4.7 times less processing time and it shows overall accuracy improvement of 115.5, 209.4, 119 and 131 % in case of KSC, Salinas, Pavia University and Indian Pines dataset respectively. The proposed framework is able to perform well when contrasted with popular spatial-spectral and spectral frameworks.

Chapter 5

Quantification of Soil Properties from the Hyperspectral Data using Long Short-Term Memory Networks

This chapter presents the framework to quantify the soil properties from spectral data using deep learning. Deep learning advancement has modified the conventional way of interpreting data in many areas. Quantity estimation of soil chemical and physical properties is performed in this chapter by utilising the LSTM. The efficient prediction model is designed, by training the LSTM on the spectral data obtained LUCAS spectral library as LSTMs learning is more effective when the input is sequential. The spectral dataset constitutes of high dimensional signatures, because of which we employed PCA for decreasing the dimensions.

5.1 Background of the Proposed Work

5.1.1 Recurrent Neural Networks

Recurrent Neural Network (RNN) is a category of artificial neural network which typically form meaningful connections in the temporal domain and takes sequential data as input. This is not a feedback network. It includes the loop structure, allowing sequence information to be continued at a certain time, meaning that one layer will process the data in sequence at various times (Williams and Zipser 1989), as shown in figure 5.1. With the integrated memory, RNN can quickly remember crucial details that allow to effectively predict the future output. As a consequence, RNNs are mainly employed for problems such as time series, speech recognition, *etc.*

The contents of the work presented in this chapter have been published in *Geoderma Regional*, Vol. 18, pp. E00233 (1-12), 2019. (SCI Indexed)

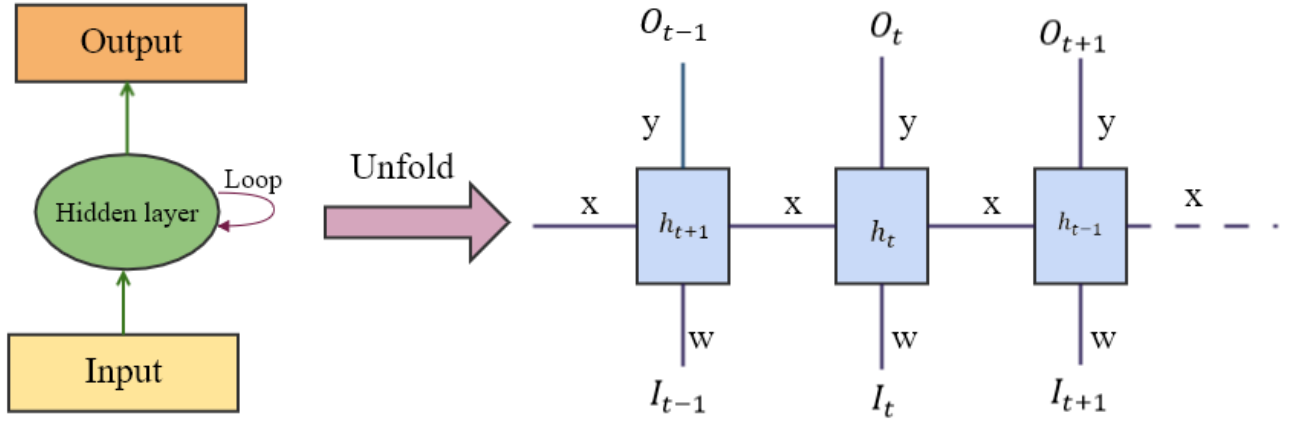


Figure 5.1: Recurrent Neural Network Unrolled

One pixel of hyperspectral data is denoted as $H = \langle i_1, i_2, i_3, \dots, i_n \rangle$. Using the following equation, the hidden RNN layer is modified.

$$h_t = \begin{cases} 0, & \text{if } x \geq 1 \\ \alpha(d_{t-1}, i_t), & \text{else} \end{cases} \quad (5.1)$$

Where d_t is the current hidden state which requires d_{t-1} (previous state) and current input i_t for the calculation of present hidden state by activation function α . α can be sigmoid or hyperbolic tangent activation function. In most cases, the hyperbolic tangent function is preferred, which is defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5.2)$$

The RNN output may be a single output or a series. In the event of hyperspectral data, only one value for each input vector can be obtained. As shown in the Figure 5.1, different weights u , v and w are allocated for effective backpropagation.

In theory, RNNs could be utilised to recall long-time dependencies, but RNNs cannot understand them in reality. Bengio (Bengio et al. 1994) first encountered this problem in 1994. It occurs because of the vanishing gradient (Pascanu et al. 2013). Therefore, in order to solve this issue, a more sophisticated neural network such as LSTMs is required.

5.1.2 Long Short-Term Memory Networks

Long Short Term Memory Network (LSTM) is the extended form of RNNs, which is mostly applicable for sequence based problems like speech recognition, time series, *etc.* (Hochreiter and Schmidhuber 1997). It performs better than conventional RNN. The presence of forget gate in LSTM enables it to selectively remember a degree of the previous information due to which it easily maintains short- and long-term dependencies. It even addresses the drawback of the vanishing gradient, that is present in

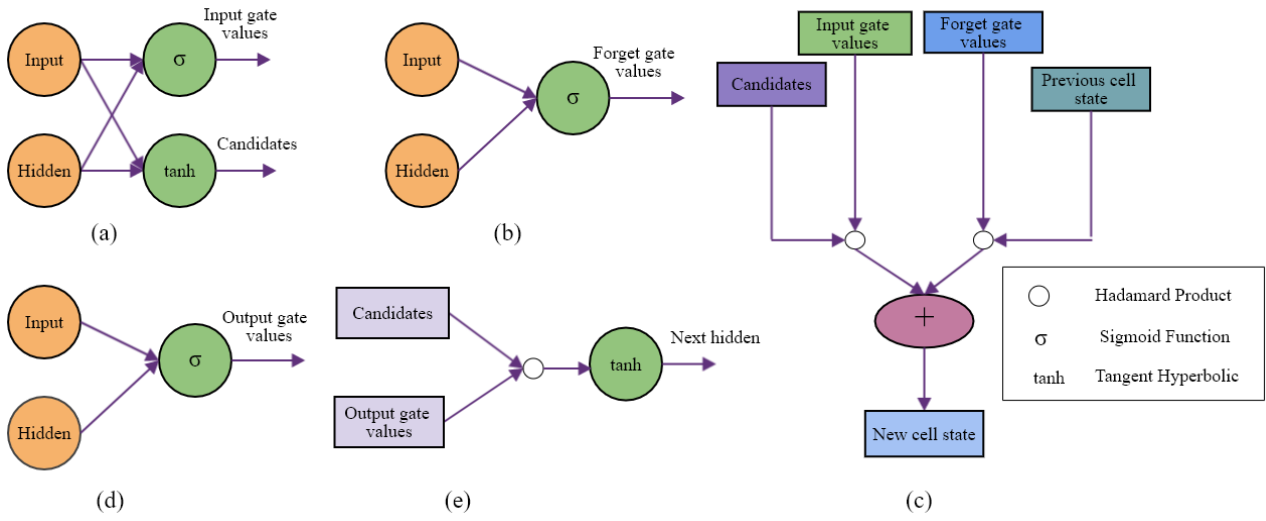


Figure 5.2: Independent operations performed on (a) Input gate (b) Forget gate (c) New cell state (d) Output gate (e) Hidden layer

conventional RNN. LSTM with two layers is shown in Figure 5.3 and gates operations are shown in Figure 5.2.

Input Gate

Its operation is performed in two steps described as:

1. In the first step, the operation is performed on the input and earlier hidden value as:

$$i_t = \sigma(\alpha_{input}) \quad (5.3)$$

$$\alpha_{input} = w_{i1} \cdot I_{it} + w_{i2} \cdot h_{t-1} + b_{input} \quad (5.4)$$

Where b_{input} denotes the bias, w is the weight associated, and σ is the activation. i_t represent the degree of data passed to the memory cell.

2. In the second step, the operation is performed input and earlier hidden value to obtain the candidates as shown in Figure 5.2 (a). It is described as:

$$\hat{C}_t = \tanh(\alpha_{\hat{C}}) \quad (5.5)$$

$$\alpha_{\hat{C}} = w_{c1} \cdot I_{it} + w_{c2} \cdot h_{t-1} + b_c \quad (5.6)$$

Where $\hat{C}_t$ represents the new candidates and $\tanh$ is the activation function employed.

Forget gate

The operation of this gate can be seen from the Figure 5.2 (b). It is employed to selectively forget useless information. Its operation can be defined as:

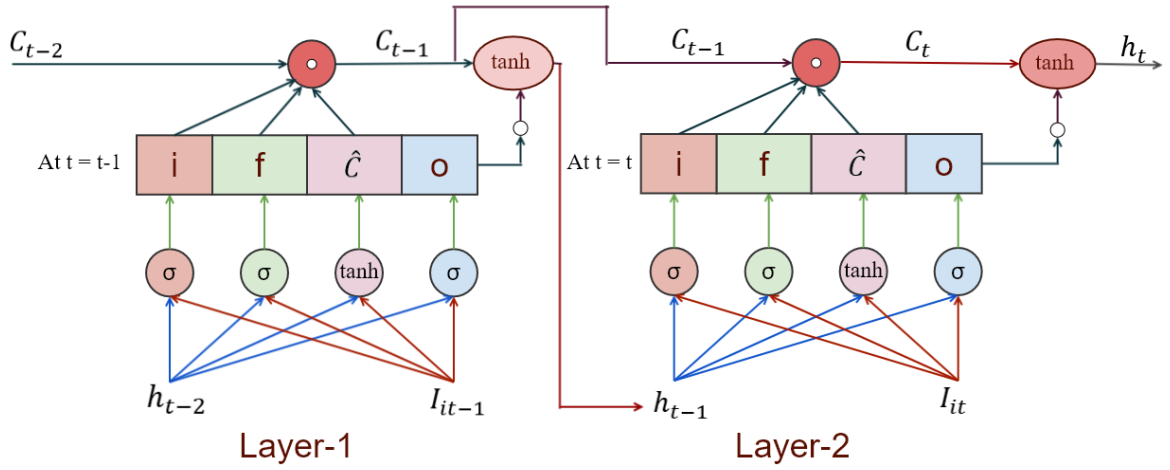


Figure 5.3: LSTM with two layers

$$f_t = \sigma(\alpha_f) \quad (5.7)$$

$$\alpha_f = w_{f1} \cdot I_{it} + w_{f2} \cdot h_{t-1} + b_f \quad (5.8)$$

Based on the value of f_t information is dropped.

Memory Cell

Its operation includes previous cell state, candidate, forget gate and input gate values as shown in Figure 5.2 (c). Its operation is described as:

$$C_t = I_{it} \odot \hat{C}_t + f_t \odot C_{t-1}$$

Where C_{t-1} represents cell state (Previous) and $\odot$ is the Hadamard product

Output Gate

Its operation is performed on input (present) and hidden value (previous) as shown in Figure 5.2 (d). Its operation can be described as:

$$O_t = \sigma(\alpha_{output}) \quad (5.9)$$

$$\alpha_{output} = w_{o1} \cdot I_{it} + w_{o2} \cdot h_{t-1} + b_{output} \quad (5.10)$$

Hidden Layer

Hidden value is calculated with the help of output and cell state as shown in Figure 5.2 (e). Its operation can be described as:

$$h_t = O_t \odot \tanh(C_t) \quad (5.11)$$

5.2 Materials and Methods

5.2.1 LUCAS Dataset

Gergely Toth conducted a large-scale project named Land Use / Land Use Area Frame Survey (LUCAS) to monitor the land surface character (Tóth et al. 2013). It captured the GPS Coordinates in the EU Area along with the real-time information of soil's chemical and physical characteristics. The sample locations of the EU is presented in Figure 5.4. Samples were drawn at a 2×2 kilometre crossing that encompasses the EU area. The goal of this project is to form a consistent dataset in the EU that will lead to informed decision-making on agriculture policies. The creation of 'LUCAS-Topsoil' was partnered by Eurostat with the General Directors of the European Commission and Joint Research Center.

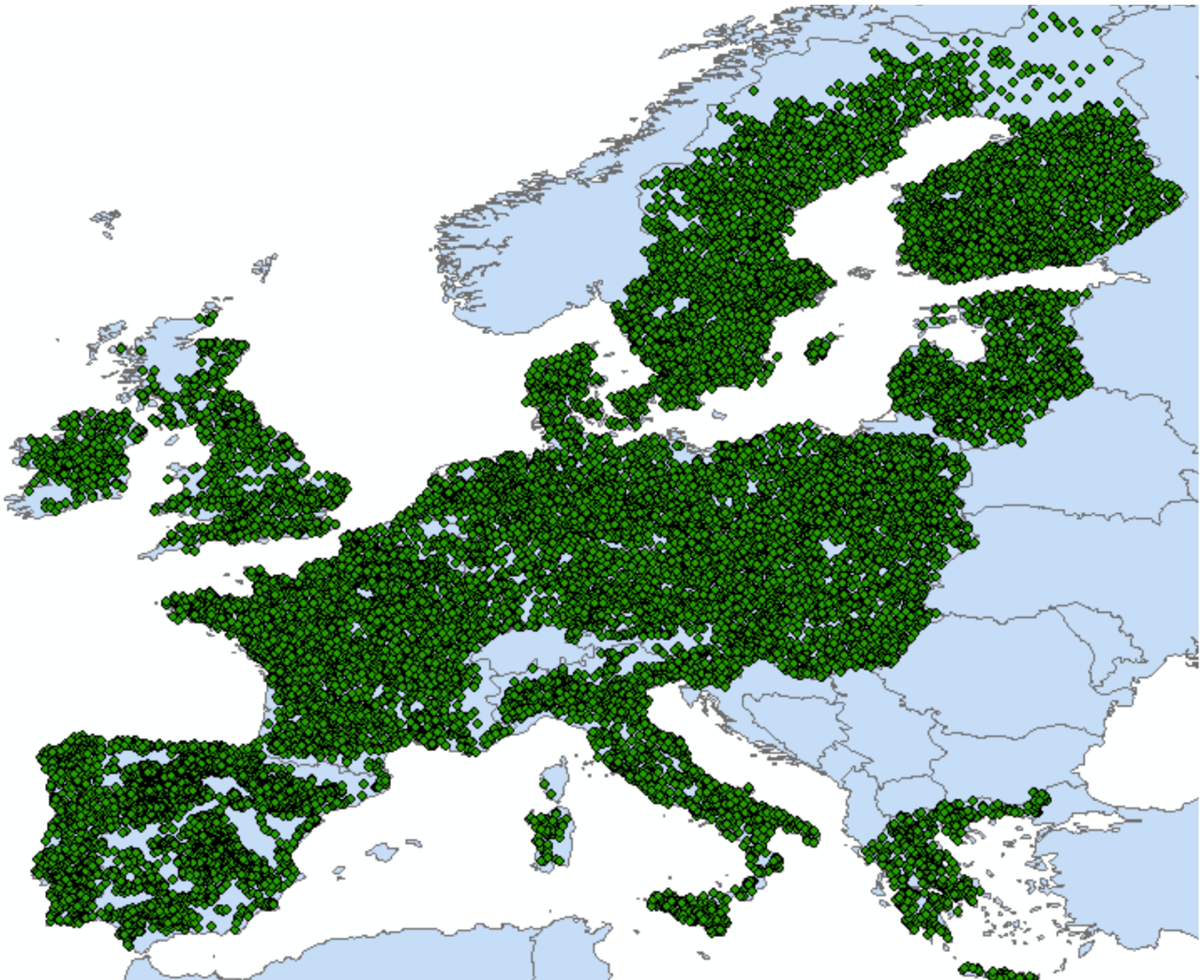


Figure 5.4: Sampling points of European Union

5.2.2 Field Sampling

The sampling of the dataset followed the appropriate research design and was split into two components. In the first section, a normal grid was accompanied by a structural set having approximately 10% of the total points with a similar geometric distribution. This process is standard (Van-Camp et al. 2004) and set by the national soil monitoring system. The second component is based on a stratified sampling of random nature for gathering information on land usage and geography. It is most widely advocated because the randomisation of this sampling system is not biased (Oliver 2010). All types of land are classified into eight grades. A total of 20,000 samples from various kinds

Table 5.1: % of samples acquired by countries and land uses

Country	Grassland	Woodland	CRC	Shrubland	CPR	TS
Luxembourg	0.05	0.01	0.01	0.26	–	0.02
Belgium	0.09	0.08	0.18	0.07	0.01	0.36
Cyprus	0.13	0.07	0.13	0.01	0.05	0.45
Slovakia	0.16	0.34	0.56	0.11	0.01	1.34
Denmark	0.17	0.13	0.83	0.03	0.14	1.16
Estonia	0.27	0.52	0.27	0.11	–	1.1
Portugal	0.32	0.42	0.23	0.24	0.36	2.38
Greece	0.44	0.32	0.75	0.22	0.5	2.46
Malta	0.44	0.11	0.01	0.04	0.01	0.1
Finland	0.47	0.63	1.57	0.27	0.01	8.59
Czech Republic	0.48	0.44	1.14	0.01	0.03	2.16
Poland	0.5	0.97	4.15	0.53	0.11	8.25
Hungary	0.52	0.3	1.57	0.05	0.03	2.49
Latvia	0.66	0.63	0.39	0.01	0.01	1.75
Austria	0.67	0.61	0.73	0.03	0.02	2.1
UK	0.67	0.61	1.77	0.01	–	4.72
Lithuania	0.71	0.35	0.69	0.06	–	1.78
Spain	0.73	9.02	6.62	0.07	2.1	13.5
Ireland	0.87	0.1	0.06	0.2	–	1.17
Italy	1.43	0.64	2.75	0.04	1.34	6.68
Slovenia	1.75	1.08	0.04	0.03	0.01	0.56
Netherlands	2.23	1.52	0.44	0.02	–	1.06
Sweden	2.29	0.36	0.93	0.01	–	11.29
Germany	2.75	2.05	4.65	0.3	0.14	9.75
France	4.16	1.9	7.64	0.02	0.44	14.78
Total Samples	4449	5643	7601	475	1026	19967

CPC–Cropland permanent crops, TS–Total Samples, CAC–Cropland annual crops,

of land were taken. The samples within the strata are similar, but inter-strata samples are non-similar in nature. Five samples were obtained from one stratum in total and mixed together to develop one sample, which is then taken for lab examination. The percentage obtained from each nation and the

land uses are shown in Table 5.1.

5.2.3 Soil Testing

The samples were examined for particle size, Potassium, clay, Cation Exchange Capacity (CEC), pH, Nitrogen, coarse fragment content, Exchangeable acidity and Phosphorous with hyperspectral/multispectral reflectance signatures. Six attributes have been identified, as shown in Table 5.2. Around 20,000 of these samples have been tested. All samples are tested using the standard guidelines with reference to SPS 2011. The procured readings from the spectroscope contains the reflectance values acquired between 400 to 2500 nm having 0.5 spectral interval leading to total 4199 reflectance points. These points were procured by utilising the spectroscopic FOSS instrument. During 2009-2011, all laboratory work was done (Tóth et al. 2013).

Table 5.2: Statistics of different soil properties from samples ($n=19036$)

	pH (H_2O)	Nitrogen g/Kg	Clay %	CEC cmol(+)/Kg	OC g/Kg	Sand %
Minimum	3.21	0	0	0	0	1
Mean	6.20	2.92	17.8	15.76	50	40.41
Maximum	10.08	38.6	79	234	586.8	99
St.dev	1.35	3.76	13.37	14.48	91.31	27.24
Median	6.21	1.7	16	12.4	20.8	39

5.2.4 Spectral Acquisition and Pre-treatments

In the first step, the samples of soil are dried and filtered ($<2\text{mm}$) for determination of the absorption spectra procured via XDS analyser. The analyser's inbuilt spectroscope acquire a very precise reading from 400-2500 nm at spectral intervals of 0.5. The spectrometer detector is made of silicone and lead sulphide having 8.75 ± 0.1 nm as optical bandwidth, which allows the electromagnetic area to acquire 4200 wavelengths.

The captured readings from spectrometer constitute of instrumental artifacts in the range of 400-500nm, which are eliminated. Afterward, Savitzky-Golay filter is employed on LUCAS for smoothing out the noisy readings (Savitzky and Golay 1964). SG illustrates that in precisely the same way as a weighted moving average, a moving polynomial function may be treated numerically. It can be employed for procuring the 1st and 2nd derivative of the smoothed signal. The smoothed signal output achieved from the SG by utilising the 1st derivative with the window of 101 points. In addition to this, SG also employed 3rd order polynomial.

Afterward, the smoothed data acquired *via* SG-filter is then transformed with Simple Normal Variate (SNV). SNV has a gaussian distribution, which performs rescaling. The output of SNV has

an average (μ)=0 and the standard deviation (σ)=1. It can be defined as:

$$Z = \frac{X - \mu}{\sigma} \quad (5.12)$$

where X belongs to $N(0,1)$ and N is normal distribution.

5.3 Proposed Framework

The soil properties quantitative framework, as presented in Figure 5.5 is proposed in this section. We call this the PCA-LSTM (PPLF) framework. The LUCAS dataset comprises 4199 values obtained from 400-2500 nm region. The captured spectral values also include the device artifacts within a range of 400-500 nm, which are omitted. Pre-treatments for the artifact-free dataset are then added and the resulting dataset reduced to 3999. LUCAS comprises of hyperspectral signature that has high dimensions due to the large spectral values. PCA is then employed to decrease the LUCAS dimensions to minimise processing and analytic time. The goal of the PCA is to minimise dimensions and optimise the variance of the resulting components without much loss of information. 100 % variance of LUCAS is represented by the selected PCA components. The processed PCA data are then partitioned randomly into training, validation and testing set.

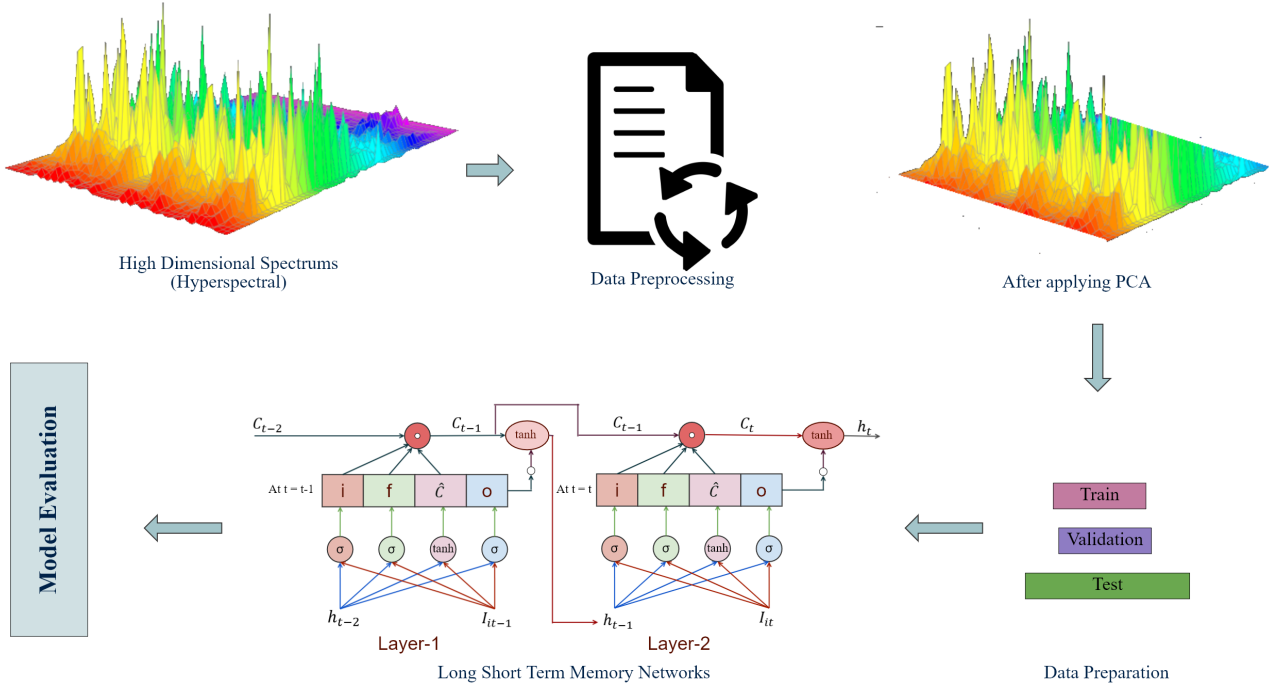


Figure 5.5: Proposed PCA-LSTM Framework

A deep learning architecture called LSTM is then employed. This is primarily seen in case of sequential problems (Mou et al. 2017). Afterward, LSTM networks utilises training and validation

samples for supervised learning. After processing, a trained learner based on LSTM is procured. Then, test set for prediction of quantities is presented to the trained model. The predicted quantity values obtained from this trained model is then contrasted with the ground quantity values for determining the predictive ability of the trained model.

Algorithm Prediction with the Proposed PCA-LSTM framework

Input-A hyperspectral point H_i^k which has total i points, each of k dimensions. Each point is associated with the quantity of minerals V_i

Output-Predicted quantification values V_i .

- 1: Take hyperspectral points from LUCAS dataset and store it into K .
 - 2: Remove the artifacts from K and store it in S .
 - 3: Evaluate SG on S and store it in N .
 - 4: Evaluate SNV on N and store it in Z .
 - 5: Evaluate PCA on Z and store it in H .
 - 6: Select p components from H such that these p components have maximum variance and input these components into F .
 - 7: Prepare training, validation and testing sets from F and store it in I , J and L respectively.
 - 8: Input point vectors from I and J to LSTMs (Optimiser=Adam, Epoch=100).
 - 9: Associate the respective value V_i with every point vector for supervised training.
 - 10: Trained prediction model TP is obtained.
 - 11: Input L to TP to predict the quantification values V_i .
-

5.4 Experimental Results and Analysis

LUCAS hyperspectral samples are considered for performing experiments on a device with GeForce GTX 1080 edition Nvidia graphics processor having 2560 cores of CUDA and a 1607 MHz Graphics clock. After pre-processing, the obtained 19036 hyperspectral signatures are then passed to PCA. It decreases the surplus dimensions by removing redundant information. The resultant data obtained from PCA can explain the 100% variance of the input LUCAS (2000 components). This practice is also followed in PCR.

The resultant dataset obtained *via* PCA is partitioned as training, validation, and testing samples. LSTM is 25% training samples, and the validation is performed by 25% samples of LUCAS. The validation set is employed to avoid the overfitting of the trained model. During the training process, the validation set samples are also evaluated, to ensure the supervised training is going smoothly. If the validation set performs badly, it means the training process is not effective, which can be further improved by hyperparameter optimisation i.e. trying various parameter to avoid overfitting (Kågebäck and Salomonsson 2016).

The PPLF is executed 10 times, and then the mean of evaluated results is taken. The hyperspectral signatures are seen as a sequence problem, so LSTMs are favoured. The LSTMs are more efficient in this scenario because they can remember both long and short-term dependency. It is developed in

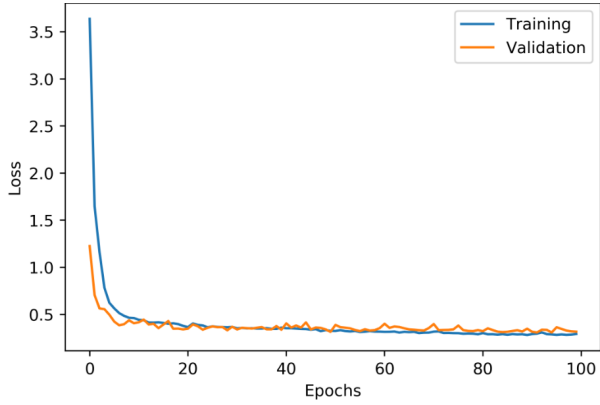
Table 5.3: Parameters/ Hyperparameter set in LSTM

Parameters/ Hyperparameter	Description	Activation
Optimiser	Adam	—
Learning rate	0.001	—
Epoch	100	—
Hidden layer	—	Tanh
Loss Function	Mean Absolute Error	—
LSTM units/ hidden neurons	2500	—
Input gate	—	Sigmoid and Tanh
Batch size	100	—
No. of Dense layers	1	Rectified Linear Unit
No. of LSTM layers	1	—
Dense Layer neurons	1	—
Output gate	—	Sigmoid
Dense Layer parameters	2501	—
Forget gate	—	Sigmoid
Dropout	0.3	—
Training %	25	—
Validation %	25	—
Testing %	50	—

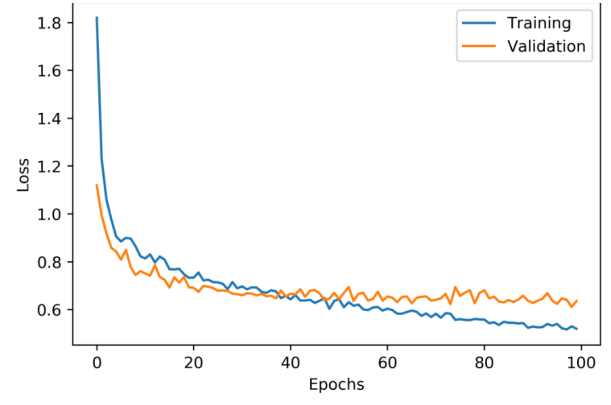
python with the help of keras library. The developed framework requires low training time due to a reduced feature set obtained from PCA. The parameter set used in designing LSTM is shown in Table 5.3. One layer of LSTM with 2500 neurons is employed in the model. Error loss function ‘Mean Absolute Error’ is employed with effective ‘Adam’ optimiser. One dense layer is employed at the end as the model outputs the one quantity value. The model is trained with 100 epochs. The parameter and hyperparameter set are fixed for quantifying the selected soil characteristics. The loss of training and validation curves of LSTM over 100 epochs are shown in Figure 5.6. The loss rate of the training and validation is nearly similar, indicating a good fit in all cases, particularly in the context of pH (H_2O) and OC.

The PPLF is contrasted to other commonly employed models such as PLSR, Multiple Linear Regression (MLR), PCR, LSTM, Support Vector Regression (SVR), (Ben-Dor and Banin 1995) and Stepwise Regression (SWR). In Python 3.6, the PPLF is implemented by the TensorFlow GPU along with PLSR, MLR, PCR, LSTM and SVR. SWR is designed in matrix laboratory version 2018a. In order to perform a fair assessment, all models received the reduced-dimensional data from PCA with the exception of LSTM and MLR. The efficacy of PCA components is shown by comparing LSTM with the PPLF framework. Table 5.4 shows the comparison of PPLF with other models on testing set on the basis of RMSE, Spearman correlation coefficient (ρ), Mean Absolute Error (MAE), Pearson correlation coefficient (r), Lin’s concordance coefficient (ρ_C) and Coefficient of Determination (R^2).

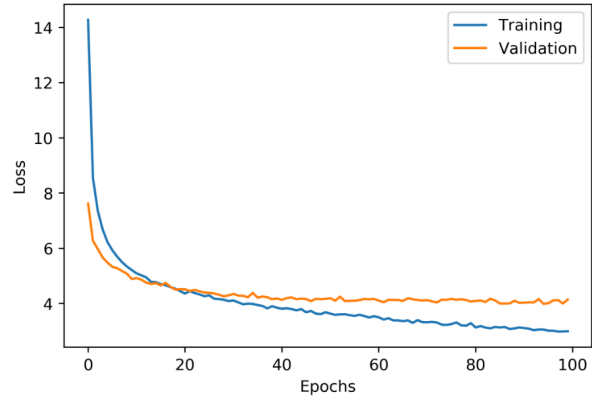
From Table 5.4, it can be inferred that based on six metrics, PPLF is able to surpass other conven-



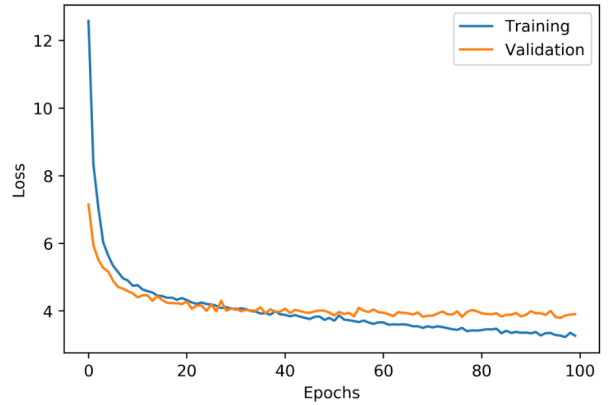
(a)



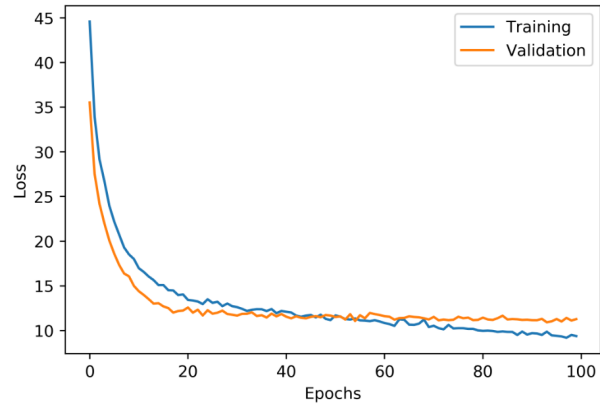
(b)



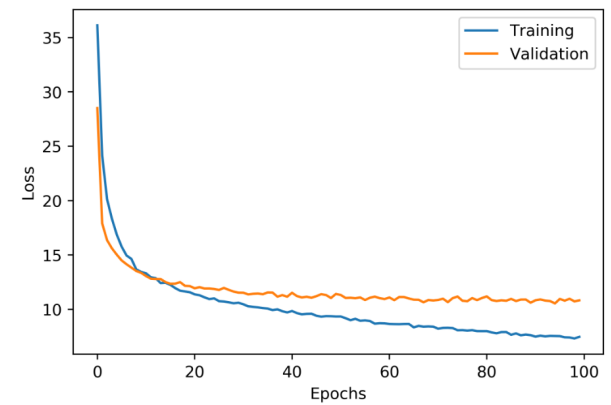
(c)



(d)



(e)



(f)

Figure 5.6: Mean absolute Error (loss function) in training and validation (a) $\text{pH}(\text{H}_2\text{O})$ (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand

Table 5.4: PPLF comparison with other models based on six evaluation metrics.

		PPLF	SWR	PLSR	LSTM	MLR	PCR	SVR
pH(H_2O)	RMSE	0.42	0.58	0.46	0.51	0.53	0.46	0.46
	R^2	0.90	0.81	0.89	0.86	0.85	0.89	0.88
	MAE	0.32	0.45	0.35	0.38	0.41	0.35	0.35
	ρ	0.96	0.91	0.94	0.93	0.93	0.95	0.94
	r	0.96	0.90	0.94	0.92	0.92	0.94	0.94
	ρ_C	0.95	0.90	0.94	0.93	0.92	0.94	0.94
Nitrogen	RMSE	1.15	1.77	3.24	1.64	1.93	1.63	2.03
	R^2	0.91	0.78	0.25	0.81	0.73	0.81	0.71
	MAE	0.64	1.05	2.43	0.90	1.41	1.11	0.98
	ρ	0.89	0.69	0.31	0.76	0.66	0.74	0.76
	r	0.95	0.89	0.57	0.90	0.87	0.90	0.87
	ρ_C	0.95	0.88	0.56	0.89	0.87	0.90	0.80
Clay	RMSE	5.95	8.74	9.53	9.19	8.49	7.34	8.99
	R^2	0.80	0.57	0.49	0.52	0.59	0.69	0.54
	MAE	4.14	6.36	7.39	5.92	6.40	5.40	6.17
	ρ	0.92	0.79	0.67	0.80	0.82	0.86	0.79
	r	0.86	0.75	0.72	0.75	0.80	0.84	0.75
	ρ_C	0.85	0.73	0.71	0.70	0.80	0.84	0.70
OC	RMSE	23.25	32.35	78.30	28.26	38.20	32.47	84.68
	R^2	0.94	0.87	0.28	0.91	0.83	0.88	0.16
	MAE	11.24	18.52	59.36	13.90	27.58	22.19	31.85
	ρ	0.89	0.70	0.08	0.84	0.62	0.69	0.69
	r	0.97	0.93	0.58	0.95	0.92	0.94	0.78
	ρ_C	0.97	0.93	0.57	0.95	0.92	0.94	0.23
CEC	RMSE	6.75	9.22	10.22	9.43	10.22	8.68	10.28
	R^2	0.77	0.59	0.47	0.55	0.47	0.62	0.47
	MAE	3.89	5.68	7.17	5.05	7.38	5.83	5.23
	ρ	0.90	0.76	0.73	0.76	0.71	0.80	0.82
	r	0.90	0.77	0.72	0.82	0.74	0.80	0.72
	ρ_C	0.89	0.75	0.71	0.67	0.74	0.80	0.59
Sand	RMSE	14.59	19.80	18.85	18.20	21.57	20.96	20.49
	R^2	0.71	0.48	0.52	0.56	0.38	0.41	0.43
	MAE	10.80	14.90	14.72	14.23	16.79	16.80	16.21
	ρ	0.86	0.69	0.75	0.77	0.69	0.65	0.67
	r	0.88	0.69	0.51	0.76	0.68	0.64	0.66
	ρ_C	0.87	0.66	0.19	0.75	0.68	0.57	0.62

tional methods. It indicates the least RMSE and MAE values among the ground truth and predicted values in nitrogen. The use of PCA also helps to construct an effective prediction model compared to the LSTM without PCA. The PPLF indicates atmost 3.36 times lower RMSE in the context of organic carbon (OC). Overall least error is reached in quantifying nitrogen with MAE of $0.64g/kg$. The RMSE results are highly correlated to MAE. In the case of OC, the PPLF Model displayed a R^2 of 0.94. OC is one of the principal ingredients of the soil and is the subject of several studies (Stevens et al. 2013; Vaudour et al. 2016).

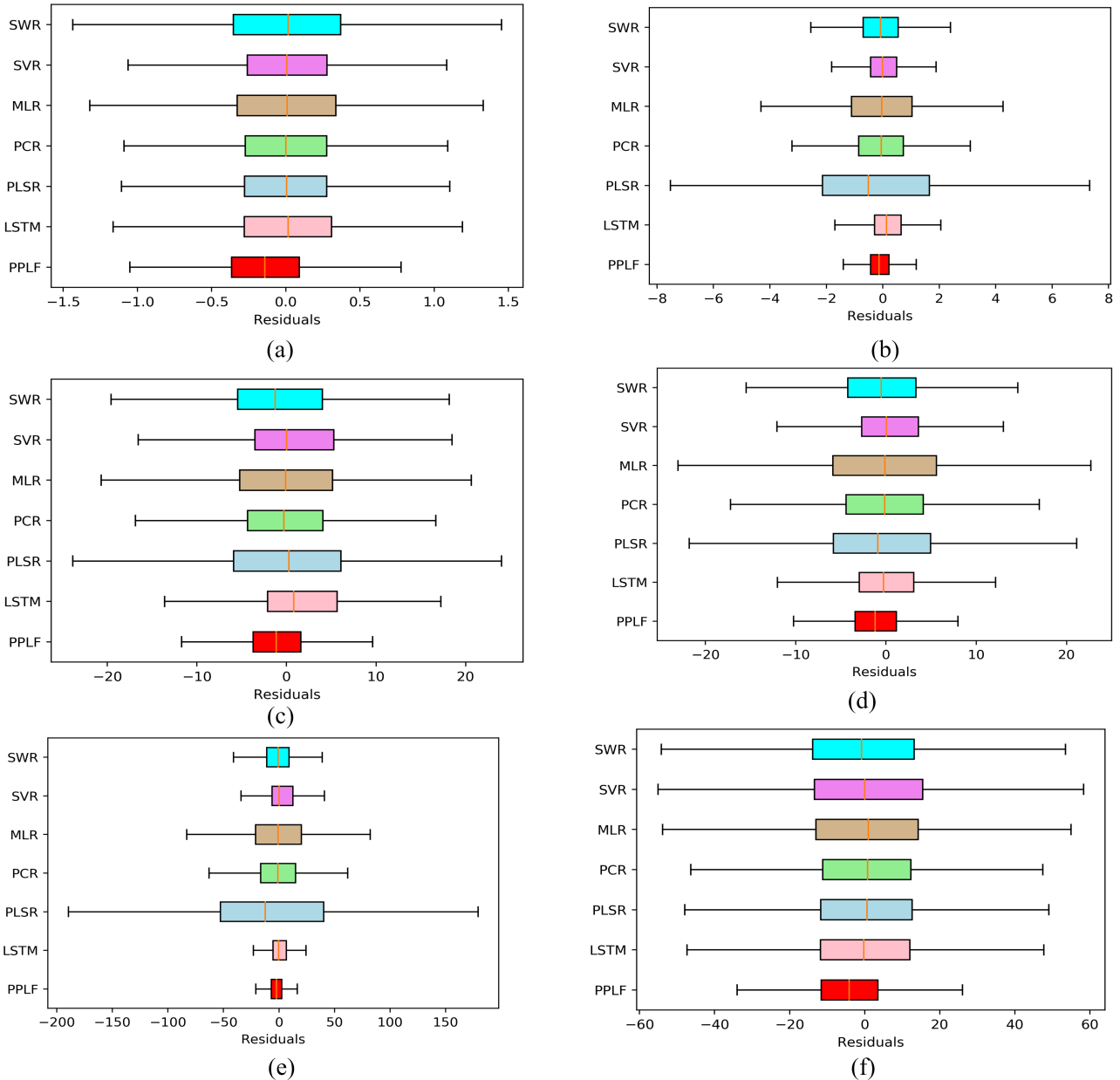


Figure 5.7: Residuals (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand

The readings of PPLF's R^2 indicate that the PPLF can perform better than most of the conventional approaches. r is employed to show the association between the ground truth and PPLF prediction.

It can show a positive or negative correlation between variables. PPLF's r indicates a significant linear association among the prediction and ground truth quantity with the highest r of 0.97 is seen in the case OC. We employed Spearman's coefficient for examining the monotonic relation among the prediction and ground truth quantity. The resulting spearman values of PPLF indicate a stronger monotonic connection.

ρ_C is also employed to examine the concord among the ground truth and the prediction. A concordance coefficient of 0.97, 0.95 and 0.95 is seen in OC, Nitrogen and pH (H_2O) in PPLF. Not only was the PPLF superior to all common regression models such as MLR, SWR, PLSR and PCR, but it also performed better in comparison to ML based SVM. In Figure 5.7, the residual graph of the test set is shown, and in each case, the PPLF has the least residual values. Figure 5.8 shows the ground truth vs predicted values observed from the 50% testing samples.

The PPLF is also contrasted with earlier efficient models such as Wadoux's 1-D CNN, Ballabio's Gaussian Process Model and Yang's combined CNN and RNN model based on R^2 as shown in Table 5.5. These models have also used LUCAS and have randomly selected samples for training and testing, which is similar to our work. PPLF outperforms other works due to the presence of efficient LSTM model which constitutes of forget gate, that filters only the important sequential information. PPLF shows the most improvement of R^2 in comparison to Wadoux's 1-D CNN model. It also shows significant improvement compared to Yang's and Wadoux's model, which are also based on deep learning. PPLF demonstrates 2.2 times better values for R^2 as compared to Ballabio's model in case of OC. Also, PPLF is trained using 25% data as compared to 75% and 90% data used for training by Yang's and Ballabio's models, respectively.

Table 5.5: Earlier studies comparison with HF on Test set and R^2

Parameters	PPLF	Wadoux (2019)	Ballabio et al. (2019)	Yang et al. (2020)
pH (H_2O)	0.90	–	0.65	–
Nitrogen	0.91	0.24	0.60	0.70
Clay	0.80	0.22	–	–
CEC	0.77	–	0.35	0.73
OC	0.94	0.15	0.41	0.73
Sand	0.71	0.36	–	–

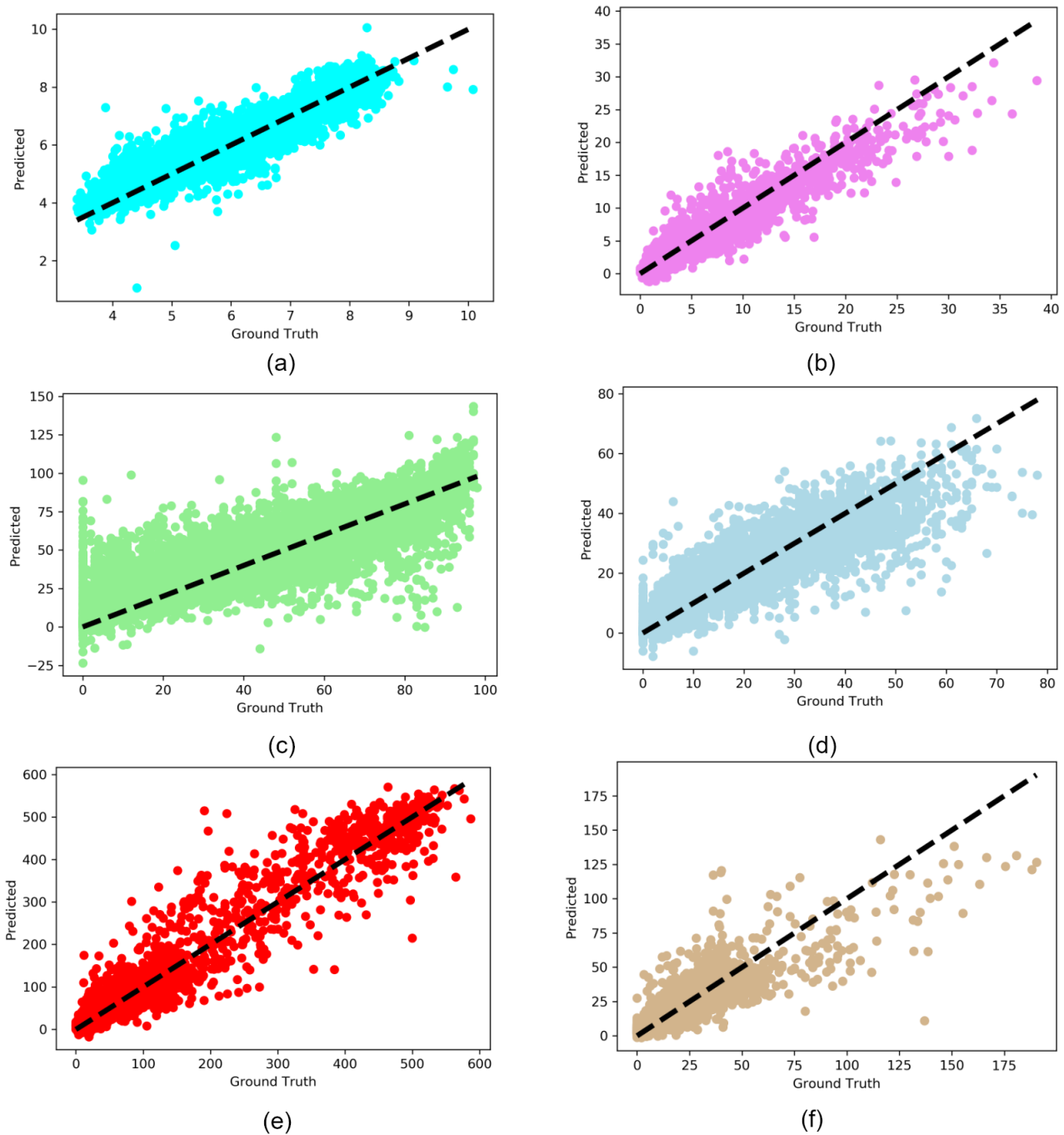


Figure 5.8: Predicted vs Ground Truth graphs of PPLF (a) $\text{pH}(\text{H}_2\text{O})$ (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand

5.5 Conclusion of the Chapter

A novel approach for quantification of soils using the hyperspectral data is presented in this chapter. The PPLF relies on PCA and LSTM to determine the spectral characteristics of the quantities of European soil resources. After the pre-processing of the dataset, PCA processes its high dimensional

hyperspectral data to minimise their dimensions. LSTMs are also supplied with PCA processed features to produce an effective model. Because of the hyperspectral data's temporal nature, LSTMs are utilised for effective quantity estimation of soil parameters. PPLF showed atmost improvement of 111%, 364%, 163%, 164%, 588%, 187% in R^2 in case of $\text{pH}(H_2O)$, nitrogen, clay, CEC, OC and sand respectively. This chapter reveals that deep learning is an effective method for the processing of the spectral data. The developed model proved to be efficient in estimating the soil characteristics as compared to conventional models.

Chapter 6

Quantification of Soil Properties Using Hybrid features and Deep Learning

This chapter presents the deep learning-based hybrid framework to quantify the soil minerals. It quantifies clay, CEC, pH of H_2O , nitrogen, OC and sand. Quantity estimation of soil's important properties is crucial for sustainable farming. Most of the existing work for soil characteristics estimation do not consider hyperspectral data as sequential. In this chapter, we proposed a novel framework named hybrid framework is proposed, in which hybrid features are employed. These hybrid features are formed by combining the output obtained from LPP and PCA. In addition to this, LSTM is utilised for designing an effective prediction model due to the fact that it generates efficient results in case of a sequential problem. The hybrid framework is compared to conventional regression-based models to show its efficiency.

6.1 Background of the proposed work

6.1.1 Principal Component Analysis and Locality Preserving Projections

Most common dimensionality reduction techniques employed are PCA and LPP. The application of PCA and LPP preserve the overall structure of the input by providing the global and local approximation, respectively. The features obtained from these techniques are put together to form hybrid features that are given to the LSTM for supervised training.

PCA- The use of PCA provide the global approximation of the input and preserves the spatial information in the output. It maximized the variance in the output. The objective function of PCA is

The contents of the work presented in this chapter have been published in 16th *International Conference on Environmental Science and Technology*, Greece.

given as:

$$O = \underset{O}{\operatorname{argmax}} \sum_{i=1}^n O^T \cdot Q \cdot O \quad (6.1)$$

Where O represent the transformation vector, and $Q = \frac{1}{n} \sum_{i=1}^n (E - \bar{E}) \cdot (E - \bar{E})^T$ and E denotes information point. More details are given in chapter 4.

LPP- The use of LPP provide local approximation of the input and preserve the locality information of neighbors. LPP objective function is given below:

$$O = \underset{O}{\operatorname{argmin}} \sum_{jk} O^T W M W^T O \quad (6.2)$$

Where $M = R - I$ and $R_{jj} = \sum_k Y_{jk}$. More details are given in chapter 3.

6.1.2 Long Short-Term Memory Networks

LSTM is the extended form of RNNs, as shown in Figure 5.3 that avoids the problem of vanishing gradient (Hochreiter and Schmidhuber 1997). It is ideal for problems like stock prediction, signal analysis, time series, speech analysis, *etc* as they are based on the sequential analysis. The presence of forget gate in LSTM enables it to selectively remember a degree of the previous information due to which it easily maintains short and long-time dependencies. It constitutes input, forget and output gates. All the gates operations are discussed in chapter 5.

6.2 Proposed Framework

A hybrid framework for the quantity prediction of soil characteristics is proposed, as shown in Figure 6.1. Hyperspectral signatures in the LUCAS dataset contained instrumental artifacts in 400 to 500 nm region of the electromagnetic spectrum, which is discarded. Then, two pre-processing techniques named Savitzky-Golay and Simple Normal Variate are applied to the resultant dataset. The details regarding pre-processing are discussed in Chapter 5. Afterward, PCA and LPP are employed on the pre-processed LUCAS dataset. Then, the resultant output of both dimensionality reduction algorithms is combined to form the new hybrid features which comprise of local as well as global spatial information. Then, the spectral vectors are randomly partitioned into training, validation, and testing vector set.

Afterward, the training and validation data is passed to LSTM, a deep learning framework to learn the short and long-time dependencies of the input. Finally, the trained model with learned dependencies is obtained to which the testing data is passed. The trained model predicts new values which are compared with the original ground truth values to examine the robustness of the built model.

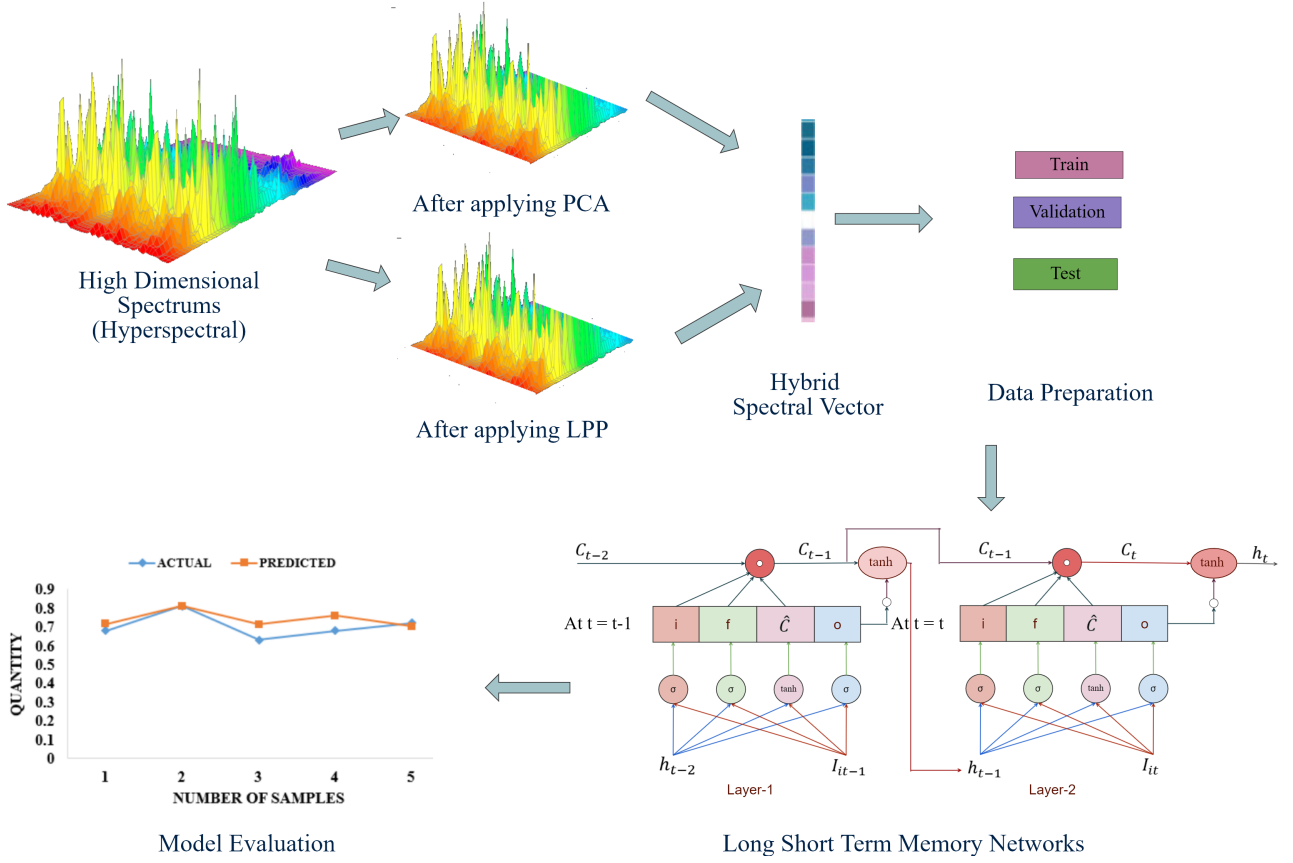


Figure 6.1: Proposed Hybrid Framework

6.3 Results and Analysis

The designed experiments are tested *via* Google Colaboratory environment with GPU enabled. It is a free jupyter notebook environment which runs completely on the cloud. 19036 samples have been collected from LUCAS. The dataset contains 4199 spectral values in the range of 400-2500 nm. The readings that are captured contain instrumental artifacts from 400-500nm. So, the values contained in this region are removed. Now, the high dimensional HSD containing 3999 spectral values in each signature is passed through PCA and LPP. These reduced the dimensions output 1000 values each. The output values obtained from PCA and LPP provides the global and local approximation of the LUCAS dataset. Afterward, both outputs are merged together to design hybrid features, which contain total 2000 values. Then, these features are partitioned into training, validation and testing sets. 50% Training and 25% validation sets are passed to LSTM for effective training. Finally, the trained model is achieved to which the remaining 25% testing data is passed. On the basis of input testing data, the model predicts the properties of soil. The predicted output is contrasted with the actual value for evaluating the performance of the model. Parameter set used in LSTM is shown in Table 6.1

The loss of training and validation curves of LSTM over 100 epochs are shown in Figure 6.2. The loss rate of the training and validation is nearly similar, indicating a good fit in all cases, particularly

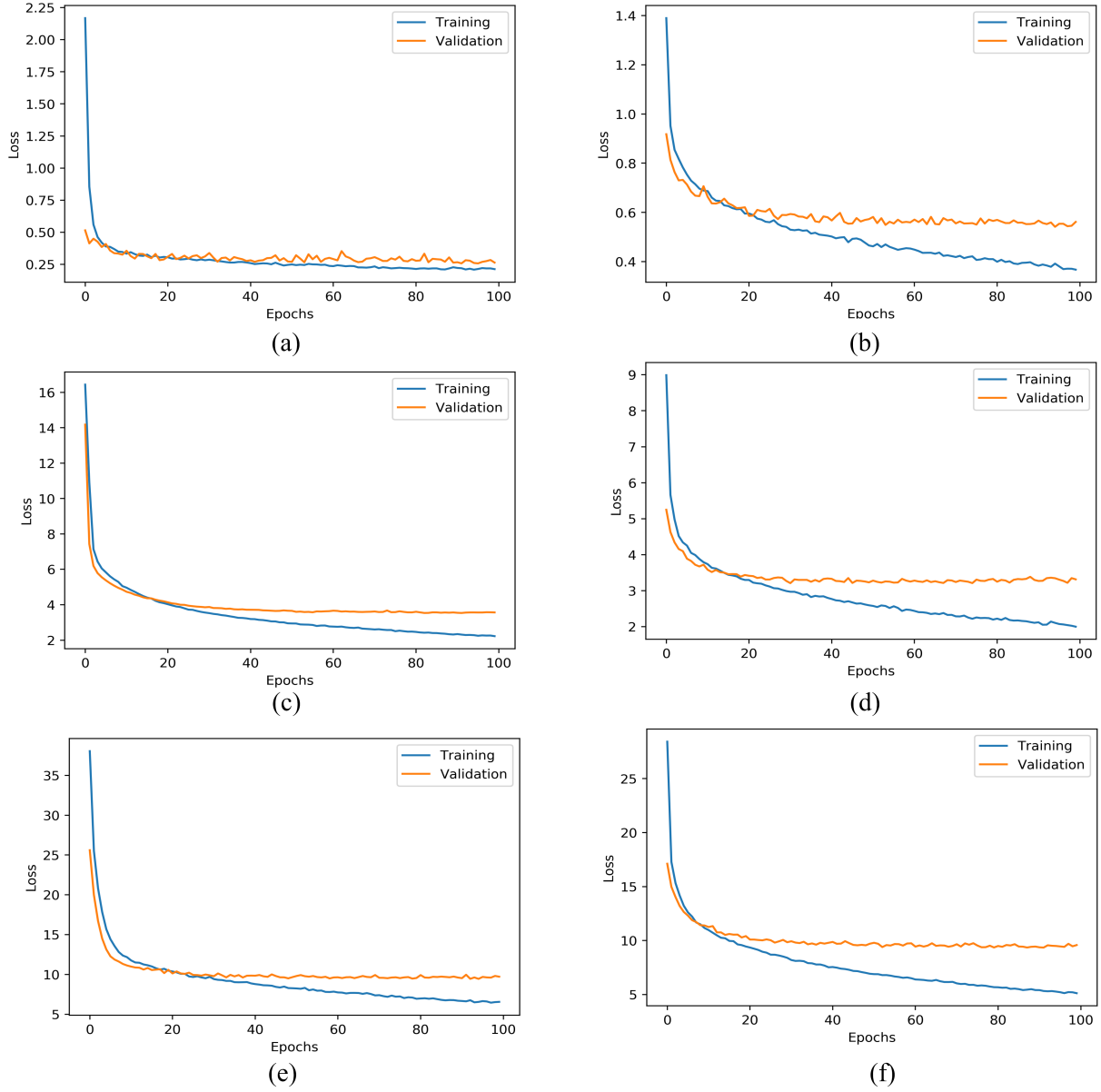


Figure 6.2: Mean absolute Error (loss function) in training and validation (a) pH(H_2O) (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand

Table 6.1: Parameters/ Hyperparameter set in LSTM

Parameters	Description	Activation
Output gate	—	Sigmoid
Epoch	100	—
No. of LSTM layers	1	—
No. of Dense layers	1	Rectified Linear Unit
Dropout	0.1	—
LSTM units/ hidden neurons	2500	—
Optimiser	Adam	—
Loss Function	Mean Absolute Error	—
Dense Layer neurons	1	—
Hidden layer	—	Tanh
Input gate	—	Sigmoid and Tanh
Dense Layer parameters	2501	—
Learning rate	0.001	—
Forget gate	—	Sigmoid
Batch size	100	—
Training %	50	—
Validation %	25	—
Testing %	25	—

in the case of OC and pH (H_2O).

The results predicted by the proposed HF is recorded and contrasted with actual values for model assessment. The metrics employed for evaluation are RMSE, ρ , R^2 , r , MAE and ρ_C . The proposed HF performance is contrasted with other techniques such as PLSR, ANN, SVR and PPLF. Table 6.2 shows that the properties selected for quantity estimation are OC, CEC, Clay, Nitrogen, $pH(H_2O)$ and Sand. HF performance exceeds other models and it has the least error value of 0.36 in case of $pH(H_2O)$ and best R^2 of 0.95 in case of OC. Mean absolute readings are also in line with the RMSE readings with the least absolute error 0.27. The resultant values of ρ , r and ρ_C are very similar to PPLF. The proposed HF also outperforms the PPLF inferring that the local features extracted via LPP play a very crucial role in combination with global features extracted from PCA.

The validation of HF is done by comparing its results with the present models such as Regression-Kriging (Aksoy et al. 2016), Machine learning models (Stevens et al. 2013), Generalized Additive Model (De Brogniez et al. 2015), CNN and Multi CNN (Padarian et al. 2019b) and it is also compared with PPLF. They also employed LUCAS dataset in their prediction models, and their models also selected the data randomly for training and testing purposes. Table 6.3 presents the comparison of HF against earlier studies and R^2 evaluation metrics are employed.

Table 6.3 shows that the HF outperform the previous works because hybrid features constitute of local and global approximation of the LUCAS and the employed LSTM can understand the short- and long-term dependencies. In addition to this, previously proposed model PPLF also performed well.

Table 6.2: Hybrid Framework comparison with other models based on six evaluation metrics.

		HF	PPLF	SVR	PLSR	ANN
pH(H_2O)	RMSE	0.36	0.39	0.44	0.43	0.40
	R^2	0.928	0.917	0.894	0.898	0.913
	MAE	0.27	0.30	0.33	0.33	0.30
	ρ	0.96	0.96	0.95	0.95	0.96
	r	0.96	0.96	0.95	0.95	0.96
	ρ_C	0.96	0.96	0.94	0.95	0.95
Nitrogen	RMSE	0.99	1.05	1.58	1.51	1.23
	R^2	0.926	0.918	0.812	0.828	0.886
	MAE	0.55	0.56	0.80	1.02	0.74
	ρ	0.91	0.91	0.84	0.76	0.83
	r	0.96	0.96	0.90	0.91	0.94
	ρ_C	0.96	0.96	0.89	0.91	0.94
Clay	RMSE	5.28	5.35	8.99	6.81	6.43
	R^2	0.841	0.837	0.538	0.735	0.764
	MAE	3.57	3.63	5.95	5.07	4.65
	ρ	0.93	0.93	0.79	0.88	0.89
	r	0.92	0.92	0.74	0.86	0.88
	ρ_C	0.91	0.91	0.70	0.86	0.87
CEC	RMSE	6.08	6.12	9.61	7.80	6.95
	R^2	0.810	0.807	0.524	0.687	0.751
	MAE	3.34	3.37	5.09	5.29	4.29
	ρ	0.92	0.92	0.82	0.82	0.87
	r	0.90	0.90	0.76	0.83	0.87
	ρ_C	0.90	0.90	0.64	0.83	0.86
OC	RMSE	20.03	21.60	50.10	29.97	20.40
	R^2	0.951	0.943	0.696	0.891	0.950
	MAE	9.58	10.20	20.70	19.97	11.80
	ρ	0.91	0.90	0.81	0.72	0.83
	r	0.98	0.97	0.89	0.94	0.97
	ρ_C	0.98	0.97	0.78	0.94	0.97
Sand	RMSE	13.34	14.32	20.64	17.00	16.60
	R^2	0.758	0.721	0.421	0.608	0.626
	MAE	9.70	10.47	16.29	13.33	12.60
	ρ	0.88	0.88	0.65	0.79	0.81
	r	0.88	0.88	0.65	0.78	0.81
	ρ_C	0.88	0.86	0.60	0.77	0.80

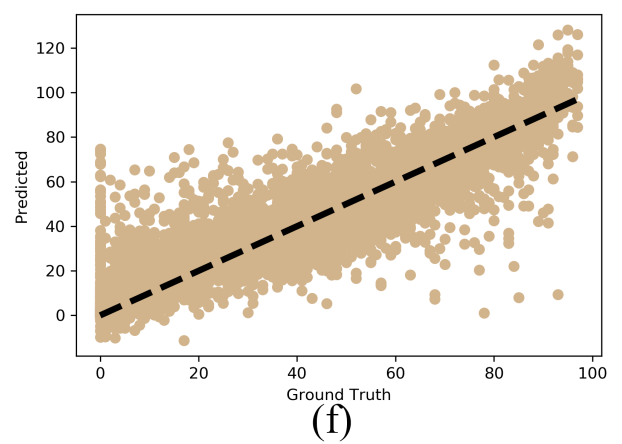
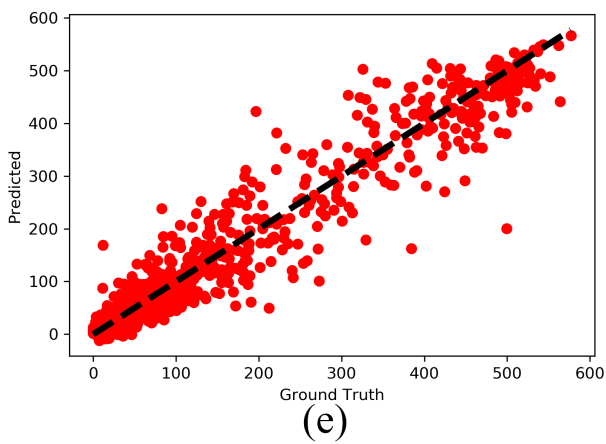
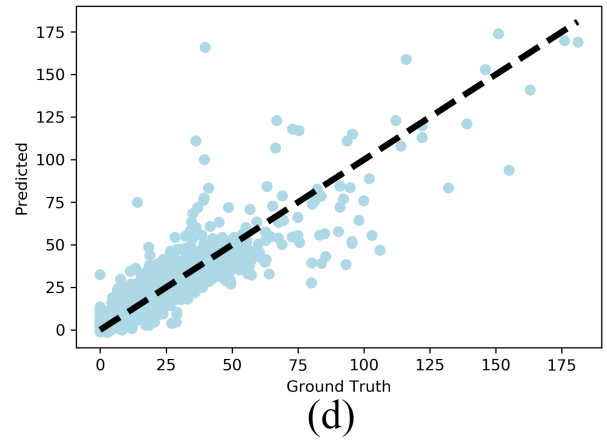
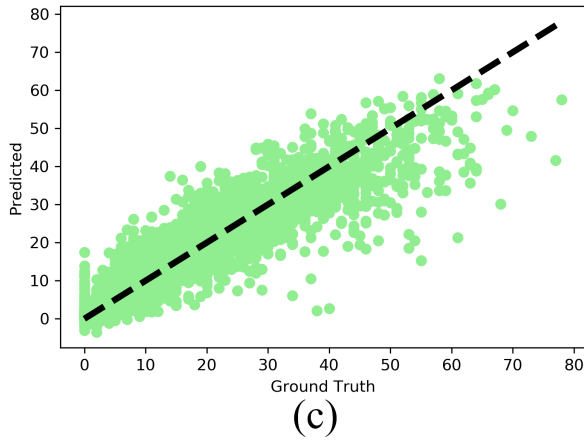
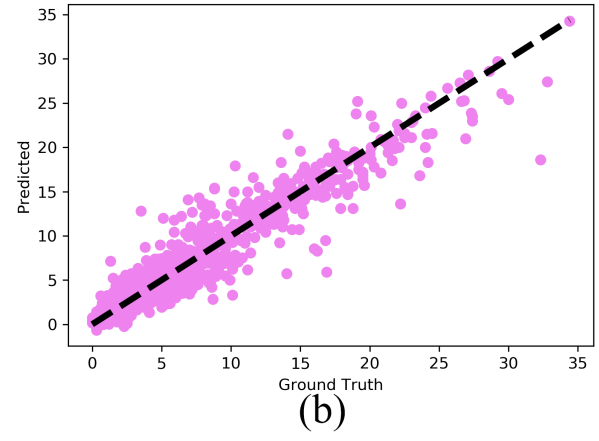
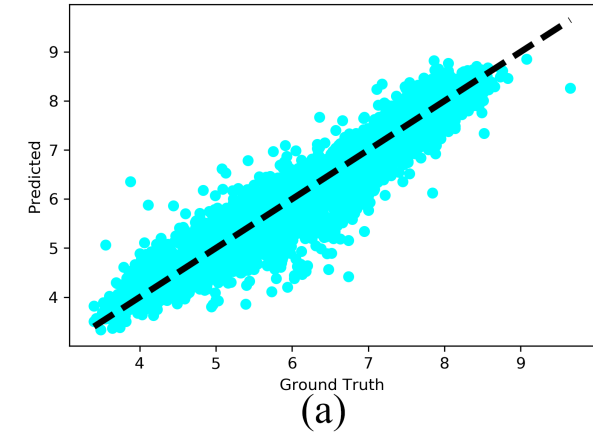
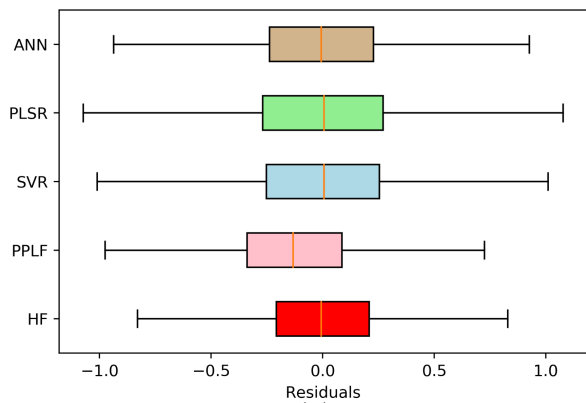
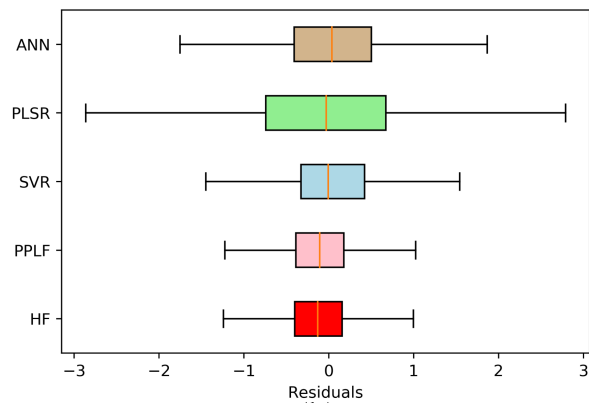


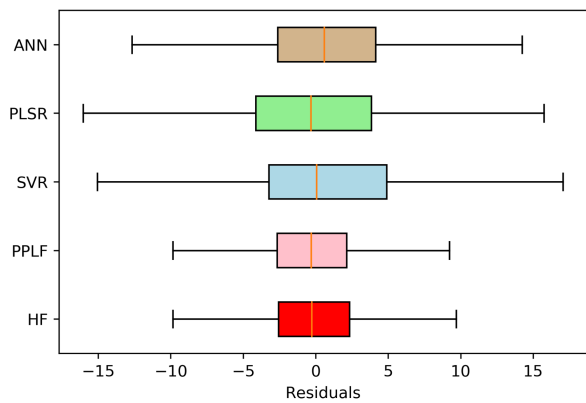
Figure 6.3: Predicted vs Ground Truth graphs calculated from HF (a) $\text{pH}(\text{H}_2\text{O})$ (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand



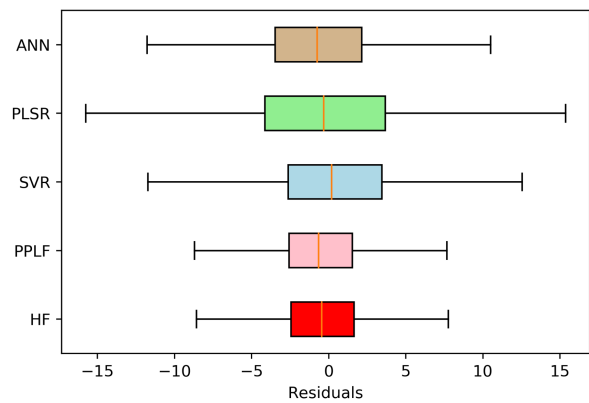
(a)



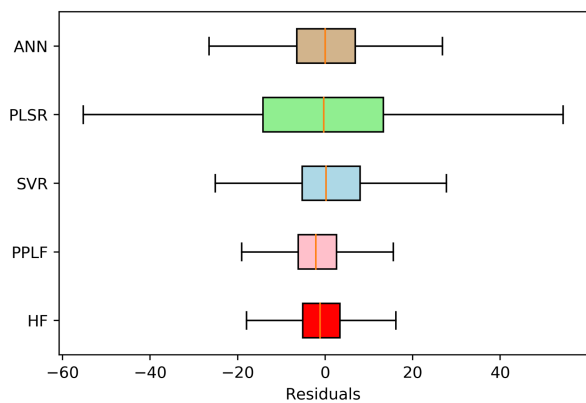
(b)



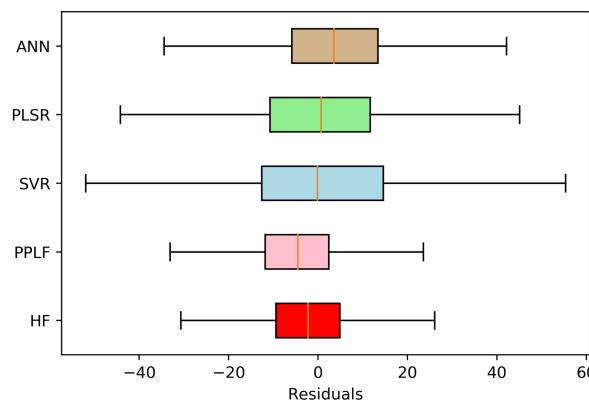
(c)



(d)



(e)



(f)

Figure 6.4: Residuals (a) $pH(H_2O)$ (b) Nitrogen (c) Clay (d) CEC (e) OC (f) Sand

Table 6.3: Earlier studies comparison with HF on Test set and R^2

Properties	HF	PPLF	PPLF	Stevens et al. (2013)	Brogniez et al. (2015)	Aksoy et al. (2016)	Padarian et al. (2019a)	Padarian et al. (2019a)
pH(H_2O)	0.928	0.917	0.90	—	—	—	0.87	0.84
Nitrogen	0.926	0.918	0.91	—	—	—	0.83	0.60
Clay	0.841	0.837	0.80	—	—	—	0.70	0.68
CEC	0.810	0.807	0.77	—	—	—	0.66	0.63
OC	0.95	0.94	0.94	0.78	0.28	0.41	0.88	0.69
Sand	0.76	0.72	0.71	—	—	—	0.53	0.59
Testing	25%	25%	50%	25%	15%	25%	25%	25%

Two versions of PPLF is compared with the existing works. The first version uses 25% testing data, and the second version uses 50% testing data. CNN and Multi-CNN are deep models, that cannot exploit the sequential nature of LUCAS. Compared to the CNN version, the HF most improvement in R^2 is observed in case of sand. Highest R^2 value of 0.95 is obtained in the case of OC. The residuals represent the error that can be obtained by subtracting the predicted quantity from the actual quantity. The HF shows the least residual error, as shown in Figure 6.4 and actual vs predicted comparison is shown in Figure 6.3.

The LUCAS topsoil data collection comprises samples from different forms of land use, regions, management and climatic influences. It is one of the most extensive soil dataset that provides geographic diversity of various soil characteristics. The PPLF is able to predict the right values of soil's various characteristics, with a high degree of precision and minimal error rate, inferring that deep learning could allow us to understand the diverse nature of European land soils. Prediction of important soil attributes allows the productive utilisation of agricultural farms. The HF and PPLF models are able to learn effectively due to the reduced characteristics. The data present in the spectral library would provide a fast approximation of discussed properties if fed into the trained model, which can assist in determining which crops will be better at a certain time for a certain field in the EU. The nitrogen-rich farm, for instance is perfect for plants like Ornamental kale, Rhubarb, spinach *etc* that allow effective utilisation of the farmland.

6.4 Conclusion of the Chapter

A hybrid framework for quantity estimation of soil attributes is proposed in this chapter. The obtained data from LUCAS is converted into hybrid data by employing PCA and LPP. Afterward, the LSTM is employed for building a prediction model based on hybrid features. The proposed HF shows the most improvement of 180.05% in R^2 and has atmost 2.5 times less RMSE. The HF is compared against the previously built PPLF to show that the features obtained from the LPP and PCA help in more

efficient learning of LSTM as these features contains both the global and local spatial information. Deep learning is mostly used in the analysis of data in other fields, and the application of DL is limited in spectral studies. We used LSTM, a deep learning framework in combination with feature manipulation achieved by PCA and LPP to show that the application of DL in spectral studies can achieve good results in comparison with the existing popular methods. This study demonstrates the efficacy of deep models in spectral analysis.

Chapter 7

Conclusions and Future Scope

This research lists varying methods to quantify the soil properties and classify the spectral data efficiently. So, this chapter presents the broad conclusions and the future scope of the study conducted.

With the recent advances in deep learning, its utilisation is ubiquitous in every domain for data analysis. It generates the best results when the dataset is big. This research work is performed mainly for two purposes. Firstly, to efficiently classify the HSIs by utilising feature manipulation and SAEs. Secondly, to quantify the soil characteristics *via* spectral data with the help of LSTM, a deep learning framework. In addition to this, dimensionality reduction algorithms like PCA, LPP are also used to efficiently reduce the higher dimensional data before passing the input for training. All the frameworks proposed are compared with the existing techniques to show their effectiveness, and all outperform others in comparison. Apart from this, the utilisation of deep learning in soil studies for estimating the soil properties is limited. This study shows that deep learning can interpret the heterogeneity of different soil profiles *via* its spectra and can accurately predict the various properties corresponding to soil with significant precision and minimal error rates.

7.1 Conclusions

This thesis comprises of 7 chapters. In Chapter 1, Introduction, along with the problem statement, is presented. In Chapter 2, the background of the research work and literature survey related to classification and soil quantification using hyperspectral data is carried out. In chapter 3, a deep learning framework constitutes of SAEs and LPP is employed for classification of HSI. The proposed framework outperforms 21 machine learning algorithms. With the utilisation of LPP, the feature set is reduced due to which the DNN processes the data quickly and the processing time is decreased drastically. The results generated by LPP with the accuracy of around 84.4 and 87.2% which leads to 236.41 and 275.08% improvement in overall accuracy in case of Indian pines and Salinas dataset respectively.

In chapter 4, a pre-processing framework for classification of HSIs is proposed. The proposed

framework takes 4.3, 5.3, 2 and 4.7 times less processing time and it shows overall accuracy improvement of 115.5, 209.4, 119 and 131 % in case of KSC, Salinas, Pavia University and Indian Pines dataset respectively. The proposed framework is able to perform well when contrasted with popular spatial-spectral and spectral frameworks.

In chapter 5, PPLF named framework is proposed for the quantification of soil characteristics. The PPLF relies on PCA and LSTM to determine the spectral characteristics of the quantities of European soil resources. The PPLF showed atmost improvement of 111%, 364%, 163%, 164%, 588%, 187% in R^2 in case of $\text{pH}(H_2O)$, nitrogen, clay, CEC, OC, sand respectively. This chapter reveals that deep learning is an effective method for the processing of the spectral data. The developed model proved to be efficient in estimating the soil characteristics as compared to conventional models.

In chapter 6, a Hybrid Framework for the quantification of soil characteristics is proposed. The proposed HF shows the most improvement of 180.05% in R^2 and has atmost 2.5 times less RMSE. The HF is compared against the previously built PPLF to show that the features obtained from the LPP and PCA help in more efficient learning of LSTM as these features contains both the global and local spatial information. Deep learning is mostly used in the analysis of data in other fields, and the application of DL is limited in spectral studies. We used LSTM, a deep learning framework in combination with feature manipulation achieved by PCA and LPP to show that the application of DL in spectral studies can achieve good results in comparison with the existing popular methods. This study demonstrates the efficacy of deep models in spectral analysis.

7.2 Future Scope

This research can be extended with more efficient classification and quantification of spectral data techniques. It would be beneficial to extend this work on the following topics further:

1. Classification studies of hyperspectral images and quantification of soil characteristics using more advanced quantum learning models.
2. Developing a quantification technique for quick estimation of soil minerals using satellite imagery.
3. Developing an affordable hyperspectral camera and embedding the trained quantum models into it to avoid the high-cost laboratory testing of soil for estimating its various nutrients.
4. Designing an efficient dimensionality reduction algorithm that can learn more efficient compressed features.

Bibliography

- AbdelFattah, M., AbdelAal, L., and El-khoribi, R. (2018). Spectral-spatial hyperspectral image classification based on randomized singular value decomposition and 3-dimensional discrete wavelet transform. *International Journal of Computer Applications*, Vol. 180(30):pages 1–10.
- Adeline, K., Gomez, C., Gorretta, N., and Roger, J.-M. (2017). Predictive ability of soil properties to spectral degradation from laboratory vis-nir spectroscopy data. *Geoderma*, Vol. 288:pages 143–153.
- Adhiwibawa, M. A., Setiawan, Y. E., Prilianti, K. R., and Brotosudarmo, T. H. (2015). Web camera as low cost multispectral sensor for quantification of chlorophyll in soybean leaves. Vol. 9444:pages 94440G (1–6).
- Agbu, P. A., Fehrenbacher, D. J., and Jansen, I. J. (1990). Soil property relationships with spot satellite digital data in east central illinois. *Soil Science Society of America Journal*, Vol. 54(3):pages 807–812.
- Aitkenhead, M., Cameron, C., Gaskin, G., Choisy, B., Coull, M., and Black, H. (2018). Digital rgb photography and visible-range spectroscopy for soil composition analysis. *Geoderma*, Vol. 313:pages 265–275.
- Aksoy, E., Yigini, Y., and Montanarella, L. (2016). f soil databases for topsoil organic carbon mapping in europe. *PloS one*, Vol. 11(3):pages E0152098 (1–17).
- Al Maliki, A., Bruce, D., and Owens, G. (2012). Capabilities of remote sensing hyperspectral images for the detection of lead contamination: a review. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences. Melbourne*, Vol. 55:pages 55–60.
- Anand, A. and Suganthan, P. (2009). Multiclass cancer classification by support vector machines with class-wise optimized genes and probability estimates. *Journal of theoretical biology*, Vol. 259(3):pages 533–540.
- Arrouays, D., Lagacherie, P., and Hartemink, A. E. (2017). Digital soil mapping across the globe. *Geoderma Regional*, Vol. 9:pages 1–4.

- Bajwa, S. and Tian, L. (2005). Soil fertility characterization in agricultural fields using hyperspectral remote sensing. *Transactions of the American Society of Agricultural and Biological Engineers*, Vol. 48(6):pages 2399–2406.
- Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., and Panagos, P. (2019). Mapping lucas topsoil chemical properties at european scale using gaussian process regression. *Geoderma*, Vol. 355:pages 1–14.
- Ballabio, C., Panagos, P., Lugato, E., Huang, J.-H., Orgiazzi, A., Jones, A., Fernández-Ugalde, O., Borrelli, P., and Montanarella, L. (2018). Copper distribution in european topsoils: An assessment based on lucas soil survey. *Science of The Total Environment*, Vol. 636:pages 282–298.
- Ballabio, C., Panagos, P., and Monatanarella, L. (2016). Mapping topsoil physical properties at european scale using the lucas database. *Geoderma*, Vol. 261:pages 110–123.
- Barnes, E. and Baker, M. (2000). Multispectral data for mapping soil texture: possibilities and limitations. *Applied Engineering in Agriculture*, Vol. 16(6):pages 731–742.
- Baumgardner, M. F., Silva, L. F., Biehl, L. L., and Stoner, E. R. (1986). Reflectance properties of soils. Vol. 38:pages 1–44.
- Behrens, T., Schmidt, K., MacMillan, R. A., and Rossel, R. A. V. (2018). Multi-scale digital soil mapping with deep learning. *Scientific reports*, Vol. 8:pages 1–9.
- Ben-Dor, E. and Banin, A. (1995). Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal*, Vol. 59(2):pages 364–372.
- Ben-Dor, E., Chabrillat, S., Demattê, J., Taylor, G., Hill, J., Whiting, M., and Sommer, S. (2009). Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*, Vol. 113:pages S38–S55.
- Bengio, Y., Simard, P., Frasconi, P., et al. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, Vol. 5(2):pages 157–166.
- Bittar, R. D., Alves, S., De Freitas, M., and Melo, F. R. D. (2018). Estimation of physical and chemical soil properties by artificial neural networks. *Revista Caatinga*, Vol. 31(3):pages 704–712.
- Blum, W. (1993). Soil protection concept of the council of europe and integrated soil research. In *Integrated soil and sediment research: a basis for proper protection*, pages 37–47.
- Bourennane, S., Fossati, C., and Cailly, A. (2010). Improvement of classification for hyperspectral images based on tensor modeling. *IEEE Geoscience and Remote Sensing Letters*, Vol. 7(4):pages 801–805.

- Bowers, S. A. and Hanks, R. (1971). *Reflection of radiant energy from soils*. PhD thesis, Kansas State University.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, Vol. 24(2):pages 123–140.
- Bricklemyer, R., Miller, P., Paustian, K., Keck, T., Nielsen, G., and Antle, J. (2005). Soil organic carbon variability and sampling optimization in montana dryland wheat fields. *Journal of soil and water conservation*, Vol. 60(1):pages 42–51.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, Vol. 2(2):pages 121–167.
- Burnase, M. S. R. and Swamy, M. S. (2002). Hyperspectral image reduction using discrete wavelet transform. *Remote Sensing*, Vol. 40(10):pages 13–16.
- Cai, Y., Guo, Y., Lang, S., Liu, J., and Hu, S. (2020). Classification of hyperspectral images by spectral–spatial dense-residual network. *Journal of Applied Remote Sensing*, Vol. 14(3):pages 1–12.
- Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., and Calpe-Maravilla, J. (2006). Composite kernels for hyperspectral image classification. *IEEE geoscience and remote sensing letters*, Vol. 3(1):pages 93–97.
- Camps-Valls, G., Tuia, D., Bruzzone, L., and Benediktsson, J. A. (2014). Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE signal processing magazine*, Vol. 31(1):pages 45–54.
- Cao, X., Xu, L., Meng, D., Zhao, Q., and Xu, Z. (2017). Integration of 3-dimensional discrete wavelet transform and markov random field for hyperspectral image classification. *Neurocomputing*, Vol. 226:pages 90–100.
- Chabrillat, S., Foerster, S., Steinberg, A., and Segl, K. (2014). Prediction of common surface soil properties using airborne and simulated enmap hyperspectral images: Impact of soil algorithm and sensor characteristic. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 2914–2917.
- Chand, N., Joshi, R. C., and Misra, M. (2006). Supporting cooperative caching in mobile ad hoc networks using clusters. *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 2(1/2):pages 58–72.
- Chang, C.-W., Laird, D. A., Mausbach, M. J., and Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal*, Vol. 65(2):pages 480–490.

- Chang, D.-H. and Islam, S. (2000). Estimation of soil physical properties using remote sensing and artificial neural network. *Remote Sensing of Environment*, Vol. 74(3):pages 534–544.
- Chen, F., Kissel, D. E., West, L. T., Rickman, D., Luvall, J., and Adkins, W. (2005). Mapping surface soil organic carbon for crop fields with remote sensing. *Journal of soil and water Conservation*, Vol. 60(1):pages 51–57.
- Chen, H., Liu, Z., Gu, J., Ai, W., Wen, J., and Cai, K. (2018). Quantitative analysis of soil nutrition based on ft-nir spectroscopy integrated with bp neural deep learning. *Analytical Methods*, Vol. 10(41):pages 5004–5013.
- Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54(10):pages 6232–6251.
- Chen, Y., Zhao, X., and Jia, X. (2015). Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 8(6):pages 2381–2392.
- Chivasa, W., Mutanga, O., and Biradar, C. (2017). Application of remote sensing in estimating maize grain yield in heterogeneous african agricultural landscapes: a review. *International journal of remote sensing*, Vol. 38(23):pages 6816–6845.
- Coleman, T., Agbu, P., and Montgomery, O. (1993). Spectral differentiation of surface soils and soil properties: Is it possible from space platforms? *Soil Science*, Vol. 155(4):pages 283–293.
- Coleman, T., Agbu, P., Montgomery, O., Gao, T., and Prasad, S. (1991). Spectral band selection for quantifying selected properties in highly weathered soils. *Soil Science*, Vol. 151(5):pages 355–361.
- Condit, H. (1970). The spectral reflectance of american soils. *Photogrammetric Engineering and Remote Sensing*, Vol. 36:pages 955–956.
- Condit, H. (1972). Application of characteristic vector analysis to the spectral energy distribution of daylight and the spectral reflectance of american soils. *Applied optics*, Vol. 11(1):pages 74–86.
- Coopersmith, E. J., Cosh, M. H., Bell, J. E., and Boyles, R. (2016). Using machine learning to produce near surface soil moisture estimates from deeper in situ records at us climate reference network (uscrn) locations: Analysis and applications to amsr-e satellite validation. *Advances in water resources*, Vol. 98:pages 122–131.
- Dalal, R. and Henry, R. (1986). Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Science Society of America Journal*, Vol. 50(1):pages 120–123.

- Daniel, K., Tripathi, N., and Honda, K. (2003). Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of lop buri (thailand). *Soil Research*, Vol. 41(1):pages 47–59.
- Das, B., Sarathjith, M., Santra, P., Sahoo, R., Srivastava, R., Routray, A., and Ray, S. (2015). Hyperspectral remote sensing: opportunities, status and challenges for rapid soil assessment in india. *Current Science*, pages 860–868.
- De Brogniez, D., Ballabio, C., Stevens, A., Jones, R., Montanarella, L., and van Wesemael, B. (2015). A map of the topsoil organic carbon content of europe generated by a generalized additive model. *European Journal of Soil Science*, Vol. 66(1):pages 121–134.
- Demattê, J., Galdos, M., Guimarães, R., Genú, A., Nanni, M., and Zullo Jr, J. (2007). Quantification of tropical soil attributes from etm+/landsat-7 data. *International Journal of Remote Sensing*, Vol. 28(17):pages 3813–3829.
- Demattê, J. A., Fiorio, P. R., and Araújo, S. R. (2010). Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. *Remote Sensing*, Vol. 2(8):pages 1998–2016.
- Deng, F., Pu, S., Chen, X., Shi, Y., Yuan, T., and Pu, S. (2018). Hyperspectral image classification with capsule network using limited training samples. *Sensors*, Vol. 18(9):pages 1–22.
- Desaules, A. (2012). Critical evaluation of soil contamination assessment methods for trace metals. *Science of the Total Environment*, Vol. 426:pages 120–131.
- Doolette, A., Smernik, R., and Dougherty, W. (2011). A quantitative assessment of phosphorus forms in some australian soils. *Soil Research*, Vol. 49(2):pages 152–165.
- Dotto, A. C., Dalmolin, R. S. D., ten Caten, A., and Grunwald, S. (2018). A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by vis-nir spectra. *Geoderma*, Vol. 314:pages 262–274.
- Fang, L., Li, S., Duan, W., Ren, J., and Benediktsson, J. A. (2015). Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. *IEEE transactions on geoscience and remote sensing*, Vol. 53(12):pages 6663–6674.
- Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., and Tilton, J. C. (2013). Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, Vol. 101(3):pages 652–675.

- Feng, J., Yu, H., Wang, L., Cao, X., Zhang, X., and Jiao, L. (2019). Classification of hyperspectral images based on multiclass spatial–spectral generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57(8):pages 5329–5343.
- Fidencio, P. H., Poppi, R. J., and de Andrade, J. C. (2002). Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Analytica Chimica Acta*, Vol. 453(1):pages 125–134.
- Forkuor, G., Hounkpatin, O. K., Welp, G., and Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: a comparison of machine learning and multiple linear regression models. *PloS one*, Vol. 12(1):pages E0170478(1–21).
- Garfagnoli, F., Ciampalini, A., Moretti, S., Chiarantini, L., and Vettori, S. (2013). Quantitative mapping of clay minerals using airborne imaging spectroscopy: new data on mugello (italy) from sim-ga prototypal sensor. *European Journal of Remote Sensing*, Vol. 46(1):1–17.
- Ge, Y., Morgan, C., Thomasson, J., and Waiser, T. (2007). A new perspective to near-infrared reflectance spectroscopy: A wavelet approach. *Transactions of the ASABE*, Vol. 50(1):pages 303–311.
- Ghamisi, P., Benediktsson, J. A., and Sveinsson, J. R. (2013a). Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52(9):pages 5771–5782.
- Ghamisi, P., Couceiro, M. S., Martins, F. M., and Benediktsson, J. A. (2013b). Multilevel image segmentation based on fractional-order darwinian particle swarm optimization. *IEEE Transactions on Geoscience and Remote sensing*, Vol. 52(5):pages 2382–2394.
- Gholizadeh, A., Borůvka, L., Saberioon, M., and Vašát, R. (2013). Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Applied spectroscopy*, Vol. 67(12):pages 1349–1362.
- Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G., and Fernandes Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in brazil. *Geoderma*, Vol. 340:pages 337–350.
- Hang, R., Liu, Q., Hong, D., and Ghamisi, P. (2019). Cascaded recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, pages 5384 – 5394.
- Hang, R., Zhou, F., Liu, Q., and Ghamisi, P. (2020). Classification of hyperspectral images via multitask generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–13.

- Harris, J., Cousins, C., and Claire, M. (2016). Spectral identification and quantification of salts in the atacama desert. In *Earth Resources and Environmental Remote Sensing/GIS Applications VII*, volume Vol. 10005, pages 100050I(1–16).
- Harris, R. K. (1986). Nuclear magnetic resonance spectroscopy.
- He, X., Cai, D., and Min, W. (2005). Statistical and computational analysis of locality preserving projection. In *Proceedings of the 22nd international conference on Machine learning*, pages 281–288.
- He, X. and Niyogi, P. (2004). Locality preserving projections. In *Advances in neural information processing systems*, pages 153–160.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, Vol. 9(8):pages 1735–1780.
- Hutchinson, J. (2003). Estimating near-surface soil moisture using active microwave satellite imagery and optical sensor inputs. *Transactions of the American Society of Agricultural and Biological Engineers*, Vol. 46(2):pages 225–236.
- Islam, K., Singh, B., and McBratney, A. (2003). Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Soil Research*, Vol. 41(6):pages 1101–1114.
- Jenny, H. (2012). *The soil resource: origin and behavior*, volume 37. Springer Science & Business Media.
- Kadupitiya, H., Sahoo, R., Ray, S., Chakraborty, D., and Ahmed, N. (2010). Quantitative assessment of soil chemical properties using visible (vis) and near-infrared (nir) proximal hyperspectral data. *Trop. Agric*, Vol. 158:pages 41–60.
- Kågeback, M. and Salomonsson, H. (2016). Word sense disambiguation using a bidirectional lstm. *arXiv preprint arXiv:1606.03568*, pages 51–56.
- Kang, X., Li, S., and Benediktsson, J. A. (2013). Spectral-spatial hyperspectral image classification with edge-preserving filtering. *IEEE transactions on geoscience and remote sensing*, Vol. 52(5):pages 2666–2677.
- Kang, X., Zhuo, B., and Duan, P. (2019). Semi-supervised deep learning for hyperspectral image classification. *Remote Sensing Letters*, Vol. 10(4):pages 353–362.
- Kearns, M. (1988). Learning boolean formulae or finite automata is as hard as factoring. *Technical Report TR-14-88 Harvard University Aikem Computation Laboratory*.

- Khedri, E., Hasanlou, M., and Tabatabaeenejad, A. (2017). Estimating soil moisture using polsar data: A machine learning approach. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 42:pages 133–137.
- Kovačević, M., Bajat, B., and Gajić, B. (2010). Soil type classification and estimation of soil properties using support vector machines. *Geoderma*, Vol. 154(4):pages 340–347.
- Krishnan, P., Alexander, J. D., Butler, B., and Hummel, J. W. (1980). Reflectance technique for predicting soil organic matter. *Soil Science Society of America Journal*, Vol. 44(6):pages 1282–1285.
- Kristof, S. (1971). Preliminary multispectral studies of soils. *Journal of soil and water conservation*, pages 15–18.
- Kumar, S. and Rani, A. (2013). Df-lda tree: a nonlinear multilevel classifier for pattern recognition. *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 25(2):pages 177–188.
- Lee, C. s., Sohn, E., Park, J. D., and Jang, J.-D. (2019). Estimation of soil moisture using deep learning based on satellite data: a case study of south korea. *GIScience & remote sensing*, Vol. 56(1):pages 43–67.
- Li, C., Ma, Y., Mei, X., Liu, C., and Ma, J. (2016a). Hyperspectral image classification with robust sparse representation. *IEEE Geoscience and Remote Sensing Letters*, Vol. 13(5):pages 641–645.
- Li, H., Leng, W., Zhou, Y., Chen, F., Xiu, Z., and Yang, D. (2014). Evaluation models for soil nutrient based on support vector machine and artificial neural networks. *The Scientific World Journal*, Vol. 2014:pages 1–7.
- Li, J., Bioucas-Dias, J. M., and Plaza, A. (2012). Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Transactions on Geoscience and remote sensing*, Vol. 51(2):pages 844–856.
- Li, L., Yang, C., and Xie, Q. (2016b). 1d embedding multi-category classification methods. *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 14(02):pages 1640006 (1–13).
- Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., and Clinton, N. (2016c). Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of african land-cover mapping. *International journal of remote sensing*, Vol. 37(23):pages 5632–5646.
- Liang, Y., Zhao, X., Guo, A. J., and Zhu, F. (2019). Hyperspectral image classification with deep metric learning and conditional random field. *IEEE Geoscience and Remote Sensing Letters*, Vol. 17(6):pages 1042–1046.

- Liao, K., Xu, S., Wu, J., and Zhu, Q. (2013). Spatial estimation of surface soil texture using remote sensing data. *Soil Science and Plant Nutrition*, Vol. 59(4):pages 488–500.
- Lin, Z., Chen, Y., Zhao, X., and Wang, G. (2013). Spectral-spatial classification of hyperspectral image using autoencoders. In *2013 9th International Conference on Information, Communications & Signal Processing*, pages 1–5. IEEE.
- Liu, K.-H., Lin, Y.-Y., and Chen, C.-S. (2014). Linear spectral mixture analysis via multiple-kernel learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53(4):pages 2254–2269.
- Liu, L., Ji, M., and Buchroithner, M. (2017a). Combining partial least squares and the gradient-boosting method for soil property retrieval using visible near-infrared shortwave infrared spectra. *Remote Sensing*, Vol. 9(12):pages 1–18.
- Liu, L., Ji, M., and Buchroithner, M. (2018). Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors*, Vol. 18(9):pages 1–18.
- Liu, P., Zhang, H., and Eom, K. B. (2016). Active deep learning for classification of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10(2):pages 712–724.
- Liu, X., Sun, Q., Liu, B., Huang, B., and Fu, M. (2017b). Hyperspectral image classification based on convolutional neural network and dimension reduction. In *2017 Chinese Automation Congress (CAC)*, pages 1686–1690.
- Liu, Y., Cao, G., Sun, Q., and Siegel, M. (2015). Hyperspectral classification via deep networks and superpixel segmentation. *International Journal of Remote Sensing*, Vol. 36(13):pages 3459–3482.
- Luce, M. S., Ziadi, N., Zebbarth, B. J., Grant, C. A., Tremblay, G. F., and Gregorich, E. G. (2014). Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. *Geoderma*, Vol. 232:pages 449–458.
- Ma, A., Filippi, A. M., Wang, Z., and Yin, Z. (2019). Hyperspectral image classification using similarity measurements-based deep recurrent neural networks. *Remote Sensing*, Vol. 11(2):pages 1–19.
- Meier, M., Souza, E. d., Francelino, M. R., Fernandes Filho, E. I., and Schaefer, C. E. G. R. (2018). Digital soil mapping using machine learning algorithms in a tropical mountainous area. *Revista Brasileira de Ciência do Solo*, Vol. 42:pages 1–22.

- Minasny, B. and Hartemink, A. E. (2011). Predicting soil properties in the tropics. *Earth-Science Reviews*, Vol. 106(1-2):pages 52–62.
- Morellos, A., Pantazi, X.-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebenson, J., Bill, R., and Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-nir spectroscopy. *Biosystems Engineering*, Vol. 152:pages 104–116.
- Morra, M., Hall, M., and Freeborn, L. (1991). Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. *Soil Science Society of America Journal*, Vol. 55(1):pages 288–291.
- Mou, L., Ghamisi, P., and Zhu, X. X. (2017). Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55(7):pages 3639–3655.
- Mouazen, A., Kuang, B., De Baerdemaeker, J., and Ramon, H. (2010). Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, Vol. 158(1-2):pages 23–31.
- Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., and Mouazen, A. M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-nir diffuse reflectance spectroscopy. *Soil and Tillage Research*, Vol. 155:pages 510–522.
- Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., and Hossard, L. (2018). Deep recurrent neural network for agricultural classification using multitemporal sar sentinel-1 for camargue, france. *Remote Sensing*, Vol. 10(8):pages 1217.
- Odhiambo, L. O., Freeland, R., Yoder, R., and Hines, J. (2004). Investigation of a fuzzy-neural network application in classification of soils using ground-penetrating radar imagery. *Applied engineering in agriculture*, Vol. 20(1):pages 109.
- Oliver, M. (2010). Nj mckenzie, mj grundy, r. webster, and aj ringrose-voase (eds): Guidelines for surveying soil and land resources. *Precision Agriculture*, Vol. 11(1):pages 100–102.
- O'Rourke, S. and Holden, N. (2011). Optical sensing and chemometric analysis of soil organic carbon—a cost effective alternative to conventional laboratory methods? *Soil Use and Management*, Vol. 27(2):pages 143–155.
- Özdemir, A. O. B., Gedik, B. E., and Çetin, C. Y. Y. (2014). Hyperspectral classification using stacked autoencoders with deep learning. In *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. IEEE.

- Padarian, J., Minasny, B., and McBratney, A. (2019a). Transfer learning to localise a continental soil vis-nir calibration model. *Geoderma*, 340:pages 279–288.
- Padarian, J., Minasny, B., and McBratney, A. (2019b). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, Vol. 16:pages E00198 (1–9).
- Padarian, J., Minasny, B., and McBratney, A. B. (2019c). Using deep learning for digital soil mapping. *Soil*, Vol. 5(1):pages 79–89.
- Pan, E., Mei, X., Wang, Q., Ma, Y., and Ma, J. (2020). Spectral-spatial classification for hyperspectral image based on a single gru. *Neurocomputing*, Vol. 387:pages 150–160.
- Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318.
- Phillips, R. D., Blinn, C. E., Watson, L. T., and Wynne, R. H. (2009). An adaptive noise-filtering algorithm for aviris data with implications for classification accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47(9):pages 3168–3179.
- Plaza, A., Martinez, P., Perez, R., and Plaza, J. (2004). A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles. *Pattern Recognition*, Vol. 37(6):pages 1097–1116.
- Plaza, A., Martinez, P., Plaza, J., and Perez, R. (2005). Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Transactions on Geoscience and remote sensing*, Vol. 43(3):pages 466–479.
- Pompilio, L., Pepe, M., Pedrazzi, G., and Marinangeli, L. (2014). Informational clustering of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7(6):pages 2209–2223.
- Puno, J. C., Sybingco, E., Dadios, E., Valenzuela, I., and Cuello, J. (2017). Determination of soil nutrients and ph level using image processing and artificial neural network. In *Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2017 IEEE 9th International Conference on*, pages 1–6.
- Qi, Z., Wang, B., Tian, Y., and Zhang, P. (2016). When ensemble learning meets deep learning: a new deep support vector machine for classification. *Knowledge-Based Systems*, Vol. 107:pages 54–60.
- Qiao, T., Ren, J., Wang, Z., Zabalza, J., Sun, M., Zhao, H., Li, S., Benediktsson, J. A., Dai, Q., and Marshall, S. (2016). Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis. *IEEE transactions on geoscience and remote sensing*, Vol. 55(1):pages 119–133.

- Rani, A., Kumar, S., Micheloni, C., and Foresti, G. L. (2013). Incorporating linear discriminant analysis in neural tree for multidimensional splitting. *Applied Soft Computing*, Vol. 13(10):pages 4219–4228.
- Robila, S. (2005). An investigation of spectral metrics in hyperspectral image preprocessing for classification. In *Geospatial goes global: from your neighborhood to the whole planet. ASPRS Annual Conference, Baltimore, Maryland*, pages 7–11.
- Rossel, R. V. (2007). Robust modelling of soil diffuse reflectance spectra by "bagging-partial least squares regression". *Journal of Near Infrared Spectroscopy*, Vol. 15(1):pages 39–47.
- Rossel, R. V., Walvoort, D., McBratney, A., Janik, L. J., and Skjemstad, J. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, Vol. 131(2):pages 59–75.
- Safanelli, J. L., Chabrillat, S., Ben-Dor, E., and Dematté, J. A. M. (2020). Multispectral models from bare soil composites for mapping topsoil properties over europe. *Remote Sensing*, Vol. 12(9):pages 1–19.
- Sanasam, R., Murthy, H., and Gonsalves, T. (2010). Feature selection for text classification based on gini coefficient of inequality. In *Feature Selection in Data Mining*, pages 76–85.
- Sarajlic, S. (2012). Land cover change and mineral composite assessment of tushka depression, in egypt, using remote sensing and gis. *Sanford Bederman Research Award*, pages 1–14.
- Sarathjith, M., Das, B. S., Wani, S. P., and Sahrawat, K. L. (2016). Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma*, Vol. 267:pages 1–9.
- Savitzky, A. and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, Vol. 36(8):pages 1627–1639.
- Selige, T., Böhner, J., and Schmidhalter, U. (2006). High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, Vol. 136(1-2):pages 235–244.
- Seydgar, M., Alizadeh Naeini, A., Zhang, M., Li, W., and Satari, M. (2019). 3-d convolution-recurrent networks for spectral-spatial classification of hyperspectral images. *Remote Sensing*, Vol. 11(7):pages 1–21.
- Shahid, S. A., Abdelfattah, M. A., Omar, S. A., Harahsheh, H., Othman, Y., and Mahmoudi, H. (2010). Mapping and monitoring of soil salinization remote sensing, gis, modeling, electromagnetic induction and conventional methods—case studies. In *Proceedings of the International Conference on Soils and Groundwater Salinization in Arid Countries*, volume Vol. 59, pages 59–97.

- Shalizi, C. (2009). The truth about principal components and factor analysis.
- Shepherd, K. D. and Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil science society of America journal*, Vol. 66(3):pages 988–998.
- Singh, S. and Kasana, S. S. (2019). Spectral-spatial hyperspectral image classification using deep learning. In *2019 Amity International Conference on Artificial Intelligence (AICAI)*, pages 411–417. IEEE.
- Slavkovikj, V., Verstockt, S., De Neve, W., Van Hoescke, S., and Van de Walle, R. (2015). Hyperspectral image classification with convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1159–1162.
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., and van Wesemael, B. (2013). Prediction of soil organic carbon at the european scale by visible and near infrared reflectance spectroscopy. *PloS one*, Vol. 8(6):pages E66409 (1–13).
- Sujithra, D., Manickam, T., and Sudheer, D. (2013). Compression of hyperspectral image using discrete wavelet transform and walsh hadamard transform. *International Journal of Advanced Research in Electronics and Communication Engineering*, Vol. 2(3):pages 314–319.
- Sullivan, D. G., Shaw, J., and Rickman, D. (2005). Ikonos imagery to estimate surface soil property variability in two alabama physiographies. *Soil Science Society of America Journal*, Vol. 69(6):pages 1789–1798.
- Sun, L., Wu, Z., Liu, J., Xiao, L., and Wei, Z. (2014). Supervised spectral–spatial hyperspectral image classification with weighted markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53(3):pages 1490–1503.
- Tarabalka, Y., Benediktsson, J. A., and Chanussot, J. (2009). Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47(8):pages 2973–2987.
- Thomasson, J., Sui, R., Cox, M., and Al-Rajehy, A. (2001). Soil reflectance sensing for determining soil properties in precision agriculture. *Transactions of the American Society of Agricultural and Biological Engineers*, Vol. 44(6):pages 1445–1453.
- Tian, Y., Zhang, J., Yao, X., Cao, W., and Zhu, Y. (2013). Laboratory assessment of three quantitative methods for estimating the organic matter content of soils in china based on visible/near-infrared reflectance spectra. *Geoderma*, Vol. 202:pages 161–170.

- Tiwari, S. K., Saha, S. K., and Kumar, S. (2015). Prediction modeling and mapping of soil carbon content using artificial neural network, hyperspectral satellite data and field spectroscopy. *Advances in Remote Sensing*, Vol. 4(01):pages 63–72.
- Tóth, G., Jones, A., and Montanarella, L. (2013). The lucas topsoil database and derived information on the regional variability of cropland topsoil properties in the european union. *Environmental monitoring and assessment*, Vol. 185(9):pages 7409–7425.
- Tsai, F. and Lai, J.-S. (2013). Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence. *IEEE transactions on geoscience and remote sensing*, Vol. 51(6):pages 3504–3513.
- Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., and Zalidis, G. (2019). A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved vis-nir-swir soil properties estimation. *Geoderma*, Vol. 340:pages 11–24.
- Udelhoven, T., Emmerling, C., and Jarmer, T. (2003). Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant and soil*, Vol. 251(2):pages 319–329.
- Van-Camp, L., Bujarrabal, B., Gentile, A. R., Jones, R. J., Montanarella, L., Olazabal, C., and Selvaradjou, S.-k. (2004). Technical working groups established under the thematic strategy for soil protection. *CiteSeer*, pages 100–102.
- Vasques, G., Demattê, J., Rossel, R. A. V., Ramírez-López, L., and Terra, F. (2014). Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths. *Geoderma*, Vol. 223:pages 73–78.
- Vaudour, E., Gilliot, J.-M., Bel, L., Lefevre, J., and Chehdi, K. (2016). Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. *International Journal of applied earth observation and geoinformation*, Vol. 49:pages 24–38.
- Veres, M., Lacey, G., and Taylor, G. W. (2015). Deep learning architectures for soil property prediction. In *2015 12th Conference on Computer and Robot Vision*, pages 8–15.
- Viscarra Rossel, R. and Lark, R. (2009). Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *European Journal of Soil Science*, Vol. 60(3):pages 453–464.
- Vohland, M., Ludwig, M., Thiele-Bruhn, S., and Ludwig, B. (2017). Quantification of soil properties with hyperspectral data: Selecting spectral variables with different methods to improve accuracies and analyze prediction mechanisms. *Remote Sensing*, Vol. 9(11):pages 1–24.

- Wadoux, A. M.-C. (2019). Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma*, Vol. 351:pages 59–70.
- Wang, C., Zhang, T., and Pan, X. (2017a). Potential of visible and near-infrared reflectance spectroscopy for the determination of rare earth elements in soil. *Geoderma*, Vol. 306:pages 120–126.
- Wang, M., Gao, K., Wang, L.-j., and Miu, X.-h. (2012). A novel hyperspectral classification method based on c5.0 decision tree of multiple combined classifiers. In *2012 Fourth International Conference on Computational and Information Sciences*, pages 373–376.
- Wang, Y., Tang, Y. Y., and Li, L. (2017b). Spectral-spatial destriping of hyperspectral image via corentropy based sparse representation and unidirectional huber–markov random fields. *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 15(06):pages 1750056 (1–18).
- Ward, K. J., Chabrilat, S., Neumann, C., and Foerster, S. (2019). A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered lucas soil database. *Geoderma*, Vol. 353:pages 297–307.
- Whiting, M. L., Ustin, S. L., Zarco-Tejada, P., Palacios-Orueta, A., and Vanderbilt, V. C. (2006). Hyperspectral mapping of crop and soils for precision agriculture. In *Remote Sensing and Modeling of Ecosystems for Sustainability III*, volume Vol. 6298, pages 62980B (1–15).
- Williams, R. J. and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, Vol. 1(2):pages 270–280.
- Xia, J., Chanussot, J., Du, P., and He, X. (2014). Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53(5):pages 2532–2546.
- Xu, C., Zeng, W., Huang, J., Wu, J., and van Leeuwen, W. (2016). Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. *Remote Sensing*, Vol. 8(1):pages 1–20.
- Yang, J., Wang, X., Wang, R., and Wang, H. (2020). Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using vis–nir spectroscopy. *Geoderma*, Vol. 380:pages 1–16.
- Yang, X., Ye, Y., Li, X., Lau, R. Y., Zhang, X., and Huang, X. (2018). Hyperspectral image classification with deep learning models. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56(9):pages 5408–5423.

- Yao, F., Qian, Y., Hu, Z., and Li, J. (2010). A novel hyperspectral remote sensing images classification using gaussian processes with conditional random fields. In *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering*, pages 197–202.
- Ye, C.-m., Liu, X., Xu, H., Ren, S.-c., Li, Y., and Li, J. (2020). Classification of hyperspectral images based on a convolutional neural network and spectral sensitivity. *Journal of Zhejiang University-SCIENCE A*, Vol. 21(3):pages 240–248.
- Ye, Z., Li, H., Song, Y., Wang, J., and Benediktsson, J. A. (2016). A novel semi-supervised learning framework for hyperspectral image classification. *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 14(02):pages 1640005 (1–17).
- Yuan, H., Tang, Y. Y., Lu, Y., Yang, L., and Luo, H. (2014). Hyperspectral image classification based on regularized sparse representation. *IEEE journal of selected topics in applied earth observations and remote sensing*, Vol. 7(6):pages 2174–2182.
- Yue, J., Mao, S., and Li, M. (2016). A deep learning framework for hyperspectral image classification using spatial pyramid pooling. *Remote Sensing Letters*, Vol. 7(9):pages 875–884.
- Zabalza, J., Ren, J., Zheng, J., Han, J., Zhao, H., Li, S., and Marshall, S. (2015). Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. *IEEE transactions on geoscience and remote sensing*, Vol. 53(8):pages 4418–4433.
- Zanaty, E. (2012). Support vector machines (svms) versus multilayer perception (mlp) in data classification. *Egyptian Informatics Journal*, Vol. 13(3):pages 177–183.
- Zhao, W. and Du, S. (2016). Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54(8):pages 4544–4554.
- Zhong, Y., Lin, X., and Zhang, L. (2014). A support vector conditional random fields classifier with a mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery. *IEEE journal of selected topics in applied earth observations and remote sensing*, Vol. 7(4):pages 1314–1330.
- Zhou, F., Hang, R., Liu, Q., and Yuan, X. (2019). Hyperspectral image classification using spectral-spatial lstms. *Neurocomputing*, Vol. 328:pages 39–47.
- Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., and Lausch, A. (2020). High-resolution digital mapping of soil organic carbon and soil total nitrogen using dem derivatives, sentinel-1 and sentinel-2 data based on machine learning algorithms. *Science of The Total Environment*, pages 1–13.

- Zhou, Y., Peng, J., and Chen, C. P. (2014a). Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53(2):pages 1082–1095.
- Zhou, Y., Peng, J., and Chen, C. P. (2014b). Extreme learning machine with composite kernels for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 8(6):pages 2351–2360.