A Machine Learning Based Approach for Soil Property Estimation using Spectral Images

Thesis Submitted
in Partial Fulfillment of the Requirements for the
Degree of

MASTER OF TECHNOLOGY
IN
DATA SCIENCE

Submitted by

SOURAV SEAL (2K22/DSC/16)

Under the supervision of

Dr. DIVYASHIKHA SETHIA, Delhi Technological University

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

JUNE, 2022

DEPARTMENT OF MECHANICAL ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "A Machine Learning Based Approach for Soil Property Estimation using Spectral Images" which is submitted by Sourav Seal, Roll No – 2K22/DSC/16, Department of Software Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Dr. Divyashikha Sethia
Assistant Professor

Date: 24.05,2024

Department of Software Engineering,
DTU

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Sourav Seal, Roll No's –2K22/DSC/16 students of M.Tech (Data Science), hereby certify that the work which is being presented in the thesis entitled "A Machine Learning Based Approach for Soil Property Estimation using Spectral Images" in partial fulfilment of the requirements for the award of degree of Master of Technology, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from Jan 2024 to May 2024 under the supervision of Dr. Divyashikha Sethia .

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Sourav Seal
Candidate's Signature

Date: 24.05.2024

Abstract

Soil is essential to the environmental process and growing crops. An accurate assessment of soil moisture and chemical characteristics is essential for agricultural and land management decisions to be well-informed. Precision agriculture has advanced significantly because of machine learning.

Firstly the paper address the study of hyperspectral images with the help of Riese and Keller's dataset on Very Near Infrared Rays (VNIR) dataset captured by a Cubert UHD-285 snapshot camera for estimating soil moisture that measures real-time band reflectance values. They proposed a framework of Self Organizing Maps (SOM) for regression to estimate soil moisture. Results indicate that MLP performs better than all the machine learning regression-based techniques and SOM framework. It shows promising results and provides a new and suitable regression method to predict soil moisture from the hyperspectral soil moisture dataset. Results indicate that MLP performs better than all the machine learning regression-based techniques and SOM framework. It shows promising results and provides a suitable regression method to predict soil moisture from the hyperspectral soil moisture dataset.

Secondly the paper address the study of multispectral images with the help of the "Land Use/Cover Area Frame Statistical Survey Soil" (LUCAS), a comprehensive and frequent topsoil survey conducted throughout the European Union to get data pertinent to policy about how land management affects soil properties. The data covers 28 European Union States. This work analyses and predicts the chemical properties of the soil of Hungary based on the LUCAS 2015 dataset that includes $CaCO_3$, N, P, K, EC, pH. It estimates them using LUCAS and Landsat 8 satellite images using different regression-based algorithms like GPR, SVR, MLP, AdaBoost, Ridge and compares them. The Lucas survey data points and Landsat 8 satellite images (multispectral) are integrated for forecasting different soil nutrients.

DEPARTMENT OF MECHANICAL ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express our sincerest gratitude to Dr. Divyashikha Sethia for the continuous guidance and mentorship that she provided us during the project. She showed us the path to achieve our targets by explaining all the tasks to be done and explaining to us the importance of this project as well as its industrial relevance. She was always ready to help us and clear our doubts regarding any hurdles in this project. Without her constant support and motivation, this project would not have been successful.

Place: Delhi Sourav Seal

Date: 24.05.2024 (2K22/DSC/16)

DECLARATION

I hereby certify that the work which is presented in the Major Project-II entitled **A Machine Learning Based Approach for Soil Property Estimation using Spectral Images** in fulfillment of the requirement for the award of the Degree of Master of Technology in Software Engineering and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my own, carried out during a period from January to May 2024, under the supervision of **Dr. Divyashikha Sethia.**

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: **Soil Moisture Prediction Using Deep Learning on Hyperspectral Data**
Author names (in sequence as per research paper): Sourav Seal, Dr. Divyashikha Sethia
Name of Conference/Journal: 3rd International Conference on AI-ML Systems
Conference Dates with venue: 25-28th October, 2023, Bangalore, India
Status of paper (Accepted/Published/Communicated): Accepted
Date of paper communication: July 11, 2023
Date of paper acceptance: September 14, 2023
Date of paper publication: May 17th, 2024

Sourav Seal
Sourav Seal
Roll No. 2K22/DSC/16

Student Roll No., Name and Signature

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I further certify that the publication and indexing information given by the students is correct.

Place: Delhi

Date: May 24 2024

Dr. Divyashikha Sethia

Supervisor Name and Signature

DECLARATION

I hereby certify that the work which is presented in the Major Project-II entitled **A Machine Learning Based Approach for Soil Property Estimation using Spectral Images** in fulfillment of the requirement for the award of the Degree of Master of Technology in Software Engineering and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my own, carried out during a period from January to May 2024, under the supervision of **Dr. Divyashikha Sethia.**

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: **A Machine Learning Based Approach for Soil Chemical Property Estimation using Multispectral Data in Hungary**
Author names (in sequence as per research paper): Sourav Seal, Dr. Divyashikha Sethia
Name of Conference/Journal: 4th INTERNATIONAL CONFERENCE ON EMERGING TRENDS AND TECHNOLOGIES ON INTELLIGENT SYSTEMS (ETTIS 2024)
Conference Dates with venue: 27-28th March, 2024 , Noida India
Status of paper (Accepted/Published/Communicated): Accepted
Date of paper communication: December 8, 2023
Date of paper acceptance: March 1, 2024
Date of paper publication: N/A

Sourav Seal
Sourav Seal
Roll No. 2K22/DSC/16

Student Roll No., Name and Signature

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I further certify that the publication and indexing information given by the students is correct.

Place: Delhi

Date: May 24, 2024

Dr. Divyashikha Sethia

Supervisor Name and Signature

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

1.1.1 Hyperspectral

The Cubert UHD 285 snapshot camera's high-resolution hyperspectral data presents an exclusive chance to investigate and comprehend the changes in space and time that occur in soil moisture within Germany. Nevertheless, there are still many things about the use of this hyperspectral data for producing precise and dependable forecasts of soil moisture that we do not know. Producing reliable predictive algorithms that can distinguish patterns in the hyperspectral data and correlate them with soil moisture is very difficult. This research address this gap by predicting with the help of machine learning and remote sensing techniques for predicting soil moisture of Germany with the help of Cubert UHD 285 hyperspectral .

1.1.2 Multispectral

Hints about the spatial distributions of soil attributes are available from multispectral satellite images like those from Landsat 8 and the LUCAS 2015. This study intends to predict soil chemical properties in a particular region by integrating multispectral satellite data with ground truth information derived from LUCAS 2015. To determine soil chemical properties, this work simply needs to come up with reliable predictive models that can accurately incorporate spectral information from Landsat 8 and other contextual variables available in LUCAS 2015. The outcomes of this research generates more accurate forecasting of soil parameters through the better understanding of how satellite data is related to measurements on ground surface.

1.2 Background

Multispectral imaging involves capturing and analyzing data at different wavelengths across the electromagnetic spectrum. Multispectral sensors capture information in several distinct bands, allowing for the extraction of particulars pertaining to vegetation health, land cover, and other environmental variables. Multispectral imaging involves capturing data in multiple bands or channels, each sensitive to a specific range of wavelengths. Nevertheless, the accuracy could be more promising due to spectral resolution limitations, which do not help detect early signs of crop diseases or nutrient deficiency.

This constraint is mitigated by hyperspectral imaging, which offers hundreds of bands. With the ability to capture more precise spectral responses, hyperspectral images—like those produced by satellites like Hyperion, Comprehensive Adolescent Severity Inventory (CASI), and Headwall Micro-Hyperspec—may be better suited to detect minute variations in ground covers and how those variations change over time. Hyperspectral photography helps overcome the mentioned difficulties and enables more precise and fast identification of crop physiological state [3]. Table 1.1 presents the details of different types of sensors used for hyperspectral imaging. Hyperspectral pictures can be used for a variety of purposes in vegetation trait monitoring, including the calculation of the Leaf Area Index (LAI), crop type differentiation, crop biomass extraction, and leaf nitrogen content determination. Despite its exceptional performance, hyperspectral imaging has been comparatively less used in operational agricultural applications over the past two decades due to high costs associated with sensors and imaging missions, low signal-to-noise ratios, and enormous data sets.

Pre-processed hyperspectral images can further be examined using powerful and effective analytical techniques for the vast quantity of data in the images (such as spectral, spatial, and textural variables) and extract target qualities (e.g., crop and soil characteristics). Previous studies propose different methods, such as Radiative Transfer Modeling (RTM, such as PROSPECT and PROSAIL), empirical regression (e.g., linear regression, Partial Least Square Regression (PLSR), and Multi-variable Regression (MLR), machine learning (e.g., Random Forest (RF), deep learning (e.g., Convolutional Neural Network (CNN)) [4–7].

Table 1.1: Different types of hyperspectral sensors [1]

	Satellite-Mounted		**Aircraft-Mounted**				**UAV-Based**	
Sensor Name	**Hyperion**	**FTHSI on MightySat II**	**AVIRIS**	**HYDICE**	**AISA**	**Hymap**	**Headwall Hyperspec**	**UHD 285-firefly**
Spectral Range(nm)	357–2576	475–1050	400–2500	400-2500	400–970	440–2500	400–1000	450–950
number of spectral bands	220	256	224	210	244	128	270 (Nano) 324 (Macro)	125
Spectral Resolution (nm)	10	> 1.7	17	8-15	<3.5	3.3	15	8
Operational altitudes (km)	705	500-600	1-20				<0.15	
Spatial resolution altitudes (km)	30	30	1-20				0.01–0.5	
Organization	NASA, USA	Air Force Research (OH, USA)	Jet Propulsion Laboratory, USA	Naval Research Lab (Washington, DC, USA)	Specim, Finland	Integrated Spectronics Australia	Headwall Photonics, USA	Cubert GmbH, Germany

1.3 Motivation

The agricultural sector is increasingly facing many challenges worldwide, these include crop diseases, climate change, environmental pollution, fast growing population and depletion of natural resources. Precision agriculture is a viable solution for addressing these problems by improving farming practices for example adjustable inputs promised results, such as crop productivity and biomass production, like water and fertilizer as well as minimal environmental effects.Remote sensing can be used to identify crop and soil variability within a field, which is useful information for site-specific management strategies. [8]. Precise application of water, fertilizers, and pesticides is made possible in precision agriculture through the assessment of soil parameters. Farmers can minimize waste and expenses by optimizing resource utilization by customizing inputs to the unique requirements of every section within a field. Estimating soil qualities mitigates the negative environmental impacts of farming operations and aids in delaying the loss of natural resources. [8]. They are using techniques for remote sensing, such as satellite imagery. Farmers may monitor crop conditions and spot pests using drones and a field's hot spots for stress or illness. This knowledge allows for prompt measures like targeted irrigation or pest control, which may increase crop yields and decrease losses.

There are two categories of remote sensing technologies based on the energy source: passive (like optical) and active (like radar and LiDAR). Imaging from multispectral and hyperspectral sensors is included in passive remote sensing.

1.4 Proposed Solution

In the hyperspectral domain, such a benchmark dataset for soil moisture was presented by Riese and Keller in [2]. Images were acquired during a multi-sensor field campaign at Cubert, Germany on the pedon-scale. As measures, they used 125 spectral bands in two different wavelengths with soil temperature and soil moisture. Soil surface reflectance values in this dataset was collected using visible and near infrared (VNIR) method.

In light of this information, the researchers propose that a SOM based regression framework could be used to forecast soil moisture contents in an efficient manner. It should be compared to other popular models in soil temperature prediction, such as RF or SVR. From all of them, SOM model obtained significant performance which its RMSE is equal to 0.66 and R^2 value of 96.78%. These are some of the advantages of neural network-based regression models which take less computation time or may even provide much better outcomes while predicting soil moisture on this data set.

For instance, a feed-forward neural network known as Multilayer Perceptron (MLP), is used to improve accuracy for predicting the soil moisture content in Hyperspectral Benchmark Soil Moisture dataset by Riese and Keller [2]. On the other hand, comparisons were made between SOM regression framework and different machine learning methods. Improved results were obtained through MLP. R^2 of 97.47% and an RMSE value of 0.58 are obtained. For the purpose of achieving ideal testing parameters, the model fine-tunes using the grid-search method.

Bing Lu et al. [1] reviewed different machine learning deep learning-based approaches adopted by researchers to estimate and classify different soil and plant properties. Shagun et al. [9] reviewed various machine learning approaches for soil properties estimation.

Ballabio et al. [10] focused their research on estimating the chemical properties of all 25 countries in the LUCAS dataset using Modis Satellite using LUCAS data points of 2012 survey. In contrary, the current study directs its focus to capture the chemical soil properties of a particular country, Hungary, from the LUCAS 2015 dataset using Landsat8 Satellite. The LUCAS 2018 survey is just an extended version of the 2015 dataset, and some of the land classification is added without any modification or addition of data points.x‘ This study focuses on a particular region for better estimation of the chemical properties in Hungary and also captures enhanced information introduced in the LUCAS 2015 dataset [11].

This work utilizes multispectral data from the Landsat8 satellite to estimate the chemical soil nutrients for Hungary. It focuses on remote sensing and machine learning to estimate chemical properties, helping agricultural professionals and researchers make well-informed decisions without physically testing the soil in the laboratory. It estimates $CaCO_3$, N, P, K, EC, pH from LUCAS topsoil data 2015 [11]. It consists of data from 28 EU countries consisting of ground truth values of soil properties, coordinates of points, and a shape file of the points collected. It is crucial for reliable prediction models.

1.5 Contributions

The contributions of this work are :

1. Hyperspectral Image
    (a) Presents a novel neural network-based MLP regressor to determine soil moisture using VNIR hyperspectral imaging on UHD-285 firefly.
    (b) It compares it with popular algorithms like XGBoost, AdaBoost, and Multilayer Perceptron (MLP).
    (c) It contrasts the suggested model's performance with that of other machine learning models and SOM for regression.
2. Multispectral Image
    (a) Using ground truth values from the LUCAS 2015 dataset for Hungary and publicly available Landsat8 satellite image data, it provides a comparative analysis of many machine learning algorithms.
    (b) Estimation of the chemical soil properties in Hungary Region.

1.6 Thesis Layout

The following sections go into detail about each step. Chapter 2 explores the Technical Background of this study including Smart agriculture, Chemical Properties, Landsat 8 satellite, Hyper and multi spectral images, pre-propcessing of the same and many of the machine learning and deep learning models used in this study. Chapter 3 explores the Hyperspectral Image, with deep dive into 3.1 as the objective, 3.2 as the related work in this field, and 3.3 deep dives into the methodology describing the dataset , models used , the model framework and regression models. Similarly Chapter 4 deep explores the Multispectral Image with deep dive into 4.1 as the objective, 4.2 as the related work in this field, and 4.3 deep dives into the methodology describing the dataset , pre-processing satellite images and describing LUCAS data, models used , the model framework and regression models. Chapter 5 deep dives into the discussions of the results achieved by the paper by fulfilling the objectives. Chapter 6 focuses on concluding the study and lays out a plan line for the work to be further done in this area.

Chapter 2

Technical Background

2.1 Smart Agriculture

Maximizing farming technique depends on smart agriculture that incorporates technology, notably remote sensing.

Decision support systems, monitoring of crops and precision farming are some of the agriculture functions for which drones, satellite imaging and other technologies are used. Farmers are able to make well-informed decisions about pest control, fertilization, and irrigation thanks to real-time data from remote sensing. Customizing farming methods to particular field conditions is known as precision farming. Precise planting, fertilizing, and harvesting are made possible by GPS-guided tractors and equipment, which maximize resource efficiency and reduce waste. Drones and satellite images are commonly used in keeping an eye on large areas of agriculture to improve food security.

These technologies are useful for tracking crops, identifying diseases and even predicting yields. Smart farming is a way of operating that helps achieve sustainability targets by optimizing resource utilization and minimizing the environmental footprint. Reduced use of water, fertilizers, and pesticides helps protect natural ecosystems and promotes long-term agricultural viability.

2.2 Chemical Properties of Soil

The chemical properties of soil, including nutrient levels, pH, and organic matter, significantly impact crop growth. Remote sensing technologies, such as hyperspectral imaging, can be used to infer soil properties based on the spectral reflectance of the soil surface. pH measures the acidity or alkalinity of the soil on a scale from 0 to 14. CEC represents the soil's ability to hold and exchange cations (positively charged ions) such as calcium, magnesium, potassium, and sodium. Soil nutrients include essential elements like N, P, K, S, Ca, Mg and others. Organic matter in soil comes from decomposed plant and animal residues. The relative amounts of sand, silt, and clay particles in the soil are referred to as its texture. EC measures the soil's ability to conduct electrical currents, indicating its salinity. The ratio of carbon to nitrogen in organic matter influences its decomposition and nutrient release.

2.3 Satellite Images for Soil

In order to obtain data on different soil qualities without the need for physical sample, satellite photos are used in the analysis of soils.

Some spectral bands in satellite imaging can be correlated to properties of soils like moisture content, organic matter content and mineral composition.

Remote sensing technology enables a wide range view of large agricultural regions and allows for quick and efficient monitoring of soil. When viewed from space, various earth materials show dissimilar spectral patterns. Spectral features are created through the processes that take place between soils and incoming energy of different wave lengths.

Such marks can be affected by a number of soil properties: for instance moisture content, organic matter level, mineral types present as well as texture among others.

Spatial resolution is a crucial aspect in satellite-based soil investigation. The ability to capture finer details of even small variations in soil attributes can be enhanced with higher quality images.

There exist some satellites including Sentinel series and commercial high resolution ones which have spatial resolutions that fit well into comprehensive soil mapping.

2.4 Landsat8

Featuring the "Thermal Infrared Sensor" (TIRS) and "Operational Land Imager" (OLI), Landsat 8 is a satellite that provides multispectral data in many bands. High-resolution imagery from Landsat 8 is available worldwide and is useful for tracking changes in land cover, evaluating crop health, and researching environmental trends. Following in the footsteps of its predecessors, For a variety of uses, including forestry, urban planning, environmental research, agriculture, and land cover monitoring, Landsat 8 seeks to provide essential data. The two main instruments included with Landsat 8 are the Thermal Infrared Sensor (TIRS) and the Operational Land Imager (OLI). Data collected by Landsat 8 in various spectral bands spans a significant section of the electromagnetic spectrum. Visible (blue, green, and red), near-infrared, shortwave infrared, and thermal infrared are among the bands. Because Landsat 8 orbits around the sun, it can take pictures of the whole planet at around the same local solar time on each pass. It covers the entire planet every 16 days and completes an orbit in around 99 minutes.

2.5 Hyperspectral Images

The Cubert UHD 285 snapshot camera's high-resolution hyperspectral data presents an exclusive chance to investigate and comprehend the changes in space and time that occur in soil moisture within Germany. Nevertheless, there are still many things about the use of this hyperspectral data for producing precise and dependable forecasts of soil moisture that we do not know. Producing reliable predictive algorithms that can distinguish patterns in the hyperspectral data and correlate them with soil moisture is very difficult. This research address this gap by investigating cutting-edge machine learning and remote sensing techniques for predicting soil moisture using Cubert UHD 285 data over Germany.

2.6 Multispectral Images

Multispectral satellite images (like those collected by Landsat 8 or LUCAS2015 dataset) already contain information about the spatial distributions of soil attributes. In this regard, accurate prediction of soil chemical properties was further aimed by combining multispectral satellite observation data with ground truth information obtained from LUCAS 2015 for a certain region. Our aim is to develop suitable predictive models that accurately integrates spectral information of Landsat 8 with other spatial contextual variables provided in LUCAS 2015 in order ot infer soil chemical properties. 'The results of this research give us more accurate predictions of soil parameters because we better understand how satellite data relates to measurements on ground surface.

2.7 PreProcessing of Hyperspectral And Multispectral Images

Common procedures in the processing of hyperspectral data include atmospheric correction, radiometric correction, orthorectification, and geometric correction. The geometry and orthorectification correction for satellite- and aircraft-based hyperspectral pictures is often handled by data providers, while the radiometric and atmospheric adjustments can be handled by following normal image processing procedures made accessible by remote sensing software. Contrarily, with UAV-based photos, users must carry out these processing stages and select the best processing strategies and related parameters. To convert digital values of images to radiance, radiometric correction is applied using calibration coefficients provided by the sensor manufacturer. Over time, these coefficients may need to be modified due to the spectral materials employed in the construction of the hyperspectral sensors deteriorating.

The signals are nonetheless affected by various atmospheric absorptions and scatterings even when the UAVs are flown at low altitudes, necessitating atmospheric adjustment. Numerous bands in hyperspectral images—which frequently contain hundreds of them are closely connected. As a result, dimension reduction is a crucial step in the pre-processing of hyperspectral images. Numerous prior investigations utilizing hyperspectral images have examined the difficulties associated with redundant data and have employed various techniques for dimension reduction. For instance, Miglani et al. [12] 's PCA of hyperspectral pictures revealed that the top 10 main components could account for 99% of the information.

Numerous algorithms have been proposed in previous studies for band selection: a column subset selection-based method that maximizes the volume of the selected subset of columns (i.e., bands) while being robust to noisy bands; an unsupervised fast volume gradient-based method that removes the most redundant bands sequentially based on the gradient of volume; and so on. Pre-processing is a crucial step in enhancing the quality of hyperspectral pictures and setting up subsequent data analysis, in general. After pre-processing, the analytical techniques that is covered in the next section may be utilised to analyse the hyperspectral data and look at various agricultural aspects on the ground.

2.8 Machine Learning Models

1. *PLSR*: PLSR is a regression method combining features from PCA and multiple linear regression. PLSR creates new variables, known as latent variables or components, that are linear combinations of the original predictors and are optimized to explain the variance in the target variable [13].

2. *SVR*: Support vector machines (SVM) are used in SVR, a regression technique, to predict the hyperlink between the target variable and the data that was provided. When addressing non-linear correlations between predictors and the target variable, it is advantageous. SVR looks for the hyperplane that most closely matches the data in order to minimize the size of the coefficients and allow for an error margin (epsilon tube) [14].

3. *MLP*: A form of artificial neural network (ANN) called a Multilayer Perceptron (MLP) has several layers of linked nodes or neurons. Regression and classification are two prominent machine learning tasks that employ it. Using hidden layers and activation functions to restructure the input data, MLP can simulate intricate non-linear connections in data [15].

4. *GPR*: A non-parametric probabilistic regression method called GPR represents the target variable as a Gaussian process. It offers estimates of prediction uncertainty in addition to point forecasts. When working with tiny datasets and accurately expressing uncertainty in forecasts is crucial, GPR is helpful [10].

5. *AdaBoost*: AdaBoost is an ensemble learning technique used mainly for classification tasks but can be adapted for regression. It combines several weak learners to produce a powerful prediction model (usually decision trees). AdaBoost is particularly efficient at enhancing model performance because it weights data points differently and concentrates on samples incorrectly identified in each iteration [16].

6. *Ridge*: Ridge regression is a linear regression technique that introduces L2 regularization to the linear regression model. Forbidding high coefficients improves the cost function by adding a penalty term that helps avoid overfitting. Ridge regression is helpful when multi-collinearity is present, and a more straightforward, understandable model is sought [17].

2.9 Applications of Hyperspectral And Multispectral Images

In smart agriculture, hyperspectral and multispectral imaging has a variety of uses,

1. Crop Biochemical and Biophysical Property Estimation.

   Retrieval of crop biochemical and biophysical properties includes chlorophyll, nitrogen content, Leaf Area Index(LAI). Agricultural soil characteristics, such as moisture content, organic matter, salinity, and roughness, significantly impact crop growth and final output [18]. Additionally, each of these aspects of vegetation (such as chlorophyll, water, and LAI) affects crop production. In order to understand agricultural productivity and apply effective management strategies. Remote sensing is also essential for yield prediction and crop biomass estimation. Yang [19].estimated agricultural yield using both multispectral and hyperspectral data, and discovered that the hyperspectral imagery-based model outperformed the other two.

2. Evaluating Crop Nutrient Status Evaluating plant biochemical qualities.

   In precision farming, crop nutrient status is assessed, and suggestions for resource management that match the demands of the crop are made.Hyperspectral pictures have been used in previous research to estimate the nitrogen content of various crop kinds. Akhtman et al. [20] analysed the temporal fluctuation of these features while assessing the nitrogen content and phytomass in maize and wheat fields using UAV-based hyperspectral pictures. Cilia et al. [21] in an experimental maize field, measured estimates of nitrogen content and dry mass were obtained using airborne hyperspectral pictures, with the goal of determining the nitrogen deficiency and offering a map of variable rate fertilisation. The authors also offered a method for determining how little nitrogen should be used to maintain crop output and prevent over fertilization.

3. Classifying Imagery to Identify Crop Types, Speicies or Disease

   Hyperspectral and multispectral images find their use to categorise crops and measure their characteristics, such as weed or invasive species classification, crop type differentiation, crop growth phases, and disease detection. Infestation of weeds is a serious problem in agricultural areas and may have a significant impact on crop development and productivity. Using remote sensing to locate and map weeds in agricultural fields will be very helpful for the field's variable rate treatment. Since many weeds are small and mixed with crops, weed identification generally calls for a high spatial resolution. To minimise financial and produce losses, farmers must pay close attention to crop disease. Hyperspectral imaging might potentially detect early signs of crop disease and help prompt treatments since it gathers signals at fine spectral resolutions (e.g., fewer than 10-nm intervals).

4. Obtaining Fertility, Soil Moisture, and Additional Physical and Chemical Properties.

   By combining airborne sensor imagery with linear regression to estimate soil moisture at three distinct depths, Finn et al. [22] also explored the benefits and drawbacks of employing hyperspectral remote sensing to study soil moisture. Many other soil features are investigated, like soil texture [23], soil nitrogen [24], copper concentration [25], potassium content [26], and CO_2 leaks [27]. SOC is an essential part of soil fertility, which has a significant impact on crop development and output. Fine spectral features from hyperspectral data are essential for estimating SOC content.For examining SOC, earlier research have utilised hyperspectral pictures gathered by various platforms. Overall, there is great potential for using hyperspectral imaging to measure soil organic matter and carbon. In total, the evaluation of organic matter and carbon content in soil is highly promising by using hyperspectral imaging. The analysis of soil moisture and plant cover has a great impact on the study of soil organic matter and carbon. In this regard, one possible approach could be to accumulate hyperspectral images in non-growing periods.

Chapter 3

Hyperspectral Image for Soil Moisture Prediction

3.1 Objective

The article by Riese and Keller presents a Very Near Infrared Rays (VNIR) data set from the Cubert UHD-285 snapshot camera for real-time measurement of soil moisture estimation using band reflectance values. They suggested an architecture based on Self Organizing Maps (SOM) for soil moisture prediction through regression.

This study proposes a Neural Network based Multilayer Perceptron (MLP) approach for estimating soil moisture values using regression analysis on the dataset. It compares it with machine learning regression models and the SOM framework.

The work proposes a forward neural network, MLP to improve accurate predictions of soil moisture in Riese and Keller's Hyperspectral Benchmark Soil Moisture dataset [2]. This is being compared to other machine learning algorithms as well as the SOM framework in regression. Its MLP has better R^2 of 97.47% and RMSE values at 0.58 than MLPs. Optimal test parameters were achieved by fine tuning of this model with grid search method.

MLP is a computationally effective algorithm as the parallelization is done efficiently. It is an adaptive learning technology that does tasks based on given data based on training data. In contrast, SOM needs adequate and relevant data to create meaningful clusters. The weight vectors work based on information that correctly classifies and separates inputs. The groupings will become random if there is no data or redundant data in the weight vectors [28].

3.2 Motivation

The agricultural sector is increasingly facing many challenges worldwide, these include crop diseases, climate change, environmental pollution, fast growing population and depletion of natural resources. Precision agriculture is a viable solution for addressing these problems by improving farming practices for example adjustable inputs promised results, such as crop productivity and biomass production, like water and fertilizer as well as minimal environmental effects.Remote sensing can be used to identify crop and soil variability within a field, which is useful information for site-specific management strategies. [8]. Precise application of water, fertilizers, and pesticides is made possible in precision agriculture through the assessment of soil parameters. Farmers can minimize waste and expenses by optimizing resource utilization by customizing inputs to the unique requirements of every section within a field. Estimating soil qualities mitigates the negative environmental impacts of farming operations and aids in delaying the loss of natural resources. [8]. They are using techniques for remote sensing, such as satellite imagery. Farmers may monitor crop conditions and spot pests using drones and a field's hot spots for stress or illness. This knowledge allows for prompt measures like targeted irrigation or pest control, which may increase crop yields and decrease losses.

3.3 Research Gap

Since its inception, a number of challenges have made it challenging to analyse and work with hyperspectral images. It was initially plagued by spectroscopic technology because of poor hyperspectral sensor quality and inadequate data quality. Though things have gotten easier as applied science has advanced, there are still a few well-known nondispersible obstacles to be solved. Here are a few of them as stated:

- Absence of high-resolution, noise-free Earth observation (EO) photos When spectrometers were originally discovered, they were not very effective. As a result, for Earth observation purposes, sounds from water vapour, air pollution, and other atmospheric perturbations alter the signals originating from the surface of the Earth. Over the past few decades, numerous attempts have been undertaken to provide high-quality hyperspectral data for Earth observation and to create a variety of high-performance spectrometers that combine the capabilities of spectroscopy, digital photography, and the extraction of numerous embedded spatial-spectral properties.
- Obstacles to feature extraction: duplication between adjacent spectrum bands during data collection leads to the availability of duplicated information, both spectrally and geographically, making it difficult to retrieve spatial-spectral characteristics in an optimal and discriminative manner.
- The large spatial variability and interclass similarity: The hyperspectral dataset collected contains unusable noisy bands due to mistakes in the acquisition that result in information loss in terms of the unique identity, that is, the spectral signatures and excessive intraclass variability. Furthermore, with poor resolution, each pixel comprises broad spatial regions on the Earth's surface, generating spectral signature mixing, contributing to the enhanced interclass similarity in border regions, thus creating inconsistencies and uncertainties for employed classification algorithms.

3.4 Related Work

Deep learning is widely used in remote sensing for image categorization, such as the identification of land cover [29] and land cover categorization [30]. Examples of current research include detecting plant diseases using CNN and photos from a smartphone [31], as well as estimating agricultural yields using CNN and multispectral photographs along with climatic data [32] and crop categorization using 3-D CNN and multi-temporal multispectral images [33]. Siegmann and Jarmer [34] assessed the effectiveness of RF, SVM, and PLSR for estimating crop LAI using ground-based hyperspectral reflectance data collected by an ASD spectroradiometer and validated RF's strong performance. Table 4.1 presents various techniques for classifying crops or predicting soil properties of hyperspectral data using machine learning and deep learning applications. A few examples of previous research include the classification of crops using 3-D CNN and multi-temporal multispectral images, the estimation of crop yield using CNN and multispectral images along with climate data, and the classification of agricultural land cover using deep recurrent neural network and multi-temporal SAR images. In conclusion, there are varying degrees of complexity, performance, and transferability across various analytical techniques (such as RTM, advanced regression, machine learning, and deep learning). Overall, linear regression is the simplest approach to apply, and it performs rather well, albeit the quality of the sample data and the selection of predictor variables can have a significant impact on this method's effectiveness. Since advanced regression (like PLSR) incorporates multiple variables and is less sensitive to data noise than linear regression, it generally outperforms linear regression. RTM (for instance, PROSAIL) has the ability to provide a variety of data outputs with reassuringly high accuracy, including chlorophyll, water, and LAI. The great transferability of this approach is one of its key benefits. However, due to its extensive programming and requirement for a wide range of parameters, this approach is the most complex. Numerous well-known machine learning algorithms, such as RF and SVM, have shown good performance in prior research. To get the most out of this strategy, some programming and model modifications are required. Deep learning is a relatively new technique that has grown in popularity recently. For this strategy, appropriate model design and programming are essential. To obtain a decent model performance, it also needs a lot of training data and computer power.

3.5 Methodology

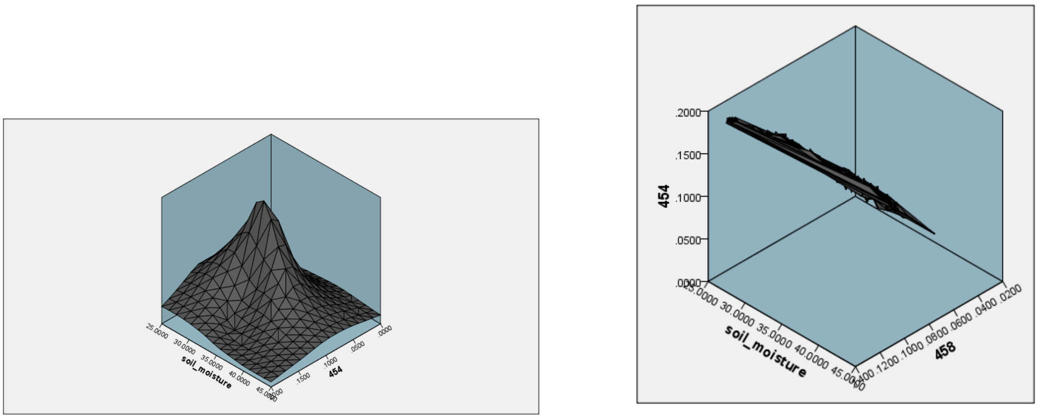
3.5.1 DataSet Description

A Cubert UHD 285 sensor captures the image used in the dataset proposed by Riese and Keller [2]. This dataset contains actual measurements of soil moisture in real-time. The measured reflectance contains the spectral signatures of plants growing on the soil's surface. Each hyperspectral image has 50 x 50 pixels and 125 spectral channels with a wavelength range of 450 nm to 950 nm and a spectral resolution of 4 nm [2]. Fig.4.4 and Fig.4.3 show the visual representation of the features. Riese and Kellar [2] documented the pre-processed hyperspectral image's reflectance values in an Excel file.

Table 3.1: Comparison of Machine Learning and Deep Learning Methods on hyperspectral data

Application	Previous Study	Research Focuses	Accuracy
Weed Detection	Bing Lu [6]	Estimating Vegetation Chlorophyll Content	R^2=0.86,RMSE=12.1
Species classification	Bing Lu [35]	Classification using UAV	86%
Land Cover	Sharma [29]	Pixel Based Recurrent Neural Network	97.21%
Plant Disease	Mohanty [31]	Image Based Plant Disease Detection	99.34%
Crop Categorisation	Shunping Ji [33]	3D CNN with Multi-Temporal Remote Sensing Images	79.4%
Wheat leaf area index	Siegmann [34]	Comparing different regression models to determine the leaf area index	R^2=0.924,RMSE=0.367
Soil Nutrients	Riese and Keller [36]	Soil Moisture Estimation using UAV	R^2=96.78%,RMSE=12.1

The dataset has 679 records of high-dimensional data, defined by 125 spectral bands ranging from 450-950 nm (Table 1.1).



(a) Soil Moisture vs 1 band

(b) Soil Moisture vs 2 band

Figure 3.1: Visualization of soil moisture with bands

3.5.2 Model Description

This work proposes and compares the MLP model with other machine-learning-based regression analyses to approximate soil moisture's value. The regression performs specific steps, including data cleaning, feature selection, data pre-processing, normalization, and splitting data into train and test sets. Further steps include an algorithm selection following training and scoring the model by evaluation metrics (R-squared and Root Mean Squared Error). Then it is repeated for all the algorithms tested in this work (Fig.4.2).

The MLP model comprises two hidden layers with sizes of 300 and 100, obtained by the primary grid search method, with an activation function of relu, a learning rate of 0.05, and a solver of lbfgs. Fig. 3.4 shows the visual representation of the implemented MLP (not to scale).

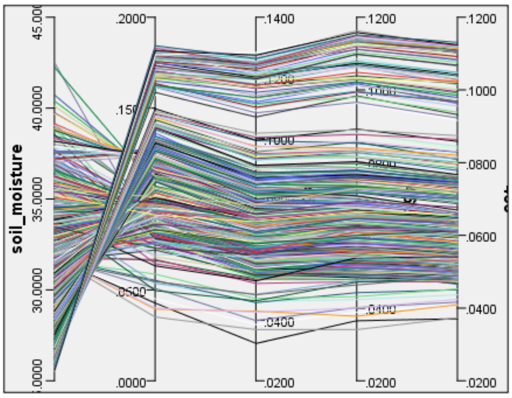


Figure 3.2: Parallel graph with soil moisture and bands

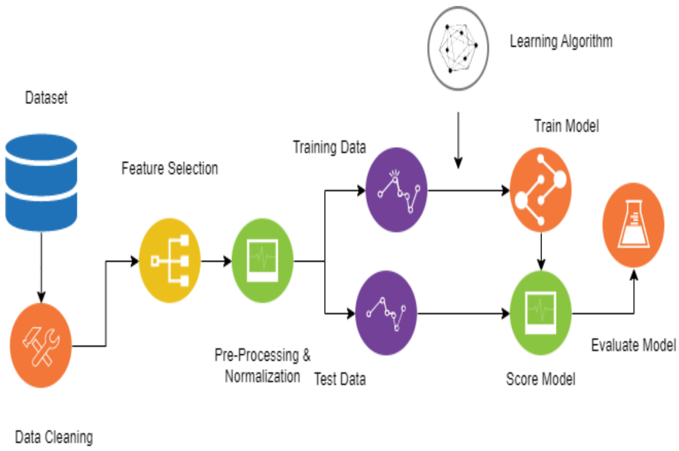


Figure 3.3: Flow diagram for regression

3.5.3 Feature Selection

In the hyperspectral soil-moisture dataset proposed by Riese and Keller [2], 679 data points capture high dimensional data consisting of 125 spectral bands and soil temperature. The features consist of 125 spectral bands along with the soil temperature. This work predicts the value of soil moisture by different regression models. It compares the regression models with reduced features to the model without feature selection. The model's accuracy is compared by reducing the number of features to 115 spectral bands and removing the first and final five bands [37]. The model's accuracy declines by reducing the number of features to 115 spectral bands.

3.5.4 Data Pre-Processing and normalization

PCA reduces the dataset dimension in data pre-processing. The layer of the VNIR reflectance values undergoes PCA. Regression is applied on the first 20 main components since they account for the majority of the variances in the dataset. The dimensionality of the band is 20, retaining the feature importance without lowering the number of

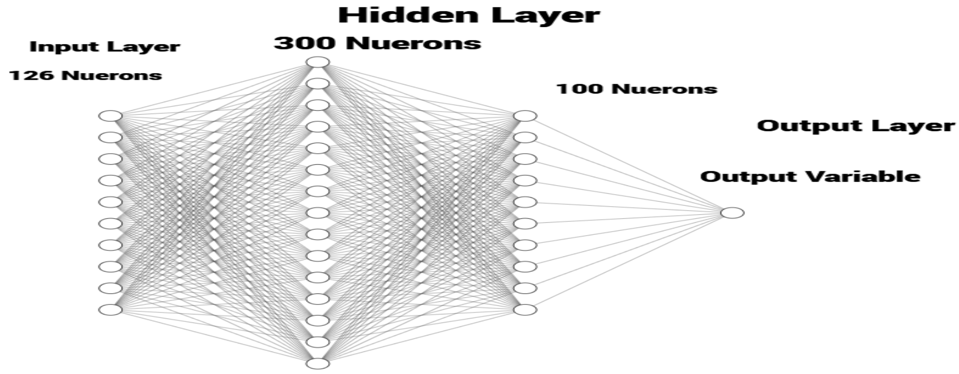


Figure 3.4: MLP Neural Network (not to scale)

features. Also, the model uses min-max scaling to normalize the data values to a range of 0 to 1 by applying them to the entire dataset without any feature selection.

This study uses PCA or min-max scaling and then compares the results of predicting soil moisture with different pre-processing and normalization techniques on the whole and reduced datasets (with 115 features). Then the dataset is split into training and test subset consisting of 339 and 340 data points, respectively.

3.5.5 Different Regression Models

Different regression models are applied in this study to predict soil moisture based on the given dataset. They include Linear regression, PLS regression, Ridge, Lasso, Elastic Net Regression, K-NN, Decision Tree Regression (DT), RF regression, Extreme Gradient Boosting (XGB), Adaptive Boosting (AdaBoost) and Multilayer Perceptron (MLP). All the Learning Algorithms are supervised and implemented using the sklearn python library [38].

Hyperparameters and model parameters make up a regression model's parameters. Hyperparameters are predetermined, whereas the training step involves fine-tuning the model parameters. Depending on the pre-processing techniques used in step 2 of the regression framework, the optimal setting of the hyperparameters varies. The method used in this study obtains the optimal hyperparameter setting by a primary grid search method and extracts the best parameter.

In the model, the evaluation metrics chosen are the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) for evaluating the model performance. (R^2) is the percentage of variation in the dependant variable that the statistical model predicts.The RMSE is a quadratic scoring mechanism used to determine the amount of the average mistake. A good performance indicates a high value of R^2 and a low value of RMSE.

Chapter 4

Multispectral LUCAS Image for Soil Chemical Property Estimation

4.1 Objective

This paper analyses and predicts the chemical properties of the soil of Hungary based on the LUCAS 2015 dataset that includes Calcium Carbonates ($CaCO_3$), Nitrogen (N), Phosphorus (P), Potassium (K), Electrical Conductivity (EC), pH. It estimates them using LUCAS and Landsat 8 satellite images by using different regression-based algorithms like Partial Least Square Regression (PLSR), Gaussian Process Regression (GPR), Support Vector Regression (SVR), Multilayer Perceptron (MLP), AdaBoost, Ridge and compares them. Ballabio et al. [10] focused their research on estimating the chemical properties of all 25 countries in the LUCAS dataset using Modis Satellite using LUCAS data points of 2012 survey. In contrary, the current study directs its focus to capture the chemical soil properties of a particular country, Hungary, from the LUCAS 2015 dataset using Landsat8 Satellite. The LUCAS 2018 survey is just an extended version of the 2015 dataset, and some of the land classification is added without any modification or addition of data points. This study focuses on a particular region for better estimation of the chemical properties in Hungary and also captures enhanced information introduced in the LUCAS 2015 dataset [11]. This paper utilizes multispectral data from the Landsat8 satellite to estimate the chemical soil nutrients for Hungary. It focuses on remote sensing and machine learning to estimate chemical properties, helping agricultural professionals and researchers make well-informed decisions without physically testing the soil in the laboratory. It estimates $CaCO_3$, N, P, K, EC, pH from LUCAS topsoil data 2015 [11]. It consists of data from 28 EU countries consisting of ground truth values of soil chemical and physical properties, coordinates of the points, and a shape file of the points collected. It is crucial for definative prediction models.

4.2 Motivation

The agricultural sector is increasingly facing many challenges worldwide, these include crop diseases, climate change, environmental pollution, fast growing population and depletion of natural resources. Precision agriculture is a viable solution for addressing these problems by improving farming practices for example adjustable inputs promised results, such as crop productivity and biomass production, like water and fertilizer as well as minimal environmental effects.Remote sensing can be used to identify crop and soil variability within a field, which is useful information for site-specific management strategies. [8]. Precise application of water, fertilizers, and pesticides is made possible in precision agriculture through the assessment of soil parameters. Farmers can minimize waste and expenses by optimizing resource utilization by customizing inputs to the unique requirements of every section within a field. Estimating soil qualities mitigates the negative environmental impacts of farming operations and aids in delaying the loss of natural resources. [8]. They are using techniques for remote sensing, such as satellite imagery. Farmers may monitor crop conditions and spot pests using drones and a field's hot spots for stress or illness. This knowledge allows for prompt measures like targeted irrigation or pest control, which may increase crop yields and decrease losses.

4.3 Research Gaps

Numerous important applications have profited from multispectral imageries' distinctive qualities. However, because of their fluctuating temporal aspects, these qualities also present practical implementation challenges. Significant adjustments are made to the spatial, spectral, temporal, and data sources (single or many data sources) of multispectral images based on its temporal feature. Although multispectral data has several benefits, there are a few issues that have also been discussed in the literature.

- Spatial : High resolution data and less accuracy of super resolution.
- Temporal: Change detection when the detection outline shifts with respect to time falls within the spatial-temporal pattern. To identify the shift in multi-temporal remote sensing image detection
- Spectral: Selection of Spectral bands for distinguished applications (the individual spectral band has different kinds and levels of details). Such as land-use mapping, land-cover mapping, forest inventory, and urban-area monitoring
- multi Data sources: A potent method for producing more detailed features is the integration of data from multiple heterogeneous data sources.

4.4 Related Work

Many researchers have attempted to explore multispectral data from satellites and used it for soil analysis and prediction of soil features and crop type classification. The LUCAS dataset is a significant leap due to its vast coverage of 28 countries and actual ground truth values. This section attempts to capture all the significant work previously done in this area. Table 4.1 refers to the researchers' previous work using the LUCAS dataset and satellite images combined to produce significant predictions about the soil properties. Ballabio et al.(2019). [10] provided a GPR regression technique to estimate the chemical properties of the soil based on Modis satellite data and LUCAS 2012 topsoil data. The chemical properties include $CaCO_3$, N, P, K, EC, and pH. Ballabio et al.(2016) also provided Multivariate Adaptive Regression Splines (MARS) to predict the physical soil properties with the help of Modis satellite data points and LUCAS 2012 database. The physical properties include sand, clay, silt, and gravel. In both the above papers, the authors achieved good accuracies. Pflugmacher et al. (2019). [39] used Landsat8 satellite data and LUCAS 2015 data to classify land/cover classes across Europe using Random Forest. The overall accuracy achieved was 75.1%. Various variables can affect crop development and yield, including soil moisture, soil type, and topographic conditions. Due to the extensive spectrum data in hyperspectral photography, crop nutrient status may be assessed accurately. Crop output can also be improved with a more comprehensive management strategy considering the crop's nutritional status and other contributing variables. The analysis of these aspects can benefit tremendously from hyperspectral remote sensing. While some of these variables may overlap spectrally, different soil characteristics impact the spectral signals in various bands and varying degrees. So, while analyzing a specific soil characteristic, it is crucial to collect adequate soil samples while controlling other soil factors.

Table 4.1: Comparison of Machine Learning and Deep Learning Methods on LUCAS using Multispectral data

Previous Study	**Lucas Dataset Year**	**Research Focuses**	**Accuracy**
Pflugmacher [39]	2015	Mapping Pan European Data using Landsat8 data and LUCAS Dataset	Random Forest:75.1%
Ballabio [40]	2012	Predicting LUCAS Topsoil Physical Properties using Modis Sattelite	Multiplicative Adaptive Regression Splines(MARS): R^2= 0.47 - 0.5
Ballabio [10]	2012	Predicting LUCAS Topsoil Chemical Properties using Modis Sattelite	Gaussian Process Regression(GPR): RMSE= 0.78 - 121.89
Ghassemi [41]	2018	Crop type classification using Sentinel 2 and LUCAS dataset	SVM : 76.8%
Riese [42]	2012	LUCAS dataset-based hyperspectral data-based soil texture classification	CNN : 74%
Castaldi [43]	2012	SOC Prediction using XDS Rapid content analyzer spectroradiometer	PLSR : RMSE - 3.7
Castaldi [44]	2012	SOC Prediction using Normalised Burn Ratio(NBR) for North-Eastern Germany	NBR : RMSE - 6.8 - 26.6 RPD - 0.9 - 4.4

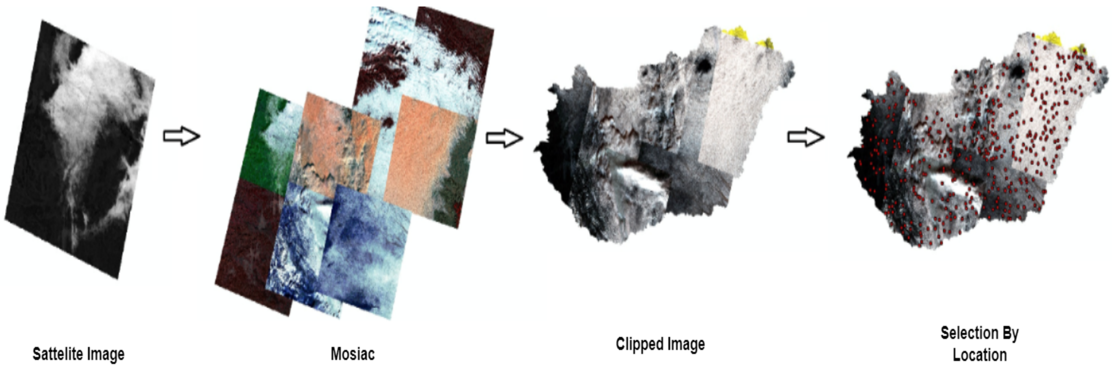


Figure 4.1: Image Processing Step

4.5 Methodology

This section covers the approach for the models for estimating soil nutrients. Fig. 4.2 describes the methodology adopted in this work.

4.5.1 DataSet Description

Land Use/Cover Area frame Statistical Survey (LUCAS) 2015 dataset [11] is used in this survey. Also Landsat 8-9 OLI/TIRS C2 L2 satellite imagery captures multispectral images. This study uses the dataset for Hungary. It estimates the chemical properties $CaCO_3$, N, P, K, EC, and pH from the LUCAS dataset.

LUCAS 2015 topsoil dataset covers 28 European Union countries with ground truth data. The LUCAS dataset consists of a ground truth CSV file and a shp file comprising all the collected data points. Approximately 22000 data samples are collected from all of the 28 European countries. In this study, we only use a subset of the data of Hungary that consists of 412 points. It consists of all the ground truth values of soil texture, pH levels, organic carbon content, nutrient concentrations, and soil type. It offers extensive information about the soil features in Europe.

The Landsat 8-9 OLI/TIRS C2 L2 satellite captures image from Hungary. The data used is a satellite image from 2015 (same as LUCAS). It is a multispectral dataset that captures different electromagnetic spectrum bands [45].

Fig. 4.4 represents the key maps of Hungary using the Inverse Distance Weighting (IDW) interpolation technique, which shows all the chemical properties of the soil. The maps show the distribution of the nutrient levels from low to high. The lighter tone represents a lower range, while a brighter tone represents a high range. A specific color represents the range of the nutrient value present in that region.

4.5.2 Satellite Data Processing

Data processing for Landsat 8-9 OLI/TIRS C2 L2 satellite imagery is critical for accurate soil properties prediction. This section discusses the essential steps to process

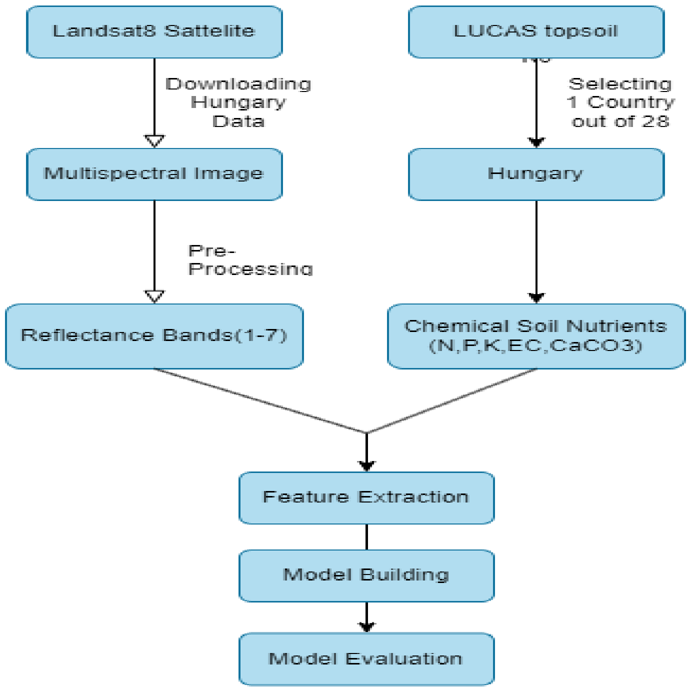


Figure 4.2: Flow diagram for Research Methodology

satellite images: band composition, mosaicing, clipping, selection by location, and extracting multi-value to points. Fig. 4.1 describes Landsat 8 multispectral data processing steps for a broad area and combining it for a target area.

- *Band Composite*: The multispectral Landsat 8 data consists of 7 bands. For each band, a tiff image is present for a particular tile. The seven different images are combined into a single image using this step.
- *Mosaicing*: Multiple tiles are needed for a broad coverage area. Each tile has data that has seven images that are combined using the previous step. The total tiff images are the same as the number of tiles. This step combines all the tiff images into one single tiff file.
- *Clipping*: The data obtained from the previous step is raw data from the specified region with outliers. The shape file of Hungary is superimposed on the raw data to trim out the target region (Hungary). To obtain the area of the excess portion is clipped out to get only the area of interest.
- *Selection By Location*: The clipped data is to choose the shape file from the LUCAS topsoil 2015 data. The collected data points are marked on the image and spatially aligned. It ensures optimal comparison and integration of the satellite image with the ground data.
- *Extracting MultiValue to points*: The final tiff image now consists of the relevant data and features of the LUCAS data and the band information of Landsat 8 satellite data. This work converts the LUCAS data and the bands into point values, where each tuple consists of the soil features and the reflectance values

of each of the seven bands.

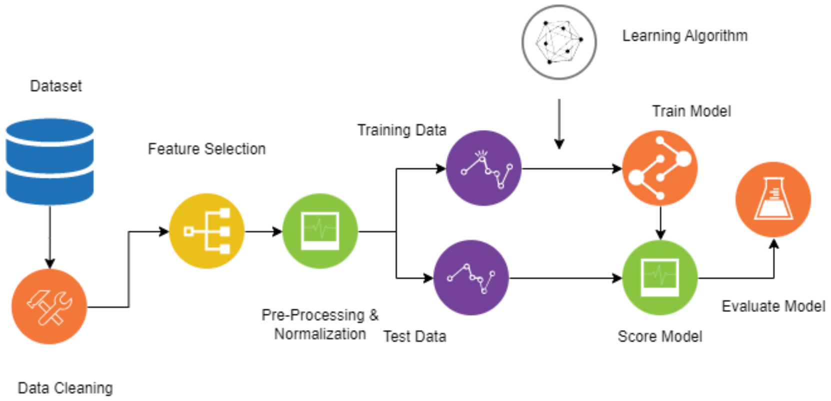


Figure 4.3: Regression Algorithm Flow

4.5.3 Materials and Methods

Reflectance Calculation and Normalization

This work calculates the reflectance index from Landsat 8 data. The OLI sensor in Landsat 8 captures the images for 7 bands (Band 1-7). The data is downloaded from the United States Geological Service (USGS). The data is converted into 16 bits DNs from 0-65536. The reflectance is used to determine Normalized Difference Vegetation Index(NDVI) and Proportion of Vegetation (P_v). NDVI and P_v for Landsat 8-9 is defined as

$$NDVI = (Band5 - Band4)/(Band5 + Band4)$$

$$P_v = ((NDVI - NDVI_{min})/(NDVI_{max} - NDVI_{min}))^2$$

The dataset is normalized using Standard Scalar, which normalizes the bands in the range of 0-1. This method facilitates the data to be converged and free from outliers for the model to be applied.

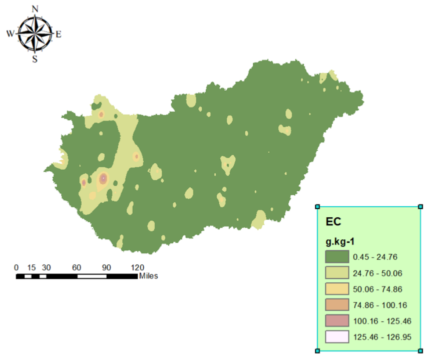
Feature Selection

Landsat 8-9 OLI/TIRS C2 L2 has 7 bands and a spatial resolution of 30m. Band 2, Band 3, and Band 4 represent Blue, Green, and Red, respectively; Band 1 represents indigo for Coastal/Aerosols, and Band 5 represents Near Infrared Rays(NIR). Water in the leaves of healthy plants reflects this portion of the spectrum. Bands 6 and 7 represent Short Wave Infrared Rays(SWIR) - I and II for identifying dry and moist regions. This study utilizes every band within this dataset.

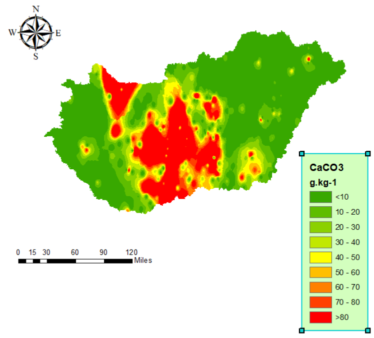
LUCAS data extraction

The primary nutrients for this study are the chemical properties $CaCO_3$, N, P, K, EC, and pH. The Lucas topsoil dataset from 2015 [11] served as this study's ground truth data source. It estimates the soil fertility and other important aspects of the ground.

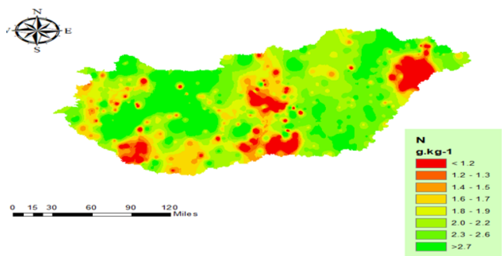
The area of interest for this study is Hungary, so out of 22,000 soil sample data, 412 data are extracted and combined with the reflectance values of bands extracted from Landsat data. The combined data is the data source to develop and validate the model.



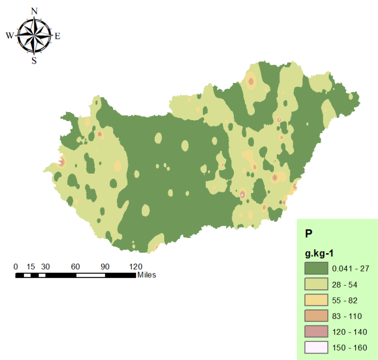
(a) Topsoil Map EC



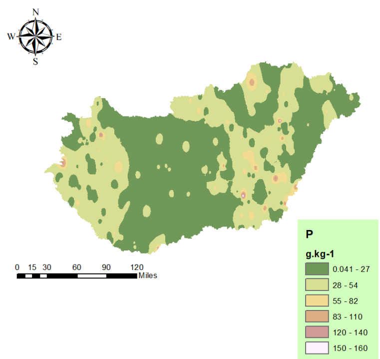
(b) Topsoil Map $CaCO_3$



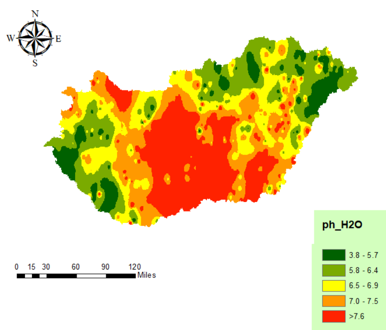
(c) Topsoil Map N



(d) Topsoil Map K



(e) Topsoil Map P



(f) Topsoil Map pH

Figure 4.4: Topsoil Map of Hungary

4.5.4 Model Description

This work proposes and compares different machine learning algorithms such as PLSR, GPR, SVR, MLP, AdaBoost and Ridge. The regression algorithm performs specific steps, including data cleaning, feature selection, data preprocessing, normalization, and splitting data into train and test sets. Further steps include an algorithm selection following training and scoring the model by evaluation metrics (RMSE and RPD). The methodology adopted in this work is to repeat the steps for all the tested algorithms. Fig. 4.3 shows the regression methodology adopted in this work. All the Learning Algorithms are supervised and implemented using the sklearn python library [38].

- *PLSR*: PLSR is a regression method combining features from PCA and multiple linear regression. PLSR creates new variables, known as latent variables or components, that are linear combinations of the original predictors and are optimized to explain the variance in the target variable [13].
- *SVR*: SVR is a regression approach that models the association between input data and the target variable using support vector machines (SVM). It is beneficial when dealing with non-linear correlations between predictors and the target variable. In order to minimise the size of coefficients and account for an error margin (epsilon tube), SVR seeks to identify the hyperplane that best matches the data [14].
- *MLP*: A form of artificial neural network (ANN) called a Multilayer Perceptron (MLP) has several layers of linked nodes or neurons. Regression and classification are two prominent machine learning tasks that employ it. Using hidden layers and activation functions to restructure the input data, MLP can simulate intricate non-linear connections in data [15].
- *GPR*: A non-parametric probabilistic regression method called GPR represents the target variable as a Gaussian process. It offers estimates of prediction uncertainty in addition to point forecasts. When working with tiny datasets and accurately expressing uncertainty in forecasts is crucial, GPR is helpful [10].
- *AdaBoost*: AdaBoost is an ensemble learning technique used mainly for classification tasks but can be adapted for regression. It combines several weak learners to produce a powerful prediction model (usually decision trees). AdaBoost is particularly efficient at enhancing model performance because it weights data points differently and concentrates on samples incorrectly identified in each iteration [16].
- *Ridge*: Ridge regression is a linear regression technique that introduces L2 regularization to the linear regression model. Forbidding high coefficients improves the cost function by adding a penalty term that helps avoid overfitting. Ridge regression is helpful when multi-collinearity is present, and a more straightforward, understandable model is sought [17].

4.5.5 Hyper parameter Tuning

Hyperparameters and model parameters make up a regression model's parameters. Hyperparameters are predetermined, whereas the training step involves fine-tuning the model parameters. The hyperparameters' optimal setting varies depending on the soil nutrient and its variation with the bands' reflectance value. The method used in this work obtains the optimal hyperparameter setting by a primary grid search method and extracts the best parameter. Table 5.3 refers to all the hyperparameters used in this work for the different models and all the chemical properties.

4.5.6 Performance Metrics

The evaluation metrics chosen in the model are the RMSE and RPD for evaluating the model performance (RPD). The RMSE is a quadratic scoring rule gauges the average error's magnitude. RPD is a relationship between range and standard deviation. A good performance indicates a low value of RMSE while a high value of RPD. RMSE ranges from 0 - any finite range [46], while RPD values usually range above 2 [46].

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(predicted - actual\right)^2}$$

$$RPD = RangeReference/StandardDeviation$$

Chapter 5

Results And Discussion

5.1 Hyperspectral Image

This study consists of three techniques :

- Using min-max scaling without any feature selection (Baseline with scaling).
- PCA without any feature selection as a dimensionality reduction technique.
- Reducing features (115) and applying min-max scaling (Reduced Features).

Table 5.1: Results of Regression for Estimating Soil Moisture (Test Set)

Model	**Baseline with scaling**		**with PCA**		**Reduced Features**		
(sklearn)	**R^2(%)**	**RMSE**	**R^2(%)**	**RMSE**	**R^2(%)**	**RMSE**	**Hyperparameter Setup**
Linear	82.59	1.59	87.87	1.29	82.08	1.56	-
PLS	86.59	1.29	86.59	1.35	87.87	1.66	n_components=11,max_iter=10000
Ridge	88.8	1.1	80.42	1.63	87.62	1.3	alpha=0.05,max_iter=15000
Lasso	81.63	1.25	75.09	1.84	79.75	1.66	alpha=1,max_iter=15000,solver=lsqr
Elastic Net	80.12	1.28	73.95	1.88	78.56	1.71	alpha=0.1, l1_ratio=0.5
Decision Tree	74.74	1.64	77.76	1.57	75.33	1.64	max_depth=3
KNN [47]	96.55	0.62	81.36	1.59	96.52	0.69	k=3
XGBoost [48]	94.67	0.84	94.44	0.87	94.93	0.91	colsample_bytree= 0.3, gamma= 0.1, learning_rate= 0.15 max_depth=8, min_child_weight=5,n_estimators=600
AdaBoost [49]	93.92	0.91	400–2500	23	93.53	0.93	DecisionTreeRegressor(max_depth=6), learning_rate=0.05,n_estimators=2000, random_state=1
MLP [50]	**97.47**	**0.58**	91.59	1.07	94.34	0.87	hidden_layer_sizes=(300,100,),activation=relu, max_iter=20000,solver=lbfgs, learning_rate_init=0.05, random_state=1
SOM [36]	96.65	0.65	-	-	-	-	50 × 50 grid with 5000 iterations of the input SOM and 10 000 iterations of the output SOM [36]

Table 5.1 contains all the regression results. The baseline model with all the features gives the best performance. Linear Regression, Partial Least Squares (PLS), Ridge, Lasso, and Elastic Net perform the worst, along with Decision Tree Regression. The algorithms are not capable of solving this high-dimensional problem. Among other algorithms, XGBoost and AdaBoost give promising results, along with KNN, which

gives the optimal value at k=3. MLP gives the best working model with an R^2 of 97.47% and the least RMSE of 0.58.

This data shows that using PCA or reducing features hampers accuracy and increases the error. So applying min-max scaling with no feature selection yields the best model.

This study implements and compares the various regression models for the hyperspectral data-based regression of soil moisture, such as Random Forest (RF), Support Vector Regression (SVR), and Self Organizing Maps (SOM) on the Hyperspectral Moisture dataset [36]. The MLP outperforms all the other algorithms although by a small margin but in a cost efficient manner with respect to the SOM. All models undergo optimization of the hyperparameter. Table 5.1 compares the regression performance using the metrics R^2 and RMSE.

Table 5.2: Comparing results with the previous study

Model	**My Model**		**Riese and Keller [36]**	
	R^2(%)	**RMSE**	**R^2(%)**	**RMSE**
KNN	96.55	0.62	-	-
XGBOOST	94.67	0.84	-	-
SOM	96.65	0.65	**96.78**	**0.66**
SVR	96.65	0.64	96.03	0.74
RF	94.01	0.9	93.06	0.94
AdaBoost	93.92	0.91	-	-
MLP	**97.47**	**0.58**	-	-

Results indicate that MLP performs better than all the models. It also provides a new method of predicting hyperspectral data. Fine-tuning the model parameters increases the performance of MLP. It also shows that MLP regressor can handle high dimensional data with precision. Fine-Tuning the parameters of MLP can be time-consuming and requires computational power. Table 5.2 compares the algorithms used in this dataset with the previous work on this dataset.

5.2 Multispectral LUCAS Image

This work consists of three techniques:

- Processing Landsat 8 data for Hungary.
- Combining it with LUCAS data for Hungary.
- Develop and compare the different machine learning models.

The regression models, namely Partial Least Squares Regression (PLSR), Support Vector Regression (SVR), Gaussian Process Regression (GPR), MultiLayer Perceptron

(MLP), AdaBoost and Ridge predict the chemical properties of soil, Electrical Conductivity(EC), Calcium Carbonate ($CaCO_3$), Potassium(K), Nitrogen(N), Phosphorus(P), pH_h_2O and pH_$CaCl_2$. RMSE and RPD are the two methods used to evaluate and validate the models. This study reveals the prediction of the chemical properties of soil in Hungary. Adaboost is a boosting algorithm that performs better by combining all the weak learners and outperforms all the traditional algorithms by a significant margin. The R^2 of the models are not at par with the base paper as there is less significant amount of data in comparison with the base paper of this study. The data fetched only depicts Hungary as a focus and estimates all the chemical properties of Hungary soil. This provides a more focused research although a few more data points could significantly improve the accuracy.

Table 5.3 contains all the regression results for different soil nutrient predictions. PLSR, GPR, AdaBoost, and Ridge Regressors perform comparably. The regression results show that the soil nutrients EC, P, N, pH_H_2O, pH_$CaCl_2$ can be predicted with reasonable accuracy as the RMSE value is low and the RPD value is high. In comparison, SVR and MLP do not perform consistently for all the properties. EC (RMSE = 16.73 and RPD = 7.19), P(RMSE=23.69 and RPD=4.89), N(RMSE=0.93 and RPD=7.07), K(RMSE=197.16 and RPD=7.33), pH_H_2O(RMSE=0.93 and RPD=4.69) achieved best results using Ridge regressor. While $CaCO_3$(RMSE=72.45 and RPD=4.89) and pH_$CaCl_2$(RMSE=0.87 and RPD=4.24) achieved best results using AdaBoost regressor. The study fully explains the various regression models' propensities to predict soil properties. This work enhances the science of soil analysis by creating prediction tools with substantial implications for agricultural and environmental research.

GPR models are less interpretable in this study due to their non-parametric nature and complex function representations.

The study fully explains the various regression models' propensities to predict soil properties. Ballabio et al. [40] accuracies mentioned in table 5.3 are of the full LUCAS dataset consisting of approximately 22,000 points. At the same time, this paper focuses on predicting soil chemical properties in Hungary only with only 400 points. This study proposes an AdaBoost algorithm to predict the chemical properties when working with a smaller dataset. The GPR algorithm, due to its complex function, is unable to perform well on a smaller dataset. AdaBoost performs better and more efficiently than GPR on a region-specific survey with minimal data points.

Table 5.3: Results of chemical property estimation by regression models

Electrical Conductivity(EC)				**Calcium Carbonate (CaCO3)**			
Model	**RMSE**	**RPD**	**Parameters**	**Model**	**RMSE**	**RPD**	**Parameters**
PLSR	16.86	7.25	n_components=5, scale=False,max_iter=200	**PLSR**	75.37	4.93	n_components=2,scale=False, max_iter=100
SVR	19.15	6.37	kernel = ’rbf’,C=100,gamma=1.15	**SVR**	73.91	5.11	kernel = ’rbf’,C=100,gamma=1.15
MLP	19.36	7.17	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05	**MLP**	78.37	4.94	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05
GPR	17.07	7.16	normalize_y=False, random_state=1, alpha=10,length_scale=3.0, length_scale_bounds="fixed"	**GPR**	72.48	4.91	normalize_y=False, random_state=1, alpha=0.8,length_scale=10.0, length_scale_bounds="fixed"
AdaBoost	16.92	7.19	n_estimators=8000,learning_ rate=0.0001,random_state=5, base_estimator=DTR	**AdaBoost**	72.45	4.99	n_estimators=3000,learning_rate=0.01, random_state=1, base_estimator=DTR
Ridge	16.73	7.27	alpha=0.01,max_iter=15000, solver=’lsqr’	**Ridge**	73.9	4.89	alpha=10000,max_iter=50000,solver=’svd’
Ballabio [3] GPR	11.02	-	-	**Ballabio [3] GPR**	78.29	-	
Pottasium(K)				**Phosphorus(P)**			
Model	**RMSE**	**RPD**	**Parameters**	**Model**	**RMSE**	**RPD**	**Parameters**
PLSR	201.41	7.32	n_components=5, scale=False,max_iter=500	**PLSR**	23.77	4.88	
SVR	207.37	7.19	kernel = ’rbf’,C=100,gamma=1.15	**SVR**	24.37	4.83	kernel = ’rbf’,C=100,gamma=1.15
MLP	208.56	7.37	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05	**MLP**	23.75	3.56	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05
GPR	199.42	7.3	normalize_y=False, random_state=1, alpha=18,length_scale=5.0, bounds="fixed"	**GPR**	23.97	4.86	normalize_y=False, random_state=1, alpha=15 length_scale=13.0, length_scale_bounds="fixed"
AdaBoost	199.81	7.24	n_estimators=3000,learning_ rate=0.001, random_state=1, base_estimato r=DTR	**AdaBoost**	23.85	4.88	n_estimators=3000,learning_rate=0.01, random_state=1, base_estimator=DTR
Ridge	197.16	7.33	alpha=0.01,max_iter=15000, solver=’lsqr’	**Ridge**	23.69	4.89	alpha=0.5,max_iter=15000,solver=’lsqr’
Ballabio [3] GPR	199.17	-	-	**Ballabio [3] GPR**	17.52	-	
Nitrogen(N)				**pH H2O**			
Model	**RMSE**	**RPD**	**Parameters**	**Model**	**RMSE**	**RPD**	**Parameters**
PLSR	1.21	7.3	n_components=1, scale=False,max_iter=200	**PLSR**	0.93	4.54	n_components=2, scale=False,max_iter=200
SVR	1.31	5.91	kernel = ’rbf’,C=100,gamma=1.15	**SVR**	1.28	3.49	kernel = ’rbf’,C=100,gamma=1.15
MLP	1.20	7.45	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05	**MLP**	0.98	4.69	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05
GPR	1.05	7.44	normalize_y=False, random_state=1, alpha=18,length_scale=10.0, length_scale_bounds="fixed"	**GPR**	0.99	4.52	length_scale=10.0, length_scale _bounds="fixed",normalize_y=False, random_state=1, alpha=18
AdaBoost	1.08	7.22	n_estimators=3000,learning_ ’rate=0.0005, random_state=1, base_estimator =DTR	**AdaBoost**	0.95	4.59	n_estimators=4000 ,learning_rate=0.01, random_state=1, base_estimator =DTR
Ridge	1.12	7.07	alpha=1,max_iter=15000,solver=’svd’	**Ridge**	0.93	4.69	alpha=0.5,max_iter=1000,solver=’saga’
Ballabio [3] GPR	0.78	-	-	**Ballabio [3] GPR**	0.78	-	-
pH CaCl2							
Model	**RMSE**	**RPD**	**Parameters**				
PLSR	0.88	4.22	n_components=7, scale=False,max_iter=200				
SVR	1.22	3.1	kernel = ’rbf’,C=100 ,gamma=1.15				
MLP	0.89	4.15	hidden_layer_sizes=(3,2,), activation=’relu’, max_iter=20000,solver=’lbfgs’, learning_rate_init=0.05				
GPR	0.91	4.06	length_scale=9.0, length_scale_ bounds="fixed",normalize_y=False, random_state=1, alpha=10, n_restarts_optimizer=10				
AdaBoost	0.87	4.24	n_estimators=3000,learning_rate=0.1, random_state=1, base_estimator=DTR				
Ridge	0.88	4.22	alpha=0.1,max_iter=15000,solver=’svd’				
Ballabio [3] GPR	0.68	-	-				

Chapter 6

Conclusion

This study discusses the estimation of soil moisture by regression using hyperspectral data. It aims to improve the accuracy score of the SOM framework on the hyperspectral soil moisture dataset proposed by Riese and Keller. Different machine learning algorithms perform the regression on high dimensional hyperspectral pre-processed data. MultiLayer Perceptron Regressor (MLP) without any feature selection and applying normalization outperforms all the other regression models. MLP can handle high-dimensional data. Nevertheless, the time required to fine-tune the parameters to obtain the optimal results is time-consuming with a high-dimensional dataset. The regression framework's findings show the potential of various data-driven models when combined with the input data utilized in various real-world measurement scenarios. It also shows the potential for predicting soil/crop properties using VNIR hyperspectral data using different machine-learning approaches.

This work discusses estimating soil chemical properties by regression using Landsat 8 multispectral data. Different machine learning algorithms perform the regression on multispectral data combined with LUCAS. It accurately predicts the chemical properties of the soil of Hungary. Such region-specific analyses can yield valuable insights and recommendations tailored to that area's unique soil and environmental conditions. Ridge and AdaBoost outperform all the machine learning models and previous studies [10]. The models' predictive capacity may be useful for monitoring land, the environment, and precision agriculture. This work showcases the detailed processing of combining and processing satellite images with soil properties.

This study can address the specific challenges and opportunities related to soil and land management of the Hungary region along with digital soil mapping and spatial analysis of the region. Future Research may examine long-term trends of soil properties, the climate change's effects on Hungary's soils, or the development of site-specific management practices.

Bibliography

[1] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sensing*, vol. 12, no. 16, p. 2659, 2020.

[2] F. M. Riese and S. Keller, "Hyperspectral benchmark dataset on soil moisture," in *Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain*, 2018, pp. 22–27.

[3] V. Gonzalez-Dugo, P. Hernandez, I. Solis, and P. J. Zarco-Tejada, "Using high-resolution hyperspectral and thermal airborne imagery to assess physiological condition in the context of wheat phenotyping," *Remote Sensing*, vol. 7, no. 10, pp. 13 586–13 605, 2015.

[4] M. W. Carroll, J. A. Glaser, R. L. Hellmich, T. E. Hunt, T. W. Sappington, D. Calvin, K. Copenhaver, and J. Fridgen, "Use of spectral vegetation indices derived from airborne hyperspectral imagery for detection of european corn borer infestation in iowa corn plots," *Journal of Economic Entomology*, vol. 101, no. 5, pp. 1614–1623, 2008.

[5] C. Ryu, M. Suguri, and M. Umeda, "Multivariate analysis of nitrogen content for rice at the heading stage using reflectance of airborne hyperspectral remote sensing," *Field Crops Research*, vol. 122, no. 3, pp. 214–224, 2011.

[6] B. Lu and Y. He, "Evaluating empirical regression, machine learning, and radiative transfer modelling for estimating vegetation chlorophyll content using bi-seasonal hyperspectral images," *Remote Sensing*, vol. 11, no. 17, p. 1979, 2019.

[7] Y. Fenghua, X. Tongyu, D. Wen, M. Hang, Z. Guosheng, and C. Chunling, "Radiative transfer models (rtms) for field phenotyping inversion of rice based on uav hyperspectral remote sensing," *International Journal of Agricultural and Biological Engineering*, vol. 10, no. 4, pp. 150–157, 2017.

[8] M. Weiss, F. Jacob, and G. Duveiller, "Remote sensing for agricultural applications: A meta-review," *Remote sensing of environment*, vol. 236, p. 111402, 2020.

[9] S. Jain and D. Sethia, "A review on applications of artificial intelligence for identifying soil nutrients," in *International Conference on Agriculture-Centric Computation*. Springer, 2023, pp. 71–86.

[10] C. Ballabio, E. Lugato, O. Fernández-Ugalde, A. Orgiazzi, A. Jones, P. Borrelli, L. Montanarella, and P. Panagos, "MAPPING LUCAS TOPSOIL CHEMICAL PROPERTIES AT EUROPEAN SCALE USING GAUSSIAN PROCESS REGRESSION," *Geoderma*, vol. 355, p. 113912, 2019.

[11] G. Tóth, A. Jones, and L. Montanarella, "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, pp. 7409–7425, 2013.

[12] A. Miglani, S. Ray, R. Pandey, and J. Parihar, "Evaluation of eo-1 hyperion data for agricultural applications," *Journal of the Indian Society of Remote Sensing*, vol. 36, pp. 255–266, 2008.

[13] G. P. Asner, R. E. Martin, C. B. Anderson, and D. E. Knapp, "Quantifying forest canopy traits: Imaging spectroscopy versus field survey," *Remote Sensing of Environment*, vol. 158, pp. 15–27, 2015.

[14] L. Pasolli, C. Notarnicola, and L. Bruzzone, "Estimating soil moisture with the support vector regression technique," *IEEE Geoscience and remote sensing letters*, vol. 8, no. 6, pp. 1080–1084, 2011.

[15] L. Estep, G. Terrie, and B. Davis, "Crop stress detection using aviris hyperspectral imagery and artificial neural networks," 2004.

[16] N. Lin, R. Jiang, G. Li, Q. Yang, D. Li, and X. Yang, "Estimating the heavy metal contents in farmland soil from hyperspectral images based on stacked adaboost ensemble learning," *Ecological Indicators*, vol. 143, p. 109330, 2022.

[17] M. Imani and H. Ghassemian, "Ridge regression-based feature extraction for hyperspectral data," *International Journal of Remote Sensing*, vol. 36, no. 6, pp. 1728–1742, 2015.

[18] M. Shoshany, N. Goldshleger, and A. Chudnovsky, "Monitoring of agricultural soil degradation by remote-sensing methods: A review," *International Journal of Remote Sensing*, vol. 34, no. 17, pp. 6152–6181, 2013.

[19] C. Yang, "Airborne hyperspectral imagery for mapping crop yield variability," *Geography Compass*, vol. 3, no. 5, pp. 1717–1731, 2009.

[20] Y. Akhtman, E. Golubeva, O. Tutubalina, and M. Zimin, "Application of hyperspectural images and ground data for precision farming," *Geography, environment, sustainability*, vol. 10, no. 4, pp. 117–128, 2017.

[21] C. Cilia, C. Panigada, M. Rossini, M. Meroni, L. Busetto, S. Amaducci, M. Boschetti, V. Picchi, and R. Colombo, "Nitrogen status assessment for variable rate fertilization in maize through hyperspectral imagery," *Remote Sensing*, vol. 6, no. 7, pp. 6549–6565, 2014.

[22] M. P. Finn, M. Lewis, D. D. Bosch, M. Giraldo, K. Yamamoto, D. G. Sullivan, R. Kincaid, R. Luna, G. K. Allam, C. Kvien *et al.*, "Remote sensing of soil moisture

using airborne hyperspectral data," *GIScience & Remote Sensing*, vol. 48, no. 4, pp. 522–540, 2011.

[23] R. Casa, F. Castaldi, S. Pascucci, A. Palombo, and S. Pignatti, "A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing," *Geoderma*, vol. 197, pp. 17–26, 2013.

[24] X. Song, G. Yan, J. Wang, L. Liu, X. Xue, C. Li, and W. Huang, "Use of airborne hyperspectral imagery to investigate the influence of soil nitrogen supplies and variable-rate fertilization to winter wheat growth," in *Remote Sensing for Agriculture, Ecosystems, and Hydrology IX*, vol. 6742. SPIE, 2007, pp. 216–225.

[25] F. Antonucci, P. Menesatti, N. Holden, E. Canali, S. Giorgi, A. Maienza, and S. R. Stazi, "Hyperspectral visible and near-infrared determination of copper concentration in agricultural polluted soils," *Communications in soil science and plant analysis*, vol. 43, no. 10, pp. 1401–1411, 2012.

[26] W.-j. Wang, Z.-w. Li, C. Wang, D. Zheng, and H. Du, "Prediction of available potassium content in cinnamon soil using hyperspectral imaging technology," *Spectrosc. Spect. Anal*, vol. 39, pp. 1579–1585, 2019.

[27] C. McCann, K. S. Repasky, R. Lawrence, and S. Powell, "Multi-temporal mesoscale hyperspectral data of mixed agricultural and grassland regions for anomaly detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 131, pp. 121–133, 2017.

[28] S. B. Wankhede, "Analytical study of neural network techniques: Som, mlp and classifier-a survey," *IOSR Journal of Computer Engineering*, vol. 16, no. 3, pp. 86–92, 2014.

[29] A. Sharma, X. Liu, and X. Yang, "Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks," *Neural Networks*, vol. 105, pp. 346–355, 2018.

[30] Y. Xu, L. Wu, Z. Xie, and Z. Chen, "Building extraction in very high resolution remote sensing imagery using deep learning and guided filters," *Remote Sensing*, vol. 10, no. 1, p. 144, 2018.

[31] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.

[32] K. Kuwata and R. Shibasaki, "Estimating crop yields with deep learning and remotely sensed data," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 858–861.

[33] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, "3d convolutional neural networks for crop classification with multi-temporal remote sensing images," *Remote Sensing*, vol. 10, no. 1, p. 75, 2018.

[34] B. Siegmann and T. Jarmer, "Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral

data," *International journal of remote sensing*, vol. 36, no. 18, pp. 4519–4534, 2015.

[35] B. Lu and Y. He, "Species classification using unmanned aerial vehicle (uav)-acquired high spatial resolution imagery in a heterogeneous grassland," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 128, pp. 73–85, 2017.

[36] F. M. Riese and S. Keller, "Introducing a framework of self-organizing maps for regression of soil moisture with hyperspectral data," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 6151–6154.

[37] S. Keller, F. M. Riese, J. Stötzer, P. M. Maier, and S. Hinz, "Developing a machine learning framework for estimating soil moisture with vnir hyperspectral data," *arXiv preprint arXiv:1804.09046*, 2018.

[38] T. E. Oliphant, "Python for scientific computing, computing in science & engineering," 2007.

[39] D. Pflugmacher, A. Rabe, M. Peters, and P. Hostert, "Mapping pan-european land cover using landsat spectral-temporal metrics and the european lucas survey," *Remote sensing of environment*, vol. 221, pp. 583–595, 2019.

[40] C. Ballabio, P. Panagos, and L. Monatanarella, "Mapping topsoil physical properties at european scale using the lucas database," *Geoderma*, vol. 261, pp. 110–123, 2016.

[41] B. Ghassemi, A. Dujakovic, M. Żółtak, M. Immitzer, C. Atzberger, and F. Vuolo, "Designing a european-wide crop type mapping approach based on machine learning algorithms using lucas field survey and sentinel-2 data," *Remote sensing*, vol. 14, no. 3, p. 541, 2022.

[42] F. M. Riese and S. Keller, "Soil texture classification with 1d convolutional neural networks based on hyperspectral data," *arXiv preprint arXiv:1901.04846*, 2019.

[43] F. Castaldi, S. Chabrillat, A. Don, and B. van Wesemael, "Soil organic carbon mapping using lucas topsoil database and sentinel-2 data: An approach to reduce soil moisture and crop residue effects," *Remote Sensing*, vol. 11, no. 18, p. 2121, 2019.

[44] F. Castaldi, S. Chabrillat, C. Chartin, V. Genot, A. Jones, and B. van Wesemael, "Estimation of soil organic carbon in arable soil in belgium and luxembourg with the lucas topsoil database," *European Journal of Soil Science*, vol. 69, no. 4, pp. 592–603, 2018.

[45] T. D. Acharya and I. Yang, "Exploring landsat 8," *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, vol. 4, no. 4, pp. 4–10, 2015.

[46] A. Gholizadeh, D. Žižala, M. Saberioon, and L. Borvka, "Soil organic carbon and texture retrieving and mapping using proximal, airborne and sentinel-2 spectral imaging," *Remote Sensing of Environment*, vol. 218, pp. 89–103, 2018.

[47] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[48] L. Breiman, "Arcing the edge," Citeseer, Tech. Rep., 1997.

[49] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[50] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5-6, pp. 183–197, 1991.