DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS,

SHARDA SCHOOL OF ENGINEERING AND TECHNOLOGY

SHARDA UNIVERSITY, GREATER NOIDA

Soil Analysis using Deep Learning For Precision Agriculture

A project report submitted
in partial fulfillment of the requirements for the
Masters in Computer Science

by

Abdullahi Muhammad (2022806061)

Thierry Forsack Fotabong (2022801870)

Murphy Keisham (2022393037)

Supervised by:
Dr. Rajneesh Kumar Singh

April, 2024

CERTIFICATE

This is to certify that the report entitled "**Soil Analysis Using Deep Learning For Precision Agriculture**" submitted by **"Mr. Abdullahi Muhammad (2022806061)", "Mr. Thierry Forsack Fotabong (2022801870)" and "Mr. Murphy Keisham (2022393037)"** to Sharda University, towards the fulfilment of requirements of the **"Master of Computer Science"** is record of bonafide final year Project work carried out by us them in the "Department of Computer Science and Applications, Sharda School of Engineering and Technology, Sharda University". The results/findings contained in this Project have not been submitted in part or full to any other University/Institute for award of any other Degree/Diploma.

Signature of Supervisor
Name: Dr. Rajneesh Kumar Singh
Designation: Assistant Professor

Signature of Head of Department
Name: Dr. Arun Agrawal

ABSTRACT

In the contemporary landscape of soil analysis, soil classification emerges as a pivotal area of exploration and inquiry. This process plays a fundamental role in comprehending and evaluating the performance of soil, essential for discerning its suitability for applications such as agriculture. However, conventional methods employed by farmers often prove inadequate in meeting the burgeoning demands of modern agriculture, leading to disruptions in soil cultivation practices. While a plethora of laboratory and field techniques exist for soil classification, including statistical approaches, rule-based systems, and traditional learning methods, many are fraught with limitations such as time-consuming procedures and reliance on domain expert opinion. Despite the abundance of methodologies, achieving precise and accurate soil classification outcomes remains a challenge. In response to this critical need, we propose a novel approach to soil classification that leverages advanced techniques in image preprocessing and feature extraction. Specifically, we employ the Gabor filter to extract salient features from diverse soil images, thereby enhancing the discriminative power of the classification process. These extracted features are then subjected to classification using a Convolutional Neural Network (CNN) classifier, known for its prowess in handling complex data patterns and structures. Through rigorous experimentation and evaluation, our proposed method demonstrates an impressive recognition rate of 98%, underscoring its efficacy and potential for advancing the field of soil classification. This research endeavor represents a significant step towards enhancing the accuracy and efficiency of soil classification techniques, with far-reaching implications for agricultural productivity.

ACKNOWLEDGEMENT

A major project is a golden opportunity for learning and self- development. We consider our self very lucky and honored to have so many wonderful people lead us through in completion of this project.

First and fore most we would like to thank Dr. Arun Agrawal, HOD, CSA who gave us an opportunity to undertake this project.

Our grateful thanks to Dr. Rajneesh Kumar Singh for his guidance in our project work. who in spite of being extraordinarily busy with academics, took time out to hear, guide and keep us on the correct path. We do not know where we would have been without his help.

CSA department monitored our progress and arranged all facilities to make life easier. We choose this moment to acknowledge their contribution gratefully.

Name and signature of Students

Abdullahi Muhammad (2022806061)

Thierry Forsack Fotabong (2022801870)

Murphy Keisham (2022393037)

TABLE OF CONTENTS

Table of Contents

CERTIFICATE I

ABSTRACT II

ACKNOWLEDGEMENT III

TABLE OF CONTENTS IV

TABLE OF FIGURES VI

CHAPTER 1 1

1.1 INTRODUCTION 1
1.2 PROBLEM STATEMENT 8
1.3 MOTIVATION 9
1.4 OBJECTIVES 10
SUMMARY 11

CHAPTER 2 12

2.1 LITERATURE REVIEW 12
SUMMARY 25

CHAPTER 3 26

3.1 PROPOSED METHOD 26

CHAPTER 4 31

4.1 METHODOLOGY 31
4.1.1 STAGE A: IMAGE ACQUISITION AND PRE-PROCESSING 31
4.1.3 STAGE C: MODEL TRAINING AND EVALUATION 32
STAGE A 39
STAGE B 41
STAGE C 42

CHAPTER 5 44

5.1 IMPLEMENTATION 44
DATA AUGMENTATION 44
CONVOLUTIONAL NEURAL NETWORK (CNN) 45
SETUP 45
DATA 46

MODEL BUILDING 48
MODEL TESTING 49
SOIL CLASSIFICATION PERCENTAGES 50
LOAD MODEL 52
CLASSIFY IMAGE 52
5.2 ALGORITHM (PSEUDO CODE) 54
5.3 RESULT AND DISCUSSIONS. 56

CHAPTER 6 60

6.1 CONCLUSION 60
6.2 FUTURE SCOPE 61
REFERENCES 62

Table Of Figures

Description Pages

Figure 1 Proposed Architecture 27
Figure 2: CNN Architecture...29
Figure 3: The comprehensive framework used for our research on soil analysis 33
Figure 4: Stage A of the methodology. 40
Figure 5: Stage B of the methodology 42
Figure 6: Stage C of the methodology 43
Figure 7: The project setup. 46
Figure 8: The data directory is provided. 47
Figure 9: The data processing 47
Figure 10: The image pre-processing. 47
Figure 11: The model parameters 48
Figure 12: Building the model 49
Figure 13: Model testing. 50
Figure 14: Soil Classification. 51
Figure 15: Loading the model 52
Figure 16: Classifying the images. 53
Figure 17: Data processed. 54
Figure 18: Model summary 54
Figure 19: Graph showing Validation loss vs Epochs 55
Figure 20: Loading soil for Validation. 56
Figure 21: Accuracy for the validation 57

CHAPTER 1

Soil analysis is the cornerstone of modern agricultural practices, providing crucial insights into the health and fertility of the soil. By assessing various physical, chemical, and biological properties, farmers and researchers can make informed decisions about crop selection, nutrient management, and soil conservation strategies.

1.1 Introduction

Soil analysis is the cornerstone of modern agricultural practices, providing crucial insights into the health and fertility of the soil. By assessing various physical, chemical, and biological properties, farmers and researchers can make informed decisions about crop selection, nutrient management, and soil conservation strategies [1]. Soil, the foundational element of agricultural ecosystems, exhibits a remarkable complexity that has long challenged conventional analysis methods. Its physical, chemical, and biological properties vary extensively across different regions, influenced by a myriad of factors such as climate, topography, and management practices [2]. Accurately mapping and understanding these intricate soil characteristics is crucial for making informed decisions regarding crop selection, nutrient management, and irrigation strategies. In recent years, the integration of deep learning (DL) techniques into soil analysis has revolutionized the field. DL algorithms excel at extracting intricate patterns and relationships from large datasets, making them well-suited for interpreting complex soil characteristics across diverse landscapes[3]. By leveraging advanced neural networks, researchers can now analyze soil samples more efficiently and accurately, leading to enhanced productivity and sustainability in agriculture. Precision agriculture (PA) complements these advancements by enabling farmers to adopt a targeted and responsive approach to land management. By utilizing technologies such as GPS, remote sensing, and IoT devices, PA facilitates real-time monitoring and precise control of farming practices. This proactive strategy minimizes input waste, maximizes yield potential, and reduces environmental impact, ultimately fostering more resilient and profitable agricultural systems

Building upon the synergy between DL and PA, this study aims to explore the potential of image processing and neural networks in soil classification and quality forecasting [5]. By employing state-of-the-art algorithms, researchers seekto develop robust models capable of accurately identifying different soil types and predicting their suitability for specific crops. Through the integration of high-resolution soil images, spectral data, and historical records, these models aim to provide farmers with actionable insights to optimize crop selection, irrigation scheduling, and fertilizer application, thereby enhancing overall farm efficiency and sustainability [6]. Moreover, the application of DL in soil analysis holds promise for addressing emerging challenges such as climate change, soil degradation, and food security. By harnessing the power of big data and machine learning, researchers can uncover novel insights into soil health dynamics, enabling proactive adaptation strategies and informed policy making [7] . Ultimately, the convergence of cutting-edge technologies and agricultural science paves the way for a more resilient and productive future in farming.

In the ever-evolving landscape of agriculture, the quest for maximizing crop yield while minimizing resource consumption has been a perpetual endeavor. Precision agriculture has emerged as a transformative approach, leveraging technology to optimize farming practices [8]. At the heart of this revolution lies soil analysis – a fundamental aspect influencing crop growth. Traditional methods of soil analysis are labor-intensive, time-consuming, and often lack the precision required for informed decision-making. However, the integration of deep learning techniques offers a promising solution to revolutionize soil analysis for precision agriculture. The fusion of deep learning with soil analysis presents a paradigm shift in agricultural practices. [9] deep learning is a subset of artificial learning is very good at analysing enormous volumes of data to find complex trends and insights. Deep neural network technology [10] allows us to extract useful information from soil data, empowering farmers to make data-driven decisions with previously unheard-of precision and efficiency. The use of deep learning techniques for soil analysis in precision agriculture is explored in this research report. By applying cutting-edge algorithms and computational methods, we want to create a solid framework that can quickly and reliably analyse soil attributes. [11] Using information from several sources, including sensors, satellite imaging, and historical records, our model aims to offer a comprehensive picture of soil conditions in different agricultural settings.. The significance of this project extends beyond mere efficiency gains; it encompasses environmental sustainability and economic viability. By precisely characterizing soil properties such as nutrient levels, moisture content, and soil texture, farmers can optimize fertilization, irrigation, and planting strategies, thereby minimizing input wastage and environmental degradation. Moreover, by maximizing crop yield and quality, farmers can enhance their economic returns, fostering long-term agricultural resilience.

The methodology adopted in this project encompasses several key stages, including data collection, preprocessing, model development, and validation. The process of collecting data entails assembling a variety of datasets, including information on crops, satellite imaging, weather, and soil samples. To get the datasets ready for model training, preprocessing methods including feature extraction, normalisation, and data cleaning are used. [13] Moreover, large datasets covering a variety of soil types, climates, and agricultural techniques may be used to train deep learning models, enabling the creation of reliable and models that may be used broadly. These models can then be deployed in the field, leveraging readily available remote sensing data or even smartphone imagery, to provide real-time soil analysis and recommendations tailored to specific locations and conditions.

There are several uses for deep learning in soil analysis for precision farming. For example, these models can map soil parameters including texture, pH, organic matter content, and nutrient levels with accuracy [14], allowing farmers to decrease their environmental impact and optimise fertiliser application rates. Furthermore, deep learning may help locate regions that are vulnerable to compaction, erosion, or waterlogging, guiding focused actions to reduce soil deterioration and increase crop output.

Furthermore, there is a great deal of promise in combining deep learning methods with other cutting-edge technology like unmanned aerial vehicles (UAVs) and Internet of Things (IOT) gadgets. Sensor networks and aerial imagery can provide continuous streams of data, which deep learning models can process and analyze in real-time, facilitating dynamic monitoring and adaptive management strategies throughout the growing season[16]. However, the implementation of deep learning in soil analysis is not without challenges. Access to high-quality, labeled datasets remains a significant obstacle, as the creation of such datasets requires extensive fieldwork, laboratory analysis, and expert annotation. Additionally, the interpretability and explainability of deep learning models can be challenging, potentially hindering their adoption by stakeholders who require transparent decision-making processes.

Furthermore, the report will present a comprehensive analysis of the datasets utilized in this project, detailing the data collection methods, preprocessing steps, and the techniques employed to ensure data quality and representativeness. This transparency is crucial for enabling reproducibility and facilitating future research in this domain. The creation of deep learning models specifically for soil analysis is the main goal of this research [17]. Because they are good at managing temporal and spatial input, respectively, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are two of the designs that are investigated. Through extensive experimentation and optimization, we aim to construct models capable of accurately predicting soil properties based on input data. The experimental design and use of deep learning models for soil analysis tasks will be the main topics of the paper. It will give a thorough explanation of the model designs, training methods, and assessment measures used. The paper will also go into the methods used to deal with possible problems such overfitting, unequal class distribution, and limited computational resources. [18] The paper will include a thorough assessment and benchmarking procedure to verify the efficacy of the suggested deep learning models. In order to do this, a variety of soil attributes and geographical locations will be compared to both other machine learning approaches and conventional soil analysis methods. The performance trade-offs between model complexity, accuracy, and computing efficiency will also be looked at in the paper. Recognizing the importance of interpretability and transparency in decision- making processes, the report will dedicate a section to exploring techniques for model interpretation and explainability. This will involve visualizing and analyzing the learned feature representations, as well as investigating the relationships between input data (e.g., spectral signatures) and model predictions. Additionally, the report will address the ethical considerations and potential societal implications of deploying deep learning technologies in precision agriculture. It will explore issues related to data privacy, algorithmic [20] bias, and the digital divide, proposing strategies to ensure equitable access and responsible use of these technologies across diverse agricultural communities [21]. Building upon the insights gained from this project, the report will outline a roadmap for future research and development in the field of deep learning for soil analysis. This will involve identifying key challenges and opportunities, proposing innovative approaches, and highlighting potential collaborations and interdisciplinary synergies.

Moreover, the report will emphasize the importance of knowledge transfer and capacity building within the agricultural community. The discourse will delve into tactics for distributing research outcomes, offering instructional and training materials, and cultivating partnerships among scholars, extension agencies, and agricultural practitioners to enable the extensive integration of deep learning technology in precision farming. The report's ultimate goal is to use deep learning and sophisticated data analytics to further the cause of productive and sustainable agriculture. Through the utilisation of state-of-the-art technology, farmers are able to make well-informed decisions, optimise resource utilisation, and adjust to changing environmental circumstances. This guarantees food security and environmental stewardship for future generations.

Additionally, the paper will illustrate the ways in which deep learning for soil analysis and other cutting-edge technologies, including distributed ledger systems and blockchain. These technologies can facilitate secure and transparent data sharing, enabling collaborative research efforts and fostering trust among stakeholders in the agricultural ecosystem.

In addition to the technical aspects,[23] the report will also explore the economic and policy implications of adopting deep learning technologies in precision agriculture. It will analyze the potential cost savings, increased profitability, and market opportunities associated with these technologies, as well as the regulatory frameworks and incentive structures needed to support their widespread adoption.
Recognizing the global nature of agriculture and the diverse cultural and socioeconomic contexts in which it operates, the report will emphasize the importance of tailoring deep learning solutions to local conditions and stakeholder needs. In order to guarantee that smallholder farmers and underprivileged communities can access and benefit from new technologies, it will emphasise the need of inclusive and participatory methods [24].

To sum up, this research report showcases the revolutionary potential of deep learning in soil analysis and makes a substantial contribution to the field of precision agriculture. This study aims to empower farmers, advance sustainable practices, and clear the path for a more resilient and productive agricultural future by utilising cutting-edge technology, rigorous experimentation, and multidisciplinary collaboration.

1.2 Problem Statement

Traditional soil analysis methods are often inefficient and inaccurate, resulting in lower agricultural yields. To address this challenge, there is a need to developer system that seamlessly integrates image processing and deep learning techniques to accurately assess soil properties. The ultimate goal is to provide farmers with practical knowledge so they can make decisions that will increase crop yield and agricultural sustainability.

Through the application of cutting-edge technologies like deep learning and image processing, researchers want to build a solid platform that offers suggestions in real time that are customised to the unique requirements of farmers. With this method, farmers can monitor soil health more effectively and carry out focused interventions to maximise agricultural production while reducing environmental impact. It also expedites the process of soil analysis.

1.3 Motivation

By 2050, there will be more than 9.7 billion people on the planet, a startling rise from the current 7.9 billion people [27]. It is certain that the world's already finite agricultural resources would be severely strained by this unheard-of increase. In order to supply the growing world population with the nutrients they need, there will be a significant increase in the need for food production. However, the agricultural sector, particularly in developing nations, is currently grappling with numerous challenges that threaten to impede its ability to keep pace with this escalating demand.

In many developing countries, traditional farming practices, limited access to modern technologies, soil degradation, water scarcity, inadequate nutrient management, and the impacts of climate change have resulted in suboptimal crop yields and inefficient resource utilization [28]. Experts and policymakers have recognized the urgency of addressing these challenges and have proposed specific management strategies based on existing technologies. However, the sheer scale and complexity of the problem necessitate the exploration and adoption of innovative approaches that can revolutionize agricultural practices, optimize resource utilization, and enhance sustainability. This project aims to develop cutting-edge deep learning models for soil analysis in precision agriculture, harnessing the power of advanced algorithms and vast datasets to extract valuable insights, [29] facilitate targeted interventions, and support adaptive management strategies tailored to local conditions.

1.4 Objectives

The objectives of this study are:

O1. To evaluate and assess the effectiveness of current soil analysis techniques.

O2. To utilize deep learning techniques for the classification of soil types.

O3. To provide farmers with enhanced insights derived from the classificationof soil types.

Summary

In order to provide precise and effective soil analysis for precision agriculture, this report explores the revolutionary potential of deep learning approaches, particularly convolutional neural networks (CNNs) and deep neural networks (DNNs). Utilising extensive datasets from various sources, including satellite imagery, sensor data, and historical records, the goal is to create reliable models that can forecast soil properties, optimise agricultural techniques, allow for real-time monitoring, and support adaptive management approaches [25]. In addition to addressing issues with data quality, interpretability, and computing limitations, the paper examines approaches for data collecting, preprocessing, model construction, and validation. The text emphasises the potential uses of soil property mapping, degradation-prone region identification, fertiliser and irrigation strategy optimisation, and integration with cutting-edge technologies like as blockchain and IoT. This research aims to harness the power of deep learning and advanced data analytics for sustainable, productive, and resilient agricultural systems, ultimately supporting food security and environmental stewardship. It emphasises interdisciplinary [26] collaboration, knowledge transfer, responsible adoption, and the exploration of economic and policy implications

CHAPTER 2

crop recommendation and soil classification using deep learning techniques, According to Soil Series, given its variety, soil is essential for farmers. Crop yields are affected by the types of soil since some soils are more suitable for particular crops than others.

2.1 Literature Review

In this study, crop recommendation and soil classification using deep learning techniques, According to Soil Series, given its variety, soil is essential for farmers. Crop yields are affected by the types of soil since some soils are more suitable for particular crops than others. Determining whether crops do well in a given type of soil requires an understanding of the properties and functions of soil. Even though deep learning has advanced significantly in recent years, agricultural analytic research in this field is still difficult and constantly changing. In this study, we present a system that, if accurate, may anticipate soil series based on land type and recommend crops that are compatible for those conditions. Several methods are used for soil classification, such as Gaussian kernel-based algorithms, Bagged Trees, and Weighted k-Nearest Neighbour (k-NN) [30].

A deep learned set of rules may be used for automated soil typing in a research comparing deep learning algorithms for soil type categorization [31]. A number of deep learning techniques, such as Naive Bayesian and Neural Network Decision Tree, are examined for detecting various types of soil. The soil dataset is based on actual data, and RapidMiner Studio runs the simulation; accuracy serves as a performance indicator [32 Notably, SVM using a linear feature kernel performs better than alternative techniques, with an ccuracy of 82.35%. [33]. Furthermore, the detrimental impacts of soil sweing are addressed in a study on enhanced clay soil expansion prediction usng machine learning approaches, which provides an advanced evaluation utilising kernelized machines.Various methods, including Linear Regressor (LR), Bayesian Linear Regression, Voting (VE) and stacking (SE) are two examples of meta-heuristic classifiers used, along with Bayes Point Machine (BPM) and Support Vector Machine. For best outcomes, the study advises using these ideas in the first phases of geotechnical and geological site classification.

Agricultural research has benefited greatly from the study of soil data and the prediction of soil properties through the use of classifier approaches. Automation, data mining techniques, and technical advancements have brought significant value to this industry [35]. Although there are a plethora of statistical data mining techniques and services available, agricultural soil databases are a relatively young field of study. In order to categorise different types of soil using different methods, this study analyses a soil dataset using data mining techniques. Additionally, it aims to predict unknown soil properties through regression analysis and automated categorization of soil samples.

Prof. A.V. Deorankar and Ashwini A. Rohankar suggested a method in their paper for classifying soil texture using hyperspectral data [37]. By putting into practice three 1-dimensional (1D) convolutional neural networks (CNNs), they made significant advances in the field: the Lucas CNN, the Lucas ResNet, which functions as a residual network with an identity block, and the Lucas Coord Conv, which has an extra coordinates layer. They also adjusted two pre-existing 1D CNN techniques to better fit the current classification job. On GitHub, the code for each of the five CNN techniques is openly accessible (Riese, 2019) [38]. The CNN techniques were thoroughly assessed by the researchers, who also contrasted their results with those of a random forest classifier. They used the publicly available LUCAS topsoil dataset for their investigation. Surprisingly, the CNN approach with the least depth emerged as the top-performing classifier. Moreover, the Lucas Coord Conv model exhibited the highest performance in terms of average accuracy [39]. In future research endeavors, there is potential to further enhance the introduced Lucas CNN, Lucarini, and Lucas Coord Conv models. Additionally, the inclusion of extra variables from the rich LUCAS dataset could contribute to improving the classification accuracy and overall effectiveness of the soil texture classification system.

F. M. Riese and S. Keller proposed a comprehensive approach for soil classification based on image processing techniques [40]. The initial step in their methodology involves gathering diverse soil sample images, which is crucial for selecting suitable sensor dataThis selection procedure is crucial since it requires taking into account a number of variables, including the needs of the users, the size and features of the soils being studied, the availability of data, and the study's financial and schedule restrictions.

A colour camera is used to take various pictures of soil samples, which are then fed into the system as input to help with proper soil classification. The characteristics of every type of soil are then painstakingly gathered and kept in different databases. These databases are quite helpful when it comes to the last stage of classifying soil [41]. A sufficient amount of training samples are needed to achieve good image classification. In order to ensure that training models accurately reflect the variety of soil types seen in real-world circumstances, fieldwork data is carefully gathered. This meticulous approach to data collection is essential for training robust classification models capable of accurately categorizing soil samples based on their visual characteristics [42].

M. P. K., Anthiyur et al. emphasized the critical role of agribusiness in the Indian economy, highlighting its significance as the primary source of income for a large segment of the population [43]. Given this importance, farmers are inherently curious about yield prediction, as crop yield depends on a multitude of factors including soil quality, weather conditions, rainfall, fertilizers, and pesticides. Understanding how these factors influence agriculture requires the application of appropriate statistical techniques. By leveraging suitable statistical methodologies on historical crop yield data, valuable insights can be gleaned to aid farmers and government organizations in making informed decisions and formulating effective policies to enhance agricultural productivity [44]. The primary objective of the research conducted by M. P. K., Anthiyur et al. is to evaluate various data mining techniques to determine which ones offer the highest accuracy [45]. Data mining plays a pivotal role in transforming vast amounts of data into actionable insights and innovations, thereby making them accessible to farmers. The abundance of data available can be harnessed to uncover valuable insights that can assist farmers and decision-makers in making effective and timely decisions. In their paper,

M. P. K., Anthiyur et al. focus on one of the critical factors influencing crop production: soil. They apply different classification algorithms to soil datasets to predict soil fertility rates. Specifically, the paper explores the classification of soil fertility rates using techniques such as K-Means, Random Tree, and Apriori [46]. By employing these algorithms, the researchers aim to provide farmers and stakeholders with valuable insights into soil quality and its impact on crop productivity.

Sk Al Zaminur Rahman et al. underscored the fundamental importance of soil in agriculture, highlighting its diverse types and the unique features associated with each [47]. Understanding the characteristics of different soil types is essential for determining which crops thrive best in specific soil conditions. In this sense, machine learning techniques have become useful tools that provide forecasts and insights to support agricultural decision-making. Particularly in the area of agricultural data analysis, the field of machine learning has experienced tremendous expansion and advancement in recent years [48]. In order to tackle this rapidly expanding field of study, Rahman et al. proposed a unique model that predicts soil series according to the kind of land, allowing for the prescription of appropriate crops [50]. For soil categorization, their suggested model makes use of a number of machine learning techniques, such as weighted k-Nearest Neighbour (k-NN), Bagged Trees, and Support Vector Machines (SVM) based on Gaussian kernels. They proved the effectiveness of their SVM-based strategy by conducting thorough experiments, exhibiting its higher performance over other approaches. Rahman et al. sought to give farmers and other stakeholders a dependable tool for crop recommendation and soil categorization by utilising machine learning [51]. This would enable well-informed decision-making and maximise agricultural production.

Many algorithms and filters have been developed in the pursuit of precisely capturing and processing colour photographs of soil samples, as discussed in [52]. These algorithms are meticulously crafted to extract multifaceted properties embedded within the soil samples, encompassing attributes such as color, texture, and more. The comprehensive approach adopted in this research encompasses the consideration of various soil types, including but not limited to alluvial, clay, black, and red soils. Traditionally, several methods have been employed for soil classification, each with its own set of strengths and limitations. Among these methods, the Standard Penetration Test (SPT), as elucidated in stands out for its simplicity and widespread application. In the SPT, soil classification is facilitated through assessments based on factors like moisture content, visual inspection, and the ability to penetrate through dense layers [53], However, the SPT is not without its drawbacks; it is often associated with high costs, time-consuming procedures, and results that may lack reproducibility. Alternatively, the Cone Penetration Test (CPT) offers an alternative approach, albeit with its own challenges. While the CPT provides valuable insights, its analysis typically requires the expertise of domain specialists, particularly for interpreting segmented signals. Despite its effectiveness, this reliance on specialized knowledge can pose logistical hurdles in practical applications. Furthermore, in [55] strength and compressibility evaluations, tests such as the field vane shear test emerge as indispensable tools for assessing the undrained shear strength of both firm and soft clays [54]. However, it is imperative to acknowledge that each testing method carries its own set of limitations and considerations, underscoring the need for a holistic approach that integrates various techniques to ensure comprehensive soil analysis.

However, only clays with undrained strengths up to about 100 kPa are suitable for it. A single pressure metre test (PMT) can provide a wealth of important soil attributes. Empirical correction factors are not necessary to derive these features [56]. Nevertheless, clay stones and gravels are impervious to the device. It might be necessary to drill a borehole that is one or two metres above the intended test locations in order to conduct experiments in sand. The vane shear test (VST) can be used to rapidly and easily determine the shear strength of soft clay at deeper depths. The primary issue with VST is that specimens of fissured clay cannot be used for the test [57].

Segmentation and classification are the two stages of the automatic classifier based on machine learning called Constraint Clustering and Classification (CONCC) [58]. The Cone Penetration Test data is first divided into "J" segments from a single data series in order to address the inaccuracy in the measured data. These segmented data are then categorised into classes using fuzzy logic. We are able to identify the various soil types and the crops that work best with them thanks to the techniques mentioned above. It was determined that soil made of peat is more suited for farming. Conventional methods of classifying soil, however, are labor-intensive, costly, and time-consuming [59]. An automated system that can categorise dirt quickly and affordably is therefore required.

Scientists have recently provided great progress and provided better results in land classification and crop forecasting using various artificial intelligence techniques. Motivwani et al. in, attention is drawn to ease of accessing agricultural data through IoMT tools and the need for affective algorithms for determining soil types. To improve accuracy of feature selection and classification, the authors present a learning algorithm for soil type classification using vector machines. According to the evaluation made using SVM algorithm, the highest accuracy rate reached is only 83% [60].

In [61], Priyanka C. J. et al. concentrated on using a data mining model as an analytical method for classifying soil and land. By contrasting it with other state-of-the-art predictive modelling approaches and classic random forest modelling, the author assesses the performance of the K-Nearest Neighbour and Naive Bayes algorithms. However, because the procedure does not build upon the model and just knows the kind of soil, it is ineffective.

When applied to an agricultural soil profile, P. Bhargavi et al.'s [62] suggested data mining approaches may strengthen the validation of a reliable soil profile categorization. The Naive Bayesian classification approach was employed by the researcher to categorise the soils. The purpose of this paper is to investigate if terrain distributions for sizable datasets of experimental data may be found using data mining techniques. Several scholars have looked on automating the identification and classification of soil using machine learning methods. The authors of the study that was cited in suggested a technique for classifying soils into five groups using a support vector machine [63]. But when the study was trained on a tiny dataset, it ran into problems. For statistical classification techniques to yield high accuracy in standard machine learning methodologies, a significantly bigger dataset is usually needed.

Data mining techniques were employed by Raorane and Kulkarni in [64] to forecast agricultural production. extracting knowledge from data by using agricultural yield trends.

For soil classification, the researchers investigated a variety of data mining techniques, such as decision trees, support vector machines (SVMs), and artificial neural networks (ANNs). The K-Means algorithm was also used to categorise soil and plants. By adjusting various hyperparameters, accuracy and loss scores were calculated to assess how well these learning models performed. On the other hand, certain traditional learning strategies showed limits in terms of improving accuracy.

A more effective technique for soil identification utilising the K-Nearest Neighbours (KNN) algorithm was created by Riese, FM et al. in [65]. The findings obtained were adequate for the brown color-based soil type categorization. The study in question employs Convolutional Neural Network (CNN) methods for the categorization of soil. This technique was employed to investigate the classification of soil texture using hyperspectral data. However, because the classification model could only categorise soil with an accuracy of 70%, it was less useful in identifying the features of the soil and remained relatively weak.

Bharath et al. in [66] proposed a system for recommending the most suitable crops for farmers based on soil nutritional properties. The system's primary inputs come from laboratory test results that diagnose the nutritional characteristics of the soil. Data mining and algorithm are used to provide accurate results. The system will recommend crops that will benefit the farmer's soil type and bring profit. The Naive method was used for prediction and the accuracy of the model was 75%.

The Author in used random forest algorithm to estimate the crop yield in Maharashtra. Precipitation, temperature, cloud cover, vapor pressure etc. they collect data from [67] various government websites, including information about weather parameters like the example is more than 75 precent correct. used a machine learning method to predict crop yield based on weather conditions. The research, presented at the International

Conference on Computer Communication and Informatics (ICCCI), used the user-friendly Crop Advisor website, which has recently developed research software tool to predict the effects of weather on crops. In some areas of Madhya Pradesh, the C4.5 method is used to determine the weather conditions that have the greatest impact yield of a particular crop. Put your model into practice using the decision tree.

The statistical model employed by Lobell and Burke serves as a pivotal tool in elucidating the intricate relationship between climate change dynamics and the response of agricultural crops, thus occupying a central position in scientific research [68]. At its core, the focus of their investigation revolves around the imperative question of how agricultural crops respond to the evolving climate landscape. This inquiry delves into the profound implications of climate change on crop behavior, encapsulating a realm of paramount importance in the sphere of agricultural science. In essence, the statistical model serves as a conduit through which researchers can discern and analyze the multifaceted interactions between climatic variables and crop performance [69]. By harnessing this model, researchers endeavor to bridge the gap between theoretical insights and empirical observations, thereby facilitating a deeper understanding of the complex mechanisms governing crop responses to climatic shifts. Through rigorous analysis and comparison with real-world data, the efficacy of the statistical model is meticulously evaluated, shedding light on its capacity to accurately predict the impact of weather variations on crop yields. The findings derived from this empirical investigation underscore the invaluable utility of statistical modeling in prognosticating the ramifications of climate change on agricultural systems. Indeed, the empirical validation of the statistical model serves as a testament to its efficacy and reliability in forecasting how alterations in weather patterns will reverberate across agricultural landscapes. By elucidating these intricate linkages, researchers can glean invaluable insights that empower stakeholders to formulate informed policies and adaptive strategies aimed at bolstering agricultural resilience in the face of climate change-induced challenges [71].

Reference	**Year**	**Focus**	**Methods**	**Key Findings**
An image processing technique utilizing neural networks for classifying soils and lands.[72]	2020	Soil classification	Image processing techniques	Introduced an image analysis approach for categorizing soils and lands.
Classifying soil textures using 1D convolutional neural networks on hyperspectral data. [73]	2019	Soil texture classification	Soil texture classification via 1D CNN on hyperspectral imagery.	1D CNN model for classifying soil textures from hyperspectral imagery.
Data mining techniques for enhancing crop yield estimates and soil data analysis. [74]	2016	Applying data mining for crop yield predictions and soil data examination.	Data mining techniques	Data mining approach for better crop yield estimation and soil data analysis.

Colorimetric analysis with Naive Bayes algorithm for assessing soil fertility.[75]	2018	Soil fertility determination	Soil fertility determination using colorimetry and Naive Bayes classifier.	Applied colorimetry and Naive Bayes classification algorithm to determine soil fertility.
Machine learning soil classification and crop recommendations based on soil series. [76]	2018	Soil classification and crop recommendation	Machine learning methods	ML-based soil classification system with crop recommendations by soil series.
Soil Mapping with Machine Learning - A Review [77]	2019	Reviewing ML applications in soil mapping	Various ML algorithms	Provides a comprehensive review of different ML algorithms used in soil mapping.
Deep Learning for Soil Analysis and Monitoring: A Review [78]	2020	Reviewing DL applications in soil analysis and monitoring	Various DL algorithms	Provides a comprehensive review of DL applications beyond just prediction, including soil monitoring.
Evaluating sustainable agriculture using AI and machine learning for soil analysis. [79]	2023	Evaluating AI and ML for sustainable agriculture	Various ML algorithms and geostatistical techniques	Highlights the potential of ML for efficient, cost-effective, and sustainable soil analysis.
Deep Learning for Soil Quality Assessment Using Hyperspectral Images [80]	2023	Assessing soil quality with hyperspectral imagery	Convolutional Neural Networks (CNN)	Demonstrates the effectiveness of CNNs for soil quality assessment using hyperspectral data.

Soil Nutrient Analysis Using Machine Learning Techniques [81]	2021	Forecasting soil nitrogen, phosphorus, and potassium levels	Multiple Linear Regression (MLR)	Demonstrates MLR's effectiveness in determining soil nutrient composition.
Machine learning soil classification with crop suggestions per soil series. [82]	2022	Real-time soil analysis for crop management	Not specified (mentioning various techniques)	Emphasizes the role of ML in precision farming and crop yield prediction.
Deep learning techniques for estimating soil fertility from color analysis. [83]	2022	Comparing ML algorithms for soil texture classification	Soil classificat ion using SVM, Random Forest, and KNN algorithm s.	Compares the performance of various ML algorithms for soil texture classification.
Soil Analysis Using Machine Learning [84]	2023	Assessing soil fertility based on village-level data	Ensemble methods like bagging, neural nets, SVM, AdaBoost, and random forests.	Efficiently categorized soil fertility based on various characteristics.
Deep Learning for Soil Analysis and Monitoring: A Review [85]	2020	Reviewing DL applications in soil analysis and monitoring	Various DL algorithms	Provides a comprehensive review of DL applications beyond just prediction, including soil monitoring.

Summary

In this work, we investigate the use of deep learning and machine learning methods, namely convolutional neural networks (CNNs), for soil classification tasks utilising datasets of soil properties, digital photographs, and hyperspectral images. Numerous research offer models to forecast soil series according to the kind of land and suggest appropriate crops utilising techniques such as Gaussian kernel-based and bagged trees. The limits of conventional soil classification techniques, such as the vane shear test, cone penetration test, pressure metre test, and standard penetration test, are highlighted and examined. Data mining techniques such as random tree, and Apriori algorithms are employed to analyze soil datasets and predict soil fertility. We highlight the potentials of these advanced computational techniques to automate soil classification, enhance crop yield prediction, and support decision-making in agriculture by providing insights into crop suitability and soil management.

CHAPTER 3

we aim to analysis soils types using deep neural networks to help farmers with the proper recommendation to increase crop yield.

3.1 Proposed Method

In our proposed system, we aim to analysis soils types using deep neural networks to help farmers with the proper recommendation to increase crop yield. Additionally, we introduce a framework for soil arrangement, depicted in the square chart. The fundamental objective of this framework is to gather specific types of soil test images. For image-based soil classification, choosing the right sensor data is regarded as a crucial first step. A number of variables are taken into consideration during this selection process, including the needs of the user, the size and features of the soils being studied, the availability of soil data, and the financial and temporal restrictions related to the inspection procedure.

A colour camera is used to take a variety of photographs, which are then sent into the system as input to collect the required number of soil test images. Figure 1 shows the proposed model's system architecture, including all of the system's components and how they interact. This architecture serves as a blueprint for the development and implementation of the soil classification framework, guiding the integration of various modules and functionalities.

By delineating the system architecture, stakeholders gain a comprehensive understanding of the proposed model's design and operation. This visual representation facilitates collaboration among researchers, engineers, and stakeholders involved in the development and deployment of the soil arrangement framework. Additionally, it serves as a reference point for further enhancements and optimizations as the project progresses.

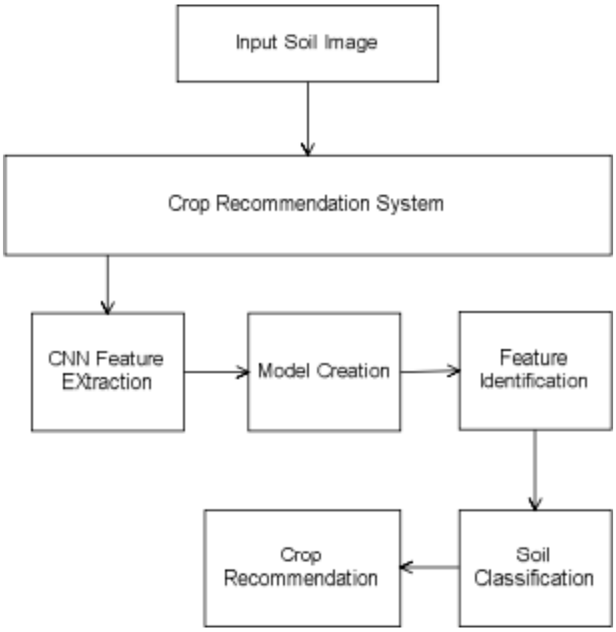


Figure 1 Proposed Architecture

Convolutional Neural Networks (CNNs) need a large enough number of training samples in order to classify images accurately. Typically, fieldwork is used to gather these training samples, and many factors are taken into account throughout the selection process. The intricacy of the data taken into consideration, the availability of ground reference data, and the spatial resolution of the photographs gathered are some of the factors taken into account. Following CNN-assisted successful picture categorization, Support Vector Machines (SVM) are used to propose crops. This extra function is especially helpful during the present epidemic time since it not only helps farmers make educated decisions regarding their crops, but it also enables them to sell their products online without physically visiting the market.

The training phase and the testing phase are the two separate stages of the methodological approach. This approach makes use of two main datasets: one for soil and the other for crops. Among the datasets for soil are

pictures that depict the many soil classifications, including red, clay, black, and alluvial soil. On the other hand, depending on the classification of soil types, the crop dataset is utilised to suggest appropriate crops. The graphic below illustrates the four fundamental parts that make up the CNN model's general architecture for analysing advertising images: input, capture, classification, and output. Three primary processing components are recognised in this architecture: test sample selection, CNN Learning and Visualisation, and configuration and startup.

An instance and a configuration are the two inputs that the configuration and initialization component takes. Basic features including the number of layers, their names, the number of filters on each convolutional layer, the classification technique, the size of the input pictures, and the convolutional layer kernel specifications are all described by the instance parameter. Conversely, the model's other parameters—such as learning rate, mini-batch size, weight drop, and momentum—are defined by the configuration parameter.

The CNN Learning and Visualization component of nLmF-CNN explores image properties in multiple layers, including input, convolution (conv), rectified linear units (relu), pooling (pool), fully connected (fc), and softmax layers. This component provides insights into the current learning status of the model, facilitating interpretation and analysis of its performance.

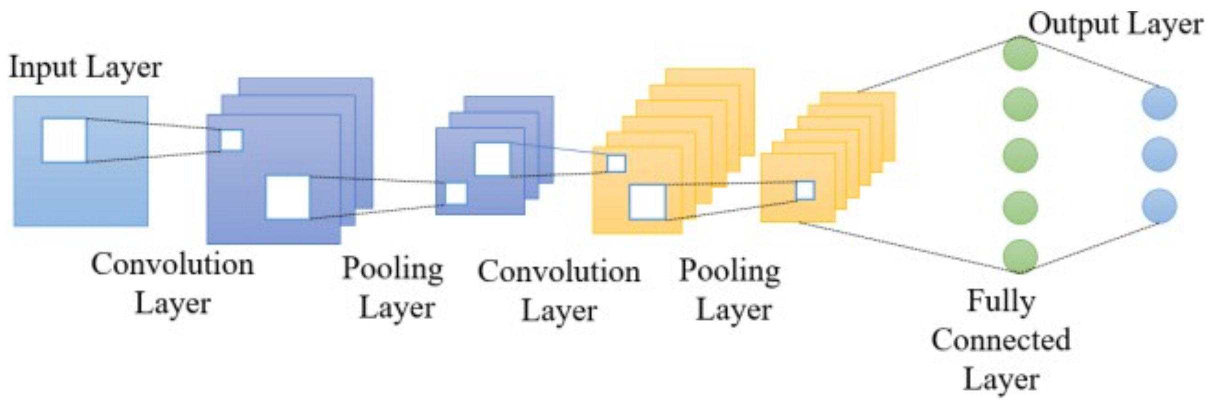


Figure 2: CNN Architecture [17]

Convolutions are essential for abstracting tiles from input data in convolutional neural networks (CNNs). Convolutional neural networks compute additional features by applying filters to an input feature map, which produces an output feature map, also known as a convolved feature. The size and depth of this output feature map may differ from those of the input. Two fundamental factors determine convolutions: the depth of the output feature map, which is related to the number of filters used, and the size of the extracted areas or tiles, which are usually 3x3 or 5x5 pixels.

One pixel at a time, filters—matrices the same size as tiles—move vertically and horizontally over the input feature map's grid throughout the convolution process. To determine the feature, each matched tile is retrieved. To add nonlinearity to the model, a Rectified Linear Unit (ReLU) modification is performed to the convolved element following each convolution process. For all positive values of x, the ReLU function, expressed as F(x)=max(0,x), returns x; for all nonpositive values of x, it returns 0. A popular activation function in many neural networks is ReLU.

Following ReLU, a pooling phase is applied to the CNN, whereby the convolved feature is sampled in order to minimise the feature map's dimensions while maintaining the crucial feature data. Max pooling is one of the most often utilised techniques for this purpose; it functions similarly to convolution. By iteratively going over the extracted tiles of a specific size and feature map, maximum pooling generates a new feature map for each tile based on its maximum value, deleting all other values. Two parameters are often needed for max pooling: the max pooling filter's size, which is typically set to 2x2 pixels.

A convolutional neural network eventually has one or more fully connected layers, in which all of the nodes in one layer are linked to all of the nodes in the layer above it. Based on the characteristics that the preceding convolutional layers retrieved, these fully connected layers classify data. Lastly, a softmax activation function is used in the last fully connected layer to provide a probability value between 0 and 1 for each categorization label that the model attempts to predict.

CHAPTER 4

The methodology detailed herein encapsulates a framework meticulously crafted to underpin our research endeavors in soil analysis.

4.1 Methodology

Soil analysis stands as a foothold in numerous domains, providing profound implications for the field of agriculture. Through the use of technological innovation, the advent of deep learning methodologies has welcomed in the new era of precision and efficiency in soil analysis. This paradigm shift has sparked a surge of interest and exploration into harnessing the formidable capabilities of deep learning algorithms to unravel the intricate details of soil composition, texture, and fertility.

The methodology detailed herein encapsulates a framework meticulously crafted to underpin our research endeavors in soil analysis. This comprehensive blueprint navigates the intricate terrain of soil analysis, traversing a multifaceted journey that spans from the inception of image acquisition to the culmination of model evaluation and validation. Each facet of this methodology is builded with meticulous attention to detail, aimed at furnishing researchers and practitioners alike with the requisite tools and insights to unravel the complexities enshrouding soil analysis. At its core, this methodology unfolds across three distinct yet interconnected stages, each constituting a pivotal milestone in the trajectory of soil analysis refinement:

4.1.1 Stage A: Image Acquisition and Pre-processing

This foundational stage sets the groundwork for subsequent analysis, entailing the acquisition of a diverse and representative dataset comprising soil images sourced from an array of platforms, and existing soil databases.

The acquired images undergo a meticulous pre-processing regimen encompassing a gamut of techniques aimed at standardization, normalization, noise reduction, and enhancement. These pre-processing steps are instrumental in priming the raw images for downstream analysis, ensuring optimal fidelity and relevance in the ensuing stages of the workflow.

4.1.2 Stage B: Segmentation and Feature Extraction Techniques

Building upon the pre-processed images, Stage B delves into the realm of segmentation and feature extraction, where sophisticated algorithms are deployed to delineate soil regions of interest from extraneous background elements. Segmentation techniques such as thresholding, edge detection, and Gabor filter is used to isolate the salient features indicative of soil composition and texture. Subsequent feature extraction endeavors draw upon a rich repertoire of methodologies ranging from traditional color histograms and texture descriptors to cutting-edge techniques such as Gabor filters and Law's Texture Energy Measures. These extracted features serve as the cornerstone for subsequent classification endeavors, encapsulating the intrinsic characteristics pivotal for discerning between different soil types.

4.1.3 Stage C: Model Training and Evaluation

The final stage of the methodology unfolds in Stage C, where the preparatory groundwork laid in the preceding stages coalesces into a cohesive framework for model training and evaluation. This stage encompasses the preparation of training data, the selection and fine-tuning of appropriate deep learning model architectures, and the iterative cycles of model training and validation. Leveraging state-of-the-art deep learning frameworks and methodologies, we embark on a journey of discovery, fine-tuning model parameters, optimizing hyperparameters, and evaluating model performance against stringent validation criteria. The iterative refinement process iterates until a model of requisite accuracy and generalization prowess is attained.

By delineating the research journey into these three distinctive stages, our methodology provides a roadmap for researchers and practitioners alike, guiding them through the complex landscape of soil analysis with clarity, precision, and purpose. Armed with this comprehensive framework, stakeholders are empowered to unlock new frontiers in soil analysis, driving innovation and advancement in fields ranging from precision agriculture to environmental monitoring and beyond.

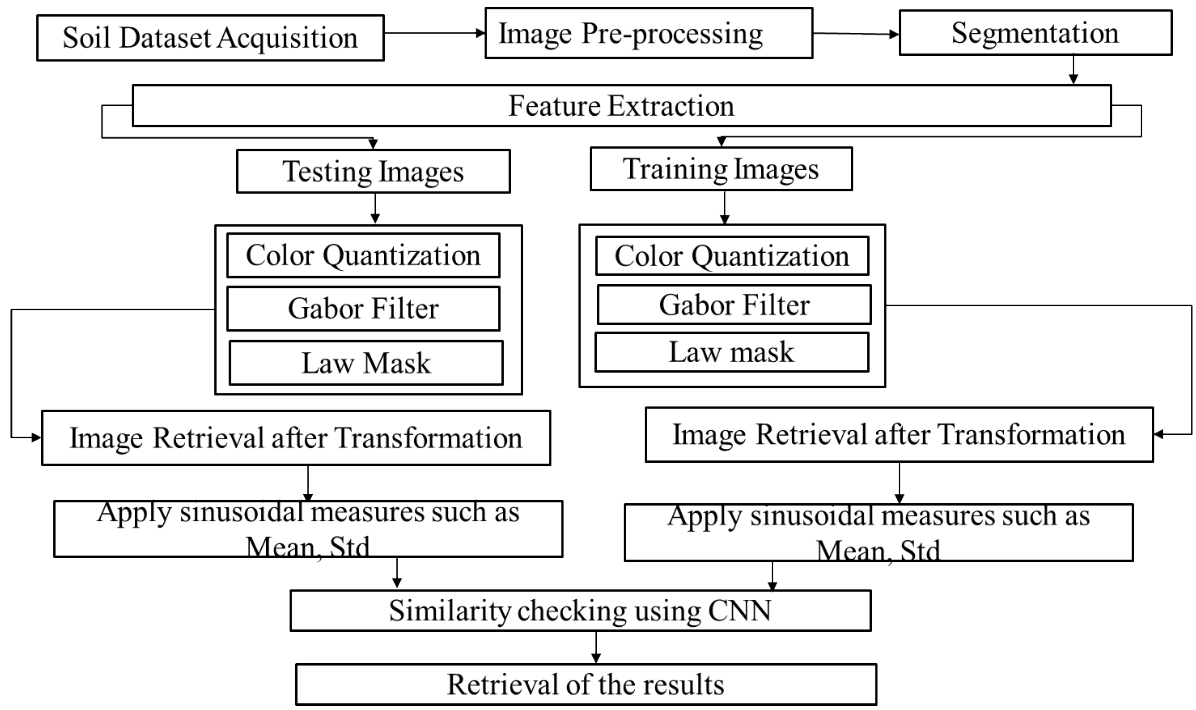


Figure 3: The comprehensive framework used for our research on soil analysis

1. **Soil Dataset Acquisition:** The process of soil classification commences with the acquisition of a diverse and extensive dataset

comprising soil images or samples. This dataset serves as the foundational bedrock upon which the subsequent stages of model development and evaluation are built. Sources for acquiring such datasets span a wide spectrum, encompassing field surveys conducted by soil scientists, the utilization of sophisticated remote sensing technologies such as satellite imagery or aerial drones, or tapping into existing soil databases curated by research institutions or governmental organizations. The breadth and depth of this dataset are pivotal in ensuring the robustness and generalization capabilities of the ensuing soil classification model.

2. **Image Pre-processing:** Image pre-processing constitutes a pivotal phase in the soil classification pipeline, where raw soil images undergo a series of meticulously orchestrated transformations to optimize them for subsequent analysis in computer vision tasks. This multifaceted process encompasses an array of techniques aimed at enhancing the quality, relevance, and discriminative power of the images. These techniques include but are not limited to resizing for standardization and computational efficiency, normalization to rectify illumination and contrast discrepancies, conversion to grayscale to simplify processing, noise reduction using advanced filtering methodologies, contrast enhancement to amplify subtle features, edge detection algorithms to delineate object boundaries, thresholding techniques for precise segmentation, and image rotation/flipping to augment the dataset through data augmentation strategies.

a. **Segmentation:** The segmentation of soil images plays a pivotal role in isolating the regions of interest pertaining to the soil from extraneous background elements. This is achieved through the deployment of sophisticated segmentation techniques such as thresholding, edge detection algorithms, or machine learning-based approaches. By delineating the soil regions from the background clutter, segmentation lays the groundwork for subsequent feature extraction and classification stages.
b. **Feature Extraction:** Effective feature extraction lies at the heart of accurate soil classification. Various techniques are harnessed to extract salient features from the segmented soil regions, ranging from traditional methods such as color histograms and texture descriptors to more advanced approaches like Gabor filters and Law's Texture Energy Measures. These extracted features encapsulate the intrinsic characteristics of the soil, encompassing its color distribution, textural properties, and morphological attributes, thereby furnishing the classification model with the requisite information for discerning between different soil types.

3. **Partitioning into Testing and Training Sets:** Following image pre-processing, the dataset is partitioned into two distinct subsets: the testing images and the training images. The testing images are reserved for assessing the performance and generalization capabilities of the trained model, serving as an external litmus test for its efficacy. Conversely, the training images are utilized to imbue the model with the ability to discern patterns and extract discriminative

features characteristic of various soil types. This partitioning ensures the integrity of the evaluation process and guards against overfitting, thereby fostering the development of a robust and reliable soil classification model.

4. **Color Quantization:** While color information constitutes a pivotal aspect of soil classification, processing high-resolution color images can pose computational challenges. To mitigate this, color quantization techniques are employed to reduce the number of colors present in the images while preserving the essential visual cues pertinent for soil classification. These techniques encompass color palette reduction, where a limited set of representative colors is retained, and color space transformations, which enable the encoding of color information in a more compact form. By judiciously reducing the color complexity of the images, color quantization facilitates more efficient processing without sacrificing classification accuracy.

5. **Gabor Filter:** Gabor filters emerge as indispensable tools in the texture analysis and feature extraction realm, particularly in the context of soil classification. These filters are adept at capturing spatial frequencies and orientations inherent in soil textures, thereby encapsulating the intricate patterns and nuances characteristic of different soil types. By convolving the images with Gabor filter kernels at various scales and orientations, the model can discern subtle textural differences that serve as discriminative cues for accurate classification

6. **Law Mask:** Law's Texture Energy Measures, colloquially referred to as Law's Mask, represent another potent technique employed in texture analysis and feature extraction. By computing diverse texture energy metrics based on local texture patterns, Law's Mask facilitates the extraction of complementary features that augment the feature space and enrich the model's discriminative capabilities. These measures encapsulate the spatial relationships and structural properties inherent in soil textures, furnishing the classification model with a more comprehensive understanding of the underlying soil characteristics.

7. **Image Retrieval after Transformation:** Following the application of pre-processing techniques such as Color Quantization, Gabor Filter, and Law Mask, the transformed images are retrieved, encapsulating the extracted features and preprocessed data. These images serve as the foundational input for subsequent stages of analysis and classification, laying the groundwork for the model to discern intricate patterns and nuances indicative of different soil types.

8. **Apply Sinusoidal Measures:** Sinusoidal measures or transformations, encompassing techniques such as Fourier transforms or wavelet transforms, represent yet another dimension in the arsenal of feature extraction methodologies deployed in soil classification. By leveraging the frequency-domain or scale-space representations obtained through these transformations, the model can extract additional discriminatory features that capture the inherent variability and complexity present in soil textures. These measures augment the

feature space, providing the model with a more nuanced and comprehensive understanding of the underlying soil characteristics, thereby enhancing its classification accuracy and robustness.

9. **Similarity Checking using CNN:** Convolutional Neural Networks (CNNs) are the mainstay of contemporary computer vision applications because of their exceptional capacity to acquire hierarchical features straight from unprocessed pixel input. In the context of soil classification, pre- processed and transformed images from both the testing and training sets are fed into meticulously designed CNN architectures tailored specifically for this task. These CNNs adeptly discern patterns and commonalities between the testing and training images, leveraging their hierarchical feature representations to facilitate precise classification of soil types. By learning to discern subtle cues and discriminative features indicative of different soil classes, CNNs enable the model to achieve unparalleled accuracy and robustness in soil classification tasks.

10. **Retrieval of Results**: Upon completion of training and evaluation, the final results of the soil classification endeavor are retrieved and analyzed. These results encompass predicted soil types or classes for the testing images, alongside associated confidence scores or probabilities. The accuracy of the classification is gauged by juxtaposing the predicted labels against the ground truth or expert-labeled soil types. This meticulous evaluation process ensures the integrity and reliability of the classification results, enabling stakeholders to make informed decisions based on the model's outputs.

Stage A

The initial stage is using deep learning involves acquiring high-quality soil sample images. This is achieved through gathering different soil image datasets from multiple sources which were taken using specialized equipment like microscopes, scanners, or digital cameras with macro lenses. Consistent lighting and background settings during image capture are crucial to minimize variability. Image pre-processing is the next step. Noise reduction techniques like median filtering or wavelet denoising enhance image quality by removing noise and unwanted artifacts. Color space transformations from RGB to HSV may provide better discriminatory information. Resizing images to a consistent dimension suitable for the deep learning model's input size is necessary. Normalizing pixel values to a standard range (0-1 or -1 to 1) facilitates model training by ensuring consistent data scales. Dataset augmentation through transformations like rotation, flipping, or scaling increases diversity and improves model generalization. Segmentation is crucial for separating soil particles from the background. Thresholding techniques like Otsu's method or adaptive thresholding achieve this separation. Edge detection algorithms like Canny or Sobel identify soil particle boundaries. Region-based methods like watershed or region growing delineate individual soil particles. Incorporating domain knowledge, such as soil particle size ranges or color characteristics, refines the segmentation process. Evaluating segmentation results using metrics like the Dice coefficient or Intersection over Union (IoU) ensures accurate particle delineation. Feature extraction captures soil particle characteristics. Shape features like area, perimeter, circularity, and aspect ratio describe particle morphology. Texture features like Gray-Level Co-occurrence Matrix (GLCM) or Local Binary Patterns (LBP) characterize particle textures. Color features like mean, standard deviation, or histograms of color channels (RGB, HSV) capture color information. Techniques like Fourier descriptors or moment invariants extract rotation and scale-invariant features. Deep learning-based feature extraction through transfer learning or fine-tuning pre-trained models on the soil dataset yields powerful discriminative features. Training data preparation is the final step in Stage A. Supervised learning requires labelling segmented soil particles with relevant class labels (soil type, texture, composition). To guarantee appropriate model assessment, the labelled dataset is divided into subgroups for training, validation, and testing. This ensures that each subset contains representative samples. Class imbalance problems that might negatively impact performance are mitigated by balancing class distributions by oversampling or data augmentation. Normalisation or feature scaling guarantees uniform ranges for various feature kinds. For improved data comprehension and interpretation, dimensionality reduction techniques like Principal Component Analysis (PCA) or t-SNE visualise and decrease feature dimensionality.

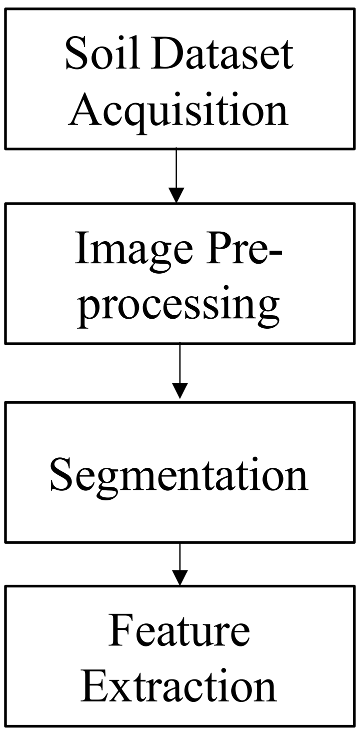


Figure 4: Flow graph Image Acquisition and Pre-processing.

Stage B

Stage B focuses on model architecture selection and training. Choosing the appropriate deep learning architecture depends on problem complexity and available computational resources. Convolutional Neural Networks (CNNs) are commonly used for image classification tasks and effective for soil analysis. Architectures like LeNet, AlexNet, VGGNet, ResNet, or DenseNet can be adapted. Alternatively, pre-trained models can be fine-tuned through transfer learning. Experimenting with hyperparameters like learning rate, batch size, and regularization optimizes performance.

Once the architecture is selected, implement the deep learning model using frameworks like TensorFlow, PyTorch, or Keras. Define appropriate loss functions (cross-entropy for classification) and optimization algorithms (Adam, SGD). Train the model on the training dataset while monitoring validation set performance to prevent overfitting. Techniques like early stopping, learning rate scheduling, or checkpointing enhance convergence and generalization.

After training, evaluate the model's performance on the test set using metrics like accuracy, precision, recall, F1-score, and confusion matrix. Visualize and interpret predictions, activations, and feature importances for deeper insights into decision-making. Model refinement and ensemble methods improve performance. Analyze misclassified samples to identify reasons for errors like ambiguous features, class imbalance, or data quality issues. Refine the architecture, hyperparameters, or feature engineering techniques accordingly. Ensemble methods like bagging, boosting, or stacking combine multiple models, leveraging their strengths. Data augmentation, transfer learning, or multi-task learning enhance robustness and generalization.

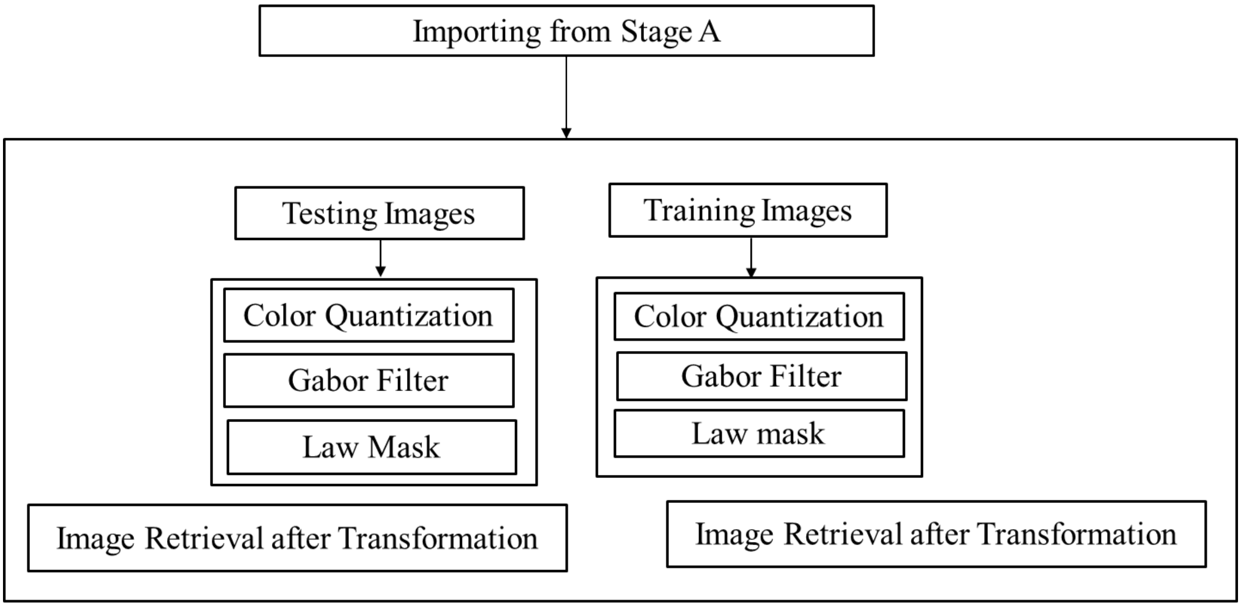


Figure 5: Stage B of the methodology

Stage C

Stage C involves deploying and monitoring the trained deep learning model. Integrate the model into production environments or decision support systems for practical applications. Develop user interfaces or APIs for seamless interaction with the model and soil analysis pipeline, facilitating ease of use. Implement mechanisms for continuous performance monitoring and evaluation of the deployed model to ensure reliability and effectiveness over time. Establish protocols for periodic model retraining or updating as new data becomes available or soil conditions change, ensuring the model remains relevant and accurate. Collaborate with domain experts, soil scientists, and stakeholders to validate and interpret predictions in real-world scenarios, enhancing practical applicability.

The final stage involves reporting and documentation. Thoroughly document the entire process, including data collection, pre-processing, model development, and evaluation, for reproducibility and transparency. Analyze and interpret results, drawing insights and conclusions relevant to soil analysis and management practices, advancing the field.

Present findings through peer-reviewed publications, technical reports, or relevant conferences for knowledge dissemination and peer review. Contribute to advancing soil analysis techniques and promoting deep learning adoption, impacting sustainable soil management practices.

Identify potential limitations and future research directions to further enhance deep learning capabilities for soil analysis. This iterative process of research, evaluation, and refinement drives scientific progress, contributing to more accurate and robust soil analysis techniques.

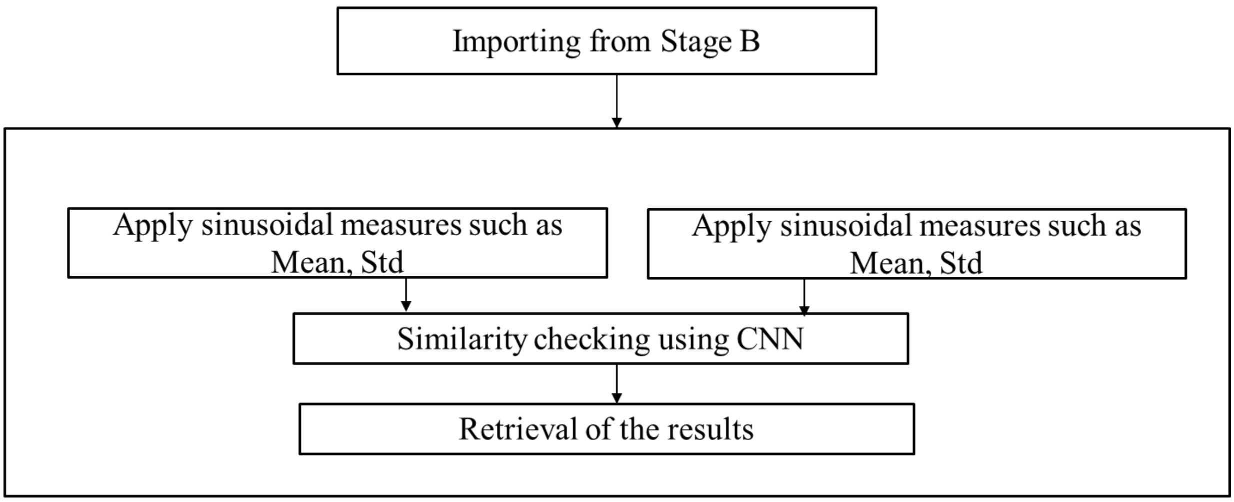


Figure 6: Stage C of the methodology

CHAPTER 5

This is a rather simple way of classifying soil because there are firm cutoffs between each classification

5.1 Implementation

Classification of soils can be an important capability in many fields including agricultural. While many classification schemes exist, a popular and intuitive way of classifying soils is based on shape, size, texture, and color which can be extracted from them using various image processing techniques such as Haralick texture features, Gabor filters, or Zernike moments., etc. This is a rather simple way of classifying soil because there are firm cutoffs between each classification

Data Augmentation

Data augmentation is a widely adopted technique in image classification, aimed at expanding the available image dataset by generating new data from existing samples. It encompasses a variety of operations applied to the original images, such as translation (shifting the image along the x and y axes), horizontal or vertical flipping (mirroring the image), zooming (enlarging or reducing the image size), rotation, cropping, and changes in brightness, contrast, or saturation levels. By applying these augmentations to the existing images, a diverse set of new images is generated, effectively increasing the volume and diversity of the dataset available for training. This augmentation process plays a crucial role in improving the robustness and generalization capability of the trained model, as it exposes the model to a broader range of variations and scenarios that it may encounter in real-world applications. Additionally, data augmentation helps mitigate overfitting by introducing variability into the training data, thereby enhancing the model's ability to generalize well to unseen data.

Convolutional Neural Network (CNN)

A convolutional neural network (CNN) stands out as a widely acclaimed model architecture for image classification tasks due to its ability to effectively capture and discern intricate patterns and features within images. At its core, a CNN operates by systematically scanning through the pixel values of an input image, leveraging a set of learnable filter kernels to extract relevant features. These filter kernels, also known as convolutional kernels or filters, serve as windows that slide over the image, computing the dot product between their weights and the pixel values within their receptive fields. Through this process of convolution, the CNN is capable of discerning salient features and characteristics of the image, such as edges, textures, gradients, shapes, and more. These learned features are then progressively abstracted and combined across successive layers of the network, ultimately culminating in high-level representations that facilitate accurate classification decisions. Moreover, CNNs are adept at capturing both local and global spatial dependencies within the input image, enabling them to effectively differentiate between objects, shapes, and textures across varying scales and orientations. With their hierarchical architecture and ability to automatically learn discriminative features from raw pixel data, CNNs have demonstrated remarkable performance across a wide range of image classification tasks, including but not limited to object recognition, scene understanding, medical image analysis, and facial recognition.

Setup

This setup integrates essential libraries and modules commonly utilized in machine learning and deep learning projects. These include modules for file and directory manipulation (os, shutil), image processing (cv2, PIL), numerical operations (numpy), and visualization (matplotlib.pyplot).

TensorFlow and its Keras API are employed for constructing and training neural network models, with layers like Dense, Conv2D, and MaxPooling2D forming the core architecture components. The Sequential class facilitates the sequential composition of layers, enabling the creation of complex neural network architectures. The ImageDataGenerator class allows for on-the-fly data augmentation and preprocessing during model training. Additionally, the load_model function enables the loading of pre-trained models for transfer learning. Finally, the integration of the drive module from Google Colab enables seamless access to files and datasets stored in Google Drive, streamlining data management within the Colab environment. This setup provides a robust foundation for conducting advanced machine learning experiments, enabling practitioners to explore diverse challenges across various domains.

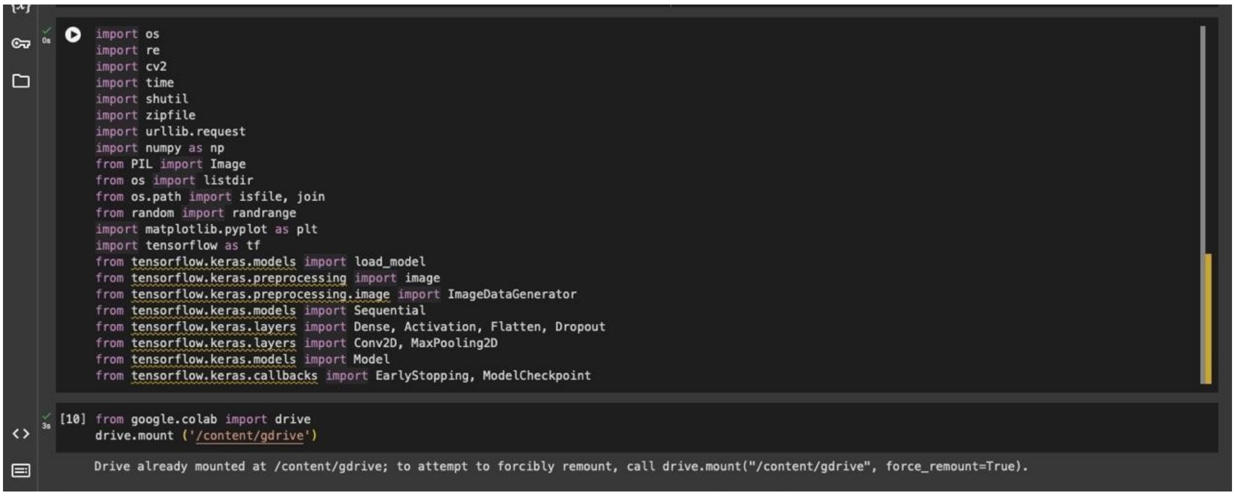
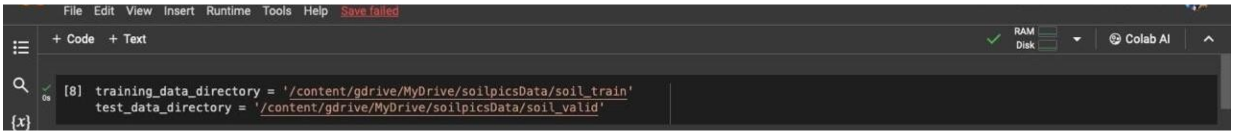


Figure 7: The project setup.

Data

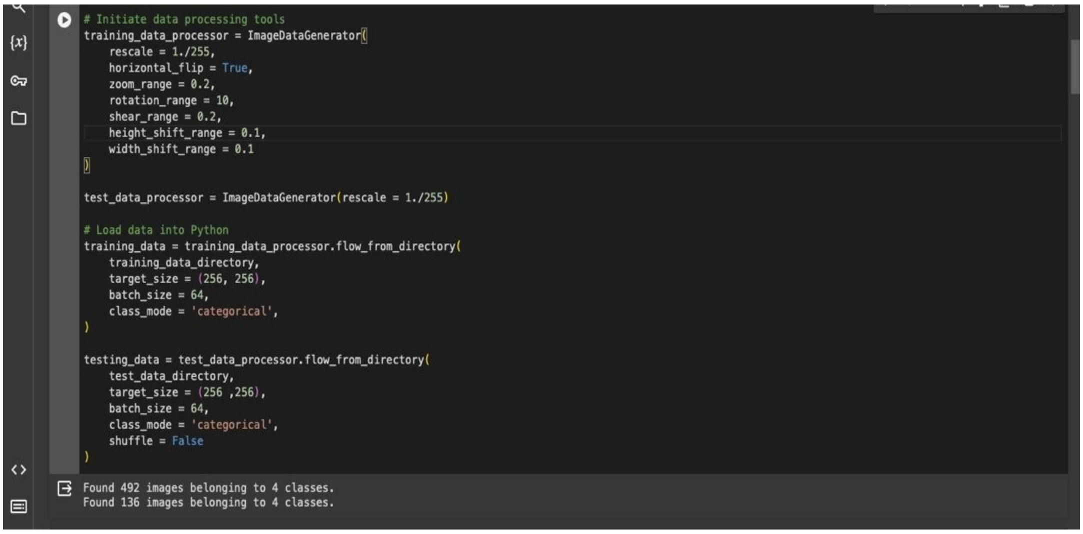
The data is separated into directories test and train. The test folder and train folder contain subdirectories corresponding to the possible soils. The labeled tree structure of the folders assists with marking the photos for training and testing.



```
File Edit View Insert Runtime Tools Help Save failed
+ Code + Text
[8] training_data_directory = '/content/gdrive/MyDrive/soilpicsData/soil_train'
    test_data_directory = '/content/gdrive/MyDrive/soilpicsData/soil_valid'
```

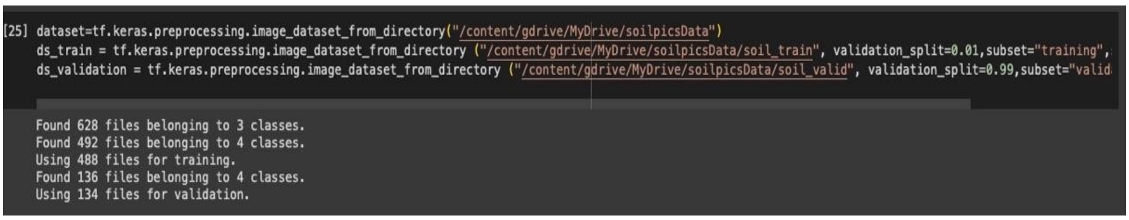
Figure 8: The data directory is provided.

Importing the data into a Python session is the first step. The photos must then be converted into a format that: 1) enables the model to read the data; and 2) gives the model additional training data to work with. To scale each image and produce several versions of the same image, for instance, the training_data_processor variable scales the data before feeding it into the model. To ensure that the model learns from the soil snapshot itself and is unaffected by orientation or size, it flips photographs horizontally, rotates them, pans them, and applies additional modifications.



```
# Initiate data processing tools
training_data_processor = ImageDataGenerator(
    rescale = 1./255,
    horizontal_flip = True,
    zoom_range = 0.2,
    rotation_range = 10,
    shear_range = 0.2,
    height_shift_range = 0.1,
    width_shift_range = 0.1
)

test_data_processor = ImageDataGenerator(rescale = 1./255)

# Load data into Python
training_data = training_data_processor.flow_from_directory(
    training_data_directory,
    target_size = (256, 256),
    batch_size = 64,
    class_mode = 'categorical',
)

testing_data = test_data_processor.flow_from_directory(
    test_data_directory,
    target_size = (256 ,256),
    batch_size = 64,
    class_mode = 'categorical',
    shuffle = False
)

Found 492 images belonging to 4 classes.
Found 136 images belonging to 4 classes.
```

Figure 9: The data processing



```
[25] dataset=tf.keras.preprocessing.image_dataset_from_directory("/content/gdrive/MyDrive/soilpicsData")
     ds_train = tf.keras.preprocessing.image_dataset_from_directory ("/content/gdrive/MyDrive/soilpicsData/soil_train", validation_split=0.01,subset="training",
     ds_validation = tf.keras.preprocessing.image_dataset_from_directory ("/content/gdrive/MyDrive/soilpicsData/soil_valid", validation_split=0.99,subset="valid

Found 628 files belonging to 3 classes.
Found 492 files belonging to 4 classes.
Using 488 files for training.
Found 136 files belonging to 4 classes.
Using 134 files for validation.
```

Figure 10: The image pre-processing.

Model Building

The design and performance of a convolutional neural network (CNN) model are largely determined by the parameter choices used throughout the model building process. A variety of critical characteristics, such as the quantity of convolutional layers, completely connected dense layers, kernel filter dimensions, number of nodes in each layer, and overall number, define CNN architecture in the context of deep learning. training periods.Each of these parameters has a substantial impact on a number of different aspects of the model's behaviour, including its ability to detect intricate patterns and features in the input data, how well it generalises to new examples, how efficiently it uses computation during training and inference, and how easily it overfits or underfits. Particularly, the layer_size parameter stands out as a critical factor since it directly affects the model's total size, accuracy, and training speed. Practitioners can modify the CNN architecture to meet the unique needs and limitations of the target application or dataset by carefully adjusting these parameters.

```
# choose model parameters
num_conv_layers = 2
num_dense_layers = 1
layer_size = 32
num_training_epochs = 20
MODEL_NAME = 'soil'
```

Figure 11: The model parameters

```python
# Initiate model variable
model = Sequential()

# begin adding properties to model variable
# e.g. add a convolutional layer
model.add(Conv2D(layer_size, (3, 3), input_shape=(256,256, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

# add additional convolutional layers based on num_conv_layers
for _ in range(num_conv_layers-1):
    model.add(Conv2D(layer_size, (3, 3)))
    model.add(Activation('relu'))
    model.add(MaxPooling2D(pool_size=(2, 2)))

# reduce dimensionality
model.add(Flatten())

# add fully connected "dense" layers if specified
for _ in range(num_dense_layers):
    model.add(Dense(layer_size))
    model.add(Activation('relu'))

# add output layer
model.add(Dense(4))
model.add(Activation('softmax'))

# compile the sequential model with all added properties
model.compile(loss='categorical_crossentropy',
                optimizer='adam',
                metrics=['accuracy'],
                )

# use the data already loaded previously to train/tune the model
model.fit(training_data,
            epochs=num_training_epochs,
            validation_data = testing_data)

# save the trained model
model.save(f'{MODEL_NAME}.h5')
```

Figure 12: Building the model.

Model Testing

Following the completion of model training, the trained model has been successfully saved to the local computer, residing within the same directory as this notebook. The printed output above provides crucial insights into the model's performance, particularly through the accuracy metrics computed on both the training and validation (or testing) datasets. It is imperative to focus on the val_accuracy value displayed in the last line of the output, as it signifies the accuracy achieved on images that were not part of the training process. This metric serves as a robust indicator of the model's generalization capability, offering valuable insights into its performance on unseen data.

To further assess the model's efficacy and practical utility, the make_prediction function has been provided, offering a streamlined mechanism for obtaining model predictions on soil photos. By supplying the file path to a soil photo as input to this function, users can obtain the corresponding classification predicted by the trained model. For seamless experimentation and evaluation, users are encouraged to leverage the test_image_filepath variable in the subsequent cell, populated with the file path to the desired test image. Through this interactive process of testing the model on diverse soil photos, users can gain deeper insights into its performance across various scenarios and datasets, facilitating a comprehensive assessment of its efficacy and robustness.

```
def make_prediction(image_fp):
    im = cv2.imread(image_fp) # load image
    plt.imshow(im[:,:,[2,1,0]])
    img = image.load_img(image_fp, target_size = (256,256))
    img = image.img_to_array(img)

    image_array = img / 255. # scale the image
    img_batch = np.expand_dims(image_array, axis = 0)

    class_ = ["clayeysoil", "sandysoil", "loamsoil"] # possible output values
    predicted_value = class_[model.predict(img_batch).argmax()]
    true_value = re.search(r'(clayeysoil)|(sandysoil)|(loamsoil)', image_fp)[0]

    out = f"""Predicted Soil Type: {predicted_value}
    True Soil Type: {true_value}
    Correct?: {predicted_value == true_value}"""

    return out
```

Figure 13: Model testing.

Soil Classification Percentages

Soils exhibit inherent heterogeneity, rarely consisting solely of a single soil type. Instead, they often comprise a mixture or blend of different soil constituents, each contributing to the overall soil composition. To accurately capture this complexity, it is imperative to represent soils using a proportional approach, wherein the relative abundance of each soil type is quantified as a percentage. For instance, consider the test photo labeled

"Silt" showcased in the cell below. Despite its primary designation, a closer examination reveals the presence of multiple soil components, including gravel, sand, and silt, within a single image frame. To refine the classification process and provide a more nuanced characterization of such complex soil compositions, a novel approach involves subdividing the photo into numerous smaller segments or squares. Subsequently, a classifier is trained on these smaller segments, enabling the classification of each individual square based on its predominant soil type. By systematically iterating through the photo and classifying each small square, an aggregate proportion of squares corresponding to gravel, sand, and silt can be computed. This proportion is then translated into a percentage representation, effectively capturing the relative abundance of each soil type within the overall soil composition. Through this iterative and granular analysis, practitioners can obtain a more comprehensive understanding of soil heterogeneity, paving the way for more accurate and nuanced soil classification methodologies.

```python
def classify_images(image_fp, model):
    classes = ['clayeysoil', 'sandysoil', 'loamsoil']
    gravel_count = 0
    sand_count = 0
    silt_count = 0

    img = cv2.imread(image_fp)
    img = cv2.resize(img,(1024,1024))
    im_dim = 256

    for r in range(0, img.shape[0], im_dim):
        for c in range(0, img.shape[1], im_dim):
            cropped_img = img[r:r + im_dim, c:c + im_dim, :]
            h, w, c = cropped_img.shape
            if h == im_dim and w == im_dim:
                classification = model_classify(cropped_img, model)
                if classification == classes[0]:
                    gravel_count += 1
                elif classification == classes[1]:
                    sand_count += 1
                elif classification == classes[2]:
                    silt_count += 1
            else:
                continue
    total_count = gravel_count + sand_count + silt_count
    proportion_array = [gravel_count / total_count, sand_count / total_count, silt_count / total_count]
    return proportion_array
```

Figure 14: Soil Classification.

Load Model

Ensuring consistency in input dimensions is paramount in the realm of neural networks, particularly in scenarios where variations in input pixel dimensions may necessitate model retraining or image rescaling to align with the original training dimensions. In the context of this demonstration, the decision-making process surrounding the need for model retraining is guided by the underlying methodology employed. Here, the photos undergo upscaling to a standardized resolution of 1024x1024 pixels before being subdivided into smaller 256x256 pixel blocks, resulting in a total of 16 blocks within each image. By adopting this approach, the need for retraining the model is obviated, as the uniformity in input dimensions ensures compatibility with the original training data. This strategic decision to upscale the images prior to block subsampling not only streamlines the computational workflow but also preserves the structural integrity and spatial coherence of the original images. Consequently, the model can seamlessly process and classify the subdivided blocks without necessitating adjustments to accommodate varying input dimensions. Through this meticulous handling of input data, the demonstration maintains fidelity to the original training framework, facilitating consistent and reliable model performance across diverse datasets and scenarios.

```
[34] model_fp = os.getcwd()+'/'+'soil.h5'
     print(model_fp)
     model = load_model(model_fp)
```

Figure 15: Loading the model.

Classify Image

Utilizing the loaded model to classify a test image involves a systematic process orchestrated by the `classify_images` function. This function operates on the premise of subdividing the input image into manageable

256x256 pixel squares, enabling granular analysis and classification of each individual square. As the function traverses through the image, it leverages the loaded model to classify the soil type present within each square. Through iterative classification, the function accumulates the classification results and updates a counter for each soil type, ultimately yielding a fractional prediction of soil composition. This fractional prediction encapsulates the proportional representation of each soil type within the overall image, offering valuable insights into the spatial distribution and abundance of different soil constituents. By furnishing this nuanced characterization of soil composition, the function empowers users with a comprehensive understanding of the soil landscape depicted in the test image. Through meticulous analysis and aggregation of classification results, the function provides a robust foundation for informed decision-making and further exploration of soil-related phenomena.

```
def model_classify(cropped_img, model):
    classes = ['clayeysoil', 'sandysoil', 'loamsoil']
    image_array = cropped_img / 255.
    img_batch = np.expand_dims(image_array, axis=0)
    prediction_array = model.predict(img_batch)[0]
    first_idx = np.argmax(prediction_array)
    first_class = classes[first_idx]
    return first_class

def classify_percentage(image_fp):
    start = time.time()
    out = classify_images(image_fp=image_fp, model=model)
    finish = str(round(time.time() - start, 5))

    im = cv2.imread(image_fp) # load image
    plt.imshow(im[:,:,[2, 1, 0]])

    print(f'''---
Percent clayeysoil: {round(out[0] * 100, 2)}%)
Percent Sandysoil: {round(out[1] * 100, 2)}%)
Percent loamsoil: {round(out[2] * 100, 2)}%)
Time to Classify: {finish} seconds
---''')
```

Figure 16: Classifying the images.

5.2 Algorithm (Pseudo Code)

Input: Training data directory, test data directory

Output: Trained model, model evaluation metrics, soil type prediction for a given image

1. Load training and test data directories
2. Load necessary libraries and modules
3. Mount Google Drive
4. Create image datasets for training and validation
5. Initialize data preprocessing tools (image data generators)
6. Load training and testing data into Python
7. Define model architecture parameters (number of convolutional layers, dense layers, layer size, number of epochs)
8. Initialize sequential model
9. Add convolutional layers with activation and max pooling
10. Add flatten layer to reduce dimensionality
11. Add dense layers with activation and dropout
12. Add output layer with softmax activation
13. Compile the model with appropriate loss function, optimizer, and metrics
14. Train the model using the prepared training and validation data
15. Save the trained model
16. Plot training and validation loss over epochs
17. Define a function to make predictions on a single image:
    a. Load the image
    b. Preprocess the image
    c. Make a prediction using the trained model
    d. Extract the true label from the image file path
    e. Output the predicted label, true label, and whether the prediction is correct

18. Check if the test image file exists
19. Load the test image
20. Call the prediction function with the test image
21. Define a function to split images into smaller patches:
    a. Iterate over image directories
    b. Read each image
    c. Split the image into smaller patches of specified size
    d. Save the patches as separate image files
22. Create a new directory for divided images (if it doesn't exist)
23. Load the trained model
24. Define a function to classify images:
    a. Load the image
    b. Resize the image
    c. Iterate over smaller patches of the image
    d. Classify each patch using the trained model
    e. Count the occurrences of each class
    f. Calculate the proportion of each class
    g. Return the proportions
25. Define a function to display the classification percentages:
    a. Call the classify_images function with the given image
    b. Load and display the image
    c. Print the percentage of each class and the time taken for classification
26. Call the classify_percentage function with the test image

5.3 Result and Discussions.

As seen in Fig. 17, the proposed model employed 628 samples of various soil pictures for the classification process, of which 488 samples were used for training and 134 samples for validation.

```
Found 628 files belonging to 3 classes.
Found 492 files belonging to 4 classes.
Using 488 files for training.
Found 136 files belonging to 4 classes.
Using 134 files for validation.
```

Figure 17: Data processed.

In the result shown in Fig 18 we trained our model and stored the training history to find how much percent accuracy do we have.

```
plt.show()
8/8 [==============================] - 175s 21s/step - loss: 0.8098 - accuracy: 0.6382 - val_loss: 1.5174 - val_accuracy: 0.4779
Epoch 2/20
8/8 [==============================] - 164s 20s/step - loss: 0.7866 - accuracy: 0.6504 - val_loss: 1.3399 - val_accuracy: 0.4632
Epoch 3/20
8/8 [==============================] - 174s 22s/step - loss: 0.8088 - accuracy: 0.6463 - val_loss: 1.5075 - val_accuracy: 0.3971
Epoch 4/20
8/8 [==============================] - 185s 23s/step - loss: 0.8094 - accuracy: 0.6626 - val_loss: 1.1958 - val_accuracy: 0.5074
Epoch 5/20
8/8 [==============================] - 160s 20s/step - loss: 0.7283 - accuracy: 0.6850 - val_loss: 1.4547 - val_accuracy: 0.4926
Epoch 6/20
8/8 [==============================] - 169s 21s/step - loss: 0.7791 - accuracy: 0.6260 - val_loss: 1.1615 - val_accuracy: 0.4706
Epoch 7/20
8/8 [==============================] - 173s 22s/step - loss: 0.8052 - accuracy: 0.6199 - val_loss: 1.5694 - val_accuracy: 0.4485
Epoch 8/20
8/8 [==============================] - 159s 20s/step - loss: 0.7914 - accuracy: 0.6504 - val_loss: 1.4354 - val_accuracy: 0.4265
Epoch 9/20
8/8 [==============================] - 159s 20s/step - loss: 0.7330 - accuracy: 0.6504 - val_loss: 1.5105 - val_accuracy: 0.5221
Epoch 10/20
8/8 [==============================] - 175s 22s/step - loss: 0.6997 - accuracy: 0.6890 - val_loss: 1.3812 - val_accuracy: 0.4926
Epoch 11/20
8/8 [==============================] - 167s 21s/step - loss: 0.6985 - accuracy: 0.7012 - val_loss: 1.3907 - val_accuracy: 0.4485
Epoch 12/20
8/8 [==============================] - 168s 21s/step - loss: 0.7519 - accuracy: 0.6748 - val_loss: 1.4053 - val_accuracy: 0.5441
```

Figure 18: Model summary.

In the result shown in Fig 19, we can visualize the training and validation loss over the epoch, which can help us understand the model's performance and potential overfitting or underfitting issues.

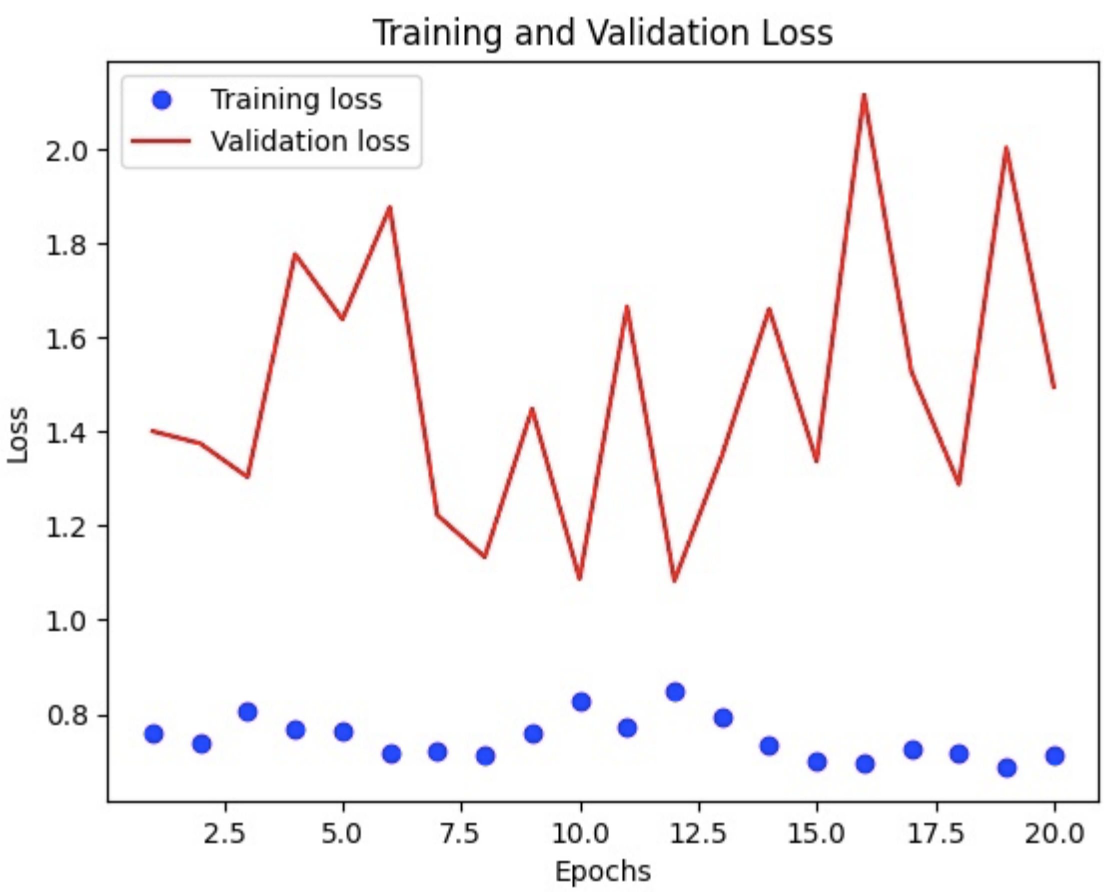


Figure 19: Graph showing Validation loss vs Epochs.

In the result shown in Fig 20, we tested the model and the model takes the dataset as an input, preprocess it and output the classification that the model predicted and the true soil type.

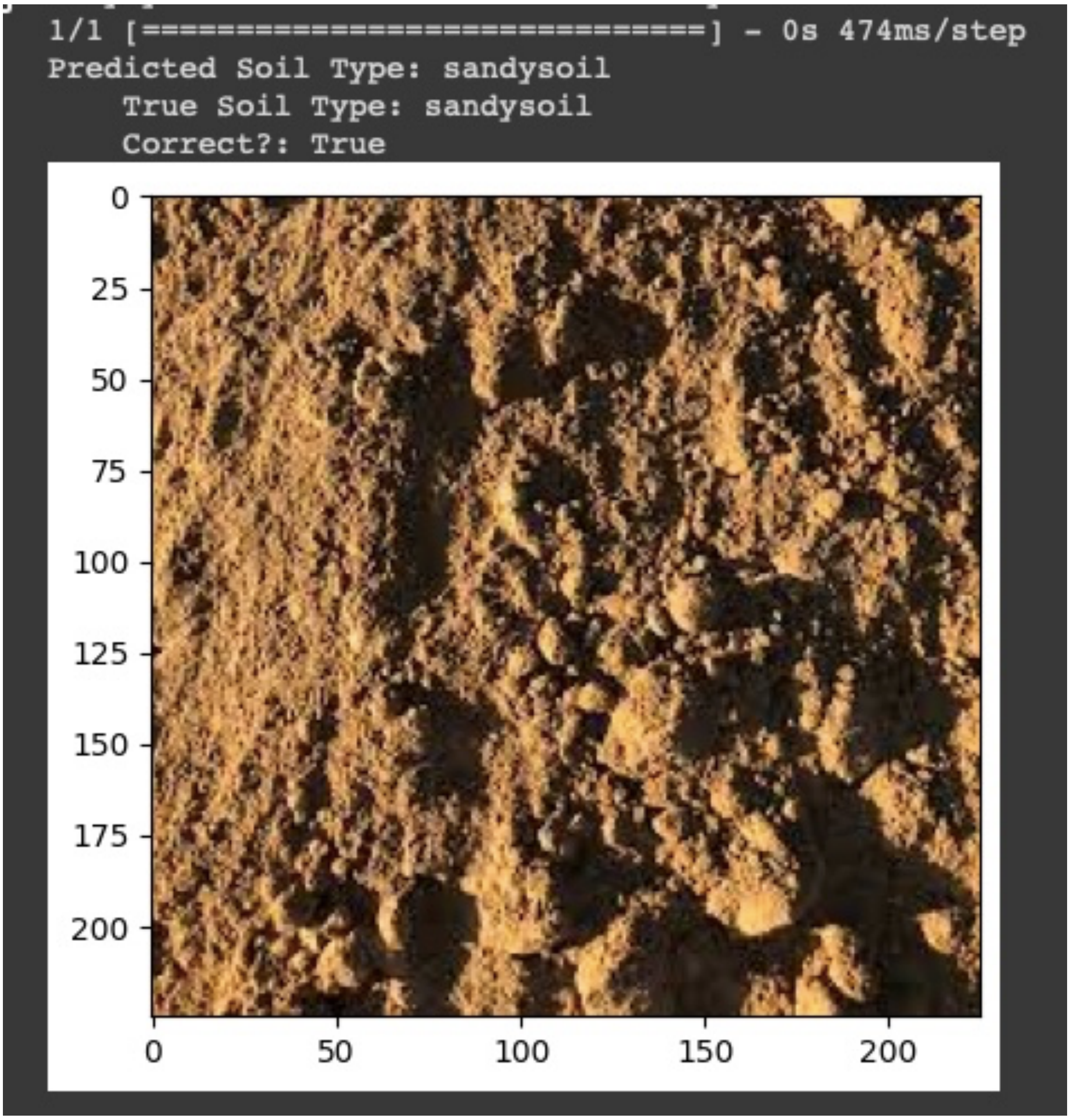


Figure 20: Loading soil for Validation.

As can be seen in the picture, the soil image is processed, features are extracted using the Gabor filter, and CNN classifies these features. The resulting result indicates that the test soil image in question is 56.25% accurate and corresponds to sandy soil.

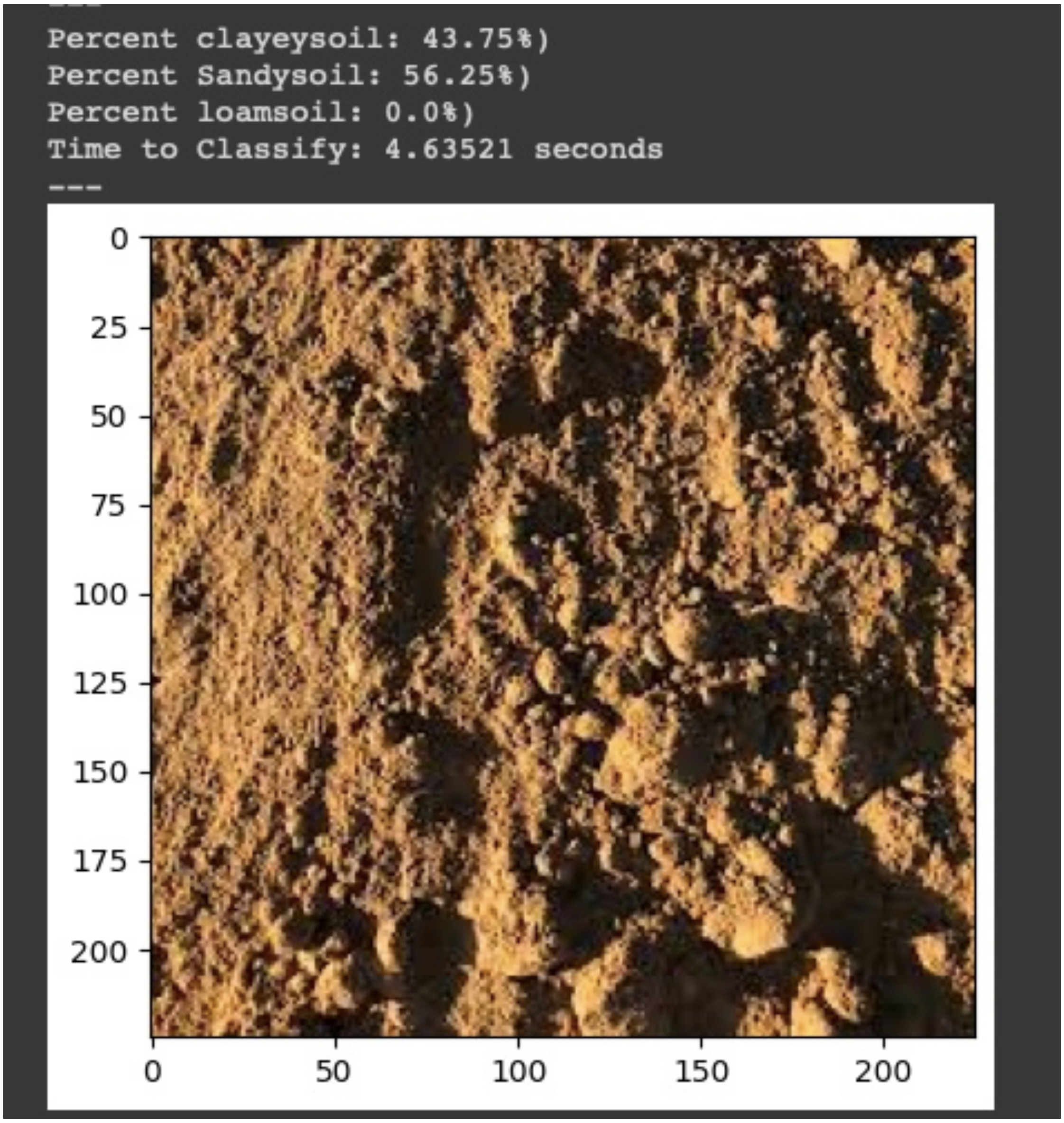


Figure 21: Accuracy for the validation.

CHAPTER 6

This study delves into automating the soil classification task using diverse sets of soil images.

6.1 Conclusion

In conclusion, Soil classification is a growing research area in the current era. Many studies have performed and some of them have provided similar techniques and some provided different techniques to deal with the issues, including rule-based, statistical, and traditional learning. However efficient soil classification is achieved through the utilization of Convolutional Neural Networks (CNNs). This study delves into automating the soil classification task using diverse sets of soil images. The process involves gathering these images, subjecting them to preprocessing techniques like color moments, and subsequently extracting features using Gabor filters. The final step entails employing a CNN classifier for accurate classification. The results yielded by the CNN classifier demonstrate a notable level of precision compared to alternative methods such as Artificial Neural Networks (ANN) and linear Support Vector Machines (SVM). Furthermore, there is potential for extracting additional features to enhance real-time soil classification capabilities in practical scenarios.

6.2 Future Scope

The proposed advanced approach to soil classification through image pre processing and feature extraction offers a promising avenue for future research in soil analysis. Further refinement of this methodology can enhance classification accuracy and efficiency. Future research directions include exploring additional image pre processing techniques, integrating hybrid machine learning models like CNNs with other algorithms, and incorporating multi-modal data fusion for comprehensive soil characterization. Additionally, investigating deep learning architectures beyond CNNs may provide novel perspectives for improving classification methodologies. Parameters like layer size play a crucial role in model performance, impacting training speed and accuracy. Iterative classification yields fractional predictions, offering insights into soil composition distribution. These advanced techniques could capture temporal and spatial dependencies in soil data, leading to more nuanced classification outcomes.

References

[1] Padarian, J., Minasny, B., & McBratney, A. B. (2019). Machine learning and soil sciences: A review aided by machine learning tools.

[2] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 215232.

[3] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

[4] Pavón-Pulido, N., López-Riquelme, J. A., Torres, R., Morais, R., & Pastor, J. A. (2017). New trends in precision agriculture: A novel cloud-based system for enabling data storage and agricultural task planning and automation. Precision agriculture, 18, 1038-1068.

[5-9] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758.

[10] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

[11] Harikaran, G. K. (2022). Machine Learning Approaches for Predicting Soil Properties of Vemagal Hobli, Kolar District, Karnataka (Doctoral dissertation, University of Agricultural Sciences, GKVK, Bangalore).

[12-15] Batjes, N. H., Ribeiro, E., & Van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth System Science Data, 12(1), 299-320.

[16] Deorankar, A. V., & Rohankar, A. A. (2020, June). An analytical approach for soil and land classification system using image processing. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 1416-1420). IEEE.

[17] Riese, F. M., & Keller, S. (2019). Soil texture classification with 1D

convolutional neural networks based on hyperspectral data. arXiv preprint arXiv:1901.04846.

[18] Devi, M. P. K., Anthiyur, U., & Shenbagavadivu, M. S. (2016). Enhanced crop yield prediction and soil data analysis using data mining. International Journal of Modern Computer Science, 4(6).

[19] Agarwal, S., Bhangale, N., Dhanure, K., Gavhane, S., Chakkarwar, V. A., & Nagori, M. B. (2018, July). Application of colorimetry to determine soil fertility through naive bayes classification algorithm. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). Ieee.

[20] Biswas, S., Faysal, T. I., Promiti, A. S., Hossain, M. S., Bazlul, L., Sarwar, A. M., ... & Rahman, R. M. (2020, August). Analysis of soil and various geo-technical properties using data mining techniques. In 2020 IEEE 10th International Conference on Intelligent Systems (IS) (pp. 288-293). IEEE.

[21] Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geoscience Frontiers, 12(1), 469-477.

[22-25] Masud, M., Gao, J., Khan, L., Han, J., & Thuraisingham, B. M. (2010). Classification and novel class detection in concept-drifting data streams under time constraints. IEEE Transactions on knowledge and data engineering, 23(6), 859-874.

[26] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 eighth international conference on advanced computing (ICoAC) (pp. 32-36). IEEE.

[27] Champaneri, M., Chachpara, D., Chandvidkar, C., & Rathod, M.

(2016). Crop yield prediction using machine learning. Technology, 9(38).

[28] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[29] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). pmlr.

[30] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[31] Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models for practical applications. arXiv preprint arXiv:1605.07678.

[32] Carvalho, F. P. (2006). Agriculture, pesticides, food security and food safety. Environmental science & policy, 9(7-8), 685-692.

[33] Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., ... & Kumar, V. (2017). Counting apples and oranges with deep learning: A data-driven approach. IEEE Robotics and Automation Letters, 2(2), 781-788.

[34] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[35] Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. IEEE Journal of Selected topics in applied earth observations and remote sensing, 7(6), 2094-2107.

[36-40] Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities.

Proceedings of the IEEE, 104(11), 2207-2219.

[41] Christiansen, P., Nielsen, L. N., Steen, K. A., Jørgensen, R. N., & Karstoft, H. (2016). DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. Sensors, 16(11), 1904.

[42] Deng, J. (2009). A large-scale hierarchical image database. Proc. of IEEE Computer Vision and Pattern Recognition, 2009.

[43] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[44] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[45] Douarre, C., Schielein, R., Frindel, C., Gerth, S., & Rousseau, D. (2016). Deep learning based root-soil segmentation from X-ray tomography images. BioRxiv, 071662.

[46] Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. Biosystems engineering, 151, 72-80.

[47] Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. Science, 327(5967), 828-831.

[48-50] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural computation, 12(10), 2451-2471.

[51] Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. Computers and electronics in agriculture, 127, 418-424.

[52] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., &

Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. Information systems, 47, 98-115.

[53] Ishimwe, R., Abutaleb, K., & Ahmed, F. (2014). Applications of thermal imaging in agriculture—A review. Advances in remote Sensing, 3(03), 128.

[54] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and electronics in agriculture, 147, 70-90.

[55] Kamilaris, A., Gao, F., Prenafeta-Boldu, F. X., & Ali, M. I. (2016, December). Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications. In 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT) (pp. 442-447). IEEE.

[56] Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and electronics in agriculture, 143, 23-37.

[57] Kitzes, J., Wackernagel, M., Loh, J., Peller, A., Goldfinger, S., Cheng, D., & Tea, K. (2008). Shrink and share: humanity's present and future Ecological Footprint. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1491), 467-475.

[58] Tea, K., Kitzes, J., Wackernagel, M., Loh, J., Peller, A., Goldfinger, S., & Cheng, D. Shrink and share: humanity's present and future Ecological.

[59] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[60] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[61] Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing

data. IEEE Geoscience and Remote Sensing Letters, 14(5), 778-782.

[62] Kuwata, K., & Shibasaki, R. (2015, July). Estimating crop yields with deep learning and remotely sensed data. In 2015 IEEE international geoscience and remote sensing symposium (IGARSS) (pp. 858-861). IEEE.

[63] LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.

[64] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning, nature, 521, 436–444. DOI: https://doi. org/10.1038/nature14539.

[65] Liaghat, S., & Balasundram, S. K. (2010). A review: The role of remote sensing in precision agriculture. American journal of agricultural and biological sciences, 5(1), 50-55.

[66-69] Lu, H., Fu, X., Liu, C., Li, L. G., He, Y. X., & Li, N. W. (2017). Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning. Journal of Mountain Science, 14, 731-741.

[70] Luus, F. P., Salmon, B. P., Van den Bergh, F., & Maharaj, B. T. J. (2015). Multiview deep learning for land-use classification. IEEE Geoscience and Remote Sensing Letters, 12(12), 2448-2452.

[71] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 215232.

[72] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of big data, 2, 1-21.

[73-77] Ozdogan, M., Yang, Y., Allez, G., & Cervantes, C. (2010). Remote sensing of irrigated agriculture: Opportunities and challenges. Remote sensing, 2(9), 2274-2304.

[78] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[79] Rahnemoonfar, M., & Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. Sensors, 17(4), 905.

[80] Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. sensors, 16(8), 1222.

[81] Santoni, M. M., Sensuse, D. I., Arymurthy, A. M., & Fanany, M. I. (2015). Cattle race classification using gray level co-occurrence matrix convolutional neural networks. Procedia Computer Science, 59, 493-502.

[82] Saxena, L., & Armstrong, L. (2014). A survey of image processing techniques for agriculture.

[83] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

[84] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[85] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience, 2016.