Soil Fertility Prediction and Fertilizer Recommendation Using Deep Learning Approaches

Mekedes Mulugeta Mekonene

A Thesis Submitted to

The Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

September 2022

Adama, Ethiopia

Soil Fertility Prediction and Fertilizer Recommendation Using Deep Learning Approaches

Mekedes Mulugeta Mekonene

Advisor: Dr. Biruk Yirga

A Thesis Submitted to Department of Computer Science and Engineering School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

September 2022

Adama, Ethiopia

DECLARATION

I hereby declare that this Master Thesis entitled "**Soil Fertility Prediction and Fertilizer Recommendation Using Deep Learning Approaches**" is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

Mekedes Mulugeta Mekonene	__________________	___________
Name of student	Signature	Date

RECOMMENDATION OF ADVISORS/ SUPERVISORS

I, the major advisor/supervisor of this research thesis, hereby certify that I have closely advised/supervised the student while developing this thesis and read the draft thesis/dissertation entitled "**Soil Fertility Prediction and Fertilizer Recommendation Using Deep Learning Approaches"** prepared under my guidance by Mekedes Mulugeta Mekonene. Therefore, I recommend the submission of the thesis to the department for further review and evaluation.

Major Advisor/Supervisor ________________ ___________________ ______________

Signature Date

Co-advisor/Co-supervisor _________________ ________________ _______________

Signature Date

APPROVAL PAGE

I/we, the advisors of the thesis entitled "**Soil Fertility Prediction and Fertilizer Recommendation Using Deep Learning Approaches**" and developed by Mekedes Mulugeta Mekonene, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

Major Advisor/Supervisor	Signature	Date
Co-advisor/ Co-supervisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by Mekedes Mulugeta Mekonene have read and evaluated the thesis entitled "**Soil Fertility Prediction and Fertilizer Recommendation Using Deep Learning Approaches**" and examined the candidate during open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in Computer Science and Engineering.

Chairperson	Signature	Date
Internal Examiner	Signature	Date
External Examiner	Signature	Date

Finally, approval and acceptance of the thesis is contingent upon the submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC)

Department Head	Signature	Date
School Dean	Signature	Date
Office of Postgraduate Studies, Dean	Signature	Date

ACKNOWLEDGMENT

First and foremost, I would like to thank God Almighty for providing me with the chance and guidance to achieve my objective and be successful in this part. Then I'd want to thank my mother Muluken Ferede and father Mulugeta Mekonene for their upbringing of me, as well as their unending efforts and support in whatever path I take to attain my goals. I'm also grateful to my other family members and friends for their continued support.

A particular thank you goes out to everyone, especially my adviser, Dr. Biruk Yirga, for his continuous support and guidance throughout the various stage of this thesis work. Without his personal involvement and intervention at critical stages, it would have been very challenging to complete this thesis according to the initial set plan. He provided critical and useful feedback and suggestions on how to approach the research problem systematically.

And I'd like to thank Dr. Tadele Amare, senior soil science researcher at Adet Agricultural Research Center, for his ongoing assistance with data assessment and dataset labeling, as well as for providing me with appropriate reference reading materials. I'd also like to thank all of my classmates, as well as others who have contributed to this thesis work either directly or indirectly.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

ANN	Artificial Neural Network
AARC	Adet Agricultural Research Center
ANRS	Amhara National Regional State
CaCO3	Calcium Carbonate
CV	Cross Validation
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DAP	Di-ammonium Phosphate
DL	Deep Learning
DNN	Deep Neural Network
EC	Electronic Conductivity
GDP	Gross Domestic Product
GOE	Government of Ethiopia
GIS	Geographic Information System
IEEE	International Electrical and Electronics Engineering
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MOP	Muriate of Potash
NPK	Nitrogen, Phosphorous, Potassium
PH	Power of Hydrogen
RF	Random Forest
SOC	Soil Organic Carbon
SFI	Soil Fertility Index

ABSTRACT

Agriculture has a significant role in defining a country's overall development, both economically and in terms of the quality of life for its citizens. We are now adopting smart farming approaches that use machine learning and data science to simplify the agricultural process and assist increase in both quantity and quality. The soil is an essential component of agricultural production because it contains the nutrients required to grow crops and adequate usage of appropriate fertilizer is a key for agricultural sustainability. So, we must first understand the properties and characteristics of various soil types to state the fertility status of the given soil samples and also type of fertilizer to apply. Therefore, this study is aimed to build predictive models for the prediction of soil fertility and suitable fertilizer by using a dataset collected from the Debre Markos soil laboratory and Adet agricultural research center which include 12,509 soil samples with 17 variables and 2 of them are dependent variables. Before feeding the data into the model it undergoes several preprocessing steps such as cleaning, handling missing and categorical values and normalization. Deep learning techniques such as Convolutional Neural Network (CNN), Artificial Neural Network (ANN), Deep Neural Network (DNN) and also Random Forest (RF) which is most previously used machine learning model is experimented in this study and results are compared to identify the most performing model. The performance of the models is evaluated under stratified 10-fold cross-validation with performance metrics such as accuracy, precision, recall, and f1-score also confusion matrix is used to assess the classification report. And the performance evaluation results revealed that the CNN model best performs than other models for both soil fertility and fertilizer prediction since it scored an accuracy of 98.82%, precision of 98.80%, recall of 98.77%, and f1-score of 98.78% for soil fertility prediction and accuracy of 97.22%, precision of 98.56%, recall of 97.72%, and f1-score of 97.77% for fertilizer prediction. While this study uses deep learning techniques to predict soil fertility and fertilizer, it would be preferable to include soil nutrient level prediction by using a large dataset.

Keywords: Soil fertility, Fertilizer, CNN, Machine learning, Deep learning

CHAPTER ONE

1 INTRODUCTION

1.1 Background of the Study

The desire for food is among the things that all people on the planet share. Producing food for human consumption therefore plays a crucial role in the world economy[1]. Having complete, wholesome, and healthy food systems is essential for achieving the goals of world development. One of the most successful strategies for eradicating extreme poverty is agricultural development.

Of all the regions of the world, sub-Saharan Africa has the fastest growing population[2]. More profits in farming will produce greater consumption by millions of individuals in country zones that fuel the move to a more differentiated and vigorous economy (Jayne et al., 2018). In many African countries, including Niger, Chad, and Uganda, more than 70% of the population is employed in agriculture[3].

One of the nations in sub-Saharan Africa is Ethiopia. As we all know, agriculture is the main driver of our nation's economy, contributing 40% of the GDP, 80% of exports, and an estimated 75% of the labor force(Mellor & Dorosh, 2010). Because market linkages are powerless, and the use of improved seeds, fertilizers, and pesticides remains restricted. Despite of these challenges, agriculture-based economic growth that is connected to improved livelihoods and nutrition can become the best and most unavoidable solution to Ethiopia's poverty and food insecurity. Moreover, the Government of Ethiopia has taken a position on this matter. It is committed to the rapid growth of agriculture as a means of accelerating the economic transformation(MoARD, 2010). So, to move forward in agricultural practices precision agriculture is needed because it merges the new technologies borne of the information age with a mature agricultural industry(Davis et al., 1998).

[1] https://howmuch.net/articles/role-agriculture-around-the-world
[2] https://blogs.worldbank.org/africacan/7-facts-about-population-in-sub-saharan-africa
[3] https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS

Soil is called the skin of the earth and is incredibly affected by topographical, natural, and climate parameters. The significance of soil fertility is a vital activity that is rather a complicated task due to the inclusion of numerous parameters(Kheyrodin, 2014). Ethiopia has incredible agricultural potential because of its tremendous regions of fertile land, diverse climate, generally adequate rainfall, and huge labor pool(Chamberlin et al., 2015). However, soil fertility is declining from time to time due to rapid population growth, the alteration in rural exercises, the absence of crop rotation, poor soil conservation or land management practices, and so on(Elias, 2019). The entire reasons lead to food insecurity in Ethiopia. Therefore, the application of artificial intelligence for the prediction of soil fertility status and determination of appropriate fertilizer is essential for sustainable soil management.

Deep learning and machine learning algorithms are widely applied in many fields. The use of artificial intelligence in agriculture enables much higher levels of precision and allows farmers to treat each plant and animal almost individually, greatly enhancing the effectiveness of their decisions[4]. The main applications of machine learning in agriculture are species management (species selection, species recognition), field conditions management (soil management, water management), crop management (yield prediction, crop quality, disease detection, weed detection), and livestock management[5]. Since agriculture is the foundation of Ethiopia's economy must concentrate on enhancing and automating all tasks in this sector. Successful agriculture depends heavily on soil and fertilizer usage. All the essential nutrients for basic plant nutrition, such as nitrogen, phosphorus, and potassium, as well as other nutrients required in smaller amounts, will be available in fertile soil (e.g., calcium, magnesium, sulfur, iron, zinc, copper, boron, molybdenum, nickel)[6]. To increase the productivity of cereal crops, this study applies and compares various deep learning algorithms in soil fertility prediction and fertilizer recommendation by taking into account various soil nutrients such as nitrogen, oxygen, phosphorus, PH value, etc. This study aims to help farmers and soil testing experts to predict the soil fertility status and the appropriate type of fertilizer to apply to make the soil more fertile for better yield production. We have trained a soil dataset that contains various attributes of the soil samples with their respective value that was collected from the soil laboratories. Then different

[4] https://www.iflexion.com/blog/machine-learning-agriculture
[5] https://encyclopedia.pub/entry/11729
[6] https://agrilifeextension.tamu.edu/asset-external/essential-nutrients-for-plants/

Deep learning algorithms are applied to classify the soil whether the soil is very highly fertile, highly fertile, moderately fertile, low fertile, or very low fertile by using soil physiochemical features. Also, the best fertilizer to use is recommended depending on the soil's macronutrient values, and soil samples with good macronutrient values are not recommended to apply any fertilizer; otherwise, the model predicts optimal fertilizer types, which include DAP, MOP, Urea DAP and MOP, and Urea and DAP for the supplied soil samples. The algorithm assessed involves a one-dimensional convolutional neural network (1D-CNN), artificial neural network (ANN), and deep neural network (DNN) and for comparing results with machine learning models Random Forest (RF) has been experimented.

1.2 Motivation of the Study

Since agriculture is the backbone of the Ethiopian economy, we must all work together to advance and enhance this sector in order to maintain national food and nutrition security. Any agricultural production is dependent on soil in both direct and indirect ways, which is why we must focus on supporting soil fertility management techniques with new technologies. Several smallholder farmers in the undeveloped nations, including Ethiopia, are impoverished, and poor soil management practices and inappropriate usage of fertilizer are one of the main causes of crop yield reduction(Zerssa et al., 2021). In the coming years, one of Ethiopia's most pressing challenges will be food security(Conijn et al., 2019). For the vast majority of farmers, agriculture is their primary source of income(Ahmed, 2016). And, because the soil is the source of nutrients used to grow crops, it is an essential component of successful and effective agriculture.

Artificial intelligence can be applied in various areas to solve many problems and simplify our lives. The current level of cereal crop production is insufficient to fulfill domestic demand(Taylor, 2020). The Government of Ethiopia (GOE) has identified key priority intervention areas to increase the productivity of smallholder farms and expand large-scale commercial farms[7]. Previous AI studies on this area, especially soil fertility prediction was experimented by using traditional machine learning techniques and they didn't include any consideration of fertilizer recommendations. So, looking at these issues we are motivated to use deep learning techniques to predict soil fertility status and appropriate fertilizer. The employment of these techniques yields

[7] https://www.privacyshield.gov/article?id=Ethiopia-Agricultural-Sector

superior results and allows farmers and soil testing professionals to make quick judgments and take essential steps, such as making appropriate suggestions regarding which sorts of crops to cultivate, crop rotation, and the application of adequate fertilizers. These prediction models save a lot of time and money.

1.3 Statement of the Problem

A key component of our environment is the soil. It allows for the growth of a variety of plants and crops that are beneficial to both humans and the environment, holds and purifies water, recycles nutrients, and gives a lot of earth's living things a place to call home. We wouldn't be able to cultivate any crops without soil, which would put an end to the cycle of life on earth(Khomiakov, 2020). Smallholder farmer's traditional farming practices gradually erode soils. In Ethiopia farmers' have gained their knowledge of farming from generations of experience and experimentation that fit local conditions since they are interacting with their soil for a long time but this traditional knowledge affects soil management. Plants will benefit from having access to the right nutrients if the soil is healthier, and this can only be done by continuously managing the soil's fertility. Crop production primarily depends on the rate of soil nutrients. Since soil properties are temporal because they might vary within a short distance and time soil fertility management is a plan for farmers and scientists(Fairhurst, 2012). The results of soil samples are manually recorded, classified, and recommend fertilizer in the soil laboratories we visited. This takes a lot of time, and there is also a high risk of misclassification because human error is so common.

Misuse of chemical fertilizers can contribute to soil acidification and soil crust, lowering organic matter, humus content, and beneficial species, limiting plant growth, altering soil pH, developing pests, and even causing greenhouse gas emissions(Rahman & Zhang, 2018). Although many prediction models were utilized in earlier studies for predicting soil fertility, Only traditional machine learning techniques such as SVM, RF, KNN, GB, and simple neural networks such as ANN were used, and all of the important soil features such as (N, P, K, PH, OC, OM, Ca, Mg, Zn, S, Fe, B, Cu) and physical characteristics such as silt, clay, and sand composition that has a direct impact on determining soil fertility were not considered, beside that fertilizer prediction was not taken into account. To our knowledge, in Ethiopia only (Abebe, 2021) proposed a land suitability prediction model for wheat and barley crops using traditional machine learning techniques, specifically Gradient Boosting; however, this paper, in addition to using only traditional machine learning, it considers only wheat and barley crops, and many soil parameters, particularly micronutrients, were not considered for land analysis. Therefore, the proposed model for soil fertility prediction and fertilizer recommendation using deep learning techniques can help agricultural experts, soil testing experts, land evaluators, and hence farmers.

1.4 Research Question

This study addresses the problem by answering the following question:

RQ1**:** Which deep learning algorithm is more accurate for the prediction of soil fertility based on physiochemical characteristics of the soil?

RQ2**:** Which deep learning model is best for the prediction of suitable fertilizer type with optimal accuracy?

RQ3: How to optimize the performance of the deep learning models that will be used in soil fertility prediction and fertilizer recommendation?

1.5 Objective of the Study

1.5.1 General Objective

The objective of this study is to build the prediction model for soil fertility and fertilizer recommendation by using deep learning techniques.

1.5.2 Specific Objective

To meet the above general objective, the following specific objectives are set:

- ✓ To review literature concerning soil fertility prediction
- ✓ To collect soil physiochemical data and prepare a dataset for model training
- ✓ To prepare and pre-process the collected data
- ✓ To build soil fertility and fertilizer recommendation model
- ✓ To choose evaluation techniques and metrics
- ✓ To evaluate the prediction model's performance
- ✓ To select a model with high performance.

1.6 Scope and Limitations

1.6.1 Scope of the Study

This research study is intended to build soil fertility prediction and fertilizer recommendation models for better soil fertility management by using deep learning algorithms using collected soil physiochemical data from Debre Markos soil laboratory and Adet agricultural research center. The dataset includes soil samples that were collected starting from 2011 up to 2013 E.C. This research covers the East Gojjam zone which is located in the Amhara region as a study area. Various deep learning and machine learning algorithms such as 1D-CNN, ANN, DNN, and RF are used and evaluated using stratified k-fold cross-validation.

1.6.2 Limitation of the Study

This thesis does not cover the following due to time and resource limitations.

- The lack of high-performance hardware capable of deep learning was a major stumbling block to the achievement of this thesis. Deep learning is a computationally difficult task that requires high-performance technology such as the Graphics Processing Unit (GPU).
- Due to lack of time and data this study focuses only on the prediction of soil fertility and suitable fertilizer but it does not consider the prediction of values of the soil attributes.
- Furthermore, budget, deadline, and resource constraints limited greater practicability.

1.7 Significance of the Study

Soil fertility prediction alone does not promise us more yield, but it does promise us the extent to which our incorporated methods have improved soil quality. Soil fertility prediction will help to condense the difficulties faced by farmers and act as a medium to bid the agriculturalists efficient evidence required to get a better yield. Once the soil fertility prediction and fertilizer recommendation are fully implemented, it assists agricultural experts such as soil experts, land evaluators, and farmers in determining the level of soil fertility and adequate type of fertilizer to apply for the particular soil this helps in achieving reliable and economic fertilizer recommendation. If this model is further extended with the application of IoT devices and deployed with Android devices, it greatly decreases farmers' confusion regarding soil status and fertilizer consumption in order to increase production.

1.8 Contribution of the Study

The contributions of this research study are:

- Preparing and preprocessing of the raw soil dataset collected from soil laboratories which include cleaning, handling missing, handling categorical data, and normalization.
- Conduct experiments and comprehensive evaluation of all applied approaches with the preprocessed soil dataset. The experiment finding shows that the proposed deep learning approaches which are Convolutional Neural Network (CNN), Deep Neural Network (DNN), and Artificial Neural Network (ANN) for soil fertility prediction and fertilizer recommendation show better accuracy as it is compared to previously used traditional machine learning methods.
- In our research, we attempted to estimate the optimal type of fertilizer to apply based on values of soil macronutrients, which were not taken into account in previous studies.

1.9 Organization of the Thesis

In this section, the organization of the whole thesis report which includes seven chapters is indicated as follows.

Chapter one describes the introductory part of this study including the background and motivation behind this study, statement of the problem that this research going to address, research questions to be answered, objectives to be achieved, scope and limitations of the study, and its significance.

Chapter two defines and discusses the theoretical background of a soil and soil resource, existing soil fertility evaluation, machine learning, deep learning, and literature related to soil fertility assessment using machine learning and deep learning algorithms.

The third chapter describes the source of data and study area, how the dataset is prepared, and labeled, various data preprocessing steps, and soil fertility prediction modeling including model evaluations that are used to achieve the goal of this research.

Chapter four describes the proposed solution for soil fertility prediction and fertilizer recommendation. In this section, the deep learning model architecture for soil fertility prediction is designed. How to achieve the proposed solution and how it works is also fully explained.

Chapter five describes all the hardware used in this study, the experimental tools, dataset feature description and implementation of dataset preprocessing, and the building of deep learning models including model testing and evaluation implementation is also done.

In chapter six, the obtained results from experimentation and implementation, dataset class distribution results, model evaluation results, and hyperparameter tuning results were discussed in detail including prediction models to figure out the most performing model.

Chapter seven is the final chapter that concludes this thesis report, then the recommendation and future research direction of this study are clearly stated.

CHAPTER TWO

2 LITERATURE REVIEW AND RELATED WORK

This chapter focuses on a survey of scholarly sources like Books, journal articles, and theses related to agriculture and soil fertility in a particular application of machine learning and deep learning approaches in agriculture and soil fertility prediction. The goals of the literature review are to summarize information, gain an understanding of the subject matter, and identify the strengths and gaps to be fulfilled in previous research on soil fertility prediction.

2.1 What is Soil and Soil Fertility

Soil is the erodible material that covers the majority of the land. It is made up of both inorganic and organic particles and it provides structural support to agricultural plants as well as a source of water and nutrients(van Es, 2017). The chemical and physical properties of soils vary greatly(McCauley et al., 2005). Our relationship with the soil has influenced our ability to cultivate crops and the success of civilizations throughout history and this interaction between humans, the environment, and food sources validates soil as the cornerstone of agricultural production[8].
Soil fertility refers to a soil's ability to support plant growth by providing essential plant nutrients as well as favorable chemical, physical, and biological properties as a habitat for plant growth[9].

2.2 Existing Soil Fertility Evaluation

The next step after identifying soil and soil fertility, is to define and understand soil fertility evaluation and its assessment. The process of calculating the quantity of native and residual nutrient elements that may be accessible for use by growing crops in a certain soil, as well as the amount of fertilizer that must be provided for profitable crop production, is known as soil fertility evaluation(A et al., 2012).

Soil productivity is limited by both physical and chemical fertility (Dobo, 2019). The study of soil fertility includes analyzing the types of plant nutrients that exist in the soil, how they become available to plants, and the conditions that influence their uptake. As a result, a study of the steps that may be taken to improve soil fertility and crop output by providing nutrients to the soil-plant

[8] https://agriculture.vic.gov.au/farm-management/soil/what-is-soil
[9] https://www.iaea.org/about/partnerships/other-international-

system is undertaken. Soil fertility evaluation was performed by different researchers worldwide as well as in Ethiopia too. Since the 1970s, various mechanical and biological soil conservation strategies have been implemented in Ethiopia to address the problem of declining soil fertility. In addition, as part of the country's national economic growth strategy, a lot of attempts were made to improve soil fertility and agricultural output (Worku & Tripathi, 2015).

2.3 Soil Characteristics

The majority of the Earth is covered in soil, which is the uppermost layer of the earth's crust. By providing plants with nutrients and water, soil sustains all life on Earth. Soil is essential for not only the plants that grow there but also for people and other organisms that depend on plants for food and shelter(Elbasit M M, 2015).

Soil characteristics like color, texture, and odor of soil are essential to know for anyone working in agriculture or horticulture. To successfully grow crops and ornamental plants, farmers, for example, must gain a fundamental awareness of what soil is and how to regulate the fertility of their soils(KUMAR et al., 2019)

2.3.1 Soil Physical and Chemical Properties

Changes in the quantity and availability of mineral nutrients caused by the addition of fertilizers, manure, compost, mulch, and lime or sulfur, as well as leaching, cause soil fertility to change throughout the growing season each year. Soil testing will assess the existing fertility status and provide information on nutrient availability in soils, which will serve as the foundation for fertilizer prescriptions aimed at improving crop yields and maintaining optimum soil fertility year after year [10]. Soil contains various types of nutrients which are known as macro and micro nutrients. Major nutrients are nutrients found in significant quantities in the soil that plants require. As a result, plants rely heavily on these nutrients. In addition, they are frequently found in high quantities in the soil. And minor nutrients are nutrients contained in the soil that plants require in minute amounts. It's possible that supplying them in big numbers will harm the plants [11]. There are about sixteen elements, which are divided into major and minor nutrients based on plants chemical elements need and some of them are explained below.

The macro nutrients are:

[10] https://www.futureoffertilizer.com
[11] https://qknowbooks.gitbooks.io/jhs_3_science-soil-and-water-conservation/content/

Carbon (C): Carbon is a crucial element and contributor to healthy soil conditions, as well as a crucial element and contributor to the production and function of soil. A key strategy for increasing agricultural yields, improving soil quality, and reducing soil loss is soil carbon management(Corning et al., 2016).

Oxygen (O): Plant growth necessitates the presence of oxygen. Reduced concentrations in the soil have an impact on plant physiological activities such as nutrition and water uptake, respiration, redox potential of soil components, and microbial activity. Diffusion is the primary mode of oxygen transport in the soil (Neira et al., 2015).

Nitrogen (N): The first most critical plant nutrient for agricultural production is nitrogen. It is a component of practically all plant structures' building blocks. N travels from the soil to the plant and from the plant (residue) back to the soil in most ecosystems thanks to microbial biomass. It's found in chlorophyll, enzymes, proteins, and other things. In soil and natural ecosystems, N occurs in many forms covering a range of valence states from –3 (in NH_4^+) to +5 (in NO_3^-) (Hofman G. and O.V. Cleemput, 2004).

Phosphorus (P): Phosphorus is the second most important macronutrient as a necessary plant nutrient, behind nitrogen. It's a crucial nutrient for increased and long-term agricultural productivity, but it also restricts plant development in many soils. Despite the vital function that soil P plays in plants, phosphorus deficit in soil is the most frequent nutritional stress in many parts of the world, impacting 42 percent of cultivated land (Mwende Muindi, 2019).

Potassium (K): Potassium, in addition to the two major components nitrogen and phosphorus, is the third essential macronutrient required for plant growth and metabolism, and its deficiency in plants results in poorly developed roots, slow growth, low resistance to disease, delayed maturity, limited seed production, and low production. Potassium is found in four various forms in the soil (Meena et al., 2016).

Calcium (Ca): Calcium is an essential plant nutrient responsible for the integrity of cells and plant structure, yet it is generally neglected, because it is available in plenty in most cultivated soils. It is therefore not applied as a fertilizer to crops except in groundnut. In acid soils, where it is limiting, it is applied in large amounts as a soil amendment as lime (Prasad & Shivay, 2020).

Magnesium (Mg): Magnesium attracts water more strongly than calcium, resulting in a wider hydrated radius. As a result, soil particles become more scattered and farther apart. As a result, soils with more magnesium have fewer water stable aggregates and pore integrity[12].

As well as macro nutrients plants need micronutrients to grow, but they only need a minimal amount of them, hence the term "micronutrient." Boron (B), chlorine (CI), copper (Cu), iron (Fe), manganese (Mn), Molybdenum (Mo), and zinc (Zn) are among them (Zn). Micronutrients can also be found in the soil. Furthermore, practically every soil contains some of each of these elements within the organic matter structure and as soluble organic acids (Voss, 1998).

Soil depth: For a plant to grow, the soil's depth is essential. Any discontinuities in the soil profile, from sand or gravel layers to bedrock, can physically restrict root penetration. When it comes to irrigation, it can also be problematic. In comparison to shallower soils, deeper soils can provide plants with more water and nutrients(Abd-Elmabod et al., 2017).

Soil Texture: One of the physical properties of soil is its texture. The fineness or coarseness of a soil is determined by the proportion of the three sizes of soil particles (sand, silt, and clay) present. A soil's texture is significant because it dictates soil qualities that influence plant growth. Water-holding capacity, permeability, and soil workability are a few of these qualities(June, 2017).

Soil PH: The soil's pH serves as an indicator for its acidity or alkalinity. It's an essential measurement for figuring out how healthy the soil is. It affects soil microbe activity, plant nutrient availability, crop yields, crop suitability, and crop suitability, all of which affect critical soil processes. Applying the proper amount of nitrogen fertilizer, liming the soil, and using cropping techniques that encourage soil organic matter and general soil health are some measures that can help maintain the pH of the soil. It's an excellent way to determine whether the soil is necessary for plant growth. The ideal range for most crops is between 6 and 7.5. Too high or too low pH values in the soil cause nutrient insufficiency, microbial activity drop, crop output decline, and soil health deterioration(Butterly et al., 2010).

[12] https://soilsolutions.net/why-gypsum-works-in-your-soils-part-4-gypsum-offsets-high-magnesium-in-soils/

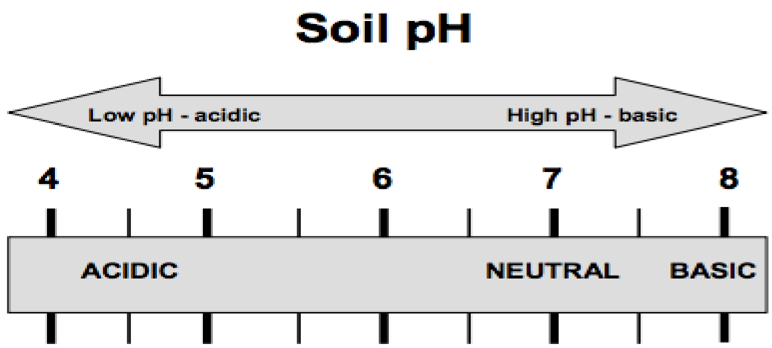


Figure 2-1 PH values and its rating

2.4 Fertilizer Management

Farmers must make sure that their soil is healthy if they want to produce nutrient rich and healthy crops. Nature tries to replace the nutrients in the soil without fertilizers but important nutrients are taken from the soil when crops are harvested because they follow the crop and are consumed during food preparation. Crop yields will gradually decrease if the soil is not fertilized to replace its nutrients[13].

Crops can be fertilized and carefully analyzed to create a food supply chain that gives people nourishing food. Planning the farming operation is also made simpler by knowing the precise nutritional contents of the mineral fertilizer. Farmers now have better access to fertilizers with major, secondary, and micronutrients[14].

2.5 Soil Fertility Rating System

Using various independent attribute of soil, the fertility class will be labeled as Very High, High, Moderate, Low and Very Low based on one of the soil fertility evaluation approaches which is soil fertility index.

A. Very Low: soil plant nutrients is deficient and the application of this element will significantly increase yield because the level of fertility is very low. To satisfy the crop's needs, account for soil interaction, and create soil reserves, a high application rate is required.

B. Low: soil plant nutrient elements level is not adequate even for moderate agronomic crop yields, but a yield response can be expected about 50% of the time from an application of this element.

[13] https://agrilifeextension.tamu.edu/?s=fertilizer
[14] https://extension.umn.edu/manage-soil-nutrients/quick-guide-fertilizing-plants

C. Moderate: The range of soil plant nutrient elements is sufficient to meet crop needs and to produce crops with consistently high yields. It is advised to apply at a maintenance application rate to account for anticipated crop removal. The essential plant nutrient element level in the subsoil won't be depleted if the surface soil is kept within the "Sufficient" range.

D. High: This soil's plant nutrient levels can have a negative impact on crop yield and product quality, and a further rise could cause crop yields to decline and plant nutrient imbalances. Therefore, it is not advised to add this element unless it is required to make up for anticipated high crop removal.

E. Very High: This level of soil plant nutrients is available excessively sometimes this might reduce plant yield, cause nutrient deficiencies as a result of imbalances, and possibly cause ecological harm to the environment(Khaki et al., 2017a).

2.6 Nutrient Requirements for Major Cereal Crops

Ethiopia is primarily concerned with cereal production. Ethiopia's crop agriculture is still dominated by the country's numerous small farms, which grow primarily cereals for both self-sufficiency and sale. Teff, wheat, maize, sorghum, and barley account for nearly three-quarters of all cultivated land (Taffesse et al., 2012). To increase the yields of the cereal crops it is necessary to work on the improvement of soil nutrient levels specially NPK values which are the most essential macro nutrients to the given specific crop requirements as indicated in the table below. And this whole soil fertility prediction and fertilizer recommendation is done to satisfy this crop nutrient requirements since the major cereal crops have almost similar nutrient requirement by applying appropriate action based on the observed soil status.

Table 2-1 nutrient requirements of wheat crop

Nutrient	Highly Suitable	Moderately Suitable	Not Suitable
T. N	>0.25	0.15-0.25	<0.1
AV. P	>15	8-15	<4
K	>1.2	0.6-1.2	<0.3
PH	5.5.0-8.0	5.2-5.5, 8.0-8.3	<4.8, >8.5

Table 2-2 nutrient requirements of Teff crop

Nutrient	Highly Suitable	Moderately Suitable	Not Suitable
T. N	>0.25	0.15-0.25	<0.1
AV. P	>8	4-8	<4
K	>1.2	0.6-1.2	<0.1
PH	5.5-7.3	5.3-5.5, 7.3-7.5	<5, >8.5

Table 2-3 nutrient requirements of Maize crop

Nutrient	Highly Suitable	Moderately Suitable	Not Suitable
T. N	>0.2	0.15-0.20	<0.1
AV. P	>15	8-15	<4
K	>1.2	0.6-1.2	<0.1
PH	5.5-6.7	5.4-5.5, 6.7-7.1	<5.0, >8.0

Table 2-4 nutrient requirements of Sorghum crop

Nutrient	Highly Suitable	Moderately Suitable	Not Suitable
T. N	>0.2	0.15-0.20	<0.1
AV. P	>15	8-15	<4
K	>1.2	0.6-1.2	<0.1
PH	6-7.5	5.5-6.0, 7.5-8.0	<5.0, >8.5

2.7 Machine Learning Techniques

Artificial intelligences of machine learning (ML) enable computers to automatically learn from data and past experiences while spotting patterns to make predictions without human intervention. Machine learning uses algorithms to identify patterns and learn in an iterative process in order to extract useful information from massive amounts of data. ML algorithms use computation techniques to learn directly from data, as opposed to relying on any predetermined equation that may serve as a model. The following are some of the most popular types of machine learning: supervised machine learning, unsupervised machine learning, semi-supervised machine learning, and reinforcement machine learning. There are other types of machines that can learn as well(Apt, 2014).

2.7.1 Supervised Machine Learning

This is a type of ML, in which machines are trained on labeled datasets and then allowed to predict outputs based on the training. Some input and output parameters are already mapped, according to the labeled dataset. As a result, the machine is trained using the input and output. In subsequent phases, a device is created to determine the outcome using the test dataset.

2.7.2 Unsupervised Machine Learning

Unsupervised learning is a learning technique that does not require supervision. In this case, the machine is trained on an unlabeled dataset and is able to predict the output without human intervention. An unsupervised learning algorithm attempts to assemble the unsorted dataset based on similarities, differences, and patterns in the input.

2.7.3 Semi-supervised Learning

Unsupervised and supervised machine learning components are combined in semi-supervised machine learning. It uses a mix of labeled and unlabeled datasets to train its algorithms. Using both types of datasets, semi-supervised learning overcomes the drawbacks of the prior options.

2.7.4 Reinforcement Learning

Reinforcement learning is a process that is based on feedback. Here, the AI component uses the hit-and-trial method to automatically assess its surroundings, take action, learn from experiences, and improve performance. The component is rewarded for each correct action and penalized for each incorrect move. As a result, the reinforcement learning component seeks to maximize rewards by performing positive actions[15].

2.8 Machine Learning Approach in Soil Fertility Prediction

Machine Learning is now used in nearly every task that requires estimation. These techniques have also benefited the non-technical sector, such as agriculture. This is true for predicting soil fertility using specific properties that vary by region. Crop rotation, cover crops, and the incorporation of organic matter into the soil have all helped this sector produce more crops. Nonetheless, they do

[15] https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/

not fully utilize the available domain knowledge. Experts and consultants have experimented with various machine learning methods to predict it. Classification algorithms were having better accuracy to deal with such a problem. For various case studies on soil fertility, ML algorithms such as KNN, DTs, SVMs, and Random Forests have been used on many previous research.

2.9 Deep Learning Approach in Soil Fertility Prediction

Deep learning is now utilized in numerous smart agriculture applications, including yield prediction, robust fruit counting, crop disease detection, weed removal, water and soil management, crop cultivation, and crop distribution. For maximum yield, picking the right crop at the right time is essential. Farmers are given a classification model that takes into account the nutrients that a crop needs from the soil to help them make decisions(Kamatchi, 2021).The prediction of soil fertility has been tested by experts using deep learning algorithms such as ANN and MLP. Also, some researchers work on soil texture prediction by using CNN model.

ANN (Multilayer Perceptron)

The Multilayer Perceptron (MLP), also known as an Artificial Neural Network (ANN), is one of the most popular machine learning techniques for prediction. It is an addition to a feed-forward neural network. Artificial Neural Networks (ANNs) are used for soil classification prediction in specific locations at various depths based on data from available site investigation from chosen sites. The input layer, the output layer, and the hidden layer are the three layers that make it up. The input layer receives the input signal that needs to be processed. The output layer is responsible for tasks such as prediction and classification. The true computational engine of the MLP is an arbitrary number of hidden layers placed between the input and output layers. In an MLP, data flows in the forward direction from input to output layer, similar to a feed forward network. The back propagation learning algorithm is used to train the neurons in the MLP. MLPs are intended to approximate any continuous function and to solve problems that cannot be solved linearly. Pattern classification, recognition, prediction, and approximation are the most common applications of MLP[16].

[16] https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron

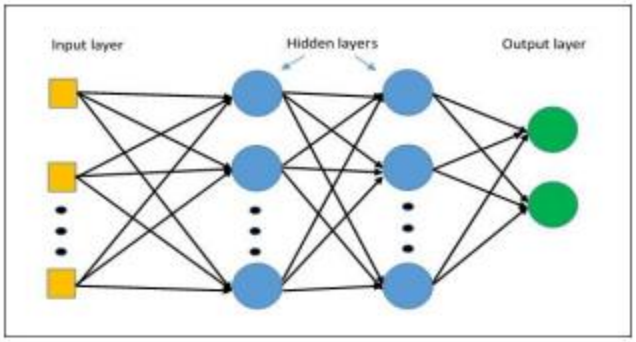


Figure 2-2 Multilevel Perceptron (Kamatchi, 2021)

2.10 Review of Related Research Works

The importance of soil in agriculture is obviously known but soil testing procedures and evaluation of soil fertility status is full of challenges and very tedious. Many researchers tried to simplify this process by applying machine learning and neural network approaches. Assessing the accuracy and cost effectiveness of available methods is crucial. And under this section we have reviewed proposed technical solution, methodology employed, and outcomes discovered in existing papers however, most of them used traditional machine learning approaches and some of them are discussed below.

The study (M S et al., 2020) Proposed a model for predicting the soil fertility and crop yield with types of the crop can grow on fertile soil. They have used various traditional machine learning algorithms like Support Vector Machine (SVM), Random Forest, Naive Bayes, Linear Regression, and ANN for soil classification and crop yield prediction. And Random Forest classifier shows good accuracy among all the classifiers but ANN has given the highest accuracy which is 95% in soil fertility prediction.

(Keerthan Kumar et al., 2019) Introduced soil fertility prediction which find out the rank of a given soil based on the previously graded soil by using a parallel random forest classifier. This study has various features that can be used in soil fertility prediction this are macro and micro soil properties such as organic content and essential plant nutrients, that affects the crop yield. They trained, tested, and measure the performance of different algorithms like RF, GNB, and SVM machine learning algorithms in order to make a comparative analysis with random forest, and RF had the best accuracy having 75.25%.

(Chaudhari et al., 2019) This paper aims to help farmers to predict the soil fertility for better yield production. To be precise and accurate in predicting fertility, it analyzes the nutrient contents present in the soil. In this system the soil database trained and they classify that particular soil sample into particular range (high, medium and low) using classification algorithms. This was achieved using unsupervised and supervised learning algorithms, like Decision tree, KNN, SVM, Naive bayes. As a, result the four algorithms have been compared and SVM shows good accuracy but among all the classifiers, Decision tree gives the highest accuracy in soil classification.

(Suchithra & Pai, 2020) The aim of this research was to categorize the soil samples by taking the macro nutrients like Nitrogen (N), Phosphorus (P) and Potassium (K) and Soil type, Soil texture, Land type, pH and Electrical conductivity as an input to predict the level of this soil nutrients they have used the fast learning classification technique known as Extreme Learning Machine (ELM) with different activation functions like gaussian radial basis, sine-squared, hyperbolic tangent, triangular basis, and hard limit.

Here in (Sekeroglu et al., 2020) among all the classification SVM has given the highest accuracy in Soil Classification. They have used several machine learning algorithms, such as neural networks, decision trees, naive Bayes, and SVM. The algorithm used to automate soil type classification with satisfactory accuracy (> 70%). And they also work on crop protection in organic agriculture. They used deep learning models that were developed, based on specific Convolutional neural network architectures. The research is developed for checking the various crop diseases to help the farmer. Summary of the related worked is shown in the following table.

(Abebe, 2021) Constructed a model that predicts land suitability for wheat and barley crops by using machine learning techniques based on the FAO guide lines. Machine learning models such as Random Forest (RF), Gradient Boosting (GB), and K-Nearest Neighbor (KNN) with feature selections methods are experimented in this study and gradient boosting shows the best result. But beside using only traditional machine learning approach crucial soil parameters which are both macro and micro nutrients were not considered in the data set preparation.

(Sivasankaran et al., 2022) Developed a model for predicting soil nutrient composition, which includes only macronutrients such as nitrogen, phosphorus, and potassium, as well as optimal fertilizer estimation using enhanced 1D-CNN DLM for groundnut crops. Although the use of this improved 1D-CNN deep learning algorithm is beneficial, only groundnut crops were considered in this work. This crop is not a major type of crop utilized in daily food, particularly in Ethiopia, and production of this crop is confined and declining with time(Sori, 2021). Since the paper focuses on the prediction of soil nutrient composition individually it is difficult to know the general status of the given soil sample. Aside from those factors, physical characteristics of the soil such as sand, silt, and clay composition were not taken into account for soil analysis.

Table 2-1 Summary of related works

S/N	Title	Author	Gap	
1	Soil Analysis and Crop Fertility Prediction Using Machine Learning	Jagdeep Yadav, Shalu Chopra, Vijayalakshmi M (2021)	• Physical properties of the soil that has crucial effect on soil were not considered. • It is binary class classification which just decides given soil sample is fertile or not. • Fertilizer recommendation was not considered.	SVM, RF, Naive Bayes, Linear Regression, and ANN
2	Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning	Keerthan Kumar T G, Shubha C, Sushma S(2019)	• Too small features were used for the soil analysis. • Only traditional ML model were applied. • Fertilizer recommendation was not considered.	RF
3	Soil Fertility Prediction Using Data Mining Techniques	Rajat Chaudhari, Saurabh Chaudhari, Atik Shaikh, RaginiChiloba,	• The dataset is gathered from agricultural books and websites. • Only two ML models were experimented.	SVM and KNN

		Prof.T.D. Khadtare (2020)		
4	Land suitability prediction for wheat and barley crops using machine learning techniques	Bikila Abebe (2021)	• Only wheat and barley crops are considered in this study. Fertilizer recommendation was not considered. • Only traditional ML models were applied.	SVM, RF, GB and KNN
5	Soil Nutrients Prediction and Optimal Fertilizer Recommendation for Sustainable Cultivation of Groundnut Crop using Enhanced-1DCNN DLM	(Sivasankaran et al., 2022)	• Physical characteristics of the soil was not considered. • Focus on prediction of macro soil nutrient the general soil fertility status is not analyzed. • They consider only groundnut crop.	1DCNN DLM

CHAPTER THREE

3 RESEARCH METHODOLOGY

3.1 Overview

This chapter describes the methodology, tools, and procedures used to analyse the research challenges. It provides a brief overview of techniques that can be used in the design, implementation, and evaluation of the proposed architecture. It went into greater detail about sample selection, data collection, and analytic methodologies. Most importantly, this section includes an academic explanation for the method chosen to achieve the primary goal of the research.

3.2 Research Design

A plan for conducting research is known as a research design. Therefore, we have created our research design as shown in Figure 3.1 in order to address our research questions. There are several different research design methodologies, including participatory, quantitative, and qualitative. This study used a quantitative or experimental research design technique, which involved defining the study's goals and creating deep learning models to support the hypothesis. In data sampling, quantitative research design methods identify the relationship between independent factors and dependent variables(Kamiri & Mariga, 2021).

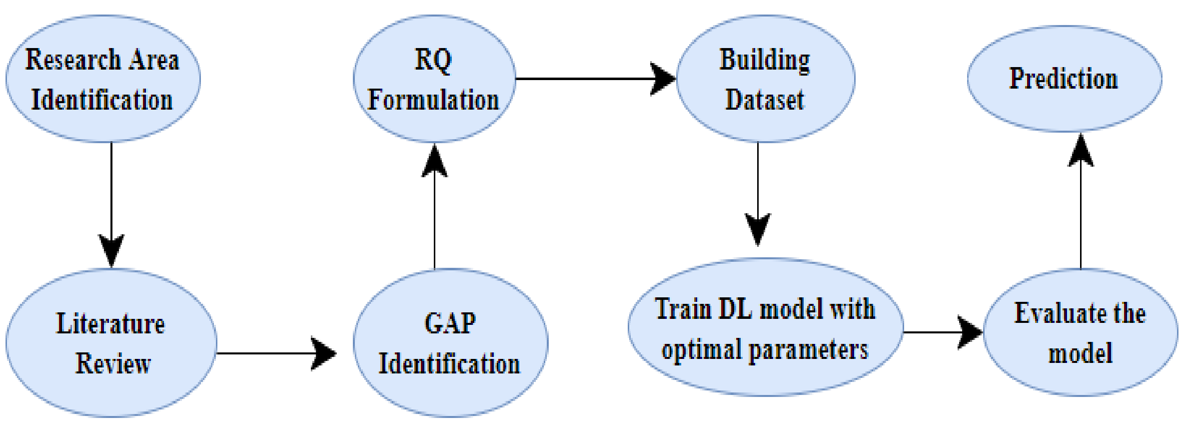


Figure 3-1 Research design procedure

3.3 Study Area

East Gojjam is one of the administrative zones of the ANRS located in Ethiopia's Blue Nile basin (Figure 3.1). It is currently divided into sixteen Weredas (Districts) and four urban administrations. The total land area of the zone is estimated to be 14,010 square kilometers (1.40 million hectares). In the highland (Dega) agroecology, the rainfall pattern is primarily unimodal, with average annual rainfall ranging from 900 to 1800 mm and a brief rainy season (Belg) February and March(Agegnehu et al., 2020). The reason behind selecting this area is that the East and West Gojjam zones provide more than 27 percent of the national yearly crop production, and the principal cereal crops cultivated in this area are teff, wheat, sorghum, maize, and barely crops (Hailu et al., 2015). It is an area with a very good climate and a large area of agricultural land as a result, if traditional soil fertility management and fertilizer application in this area are enhanced using new technology, our country's ability to meet the national requirement for cereal crop production will increase.

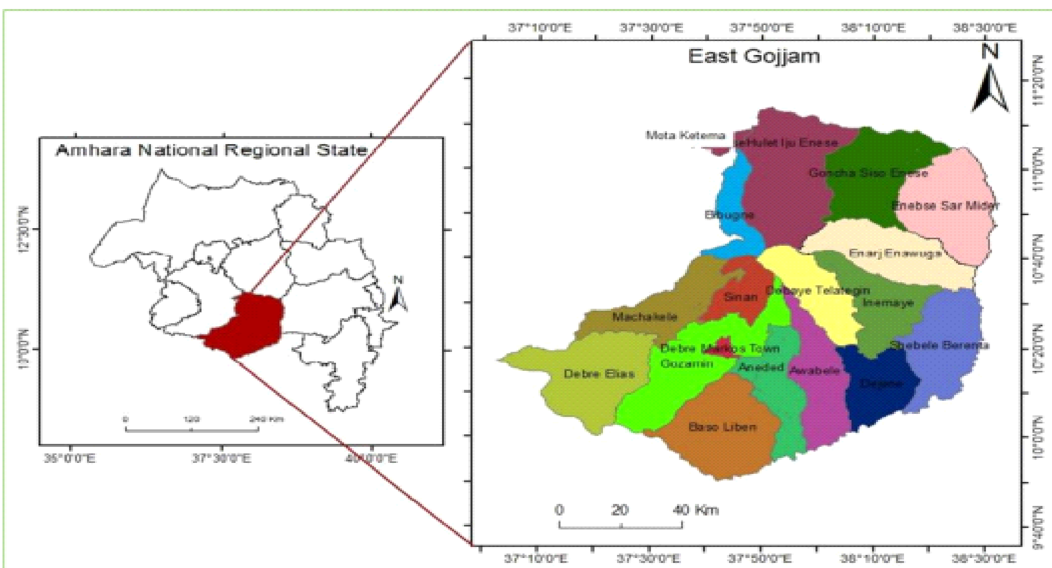


Figure 3-2 Location map of East Gojjam Zone (Gashe, 2017)

3.4 Data Collection

Soil samples for the preparation of dataset for this thesis were collected from Debre Markos soil laboratory center which aims to develop and spread best Soil conservation and fertility management practices by applying chemical analysis of soil samples gathered from farmers by assistant agricultural experts, specifically to assess the available nutrient status and soil reaction for maximizing agricultural productivity while minimizing negative environmental impacts.

And additional soil data have been gathered from Adet agricultural research center. The center aimed at conducting scientific research on major agricultural constraints in the western part of the Amhara Region in four administrative zones (West Gojjam, East Gojjam, Awi and South Gondar). As a result, the datasets contain a variety of attributes with respective values.

3.5 Dataset Labeling

The process of assigning tags or labels to raw data is known as data labeling. This study aimed at grading of given soil samples in a specific location and prediction of suitable fertilizer based on previously fed data. Experts from soil testing laboratories with experience in soil testing procedures, as well as one senior soil scientist, were involved in the process of dataset labeling that is assigning soil samples to appropriate fertility classes based on the soil fertility index and deciding on appropriate fertilizer based on the status of soil fertility. Mapping of independent features with dependent features of the dataset for this research explained as follows.

3.5.1 Fertility Class Labeling

The first model assists soil testing experts to understand the status of soil fertility by considering various soil nutrients both macro and micro nutrients and also other physical parameters. Based on these parameters content of various soil nutrient like (EC, K, pH, Mg, Zn, N, P, K, Fe, Cu, Mn, OC) are the feature variables, while the soil fertility class and fertilizer type are the target variables.

The labeling of the target variable is done based on the soil fertility index which can be calculated either by using the Parametric method or by identifying based on the given list of ranges. The soil fertility index of a study point was estimated as the product of the geometric mean of the score of each soil chemical and physical property in relation to the n^{th} root value of scores (Keshavarzi et al., 2020). The dataset has been labeled based on the following information's.

Table 3-1 The soil fertility indices and the corresponding fertility classes(Khaki et al., 2017b)

Fertility Class	**Index Value**
Very Low	0.0 - <0.25
Low	0.25 - <0.5
Moderate	0.5 - <0.75

High	0.75 - <0.90
Very High	0.90 - <1.0

Table 3-2 Five tier ratings of available primary nutrients

Organic Carbon (OC) (%)	**Nitrogen(N)**	**Phosphorus (P)**	**Potassium (K)**	**PH**	**Ratings**
<0.20	<50	<7	<100	<4.8	Very Low
0.21-0.40	51-100	7.1-14	101-150	4.8-5.5	Low
0.61-0.80	101-200	21.1-28	201-250	5.5-6.5	Moderately High
0.81-1.0	201-300	28.1-35	251-300	6.5-8.5	High
>1.0	>300	>35	>300	>8.5	Very High

The following table shows the classification of each soil nutrients into a particular range which has a great advantage in deciding the status of the soil health or fertility. matching crops to cultivate based on crop nutrient requirement and finally to decide what type of fertilizer to apply on the soil to make it more fertile.

Table 3-3 Soil nutrient fertility classification

Nutrient	**Extreme low**	**Low**	**Optimum**	**High**	**Extreme high**
pH	<4.8	4.8-5.5	5.5-6.5	6.5-8.5	>8.5
OM, g/kg	<5	5-15	15-30	>30	-
N, mg/kg	<50	50-100	100-200	>200	-
P, mg/kg	<5	5-15	15-80	>80	-
K, mg/kg	<50	50-100	100-200	>200	-
Ca, mg/kg	<200	200-1000	1000-2000	2000-3000	>3000
Mg, mg/kg	<80	80-150	150-300	300-500	>500
Fe, mg/kg	<5	5-10	10-20	20-50	>50
Mn, mg/kg	<2	2-5	5-20	20-50	>50

Zn, mg/kg	<0.5	0.5-1.0	1.0-5.0	5.0-10.0	>10.0
Mo, mg/kg		<0.05	0.05-0.20	0.20-0.30	>0.3

3.5.2 Fertilizer Type Labeling

Chemical fertilizer is primarily used in cereal production in Ethiopia. According to Ministry of Agriculture and Rural Development (MoARD) statistics, cereals account for 90 percent of the country's total chemical fertilizer application and the most used types of fertilizer are DAP, Urea and MOP(Rashid et al., 2014).

Table 3-4 nutrient ranges for fertilizer recommendation based on FAO guide line

Nitrogen(N) Kg/h	**Phosphorus(P) Kg/h**	**Potassium(K) Kg/h**	**Fertilizer Kg/h**
>150	>60	>150	Good NPK
<50	<30	>50	Urea and DAP
<50	>60	>150	Urea
>150	>60	<150	MOP
>150	<30	>150	DAP
>150	<30	<150	DAP and MOP

3.6 Data Preprocessing

The collected data was initially raw data, which include missed values, noise, inaccurate, inconsistent, and irrelevant data points. Data preprocessing is done to get the data ready for model creation. It converts raw data into a clean dataset. Stapes for data preprocessing includes data cleaning which is removal of unwanted symbols, columns and rows. The other technique is handling null or missing values by using different methods such as imputation. Finally handling categorical data and normalization is done to make the features in similar scale by using feature scaling. Preparation of the dataset for training of the model in general has been done under data preprocessing.

3.7 Prediction Model

This study is intended to predict soil fertility and suitable fertilizer using deep learning techniques to ensure soil stability for enhancing crop productivity using an already prepared, and labeled dataset. prediction models are built using deep learning algorithms. The goal is to make farming more efficient and productive with minimal impact on the environment. We propose an architectural model that evaluates the soil's fertility and productivity through context history. Therefore, a one-dimensional convolutional neural network is proposed for both soil fertility and fertilizer prediction which show the best accuracy. And other supervised deep learning algorithms such as Artificial neural network (ANN), Deep Neural Network (DNN) and machine learning algorithm such as Random Forest which are most used methodologies in the previous studies are also experimented for comparative analysis.

3.8 Predictive Model Evaluations

Model Evaluations used to measure how well the chosen model will work on unseen data. It is intended to evaluate the generalization accuracy of the model on unseen data(Pfert, 2017). Cross-validation is the technique used for evaluating model performance on a test dataset that is not seen by the model, which helps to evade overfitting in predictive models. Stratified k-fold cross validation is applied for this study.

There are numerous methods for assessing classification performance. Metrics such as accuracy, confusion matrix, and Precision-recall are popular metrics for classification problems. In this research, the performance of the designed system was evaluated using accuracy and loss metrics all are explained below.

Accuracy: Accuracy basically measures how far the classifier predicts correctly. The accuracy of a prediction can be calculated as the ratio of correct predictions to total predictions.

True positive (TP) is an outcome where the model correctly predicts positive classes. Whereas True Negative (TN) is a result obtained when the model correctly predicts negative classes. Similarly, False Negative (FP) is when the model incorrectly predicts positive classes and False Negative (FN) is the outcome obtained where the model incorrectly predicts negative classes(Li et al., 2021). And accuracy can be defined as:

$$\boldsymbol{Accuracy} = \frac{No\ of\ correct\ predictions}{Total\ No\ of\ predictions} \qquad \textit{Equation 3-1}$$

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: is the ratio of the entire number of true positive results to the number of positive results predicted by the classifier model, and it is how close a measured value is to each other(Boggs, Liam, 2017).

$$\boldsymbol{Precision} = \frac{TP}{TP+FP} \qquad \textit{Equation 3-2}$$

Recall: is the total number of true positive divided by the number of all relevant samples, or it is the ratio of true positive to a true positive plus false negative.

$$\boldsymbol{Recall} = \frac{TP}{TP+FN} \qquad \textit{Equation 3-3}$$

F1 Score: The harmonic means of the precision and recall values for a classification task is known as the F1-Score. When False Positives and False Negatives are more important than True Positives and True Negatives, F1 score is a better choice for measuring Accuracy. The following is the F1-Score formula:

$$\boldsymbol{F1 - score} = 2 * (\frac{precision*recall}{precision+recall}) \qquad \textit{Equation 3-4}$$

Categorical Cross-entropy loss*:* Loss/cost functions are used to optimize the model during training when working on a Machine Learning or Deep Learning problem. Almost always, the goal is to minimize the loss function. The better the model, the lower the loss. Categorical cross-entropy is a widely accepted loss function used in classification tasks. It is used in this research for computing the loss of multi-classification action which were encoded using on-hot approach.

The categorical cross entropy loss function calculates the loss of an example by computing the following sum:

$$CCE = -\sum_{i=0}^{n} Y_i * log(Y_i) \qquad \textit{Equation 3-5}$$

Confusion Matrix: is a technique that sums up the classification performance of models on some test datasets, with the result being two or more classes. It is used to classify data in a matrix format that is indexed in one dimension by the true class of an object and in the other dimension by the class that the classifier assigns. The confusion matrix is used with five class classifications in this study for soil fertility prediction which is classified as very high, high, moderate, low, and very low.

Table 3-5 Confusion matrix for five class classifications

		Predicted Values				
Actual Values		VH	H	M	VL	L
	VH	True VH	False H	False M	False VL	False L
	H	False VH	True H	False M	False VL	False L
	M	False VH	False H	True M	False VL	False L
	VL	False VH	False H	False M	True VL	False L
	L	False VH	False H	False M	False VL	True L

3.9 Development Tools

Various development tools and packages have been used in this study in order to implement the proposed system. The development tools are summarized and justified in this section[17].

3.9.1 Design Tools

Edraw-Max is used in this study for the designing purpose. It's a quick and easy way to make professional-looking flowcharts, network diagrams, and UML diagrams, among other tools

3.9.2 Hardware Tools

Hard disk is used as storage for huge datasets, GPUs will be used to expedite computation and training, and RAM used cooperatively with GPU to accelerate the training process.

3.9.3 Software Tools

Anaconda: Anaconda is a distribution of the python and R programming languages for data science, machine learning, predictive analytics etc. It comes with more than 1500 packages as well as virtual environment manager and conda package[18].

[17] https://www.edrawsoft.com/guide/edrawmax/
[18] https://www.anaconda.com/

Jupyter Notebook: Exists to develop open-source software, open-standards, and services that allows you to create and share documents that contain live code, equations, visualizations and narrative text[19].

Python: It is used to conduct and demonstrate our testing, and it will include some of the drivers needed to configure and package any system that will be installed on the computer.

Pandas: Is an open-source library that offers easy access and high-performance data structures and analysis tools for python programming language.

TensorFlow Is an open-source platform for machine learning. It has a flexible, complete system6i6i of libraries, community resources and tools and that lets researchers and developers to easily build and deploy ML powered applications.

Keras: Keras is a minimalist Python library for deep learning that can run on top of Theano or TensorFlow. It was developed to make implementing deep learning models as fast and easy as possible for research and development.

Mendeley Desktop: It's a handy reference manager that also functions as a scholarly social network for finding related works.

[19] https://jupyter.org/

CHAPTER FOUR

4 PROPOSED SOLUTION FOR SOIL FERTILITY PREDICTION

4.1 Overview

This chapter describes the model, algorithms used in the implementation of soil fertility and fertilizer prediction models, basic process, and steps used to develop this model. Additionally, the result obtained by each algorithm of the model was given under each section. The model is trained using historical data sets where the past experience is used to represent the outcome. The general process flow is shown in the following figure.

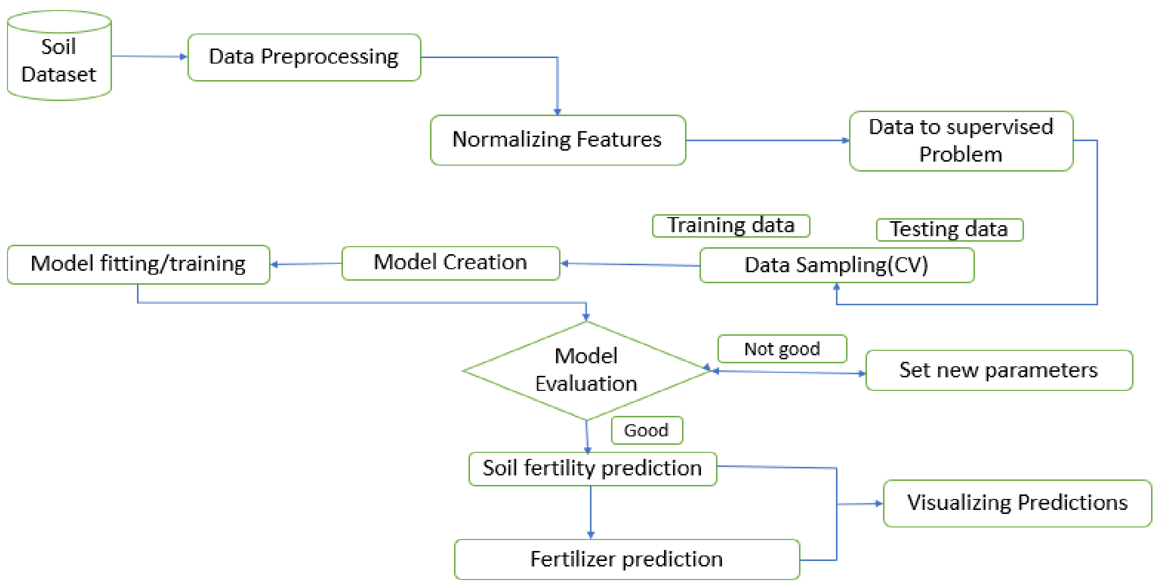


Figure 4-1 process workflow for soil fertility prediction and fertilizer recommendation

The above figure shows the algorithm flow of the research. It has mainly five steps: initialization, data pre-processing, model implementation and finally model training and model evaluation.

4.2 The Proposed System Architecture

This research aimed to build prediction model for soil fertility prediction and fertilizer recommendation using deep learning approaches. It predicts which soil samples are with very high, high, moderate, low and very low fertility rate for cereal crops production based on soil physical and chemical characteristics dataset which was collected from Debre Markos soil laboratory center and Adet agricultural research center as stated in details in chapter three. And for fertilizer prediction similar soil dataset has been used to recommend appropriate type of fertilizer such as DAP, MOP, Urea, Urea and DAP, DAP and MOP based on macronutrient values of the given soil sample.

This study's proposed architecture specifies that initially the collected data was raw data. since it is a raw data its prone to inconsistency, noisy, variety, duplication and so on. Because of this we need to undergoes data preprocessing stage such as data cleaning, handling missing values, handling categorical values to transform raw data into meaningful information and make it suitable for the deep learning algorithm. Following preprocessing of the dataset, Artificial Neural Network (ANN), Deep Neural Network (DNN) and one dimensional convolutional neural network (1D-CNN) are used to build the predictive model. and their predictive performance is assessed by using various performance metrics to determine the best performing model. Also, other machine learning algorithms such as Random Forest algorithm were used to train the dataset for comparative analysis. And based on the experimental result 1D-CNN shows the best performance for both soil fertility prediction and fertilizer recommendation.

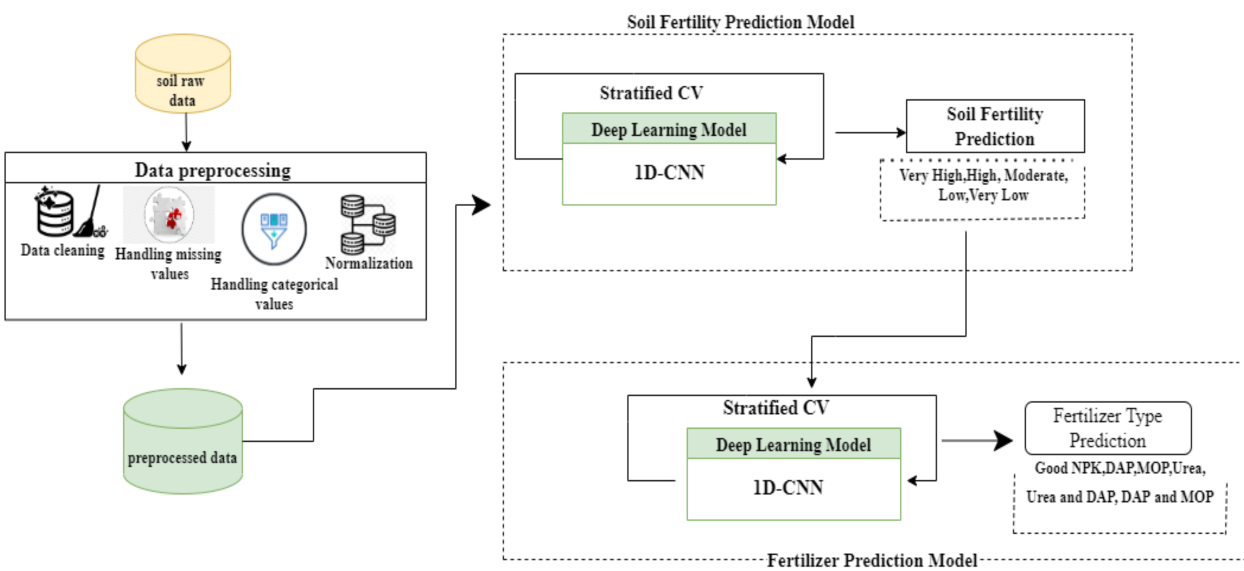


Figure 4-2 Proposed Soil fertility prediction Architecture

4.3 Proposed Data Preprocessing

Data preprocessing is a crucial stage in machine learning since the quality of the data and the information that can be gained from it have a direct impact on our model's capacity to learn. The preprocessing techniques were applied to the already prepared dataset file and converted into a common separated value (CSV) format with .csv.

4.3.1 Missing Values

The first problem that we have to deal with is missing values. We cannot erase a row with missing values because it might hold necessary information in it. Missing values in a dataset handled in a variety of ways, including data imputation, which replaces missing values with their mean, mode, or standard deviation, predicting missing values from an available dataset. We have used mean method to fill the missing values. Mean imputation is a method that replaces a missing value on a variable with the mean of the available cases. This method maintains the sample size and is simple to use, and it is the most commonly used method in most studies to deal with numerical datasets for missing values. Imputer, which is a class in sklearn.preprocessing is used to do this task.

4.3.1 Encoding Categorical Data

The data set contains some categorical data which are not numbers and before we run a model, we must convert our data into a model-understandable numerical format. This task is performed by label encoder, which is also a class in sklearn. preprocessing. It converts categorical information into numerical variables.

4.3.1 Normalization

All features are in different range so feature scaling can assist us in having features with the same scale that contribute equally to the result. The min-max scaler, also known as the normalization method, is used in this study among other feature scaling techniques. Where the minimum value gets transformed into 0 and maximum value gets transformed in to 1. The formula of minmax normalization is as follows:

$$x_n = \frac{x_{i-} \quad x_{min}}{x_{max-}x_{min}} \qquad \textit{Equation 4-1}$$

Where, Xn is normalized value of a variable X

Xi is current value of variable X

X_{Min} minimum value of variable X and,

X_{Max} is the maximum value of variable X

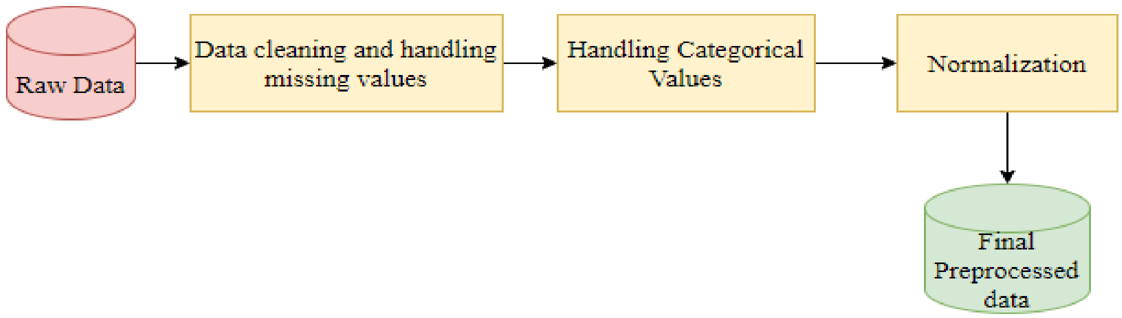


Figure 4-3 steps of data preprocessing

4.4 Prediction Models Building

The goal of this research is to create a prediction model for soil fertility prediction and fertilizer recommendation system to identify fertility status of a given soil sample and also to determine appropriate fertilizer to use based on soil physical and chemical characteristics using deep learning algorithms. So that we had built deep learning models using deep learning algorithms such as ANN, DNN and 1D-CNN algorithm. These models are chosen for this study for the soil fertility and fertilizer prediction which are multiclass classification, and their performance is evaluated using stratified 10-fold cv, and others performance metrics are used to identify the best performing model.

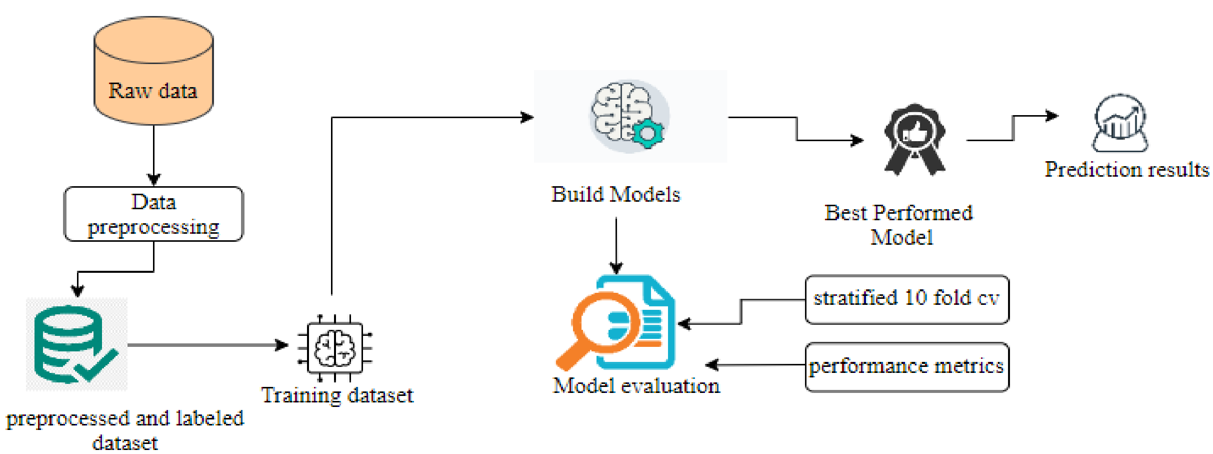


Figure 4-4 Model Building Flow Diagram

4.4.1 One Dimensional Convolutional Neural Network(1D-CNN)

1Dimensional CNN and 2Ddimensional CNN has almost similar principles their main difference is that 2D CNN is used to process matrix information where as in 1D CNN the input will be one dimensional array or it processes one dimension vector information(Mitiche et al., 2020). And 1D-CNN can be expressed as:

$$X_{i=conv1d}(w_{i-1}, x_{i-1}) + b_i \qquad \textit{Equation 4-2}$$

where X_i is the output of x_{i-1} and b_i is the bias

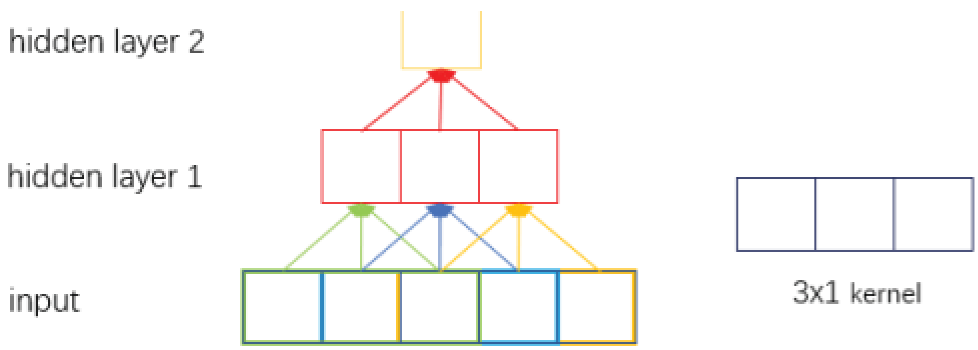


Figure 4-5 1d convolution work process diagram

1D CNN has a great advantage since it has a compact architecture configuration which makes it better than the complex deep architectures and it also had the capability to work without any handcrafted feature selection and this makes it efficient at training of the classifier. 1D CNN based deep learning model were built to train the prepared dataset. Experimental results demonstrated that the proposed model shows better performance and capability in prediction of soil fertility.

we use Batch normalization to make training convergence faster and this can improve performance result. It is a method for increasing neural network performance and accuracy while avoiding gradient vanishing and over-fitting. By subtracting the batch mean and dividing by the batch standard deviation, BN normalizes the previous activation output. As a result, each network layer will learn independently of the others. We can use much higher learning rates with BN and worry less about initialization. It also acts as a regularizer, eliminating the need for dropout in some cases and increasing the stability of a neural network (Meliboev et al., 2020).

$$BN(X_K) = \alpha\left(\frac{X_{K-\mu B}}{\sqrt{\sigma B2 + \varepsilon}}\right) + \beta \qquad \textit{Equation 4-3}$$

where, ε is a random noise for (stability), μB represents the mini-batch mean, α is the mini-batch variance, σ represents a scale parameter, and β is the shifting parameter (offset) for the layer.

The proposed 1D CNN architecture is shown in fig 4. The model consists of 1D convolutional layers. And the 1D convolutional layer is followed by one max-pooling layer which is like that of convolutional layer and adopted in our network for down sampled output to reduce the number of parameters while retaining the key features in order to accelerate the next stage and this layer provides regularization this makes it similar to dropout layer. This layer also intended to solve the issue of overfitting.

Also, we add dropout technique before the fully connected layer with rate of 0.5. and before values are fed into the fully connected layers. The dropout method is frequently used to address the problem of over-fitting during the training step. Network input data is transmitted from the input layer to the output layer via various hidden layers.

At the end we add fully-connected (FC) layer to perform the classification. The output of the network's earlier layers is reshaped (flattened) into a single vector. Each one represents the possibility that a specific feature is a class label. To determine the correct label, inputs from the feature map are used, and weights are applied. FC's output layer provides the final probabilities for each label.

In addition, with the exception of the output node, we used a non-linear activation function called Rectified Linear Unit (ReLu). In order to avoid overfitting, we have used softmax function which computes the probability for each class. Since our problem is multiclass classification in a multi-class problem, Softmax assigns decimal probabilities to each class. The softmax equation can be expressed as follows:

$$softmax(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)} \qquad \textit{Equation 4-4}$$

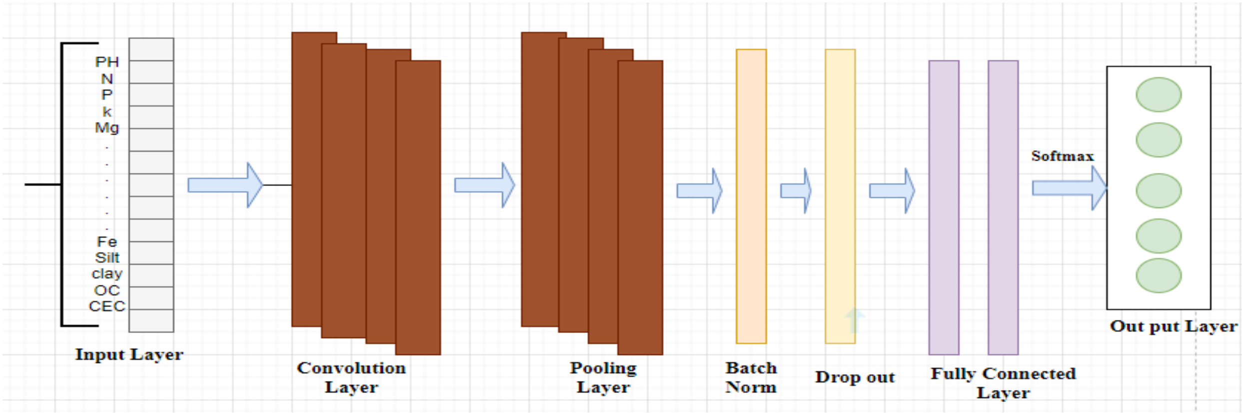


Figure 4-6 1D-CNN Model [20]

4.4.2 Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are pliable mathematical structures capable of recognizing complex non-linear relationships between input and output data sets (Ayoubi et al., 2011). Given that all of the input variables in our dataset are numerical and have the same scale thanks to normalization, it makes it an excellent problem for neural network training. It is a multi-class classification problem, which means that there are five different classes of soil fertility and six different classes of fertilizer types that can be predicted rather than just two. Given that the multi class values call for specialized handling, this is a crucial classification problem for neural network training.

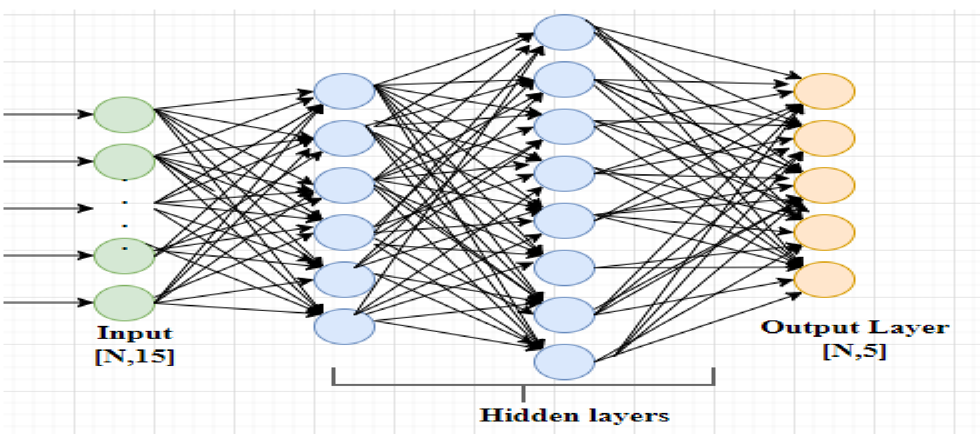


Figure 4-7 ANN model architecture[21]

[20] 2 https://www.researchgate.net/figure/Standard-CNN-on- classification_fig1_333752473
[21] https://www.researchgate.net/figure/The-artificial-neural-network-ANN-designed-for-the-classification

4.4.3 Deep Neural Network (DNN)

Deep neural networks (DNNs) are enhanced versions of traditional ANNs that have multiple layers. DNN models have recently gained popularity due to their superior ability to learn not only nonlinear input-output mappings but also the underlying structure of input data vectors. contains more than three layers It inherently integrates feature extraction and classification into learning and enables decision making.

4.4.4 Random Forest (RF)

A soil fertility prediction and fertilizer recommendation is also experimented by using Random Forest model, it is an ensemble of classifiers that is typically made up of decision trees used for both classification and regression problems. As its name suggests, this algorithm generates a forest of various decision trees at random. The accuracy increases with the number of trees in the forest. The way RF operates is that it first chooses K data points at random from the training dataset. It then creates decision trees linked to the chosen data points. It then repeats steps 1 and 2 after selecting the desired number N for the decision trees it wants to build. The new data instances are then assigned to the category with the most votes after finding the predictions of each decision tree majority votes (Kiranyaz et al., 2021).

4.5 Optimal Hyperparameter

The hyperparameters function as knobs that can be adjusted during model training which would lead to improved accuracy and reduce the risk of over-fitting. There are two kinds of hyperparameters. The first defines the network structure, including the number of filters, kernel size and the hidden layer between input and output. The other indicates network training parameters such as the number of epochs, batch size, and learning rate(Lateef & Abbas, 2022). Selecting the best collection of hyperparameters for a learning algorithm is known as hyperparameter tuning. A model input called a hyperparameter has its value predetermined before the learning process even starts. Hyperparameter tuning is essential for machine learning algorithms to work. And for the essential parameters for the model performance such as epoch, batch size, number of neuron and batch size has been tuned for 1D-CNN model.

4.6 Model Training

The proposed learning system comprises three essential components: model development, model compilation, and model training. First, create a model that fits the proposed goal, and then compile it by changing the hyperparameters it requires. After detailing the first two phases, the model will fit with the dataset to begin learning.

4.7 Model Evaluation

Model evaluation is used to estimate the generalization accuracy of the predictive model on the out-of-sample dataset. Soil fertility prediction model is trained with the entire dataset and its performance is assessed using stratified 10-fold cross-validation with others performance metrics which are appropriate to the model such as confusion matrix, accuracy, precision, recall, and f1-score which were discussed in detail in chapter three section 3.7. The RMSE and MAE are also used as evaluation metrics to measure the efficiency of the soil organic carbon prediction model.

CHAPTER FIVE

5 IMPLEMENTATION AND EXPERIMENTATION

5.1 Overview

In this chapter, the implementation of the proposed approach and the tools needed for experimentation to achieve the proposed soil fertility prediction and fertilizer recommendation is discussed in detail. For implementing the proposed approach, the environment was set as described in chapter three. This setup includes: python programming language, keras framework backend as tensorflow, pandas, Scikitilearn and etc. Significant code segments were presented in this section, while important detail codes were depicted in the appendix section.

5.2 Working Environment

Several development tools and packages has been used to implement the proposed approach. A python programming language is used for implementing and experimenting with each proposed solution starting from preprocessing phase to model building. It is also used to evaluate the implemented models to identify the best model.

5.2.1 Desktop Computers with Hardware Utilities

- Hard Disk: configured terabyte added two-tetra bytes of storage.
- SSD: configured SSD is 128GB.
- Processor: Inter(R) Core (TM) i7-7700HQ CPU @ 2.80GHZ (8 CPUs), ~2.8GHz
- Graphical purpose processing Unit: for high-speed time computation during training. NVIDIA GeForce GTX 1050 super installed.
- Installed Memory: Physical Memory (RAM) 16.00 GB.

5.2.2 Desktop Computers with Software Configuration

- OS: Microsoft Windows 10 Pro, x64-based PC, Version 10.0.19041 Build 1904.
- System Model: - HP Pavilion Power Laptop 15-cb0xx.
- GPU optimizer: the latest version of the NVIDIA GPU optimizer and scheduler.
- CUDA toolkit.

5.3 Setup Environments

5.3.1 Application Software

Anaconda: The proposed approach is implemented using an anaconda application version 4.8 with 64-bit support which helps us to launch developed applications, and can easily manage conda packages, environments, and channels without using commands in command-line.

5.3.1 Integrated Development Environment and Editors

Jupyter Notebook: is a popular and highly recommended IDE for academics who need to update Python code, display diagrams, training and testing charts, and track processing progress. It is open-source web application that assist us for creating and sharing documents, also for visualizations, cleaning, transformation, modeling, machine learning, and more. The proposed solution is implemented using the Jupyter notebook version 6.1.5, which is installed via anaconda installation.

5.3.1 Programming Language and Module Libraries

Python: Python is object-oriented, high-level, and flexible programming language with dynamic semantics. It is used for teaching machines to learn, and for analyzing an extensive amount of data.

Kera's: with version 2.3.0, which used to load the pre-trained model, used to apply convolution, pooling, time series generator, and other deep learning processes.

TensorFlow: with version 2.3.0 and above is used to connect the retrained model and feature fusion model graphs Tensor board with version 2.3.0 and above to visualize training progress in deep learning.

5.4 Training Procedure

- The training procedure in this proposed research
- Collect massive dataset from soil and agricultural laboratories
- Pre-processing dataset
- Scaling and transforming the data by train and test split
- Building the model
- Compile the model
- Train the model

5.5 Dataset Description

The dataset used for this study is collected from Debre Markos soil laboratory center located in Debre Markos, Ethiopia and also from Adet agricultural research center located in Bahirdar, Ethiopia. The report helps the researcher for a detailed understanding of each parameter and dataset labeling. Therefore, East Gojjam is used as the study area for this thesis. Data was collected by targeting to identify the status of the soil fertility class and suitable type of fertilizer. Soil physical and chemical characteristics included in this dataset were starting from 2011-2013 E.C, and contain 12,509 rows with 17 columns including two target variables one is the soil fertility class and the other is fertilizer prediction. The columns of the dataset are PH, Mg, N, P, K, Zn, Fe, Cu, Mn, Sand, Silt, Clay, CaCO3, CEC, OC, Fertility classes and Fertilizer types.

Table 5-1 Dataset Features Description

No	***Symbol***	***Feature full name***	***Type***	***Description***
1	PH	Soil Reaction Power	Numeric	It is used to measure the amount of alkalinity and acidity of the soil.
2	Mg	Magnesium	Numeric	In plant tissue, magnesium is the central core of the chlorophyll molecule. And it is vital for plant growth
3	N	Nitrogen	Numeric	This essential nutrient can even be found in the roots, where proteins and enzymes regulate water and nutrient uptake.
4	P	Phosphorous	Numeric	t is the amount of phosphorus readily available for nutrient absorption by the plant roots. Measured in parts per million (ppm).
5	K	Potassium	Numeric	Essential element that increases root growth and resistance to drought.

6	Zn	Zinc	Numeric	Is recognized as an essential component in several dehydrogenases, proteinases, and peptidases.
7	Fe	Iron	Numeric	It is required for photosynthesis and chlorophyll and it is measured in ppm.
8	Cu	Copper	Numeric	It facilitates respiration and photosynthesis and is important for plant metabolism.
9	Mn	Manganese	Numeric	sustains metabolic roles within different plant cell compartments.
10	Sand	Sand	Numeric	Amount of sand particles in the soil Measured in percent.
11	Silt	Silt	Numeric	Amount of silt particles in the soil Measured in percent.
12	Clay	Clay	Numeric	Amount of clay particles in the soil Measured in percent.
13	CaCO3	Calcium carbonate	Numeric	It is the carbonate that defines the soil.
14	CEC	Cation Exchange Capacity	Numeric	It measures the ability of soil to grip positively charged ions. Measured in cent moles of charge per kilogram of the exchanger (cmol (+)/kg).
15	OC	Organic Carbon	Numeric	which is critical nutrient element it determines the composition of other soil nutrients.
16	Fertility classes	Soil fertility classes	Nominal	Target classes containing VH, H, M, VL and L based on soil fertility index.

17	Fertilizer types	Fertilizer type to apply	Nominal	Target classes containing Good NPK, DAP, MOP, Urea, Urea and DAP, and DAP and MOP

5.6 Implementation Initialization

Before starting all the necessary packages and libraries like NumPy, Pandas, Matplotlib, Sckit-Learn, TensorFlow, keras are installed and imported. Each imported libraries have specific applications for example we imported Pandas NumPy and pandas to Jupiter notebook in order to feed our csv data into pandas Dataframe through the function called read_csv. Sckit-learn is built upon the SciPy (scientific python) which has to be installed before installing Sckit-learn. This library has pre-built functionality in sklearn.preprocessing. Some of the imported libraries are shown in the figure below.

```
from keras.models import Sequential
import pandas as pd
from keras.layers import Dense, Conv1D, Flatten, MaxPooling1D,BatchNormalization,Dropout
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from numpy import unique
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn import preprocessing
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import KFold
```

Figure 5-1 implementation of importing packages and libraries

5.7 Preprocessing Implementations

Preprocessing implementation includes handling missing values, handling categorical values, and normalization. All of them are implemented using a python programming language as discussed in chapters three and four to make it suitable for the deep learning algorithm to build predictive learning models. The prepared dataset is converted into a common separated value (CSV) format with a .csv extension as the CSV python module gives the python programmer the ability to parse the file with CSV format for further preprocessing using python code in machine learning.

```
import pandas as pd
data=pd.read_csv('Soil_Ferrtlity.csv')#Loading dataset
df.head()
```

Figure 5-2 Implementation of loading dataset using panda's library

5.7.1 Handling missing Values Implementation

Since we have a raw data mostly, they are incomplete: they lack attribute values or attributes of interest, noisy: they contain errors and outliers, inconsistent so we need to transform this raw data into machine understandable format. To deal with missing values, the imputation method is used, in which missing values in the dataset are replaced with the mean or median values of the dataset's non-missing data. This study employs mean imputation.

```
from sklearn.impute import SimpleImputer
imputer= SimpleImputer(missing_values=np.nan,strategy='mean',verbose=0)
imputer=imputer.fit(data.iloc[:,1:15])
data.iloc[:,1:15]=imputer.transform(data.iloc[:,1:15])
```

Figure 5-3 Implementation of Handling missing values

5.7.2 Handling Categorical Values Implementation

After we check out for the missing values which is none and in the next step is to change the categorical values. In our dataset we have a categorical feature such as fertility class. Label encoding provides very efficient tool for encoding labels of categorical data to numeric values. Then the numerical values are converted to float by using astype(float) method.

```python
le = preprocessing.LabelEncoder()
df['Fertility_Class'] = le.fit_transform(df['Fertility_Class'])
label_encoder = LabelEncoder()
Y = label_encoder.fit_transform(Y)
print(unique(Y))
```

```
[0 1 2 3 4]
```

```python
le = preprocessing.LabelEncoder()
df['Fertilizer'] = le.fit_transform(df['Fertilizer'])
Y=df['Fertilizer'].unique()
print(unique(Y))
```

```
[0 1 2 3 4 5]
```

Figure 5-4 encoding of nominal values

5.7.3 Scaling and Transforming the Data

Before putting the data into the suggested model, it goes through many rounds of pre-processing. Minmax scaling is considered as a way to normalize data in which we scale each feature to a fixed range. It is probably the most famous scaling algorithm. In this scaling technique values will end up ranging from 0 to 1 by shifting and rescaling. This is done by subtracting the minimum value and dividing by the maximum value minus the minimum values.

```python
values1= data.values
scaler = MinMaxScaler(feature_range=(0,1))
# fit and transform in one step
normalized = scaler.fit_transform(values1)
# inverse transform
inverse = scaler.inverse_transform(normalized)
```

Figure 5-5 implementation of normalizing features

5.7.4 Specify Features and Target

After dealing with the data preprocessing steps such as handling missing values, and handling categorical values the next step is to divide the dataset into features, and target which means independent, and dependent features respectively in order to build the proposed models.

```python
#Dividing Dataset into independent&dependent for soil fertility
X = df.drop(['Fertility_Class'],axis=1)
y = df[['Fertility_Class']]
```

```python
X=df[['N','P','K']]
Y=df[['Fertilizer']]
```

Figure 5-6 Dividing dataset into independent and target feature

5.8 Prediction Models Implementation

Various deep learning and machine learning prediction models are experimented and evaluated with the same prepared soil dataset. 1D-CNN shows better performance in soil fertility classification and other models which are frequently used by previous researchers are selected to compare the result with the proposed model.

In the proposed 1D-CNN model We have define a model which has two 1D CNN layers with 64 filter and 3 kernel size, other several features including an optimizer, a loss, and metrics, are used to form the sequential model. Dropout and batch normalization are also used for regularization and overcome overfitting problem. All the available hidden layer except for the output layer uses a rectifier (ReLu) activation function the output layer must create 5 output values, one for each class. The output value with the largest value will be taken as the class predicted by the model.

```
model = Sequential()
model.add(Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(15,1)))
model.add(BatchNormalization())
model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
model.add(Dropout(0.5))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())
model.add(Dense(64, activation='relu'))
model.add(Dense(5, activation='softmax'))
model.compile(loss='sparse_categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
model.summary()
```

Figure 5-7 Python code to create 1D-CNN model

5.8.1 Compile the Model

Now the model has defined, we compiled it in this step. Model compilation uses efficient numerical libraries such as Theano or TensorFlow. The backend has automatically chosen the best way to represent the network for training and making predictions to run on our hardware, such as CPU or GPU or even distributed. In the process of compiling, additional properties have been specified which are required when training the network. In this model, we used crossentropy as the loss argument. This loss is shown for categorical classification problems and is defined in Keras as "categorical_crossentropy". We defined the optimizer which is called "adam". It is a popular version of gradient descent because it automatically tunes itself and gives good results in a wide range of problems. Hence it is a classification problem; we collected and reported the classification accuracy, defined via the metrics argument.

```
model.compile(loss = 'sparse_categorical_crossentropy', optimizer = "adam", metrics = ['accuracy'])
model.summary()
```

Figure 5-8 Compiling the model

5.8.2 Model Training

The next step after the model has been compiled is effective computation. On a set of data, the model was run. By using the fit() function on the model, we can train and fit our model using the loaded data. Each epoch in the training process is divided into batches. Epoch: One runs through every row in the training dataset. Batch: A sample or samples that the model takes into account during an epoch before updating its weights. Depending on the chosen batch size, an epoch may consist of one or more batches, and the model is applicable across a wide range of epochs. The training procedure cycled through the dataset, known as epochs, a predetermined number of times. The batch size, also known as the number of dataset rows taken into account prior to the updating of the model weights within each epoch, is also set using the batch size argument. In this model, we used a relative default batch size of 64 and ran for a certain number of epochs (60). Trial and error can be used experimentally to select these configurations. We gave the model enough training to enable it to develop an accurate mapping of input data rows to output classification. Although the model will always contain some error, for a given model configuration, the error would eventually level off. We refer to this as model convergence.

```
history=model.fit(xtrain, ytrain, validation_data=(xtest, ytest), batch_size=64,epochs=60)
```

Figure 5-9 Model Training

Other deep learning models which are commonly used for classification problems such as ANN, DNN and machine learning algorithms like Random Forest were also implemented and the models are evaluated using stratified 10-fold cross-validation along with performance metrics. The following code snippets shows the implementation of ANN, DNN and RF for the prediction of soil fertility respectively.

ANN: The function to build a foundational artificial neural network for the soil fertility classification problem is provided below. It builds a straightforward fully connected network with one hidden layer, containing 28 neurons. The rectifier activation function used by the hidden layer is a good practice. For our soil dataset, we used a one-hot encoding, so the output layer needs to generate 5 output values one for each class. The output value with the highest value will be considered to represent the class the model predicted. In the output layer, we employ a "softmax" activation function. The network then employs the effective Adam gradient descent optimization algorithm, also known as "categorical crossentropy" in Keras, which has a logarithmic loss function.

```
model= Sequential()
model.add(Dense(64,input_shape=(15,), activation="relu"))
model.add(Dense(28, activation="relu"))
model.add(Dense(5, activation='softmax'))
model.compile(optimizer="adam", loss="sparse_categorical_crossentropy", metrics=["accuracy"])
history=model.fit(X[train], Y[train], epochs=60, batch_size=64, verbose=0)
```

Figure 5-10 implementation of ANN for soil fertility prediction

DNN: Below is the function to build a basic deep neural network for the soil fertility classification problem. It builds a straightforward fully connected network with three hidden layers, each containing 28 neurons. The rectifier activation function used by the hidden layers. We used a one-hot encoding, so the output layer needs to generate 5 output values one for each class. The output value with the highest value will be considered to represent the class the model predicted. In the output layer, we employ a "softmax" activation function. The network then employs the effective Adam gradient descent optimization algorithm, also known as "categorical crossentropy" in Keras, which has a logarithmic loss function.

```
model = Sequential()
model.add(Dense(68, input_dim=15, activation='relu', name='layer_1'))
model.add(Dense(28, activation='relu', name='layer_2'))
model.add(Dense(28, activation='relu', name='layer_3'))
model.add(Dense(28, activation='relu', name='layer_4'))
model.add(Dense(5, activation='softmax', name='output_layer'))
model.compile(optimizer="adam", loss="sparse_categorical_crossentropy", metrics=["accuracy"])
model.summary()
```

Figure 5-11 implementation of DNN for soil fertility prediction

RF: Random Forest is a flexible and easy-to-use supervised machine learning algorithm that fits several decision tree classifiers on several sub-samples of the dataset and uses averaging to improve predictive accuracy and resist over-fitting. It is implemented with the RandomForestClassifier () method from the ensemble package's sklearn, which is used to fit and train the classifier.

```
#Import Random Forest Model
from sklearn.ensemble import RandomForestClassifier
clf=RandomForestClassifier(n_estimators=100,max_features='auto',
                                class_weight='balanced',
                                max_depth=10,criterion='entropy',
                                min_samples_split=2,random_state=42,n_jobs=-1)

clf.fit(X_train,y_train)
y_pred=clf.predict(X_test)
```

Figure 5-12 implementation of random forest

5.9 Implementation of Model Evaluation

Now we have trained our models on the entire dataset and we can evaluate the performance of the model on the same dataset by using different performance metrics the implementation of some of them are shown in the following code snippet.

5.9.1 Stratified K-fold cross validation

We can begin by defining the model evaluation procedure. In this case, we set the number of folds to 10 (a good starting point) and shuffle the data before partitioning it.

```
kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=7)
cvscores = []
```

Figure 5-12 Implementation of kfold cv

Additionally, we evaluated our model's performance using evaluation metrics such as the confusion matrix, precision, recall, and F1-score.

CHAPTER SIX

6 RESULTS, EVALUATION, AND DISCUSSIONS

This chapter describes the experiments carried out to evaluate four different prediction models, namely 1D-CNN, ANN, DNN and RF. First, the overview of used dataset and its class distribution is discussed. Second, models performance, and its evaluation results, different performance metric results for the selected models are discussed. Finally, the comparison of the selected models for soil fertility classifications and suitable fertilizer prediction is provided. A detailed comparison of this research with the previous studies is also provided to assure as this study fills the gaps identified in the literature review, answers the research question, and achieves stated research specific objectives.

6.1 Dataset Class Distributions Result

The dataset used for this study contains 12,509 rows with 17 columns including two target variables for soil fertility prediction and the other for suitable fertilizer prediction after data preprocessing. Soil fertility prediction for cereal crops is a five-class classification problem and suitable fertilizer prediction also has six classes. Since a supervised machine learning algorithm is used in this study, dataset labeling is a very important task that should be carried out. The details of dataset class distribution shown in the figure below.

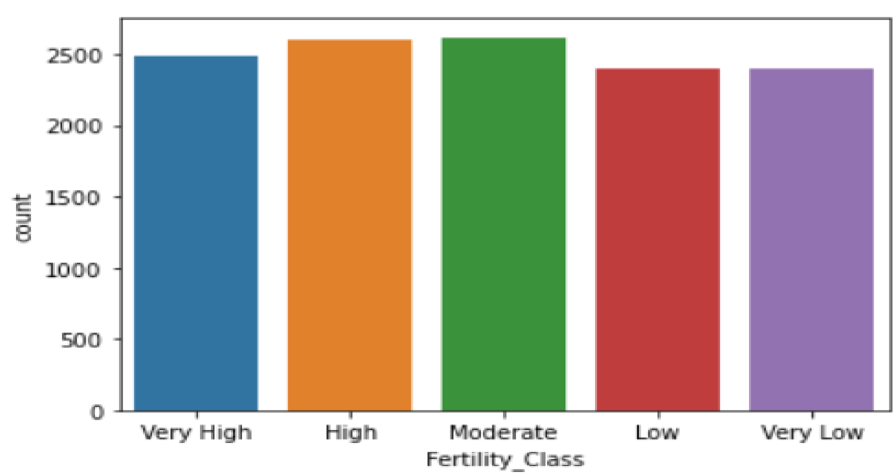


Figure 6-1 Dataset target classes distribution

When we come to the machine learning approach, training the model with class imbalance dataset distributions makes the model to be biased toward the majority class that causes prediction problems toward the minority class, and reduces the general accuracy of the model. And as shown in the above figure Fig 6.1 when we were labeling the dataset, we consider about the class imbalance and as it is clearly shown the instances are equally distributed among the target classes. There is a slight difference in very low and low fertility classes and this is due to more of the soil samples taken in the study area were having a moderate and high fertility status.

6.2 Experimental Setup

Jupyter notebook environment was used to conduct the experiment. Python and the Keras package were utilized as the implementation and simulation tools. Experiments were conducted on the preprocessed dataset by applying various models for the classification of soil fertility and to predict the level of soil organic carbon.

Various performance metrics for the evaluation of the model have been applied like accuracy, precision, recall, f1 score and confusion matrix for the classification problem. The findings and comparative analysis of the models discussed in detail in the following sections.

6.3 Model Learning Curves

Learning curve are widely used as diagnostic tool in machine learning for algorithms such as deep learning that learn incrementally. During training time, we evaluate model performance on both the training and test dataset and we plot this performance for each training step (i.e., each epoch of deep learning model). 1D-CNN model for soil fertility prediction scored about 98.82% accuracy on classifying the given soil samples and fertilizer prediction scored about 97.65% accuracy. The model was compared with some other previously proposed models, and worked better than those. 60 epochs were completed during the training period to check the model's accuracy and loss. Reviewing learning curves of models during training can be used to diagnose problems with learning such as an underfit or overfit model.

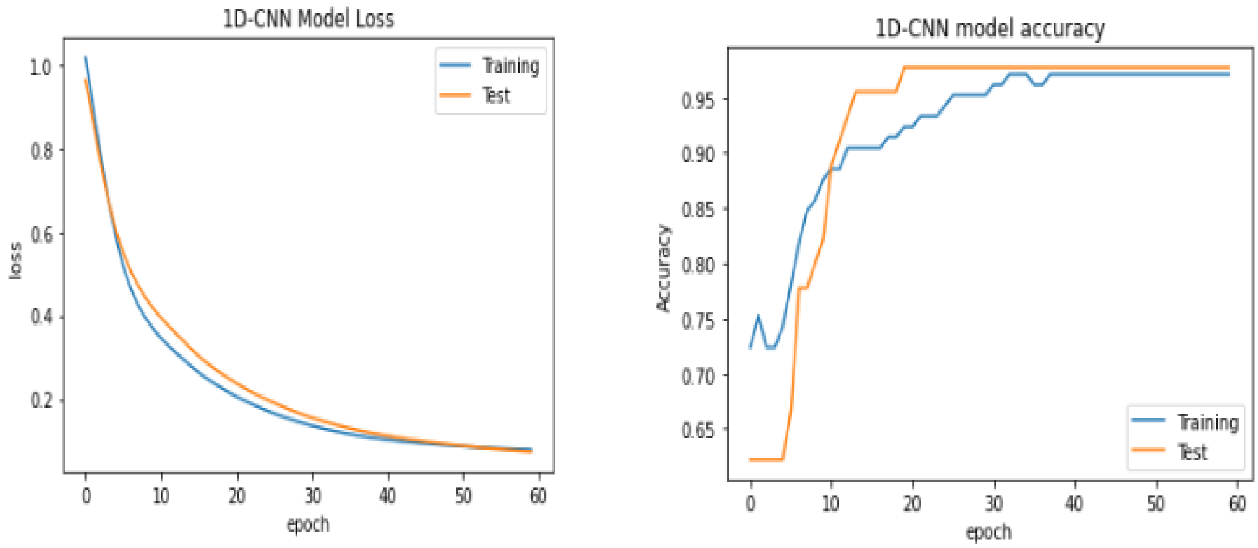


Figure 6-2 Model learning curves

6.4 Models Performance and Evaluation Result

6.4.1 Soil Fertility Prediction Model Result

Here the performance and evaluation result of the 1D-CNN, ANN, DNN and RF based models on soil fertility prediction is discussed with the confusion matrix and with others performance metrics of the models under stratified 10-fold cross-validation.

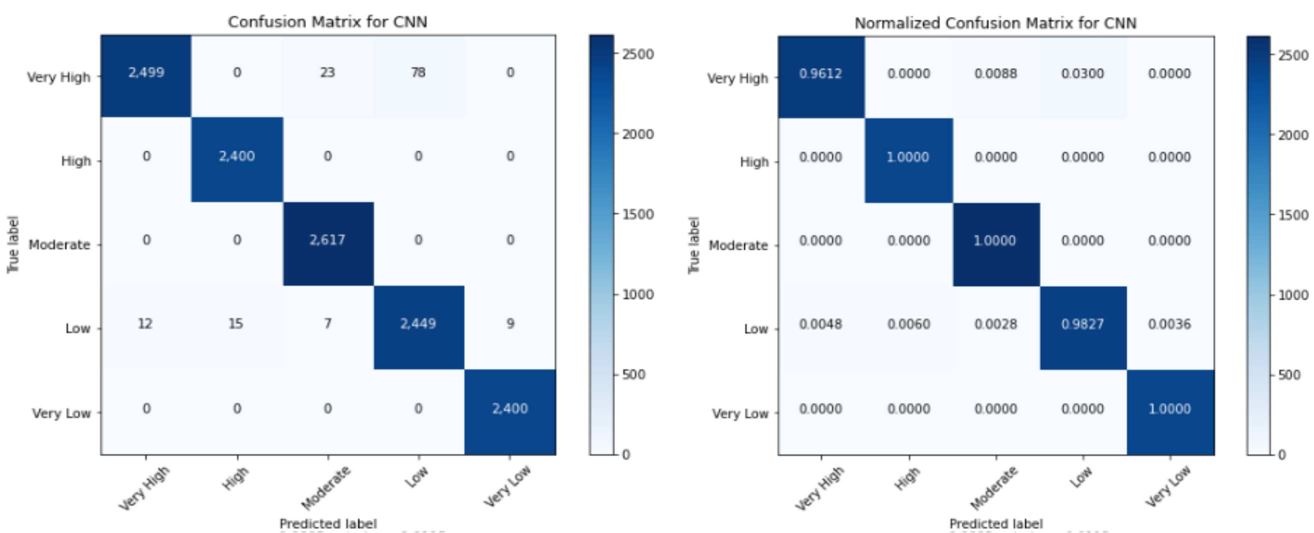


Figure 6-3 confusion matrix of CNN model

As can be seen from Fig 6-3 above, the performance of the 1D-CNN model is examined on soil fertility dataset using confusion matrix, the target class are seen as five target classes very high (VH), high(H), moderate(M), low(L) and very low (VL) respectively. As such 2499 out of 2600, 2400 out of 2400, 2617 out of 2617, 2449 out of 2492, and 2400 out of 2400 of dataset instances are correctly classified as Very High, High, Moderate, Low, and Very Low respectively. Where the model misclassifies only 23 instances of Very high as Moderate and 39 instance as Low class. And 12,15,7 instances of low class in to Very High, High and Moderate respectively. Else the rest instances are correctly classified so it shows a great accuracy. The normalized confusion matrix for 1D-CNN is also illustrated on the right side in fig 6.3 above which is the same as the left, except it indicates correctly classified instances in the form of a percentage.

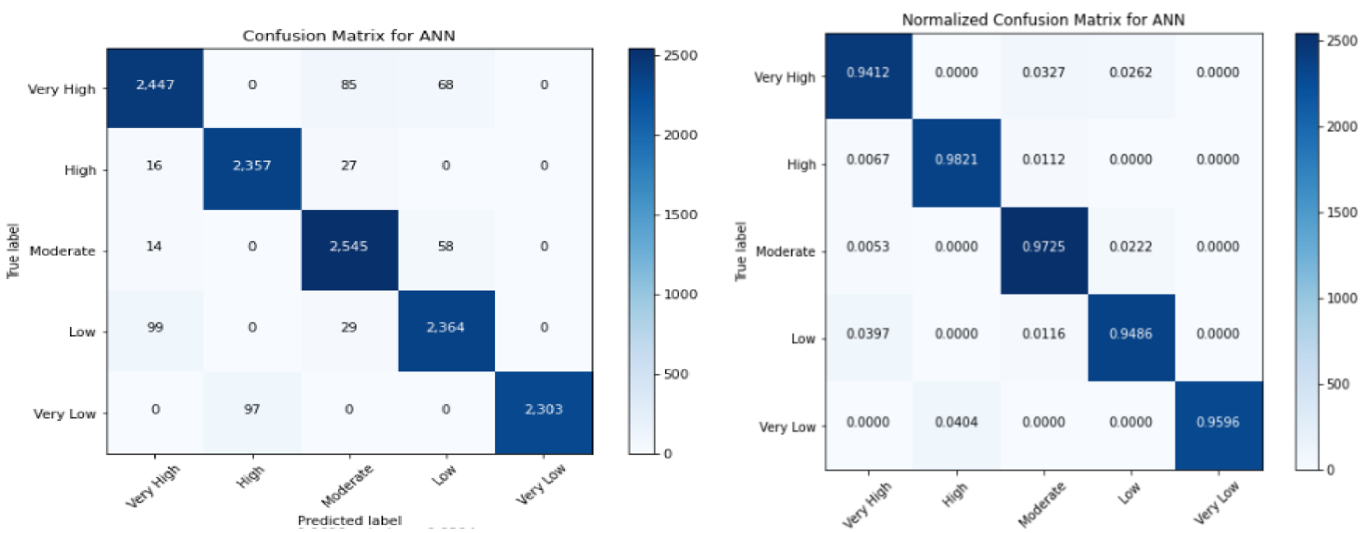


Figure 6-4 Confusion matrix of ANN model

Similarly, another model used for this study is ANN and its performance is also assessed using a confusion matrix as shown in Fig 6.4 above. Accordingly, the model correctly classifies instances of 2447 out of 2600, 2357 out of 2357, 2545 out of 2617, 2364 out of 2492, 2303 out of 2400 as Very High, High, Moderate, Low, and Very Low respectively. The model misclassifies 31 instances of very high class as Low, 11,8 instance of Moderate class as very high and low respectively and 25 instances of Low class as Very High. However, the model classifies the dataset instances with a good accuracy and the above figure shows Confusion Matrix for ANN without normalization on the (left) and with Normalization(right) they are the same except the right one is in the normalized form.

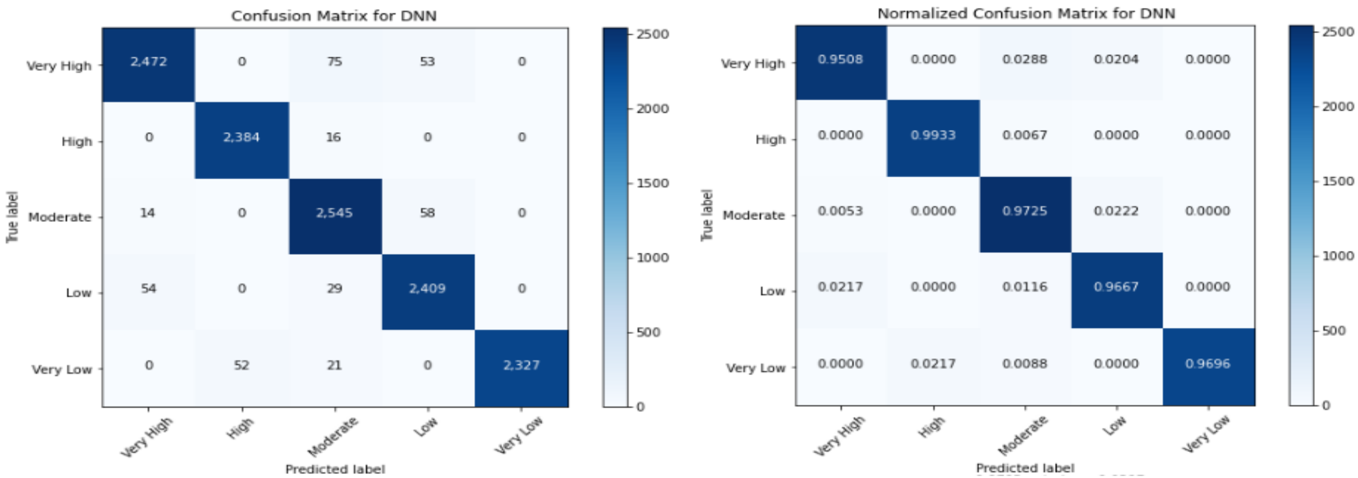
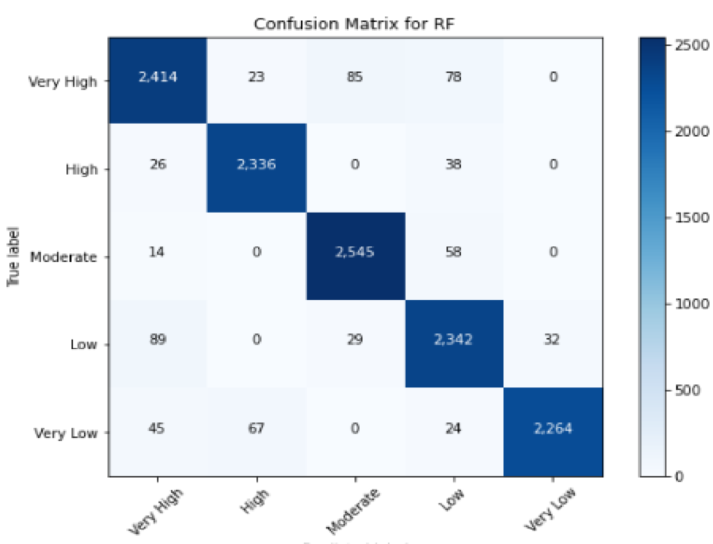


Figure 6-5 Confusion matrix of DNN model

Another model used for this study is DNN and its performance is also assessed using a confusion matrix as shown in Fig 6.5 above. Accordingly, the model correctly classifies instances of 2472 out of 2600, 2384 out of 2357, 2545 out of 2617, 2409 out of 2492, 2327 out of 2400 as Very High, High, Moderate, Low, and Very Low respectively. The model misclassifies 75,53 instances of very high class as Moderate and Low respectively, 16 instance of High class as Moderate and 25 instances of Low class as Very High. However, the model classifies the dataset instances with a good accuracy and the above figure shows Confusion Matrix for ANN without normalization on the (left) and with Normalization(right) they are the same except the right one is in the normalized form.



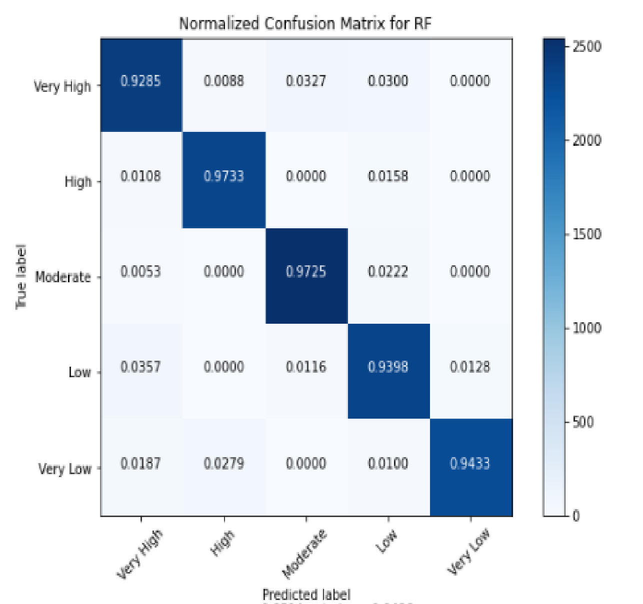


Figure 6-6 Confusion Matrix of RF model

The above figure Fig 6. illustrates confusion matrix for Random Forest model. We applied this model in order to compare classification accuracy of CNN-based deep learning with machine learning algorithms, which are a more conventional approach for classification problems. The same learning algorithm that we used for 1D-CNN is also applied here like cleaning of noisy, handling of categorical data, normalization before applying the classifier and evaluate the result. However, the result shows that it has lower accuracy compared to the other deep learning models.

For soil fertility prediction classification performance evaluation results of the four models under stratified 10-fold cross-validation is shown below.

Table 6-1 Stratified 10-fold Accuracy result for soil fertility prediction

Models	**Stratified 10-fold Accuracy result for each fold (%)**										**Accuracy**
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	(%)
1D CNN	98.96	99.12	98.24	98.64	99.60	98.88	98.00	98.56	99.04	99.20	**98.82**
DNN	97.20	97.28	97.04	95.20	91.05	97.84	98.88	99.20	95.04	99.52	**96.83**
ANN	96.32	98.64	96.56	95.76	96.24	94.88	97.52	96.48	92.33	97.04	**96.18**
RF	95.50	96.03	96.03	93.54	98.02	96.52	95.01	94.03	94.66	93.42	**95.27**

As shown in the table, among the four models, 1D-CNN outperforms ANN and the traditional machine learning approach Random Forest in all performance metrics. The table below shows the results of 1D-CNN, DNN, ANN, and RF models with stratified 10-fold CV and other performance metrics such as precision, recall, and F1 score.

Table 6-2 Performance metrics report

Models	**Stratified 10-fold CV result of performance metrics for each model**		
	Precision (%)	**F1-score (%)**	**Recall (%)**
1D-CNN	98.80	98.78	98.77
DNN	96.81	96.76	96.75
ANN	96.07	96.05	96.16
RF	95.16	95.14	95.17

6.4.2 Suitable Fertilizer Prediction Model Result

Like that of soil fertility prediction the performance and evaluation result of the 1D-CNN, ANN, DNN and RF models on suitable fertilizer prediction is discussed with the confusion matrix and with others performance metrics of the models under stratified 10-fold cross-validation.

CNN confusion matrix for fertilizer prediction

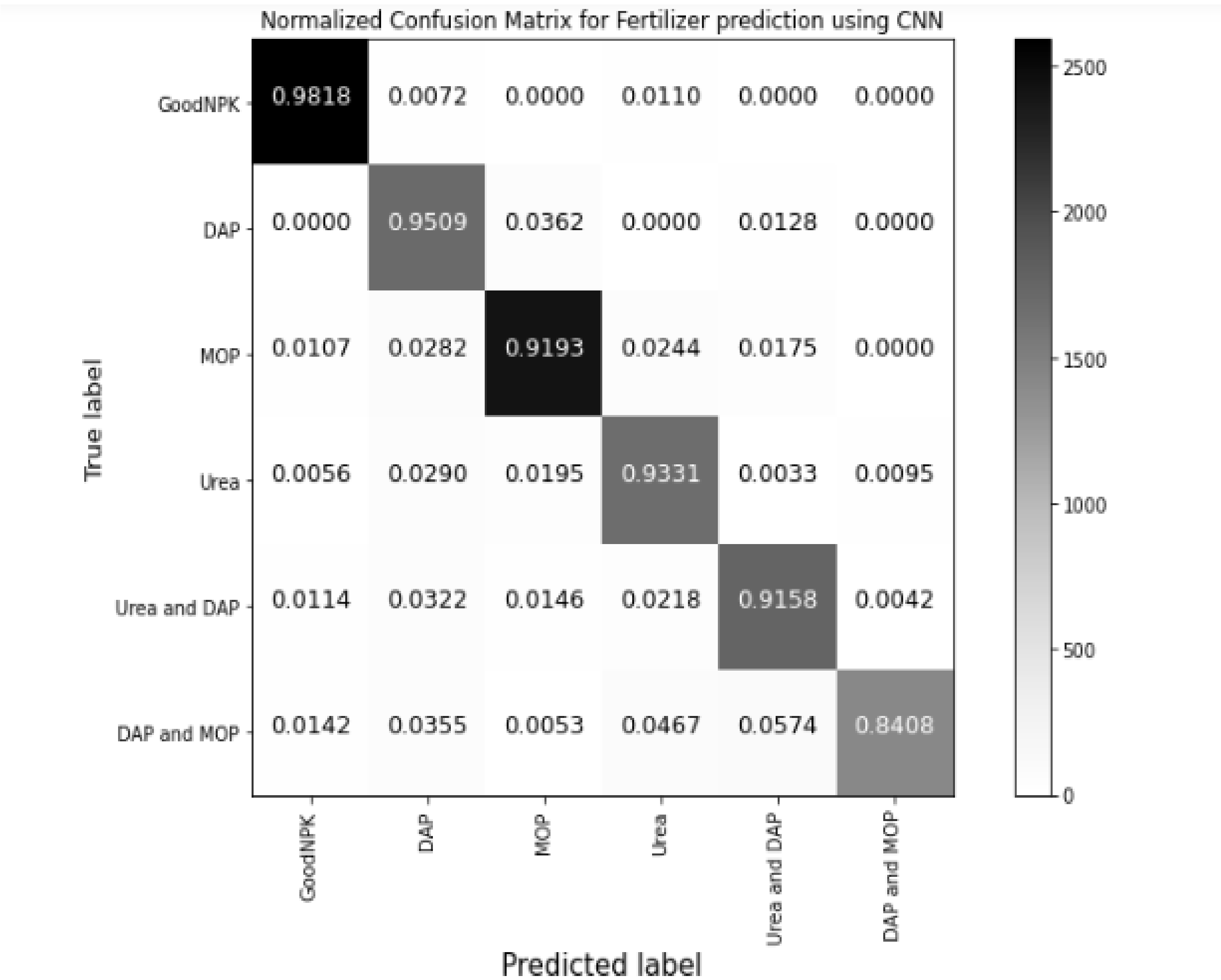


Figure 6-7 confusion matrix for fertilizer prediction

Below are the evaluation findings of the four models using stratified 10-fold cross-validation for appropriate fertilizer prediction.

Table 6-3 ***Stratified 10-fold Accuracy result for fertilizer prediction***

Models	Stratified 10-fold Accuracy result for each fold (%)	Accuracy

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	(%)
1D CNN	97.94	98.01	96.22	97.64	97.50	98.64	95.86	97.56	94.96	98.94	**97.22**
DNN	97.28	94.68	95.35	95.28	91.05	96.84	95.67	93.66	94.94	95.52	**95.02**
ANN	94.52	91.84	96.66	92.57	96.94	94.98	96.52	96.48	92.33	94.64	**94.74**

Similar to that of soil fertility prediction precision, recall and F1-score has been used as evaluation metrics for suitable fertilizer prediction. The experimental results are shown in the following table.

***Table 6-4* Performance metrics report for fertilizer prediction**

Models	**Stratified 10-fold CV result of performance metrics for each model**		
	Precision (%)	**F1-score (%)**	**Recall (%)**
1D-CNN	98.56	97.23	97.77
DNN	96.84	95.28	95.64
ANN	95.26	95.55	95.55

6.5 Tuning Hyperparameters

Because neural networks have numerous parameters and can be a little challenging to set up so they can perform well on a particular problem hyperparameter optimization was developed. The grid search optimization method was applied. It takes a little bit of time because many trials must be done, but it is the best method for optimizing hyperparameters. We tweaked parameters for our CNN model starting from several neurons. The quantity of neurons in a layer is a crucial variable to adjust. The representational capacity of the network is often controlled by the number of neurons in a layer, at least at that point in the architecture. Epochs and batch size were also tuned. They specify how many patterns to read at once and how frequently the network will see the training dataset. Using the same methodology, dropout and learning rates were optimized.

```
Best: 0.983292 using {'batch_size': 30, 'epochs': 60}
0.846350 (0.009645) with: {'batch_size': 10, 'epochs': 10}
0.896397 (0.042342) with: {'batch_size': 10, 'epochs': 30}
0.967302 (0.009922) with: {'batch_size': 10, 'epochs': 60}
0.849707 (0.010795) with: {'batch_size': 20, 'epochs': 10}
0.926054 (0.005844) with: {'batch_size': 20, 'epochs': 30}
0.964587 (0.019092) with: {'batch_size': 20, 'epochs': 60}
0.844912 (0.015125) with: {'batch_size': 30, 'epochs': 10}
0.934288 (0.009499) with: {'batch_size': 30, 'epochs': 30}
0.983292 (0.004703) with: {'batch_size': 30, 'epochs': 60}
0.825165 (0.009679) with: {'batch_size': 64, 'epochs': 10}
0.933169 (0.008607) with: {'batch_size': 64, 'epochs': 30}
0.979855 (0.001369) with: {'batch_size': 64, 'epochs': 60}
```

Figure 6-8 Parameter tuning for the number of epochs and batches

6.6 Comparison

Today, practically all industries that require estimation uses machine learning as a tool. and these strategies also aid non-technical industries like agriculture in order to increase crop production and quality, battle poverty, improve food security, and improve countries' economies generally, as well as to change farmer’s life style. And several research studies were conducted in the agricultural domain by using this machine learning approaches. Predicting soil fertility will help to simplify the challenges faced by farmers and serve as a conduit for providing the agriculturalist with the effective proof needed to obtain a higher yield. Therefore, in this study deep learning techniques in predicting the status of soil fertility and the prediction of suitable type of fertilizer is proposed to improve crop productivity. Some studies that used a machine learning approach to predict the soil fertility a given soil samples are documented in literature reviews as related work.

To make sure that this research study outperforms the previous study on soil fertility prediction, it is compared to a prior study that is discussed in related work on soil fertility prediction. In most research papers they proposed a model that determine whether a soil fertile or not which is limited in binary classification. And for soil classification and agricultural yield a variety of machine learning algorithms are utilized including Support Vector Machine (SVM), Random Forest, Naive Bayes, Linear Regression, Multilayer Perceptron (MLP), and ANN. However, they have used traditional machine learning algorithms and the dataset was too small and most of them labeled it with a binary (fertile or not) target classes to classify the soil samples fertility level which make it difficult to determine the exact status of the soil fertility. But in our study enough soil samples with various physiochemical characteristics of the soil were taken to train the machine. Using multiclass classification, this dataset is divided into Very Highly Fertile, Highly Fertile, Moderately Fertile, Low Fertile, and Very Low Fertile for prediction of soil fertility. In our study, suitable fertilizer prediction is also taken into account. This predicts an appropriate type of fertilizer to apply for a given soil sample based on the values of macro nutrients of the soil. DAP, MOP, Urea, Urea and DAP, and DAP and MOP are different types of fertilizer. In the majority of research papers, the features used to predict soil fertility only take into account the soil's macronutrient content, however this study aims to take into account all the significant characteristics of a soil that directly affect soil fertility. Our proposed model has performed extremely well in terms of accuracy, with accuracy ratings of 98.82% for predicting soil fertility and 97.22% for predicting appropriate fertilizer.

6.7 Discussion

As a result, the goal of this study to build a deep learning model that can predict the level of soil fertility based on the soil fertility index by analyzing the soil physiochemical characteristics and also for the prediction of appropriate type of fertilizer to apply on a given soil sample. This reduces time and enhances the accuracy of soil fertility prediction for agricultural professionals and soil science experts.

To achieve this, soil survey reports were collected from Debre Markos soil laboratory center and Adet agricultural research center to perform soil fertility evaluation, soil data like soil physical and chemical parameters included in the dataset. Therefore, data were collected by considering those features individually. The collected, data was raw data with a lot of missing and incorrect values and by having direct communication with soil laboratory experts to understand more about the data and how to label it with appropriate fertility class and suitable fertilizer. As a result, we label the dataset based on the soil fertility index for soil fertility classification and also we have used FAO guide line for the labeling of suitable fertilizer based on attribute values and the labeled results were again reviewed by agricultural experts to get rid of major and slight mistakes.

Several experiments were carried out in this study to develop deep learning models and answer the formulated research questions. various deep learning models like 1D CNN, ANN, DNN and machine learning RF have been experimented and their prediction performance assessed with different evaluation metrics for both soil fertility and suitable fertilizer prediction. Models with excellent accuracy on the soil dataset were selected and 1D-CNN in both cases shows a great performance. According to Table 6-2 and Table 6-3, which highlights the model's performance evaluation result, 1D-CNN outperforms all other employed models in all performance criteria.

To achieve optimal accuracy, we tuned CNN main parameters that have a large impact on model performance, such as epoch, dropout rate, batch size, and number of neurons. We achieved the best results by configuring all of these parameters with the results of hyper parameter tuning. The figure below displays graphical perspectives of model comparison in terms of accuracy for soil fertility prediction.

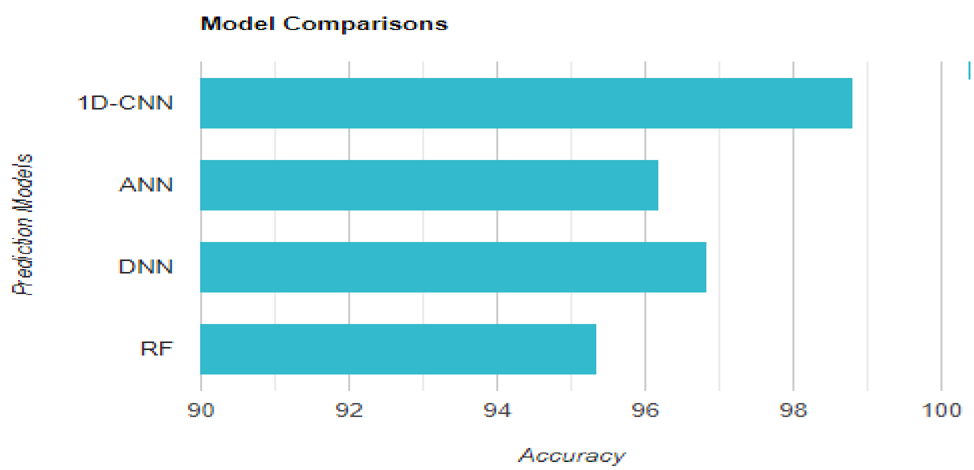


Figure 6-9 Accuracy comparison for soil fertility prediction

CHAPTER SEVEN

7 CONCLUSION AND FUTURE WORK

7.1 Conclusion

Knowing the exact fertility status of a particular soil sample and also identifying the appropriate type of fertilizer to apply to make the soil more fertile are critical components for successful agricultural productivity. As a result, this study presents two prediction models that were applied to newly obtained soil samples from soil testing laboratories. The dataset contains 12,509 instances with 17 variables, including two target variables. Experts in soil testing reviewed the labeled dataset. We compared deep learning models for classification and tested each one using a 10-fold stratified cross-validation technique for both soil fertility and fertilizer prediction.

The best hyperparameters for the CNN, DNN, and ANN models were found using the grid search cv hyperparameter tuning technique. Under stratified 10-fold cross-validation, the best performing model on both soil fertility and suitable fertilizer prediction was determined to be a CNN, with scores of 98.82% accuracy, 98.0% precision, 98.78% recall, and 98.77 percent f1-score for soil fertility prediction and 97.22% accuracy, 98.56 precision, 97.23 recall, and 97.77 F1-score for suitable fertilizer prediction. However, it would be preferable if this model was deployed with Android to make it more suitable for farmers to use.

7.2 Future work

Even though this study significantly improves on previous studies on soil fertility prediction, it is preferable to extend this study into an IoT-based system to obtain real-time soil data where sensors are installed in the farming to directly detect soil nutrients values, which are then used as input to the model for prediction. So that farmers can make an informed judgment on the fertility of their soil. Soil nutrient testing is an expensive and time-consuming operation that takes more than 21 hours to complete, hence it is preferable if predicting values of this soil nutrient are considered.

* * *

This research was funded by Adama Science and Technology University with a grant

number ASTU/SM-R/556/22

REFERENCES

A, K. N., M, O. A., I, M. O., & V, O. C. (2012). Soil fertility evaluation: a potential tool for predicting fertilizer requirement for crops in Nigeria. *African Journal of Agricultural Research*, *7*(47), 6204–6214. https://doi.org/10.5897/ajar12.210

Abd-Elmabod, S. K., Jordán, A., Fleskens, L., Phillips, J. D., Muñoz-Rojas, M., van der Ploeg, M., Anaya-Romero, M., El-Ashry, S., & de la Rosa, D. (2017). Modeling Agricultural Suitability Along Soil Transects Under Current Conditions and Improved Scenario of Soil Factors. *Soil Mapping and Process Modeling for Sustainable Land Use Management*, 193–219. https://doi.org/10.1016/B978-0-12-805200-6.00007-4

Abebe, B. (2021). September 2021 Adama, Ethiopia. *Thesis*.

Agegnehu, G., Ashagrie, A. K., & Yirga, C. (2020). *Farming Systems Characterization and Analysis in East Gojjam Zone : Implications for research and Research Report No . 127 Getachew Agegnehu Ethiopian Institute of Agricultural Research* (Issue October).

Ahmed, B. (2016). What Factors Contribute to the Smallholder Farmers Farm Income Differential? Evidence from East Hararghe, Oromia, Ethiopia. *Journal of Asian Scientific Research*, *6*(7), 112–119. https://doi.org/10.18488/journal.2/2016.6.7/2.7.112.119

Apt, W. (2014). Introduction. *Demographic Research Monographs*, 1–13. https://doi.org/10.1007/978-94-007-6964-9_1

Ayoubi, S., Pilehvar, A., Mokhtari, P., & L., K. (2011). Application of Artificial Neural Network (ANN) to Predict Soil Organic Matter Using Remote Sensing Data in Two Ecosystems. In *Biomass and Remote Sensing of Biomass*. InTech. https://doi.org/10.5772/18956

Boggs, Liam, K. (2017). *Performance Measures for Machine Learning 1*. 1–32.

Butterly, C., Baldock, J., & Tang, C. (2010). Chemical mechanisms of soil pH change by agricultural residues. *19th World Congress of Soil Science, August*, 43–46.

Chamberlin, J., Schmidt, E., Chamberlin, J., & Schmidt, E. (2015). *Ethiopian Agriculture : A Dynamic Geographic Perspective ESSP II Working Paper 17 Ethiopian Agriculture : A Dynamic Geographic Perspective*. *April*, 30.

Chaudhari, R., Chaudhari, S., Shaikh, A., & Chiloba, R. (2019). *Soil Fertility Prediction Using Data Mining Techniques*. *IX*(Vi), 202–209.

Conijn, S., Hermelink, M., Deolu-Ajayi, A., Kuiper, M., & Cervi, W. R. (2019). Food System Challenges for Ethiopia. *Project Progress Report*, 1–59.

Corning, E., Sadeghpour, A., Ketterings, Q., & Czymmek, K. (2016). The Carbon Cycle and Soil Organic Carbon. *Agronomy Fact Sheet Series*, *C*, 2.

Davis, G., Casady, W., & Massey, R. (1998). Water Quality Precision Agriculture : An Intoduction. *Precision Agriculture: An Introduction*, 1–7.

Dobo, B. (2019). *FERTILITY EVALUATION OF SOIL FOR CEREAL*. *7*(3), 1007–1037.

Elbasit M M, A. M. (2015). *Journal of Soil Science and Environmental Management Spatio-temporal variations of rainfall erosivity in Rwanda*. *6*(4), 72–83. http://www.academicjournals.org/JSSEM

Elias, E. (2019). *Eyasu E. Is soil fertility declining? 1*. *October*.

Fairhurst, T. (2012). Fairhurst, Thomas. Handbook for Integrated Soil Fertility Management (. *Handbook of Soil Fetility Management*.

Gashe, A. (2017). *Environmental & Analytical Toxicology Feed Resources Gozamen District , East Gojjam Zone , Amhara Region*. *7*(2). https://doi.org/10.4172/2161-0525.1000437

Hailu, G., Weersink, A., & Minten, B. (2015). Rural Organizations, Agricultural Technologies and Production Efficiency of Teff in Ethiopia. In *International Conference of Agricultural Economists*.

Hofman G. and O.V. Cleemput. (2004). Soil and Plant Nitrogen. First version. *IFA, Paris, France, September*, 31.

June, A. (2017). *Soil Texture and Water*. 1–9.

Kamatchi, B. (2021). *Crop Recommendation System by Artificial Neural Network*. 1–18.

Kamiri, J., & Mariga, G. (2021). Research Methods in Machine Learning: A Content Analysis. *International Journal of Computer and Information Technology(2279-0764)*, *10*(2), 78–91. https://doi.org/10.24203/ijcit.v10i2.79

Keerthan Kumar, T. G., Shubha, C., & Sushma, S. A. (2019). Random forest algorithm for soil fertility prediction and grading using machine learning. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 1301–1304. https://doi.org/10.35940/ijitee.L3609.119119

Keshavarzi, A., Tuffour, H. O., Bagherzadeh, A., Tattrah, L. P., Kumar, V., Gholizadeh, A., & Rodrigo-Comino, J. (2020). Using fuzzy-AHP and parametric technique to assess soil fertility status in Northeast of Iran. *Journal of Mountain Science*, *17*(4), 931–948. https://doi.org/10.1007/s11629-019-5666-6

Khaki, B. D., Honarjoo, N., Davatgar, N., Jalalian, A., & Golsefidi, H. T. (2017a). Assessment of two soil fertility indexes to evaluate paddy fields for rice cultivation. *Sustainability (Switzerland)*, *9*(8). https://doi.org/10.3390/su9081299

Khaki, B. D., Honarjoo, N., Davatgar, N., Jalalian, A., & Golsefidi, H. T. (2017b). Assessment of two soil fertility indexes to evaluate paddy fields for rice cultivation. *Sustainability (Switzerland)*, *9*(8), 1–13. https://doi.org/10.3390/su9081299

Kheyrodin, H. (2014). Important of soil quality and soil agriculture indicators. *Academia Journal of Agricultural Research*, *2*(11), 231–238.

Khomiakov, D. M. (2020). Soil Is an Essential Component of the Biosphere and the Global Food System (Critical Assessment of the Situation). *Moscow University Soil Science Bulletin*, *75*(4–5), 147–158. https://doi.org/10.3103/s0147687420040055

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*, 1–20. https://doi.org/10.1016/j.ymssp.2020.107398

KUMAR, M., NAMEI, A., MEENA, K. L., RAJKHOWA2, D. J., G, R., DAS, A., MAHAPATRA, K., CHOWDHURY, P., RANGNAMEI, K. L., & KUOTSU, R. (2019). Soil

Properties as Influenced by Various Land Use Systems under Hill Ecosystem of North East India. *Journal of AgriSearch*, *6*(02). https://doi.org/10.21921/jas.v6i02.15752

Lateef, R., & Abbas, A. (2022). Tuning the Hyperparameters of the 1D CNN Model to Improve the Performance of Human Activity Recognition. *Engineering and Technology Journal*, *40*(4), 547–554. https://doi.org/10.30684/etj.v40i4.2054

Li, Q., Yang, Y., Guo, Y., Li, W., Liu, Y., Liu, H., & Kang, Y. (2021). Performance Evaluation of Deep Learning Classification Network for Image Features. *IEEE Access*, *9*(January), 9318–9333. https://doi.org/10.1109/ACCESS.2020.3048956

M S*, Prof. M., Gracias, Ms. A. A., Neha, Ms. C. G., Ms. Preethi, & S, Mr. A. (2020). Soil Fertility Analysis and Crop Prediction using Machine Learning. *International Journal of Innovative Technology and Exploring Engineering*, *9*(6), 380–383. https://doi.org/10.35940/ijitee.f3609.049620

McCauley, A., Jones, C., & Jacobsen, J. (2005). Basic Soil Properties. *Soil and Water*, 1–12. http://landresources.montana.edu/SWM/PDF/Final_proof_SW1.pdf

Meena, V. S., Maurya, B. R., Verma, J. P., & Meena, R. S. (2016). Potassium solubilizing microorganisms for sustainable agriculture. *Potassium Solubilizing Microorganisms for Sustainable Agriculture*, *June*, 1–331. https://doi.org/10.1007/978-81-322-2776-2

Meliboev, A., Alikhanov, J., & Kim, W. (2020). *1D CNN Based Network Intrusion Detection with Normalization on Imbalanced Data*. https://doi.org/10.48550/arxiv.2003.00476

Mellor, J. W., & Dorosh, P. (2010). Agriculture and the economic transformation of Ethiopia. *ESSP II Working Paper*. http://www.ifpri.org/sites/default/files/publications/esspwp010.pdf

Mitiche, I., Nesbitt, A., Conner, S., Boreham, P., & Morison, G. (2020). 1D-CNN based real-time fault detection system for power asset diagnostics. *IET Generation, Transmission and Distribution*, *14*(24), 5816–5822. https://doi.org/10.1049/iet-gtd.2020.0773

MoARD. (2010). Federal Democratic Republic of Ethiopia Ministry of Agriculture and Rural Development. *Ministry of Agriculture and Rural Development, Ethiopia's Agricultural Sector Policy and Investment Framework*, *2010*(June), 2009–2012.

Mwende Muindi, E. (2019). Understanding Soil Phosphorus. *International Journal of Plant & Soil Science*, *December*, 1–18. https://doi.org/10.9734/ijpss/2019/v31i230208

Neira, J., Ortiz, M., Morales, L., & Acevedo, E. (2015). Oxygen diffusion in soils: Understanding the factors and processes needed for modeling. *Chilean Journal of Agricultural Research*, *75*(July), 35–44. https://doi.org/10.4067/S0718-58392015000300005

Page, D. (2017). *Evaluating Machine Learning Methods*.

Prasad, R., & Shivay, Y. S. (2020). *Calcium as a Plant Nutrient October 2020 Calcium as a Plant Nutrient*. *February 2021*. https://doi.org/10.23910/1.2020.2075a

Rahman, K. M. A., & Zhang, D. (2018). Effects of fertilizer broadcasting on the excessive use of inorganic fertilizers and environmental sustainability. *Sustainability (Switzerland)*, *10*(3). https://doi.org/10.3390/su10030759

Rashid, S., Tefera, N., Minot, N., & Ayele, G. (2014). Fertilizer in Ethiopia: An Assessment of Policies, Value Chain, and Profitability. *SSRN Electronic Journal*, *December*. https://doi.org/10.2139/ssrn.2373214

Sekeroglu, B., Hasan, S. S., & Abdullah, S. M. (2020). Comparison of Machine Learning Algorithms for Classification Problems. *Advances in Intelligent Systems and Computing*, *944*, 491–499. https://doi.org/10.1007/978-3-030-17798-0_39

Sivasankaran, S., Mohan, K. J., & Nazer, G. M. (2022). Soil Nutrients Prediction and Optimal Fertilizer Recommendation for Sustainable Cultivation of Groundnut Crop using Enhanced-1DCNN DLM. *International Journal of Advanced Computer Science and Applications*, *13*(4), 162–171. https://doi.org/10.14569/IJACSA.2022.0130419

Sori, O. (2021). Factors affecting groundnut market supply in Western Oromia, Ethiopia. *Heliyon*, *7*(1), e05892. https://doi.org/10.1016/j.heliyon.2020.e05892

Suchithra, M. S., & Pai, M. L. (2020). Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. *Information Processing in Agriculture*, *7*(1), 72–82. https://doi.org/10.1016/j.inpa.2019.05.003

Taffesse, A. S., Dorosh, P., & Gemessa, S. A. (2012). *Crop Production in Ethiopia: Regional Patterns and Trends. Summary of Report Ethiopian Strategy Support Program (ESSP II), Research Note 11, IFPRI and EDRI, Addis Ababa, Ethiopia*. *Essp Ii*.

Taylor, L. (2020). From food crisis to nutrition: challenges and possibilities in Ethiopia’s nutrition sector. *Institute of Development Studies, February*, 1–22.

van Es, H. (2017). A New Definition of Soil. *CSA News*, *62*(10), 20–21. https://doi.org/10.2134/csa2017.62.1016

Voss, R. (1998). *Micronutrients Micronutrients Introduction*.

Worku, T., & Tripathi, S. K. (2015). *Watershed Management in Highlands of Ethiopia : A Review Watershed Management in Highlands of Ethiopia : A Review*. *April 2019*. https://doi.org/10.4236/oalib.1101481

Zerssa, G., Feyssa, D., Kim, D. G., & Eichler-Löbermann, B. (2021). Challenges of smallholder farming in Ethiopia and opportunities by adopting climate-smart agriculture. *Agriculture (Switzerland)*, *11*(3), 1–26. https://doi.org/10.3390/agriculture11030192

APPENDICES

Appendix A: Statistical Results of the Collected Soil Samples

Nutrient	**Min**	**Max**	**Mean**	**Std**
PH	5	7.89	6.29	0.86
OC	0.15	11.5	2.35	1.38
Mg	0.11	8.77	2.99	1.75
N	18	499	167.76	137.02
P	20	60	22.51	13.02
K	29	480	153.61	116.5
Zn	0.01	2.9	0.83	0.50
Fe	1.26	15.62	4.89	2.6
Cu	0.2	0.99	0.58	0.23
Mn	0.27	9.6	3.96	2.01
Sand	40.2	561.45	75.74	20.46
Silt	1.8	20	8.59	3.74
Clay	2.08	18.4	9.21	2.76
CaCo3	1.06	19.97	8.22	4.13
CEC	1.4	24.92	9.92	5.59

Appendix B: Supporting Results

Appendix B:1 Stratified 10-fold CV result

1D-CNN

```
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import KFold
import numpy

kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=7)
cvscores = []
for train, test in kfold.split(X, Y):
    model = Sequential()
    model.add(Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(15,1)))
    model.add(BatchNormalization())
    model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
    model.add(Dropout(0.5))
    model.add(MaxPooling1D(pool_size=2))
    model.add(Flatten())
    model.add(Dense(64, activation='relu'))
    model.add(Dense(5, activation='softmax'))
    model.compile(loss='sparse_categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    model.fit(X[train], Y[train], epochs=60, batch_size=64, verbose=0)
    # evaluate the model
    scores = model.evaluate(X[test], Y[test], verbose=0)
    print("%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
    cvscores.append(scores[1] * 100)
print("%.2f%% (+/- %.2f%%)" % (numpy.mean(cvscores), numpy.std(cvscores)))
```

```
accuracy: 98.96%
accuracy: 99.12%
accuracy: 98.24%
accuracy: 98.64%
accuracy: 99.60%
accuracy: 98.88%
accuracy: 98.00%
accuracy: 98.56%
accuracy: 99.04%
accuracy: 99.20%
98.82% (+/- 0.45%)
```

DNN

```
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import KFold
import numpy

kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=7)
cvscores = []
for train, test in kfold.split(X, Y):
    model = Sequential()
    model.add(Dense(64, input_dim=15, activation='relu', name='layer_1'))
    model.add(Dense(28, activation='relu', name='layer_2'))
    model.add(Dense(28, activation='relu', name='layer_3'))
    model.add(Dense(28, activation='relu', name='layer_4'))
    model.add(Dense(5, activation='softmax', name='output_layer'))
    model.compile(optimizer="adam", loss="sparse_categorical_crossentropy", metrics=["accuracy"])
    history=model.fit(X[train], Y[train], epochs=70, batch_size=64, verbose=0)
    # evaluate the model
    scores = model.evaluate(X[test], Y[test], verbose=0)
    print("%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
    cvscores.append(scores[1] * 100)
print("%.2f%% (+/- %.2f%%)" % (numpy.mean(cvscores), numpy.std(cvscores)))
```

```
accuracy: 97.20%
accuracy: 97.28%
accuracy: 97.04%
accuracy: 95.20%
accuracy: 91.05%
accuracy: 97.84%
accuracy: 98.88%
accuracy: 99.20%
accuracy: 95.04%
accuracy: 99.52%
96.83% (+/- 2.40%)
```

ANN

```
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import KFold
import numpy

kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=7)
cvscores = []
for train, test in kfold.split(X, Y):
    model= Sequential()
    model.add(Dense(64,input_shape=(15,), activation="relu"))
    model.add(Dense(28, activation="relu"))
    model.add(Dense(5, activation='softmax'))
    model.compile(optimizer="adam", loss="sparse_categorical_crossentropy", metrics=["accuracy"])
    history=model.fit(X[train], Y[train], epochs=80, batch_size=64, verbose=0)
    # evaluate the model
    scores = model.evaluate(X[test], Y[test], verbose=0)
    print("%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
    cvscores.append(scores[1] * 100)
print("%.2f%% (+/- %.2f%%)" % (numpy.mean(cvscores), numpy.std(cvscores)))
```

```
accuracy: 96.32%
accuracy: 98.64%
accuracy: 96.56%
accuracy: 95.76%
accuracy: 96.24%
accuracy: 94.88%
accuracy: 97.52%
accuracy: 96.48%
accuracy: 92.33%
accuracy: 97.04%
96.18% (+/- 1.60%)
```

RF

```
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import StratifiedKFold
from statistics import mean
cv=StratifiedKFold(n_splits=10, random_state=20, shuffle=True)
acc_scores=cross_val_score(rf_model,X,y,cv=cv,n_jobs=-1)
print("Accuracy")
print("======")
fold=1
for i in acc_scores:
    print("accuracy:",round(i*100,2))
    fold=fold+1
print("mean accuracy:", acc_scores.mean()*100)
```

```
Accuracy
======
accuracy: 95.5
accuracy: 96.0
accuracy: 96.0
accuracy: 93.5
accuracy: 98.0
accuracy: 97.0
accuracy: 95.0
accuracy: 99.0
accuracy: 89.5
accuracy: 94.0
mean accuracy: 95.34999999999998
```

Appendix C: Sample Code

Appendix C:1 Sample code for Preprocessing Implementations

#Feature scaling

```
X = df.iloc[:, :-1].values
scaler = MinMaxScaler(feature_range=(0,1))
# fit and transform in one step
normalized = scaler.fit_transform(X)
# inverse transform
inverse = scaler.inverse_transform(normalized)
print(normalized)
```

```
[[0.72258917 1.         0.1039261  ... 0.41789216 0.14066631 0.05994898]
 [0.7661823  0.52264808 0.12933025 ... 0.7120098  0.18773136 0.37372449]
 [0.67767503 0.43554007 0.10623557 ... 0.64460784 0.13061872 0.46428571]
 ...
 [0.68692206 0.17595819 0.45381062 ... 0.40379902 0.42728715 0.40263605]
 [0.90620872 0.04703833 0.4965358  ... 0.32720588 0.09360127 0.71258503]
 [0.78071334 0.2325784  0.13048499 ... 0.38664216 0.5483871  0.66454082]]
```

#Handling categorical values

```
le = preprocessing.LabelEncoder()
df['Fertility_Class'] = le.fit_transform(df['Fertility_Class'])
df['Fertility_Class'].unique()
```

```
array([3, 0, 2, 1, 4])
```

Appendix C:2 Sample Code for Building, Creating and Compiling the Model

```
from keras.models import Sequential
import pandas as pd
from keras.layers import Dense, Conv1D, Flatten, MaxPooling1D,BatchNormalization,Dropout
from numpy import unique
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn import preprocessing
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
data=pd.read_csv('Soil_Ferrtlity.csv', encoding='cp1252')
X = df.iloc[:, :-1].values
X = X.reshape(X.shape[0], X.shape[1], 1)
Y = df.iloc[:, -1].values
Y= le.fit_transform(Y)
model = Sequential()
model.add(Conv1D(128, 2, activation="relu", input_shape=(15,1)))
model.add(BatchNormalization())
model.add(Dense(64, activation="relu"))
model.add(Dropout(0.5))
model.add(MaxPooling1D())
model.add(Flatten())
model.add(Dense(5, activation = 'softmax'))
model.compile(loss = 'sparse_categorical_crossentropy', optimizer = "adam", metrics = ['accuracy'])
model.summary()
```

Appendix C:3 Sample Code for Hyper Parameter Tuning

```
import numpy as np
import tensorflow as tf
from sklearn.model_selection import GridSearchCV
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from scikeras.wrappers import KerasClassifier
def create_model():
    # create model
    model = Sequential()
    model.add(Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(15,1)))
    model.add(BatchNormalization())
    model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
    model.add(Dropout(0.5))
    model.add(MaxPooling1D(pool_size=2))
    model.add(Flatten())
    model.add(Dense(64, activation='relu'))
    model.add(Dense(5, activation='softmax'))
    model.compile(loss='sparse_categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model
seed = 7
tf.random.set_seed(seed)

model = KerasClassifier(model=create_model, verbose=0)
# define the grid search parameters
batch_size = [10,20,30,64]
epochs = [10, 30, 60]
param_grid = dict(batch_size=batch_size, epochs=epochs)
grid = GridSearchCV(estimator=model, param_grid=param_grid, n_jobs=-1, cv=3)
grid_result = grid.fit(X, Y)
# summarize results
print("Best: %f using %s" % (grid_result.best_score_, grid_result.best_params_))
means = grid_result.cv_results_['mean_test_score']
stds = grid_result.cv_results_['std_test_score']
params = grid_result.cv_results_['params']
for mean, stdev, param in zip(means, stds, params):
     print("%f (%f) with: %r" % (mean, stdev, param))
```