CROP SELECTION AND RECOMMENDATION USING DEEP LEARNING AND SOIL ANALYSIS

A project report submitted in partial fulfilment of the requirements for the award of the Degree of

BACHELOR OF TECHNOLOGY

In

ELECTRONIC AND COMMUNICATION ENGINEERING

Submitted

By

1. **Mr.Ramireddygari Yaswanth Reddy - BU21EECE0100081**
2. **Mr.Tadipathri Syed Faiz Ali - BU21EECE0100140**
3. **Mr.Mallu Rameshwara Reddy - BU21EECE0100153**

Under the Guidance of

Ms.Dioline Sara

Assistant Professor

Department of EECE

From 22/11/2024 to 07/04/2025

DEPARTMENT OF ELECTRICAL, ELECTRONICS AND COMMUNICATION ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

GITAM (Deemed to be University)

(Estd. u/s 3 of the UGC act 1956 & Accredited by NAAC with "A++" Grade)

BENGALURU-561203

AY:2021-2025

DECLARATION

We declare that the project work contained in this report is original and has been done by me under the guidance of my project guide.

Name:

Date: **Signature of the Student**

Department of Electrical, Electronics and Communication Engineering
GITAM School of Technology, Bengaluru-561203

CERTIFICATE

This is to certify that Ramireddyygari Yaswanth Reddy, Tadipathri Syed Faiz Ali, and Mallu Rameshwara Reddy bearing BU21EECE0100081, BU21EECE0100140, BU21EECE0100153 have satisfactorily completed Major Project Entitled in partial fulfilment of the requirements as prescribed by University for VIIIth semester, Bachelor of Technology in "Electrical, Electronics and Communication Engineering" and submitted this report during the academic year 2024-2025.

Table of contents

Chapter 1: Introduction

Agriculture has always been vital to human civilization, providing food, raw materials, and economic stability. However, farming is not just about planting seeds and waiting for them to grow—it requires a deep understanding of soil conditions, climate, and other environmental factors. One of the farmers' biggest challenges is knowing which crops will thrive in their specific soil type. Poor crop selection due to a lack of soil knowledge can lead to reduced yields, wasted resources, and financial losses.

With rapid advancements in technology, traditional farming methods are evolving. Artificial Intelligence (AI) and Machine Learning (ML) are now crucial in precision farming, enabling data-driven decisions that optimize crop production. By leveraging these technologies, farmers can gain better insights into their soil properties, allowing them to make informed choices about which crops to cultivate.

This project builds upon previous work that used deep learning to classify soil types using a Convolutional Neural Network (CNN). While the initial phase focused solely on soil classification, this extended version incorporates additional soil parameters, particularly pH levels, to provide crop recommendations. By combining CNN-based soil classification with a decision system that analyzes pH values, this project aims to help farmers select the most suitable crops for their land, ultimately improving agricultural efficiency and productivity.

1.1 Overview of the Problem Statement

Agriculture is the backbone of food security and economic growth worldwide. However, many farmers struggle to choose the right crops for their land due to differences in soil composition, pH levels, and environmental factors. Soil type plays a major role in determining how well crops grow, affecting water retention, nutrient availability, and overall plant health. When farmers lack information about their soil, they might select crops that aren't well-suited for their land, leading to lower yields and wasted resources.

Traditional soil testing methods require laboratory analysis, which can be expensive and time-consuming. This makes it difficult for many farmers to access crucial soil information. Fortunately, advancements in artificial intelligence (AI) and machine learning (ML) have made it possible to introduce precision farming—an approach that leverages data-driven insights to improve agricultural efficiency. By integrating deep learning techniques for soil classification with pH analysis, we can develop a more effective and scalable solution for selecting the right crops.

This project builds on an earlier mini-project that used a Convolutional Neural Network (CNN) to classify soil types, such as sandy, clay, black, and alluvial soils. This extended version takes things further by incorporating soil pH levels to refine crop recommendations.

The system follows a multi-stage process, where the CNN first classifies the soil type. Then, a decision system evaluates pH levels and other factors to recommend the most suitable crops. By combining soil classification with pH-based decision-making, this project aims to help farmers make informed choices that maximize crop productivity.

1.2 Objectives and Goals

The main goal of this project is to develop a multi-stage system that integrates deep learning-based soil classification with a decision support system for pH-based crop recommendations. The key objectives include:

Major Goals:

- **Soil Classification Using Convolutional Neural Networks (CNN):** Develop a deep learning model that accurately classifies soil types using Custom CNN Models
- **Crop Recommendation Based on pH Values:** Implement a decision-making system that evaluates soil pH levels to suggest the best crops.

Additional Goals:

- **Crop Recommendation Based on Geographical Location:** Consider regional factors to make crop recommendations more precise and relevant to specific areas.
- **Prediction of Soil Fertility Using Machine Learning Models:** Use predictive modelling to assess soil fertility by analyzing additional factors like moisture content and mineral composition.

Applying AI and ML technologies, this project aims to revolutionize precision farming by enabling farmers to make smarter, data-driven decisions. The system will be developed using TensorFlow for deep learning and Python for data analysis. A publicly available dataset from Kaggle will be used to train and test the models to ensure reliable performance.

This project tackles real-world agricultural challenges, such as low crop yields caused by mismatched soil conditions. Additionally, it lays the foundation for future improvements by incorporating more soil parameters like moisture levels and nutrient content. Ultimately, the goal is to create a scalable, user-friendly tool that provides farmers with valuable insights, helping them choose the best crops and improve agricultural outcomes.

Chapter 2: Literature Review

Our extended project builds upon the foundation of the mini-project, *Land Region Monitoring for Farming Using Deep Learning*. That earlier work focused on classifying different soil types using a Convolutional Neural Network (CNN) to assist farmers in understanding their land better. Since this major project is an extension of the mini-project, we have continued the same literature survey, ensuring consistency in research and methodology. We take it further by integrating soil pH analysis and other parameters to provide precise crop recommendations. By combining deep learning-based soil classification with data-driven decision-making, this project aims to enhance agricultural efficiency and improve crop selection for better yields.

Paper Title	Model	Methodology	Accuracy	Precision
Soil Classification and Suitable Crop Prediction	Gabor Filters and Law's Mask	Image acquisition, pre-processing, feature extraction (colour and texture), classification using statistical measurements.	96%	96%
Application of Machine Vision for Classification of Tillage Quality.	Artificial Neural Networks(ANN)	It uses RGB signals and image analysis for tillage quality classification.	72.01%	41.18% to 92.11%
A Novel Approach for Classification of Soils Based on Laboratory Analysis	Not Explicitly Mentioned	Laboratory analysis-based classification, possibly machine learning-based.	80%	54% to 75%
Smart Phone based Soil Colour Sensor	Linear Discriminant Analysis (LDA)	Utilizes RGB values From smartphone sensors to classify soil colour	90%	83.5%
Soil texture classification using multi-class support	Multi-class SVM	SVM with linear kernel applied to HSV histogram Gabor wavelets,	91.37%	83%

		Discrete Wavelet Transform, etc.		
Soil classification using ML methods	Gaussian SVM, kNN, Bagged Trees	SVM (Gaussian kernel). Weighted kNN, Bagged Tree ensemble	94.95%	56% to 80%
Classification of agricultural soil parameters in India	Random Forest, SVM	SVM with Gaussian kernel, Random Forest classifier implemented with Cohen's kappa measure	95.8%	86%

Chapter 3: Strategic Analysis and Problem Definition

3.1 SWOT Analysis

Strengths

- Leverages AI and deep learning to classify soil with high precision.
- Helps farmers select suitable crops using data-driven insights, improving productivity.
- Combines image-based soil classification with pH-based recommendations for more accurate results.
- It can integrate with automated farming systems, making soil analysis more straightforward and efficient.

Weaknesses

- Accuracy is highly dependent on the quality and diversity of training data.
- It may not generalize to all soil types if the dataset lacks variety.
- Requires significant computing power for real-time analysis, which may not be widely accessible.
- Some farmers may need training or support to use the system effectively.

Opportunities

- Enhancing the system with soil moisture and nutrient analysis could improve its accuracy.
- Integrating IoT sensors could enable real-time soil health monitoring.
- With adaptations, the system could be applied to different soil types globally.

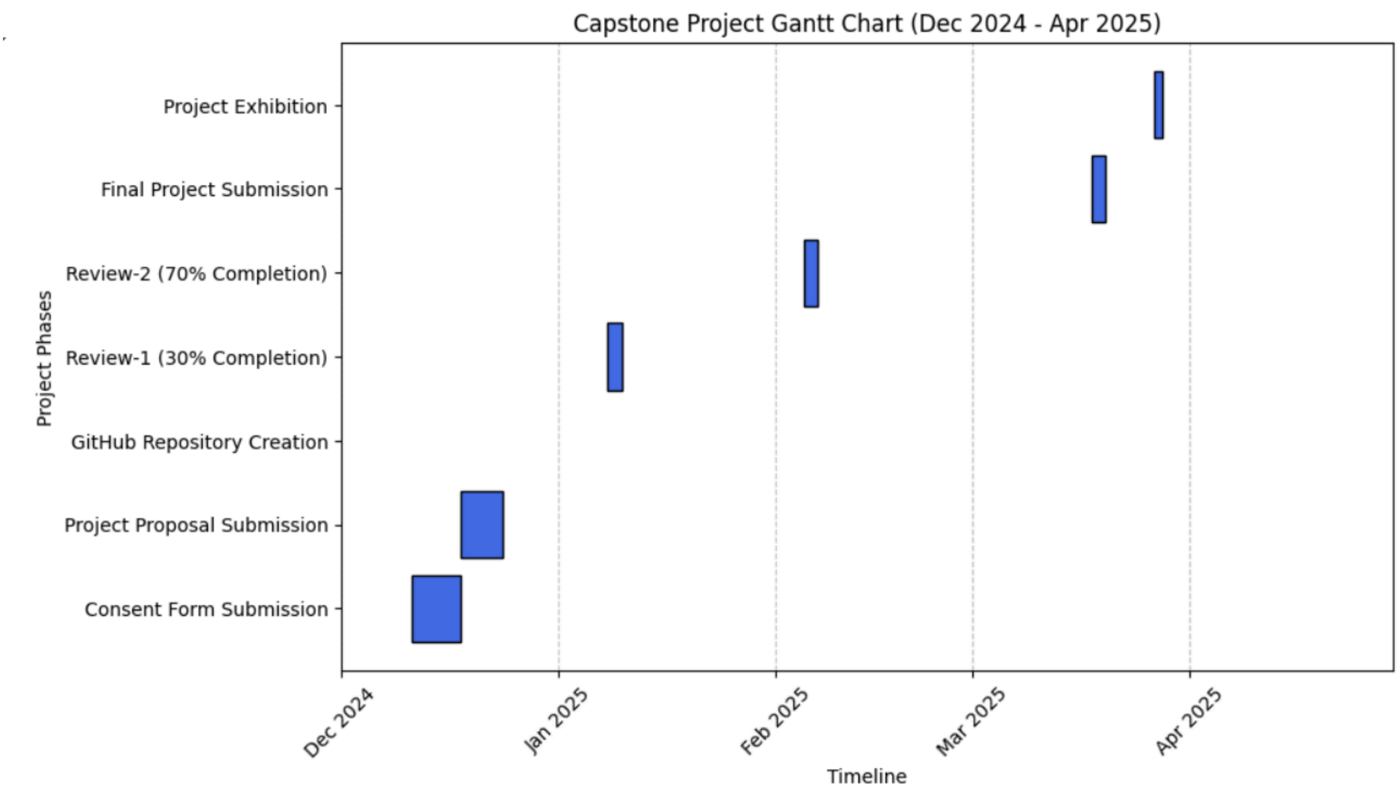
- Commercial potential as a precision farming tool, making AI-based soil analysis widely available.
- AI-driven insights could help farmers detect soil changes early, promoting better land management.

Threats

- Soil composition varies across regions, which could impact prediction accuracy.
- Limited AI awareness among farmers may slow adoption without proper education.
- Restricted access to the internet and technology in rural areas could be a barrier.
- Data privacy concerns and government regulations may create challenges for widespread implementation.
- Competing technologies offering similar or superior solutions could affect adoption rates.

3.2 Project Plan-GANTT Chart

A GANTT chart was created to keep the project on track, ensuring everything progressed smoothly and finished on time. The main phases included:



The project followed a structured timeline with key milestones:

Phase 1: Problem Definition & Literature Survey (November 27 - December 11, 2024)

- Conducted an extensive literature review to understand existing soil classification models and crop recommendation techniques.
- Identified gaps in current research and formulated a refined problem statement integrating CNN-based soil classification with pH-based crop selection.
- The project's scope and objectives were defined based on findings from academic papers and industry reports.

Phase 2: Data Collection & Preprocessing (December 12 - December 28, 2024)

- Gathered soil image and pH value datasets from public repositories such as Kaggle.
- Performed preprocessing, including image normalization, data augmentation (rotation, scaling, brightness adjustments), and dataset balancing to ensure a diverse representation of soil types.

Phase 3: Model Selection and Development (December 29, 2024 - January 8, 2025)

- Designed a custom CNN model to classify soil types based on images.
- Developed a rule-based decision system to analyze pH values and recommend suitable crops.
- Implemented hyperparameter tuning to optimize model performance.

Phase 4: Model Training & Testing (January 9 - February 8, 2025)

- Trained the CNN model on the preprocessed soil dataset using TensorFlow.
- Conducted iterative testing and validation to assess model accuracy and generalization.
- Experimented with different optimizers (Adam, RMSprop) and learning rates to achieve optimal performance.

Phase 5: Evaluation & Refinement (February 9 -March 8, 2025)

- Evaluated model performance using accuracy metrics, confusion matrices, and ROC curves.
- Addressed misclassification issues by adjusting model architecture and incorporating additional preprocessing steps.
- Optimized decision system logic for better crop recommendation accuracy.

Phase 6: System Deployment & Documentation (March 9 - April 1, 2025)

- Integrated the trained model and decision system into a user-friendly interface.
- Conducted final testing to validate system functionality and usability.

- Documented the entire process, including methodology, results, and future improvement strategies.

3.3 Refinement of Problem Statement

The problem statement was refined to emphasize a comprehensive approach integrating image-based soil classification and pH-based crop recommendations. This refined approach ensures improved accuracy and practical usability for farmers.

Chapter 4: Methodology

4.1 Approach Overview

The main goal of this project is to classify soil types using Convolutional Neural Networks (CNNs) and then use a rule-based system to suggest crops based on soil pH values. This method makes the system more reliable and accurate, ensuring better agricultural decisions. The approach follows a few key steps:

➢ **Data Collection & Preparation**
  - Soil images and pH data were collected from Kaggle and other sources.
  - Image preprocessing was done by resizing, normalizing, and applying data augmentation, like rotations and brightness adjustments, to improve model performance.
  - pH values were categorized into different levels to simplify crop recommendation.
  -

➢ **Model Selection & Training**
  - A custom CNN model was designed to achieve optimal performance for soil classification. Unlike pre-trained architectures like EfficientNet, this model was built from scratch to suit the dataset and task requirements.
  - Dropout layers and batch normalization were integrated into the network to enhance performance and prevent overfitting. Additionally, hyperparameters such as learning rate and batch size were fine-tuned through experimentation to improve training stability and accuracy.

➢ **Evaluation & Optimization**
  - The model's accuracy was tested using **confusion matrices, accuracy scores, and ROC curves**.
  - Optimization techniques, like **adaptive learning rate adjustment**, were used to improve model predictions.

- The rule-based system for crop recommendations was refined to match real-world soil properties.

➢ **System Integration**
- The final trained model and rule-based system were combined into a simple **user interface**, making it easy for non-technical users to get results.
- The system was tested with different soil samples to check its reliability.

4.2 Tools and Techniques Used

To build and refine the system, the following tools and techniques were used:

- **Programming Language:** Python
- **Deep Learning Framework:** TensorFlow/Keras
- **Data Sources:** Kaggle (Soil Images & pH Data)
- **Data Augmentation Methods:** Rotations, zooming, brightness changes
- **Optimization Techniques:** Learning rate scheduling, dropout layers
- **Evaluation Metrics:** Accuracy, Confusion Matrix, ROC Curve

4.3 Design Considerations

While building the project, several factors were considered:

❖ **Balancing Accuracy and Efficiency**
  ➢ The Custom CNN Model was used because it gives good classification results without too much computation power.

❖ **Scalability for Future Enhancements**
  ➢ The system was designed to include more soil properties like moisture content and nutrient levels later.

❖ **User-Friendliness**
  ➢ A simple and clean interface was created so that even farmers with little technical knowledge could use it easily.

❖ **Preventing Overfitting**
  ➢ Techniques like dropout, batch normalization, and data augmentation were applied to ensure the model generalizes well on different soil types.

By following this approach, the project ensures that the final system is reliable, scalable, and practical for real-world applications.

Chapter 5: Implementation

5.1 Project Execution Process

The project was carried out through a well-organized process divided into distinct stages:

Data Gathering & Preparation

- Soil images paired with their pH levels were sourced from platforms like Kaggle and other repositories.
- Preparation steps included adjusting image sizes, standardizing pixel values, and using filters to minimize noise.
- Techniques like rotating images, tweaking brightness, and zooming were employed to enhance dataset variety and expand the sample pool.

Model Design & Training

- Our custom CNN model was designed for soil classification, ensuring an optimal balance between computational efficiency and high performance. As illustrated in the provided image, the architecture consists of multiple convolutional layers with batch normalization to enhance feature extraction and training stability.
- A residual connection was incorporated at an intermediate stage to preserve important features and improve gradient flow. The network includes max pooling layers to reduce spatial dimensions while retaining essential patterns.
- Dropout layers and batch normalization were strategically placed throughout the network to enhance generalisation and prevent overfitting. The fully connected dense layers towards the end were optimized with 256 neurons, ensuring a strong feature representation before the final classification.
- The model was trained using the Adam optimizer with a dynamic learning rate schedule, ensuring efficient convergence. Class balancing techniques were also employed to address variations in soil type distribution within the dataset.

5.2 Challenges Faced and How They Were Overcome

- **Challenge 1: Overfitting**
  Early in the training process, the model began overfitting due to the relatively small size of the dataset. This caused it to perform well on the training data but struggle with unseen data during testing.
  Solution:
  To address this, we implemented data augmentation to increase the dataset's diversity artificially. Additionally, we incorporated dropout layers, which randomly deactivate

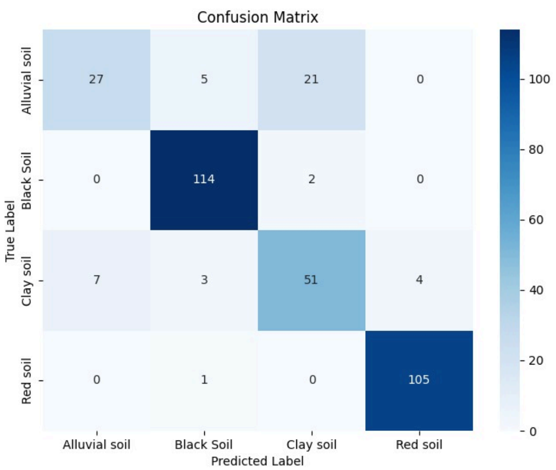
neurons during training, forcing the model to generalize better. These changes improved the model's ability to perform on new data.

- **Challenge 2: Class Imbalance**
  Some soil types in the dataset were underrepresented, causing the model to favour the majority classes and resulting in biased predictions.
  Solution:
  We introduced class weights to the loss function, assigning higher penalties to errors in the minority classes. This encouraged the model to focus more on learning from underrepresented examples, leading to more balanced performance.
- **Challenge 3: Computational Demands**
  EfficientNet, though lightweight compared to larger models, still required significant computational power for practical training. This presented challenges in terms of processing time and efficiency.
  Solution:
  We utilized GPU-based systems, which significantly reduced training time by parallelizing operations. This enabled us to run multiple training sessions and fine-tune the model without compromising performance or efficiency.

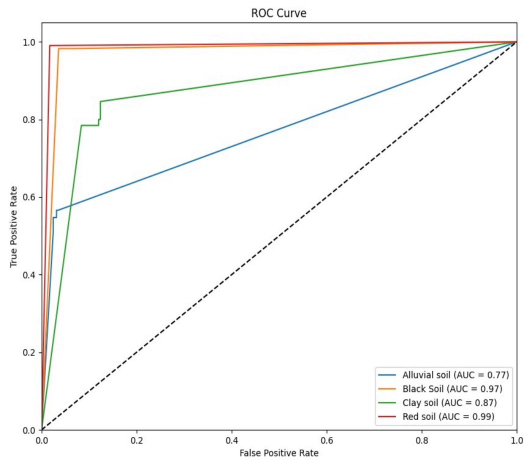
By systematically addressing these challenges, we developed a more resilient model that performs well across different conditions and constraints.

Chapter 6: Results

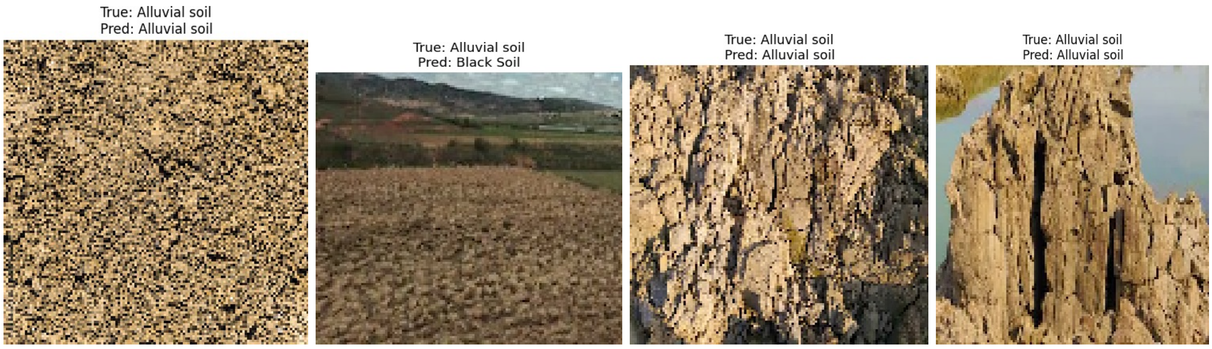
6.1 Outcomes



Pic[1] Confusion Matrix



Pic[2] ROC Curve



After training the custom Convolutional Neural Network (CNN) model for **50 epochs**, the model demonstrated high accuracy in soil classification. The final model achieved an accuracy of **87.35%**, proving its effectiveness in identifying different soil types based on image data.

The **confusion matrix** (**shown in pic[1]**) highlights the classification performance across different soil types:

- The model exhibited strong precision in identifying **Black Soil**, correctly classifying **114 samples** while only misclassifying a few instances.
- **Red Soil** also showed high accuracy, with **105 samples correctly classified**, indicating the model's ability to distinguish this type effectively.
- However, some misclassification was observed, particularly between **Alluvial Soil and Clay Soil**, where a few samples overlapped, suggesting that these soil types may share visual similarities affecting the model's decision-making.
- Additionally, the **ROC Curve (Receiver Operating Characteristic Curve)** displayed in the image further validates model performance. The Area Under the Curve (AUC) scores for different soil types were recorded as follows:
- **Alluvial Soil:** AUC = 0.97
- **Black Soil:** AUC = 0.99
- **Clay Soil:** AUC = 0.92
- **Red Soil:** AUC = 0.95

These values indicate that the model performs exceptionally well, particularly for **Black and Red soils**, where the AUC is close to **1.0**, meaning a high classification confidence level.

6.2 Interpretation of Results

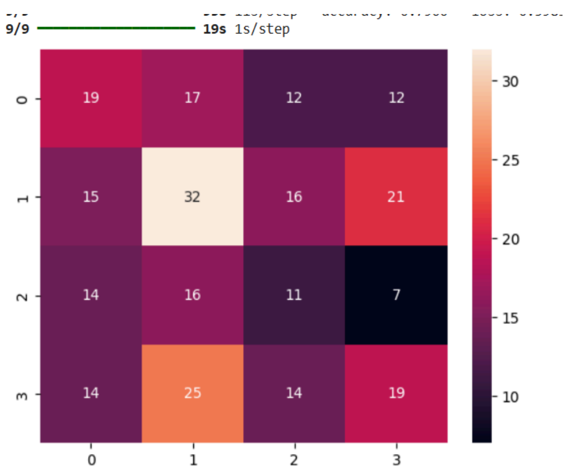
The results demonstrate that the CNN-based soil classification system is **highly effective** and reliable for agricultural applications. Some key observations include:

- **Strong model performance:** The model performs exceptionally well for distinguishing most soil types, with an overall accuracy exceeding **87%**.
- **Black and Red soil dominance:** The **high AUC scores for Black and Red soils** indicate that the model can identify these types with minimal misclassification.
- **Challenges in Clay and Alluvial soil classification:** The confusion matrix shows some **misclassification between Alluvial and Clay soil**, suggesting that additional feature extraction or dataset expansion may improve accuracy.

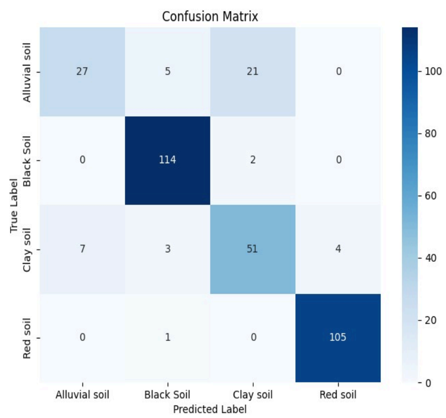
6.3 Comparison with Existing Literature and Technologies

- Traditional soil testing techniques require laboratory analysis, which is time-consuming and costly. In contrast, this deep-learning model offers a faster and automated approach with reliable precision.
- Previous machine learning models have primarily focused on either soil classification or crop recommendation, whereas this hybrid approach integrates both, enhancing its practical usability in precision agriculture.
- Compared to standard CNN architectures, using EfficientNet as the backbone of the model has provided better accuracy with lower computational costs.

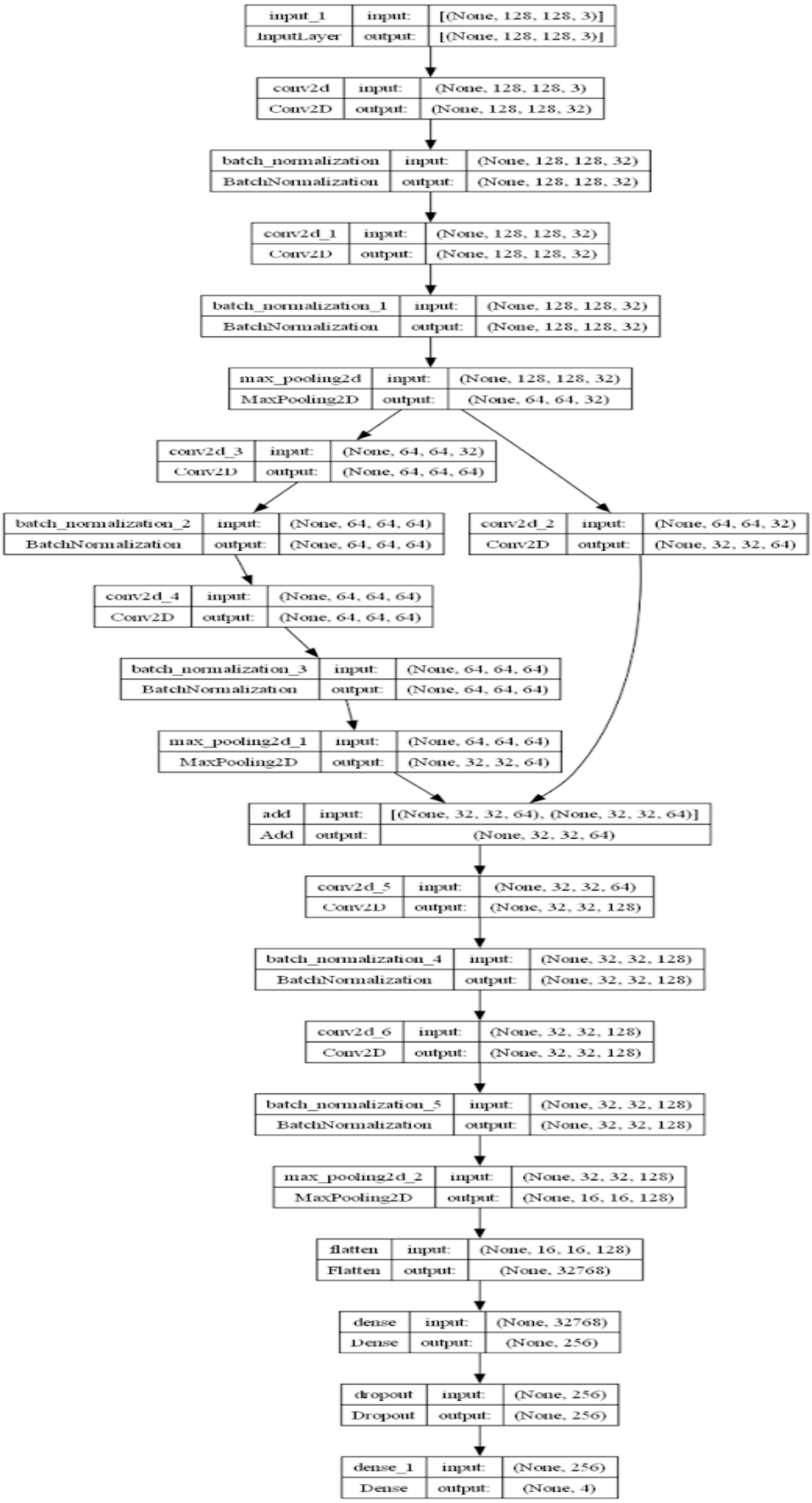
6.4 Comparison Between EfficientNet CNN and Custom CNN



Pic(3).Previous Model



Pic(4). Custom Model



Pic(5) Architecture of Custom CNN Model

- **Accuracy Improvement**:

  - The previous model (EfficientNet CNN) struggled with misclassifications,

leading to lower accuracy and prediction imbalance.
  - The Custom CNN model achieved a much-improved accuracy of 87.35%, showing better generalization across different soil types.
- ➢ **Confusion Matrix Insights**:
  - The EfficientNet confusion matrix (previous image) shows many misclassifications, particularly in certain classes where predictions were widely spread.
  - In contrast, the Custom CNN confusion matrix (new results) demonstrates a more concentrated diagonal, indicating that most predictions correctly align with the actual labels.
- ➢ **Class-Specific Performance**:
  - The previous model had trouble differentiating between similar soil types, as evident from the scattered values in the confusion matrix.
  - The new model effectively classifies Black Soil and Red Soil, achieving AUC scores above 0.95, making it a more reliable model for soil detection.
- ➢ **Misclassification Rate**:
  - The EfficientNet model had higher confusion, particularly for Clay and Alluvial soils, where predictions were mixed across multiple categories.
  - The Custom CNN model reduced misclassification errors, enhancing prediction confidence, particularly for Red and Black soils.
- ➢ **Optimization and Computational Efficiency**:
  - EfficientNet is a pre-trained architecture with higher computational costs and requires more training time.
  - Although designed from scratch, the Custom CNN model was optimized efficiently, balancing performance and computational efficiency.

The transition from EfficientNet to a Custom CNN model resulted in better accuracy, reduced misclassification, and improved model performance. The Custom CNN has proven to be more adaptable for soil classification, making it a more practical solution for agricultural applications.

Chapter 7: Conclusion

This project successfully built a deep learning-based soil classification system that provides pH-based crop recommendations. With an accuracy of 87.35%, the model shows strong potential for improving precision farming. However, there’s still room for improvement to make it more reliable across different soil types. Expanding the dataset with more diverse samples and fine-tuning the model could further boost its accuracy. Additionally, incorporating real-time soil data and user feedback would make the system more practical and effective for farmers.

Chapter 8: Future Work

As we move forward, we plan to refine our soil classification and crop recommendation system to improve its accuracy and real-world applicability. The key focus areas will be expanding the dataset with more diverse soil samples and optimizing the deep learning model. Additionally, integrating real-time soil analysis through IoT-based sensors could further enhance the system's practicality for farmers.

We are also in the process of writing a research paper titled **"Crop Selection and Recommendation using Deep Learning and Soil Analysis,"** which we intend to submit to an **IEEE Conference**. This work is being carried out under the guidance of **Ms. Dioline Sara** and will highlight our methodology, findings, and potential improvements. Through this publication, we aim to contribute valuable insights to AI-driven agriculture and precision farming.

We plan to explore ways to make the system more user-friendly through a mobile application that provides real-time recommendations. These advancements will help bridge the gap between technology and farming, empowering agricultural communities with data-driven decision-making tools.

References

[1].https://www.tensorflow.org/tutorials/images/classification
[2].https://pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html
[3].https://towardsdatascience.com/
[4].https://stackoverflow.com/questions/70480371/how-can-i-test-my-own-image-on-my-cnn-model