International Journal of Research Publication and Reviews, Vol (5), Issue (11), November (2024), Page – 3948-3954

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Soil Classification Using Convolutional Neural Network

S.Yaswanth Kiran[1],T.Prasuna[2],T.Chaitanya Laxmi[3],V.Bharath[4],S.Kethan Babu[5]

GMR Institute of Technology,4th year civil,vizianagaram,532127,India

ABSTRACT :

Soil classification plays a crucial role in identifying different soil types. Traditional classification methods are often slow and prone to human error, so this project uses Convolutional Neural Networks (CNNs) for automated soil classification through image recognition. By training the CNN on a dataset of labeled soil images, it can accurately classify soils based on visual features such as texture and color. This approach provides a faster, more consistent alternative to manual classification, enhancing efficiency and accuracy in soil type identification. The project begins with the collection and preprocessing of a diverse dataset of soil images, covering various soil types such as sandy, clayey, loamy, and silty soils. The images are labeled and augmented to enhance model generalizability. A CNN architecture is then designed and trained on this dataset, with parameters optimized to achieve high accuracy in soil type identification. Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate the model, ensuring its reliability in real-world application. This CNN-based soil classification system aims to provide a faster and more accurate alternative to traditional methods, reducing manual effort and increasing consistency.

Keywords: Soil Classification, Convolutional Neural Networks (CNN), Image Recognition, Deep Learning, Soil Types, Soil Image Dataset, Machine Learning, Soil Identification, Model Optimization

INTRODUCTION :

- Traditional soil classification methods often rely on manual sampling and laboratory analysis, which can be time-consuming, labour-intensive, and subject to human error. These methods may also lack the spatial resolution needed to make comprehensive assessments across large geographic areas. Furthermore, they typically require expertise in soil science and may not always accommodate the variability found in complex landscapes. To address these challenges, there is a growing need for automated and efficient soil classification techniques that can leverage advances in machine learning and remote sensing technologies.

- Convolutional Neural Networks (CNNs) are a type of deep learning architecture particularly well-suited for image analysis tasks. They excel at identifying patterns and features in visual data through a process called convolutions. In the context of soil classification, CNNs can analyse images of soil profiles, core samples, or aerial photographs to classify soils based on visual characteristics. Their ability to learn features directly from data without the need for explicit feature engineering sets CNNs apart from traditional algorithms, making them a powerful tool for automating the soil classification process.

- The application of CNNs in soil classification involves several steps. First, high-quality images of soil samples are collected, often from different regions and under various conditions. These images are then pre-processed to normalize lighting, scale, and resolution. Next, a convolutional neural network is designed and trained using labelled datasets, allowing it to learn the features associated with different soil types. Once trained, the CNN can accurately classify unseen soil images based on the patterns it has recognized, significantly enhancing the speed and accuracy of soil classification processes.

- Implementing CNNs for soil classification presents numerous advantages, including increased efficiency, improved accuracy, and the ability to analyse large datasets in real-time. The integration of CNNs with remote sensing technologies, such as drones and satellite imagery, allows for large-scale soil mapping and monitoring applications that were previously unattainable. As research in this field continues to advance, future developments may focus on refining CNN architectures, improving data acquisition methods, and expanding to 3D image analyses. Moreover, the results from CNN-based classifications can be integrated with geographic information systems (GIS) and other data sources to enhance environmental modelling and sustainable land management practices.

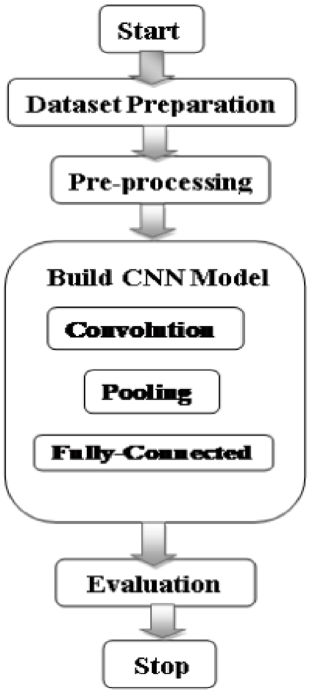
Literature Review :

- Pallavi srivastava (2021) reported that; the study on how A comprehensive review of soil classification using deep learning and computer vision techniques emphasizes the effectiveness of convolutional neural networks (CNNs) in automating soil analysis. The methodology begins with the collection of high-resolution soil images from various locations, capturing diverse soil types and conditions. These images undergo preprocessing, which includes normalization, data augmentation, and noise reduction to enhance image quality and ensure robust model training. Next, CNN architectures are designed or selected based on the complexity of the classification task, focusing on feature extraction capabilities. The selected models are trained on labelled datasets, allowing them to learn distinctive visual patterns associated with different soil classes. Performance is evaluated using metrics like accuracy, precision, and recall. Advanced techniques, such as transfer learning, may also be employed to improve classification performance with limited datasets. This integration of deep learning and computer vision significantly enhances the efficiency and accuracy of soil classification, paving the way for better agricultural and environmental management practices.

- E.P.B. Guidang (2019) reported that; the study on how utilizing pre-trained deep learning models to classify soil texture images effectively. The process begins with the collection of a dataset containing high-quality images of various soil textures, ensuring a diverse representation of soil types. The images are then pre-processed to standardize dimensions and enhance quality through normalization and data augmentation techniques, such as rotation and flipping, to increase the dataset's diversity. The study employs transfer learning by leveraging established models like VGG16 or ResNet, which have been trained on extensive image datasets. These models are fine-tuned using the soil texture dataset, allowing them to adapt to the specific features of soil images. The performance of the model is assessed using metrics such as accuracy and F1 score, demonstrating the effectiveness of transfer learning in improving the classification of soil textures while reducing training time.
- Shinya inazumi (2020) reported that; the study on how in the study involves a systematic approach to developing an AI-based system for soil classification. The process starts with the collection of diverse soil samples, ensuring representation across various types and conditions. Soil characteristics are assessed using various techniques, including physical, chemical, and biological analyses, which create a robust dataset of features relevant to classification. Next, machine learning algorithms, such as decision trees, support vector machines, or neural networks, are deployed to model the relationships between soil characteristics and their classifications. The system incorporates feature selection and engineering to identify the most significant attributes for accurate classification. After training the models on the dataset, their performance is evaluated using metrics like accuracy, precision, and recall. The developed AI system aims to assist soil scientists and agricultural practitioners by providing reliable classifications, enhancing decision-making processes in soil management and utilization.
- Matthew veres (2017) reported that; the study focuses on utilizing deep learning techniques to predict various soil properties based on diverse datasets. The process begins with the collection of extensive soil data, including physical, chemical, and biological characteristics from various locations. Data preprocessing is conducted to clean and normalize the dataset, ensuring that missing values are addressed and features are scaled appropriately. The study employs several deep learning architectures, such as feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), to explore their effectiveness in predicting soil properties. Training is performed using labelled datasets, and model performance is evaluated through metrics like mean absolute error (MAE), root mean square error (RMSE), and R-squared values. The results indicate that deep learning models can significantly enhance soil property prediction accuracy, providing valuable insights for agricultural and environmental applications.
- Shahbaz Khan Baloch (2020) reported that; the study focuses on utilizing conventional neural networks (CNNs) for the accurate classification of soil types. The process begins with the collection of a comprehensive dataset consisting of soil samples, including images and associated features. Data preprocessing is crucial, involving image enhancement techniques, normalization, and segmentation to improve the overall dataset quality. Subsequently, CNN architectures are implemented to learn complex patterns and features within the soil images. The models are designed with multiple layers to extract hierarchical features, facilitating effective classification. Training involves splitting the dataset into training, validation, and testing subsets, allowing the model to learn and validate its performance iteratively. The study evaluates model effectiveness using metrics such as accuracy, precision, and confusion matrices. The results demonstrate that conventional neural networks can significantly improve soil classification accuracy, promoting more efficient soil management practices.
- Ritual tharkur (2018) reported that; the study focuses on creating a robust classification model for Indian soils using diverse machine learning algorithms. The research begins with the collection of a detailed dataset, which includes soil samples sourced from various regions across India, encompassing a wide range of soil properties such as texture, pH, organic carbon, and nutrient content. Data preprocessing steps are implemented to clean the dataset, handle missing values, and normalize the features for consistency. Various machine learning techniques, including decision trees, random forests, support vector machines, and k-nearest neighbors, are employed to develop classification models. The models are trained and validated using cross-validation techniques to optimize performance. Evaluation metrics, such as accuracy, precision, recall, and F1 score, are used to assess the effectiveness of each algorithm. The findings reveal that machine learning can significantly enhance the accuracy and efficiency of Indian soil classification.
- Emmanuel Kwabena Gyasi and Swarnalatha Purushotham (2011) reported that; the study focuses on employing convolutional neural networks (CNNs) for effective soil classification. The process begins with the collection of soil samples accompanied by high-resolution images, which capture various soil morphologies and their geospatial contexts. Data preprocessing is undertaken to prepare images for analysis, including normalization, resizing, and data augmentation techniques to enhance model robustness and generalization. The CNN architecture is designed to automatically extract features from the input images, with multiple convolutional and pooling layers capturing intricate patterns associated with different soil types. The model is trained on labelled datasets using cross-entropy loss, optimizing performance through techniques like dropout for regularization. Evaluation is conducted using metrics such as accuracy, sensitivity, and specificity, demonstrating the CNN model's effectiveness in classifying soils based on morphology and spatial locations.
- Chandan (2017) reported that; the study focuses on developing an effective machine learning model for classifying soil images. The research begins with the collection of a diverse dataset of soil images, ensuring representation of various soil types and conditions. Data preprocessing is critical, involving steps such as image resizing, normalization, and enhancement to improve image quality and consistency. Additionally, data augmentation techniques like rotation, flipping, and scaling are applied to increase dataset variability and robustness. The study employs several machine learning algorithms, including support vector machines (SVM), decision trees, and random forests, to assess

their performance in soil image classification. Models are trained and evaluated using predefined metrics such as accuracy, precision, and recall, with cross-validation to ensure reliability. The results indicate that the intelligent machine learning model significantly improves the accuracy and efficiency of soil image classification, aiding in soil management practices.

Methodology :



1. **Data Collection**:
   - Soil Sample Collection: Collect samples from various locations to ensure a diverse dataset.
   - Image Acquisition: Capture images of the soil samples using high-quality cameras. It's important to standardize the lighting conditions and background to ensure image consistency.

2. **Data Preprocessing**:
   - Image Processing: Apply techniques such as resizing, normalization, and augmentation (including rotations, flipping, and color adjustments) to improve the dataset's robustness.
   - Labeling: Ensure that each image is accurately labeled according to the classification criteria (e.g., soil type, texture).

3. **Dataset Splitting**:
   - Divide the dataset into training, validation, and testing subsets, often following a common split such as 70% training, 15% validation, and 15% testing.

4. **Model Selection and Architecture**:
   - CNN Architecture: Choose an appropriate CNN architecture (e.g., Alex Net, VGG16, ResNet). Some studies may also create custom architectures tailored to the specific characteristics of the soil image data.
   - Transfer Learning: If necessary, leverage pre-trained models to take advantage of previously learned features, which can help in cases where the dataset is small.

5. **Training the Model**:
   - Hyperparameter Tuning: Set and optimize hyperparameters such as learning rate, batch size, number of epochs, etc.
   - Loss Function and Optimizer: Select appropriate loss function (e.g., categorical cross entropy for multi-class classification) and optimizer (e.g., Adam, SGD).
   - Use the training dataset to train the model while validating performance on the validation set to avoid overfitting.

6. **Model Evaluation**:
    - After training, evaluate the model's performance using the testing dataset. Common metrics include accuracy, precision, recall, F1-score, and confusion matrix analysis.

7. **Results Analysis**:
    - Analyze the classification results to identify which classes are accurately predicted and which are confused with others.

8. **Discussion and Conclusion**:
    - Discuss the implications of the findings, compare results with other existing methods, and consider potential applications in fields such as agriculture, environmental monitoring, or urban planning. Suggestions for future work may also be included.

9. **Visualization**:
    - Include visualizations of the training/validation losses, sample predictions versus true labels, and potentially class activation maps to understand what the model is focusing on.

RESULTS AND DISCUSSIONS:

The implementation of a convolutional neural network (CNN) for soil classification using TensorFlow and Keras. The model is designed to classify images of soil into various categories. It highlights the effectiveness of CNNs in classifying soil types based on image data. Summarize the achieved accuracy and other performance metrics, demonstrating how the model performed on the test dataset compared to traditional classification methods.

```
In [5]: train_ds.class_names
Out[5]: ['Clayey_soils', 'Gravel', 'Soil_Family']

In [6]: (train_ds)
Out[6]: <_PrefetchDataset element_spec=(TensorSpec(shape=(None, 128, 128, 3), dtype=tf.float32, name=None), TensorSpec(shape=(None,), d
        type=tf.int32, name=None))>

In [7]: len(train_ds)
Out[7]: 16
```

Emphasize how this automated approach can save time and improve accuracy in soil classification compared to manual methods.

```
# Retrieve a batch of images and labels from train_ds
for images, labels in train_ds.take(1):  # Take one batch from the dataset
    plt.figure(figsize=(5, 5))
    plt.imshow(images[0].numpy().astype("uint8"))  # Convert image to numpy format and plot
    plt.title(train_ds.class_names[labels[0]])     # Get label name from class_names
    plt.axis("off")
    plt.show()
```

Present the contributions of the study to the existing body of knowledge in soil science, machine learning, and remote sensing. Mention any novel methodologies, techniques, or insights that were developed or discovered during the research.

To present the main results of a convolutional neural network (CNN) for soil classification effectively with images, you typically would visualize several key aspects of your model's performance.

1. **Training and Validation Accuracy and Loss**:

- Visualizing the training and validation accuracy over epochs helps in understanding how well the model is learning. A plot showing the loss can indicate whether the model is overfitting or underfitting.

Examples of Plots:

```python
import matplotlib.pyplot as plt

# Assuming history is the output of model.fit()
plt.figure(figsize=(12, 5))

# Plot training & validation accuracy values
plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'], label='Training Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.title('Model Accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(loc='upper left')

# Plot training & validation loss values
plt.subplot(1, 2, 2)
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('Model Loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(loc='upper right')

plt.show()
```

These plots will typically show how accuracy improves with each epoch and whether the loss decreases, indicating effective learning.

2. **Sample Predictions**:

- Display sample images alongside their predicted and actual labels to qualitatively assess the model's performance.

This section will aid in visually assessing how accurately the model classifies different soil types.

```python
import numpy as np
from tensorflow.keras.preprocessing import image

# Assuming `model` is trained and `x_test` is a set of test images
predictions = model.predict(x_test)
predicted_classes = np.argmax(predictions, axis=1)

# Visualizing sample results
plt.figure(figsize=(12, 10))
for i in range(9):
    plt.subplot(3, 3, i + 1)
    plt.imshow(x_test[i].astype("uint8"))
    plt.title(f"Predicted: {class_names[predicted_classes[i]]}\nTrue: {class_names[np.argmax(y_test[i])]
    plt.axis('off')
plt.show()
```

DISCUSSIONS:

- **Effectiveness of Architecture**: The chosen architecture, with multiple convolutional and pooling layers, is well-suited for image classification tasks. The use of dropout layers helps mitigate overfitting, which is a common challenge in deep learning, especially with limited datasets.
- **Batch Size**: A batch size of 8 is relatively small, which may lead to more noisy gradient estimates but can help in generalization. It allows the model to update weights more frequently, which can be beneficial in certain scenarios.
- **Image Preprocessing**: Rescaling the images ensures that the model trains effectively by standardizing the input data. This preprocessing step is crucial for the convergence of the training process.
- **Potential Improvements**: Future work could explore data augmentation techniques to artificially expand the dataset, which can improve model robustness. Additionally, experimenting with different architectures or hyperparameters (like learning rate, number of epochs, etc.) could yield better performance.

5. Conclusions:

The CNN model is well-structured for the task of soil classification, with appropriate layers for feature extraction and classification. The architecture, including dropout layers, aims to enhance generalization and reduce overfitting. The performance on a validation set to assess accuracy and other metrics. Soil classification using Convolutional Neural Networks (CNNs), the conclusion typically summarizes the key findings, implications, and potential for future research.

1.**Summary of Key Findings**:

- Highlight the effectiveness of CNNs in classifying soil types based on image data.
- Summarize the achieved accuracy and other performance metrics, demonstrating how the model performed on the test dataset compared to traditional classification methods.

2. **Implications of the Results**:

- Discuss the significance of these findings for agricultural practices, land management, and environmental monitoring.
- Emphasize how this automated approach can save time and improve accuracy in soil classification compared to manual methods.

3. **Contributions to the Field**:

- Present the contributions of the study to the existing body of knowledge in soil science, machine learning, and remote sensing.
- Mention any novel methodologies, techniques, or insights that were developed or discovered during the research.

4. **Limitations**:

- Acknowledge any limitations encountered in the study. This might include sample size, diversity of soil types, or challenges in image acquisition and processing.
- Discuss the potential biases and variability in the dataset that could affect generalizability.

5. **Future Work**:

- - Suggest areas for future research, such as exploring other neural network architectures, incorporating additional features (e.g., spectral data), or applying the model to different geographical areas.
- Recommend further validation studies or real-world applications that could enhance the robustness of the findings.
- Highlight the possibility of integrating CNNs with other machine learning techniques or deep learning frameworks.

6. **Final Remarks**:

- Conclude with a strong statement about the potential of using CNNs for soil classification and their role in advancing precision agriculture, environmental studies, and sustainability efforts.
- Encourage continued research and the adoption of advanced technologies in soil science for better decision-making and resource management.

Acknowledgements

The authors wish to acknowledge M/s GMR Institute of Technology for the moral support.

REFERENCES :

1. K. Alary et Pretreatment of soil samples rich in short-range-order minerals before particle-size analysis by the pipette method Pedosphere(2013).
2. P. Han et al.A smartphone-based soil color sensor: For soil type classification Computers and Electronics in Agriculture (2016).
3. D. Fomin et al.Energy-based indicators of soil structure by automatic dry sieving Soil and Tillage Research (2021).
4. A Gómez-Ríos et al.Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation
Expert Systems with Applications (2019).
5. P. Han et al.A smartphone-based soil color sensor: For soil type classification Computers and Electronics in Agriculture (2016).
6. M.S. Islam et al.IncePTB: A CNN based classification approach for recognizing traditional Bengali games Procedia Computer Science (2018).
7. S. Liu et al.Toward a “Green Revolution” for Soybean Molecular Plant (2020).
8. A Michele et al.Mobilenet convolutional neural networks and support vector machines for palmprint recognition Procedia Computer Science (2019).
9. A Rehmer et al.On the vanishing and exploding gradient problem in Gated Recurrent Units IFAC-PapersOnLine (2020).
10. J. von der Goltz et al.Health Impacts of the Green Revolution: Evidence from 600,000 births across the Developing World Journal of Health Economics
(2020)
11. J. Yang et al.CodnNet: A lightweight CNN architecture for detection of COVID-19 infection Applied Soft Computing (2022).
12. W.J. Zhang et al.On definition of deep learning World Automation Congress (WAC) (2018).
13. A. Krizhevsky et al.ImageNet Classification with Deep Convolutional Neural Networks Advances in Neural Information Processing Systems
(2012)
14. Azizi et al.Classification of soil aggregates: A novel approach based on deep learning Soil and Tillage Research (2020).
15. Benjamin E. Backus. (2018). Soil Hydrometer Testing, Easy Guide for Hydrometer Analysis - Gilson Co....
16. Cervantes-Godoy, D., Dewbre, J., Amegnaglo, C. J., Soglo, Y. Y., Akpa, A. F., Bickel, M., Sanyang, S., Ly, S., Kuiseu,...
17. K. Chatterjee M.S. Obaidat D. Samanta B. Sadoun S.K.H. Islam R. Chatterjee Classification of Soil Images using...
18. Z. Chen et al.A lightweight CNN-based algorithm and implementation on embedded system for real-time face recognition Multimedia Systems
(2023).
19. J. Cheng et al.Quantized CNN: A unified approach to accelerate and compress convolutional networks IEEE Transactions on Neural Networks and Learning Systems (2017).
20. N.K. Chowdhury et al.ECOVNet: A highly effective ensemble based deep learning model for detecting COVID-19 PeerJ Computer Science (2021).