

An ensemble deep learning approach for soil pH classification leveraging GLCM-based texture features and neural networks

Sujitranan Mungklachaiya¹ and Anongporn Salaiwarakul^{2*}

¹Program in Computer and Information Technology, Loei Rajabhat University, Loei 42000, Thailand

²Department of Computer Science and Information Technology, Naresuan University, Phitsanulok 65000, Thailand

ABSTRACT

***Corresponding author:**
Anongporn Salaiwarakul
anongporns@nu.ac.th

Received: 18 February 2025

Revised: 27 October 2025

Accepted: 2 November 2025

Published: 9 September 2026

Citation:
Mungklachaiya, S., &
Salaiwarakul, A. (2026). An
ensemble deep learning
approach for soil pH
classification leveraging GLCM-
based texture features and
neural networks. *Science,
Engineering and Health
Studies*, 20, 26020002.

Soil pH-level classification is a critical challenge in precision agriculture owing to data scarcity, high intra-class similarity, and subtle visual variations between pH levels. This research proposed an ensemble-feature deep learning framework that integrates convolutional neural networks with gray-level co-occurrence matrix (GLCM) texture descriptors to address fundamental limitations in pH-level classification. Through comprehensive experiments, we evaluated diverse architectures (i.e., InceptionV3, Inception ResNetV2, ResNet50, and VGG16) enhanced by GLCM features, demonstrating a substantial increase in classification accuracy, particularly under minimal visual distinctions between pH levels. The optimal Inception ResNetV2 w/GLCM configuration achieved the best overall performance, with an F1-score of 82.10%, an accuracy of 82.22%, and a precision of 82.01%, outperforming the conventional machine learning and deep learning approaches evaluated in this study. The proposed method innovatively provides a synergistic integration of hierarchical deep learning features with statistical texture analysis, achieving robust discrimination of visually similar soil sample images capturing different pH levels. These findings advance automated soil analysis and contribute to precision agriculture by establishing a robust solution for image-based soil pH-level classification.

Keywords: soil pH-level classification; deep learning; ensemble feature; gray-level co-occurrence matrix (GLCM); agricultural image analysis

1. INTRODUCTION

Recent major advancements in computer vision and deep learning are reshaping analytics in agriculture. For instance, convolutional neural networks (CNNs) have demonstrated efficacy in diverse applications, including crown flower quality evaluation (Pramote et al., 2023), plant disease classification through hybrid learning (Phiphatkamtorn & Jitanan, 2025), crop anomaly detection

(Mendoza-Bernal et al., 2024; Ramos-Ospina et al., 2024), and rice leaf blast diagnosis (Shah et al., 2023). These advances highlight the versatility and adaptability of CNNs across agricultural applications.

The soil pH level plays a pivotal role in agricultural productivity, directly influencing nutrient solubility, microbial activity, and crop yield potential. Despite their accuracy, conventional methods for soil pH measurement rely on time-consuming and resource-intensive chemical

analyses and laboratory procedures (Neina, 2019). Consequently, they hinder rapid and large-scale monitoring, an increasingly critical need given the ever-expanding demand for food security and sustainable agricultural practices (Srivastava et al., 2023).

To improve pH measurement, digital image processing and multivariate image analysis provide sustainable alternatives to laboratory-based soil testing (Sharma & Kumar, 2018). Prior studies have demonstrated the utility of machine learning models, such as support vector machines (Barman & Choudhury, 2020), and proximal sensing technologies (e.g., Nix Pro color sensing) coupled with random forest classifiers (Mancini et al., 2020). More recently, deep learning has expanded the possibilities of soil analysis. For example, MobileNet shows strong performance in soil-type classification (Salaiwarakul, 2023), while Light-SoilNet is a lightweight CNN that addresses data imbalance, which is a persistent challenge in soil science (Pandiri et al., 2024). Nevertheless, reliable pH-level classification remains difficult owing to data scarcity, high intra-class similarity, and subtle visual differences that characterize soils with varying pH levels. Although an attention-X mechanism has been adopted to mitigate issues related to inter-class similarity and intra-class variance (Mungklachaiya & Salaiwarakul, 2025), it cannot fully resolve soil-specific challenges in pH-level classification. A bag-of-visual-words method has been applied to extract discriminative features (Mungklachaiya & Salaiwarakul, 2025), but further refinement is required to address soil-specific challenges.

Unlike soil-type classification or plant disease detection, where visual differences are pronounced, soil pH variations are subtle and nuanced, being easily overlooked by conventional CNNs. Moreover, data imbalance and limited sample diversity further hinder robust and generalizable soil pH-level classification. We aimed to address these limitations by integrating the complementary strengths of deep learning and conventional texture-based analysis. Specifically, we proposed a feature ensemble method that combines CNNs with handcrafted descriptors derived from the gray-level co-occurrence matrix (GLCM). This ensemble approach is intended to capture both high-level spatial representations from CNNs and fine-grained texture variations that are critical for distinguishing visually similar soil sample images with different pH levels. For the proposed method, rigorous experimentation was conducted using multiple CNN backbones (i.e., InceptionV3, Inception ResNetV2, ResNet50, and VGG16) enriched with GLCM features in a systematic analysis of optimal configurations for soil pH-level classification.

This study may advance image-based soil analysis and contribute to precision agriculture through a scalable, technology-driven solution that complements and extends conventional laboratory techniques.

2. MATERIALS AND METHODS

2.1 The research framework

Figure 1 illustrates the proposed soil classification method based on complex image processing and multimodal feature extraction. The method establishes a systematic procedure that begins with image quantization and standardization as a foundation for robust soil analysis and preserving essential soil features during dimensionality reduction.

The architecture involves dual-stream feature extraction. The main stage employs transfer learning with pretrained deep learning models, and a final global average pooling (GAP) vector encapsulates high-level semantic information. The second phase extracts GLCM features from the data and performs feature normalization. GLCM features enable statistical texture analysis and are crucial for capturing spatial variations in gray-level intensities. As a result, valuable insights into the texture patterns potentially associated with variations in soil pH level can be unveiled. The GLCM pathway employs rigorous normalization to ensure feature consistency and accurate representation of the feature space. This facilitates the visualization of intricate interconnections across pixels.

Finally, the proposed method performs feature fusion to integrate complementary features from both sources. Hence, statistical texture descriptors calculated from the GLCM are combined with hierarchical features extracted through deep learning. This creates a comprehensive representation that encompasses both global and local features of soil sample images. For classification, the enhanced feature set is processed by advanced machine learning methods to provide reliable pH-level classification. Comprehensive metrics confirm the effectiveness of the proposed method, emphasizing the benefits of its integrated approach to soil classification.

The proposed method is expected to advance automated soil analysis by integrating texture analysis with deep learning to achieve superior classification performance.

2.2 Soil sampling

Given the spatial variability of soil pH levels over agricultural plots, we adopted a random sampling approach. Soil samples were collected from 10 discrete agricultural locations distributed across 10 subdistricts to ensure broad geographic representation and capture regional variability in soil characteristics. The selection of multiple subdistricts was aimed at minimizing location-specific biases and enhancing the generalizability of the classification model across diverse agricultural environments. Each sampling location was chosen to represent typical agricultural land use patterns within the study region, ensuring that the dataset encompassed natural variations in soil properties encountered in real-world farming. Within each designated plot, sampling points were positioned at approximately 3–5 m intervals in a grid pattern to ensure comprehensive spatial coverage and minimize sampling bias. This spacing protocol was intended to capture microscale spatial heterogeneity while maintaining feasibility for field collection in practice. Systematic sampling ensured that each plot contributed representative soil specimens reflecting inherent variability within individual agricultural fields, thereby enhancing the robustness of the constructed pH-level classification dataset. At each sampling point, soil specimens were extracted from a depth range of 0–15 cm using sterile collection tools to maintain sample integrity and prevent cross-contamination between sampling locations. Each sample was immediately labeled with unique identifiers corresponding to the subdistrict, plot, and spatial coordinates to enable traceability throughout analyses. Collection followed established soil sampling protocols to ensure consistency and reproducibility across sampling events.

Before soil sample extraction, a meticulous surface preparation protocol was implemented. It involved carefully removing surface vegetation, including grass, and clearing superficial debris such as plant fragments and small stones. Surface cleaning was conducted without disturbing the soil structure or stratification.

The soil sampling procedures followed the guidelines (OSU Extension Service, *n.d.*; Oklahoma State University Extension, 2024; Resource Management, *n.d.*). For soil sample extraction, we used a standard agricultural shovel for V-shaped excavation. The sampling depth was fixed at approximately 15 cm, corresponding to the typical tillage

zone, as illustrated in Figure 2. The excavation technique ensured a consistent sampling depth across sampling points. Sample collection involved extracting a uniform soil profile using the flat edge of the shovel. A continuous soil slice of approximately 2–3 cm in thickness was carefully extracted from the surface to the excavation base. Soil samples were collected from 17 plots, with 10 sampling points per plot. Random sampling was aimed at ensuring that the collected soil samples were well distributed. Each soil sample was immediately placed in a labeled plastic bag, with precise labels including a unique hole number and plot identification to maintain sample traceability and spatial reference.

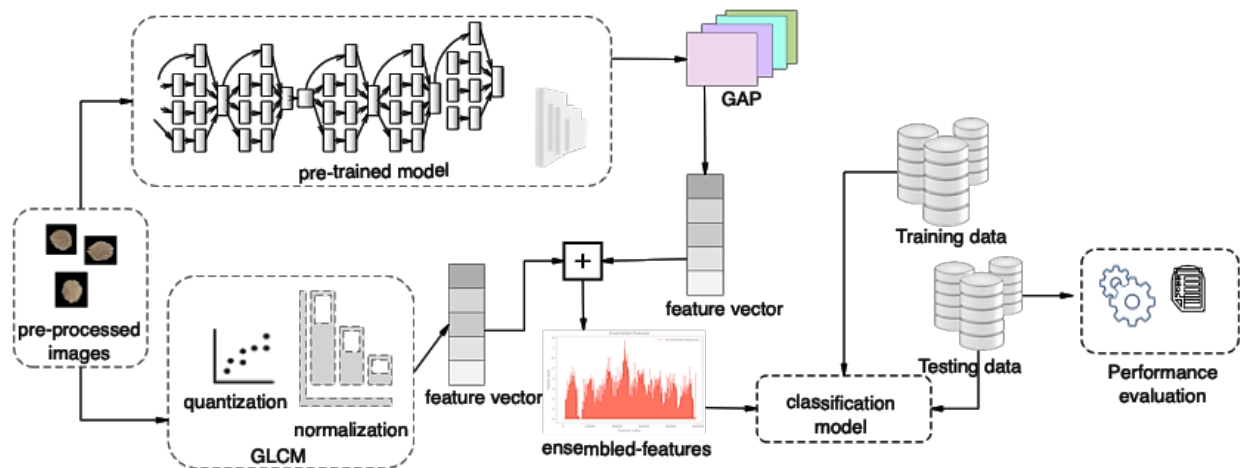


Figure 1. Research framework for an ensemble deep learning approach for soil pH classification



Figure 2. Soil sampling procedure

The systematic sampling approach enabled comprehensive spatial characterization of soil pH heterogeneity within agricultural plots, providing a robust foundation for subsequent geospatial and pedological analyses.

2.3 Image acquisition

All soil sample images (10 images per sample) were acquired using a photographic protocol intended to ensure consistency and reproducibility across the corresponding dataset by employing a Huawei JKM-LX2 smartphone camera under controlled imaging conditions. Specifically, images were captured at a resolution of 3264×2448 pixels with 24-bit color depth in the RGB (standard red–green–blue) color space and camera settings including an $f/1.8$ aperture, a $1/100$ s exposure time, ISO 59 sensitivity, a 4 mm focal length (equivalent to approximately 26 mm in a 35 mm

format), and disabled flash illumination to prevent artificial lighting artifacts that may affect soil color representation. All images were captured under natural daylight with automatic white balance to compensate for ambient lighting variations while maintaining color consistency across acquisition sessions. This setting minimized color cast variations that may interfere with subsequent color-based feature extraction and pH-level classification.

Before image acquisition, all soil samples were systematically prepared by manually removing visible impurities, including organic debris, plant residues, and foreign materials, to prevent interference with automated image processing. Each prepared sample was uniformly distributed in a circular area with an approximate radius of 15 cm on a white paper background, with corresponding soil identification labels positioned above each sample for

traceability. Perpendicular top-view images of the samples were consistently captured at 15 cm above the soil sample surface, maintaining this geometric configuration across all samples to ensure uniform spatial scaling and minimize perspective distortion that may compromise subsequent image analyses. This imaging distance and angle provided consistent pixel-to-millimeter ratios that are essential for reliable feature extraction and comparative analysis of the

collected dataset, thereby establishing the foundation for robust and reproducible image-based soil pH-level classification.

We used an Atech sensor (NPK, pH, moisture, temperature, EC, RS485), as shown in Figure 3, for soil pH-level measurement in the samples. The device yielded pH measurements in the range of 3.0 to 9.0, with a resolution of one decimal place and a measurement accuracy of ± 0.3 .



Figure 3. Atech sensor (NPK, pH, moisture, temperature, EC, RS485)

As agricultural soil management often categorizes pH levels into practical ranges (e.g., strongly acidic, moderately acidic, neutral, and alkaline), the integer-based measurements still align with meaningful decision-making in agriculture. The proposed classification method was therefore designed to operate within these ranges using integer resolution, whereby values were rounded up to the nearest whole number if the decimal exceeded or equaled 0.5 and rounded down otherwise.

2.4 Image preprocessing and standardization

Preprocessing was critical for mitigating the variability across soil sample images and ensuring robust feature extraction for soil pH-level classification. Our comprehensive preprocessing pipeline addressed three fundamental challenges in soil image analysis.

2.4.1 Padding to square methodology

Padding to a square geometry was applied as a critical preprocessing step to ensure compatibility between input data and neural network architectures. The raw acquired soil sample images exhibited varying dimensions and aspect ratios, which posed major challenges for consistent processing. To overcome these challenges, we systematically applied padding to standardize all input images to a uniform 1:1 aspect ratio before model inference. This standardization was necessary because deep learning models for image segmentation tasks, including the *rembg* architecture employed in this study, require dimensional consistency to achieve optimal stability and accuracy.

The technical necessity for this preprocessing step stems from the inherent limitations of CNNs when processing inputs with extreme aspect ratio variations. Without proper geometric normalization, either drastically elongated or compressed images would undergo non-uniform scaling during resizing, inevitably introducing geometric distortions that compromise spatial features. By padding to square regions before resizing, we maintained the original spatial relationships within each soil sample image while ensuring consistency in the input tensor dimensions required by segmentation models.

Padding mitigated variance due to heterogeneous image geometries and reduced potential artifacts that might arise from irregular scaling. The uniform input format ensured that the pretrained segmentation model operated in its optimal parameter space, thereby maximizing the segmentation accuracy and maintaining consistent performance across a diverse range of geometries from the soil sample images in the dataset. We implemented image padding to standardize the spatial dimensions, addressing potential geometric distortions inherent in agricultural image acquisition. By transforming input images into square regions of 224×224 or 299×299 pixels, we eliminated potential artifacts arising from non-uniform aspect ratios during feature extraction and model training. Padding ensured that all images conformed to consistent dimensions, which facilitated standardized analysis. Examples of soil sample images before and after padding are shown in Figure 4.

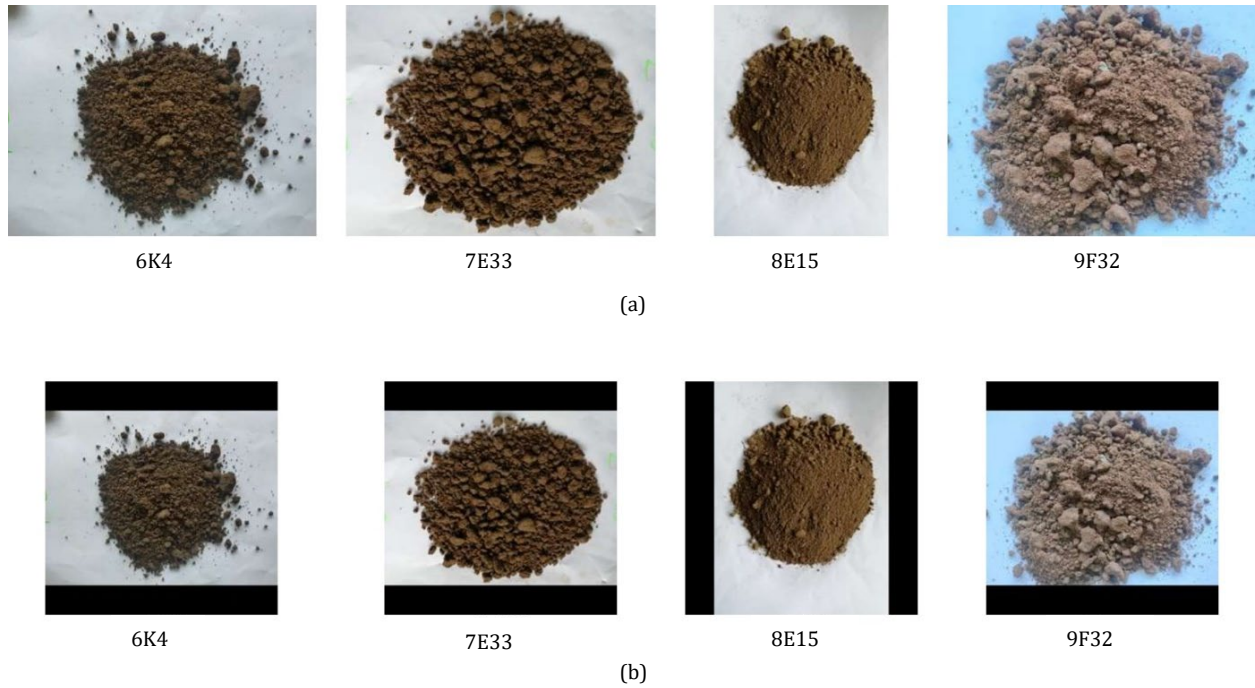


Figure 4. Soil sample images under geometric normalization (a) before and (b) after padding to square methodology

2.4.2 Background removal

Next, background removal was implemented as a critical preprocessing step to isolate soil samples from extraneous visual information that might interfere with subsequent pH-level classification in the captured images. We employed the *rembg* library with the U²-Net deep learning model for automated foreground segmentation operating directly on the RGB color space without requiring manual parameter tuning or morphological operations. This model was necessary because traditional threshold-based segmentation often fails to capture subtle variations in soil texture and color, which are essential for accurate pH-level determination, while manual segmentation would introduce subjective bias and reduce reproducibility in large datasets.

The *rembg* implementation was selected because it eliminated the need for user-defined thresholds that traditionally determine object-background boundaries, addressing a fundamental limitation of conventional segmentation. Instead of producing binary masks with hard edges and potential information loss, U²-Net provided a salient map or alpha matte—a grayscale representation

where pixel intensities range from 0 (complete background) to 1 (complete foreground). This probabilistic approach was essential for preserving transitional pixel information at object boundaries, which was important for soil sample images, where edge definition could be gradual owing to particle distribution and surface texture variations.

We recognized that the *rembg* model might exhibit suboptimal performance under specific conditions, including insufficient object-background contrast, complex lighting with harsh shadows, and highly reflective surfaces that might be misclassified as background. To enhance background removal, we implemented controlled image acquisition protocols including uniform soil distribution to eliminate surface irregularities, standardized white background photography to maximize contrast, and controlled lighting with sufficient illumination to minimize shadows. These strategies were crucial for ensuring consistent segmentation quality across samples, thereby maintaining the integrity of the dataset for reliable pH-level classification model training and evaluation. Examples of soil sample images after background removal are shown in Figure 5.

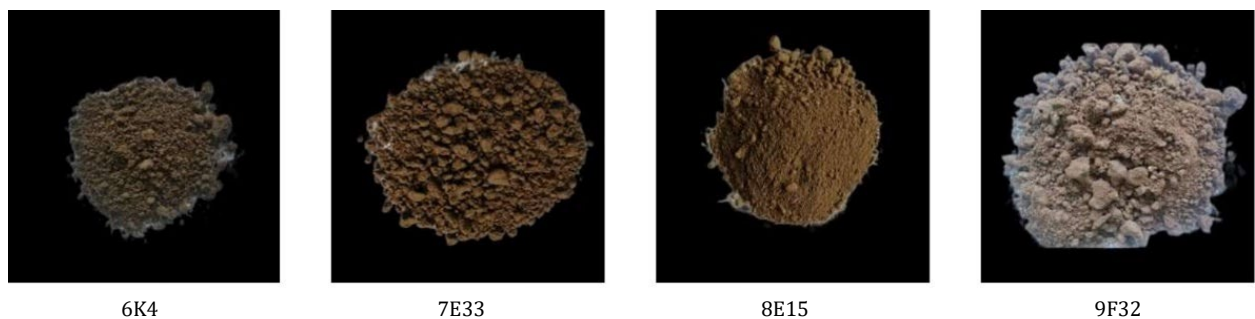


Figure 5. Soil sample images after background removal

2.4.3 Image cropping

As the collected images contained circular soil samples with surrounding background areas, image cropping was applied to eliminate residual padding areas and any remaining background artifacts following segmentation. Cropping ensured that the processed images contained exclusively the region of interest to obtain optimal data for subsequent analytical procedures. We conducted systematic cropping experiments to optimize the region of interest. Specifically, we evaluated cropping percentages of 20%, 25%, 30%, and 35% from the image borders to determine the optimal preprocessing approach that would eliminate most soil-pile boundaries while preserving essential soil characteristics. Soil sample images at cropping percentages ranging from 20% to 35% are shown in Figure 6.

The experimental analysis revealed that edge pixels containing soil color information mixed with background elements could introduce noise into classification. Through comparative evaluation, we determined that 30% border cropping was the optimal threshold because it removed boundary pixels where soil coloration became contaminated with background interference while preserving sufficient soil texture and color information for accurate pH-level classification. This cropping process ensured that the preserved image data contained primarily pure soil pixels, thereby minimizing potential classification errors from mixed pixel values at the soil-background interface while maintaining adequate spatial resolution for feature extraction and pH-level classification.

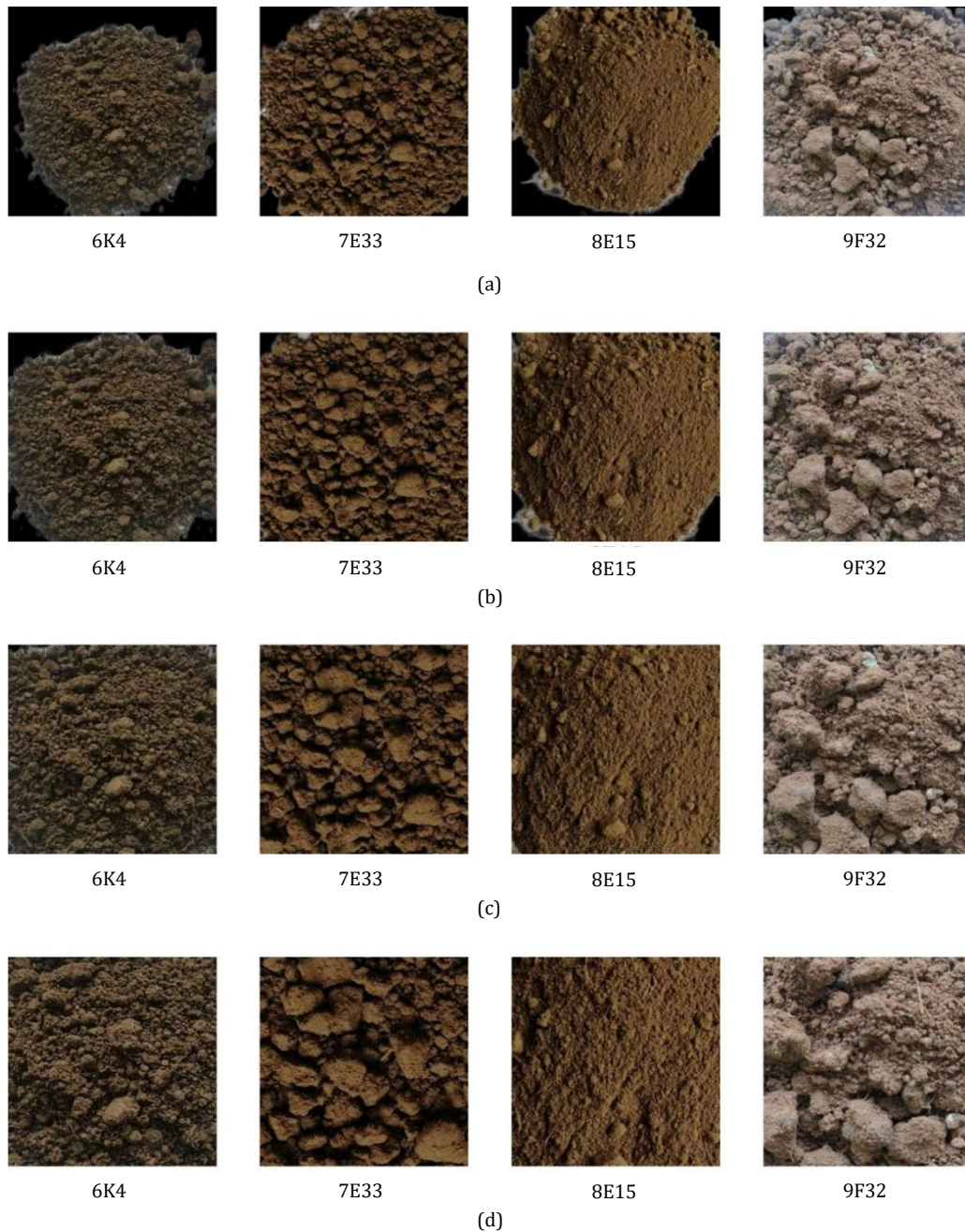


Figure 6. Soil sample images cropped at (a) 20%, (b) 25%, (c) 30%, and (d) 35%

2.5 Dataset characterization

2.5.1 Initial class distribution analysis

Before model training, the collected soil sample images underwent a screening and cleaning process. Images identified as unsuitable, such as those exhibiting angular distortions, blurring, or other visual artifacts, were

excluded to maintain the dataset quality and consistency. A preliminary assessment revealed notable class imbalance and limited sample diversity, which posed challenges for reliable soil pH-level classification. The initial distribution of soil sample images across four pH levels is summarized in Table 1.

Table 1. Initial soil sample image class distribution according to pH level

pH level	Initial image count	Proportion (%)
6	289	21.4
7	422	31.2
8	422	31.2
9	215	16.2
Total	1,348	100

Certain pH levels had substantially fewer images in the collected dataset, which could lead to biased training. The observed distribution demonstrated high variability in sample representation. Specifically, pH levels 7 and 8 exhibited identical sample counts (422 images), while pH levels 6 and 9 showed substantially lower representation. This inherent class imbalance required strategic preprocessing to ensure robust and unbiased model training.

2.5.2 Augmentation for class balancing in the dataset

A major challenge in image-based classification of soil samples is the scarcity of balanced and diverse datasets. To address class imbalance and enhance generalizability, we performed comprehensive data augmentation to achieve uniform class representation and improve trained model robustness. Augmentation involved controlled synthetic image generation through zooming and geometric transformations designed to simulate realistic variations while preserving essential soil sample features in the images.

Offline data augmentation was performed using the Keras *ImageDataGenerator* class, incorporating zoom-out operations in the range of 0.8–1.0 (maximum 20% reduction). This zooming mechanism enabled the model to recognize fine-grained local textures, such as soil particle patterns, while simultaneously capturing broader contextual information, including the overall soil structure

and the spatial arrangement of gravel. Spatial variability was introduced through horizontal and vertical flipping operations to increase orientational diversity without altering core visual properties, as soil texture and granularity are largely orientation-invariant. Rotation, translation, and color adjustment transformations were excluded to maintain authentic soil color representation, with the nearest-neighbor fill mode being retained for boundary handling.

Offline augmentation was performed after background removal and 30% border cropping, ensuring consistent preprocessing across soil sample images. The augmented images were stored in specific directories, and the dataset was augmented before splitting. Augmented images were utilized exclusively within training and validation sets to enhance model generalization, while the test set was stored in a separate directory.

Augmentation contributed to overcoming limitations associated with small and imbalanced datasets by ensuring standardized class representations across pH levels. This balancing approach reduced bias toward dominant classes and ensured an equal contribution of samples from every pH level during training, ultimately enhancing the classification reliability and robustness. Characteristics of the soil sample images—including texture, granularity, and color integrity—were preserved throughout augmentation. Table 2 lists the dataset composition after augmentation, which provided a balanced representation across pH levels.

Table 2. Augmented soil pH image class distribution

pH level	Augmented image count	Proportion (%)
6	450	25
7	450	25
8	450	25
9	450	25
Total	1,800	100

2.5.3 Data splitting and cross-validation

Data were split at the image-level using an 80/20 holdout strategy with a validation split of 0.2 in the Keras *Image DataGenerator* class and a seed of 1,234 for reproducibility. Class labels were automatically assigned according to the directory structure, and the test set was stratified to contain exactly 90 images per class, ensuring balanced representation across pH levels. The training and validation sets were obtained using Keras's built-in stratified validation splitting, while the test set was maintained in a separate directory and loaded using a distinct generator with shuffle set to False for evaluation consistency.

Individual soil sample images were captured from multiple angles, and comprehensive data augmentation was applied to ensure no identical images appeared in the training, validation, and test sets. This prevented data leakage even under image-level random splitting because augmentation increased visual diversity to prevent exact duplicate recognition between sets.

2.6 Evaluated models

We investigated four high-performance deep learning models with standardized input setups. The models processed inputs with dimensions of 224×224 pixels for VGG16 and ResNet50, and 299×299 pixels for InceptionV3 and Inception ResNetV2. These dimensions were intended to enhance the processing of medium-resolution images, notably for soil texture analysis, while reducing preprocessing demands. The uniformity of input dimensions across models enabled direct comparisons given the structural compatibility of their input layers.

The four evaluated models were initialized using pretrained weights from the ImageNet dataset, facilitating transfer learning. We meticulously selected these diverse

models to assess distinct structures for feature extraction from our constructed dataset. Each architecture provided unique attributes to the proposed pH-level classification method. We aimed to systematically examine the relationship between network architecture and feature extraction capabilities, particularly for specialized datasets such as the constructed dataset containing soil sample images. This systematic comparison offered insights into the efficacy of various models in specialized image classification.

2.7 Experimental setup

To achieve the best performance in soil pH-level classification, the evaluated models were implemented and trained utilizing a systematic configuration and technique. The input dimensions were either 224×224 or 299×299 pixels according to the model. A batch size of 32 was utilized during training, with a maximum of 150 epochs to facilitate sufficient learning while maintaining computational efficiency. The Adam optimizer was adopted to promote convergence and adaptive learning. Early stopping was also implemented to prevent overfitting and mitigate unnecessary computations. It terminated training if the validation loss exhibited no improvement over 20 consecutive epochs. The initial learning rate was set to 0.001, with dynamic modifications implemented via the *ReduceLROnPlateau* callback, which reduced the learning rate when the model performance remained unchanged over a specified number of epochs according to the validation loss. Specifically, we reduced the learning rate by a factor of 0.2 upon observing no enhancement in the validation loss over five consecutive epochs. Minimum learning rate (*min_lr*) thresholds of 0.0001 and 0.000001 were evaluated, providing adaptability in model optimization. The model configuration and training parameters are listed in Table 3.

Table 3. Model configuration and training parameters

Parameter	Description
Input Size	224×224 or 299×299 depending on model
Batch Size	32
Maximum No. Epochs	150
Optimizer	Adam
Early Stop	Training stops if <i>val_loss</i> does not improve for 20 consecutive epochs.
Initial Learning Rate	ReduceLROnPlateau callback used to dynamically adjust learning rate: - <i>lossval_loss</i> monitored for 5 epochs - Learning rate reduced by a factor of 0.2 if no improvement is observed ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=5, min_lr=0.0001)
Minimum Learning Rate (<i>min_lr</i>)	Evaluation of two values: 0.0001 and 0.000001

3. RESULTS

3.1 Baseline performance analysis of pretrained models

The experimental results of the pretrained models are listed in Table 4, providing a comprehensive evaluation that underscores the significance of model improvement for accurate classification. These performance results indicate varying efficacies among the pretrained models. For instance, Inception ResNetV2 with 299×299 -pixel inputs achieves the best performance. Less complex models, such as VGG16, show inferior performance. The large performance

disparity highlights several crucial factors that demand model improvement.

Despite being trained on large datasets (e.g., ImageNet), the pretrained models have limited effectiveness in specific classification tasks. The problem is particularly evident in the outcomes of ResNet50 and VGG16, indicating that general features may inadequately encapsulate domain-specific characteristics of soil pH-level classification. A higher input resolution (299×299 pixels) leads to consistently higher performance than a lower resolution (224×224 pixels) across the evaluated models. For example, Inception ResNetV2 demonstrates an increase in accuracy from 73.89% to 80.28% at a higher resolution,

indicating that feature extraction may be improved by additional spatial information.

The considerable accuracy disparity between models (ranging from 25.00% to 80.28%) indicates that the underlying architecture is essential for feature extraction. Hence, augmenting these architectures with supplementary features may enhance performance. The low performance of certain pretrained models indicates the need for domain-specific features that can properly capture soil sample textures. The performance gap between models suggests that various feature extraction methods may obtain

complementary information, highlighting the suitability of ensemble or hybrid strategies.

The highest-performing Inception ResNetV2 (accuracy of 80.28%) may be further enhanced, particularly for applications in agriculture and soil science. Furthermore, improved models may yield uniform and dependable results across a variety of scenarios. Hence, model enhancement should be explored through techniques including domain-specific feature extractors, ensemble methods that amalgamate diverse features, architecture modifications to capture specific characteristics, and tailored preprocessing.

Table 4. Experimental results of pretrained models

Model	Input size	Dense layer	Total No. parameters	Learning rate	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
InceptionV3	224 × 224	1 × 2048	26,007,332 (99.21 MB)	1.00×10 ⁻⁴	69.44	69.49	69.44	68.53
	299 × 299	1 × 2048	26,007,332 (99.21 MB)	2.00×10 ⁻⁴	68.89	70.47	68.89	68.75
Inception ResNetV2	224 × 224	1 × 2048	57,492,708 (219.32MB)	2.00×10 ⁻⁴	73.89	75.78	73.89	74.06
	299 × 299	1 × 2048	57,492,708 (219.32MB)	1.00×10 ⁻⁴	80.28	80.55	80.28	80.10
ResNet50	224 × 224	1 × 2048	27,792,260 (106.02MB)	1.00×10 ⁻³	28.61	14.70	28.61	19.32
VGG16	224 × 224	1 × 2048	15,773,508 (60.17 MB)	1.00×10 ⁻⁴	25.00	6.25	25.00	10.00

The experimental results indicate that pretrained models establish valuable baselines. Nonetheless, they require enhancement to achieve optimal performance on specific classification tasks. Hence, refined methods should be tailored for handling soil pH-level classification. In fact, incorporating domain-specific data and modifying the model architecture may enhance the model's ability to accurately represent critical soil characteristics while retaining the advantages of transfer learning from pretrained models.

3.2 Ensemble of pretrained and texture features

To enhance the classification performance, we employed the GLCM to extract texture features, thereby augmenting the information for integration with pretrained models, whose experimental results are listed in Table 4. The most effective backbone model is Inception ResNetV2. The GLCM provides a statistical technique that examines the spatial correlations across pixel intensities in an image, emphasizing the co-occurrence of gray levels. This method captures texture characteristics, which are necessary for unveiling structures and relationships within an image.

3.2.1 CNN backbone

We employed Inception ResNetV2 as the backbone. All input images were resized to 299 × 299 × 3, and the pixel intensities were normalized in [0, 1] using linear rescaling. For input image I with pixel intensities $I_{raw}(i, j, c) \in [0, 255]$, the normalized image is given by

$$I_{norm} = \frac{I_{raw}(i, j, c)}{255}, \forall i, j \in [1, 299], c \in \{R, G, B\} \quad (1)$$

All backbone layers were fine-tuned from the beginning (trainable set to True), and neither staged freezing nor unfreezing was performed. After the final

convolutional block, we applied GAP to generate a fixed-length embedding. Let the final feature tensor be $F \in \mathbb{R}^{h \times w \times d}$. GAP is defined as

$$Z_k = \frac{1}{h \cdot w} \sum_{i=1}^h \sum_{j=1}^w F_{i,j,k}, \quad k = 1, \dots, d \quad (2)$$

For Inception ResNetV2, this yielded a backbone embedding vector $Z \in \mathbb{R}^{1526}$.

3.2.2 GLCM-derived texture features

To capture handcrafted texture descriptors, we employed the GLCM. Each RGB image was first represented as 8-bit unsigned integers and quantized into 8 gray levels using equal-width binning (bin size of 32) as follows:

$$I_q = \left\lfloor \frac{I_{raw}}{256/8} \right\rfloor, \quad I_q \in \{0, 1, \dots, 7\} \quad (3)$$

For each channel $c \in \{R, G, B\}$, GLCM $P^{(c)}(i, j) \in P^{(c)} \in \mathbb{R}^{8 \times 8}$ was computed using a displacement of 1 pixel ($d = 1$) and orientation of 0°. By setting both *symmetric* and *normed* to True, we ensured that the matrix was symmetric and normalized, such that

$$\sum_{i=0}^7 \sum_{j=0}^7 P^{(c)}(i, j) = 1 \quad (4)$$

From the GLCM per channel, we obtained the contrast, energy, entropy, and homogeneity. Among these GLCM features, contrast was calculated using (5) to quantify the intensity variations between adjacent pixels, indicating texture diversity.

$$contrast^{(c)} = \sum_{i=0}^7 \sum_{j=0}^7 (i - j)^2 P^{(c)}(i, j) \quad (5)$$

To assess the regularity of pixel intensities, that is, the texture coherence of an image, energy was calculated as

$$energy^{(c)} = \sum_{i=0}^7 \sum_{j=0}^7 (P^{(c)}(i, j))^2 \quad (6)$$

Entropy was calculated using (7) to indicate the unpredictability or uncertainty in image texture features:

$$entropy^{(c)} = \sum_{i=0}^7 \sum_{j=0}^7 P^{(c)}(i, j) \log_2 P^{(c)}(i, j) + \varepsilon \quad (7)$$

Homogeneity evaluated the similarity of pixel intensities in neighboring regions to reflect the texture smoothness as follows:

$$homogeneity^{(c)} = \sum_{i=0}^7 \sum_{j=0}^7 \frac{P^{(c)}(i, j)}{1 + |i - j|} \quad (8)$$

Hence, a four-dimensional descriptor, $Z^{(c)} \in \mathbb{R}^4$, was obtained per channel. Concatenating across the three channels resulted in a 12-dimensional handcrafted texture vector given by

$$Z_{GLCM} = [Z^{(R)} || Z^{(G)} || Z^{(B)}] \in \mathbb{R}^{12} \quad (9)$$

3.2.3 Feature fusion

Before feature fusion, GLCM normalization was applied to ensure that the extracted features were suitably scaled for model training. Normalization was essential for preserving uniformity across feature values, guaranteeing consistent scaling of handcrafted features during learning. Moreover, normalization increased the model stability throughout training and enhanced its overall efficacy. The classification accuracy and reliability of the soil pH-level estimation improved when texture information from the GLCM was combined with high-level semantic features from the pretrained models. This combination of statistical texture and deep learning features yielded a detailed representation by capitalizing on the unique advantages of both types of features. Normalization constrained the features to the range [0, 1] as follows:

$$Z_{GLCM} = \frac{Z_{GLCM} - \min(Z_{GLCM})}{\max(Z_{GLCM}) - \min(Z_{GLCM})} \quad (10)$$

Table 5. Classification architecture settings

Layer/operation	Parameters	Output dimension
GAP	-	1,536
Concatenation	CNN: 1,536 + GLCM: 12 = 1,548	1,548
Dropout	0.5	1,548
Dense (fully connected)	Units = 2,048 Activation: ReLU L2 regularization (λ) = 0.001	2,048
Dropout	0.5	2,048
Output layer (dense + softmax)	Units = $n_{classes}$ = 4 Activation = Softmax	4

3.4 Performance evaluation

The performance of the evaluated classification models was compared using key evaluation metrics—namely, accuracy, precision, recall, and F1-score—to identify the best-performing model for classifying pH levels.

Accuracy is the ratio of correctly classified samples to the total number of samples:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Precision evaluates the model's ability to correctly predict positive results as follows:

Finally, the 1,536-dimensional CNN embedding and 12-dimensional GLCM-based descriptor with handcrafted features were concatenated to form a unified representation as follows:

$$Z_{fused} = [Z_{CNN} || Z_{GLCM}] \in \mathbb{R}^{1548} \quad (11)$$

where $Z_{CNN} \in \mathbb{R}^{1,536}$ and $Z_{GLCM} \in \mathbb{R}^{12}$. Thus, the final representation per image was a 1,548-dimensional feature vector used for downstream classification of pH levels in soil sample images.

3.3 Classification architecture

We employed the same classification architecture for all evaluated models to ensure consistency and comparability. Specifically, a GAP layer first consolidated spatial information by averaging each feature map, thus reducing the data dimensionality. GAP amalgamated GLCM features with its output for all models, augmenting the synthesis of texture features. A dropout layer with a rate of 0.5 was then implemented to mitigate overfitting by randomly deactivating 50% of the neurons during training. A dense layer comprising 2,048 units with ReLU (rectified linear unit) activation was utilized to incorporate nonlinearity and enhance model learning. L2 regularization with a penalty of 0.001 was incorporated into this layer to enhance generalization and reduce overfitting. An additional dropout layer with a rate of 0.5 was incorporated after the dense layer to enhance model resilience to overfitting.

Finally, an output layer intended for multiclass classification consisted of four units, one per pH level. This layer employed the softmax activation to provide normalized probabilities per class. The classification method integrated convolutional and GLCM features. This architecture utilized spatial and texture information, along with regularization methods including dropout and L2 regularization, to create a robust and adaptable model for accurate classification of soil pH levels. The classification architecture settings are listed in Table 5.

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

Recall evaluates the model's ability to correctly predict positive results as follows:

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

The F1-score evaluates the balance between precision and recall as follows:

$$F1\text{-score} = \frac{(2 * Precision * Recall)}{Precision + Recall} \quad (15)$$

Above, TP (true positives) is the number of accurately identified true positive cases, TN (true negatives) is the number of accurately identified true negative cases, FP (false positives) is the number of cases erroneously classified as positive, and FN (false negatives) is the number of instances mistakenly classified as negative.

The evaluation metrics offer a quantifiable representation of the model efficacy, assisting researchers in identifying the best soil pH-level classification model that enhances both accuracy and efficiency.

3.5 Experimental results and analysis of ensemble-feature classification method

Model optimization required an evaluation of diverse hyperparameters under comprehensive fine-tuning. The

learning rate, an essential hyperparameter influencing model convergence and training stability, was examined in two ranges of 0.001–0.0001 and 0.001–0.000001. This evaluation of learning rates allowed us to determine the ideal equilibrium between convergence speed and training stability for the model to proficiently learn intricate features found in soil sample images.

Regularization strategies were systematically evaluated by comparing model performance with and without the application of L_2 regularization. This evaluation allowed us to determine the model generalizability and mitigate overfitting given the uniqueness of pH-level classification in soil sample images. L_2 regularization was investigated for its potential to enhance model robustness and generalizability on unseen soil sample images.

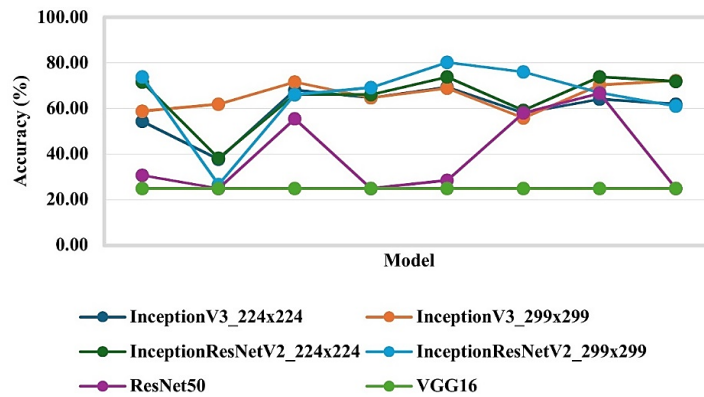


Figure 7. Performance comparison of the investigated model following hyperparameter tuning

The experimental results shown in Figure 7 highlight the effect of different hyperparameters on model accuracy. Hence, suitable hyperparameters for soil pH-level classification were obtained, as well as insights into the importance of each hyperparameter for soil image analysis. The comprehensive hyperparameter evaluation ensured model robustness, and the best model was further enhanced with ensemble features for training.

The hyperparameters that provided the highest accuracy were selected for each evaluated model for feature ensemble integration to increase model performance while evaluating GLCM feature integration. Hybrid features allowed us to integrate the advanced feature extraction of deep learning with conventional texture analysis.

The ensemble properties were included in the optimized architectures for InceptionV3, Inception ResNetV2, ResNet50, and VGG16. Thus, hierarchical feature extraction in deep learning models was combined with explicit texture features obtained from the GLCM. The ensemble method was executed under the optimal hyperparameters obtained from fine-tuning, facilitating comparison with baseline models.

Integrating complementary features increased the model's ability to capture high-level semantic information and low-level textures pertinent to soil pH-level classification. Incorporating GLCM features in soil image analysis was particularly advantageous because texture patterns provided representative information about pH levels. The texture information reflected statistical metrics of pixel intensity correlations.

The improved models were evaluated similarly to the baseline models, facilitating direct performance comparisons and measurement of the enhancements attained by feature ensembling. This evaluation elucidated the efficacy of integrating conventional computer vision with deep learning for specialized classification.

The experimental results listed in Table 6 demonstrate that among the evaluated models, Inception ResNetV2 with GLCM for a 299×299 -pixel input resolution achieves outstanding performance, with the best accuracy of 82.22%, precision of 82.01%, and F1-score of 82.10% at a learning rate of 1×10^{-4} . In fact, Inception ResNetV2, combined with GLCM features, provides efficient feature extraction and accurate pH-level classification.

InceptionV3 with GLCM was evaluated for two input configurations. The 224×224 -pixel input leads to 67.22% accuracy, whereas the 299×299 -pixel input shows inferior performance with 56.94% accuracy. Hence, the input resolution can substantially influence the performance, even for the same underlying model.

ResNet50 with GLCM was evaluated for an input resolution of 224×224 pixels, showing limited efficacy with an accuracy of 33.33%. After evaluation using a 224×224 -pixel input resolution, VGG16 with GLCM exhibits moderate performance, achieving 45.28% accuracy.

The results reveal various insights into the design and performance of deep learning models. First, no direct association appears between model size and performance, as demonstrated by VGG16 with GLCM surpassing ResNet50 with GLCM despite possessing fewer parameters. Second, the input resolution seems to be a

crucial parameter for model performance, particularly in Inception ResNetV2 with GLCM. Empirical testing indicates that a learning rate of 1×10^{-4} is the best hyperparameter because it leads to model consistency.

The confusion matrices in Figure 8 offer deeper insight into how input resolution influences the class-discrimination capability of the evaluated models. Beyond indicating correct predictions, they reveal how classification errors are distributed across pH levels and whether increasing the input size helps reduce confusion between visually similar classes. A comparison of the 224×224 and 299×299 configurations demonstrates that higher resolution does not consistently translate into better class separation. Instead, its impact varies across model architectures, suggesting that performance depends on how effectively each architecture captures and integrates the additional spatial details with GLCM-derived texture information.

For Inception ResNetV2 with GLCM, increasing the input size from 224×224 to 299×299 results in a substantially more balanced diagonal distribution. At 224×224 , the correct classification rates for pH levels 6, 7, 8, and 9 are 81.11%, 66.67%, 47.78%, and 86.67%, respectively. The major weakness occurs at pH 8, where 35.56% of the samples are incorrectly classified as pH 7. With the 299×299 input, the corresponding correct classification rates become 87.78%, 70.00%, 76.67%, and 94.44%. Most notably, the correct classification of pH 8 increases considerably, while its misclassification as pH 7 decreases from 35.56% to 16.67%. Misclassification between pH 7 and pH 8 remains the most prominent error pattern, suggesting that these two classes share similar visual or textural characteristics. Nevertheless, the 299×299 configuration provides clearer separation across all four pH levels, indicating that the higher-resolution input enables Inception ResNetV2 to utilize the complementary deep and GLCM-based texture information more effectively.

A different pattern is observed for InceptionV3 with GLCM. At 224×224 , the model produces relatively balanced correct classification rates of 61.11%, 62.22%, 71.11%, and 74.44% for pH levels 6–9, respectively, although confusion is evident between pH 6 and 7 and between pH 7 and 8. When the input size is increased to 299×299 , correct classification improves for pH 8 and pH 9 to 81.11% and 84.44%, respectively. However, this improvement is accompanied by a substantial deterioration in the recognition of pH 6, for which the correct classification rate decreases to 14.44%, while 57.78% of pH 6 samples are misclassified as pH 7. Thus, unlike Inception ResNetV2, increasing the input resolution does not produce a consistently beneficial effect for InceptionV3. Instead, it shifts the model's discrimination toward the higher pH classes at the expense of pH 6, indicating greater class-specific sensitivity to changes in input resolution.

The comparison therefore reveals that the effect of increasing input resolution is architecture-dependent rather than universally beneficial. For Inception ResNetV2 with GLCM, the 299×299 input reduces major off-diagonal errors and produces stronger and more balanced discrimination across the four pH levels. In contrast, for InceptionV3 with GLCM, the same increase in resolution improves the recognition of pH 8 and pH 9 but simultaneously introduces severe confusion between pH 6 and pH 7. These contrasting patterns suggest that higher-resolution images provide additional spatial and texture information, but the extent to which this information contributes to classification depends on the feature-learning capability of the underlying architecture. Overall, the confusion-matrix analysis supports the selection of Inception ResNetV2 with GLCM at 299×299 pixels as the more effective configuration, not only because of its aggregate performance but also because of its more consistent class-wise discrimination and reduced concentration of misclassification errors.

Table 6. Experimental results of ensemble-feature classification model

Model	Input size	Density	Total parameters	Learning rate	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
InceptionV3 w/GLCM	224×224	1×2048	26,031,908 (99.30 MB)	1.00×10^{-4}	67.22	71.50	67.22	68.23
	299×299	1×2048	26,031,908 (99.30 MB)	2.00×10^{-4}	56.94	68.34	56.94	53.54
Inception ResNetV2 w/GLCM	224×224	1×2048	57,517,284 (219.41 MB)	2.00×10^{-4}	70.56	71.75	70.56	70.22
	299×299	1×2048	57,517,284 (219.41 MB)	1.00×10^{-4}	82.22	82.01	82.22	82.10
ResNet50 w/GLCM	224×224	1×2048	27,816,836 (106.11 MB)	1.00×10^{-4}	33.33	35.59	33.33	27.51
VGG16 w/GLCM	224×224	1×2048	15,798,084 (60.26 MB)	1.00×10^{-4}	45.28	45.83	45.28	43.75

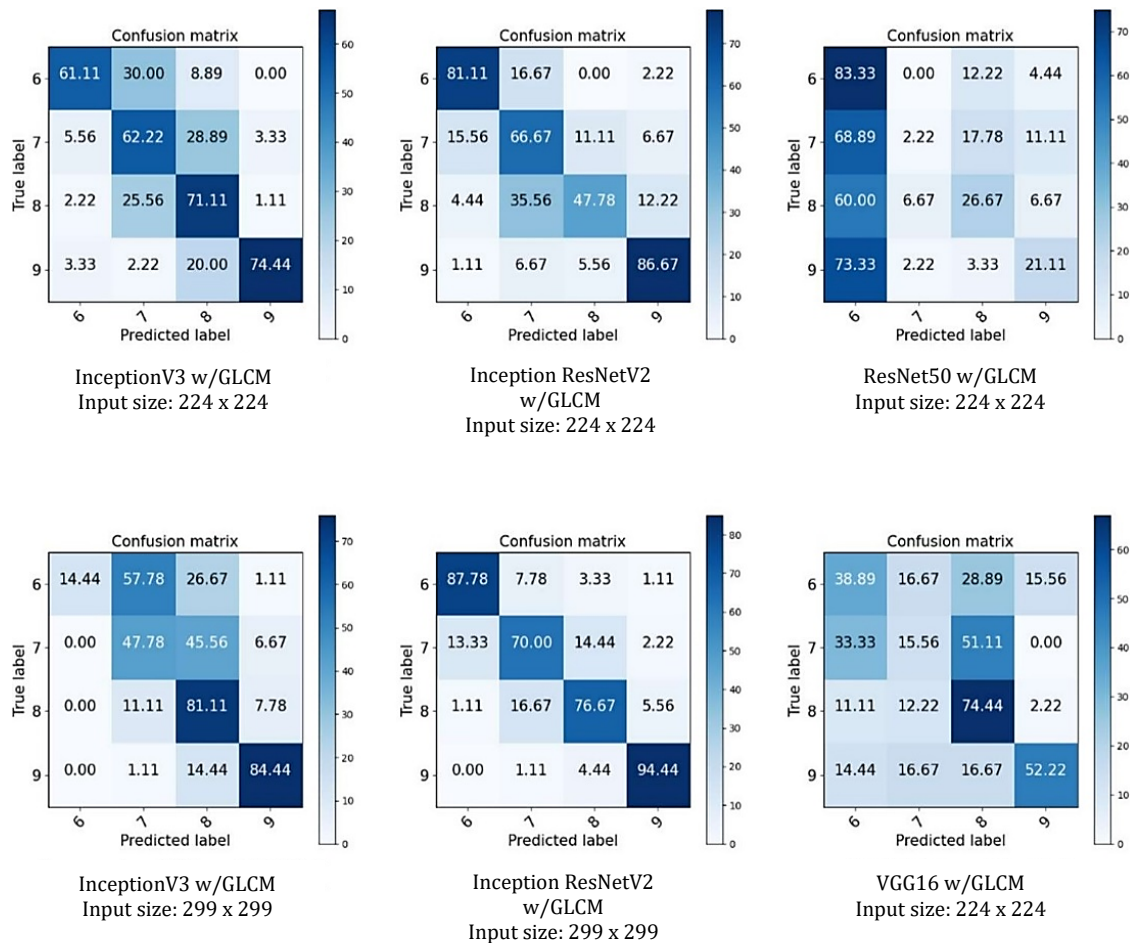


Figure 8. Confusion matrices of evaluated ensemble-feature classification models

3.6 Analysis of GLCM-enhanced feature representations for soil pH-level classification

The experimental results listed in Tables 4 and 5 demonstrate that the classification performance is substantially enhanced when ensemble features are incorporated into pretrained models. Remarkably, incorporating GLCM features into Inception ResNetV2 enhances the classification performance, increasing the accuracy from 80.28% to 82.22%. This enhancement is considerable because determining the pH level from soil sample images is challenging. In fact, even small variations in texture can indicate differing pH levels, despite the soil sample appearing uniform.

Empirical results demonstrate that the ensemble method combining deep learning and texture features consistently outperforms pretrained models. This enhancement is attributed to the synergy between global structures derived from deep learning and local textures obtained from GLCM analysis. The enhanced capacity to differentiate pH levels is particularly evident when samples exhibit similar features, which conventional single-feature methods frequently fail to distinguish.

Overall, the classification results suggest that using ensemble features is a step forward in classifying pH levels in soil sample images. The proposed method increases the accuracy and reliability in practical scenarios where intraclass similarities may disrupt classification.

An examination of feature representations, as shown in Figure 9, indicates that the baseline Inception ResNetV2 features and the enhanced ensemble using GLCM features exhibit distinct characteristics. The baseline features effectively capture global characteristics such as structural patterns and color distributions ranging from 0 to 7. However, these features are inadequate for retaining detailed textures essential for precise classification of soil pH levels.

GLCM features provide statistical metrics that substantially improve the feature space. Specifically, contrast (3.5–4.0) indicates variations in texture intensity, entropy (2.5–3.5) reflects the complexity of surface patterns, energy (0.5–1.0) demonstrates texture uniformity, and homogeneity (1.0–2.0) reveals sample surface consistency. This enhanced feature representation exhibits greater variability than the baseline, providing a comprehensive overview of soil characteristics.

The ensemble method allows for distinguishing soil sample images based on analogous colorimetric characteristics and disparate textures. Integrating global structures from Inception ResNetV2 with local textures from the GLCM yields a comprehensive feature space that encompasses both macro- and micro-characteristics in soil samples. This integration represents a major advancement over conventional feature extraction from a single source, enhancing soil pH-level classification accuracy by considering both global patterns and local textures.

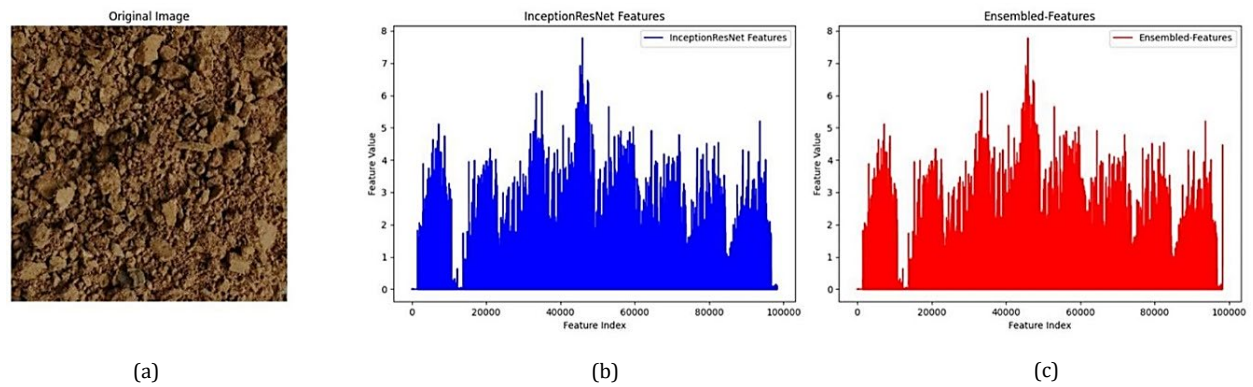


Figure 9. (a) Soil sample image and heatmaps for (b) pretrained and (c) ensemble features

Incorporating GLCM features into Inception ResNetV2 enhances classification through complementary feature representations. The baseline Inception ResNetV2 adeptly captures global features like dominant soil color and macro-structural patterns but misses small details, likely undermining classification for soil sample images with similar colorimetric properties but differing pH levels.

The GLCM introduces statistical metrics that notably enhance the model's analytical capabilities. Contrast accurately assesses texture differences, whereas entropy offers an understanding of surface pattern intricacy. Energy and homogeneity provide critical insights into particle dispersion uniformity and surface consistency. These metrics enhance the distinction between soil samples that appear identical yet possess varying structural properties.

The developed and comprehensive feature extraction technique enhances the model's ability to reliably classify the pH levels of diverse soil sample images, overcoming the deficiencies of conventional single-feature methods and establishing a robust method for soil analysis.

3.7 Comparative analysis and benchmarking

The comparison results of the proposed ensemble-feature classification method with similar solutions listed in Table 7 reveal substantial enhancements across evaluation metrics when using our proposal. All benchmark models were evaluated on the same dataset (SoilPh dataset) for a fair comparison. The proposed method achieves an accuracy of 82.22%, precision of 82.01%, recall of 82.22%, and F_1 -score of 82.10%, indicating major progress compared with existing solutions. Specifically, our proposal demonstrates a large enhancement over the implementations (Barman & Choudhury, 2019), which show accuracies of 33.89% for linear regression, 52.22% for k -nearest neighbor (KNN) regression, and 55.00% for a neural network with Levenberg–Marquardt optimization. Hence, our proposal delivers an accuracy enhancement of 27.48% compared with similar methods.

Compared with deep learning methods, our proposal notably outperforms the recent SoilNet (Pandiri et al., 2024), which achieves 45.19% accuracy with preprocessing and 58.61% accuracy without preprocessing. Our ensemble-feature method exhibits an enhancement of roughly 23.61% compared with the most effective SoilNet variant. This enhancement is particularly noteworthy because our method can manage challenging datasets characterized by few samples and large intraclass similarity.

The superior performance of the proposed Inception ResNetV2 ensemble architecture utilizing 299×299 -pixel inputs and GLCM feature integration advances pH-level classification in soil sample images. Notwithstanding the heightened computational complexity of 57,517,284 parameters, the improved accuracy and strong performance across all evaluation metrics suggest the applicability of the proposed method in practical agricultural contexts.

4. DISCUSSION

This study addresses major challenges inherent in agricultural image datasets, including limited sample diversity, class imbalance, and high intraclass similarity. Data augmentation addresses these limitations while preserving essential soil-specific characteristics for pH-level classification. Offline augmentation ensures consistent preprocessing across samples and prevents data leakage between training and evaluation sets.

However, this study has several limitations that warrant consideration. Image-level data splitting was improved through extensive augmentation but did not enforce strict physical sample separation that would strengthen generalization. Future work should implement grouped cross-validation to ensure that all images from identical soil samples remain within the same data partition, thereby preventing sample-specific memorization.

The dependency on natural lighting may limit the method's consistency under varying acquisition conditions. While an automatic white balance provides partial compensation for ambient light variations, systematic differences in color temperature and intensity across acquisition sessions may introduce unwanted variance that weakens color-based feature extraction. The subtle variations in soil color that correlate with pH levels make this limitation particularly critical for maintaining classification accuracy. Future implementations should incorporate controlled illumination with standard LED panels providing consistent color temperature and intensity. Advanced color calibration methods, including spectro-photometric reference standards and multi-illuminant normalization, may notably enhance the method's robustness to lighting variations. These improvements may lead to reliable deployment across diverse field conditions and geographic locations.

This research opens several promising avenues for future work and development. Multimodal approaches

incorporating additional soil properties (e.g., moisture content, organic matter, and nutrient levels) may expand soil characterization. The integration of spectral imaging with the established method may enhance discrimination by incorporating wavelength-specific information beyond the visible spectrum.

Advanced semantic integration approaches building on recent developments in knowledge-based computer vision may incorporate domain-specific agricultural knowledge to improve the classification accuracy for challenging samples in accordance with an autoregressive integrated moving average with semantic information (Juraphanthong &

Kesorn, 2025). The development of uncertainty quantification mechanisms may allow the method to provide confidence estimates for classification decisions, supporting informed decision-making in agricultural applications.

Our method addresses the challenging problem of pH-level classification in soil sample images, suggesting broad potential for applying advanced computer vision techniques to agricultural problems. The established principles and methods provide foundations for tackling similar problems in agricultural monitoring, crop health assessment, and environmental management that are essential for addressing global food security challenges.

Table 7. Comparative experimental results

Model	Input size	Density	Total No. parameters	Learning rate	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Linear Regression (Barman & Choudhury, 2019)	1,035 × 1,035	N/A	9 (0.000198MB)	N/A	33.89	46.31	33.89	30.76
KNN Regression knn_neighbors = 5 (Barman & Choudhury, 2019)	1,035 × 1,035	N/A	11,520 (0.35MB)	N/A	52.22	59.50	52.22	52.67
ANN with Levenberg-Marquardt and Scaled Conjugate Gradient Backpropagation Algorithms (Barman & Choudhury, 2019)	1,035 × 1,035	1 × 128, 1 × 64, 1 × 32	12,388 (0.05 MB)	1.00 × 10 ⁻⁴	55.00	63.29	55.00	55.22
SoilNet with preprocessed LR 0.01, min_lr = 0.0001) (Pandiri et al., 2024)	728 × 728	2 × 4	723,984 (2.76 MB)	1.00 × 10 ⁻²	45.19	45.44	44.81	43.14
SoilNet without preprocessed RGB image LR 0.01, min_lr = 0.0001) (Pandiri et al., 2024)	728 × 728	2 × 4	723,984 (2.76 MB)	1.00 × 10 ⁻²	58.61	58.40	58.61	58.41
Proposed ensemble-feature classification model	299 × 299	1 × 2,048	57,517,284 (219.41 MB)	1.00 × 10⁻⁴	82.22	82.01	82.22	82.10

*All benchmark models were evaluated on the SoilPh dataset (specialized soil image dataset developed for this study)

5. CONCLUSION

We present major advancements in automated soil pH-level classification through the development and validation of an ensemble-feature deep learning method. The integration of CNNs with GLCM features represents a paradigm shift in soil sample image analysis, achieving

an unprecedented accuracy of 82.22%, substantially outperforming existing methods. The main contribution of this study lies in the systematic fusion of hierarchical deep learning representations with statistical texture descriptors, creating a comprehensive feature space that captures both global structures and fine-grained texture variations critical for soil pH-level classification. Inception ResNetV2 with

GLCM emerges as the optimal architecture, demonstrating that advanced inception modules combined with texture features provide superior discriminative capabilities for challenging agricultural datasets. The systematic evaluation across multiple architectures (i.e., InceptionV3, Inception ResNetV2, ResNet50, and VGG16) provides robust evidence for the method's generalizability and establishes clear guidelines for future research on soil analysis.

A comprehensive comparison with existing methods demonstrates the substantial advancement achieved through ensemble features. The 48.33% improvement over conventional machine learning methods (Barman & Choudhury, 2019) and 23.61% improvement over deep learning methods (Pandiri et al., 2024) provide strong evidence for the superiority of our proposed framework. These improvements are remarkable given the challenging nature of pH-level classification in soil sample images and subtle visual differences that characterize samples across pH levels.

The method's ability to maintain consistent performance across pH levels (6–9) while handling challenging intraclass similarity represents an advancement in agricultural image analysis. A comprehensive evaluation using various metrics (i.e., accuracy, precision, recall, and F₁-score) confirms the method's reliability and practical applicability.

The proposed method addresses critical limitations in conventional soil pH-level classification by providing a scalable and cost-effective alternative to laboratory chemical analysis. Its capacity to process field images under natural lighting enhances its practical deployment potential, possibly supporting real-time decision-making in agricultural management. Integrating our proposal into existing precision agriculture workflows may reduce evaluation time and costs while enabling large-scale soil monitoring unattainable through traditional methods. The findings provide evidence for the importance of texture features in soil property assessment and offer a replicable method for addressing similar challenges in other agricultural classification tasks. The method's robust performance across diverse soil conditions suggests broader applicability to various agricultural contexts and geographic regions.

Overall, this study contributes to the advancement of intelligent agriculture by bridging the gap between laboratory-grade accuracy and field-deployable technology. The established method paves the way for sophisticated soil analysis tools that may enhance agricultural decision-making and support sustainable farming through enhanced soil management.

AUTHOR CONTRIBUTIONS

A.S. devised the project, established the main conceptual framework, and outlined the analytical strategy. A.S. encouraged S.M. to explore approaches for addressing interclass similarity and intraclass variance in soil classification and supervised the development of the methodology. S.M. conducted the computational analysis and performed the experiments. A.S. validated the analytical methods, supervised the study, and contributed to the interpretation of the results. A.S. led the manuscript writing, validation, and visualization, while S.M. drafted the original manuscript with guidance from A.S. A.S. also oversaw project planning and funding acquisition. All authors discussed the findings and contributed to the final

manuscript. The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

REFERENCES

- Barman, U., & Choudhury, R. D. (2019). Prediction of soil pH using smartphone based digital image processing and prediction algorithm. *Journal of Mechanics of Continua and Mathematical Sciences*, 14(2), 226–249. <https://doi.org/10.26782/jmcms.2019.04.00019>
- Barman, U., & Choudhury, R. D. (2020). Soil texture classification using multi class support vector machine. *Information Processing in Agriculture*, 7(2), 318–332. <https://doi.org/10.1016/j.inpa.2019.08.001>
- Juraphanthong, W., & Kesorn, K. (2025). Autoregressive integrated moving average with semantic information: An efficient technique for intelligent prediction of dengue cases. *Engineering Applications of Artificial Intelligence*, 143, Article 109985. <https://doi.org/10.1016/j.engappai.2024.109985>
- Mancini, M., Weindorf, D. C., Monteiro, M. E. C., de Faria, Á. J. G., dos Santos Teixeira, A. F., de Lima, W., de Lima, F. R. D., Dijair, T. S. B., Marques, F. D., Ribeiro, D., Silva, S. H. G., Chakraborty, S., & Curi, N. (2020). From sensor data to Munsell color system: Machine learning algorithm applied to tropical soil color classification via Nix™ Pro sensor. *Geoderma*, 375, Article 114471. <https://doi.org/10.1016/j.geoderma.2020.114471>
- Mendoza-Bernal, J., González-Vidal, A., & Skarmeta, A. F. (2024). A convolutional neural network approach for image-based anomaly detection in smart agriculture. *Expert Systems with Applications*, 247, Article 123210. <https://doi.org/10.1016/j.eswa.2024.123210>
- Mungklachaiya, S., & Salaiwarakul, A. (2025). Attention-X: Enhancing the classification of natural attraction scenes with advanced attention mechanisms. *Engineering and Applied Science Research*, 52(3), 337–351. <https://ph01.tci-thaijo.org/index.php/easr/article/view/259517>
- Neina, D. (2019). The role of soil pH in plant nutrition and soil remediation. *Applied and Environmental Soil Science*, 2019(1), Article 5794869. <https://doi.org/10.1155/2019/5794869>
- Pandiri, D. N. K., Murugan, R., & Goel, T. (2024). Smart soil image classification system using lightweight convolutional neural network. *Expert Systems with Applications*, 238(Part D), Article 122185. <https://doi.org/10.1016/j.eswa.2023.122185>
- Phiphatkamtorn, P., & Jitanan, S. (2025). Comparison of three learning approaches for plant disease classification. *ICIC Express Letters Part B: Applications*, 16(1), 63–71. <https://doi.org/10.24507/icicelb.16.01.63>
- Pramote, O.-U., Khruahong, S., Jitanan, S., & Kong, X. (2023). Quality grading of crown flowers using convolutional neural network. *ICIC Express Letters*, 17(2), 143–152. <https://doi.org/10.24507/icicel.17.02.143>
- Ramos-Ospina, M., Gomez, L., Trujillo, C., & Marulanda-Tobón, A. (2024). Deep transfer learning for image classification of phosphorus nutrition states in individual Maize leaves. *Electronics*, 13(1), Article 16. <https://doi.org/10.3390/electronics13010016>
- Salaiwarakul, A. (2023). Soil classification using MobileNet model. *ICIC Express Letters Part B: Applications*, 14(8), 795–802. <https://doi.org/10.24507/icicelb.14.08.795>

- Shah, S. R., Qadri, S., Bibi, H., Shah, S. M. W., Sharif, M. I., & Marinello, F. (2023). Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A case study on early detection of a rice disease. *Agronomy*, 13(6), Article 1633. <https://doi.org/10.3390/agronomy13061633>
- Sharma, H. K., & Kumar, S. (2018). Soil classification & characterization using image processing. In *Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018)* (pp. 885–890). IEEE. <https://doi.org/10.1109/ICCMC.2018.8488103>
- Srivastava, P., Shukla, A., & Bansal, A. (2023). Convolutional neural network to predict soil texture classes. In A. Rani, B. Kumar, V. Shrivastava, & R. C. Bansal (Eds.), *Signals, machines and automation* (pp. 589–595). Springer. https://doi.org/10.1007/978-981-99-0969-8_61