International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

Evaluating Soil Attributes and Recommendation of Crop Using KFSMOTEC Approach in CNN Model

Sharmila G[1], Kavitha Rajamohan[2] and Srinivasan Ramasamy[3]
[1, 2] Department of Computer Science, CHRIST (Deemed to be University)
[3]Senior Scientist (Soil Science), National Bureau of Soil Survey and Land Use Planning
[1]sharmila.g@res.christuniversity.in, [2] kavitha.r@christuniversity.in and [3]R.Srinivasan@icar.gov.in
[1]0000-0001-6179-1393 and [2]0000-0002-7803-8901

Abstract: *In modern agriculture, precise soil analysis and crop prediction are pivotal for optimizing agricultural practices and ensuring sustainable food production. This paper focuses on a comprehensive approach that integrates soil pH and nutrient detection with subsequent crop prediction, harnessing advanced image processing and machine learning techniques. First the input soil image is preprocessed. After preprocessing, Active contour models are utilized to accurately delineate the soil regions within input images. This segmentation is a fundamental step in subsequent PH and soil nutrient detection. For soil nutrient detection, an Extreme Learning Machine has been employed with Ensemble Kernels (EK-ELM). The goal of developing this novel EK-ELM method is to classify nutrients by the result obtained by aggregating different kernel functions. All the kernel functions were optimized using a weighted Ensemble by prior knowledge. This robust machine learning algorithm effectively processes the segmented soil images, extracting essential nutrient information from the soil's visual properties. Furthermore, the novel approaches K-Fold cross validation & Synthetic Minority Over-Sampling Technique implemented with Convolutional Neural Network (KFSMOTECNN) which comes into play for crop prediction. KFSMOTECNN have demonstrated their prowess in image analysis tasks. Here, they analyze the detected soil pH and nutrient information alongside additional environmental factors. By learning intricate relationships between these variables, KFSMOTECNN facilitate accurate crop yield predictions, aiding farmers in selecting suitable crop varieties. The proposed KFSMOTECNN approach provides 99% accuracy, 98.5% sensitivity, 98.62% specificity, 98% precision, and 98.21% F1-score for real soil dataset. These results unequivocally establish the superiority of our proposed approach -based crop prediction method when compared to conventional approaches.*

Keywords: *Precision agriculture, Soil analysis, Soil pH, Nutrient detection, Crop prediction, ELM and CNN*

1. INTRODUCTION

Technological developments are driving a revolutionary revolution in agriculture, the foundation of the world's food supply. The agricultural domain is a highly significant and delicate area of research. In consultation with farmers, citizens assist farmers in identifying novel solutions for current systems, conducting research, and implementing innovative farming practices. Additionally, government funding is provided to support new agricultural research and innovation, which directly benefits farmers. Nearly 54.6% of India's workforce, as reported by the 2011 census, works in agriculture and related sectors, which contributed seventeen percentage of country's Gross Value Added in 2017–2018. In 1985, the Ministry of Agriculture & Farmers Welfare announced insurance programme to protect crop against the dangers associated with farming. Issues have surfaced with the technological plan to gather information and reduce the time it takes to answer farmers' insurance claims.

For this, crop yield estimation is required, and it is measured by the Government of India's Crop Cutting Experiments (CCE), which are carried out in several state regions. In order to enhance the caliber of data gathering for Crop Cutting Experiments, Soil PH & Nutrients Prediction, and Crop Prediction based on Soil PH & Nutrients, remote sensing is used to record GPS data, including field elevation, area, latitude, and longitude [1, 2, 3, and 4]. Throughout the year, the phonological profiles of the crops are tracked by the vegetation indices obtained from the satellite photos [5, 6].

Advanced technologies like internet of things, data mining, data science, and artificial intelligence help in precision agriculture. The Internet of Things, or IOT, is a network of computationally interconnected objects, such as smart devices and sensors, which can exchange data and communicate with one another [7-10]. Among the many difficulties facing modern agriculture are the needs to maximize resource use, increase crop output, and lessen the effects of environmental variables. The integration of cutting-edge technologies has become crucial to addressing these issues; two such technologies that have completely changed the area are machine learning and deep learning [11, 12].

Machine and deep learning algorithms, especially those related to image processing, have shown to be very useful for determining the pH of soil, identifying nutrient values, and forecasting crop yields. Crop forecast, nutrient measurement, and soil pH were traditionally labour-intensive processes that relied on manual sampling and arbitrary evaluations. However, a new era of precision agriculture has begun with the advancement of techniques in machine and deep learning. Agronomic scientists and practitioners can now take advantage of large-scale image data by using neural networks, specifically Extreme Learning Machine and convolutional neural networks, to make timely and accurate decisions about soil nutrients and pH value.

While rapid training speed and good generalisation performance are two advantages of ELM, poor robustness is a drawback. In order to generate a forward extreme learning machine with the kernel (KELM), the extreme learning machine and the kernel approach are coupled to tackle this challenge. While KELM is generally fast and exhibits good generalisation, there is still need for improvement in terms of stability and generalisation performance. To solve this problem, this work developed a novel approach called as Extreme Learning Machine with Ensemble Kernels (EK-ELM) for soil nutrient detection. Despite being the most widely used method in deep learning, CNN has a number of drawbacks. Large feature maps require a very high computational cost when using CNN. CNN takes a while to train on big feature maps. To overcome the drawback this paper introduced a new approach named as K-Fold Synthetic Minority Over-sampling Convolution Neural Network (KFSMOTECNN) for crop prediction. In the process of crop prediction using image processing, soil pH value & nutrient detection this research investigates the function of machine and deep learning methods. The benefits, potential, and practical uses of these algorithms in converting agriculture into a data-driven, effective, and sustainable enterprise are explored in this work. The future of agriculture is expected to be significantly shaped by deep learning, which has the potential to optimise everything from crop production to soil health.Below are the main objectives of this study:

1. This paper presents a comprehensive approach that integrates multiple facets of modern agriculture, including soil pH and nutrient detection, along with crop prediction.
2. The utilization of active contour models results in precise and reliable soil region segmentation within input images. This contributes to the accuracy of subsequent soil nutrient detection.
3. The implementation of EK-ELM for soil nutrient detection yields impressive results. It accurately extracts nutrient information from soil visuals, enabling informed soil management decisions.
4. By integrating KFSMOTECNN into the analysis process, this work enhances crop prediction accuracy. This empowers farmers to make data-driven decisions regarding crop selection.

The remaining part of this article is structured as follows:
The remainder structure of the paper is as follows: Section 2 conducts a thorough review of existing literature concerning soil nutrient detection and crop prediction through the utilization of both machine and deep learning algorithms. Section 3 presents processing of soil pH value, soil nutrient analysis and novel crop prediction algorithm were detailed its architecture and methodologies. Subsequently, Section 4 is dedicated to the discussion of the experimental outcomes, offering insights into the practical performance of our approach. Finally, in Section 5, these works present our concluding remarks and summarize the key contributions and implications of this research.

2. RELATED WORK

The macronutrient and trace mineral levels in the soil are vital in crop growth. The term "soil" encompasses a wide range of environmental factors such as temperature, humidity, sunlight, rainfall, along with soil pH & nutrients present [13]. To predict crop types effectively, a method has been proposed that employs a combination of decision tree and support vector machine algorithms, taking into account micronutrient levels and climatic conditions. The study focused on three major crops: rice, wheat, and sugarcane. Specific data regarding micronutrients were collected and utilized as input for the classifier model, which made predictions based on these data. Given the differences in the functioning of various machine learning algorithms, employing only two models may not yield the desired results. Notably, the SVM achieved a higher accuracy score of 92% compared to the decision tree algorithm [14]. Unlike the approach in [14], this study explored alternative algorithms. Predictions of crop yield were made using a linear regression model, considering meteorological variables like humidity, temperature, and rainfall. However, each of these methods achieved an accuracy score of less than 90% [15].

A deep convolutional neural network (DCNN), including multiple layers of neurons, is trained to recognize complex patterns within accurately labeled image data. The InceptionV3 model is employed as traditional feature extractor of image to assist in the classification of background and fruit pixels within images. This classifier, capable of localizing fruits, is used to determine the number of fruits present and to categorize tomato species. In the training datasets, fruit pixels are identified using a K-nearest neighbor (KNN) classifier, with a designated pixel value threshold used to classify pixels as fruit. The Support Vector Machine (SVM) is then utilized with linear regression analysis and pattern categorization based on selected features [16-18]. To achieve a remarkable 96% accuracy in tomato berry detection, the AdaBoost model is used as a traditional classifier. It accomplishes this with linearly combining the weak classifiers that has minimal thresholding and employing Haar-like features.

For the detection of iceberg lettuce [19] and grapes [20] with edges suitable for harvesting by a Vegebot, a pre-trained "this work only look once" (YOLO) model in a darknet classifier is employed.YOLO models offer superior real-time object identification capabilities compared to faster region-based CNN (FRCNN).A multi-modal faster region-based CNN model enhances fruit yield detection by merging RGB and near-infrared images, resulting in a performance boost to an impressive 0.83 F1 score [21]. The region based R-CNN model is trained on dataset images to create a feature map for classification purposes. Additionally, a spiking neural network (SNN) is trained on spatiotemporal analysis using image of remote sensing data, specifically normalized differential vegetation indicators. This enables the planning of crop yield predictions and the estimation of winter wheat yields.

In context of apple yield prediction, a backpropagation neural network (BPNN) model employs techniques of segmenting image and canopy features to train the system. Remarkably, classifiers SVM and KNN demonstrate exceptional efficiency of accuracy rates 98.49% and 98.50%, respectively [22]. Deep convolutional neural networks are employed to detect plant diseases and forecast macronutrient deficits during flowering and fruit development stages [23]. Plant issues are detected through the Visual Geometry Group (VGG) CNN architecture using plant leaf photos and this information is conveyed to farmers via smartphones [24]. An adaptive deep CNN, trained with Image-net datasets, validates and enhances the diagnosis of endemic fungal infections in winter wheat. Soil texture classification also helps in precision agriculture greatly; these were discussed in [25-27].The models used were multiclass support vector machine, machine learning models and active contour model with neural networks. The papers [28-31] discuss the processing the soil nutrient content which is an essential component in crop prediction. The machine learning models, genetic algorithm with back propagation, extreme learning neural networks with parameters and optimal deep neural network models were discussed.

3. PROPOSED SOIL NUTRIENTS AND CROP PREDICTION SYSTEM

This work, introduce a comprehensive methodology that integrates various advanced techniques in the domain of soil pH, nutrients analysis and crop prediction. The main blocks of the proposed soil pH

and nutrients analysis and crop prediction algorithm shown in Fig. 1.

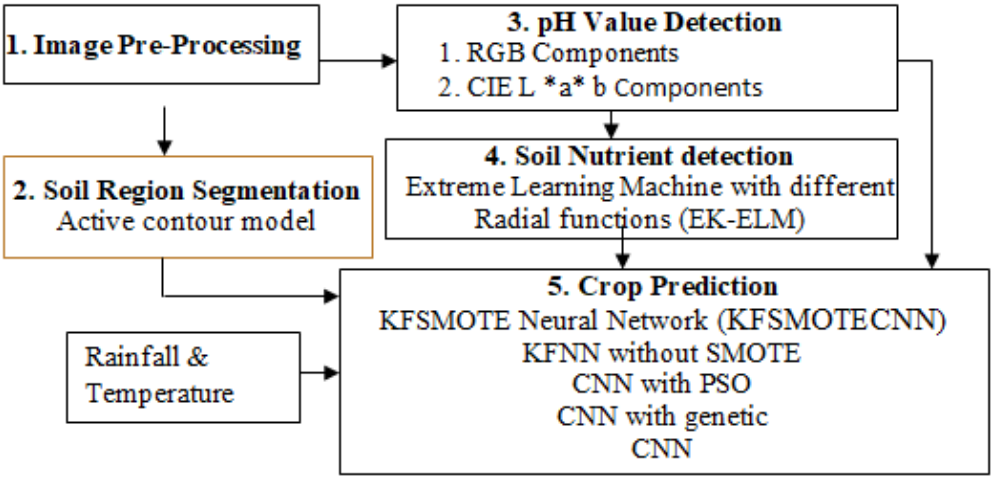


Fig. 1 Overall Diagram of Proposed Systems

3.1 Pre-Processing

Noise removal using a median filter is a common image processing technique, especially when dealing with salt-and-pepper noise, which appears as isolated bright and dark pixels in the image. Each pixel's value is replaced with the median value using median filter & the pixels in a local neighborhood around it. This process helps to reduce the impact of outliers or noisy pixels. After noise removal process completed, the next step is to enhance the noise removed image using CLAHE approach. CLAHE is a technique enhances the contrast in images while also limiting amplification of noise. It is particularly useful for improving the visibility of details in images with non-uniform lighting and varying contrast.

3.2 Soil Region Segmentation

The segmentation method most frequently employed to divide up a region of interest in the spatial domain of images is the active contour model (ACM). It is utilized in this work to separate the soil area from the previously processed picture. The suggested method divides the Fourier coefficients into bands of lower and higher frequency by employing ACM in Fourier domain rather in spatial domain. A starting contour $\hat{C}$ that is circular is used in the suggested method. The Fourier coefficients' magnitudes are subjected to the ACM. The initial contour will travel towards a point where the Fourier coefficient magnitude abruptly varies due to both internal and external influences operating within the Fourier domain. A sudden change in the magnitude coefficient of Fourier moves the contour internal that produce the smooth deformation. Let z(c) be contour representation indicated by

$$z(c) = [x(c), y(c)] \quad (1)$$

The total energy is given by the contour is

$$T_e = \int T_c\, z(c)\ dc = \int T_{int}\big(z(c)\big) + T_{ext}\big(z(c)\big)dc \quad (2)$$

The external energy of the contour in this case is T_{ext} which is the total of the external constraint force and image force. The total of the energy from the snake contour and the spline curvature is the internal energy. The evaluation results in classifying the type of the soil outperform traditional approaches.

3.3 pH detection

The Soil pH detection algorithm which is used in the work is designed to estimate soil pH value based on input soil images Figure 1. It begins by obtaining the input soil image and then proceeds to remove noise from it using a filtering technique, such as median filtering. Next, the algorithm converts the color space of RGB from noise-removed image to CIE Lab color space, which involves several

mathematical transformations. This color space conversion is essential for extracting meaningful color information from the image.

Y = 0.299 * (Red + 0.587) * Green + 0.114 * Blue

U = (Blue - Y) * 0.565

V = (Red - Y) * 0.713

L* = 116 * (Y/Yn)^1/3 - 16

a* = 500 *[f(X/Xn) - f(Y/Yn)]

b* = 200 * [f(Y/Yn) - f(Z/Zn)] (3)

where R, G, B are the components red, green, and blue of the RGB image. Y, U, and V are components of the YUV color space. (L*, a*, b*) are Lab color space components of CIE. Xn, Yn, and Zn are the reference to the white point values. f(t) is a function that maps linear light intensity to a non-linear response.

Afterward, the color-converted image is subjected to an adaptive histogram equalization approach to enhance its contrast, which improves the visibility of soil features. Subsequently, the algorithm segments the contrast-enhanced image to identify the soil region. Various image segmentation techniques can be employed for this purpose, such as thresholding or active contour models.Once the soil region is successfully segmented, the algorithm separates the L*, a*, and b* color channels from the segmented image. It then calculates the average value of the b* channel by dividing it by the L* channel. Similarly, the average value of the a* channel is computed by dividing it by the L* channel.

Average _b = Σ(b*) / N (4)

Where, Average_b is the average value of the b* channel. Σ(b*) is the sum of all pixel values in the b* channel within the soil region. N is from soil regions total number of pixels. These averages are evaluated by summing pixel values within the respective color channels within the soil region and divided by total number of pixels in that region. The algorithm then identifies maximum value between the two average values calculated for the a* and b* channels. This maximum value, referred to as "Max_Average," is considered the estimated soil pH value.

Max_Average = max(Average_a, Average_b) (5)

3.4 Soil Nutrient Detection

Soil nutrient detection in this segment employs the use of Extreme Learning Machine with Ensemble Kernels Fig. 2. Extreme Learning Machines are a versatile solution for various types of generalized feed-forward networks, encompassing feed-forward neural networks, RBF networks, and kernel learning, among others. ELM exhibits distinct capabilities for handling both regression and multi-class classification tasks. As a result, ELM presents numerous benefits, including rapid learning, straightforward implementation, and minimal need for human intervention.

ELM boasts the benefit of rapid training and excellent generalization performance. However, it does suffer from poor reliability. To address, an integrated kernel method with Extreme Learning Machine, resulting in the creation of a Kernel-based Forward Extreme Learning Machine (KELM). While KELM often exhibits speed and solid generalization, there is still room for enhancing stability and generalization performance further.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

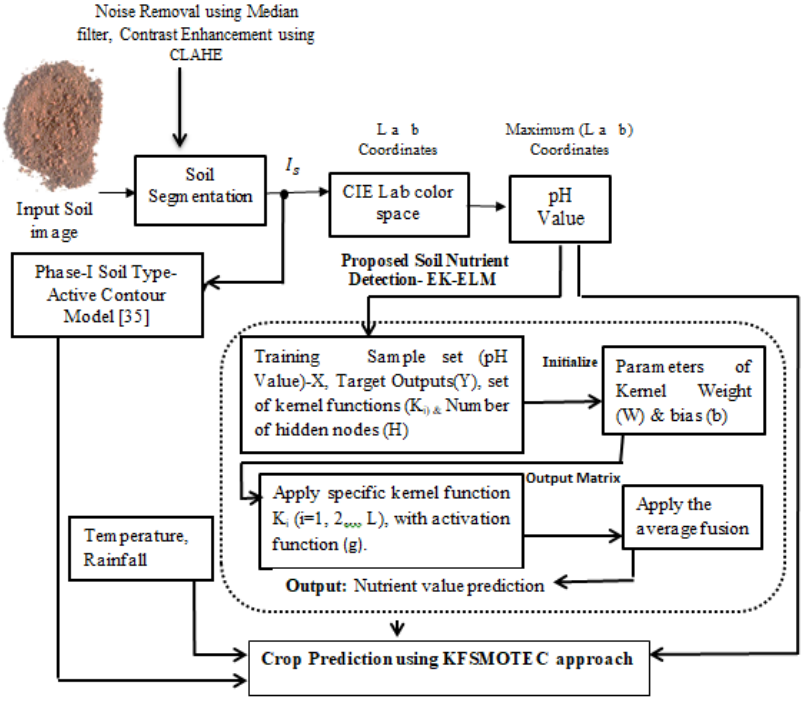


Fig. 2 Proposed System for Soil Nutrients

This approach is known as the Kernel Extreme Learning Machine (KELM), wherein the conventional activation function is substituted with a kernel function. In many instances, KELM exhibits swiftness and demonstrates robust generalization. Nonetheless, there is still room for enhancing both stability and generalization performance. To address this, a novel model known as "Extreme Learning Machine with Ensemble Kernels" (EK-ELM) has been introduced Fig. 3. The concept of an "ensemble of kernels" is central to this method, which involves the aggregation of values from multiple kernels.

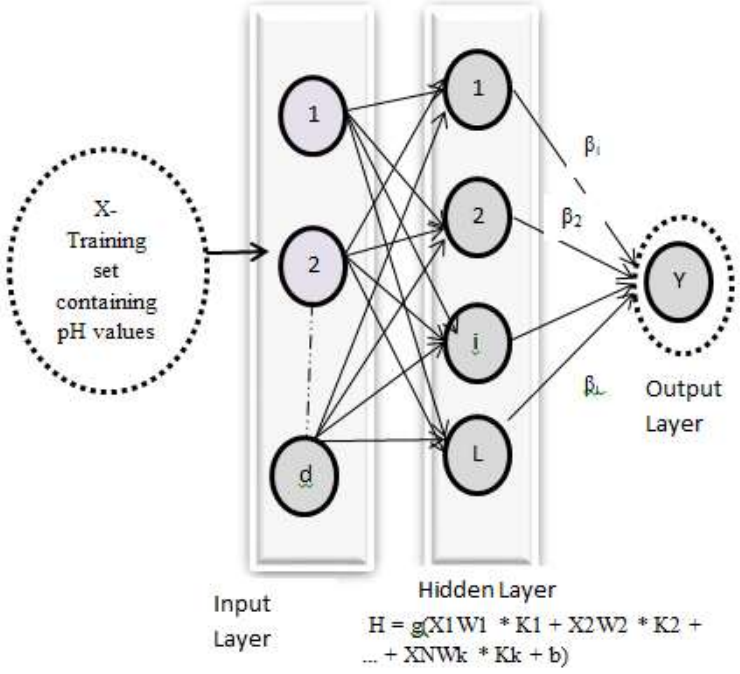


Fig. 3 EK-ELM Model for Soil Nutrient

In this approach, the kernel function is not a single entity but rather a weighted ensemble of kernel functions, with these weights being determined by knowledge prior. The "ensemble-of-kernels" model explores effects of the cumulative through a combination linear of distinct generating functions. The primary objective of this methodology is to obtain an ensemble result of kernels by incorporating functions of kernel in classification problems. EK-ELM can be outlined in the following steps.

a) **Feature Extraction**: Collect and preprocess soil-related data, which may include physical properties of the soil (e.g., temperature, pH, rain fall).

b) **Kernel Construction**: Define multiple kernels, each representing a different type of feature or information source. Calculate the kernel matrices for each feature type. These matrices capture the pairwise similarities or relationships between data points in the feature space. In this work kernel functions include the Linear, Quadratic, Radial Basis, Polynomial and Multi-Layer kernel.

c) **EK-ELM Model**: The EK-ELM model aims to combine the information from multiple kernels to create an effective predictive model. It extends the basic ELM framework to handle multiple kernels.

$$H = g(W1 * K1 + W2 * K2 + ... + Wk * Kk + b) \quad (6)$$

Where, **H** is hidden (feature mapping) layer output. **g()** is the activation function, typically radial basis function (RBF) is used. **W1, W2, ..., Wk** are the weights associated with each kernel. **K1, K2, ..., Kk** are the kernel matrices for each feature type. **b** is the bias term.

d) **Output Layer**: The EK-ELM models output layer is trained to predict soil nutrient levels based on the extracted features from hidden layer. This work used classification techniques in the output layer, depending on the nature of the nutrient data.

e) **Training**: Use a training dataset to learn the weights **W1, W2, ..., Wk** and the bias **b** in the hidden layer. This can be done using optimization techniques like gradient descent or through the use of matrix pseudo-inversion.

f) **Detection**: After training, the EK-ELM model predicts soil nutrient levels.

Algorithm 2: Ensemble Kernel - ELM Algorithm that predicts soil nutrient levels

Input: pH values- Training set, validation data, set of kernel functions, and number of hidden nodes.

Output: Soil nutrient

1: Initialize weights of input layer and output layer biases with random values.

2: Initial parameters were generated for kernel functions randomly.

3: For each kernel (k = 1, 2, ..., L):

a) Calculate output matrix of hidden layer using each kernel.

b) Calculate output weight (β) for this kernel.

c) Utilize the output weight matrix (β) to make predictions.

4: Apply the approach of average fusion to combine all output weight matrix values, resulting in the final nutrient value prediction.

This algorithm 2 takes pH values as input and leverages a Ensemble-kernel approach to estimate soil nutrient details by iteratively learning and fusing information from different kernel functions.

3.5 Crop Prediction

In this part of work, deep learning model are used for predicting crop Fig. 4. The three layers are:

1. Convolution Layer

2. Pooling Layer
3. Fully Connected Layer

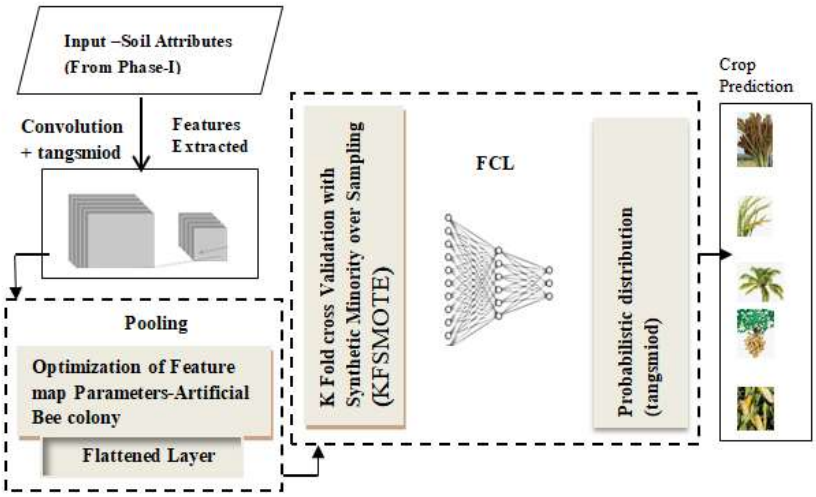


Fig. 4 Proposed System for Crop Prediction System

The convolution layer serves purpose of feature extraction in Convolutional Neural Networks (CNNs). Additionally, the pooling layers are often referred to as subsampling layers. Typically, CNNs are employed for classification tasks. The pooling layer's primary function is to conduct down sampling, reducing the computational load by condensing the features extracted in the convolutional layer. The pooling layer's output serves as input for subsequent layer. It's worth noting that CNNs can be computationally intensive, especially when working with large feature maps. Consequently, training CNNs on these large feature maps can be a time-consuming process. To overcome the drawback, the Artificial Bee Colony technique will be used. In last layer (Fully Connected layer) the prediction process will be done. In this layer result are validated by K Fold Cross Validation and final decision is made by SMOTE (Synthetic Minority Over-sampling) fusion approach. By introducing optimization technique in pooling layer and K Fold with SMOTE function in fully connected layer, it is a enhanced version of CNN. This method is called "KFSMOTECNN". Here are the key components and steps of a CNN:

a) **Convolutional Layer:** This is the core operation in a CNN. It involves applying convolutional filters (also known as kernels) to the input image to extract features. The convolution operation can be represented mathematically :

$$(I * K)(x, y) = \sum\sum I(i, j) * K(x - i, y - j) \quad (7)$$

Where, (I * K) (x, y) is the result of convolution at position (x, y). I(i, j) is value of pixel at position (i, j) from the input image. K(x - i, y - j) is kernel value at position (x - i, y - j). After convolution, an activation function is applied element-wise to each feature map. Common activation functions include tanh activation function. The activation function introduced is a non-linearity model that allows learning very much complex patterns.

$$A = \text{tansig}(X) \quad (8)$$

b) Pooling Layer: This layer used to reduce spatial dimensions of feature maps, which helps reduce computational complexity and over fitting. In this step, instead of max and average pooling, this work uses Artificial Bee Colony Approach to select the optimal features.

ABC Algorithm 3: ABC in Pooling Layer

Input: Extracted Features from Convolution Layer.

Output: Reduced Features

1: Generate a population of employed bees with random threshold values: Ti for i=1,2,...,N (where N is size of the population).

2: Evaluate objective function for each threshold:

J(Ti)=∑_(i=1)^k▒ 〖w_i ·σ_i^2 〗

3: Employed bees select a neighbor bee's solution (threshold) and modify it. The modification can be done using below equation:

Ti′=Ti+δ·(Tj−Tk)

Where, Ti′ is the updated threshold value for bee i. Tj and Tk are randomly selected threshold values from the population. δ is a random scaling factor.

(i) Calculate the fitness (objective function value) of the new solutions:

J(Ti′)=∑_(i=1)^k▒ 〖w_i ·σ_i^2 〗

(ii) Employed bees share information about their solutions with onlooker bees.

(iii) Onlooker bees select solutions based on their fitness and the information shared by employed bees. The probability of selecting solution i is proportional to its fitness:

Pi=(〖J(T〗 _i))/(∑_(i=1)^N▒ 〖 〖J(T〗 _i)〗)

4: Repeat Steps 2-4 for a predefined number of iterations or until convergence.

c) Fully Connected Layer: This layer is composed of one or more densely connected neurons that take the flattened output from the previous layer as input. A weighted sum from each of its input is computed and activation, typically tansig is used for crop prediction. In step the predicted result are validated by using K-Fold cross validation approach.

i. Partition Data into Folds:
- Divide the dataset D into K subsets, often referred to as "folds."

ii. Iterate Over Folds:
- For each iteration k from 1 to K, do the following:

a. Split Data:
- Use fold k as test set.
- Combine remaining folds to create training set.

b. Train Model:
- Train the model on training set.

c. Evaluate Model:
- Use the trained model in predicting on the test set.
- Calculate evaluation metrics using accuracy based on model's predictions and the ground truth labels for the test set.
- Record performance metrics for fold k.

iii. Aggregate Results:
- After all K iterations are complete, calculate the overall performance metrics by aggregating the results from each fold. Common aggregation methods include taking the mean or median of the fold-specific metrics. And then SMOTE algorithm is used Fig. 5.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

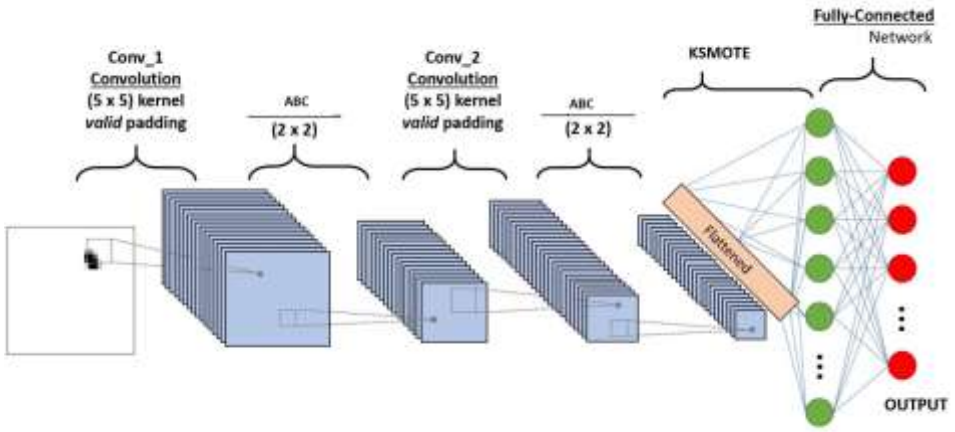


Fig. 5 Overall Diagram of KFSMOTECNN

Synthetic Minority Over-sampling Technique is a popular technique in machine learning used for addressing class imbalance problems, especially in classification tasks. Class imbalance occurs when the minority class is significantly underrepresented compared to another class majority class in a dataset. SMOTE is designed to balance class distribution by creating synthetic samples for the minority class. Here's an explanation of how SMOTE works:

4. EXPERIMENTAL RESULTS

The performance of the soil classification algorithm based on the proposed active contour model was assessed using standard evaluation metrics, including precision, specificity, sensitivity, F1-score, and accuracy. The evaluation was carried out on publicly available datasets, specifically the REAL dataset. These metrics, namely sensitivity, specificity, precision, F1-score, and accuracy, were used to gauge the algorithm's effectiveness.

4.1 Dataset Description

The REAL dataset has various soil types collected from rural area of Bangalore, Karnataka, India. The images of the soil were collected from different farm fields in 18 areas around the Chikkanayakanahalli Village, Tumkur District in the state of Karnataka. Totally 110 images of soil samples were collected around the Tumkur district. It consists of five soil types namely Sandy Clay, Clay, Loamy Sand, Sandy Loam and Sandy Clay Loam. The samples were collected at a distance of radius of approximately 3- 5 km with a depth of 1 foot in the ground. The images are captured in the size of 1280x1194 (pixels) approximately by using a compact digital camera with an effective megapixel CMOS image sensor and is stored in jpeg file format. Sample images from REAL dataset are visualized in Fig. 6. And all these categories of soil images were uploaded in kaggle repository [32] REAL Dataset. Table 1 represent the sample image number after image augmentation.

Algorithm 4: SMOTE Technique in Fully Connected Layer

Input: A dataset with imbalanced class distribution. The minority class, which needs to be oversampled. The number of synthetic samples to generate (often specified as a percentage of the original minority class size).

Output: Balanced classes

1: Select a Minority Class Sample: SMOTE starts by randomly selecting a minority class sample from the dataset. Let's call this sample A.

2: Find k Nearest Neighbors: A critical part of SMOTE is determining the neighborhood of A. It does this by calculating the Euclidean distance (or other distance metrics) between A and all other minority

class samples. The nearest neighbors k of A are then identified.

3: Randomly Select a Neighbor: From the k nearest neighbors, SMOTE randomly selects one neighbor. Let's call this neighbor B.

4: Generate Synthetic Samples: SMOTE creates synthetic samples by interpolating between A and B. It does this for each feature independently. For example, if A and B are two-dimensional points in a feature space, a synthetic sample S is created as follows for each i feature:

S(i) = A(i) + (random_number) * (B(i) - A(i))

Where random_number value between 0 and 1.

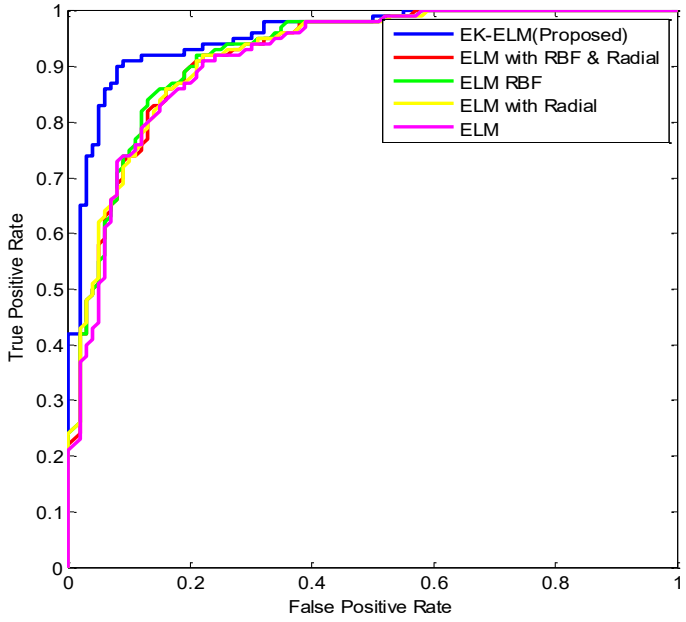
5. Repeat: Steps 1-4 are repeated until desired number of synthetic samples is generated.

Fig. 6 Sample images from REAL dataset

Table 1 Test cases used in evaluation of algorithms

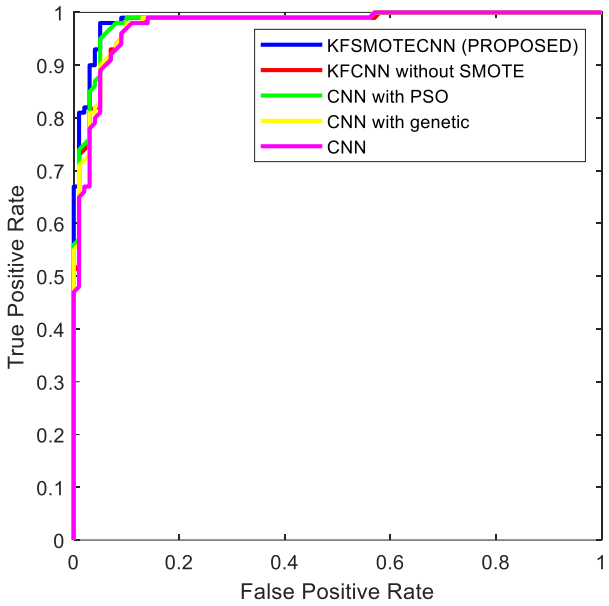
Test cases	REAL dataset	
	No of Training images	No of Test images
Case (1) 90:10	499	51
Case (2) 80:20	447	103
Case (3) 70:30	385	165

Fig. 7 (a) & (b) present a comparison of ROC curves between the proposed EK-ELM and proposed CNN approaches. In Fig. 7 (a) the area under the ROC for the proposed EK-ELM is observed to be greater than that of the other methods. Conversely, in Fig. 7 (b), the area under the ROC for the proposed KFSMOTECNN surpasses that of the other methods. Notably, both of the proposed approaches KFSMOTECNN & EK-ELM for REAL dataset consistently yield higher area under the ROC curves, indicating their superior performance in terms of predication accuracy.



(a)

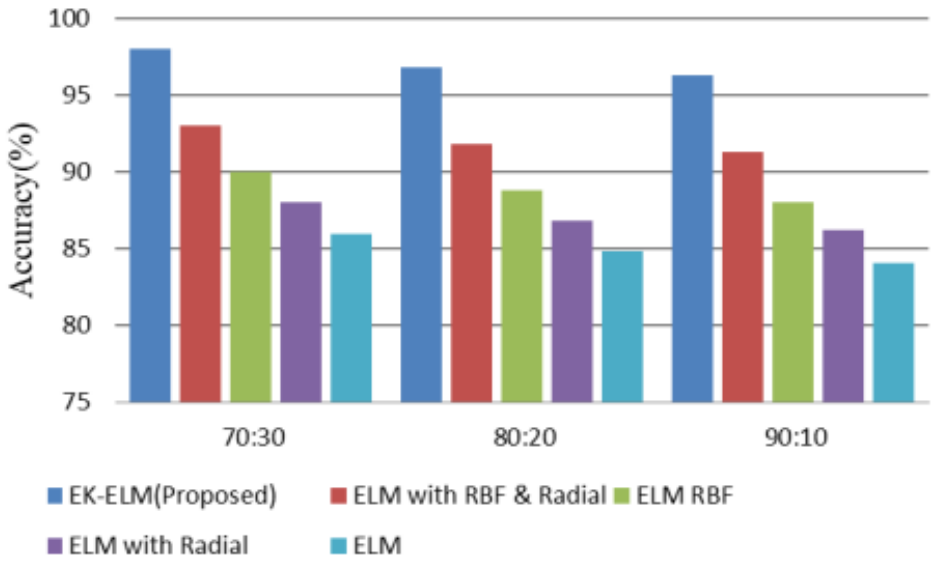
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php



(b)

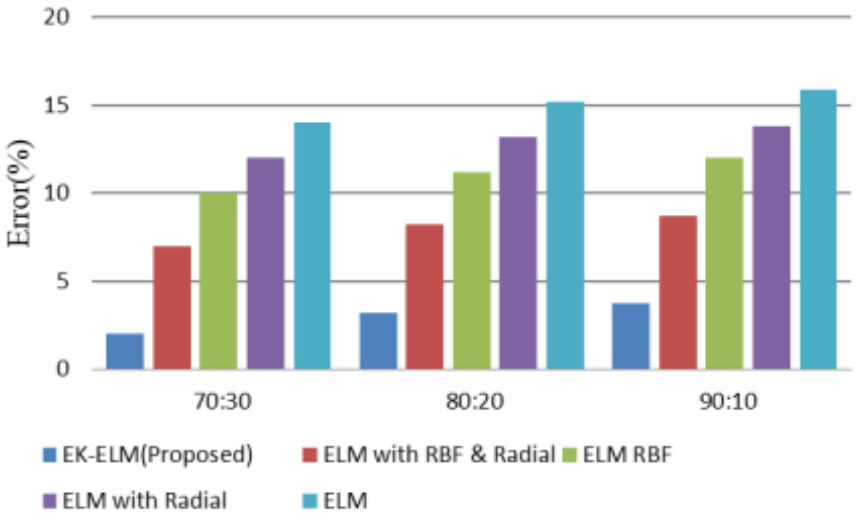
Fig. 7 ROC curve comparison between proposed schemes (a) Proposed EK- ELM-Soil Nutrient (b) Proposed KFSMOTECNN-Crop Prediction

The Fig. 8 summarizes nutrient detection accuracy values for different methods when tested on three different data splits (70:30, 80:20, and 90:10). The Fig. 8 (a) provides accuracy values for nutrient detection, with the proposed EK-ELM method consistently outperforming the other methods across all three data splits. The Fig. 8 (b) summarizes nutrient detection error values for different methods when tested on three different data splits (70:30, 80:20, and 90:10). EK-ELM (Proposed): Reports error values of 0.02, 0.032, and 0.037 for the respective data splits. ELM with RBF & Radial: Shows error values of 0.07, 0.082, and 0.087 for the three data splits. ELM RBF: Demonstrates error values of 0.1, 0.112, and 0.12 for the three data splits. ELM with Radial: Indicates error values of 0.12, 0.132, and 0.138 for the three data splits. ELM: Displays error values of 0.14, 0.152, and 0.159 for the three data splits.



(a)

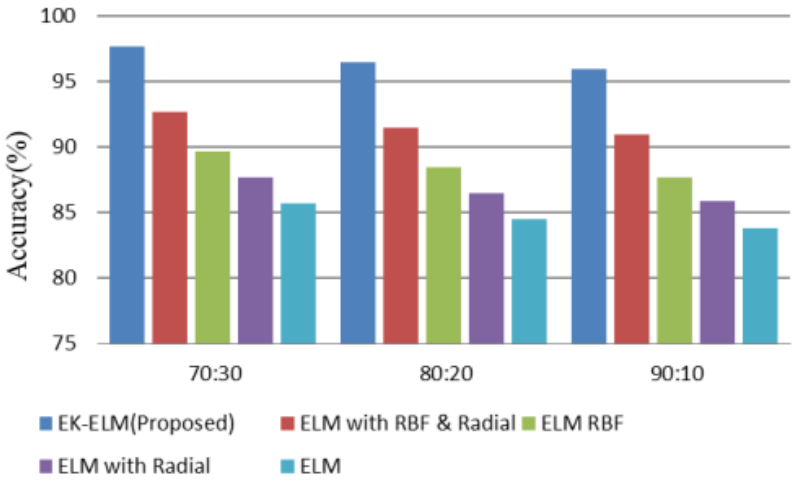
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php



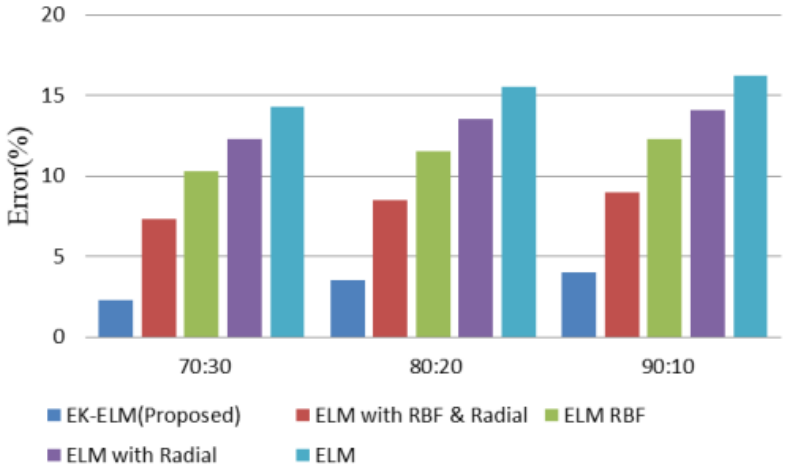
(b)

Fig. 8 Testing Accuracy and Error comparison for REAL datasets for different ratios a) Proposed EK-ELM's accuracy values (b) Proposed EK-ELM's error values

The Fig. 9 summarizes nutrient detection accuracy values for different methods during training on three different data splits (70:30, 80:20, and 90:10). The Fig. 9 (a) provides EK-ELM (Proposed): Achieves accuracy values of 0.977, 0.965, and 0.96 for the respective data splits during training. ELM with RBF & Radial: Attains accuracy values of 0.927, 0.915, and 0.91 during training for the three data splits. ELM RBF: Shows accuracy values of 0.897, 0.885, and 0.877 during training for the three data splits. ELM with Radial: Reports accuracy values of 0.877, 0.865, and 0.859 during training for the three data splits.ELM: Demonstrates accuracy values of 0.857, 0.845, and 0.838 during training for the three data splits.



(a)



(b)

Fig. 9 Training Accuracy and Error comparison for REAL datasets for different ratios (a) Proposed ELM's accuracy values (b) Proposed ELM's error values

The Fig. 9 (b) summarizes nutrient detection error values for different methods during training on three different data splits (70:30, 80:20, and 90:10). EK-ELM (Proposed): Reports error values of 0.023, 0.035, and 0.04 for the respective data splits during training. ELM with RBF & Radial: Shows error values of 0.073, 0.085, and 0.09 during training for the three data splits. ELM RBF: Demonstrates error values of 0.103, 0.115, and 0.123 during training for the three data splits. ELM with Radial: Indicates error values of 0.123, 0.135, and 0.141 during training for the three data splits. ELM: Displays error values of 0.143, 0.155, and 0.162 during training for the three data splits.

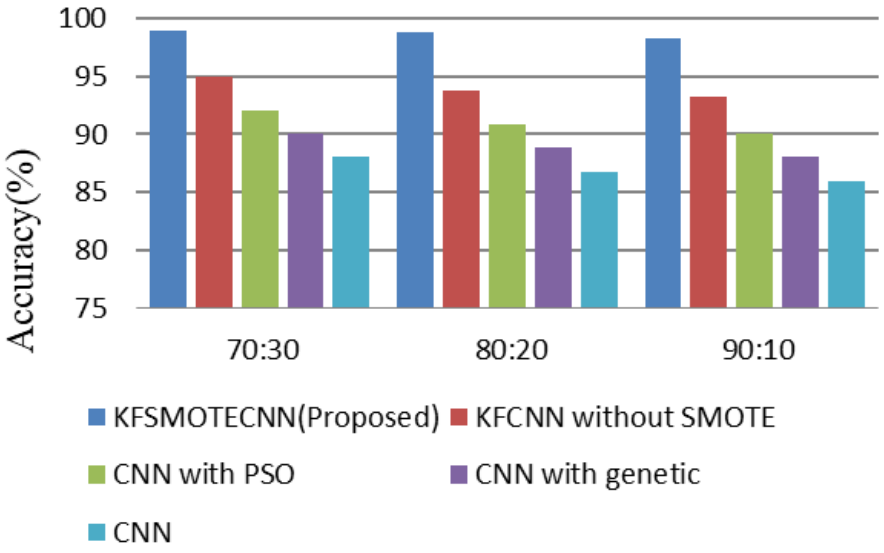
The Table 7 provides performance metrics for different soil nutrients detection methods, including EK-ELM (Proposed), ELM with RBF & Radial, ELM RBF, ELM with Radial, and ELM. In summary, the table presents a comprehensive comparison of the evaluated methods across multiple performance metrics, with EK-ELM (Proposed) consistently achieving the highest values in most of these metrics.

Table 7 Performance Comparison between the EK- ELM (proposed) & other models for the REAL dataset

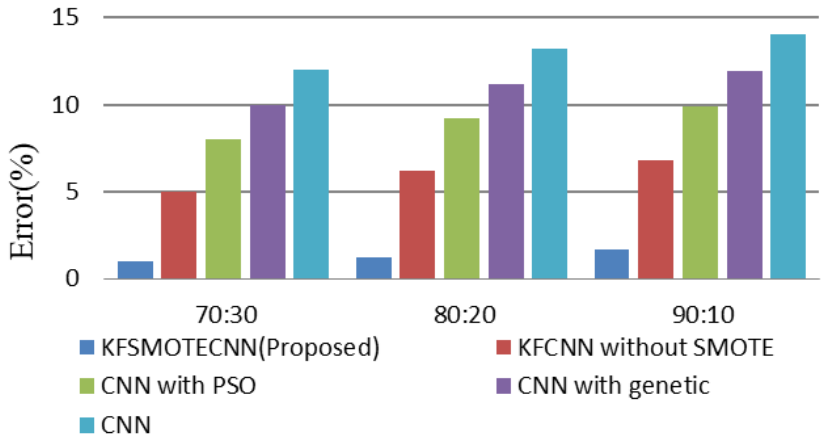
	EK-ELM (Proposed)	ELM with RBF & Radial	ELM RBF	ELM with Radial	ELM
Accuracy	0.98	0.93	0.9	0.88	0.86
Precision	0.9706	0.9206	0.8906	0.8706	0.8506
Recall	1	0.95	0.92	0.9	0.88
F1-score	0.9851	0.9351	0.9051	0.8851	0.8651
Sensitivity	0.9792	0.9292	0.8992	0.8792	0.8592
Specificity	0.998	0.948	0.918	0.898	0.878

The Fig. 10 (a) summarizes crop prediction accuracy values for different methods during testing on three different data splits (70:30, 80:20, and 90:10). KFSMOTECNN (Proposed): Achieves accuracy values of 0.99, 0.988, and 0.983 for the respective data splits during testing. KFCNN without SMOTE: Reports accuracy values of 0.95, 0.938, and 0.932 during testing for the three data splits. CNN with PSO (Particle Swarm Optimization): Shows accuracy values of 0.92, 0.908, and 0.901 during testing for the three data splits.CNN with Genetic Algorithm: Indicates accuracy values of 0.9, 0.888, and 0.881 during testing for the three data splits. CNN (Convolutional Neural Network) demonstrates accuracy values of 0.88, 0.868, and 0.86 during testing for the three data splits.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php



(a)



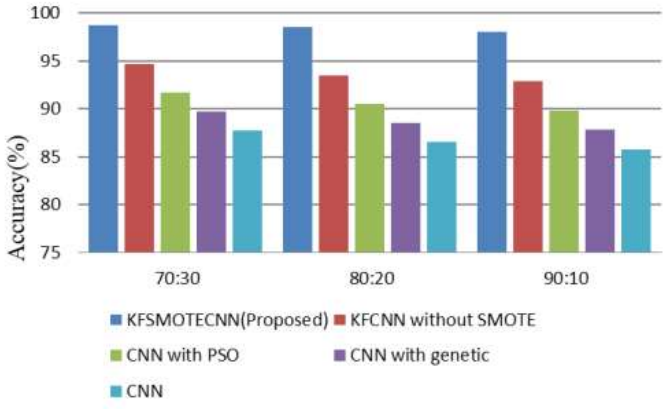
(b)

Fig. 10 Testing Accuracy and Error comparison for REAL datasets for different ratios (a) Proposed KFSMOTECNN's accuracy values (b) Proposed KFSMOTECNN's error values

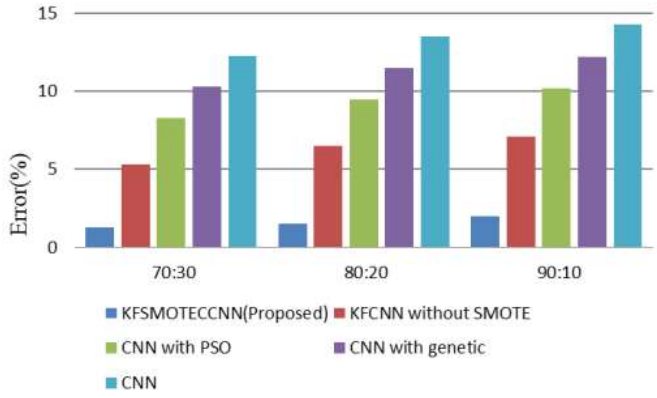
The Fig. 10 (b) provides crop prediction error values for different methods during testing on three different data splits (70:30, 80:20, and 90:10). The methods are as follows: KFSMOTECNN (Proposed): Reports error values of 0.01, 0.012, and 0.017 for the respective data splits during testing. KFCNN without SMOTE: Shows error values of 0.05, 0.062, and 0.068 during testing for the three data splits. CNN with PSO (Particle Swarm Optimization): Demonstrates error values of 0.08, 0.092, and 0.099 during testing for the three data splits. CNN with Genetic Algorithm: Indicates error values of 0.1, 0.112, and 0.119 during testing for the three data splits. CNN (Convolutional Neural Network) achieves error values of 0.12, 0.132, and 0.14 during testing for the three data splits.

The Fig. 11 (a) summarizes crop prediction accuracy values for different methods during the training phase on three different data splits (70:30, 80:20, and 90:10). The methods are as follows: KFSMOTECNN (Proposed): Achieves accuracy values of 0.987, 0.985, and 0.98 for the respective data splits during training. KFCNN without SMOTE: Reports accuracy values of 0.947, 0.935, and 0.929 during training for the three data splits. CNN with PSO (Particle Swarm Optimization) Shows accuracy

values of 0.917, 0.905, and 0.898 during training for the three data splits. CNN with Genetic Algorithm: Indicates accuracy values of 0.897, 0.885, and 0.878 during training for the three data splits. CNN (Convolutional Neural Network) demonstrates accuracy values of 0.877, 0.865, and 0.857 during training for the three data splits.



(a)



(b)

Fig. 11 Training Accuracy and Error comparison for REAL datasets for different ratios (a) Proposed CNN's accuracy values (b) Proposed CNN's error values

The Fig. 11 (b) provides crop prediction error values for different methods during the training phase on three different data splits (70:30, 80:20, and 90:10). KFSMOTECNN (Proposed): Reports error values of 0.013, 0.015, and 0.02 for the respective data splits during training. KFCNN without SMOTE: Shows error values of 0.053, 0.065, and 0.071 during training for the three data splits. CNN with PSO (Particle Swarm Optimization) demonstrates error values of 0.083, 0.095, and 0.102 during training for the three data splits. CNN with Genetic Algorithm indicates error values of 0.103, 0.115, and 0.122 during training for the three data splits. CNN (Convolutional Neural Network) achieves error values of 0.123, 0.135, and 0.143 during training for the three data splits.

Table 8 Performance Comparison between the KFSMOTECNN (Proposed) & other CNN for the REAL dataset

	KFSMOTECNN (Proposed)	KFCNN without SMOTE	CNN with PSO	CNN with genetic	CNN
Accuracy	1	0.95	0.92	0.9	0.88
Precision	0.98	0.93	0.9	0.88	0.86
Recall	0.9842	0.9292	0.8992	0.8792	0.8592
F1-score	0.9821	0.9296	0.8996	0.8796	0.8596

Sensitivity	0.985	0.935	0.905	0.885	0.865
Specificity	0.9862	0.9312	0.9012	0.8812	0.8612

The Table 8 provides performance metrics for different crop prediction methods, including KFSMOTECNN (Proposed), KFCNN without SMOTE, CNN with PSO (Particle Swarm Optimization), CNN with genetic algorithm, and CNN (Convolutional Neural Network). In summary, the table presents a comprehensive comparison of the evaluated methods across multiple performance metrics, with KFSMOTECNN (Proposed) consistently achieving the highest values in most of these metrics.

The experimental results of the proposed algorithm were subjected to a comparison with several recent algorithms, specifically Multiclass-SVM [30], MLPNN [31], IGBNN [32], OELM [33], and Deep CNN [34] in Fig. 12.

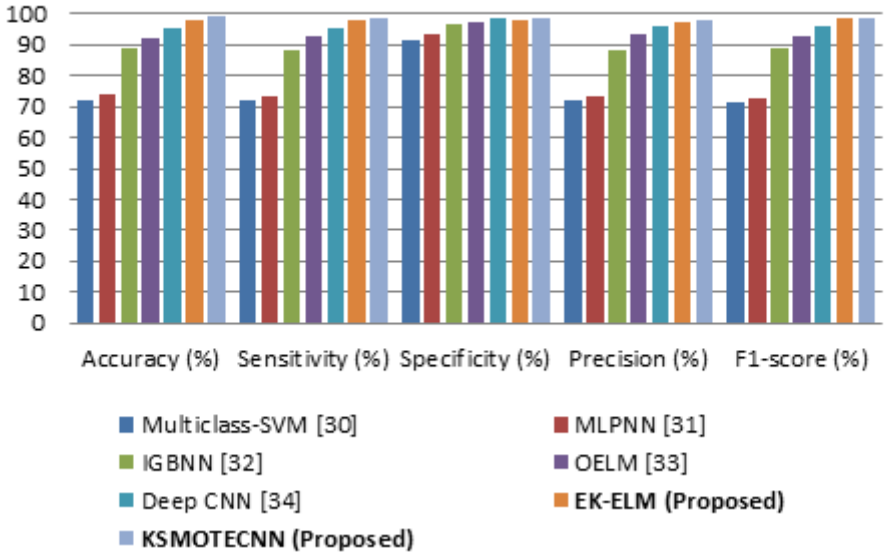


Fig. 12 Graphical comparison of the proposed method with conventional schemes

The evaluation of the proposed approach considers various values of £. It is observed that for both methods, an increase in the value of α corresponds to an improvement in performance. Notably, the highest accuracy is achieved when £ reaches 20, as indicated in Table 9. However, when α exceeds this optimal value, there is a slight reduction in performance. Specifically, with £ set to 10, 15, and 20, the proposed KFSMOTECNN's accuracy for the REAL dataset is estimated at 0.983, 0.987 and 0.991respectively.

Table 9 Proposed approach time complexity for different values £

£	Training time (s)	Testing time (s)	Accuracy (%)			
	KFSMOTECNN (Proposed)	EK-ELM (Proposed)	KFSMOTE CNN (Proposed)	EK-ELM (Proposed)	KFSMOTE CNN (Proposed)	EK-ELM (Proposed)
10	130	93	0.672	0.591	0.983	0.98
15	196	110.4	0.806	0.733	0.987	0.986
20	256.5	153.85	0.947	0.897	0.991	0.99

5. CONCLUSION

In conclusion, our research introduces a comprehensive and innovative approach to address the critical challenges in modern agriculture, namely soil analysis, nutrient detection, and crop prediction. By

seamlessly integrating advanced image processing techniques, such as median filtering, CLAHE, and active contour models, with cutting-edge machine learning algorithms like EK-ELM and KFSMOTECNNs, this work have demonstrated the potential of data-driven decision-making in the agricultural domain. Our methodology not only enhances the accuracy of soil nutrient detection but also empowers precise crop prediction, thereby contributing to sustainable farming practices, resource optimization, and improved crop yields. The evaluation results underscore the effectiveness of our proposed solutions, with impressive performance metrics on REAL dataset. Ultimately, our research paves the way for a new era of precision agriculture, where technology and data-driven insights drive efficient and sustainable farming practices, ensuring food security and environmental conservation. This work believe that the integrated approach presented in this work has the potential to revolutionize the way this work approach soil analysis, nutrient management, and crop cultivation, offering a promising path towards a more resilient and productive agricultural industry.

1.1 Declaration of Competing Interest

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

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Adamchuk, V.I.; Hummel, J.W.; Morgan, M.K.; Upadhyaya, S. On-the-go soil sensors for precision agriculture. Comput. Electron. Agric. 2004, 44, 71–91.

[2] G. Sharmila and K. Rajamohan, ―Image Processing and Artificial Intelligence for Precision Agriculture,‖ Proc. 2022 Int. Conf. Innov. Comput. Intell. Commun. SmartElectr. Syst. ICSES 2022, pp. 1–8, 2022, doi: 10.1109/ICSES55317.2022.9914148.

[3] G. Sharmila and K. Rajamohan, A Systematic Literature Review on Image Preprocessing and Feature Extraction Techniques in Precision Agriculture, vol. 114,Springer Nature Singapore, 2022.

[4] Perez-Ruiz, M.; Slaughter, D.C.; Gliever, C.; Upadhyaya, S.K. Tractor-based Real-time Kinematic-Global Positioning System (RTK-GPS) guidance system for geospatial mapping of row crop transplant. Biosyst. Eng. 2012, 111, 64–71.

[5] Pastor-Guzman, J.; Dash, J.; Atkinson, P.M. Remote sensing of mangrove forest phenology and its environmental drivers. Remote Sens. Environ. 2018, 205, 71–84.

[6] Zhang, X.; Friedl, M.A.; Schaaf, C.B.; Strahler, A.H.; Hodges, J.C.F.; Gao, F.; Reed, B.C.; Huete, A. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 2003, 84, 471–475.

[7] Mohammed Riyadh Abdmezie, D. Tandjaoui, Imed Romdhani, Architecting the Internet of Things:State of the Art, Springer, 2016.

[8] Shaikh, F.K.; Memon, M.A.; Mahoto, N.A.; Zeadally, S.; Nebhen, J. Artificial intelligence best practices in smart agriculture. IEEE Micro 2021, 42, 17–24.

[9] Chen, Q.; Li, L.; Chong, C.; Wang, X. AI-enhanced soil management and smart farming. Soil Use Manag. 2022, 38, 7–13.

[10] Wadoux, A.M.C.; Minasny, B.; McBratney, A.B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Sci. Rev. 2020, 210, 103359.

[11] Lagacherie, P.; Buis, S.; Constantin, J.; Dharumarajan, S.; Ruiz, L.; Sekhar, M. Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of

drainage and maize yield simulated by STICS in the Berambadi catchment (India). Geoderma 2022, 406, 115503.

[12] Khaledian, Y.; Miller, B.A. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model. 2020, 81, 401–418.

[13] Martinelli, G.; Gasser, M.O. Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec. Soil Sci. Soc. Am. J. 2022, 86, 1509–1522.

[14] Ritesh Dash, Dillip Ku Dash, G.C. Biswal, Classification of Crop Based on Macronutrients and Weather Data using Machine Learning Techniques, Elsevier, 2021, p. 8.

[15] A. Priyadharshini, Swapneel Chakraborty, Aayush Kumar, O. Rajendra Pooniwala, Intelligent crop recommendation system using machine learning, 2021, p. 6.

[16] Lee, J.; Nazki, H.; Baek, J.; Hong, Y.; Lee, M. Artificial intelligence approach for tomato detection and mass estimation in precision agriculture. Sustainability 2020, 12, 9138.

[17] Alajrami, M.A.; Abunaser, S.S. Type of tomato classification using deep learning. Int. J. Acad. Pedagog. Res. 2020, 3, 21–25.

[18] Koirala, A.; Walsh, K.B.; Wang, Z.; Anderson, N. Deep learning for mango (Mangifera Indica) panicle stage classification. Agronomy 2020, 10, 143.

[19] Birrell, S.; Hughes, J.; Cai, J.Y.; Iida, F. A field-tested robotic harvesting system for iceberg lettuce. J. Field Robot. 2019, 37, 225–245.

[20] Santos, T.T.; De Souza, L.L.; Dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Comput. Electron. Agric. 2020, 170, 105247.

[21] Bender, A.; Whelan, B.; Sukkarieh, S. A high-resolution, multimodal data set for agricultural robotics: A Ladybird 's-eye view of Brassica. J. Field Robot. 2020, 37, 73–96.

[22] Cheng, H.; Damerow, L.; Sun, Y.; Blanke, M. Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. J. Imaging 2017, 3, 6.

[23] Tran, T.-T.; Choi, J.-W.; Le, T.-T.H.; Kim, J.-W. A comparative study of deep CNN in forecasting and classifying the macronutrient deficiencies on development of tomato plant. Appl. Sci. 2019, 9, 1601.

[24] Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 2018, 145, 311–318.

[25] Barman, Utpal, and Ridip Dev Choudhury. "Soil texture classification using multi class support vector machine." Information processing in agriculture 7, no. 2 (2020): 318-332.

[26] G. Sharmila and K. Rajamohan, —Digital Soil Texture Classification Using Machine Learning Approaches,Lecture notes in networks and systems, 2024, pp. 133–144.doi: 10.1007/978-981-99-5015-7_12.

[27] Soil classification using active contour model for efficient texture feature extraction,Int. J. Inf. Technol., vol. 15, no. 7, pp. 3791–3805, 2023, doi: 10.1007/s41870-023-01404-6.

[28] Sarkar U, Banerjee G, Ghosh I. (2022) A Machine Learning Model for Estimation of Village Level Soil Nutrient Index. Indian Journal of Science and Technology. 15(36):1815-1822.

[29] Liu, Y., Jiang, C., Lu, C., Wang, Z., & Che, W. (2022). Increasing the Accuracy of Soil Nutrient Prediction by Improving Genetic Algorithm Backpropagation Neural Networks. Symmetry, 15(1), 151.

[30] M.S. Suchithra, Maya L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters", Information Processing in Agriculture, Volume 7, Issue 1, 2020, Pages 72-82.

[31] Sharmila.G, et al. (2023). Optimal Deep Convolutional Neural Network based Fusion Model for Soil Nutrient Analysis. International Journal on Recent and Innovation Trends in Computing and Communication, 11(10), 44–51. https://doi.org/10.17762/ijritcc.v11i10.8463.

[32] https://www.kaggle.com/datasets/gsharmila/soil-types.