Sensor based framework for soil nutrients prediction using deep learning techniques

Shwetha G.N.[1*] , Bhat Geetalaxmi Jairam[2]

(1.Research Scholar, Department of Computer Science & Engineering, National Institute of Engineering, Mysuru – 570008,India;Visvesvaraya Technological University, Belagavi-590018, India;
2. Research Supervisor, Department of Computer Science & Engineering, National Institute of Engineering,Mysuru-570008; Visvesvaraya Technological University, Belagavi-590018, India)

Abstract: A thriving economy of the country depends on agriculture. Soil is necessary for producing food in a sustainable manner as farming gets more intensive and demand rises, the quality of the soil may decrease. Smart soil prediction provides precise information on soil nutrient distribution needed for precision agriculture. Deep learning and machine learning techniques are currently the main drivers of intelligent soil prediction systems. This paper presents a hybrid deep learning based framework for identifying the soil nutrients in the sensor collected data and to predict the quality of the soil. The soil data collected from the sensors will be fed into the binary classifiers to classify the various soil quality then it will be fed into multi-class classifier to classify the soil nutrients. This is implemented to show the better suitability for any agriculture field to check its soil quality and lack of nutrients in it. The proposed one dimensional Convolutional Neural Network-LSTM framework is compared with benchmark techniques like Support Vector Machine, K-Nearest Neighbors, Random Forest, Long Short Term Memory, One-Dimension Convolutional Neural Network, and Artificial Neural Network schemes. Our proposed model shows 98% accuracy in the prediction of soil quality compared to the traditional ones.
Keyword: Classification, Convolutional Neural Network, Long Short Term Memory, Deep Learning, Machine Learning, Smart Agriculture.

Citation: Shwetha, G. N., and B. G. Jairam. 2025. Sensor based framework for soil nutrients prediction using deep learning techniques. Agricultural Engineering International: CIGR Journal, 27(4):291-305.

1 Introduction

Smart agriculture is adopting advanced technologies like data-driven farming practices to maximize and enhance sustainability in agricultural output. Artificial intelligence (AI), automation, machine learning (ML), and the internet of things (IoT) are among the technologies utilized in smart farming. Agricultural processes can be made links in a digital network in the way that these technologies facilitate more effective and sustainable farming methods, commonly known as precision agriculture, smart agriculture, and digital farming. Precision Agriculture Trains on optimizing agricultural practices using data-driven insights to manage variability in soil, crops and environmental conditions. Smart agriculture connects various IoT-AI-big data-enabled devices across the farming ecosystem. Digital Farming makes digitization of agricultural processes, including data collection, analysis, and decision-making.

The U.S. Department of Agriculture reports that in 2019, the agricultural sector and related businesses

Received date: 2024-06-11 **Accepted date:** 2025-04-15
***Corresponding author: Shwetha G. N.** Research Scholar, Department of Computer Science & Engineering, The National Institute of Engineering,, Mysuru, India, Visvesvaraya Technological University, Belagavi. Email: shwethagn@jssstuniv.in, bgj@nie.ac.in.

supported 10.9% of all jobs and added USD 1.109 trillion to the country's GDP (Chatterjee et al., 2020; Pekel, 2020). Soil quality is very important for the production of food and other necessities for the world's population. One of the most important aspects of feeding the world's expanding population is to maximize natural resources (Prakash and Sahu, 2020). Soil quality is an important component of cultivation as it increases the crop yield and quality transferred from the soil nutrients to the crop. The amount of nutrients in the soil represented as a percentage is known as soil macro and micronutrients. Over time, the soil's quality and soil content gets degrades. Soil's quality depends on a number of variables, namely the quantity of rain in a given location, irrigation, the amount of pesticides and its quality. It also has a stronger connection with insecticide usage for crop production. So, farmers can use certain nutrients as complements to increase the soil quality and have good crop health and yield (Han et al., 2018).

The use of digital technology, including robotics, electronics, sensors, and automation technologies, is associated with precision agriculture. This technology aims to reduce workloads, increase profitability, and enhance decision-making. Precision farming, commonly referred to as precision agriculture, is an agricultural control system that provides a thorough approach for working with the temporal and geographical variation of crops and soil with the goal of maximizing revenue, optimizing yield, and improving the quality of product (Kim et al., 2009). In the mountainous areas, farmers are using precision agriculture at a lesser rate than in the valley (Viscarra Rossel and Bouma, 2016). In agriculture, machine learning can be applied to predict soil properties like moisture content and organic carbon content, and also to identify weeds and diseases in crops, detect species, and estimate crop production (Shi et al., 2015).

Monitoring the physical aspects, such as weather patterns and geological features, is referred to as land management. The significance of global changes in climate could impact agricultural production. The climate of the earth is determined by rainfall, which also has an impact on agriculture, biological systems, and water management (Wu et al., 2020). For instance, fertilization is very important, especially for growing cabbage successfully (Wang et al., 2007). Like all other crops, cabbage plants have particular nutrient needs. For their growth and development, nitrogen (N), phosphorus (P), and potassium (K) are the three primary macronutrients required (Wang et al., 2007; Mmbaga et al., 2014). In order to effectively meet these nutritional needs for growing cabbage, chemical fertilizers such urea, triple superphosphate, and muriate of potash are commonly utilized (Shahena et al., 2021). Organic cabbage producers use natural resources including grass, cow dung, wood ash, rice bran, and chicken manure to meet the crop's nutritional needs (Ibukunoluwa Moyin-Jesu, 2015). It is critical to give farmers information on the state of the soil, especially with regard to important minerals, and to improve soil nutrition by implementing cutting-edge prediction models.

Artificial intelligence methods (Tantalaki et al., 2019) are useful for a variety of prediction tasks and are essentials to design global precision agriculture framework. To anticipate diverse agricultural components these models and tools make use of a variety of methodologies. The methods include expert systems, pattern recognition, non-linear simulations, data analysis, automation, decision making, and artificial neural networks (Sarker, 2022). The nonlinear parallel structure of the human brain system serves as the model for neural networks (Sajindra et al., 2023), a large-scale, parallel distributed information processing system. Its original source was the biological central nervous system (Indiveri and Liu, 2015; Prieto et al., 2016). Simulating the information processing mechanism of the human brain is the aim of deep neural networks, a class of machine learning techniques connected to artificial neural networks (Aouichaoui et al., 2021). A kind of artificial neural network known as deep learning networks, or DNNs for short, are unique in that they consist of multiple layers of artificial neurons, or interconnected nodes (Chartrand et al., 2017).

In recent times, neural networks have gained significant recognition because of their remarkable ability to tackle complex problems, especially in the domains of machine vision, natural language processing, and reinforcement learning (Liu et al., 2017). The two primary groups of soil nutrients are macronutrients and micronutrients. Plants require significant amounts of macronutrients such as phosphorus (P), nitrogen (N), potassium (K), calcium (Ca), sulfur (S), and magnesium (Mg). They also need micronutrients including zinc (Zn), molybdenum (Mo), manganese (Mn), copper (Cu), iron (Fe), chlorine (Cl), and boron (B), although in much smaller levels. The extra vital soil nutrients that are required for crop growth are soil pH, soil organic carbon, and soil organic matter (Kumar and Kaur, 2024).

Figure 1 illustrates a basic IoT sensor system consisting of a sensor interfaced with a microcontroller that intern interfaced with a radio module. The microprocessor sends commands to the microcontroller, which is run by a battery supply. The radio module communicates with a cloud, smart device, or personal computer by sending and receiving data over a network. It has an IP address, which is a distinct numerical identification. The desktop computer and the cloud web application have deep learning frameworks for yield prediction. When establishing a wireless sensor microheater network for smart agriculture, researchers also used a microcontroller and wireless radio module arrangement. Furthermore, the employment of additional actuator systems for control in conjunction with real-time soil nutrient monitoring techniques has improved the efficiency of precision agriculture's farming and crop cultivation.

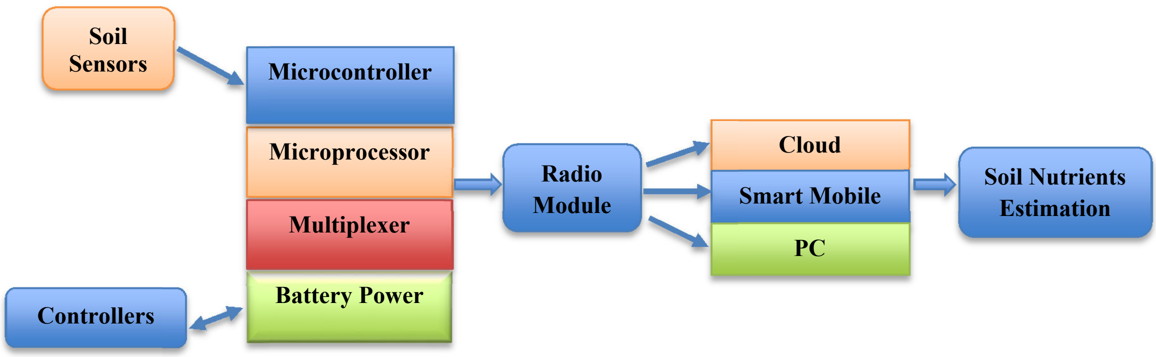

Figure 1 Schematic of a typical IoT sensor system for real-time soil monitoring

Although many studies have been conducted on soil nutrient prediction by using machine learning and data mining methods. There are some significant research gaps addressed in this research are Limited Temporal Modeling: Several state-of-the-art frameworks do not adequately learn long- or short-term dependencies within soil nutrient data, which is important for accurate long-term predictions. Feature Extraction Challenges: Traditional models generally have difficulty in extracting relevant features from raw sensor data which leads to lower performance. Insufficiency of Traditional ML Approach: Traditional machine learning approaches that have been used for soil nutrient prediction are often limited in complexity and may not adequately capture non-linear relationships in the data.

Research contributions of our proposed work include the following:

Preprocessing, cleaning, selecting features, and normalizing the data that is suitable to train and test the designed model.

Designing the hybrid techniques with fine-tuned parameters tuning in machine learning algorithms for predictive modeling.

Prediction of soil nutrients and quality of soil estimation by designing a deep learning framework.

Evaluating the performance to show that the proposed sensor-based deep learning framework had good stability and reliability.

2 Literature Survey

The availability of soil nutrients and suggested fertilizer recommendation methods have been the basis of numerous data mining and machine learning-based studies conducted in recent years, which have shown how important soil is in affecting crop output. Currently, there exist many frameworks predict soil nutrients, To say a few, J48/C4.5, Artificial Neural Network, K- nearest neighbor, ID3, Naive Bayes, and Support Vector Machine were defined by Nikam (2015) for classifications. It is more tailored to certain patterns of crop cultivation because to the range of machine learning-based schemes available. Numerous foundational works related to our study issue have been thoroughly examined in this regard.

The three categories of these classification techniques are Neural Networks, machine learning, and statistics. To define their work, Gholap et al. (2012) used three algorithms. They have used 1988 soil cases combined with the J48 and JRip algorithms. The classifier is really basic and produces a decision tree with an accuracy rate of 91.90 percent. The author also mentioned that creating a system of recommendations for fertilizer based on cropping patterns and soil test samples would be a future objective. In Virudhunagar District, Tamilnadu, India, Arunesh and Rajeswari (2017) studied and experimented using 203 soil examples with 6 variables. They have discovered that the Naive Bayes classifier beats the performance of random tree, J48, OneR and JRip.

Ramesh and Ramar (2011) have designed a framework by using variants of Bayes algorithms like, Bayes Net, Naive Bayes, Naive Bayes Updatable, Random forest, and J48. They used a dataset that included 1,500 samples of soil that were collected from the Department of Agriculture. J48 calculated 92.3 percent accuracy in the classification of soil nutrients, whereas the Naive Bayes algorithm produced 100 percent accuracy. Chiranjeevi and Nadagoudar (2018) developed a method for analyzing soil condition and nutrients, including pH, potassium, nitrogen, phosphorus, EC, OC, iron, sulfur, zinc, boron, magnesium, and copper, at Belagavi Department of Agriculture. The experimentation is done using two classification algorithms Naive Bayes and J48. Compared to the J48 algorithm the Naive Bayes algorithm yielded a superior result in accurately identifying the amount of soil nutrients present in the soil.

Bhargavi and Jyothi (2009) have done the research on the soil of Tirupati, Andhra Pradesh, by implementing the Naive Bayes Classifier. They have considered many sand classes for classification including clay, loamy sand, loam, sandy loam, sand, clay loam, and sandy clay loam for each instance of soil data. Puno et al. (2017) used MATLAB software to propose and create a completely working system utilizing IP (IP segmentation, feature extraction, IP improvement). They have used different notations for 5 classes of soil nutrient availability. The values considered are Deficient, Sufficient, High, Medium, and Low corresponds to each of the seven nutrient characteristics.

Ahmad et al. (2010) have proposed a framework for estimating soil moisture by performing trials with models 1 and 2. This model was designed based on five-year data that took into account just one feature for the classification of VIC moisture. According to them, the support vector machine model performs better than the artificial neural network and multilayer regression models. Juhi Reshma and Aravindhar (2021) proposed regression methods for future plantations and a suggestion for estimating the quantity of fertilizers required for plantations. They have worked on a particular banana crop using the Neural Networks model. This crop requires the three most crucial soil nutrients, namely phosphorus, nitrogen, and potassium.

Golicz et al. (2019) recently demonstrated that the reflectometers included with the test strips could be replaced by the smartphone app Akvo Caddisfly. The researchers replaced the standard Quantofix Relax Reflectometer with the app and got good results, especially when it came to measuring the amount of nitrate accessible in plants. A field-deployable colorimetric analyzer based on an Android mobile application was proposed by Moonrungsee et al. (2015). This model was built to identify and quantify the phosphorus content in the soil with less expense and fast. The classic method of combining potassium orthophosphate, antimonyl tartrate, and ammonium molybdate to create phosphomolybdic acid which is subsequently reduced by ascorbic acid to yield the brilliantly colored molybdenum blue was employed to do this.

Riese and Keller (2019) have designed the framework with two stages. The first stage of the process involves gathering different types of soil test pictures. They have preprocessed it to consider factors like user needs, the size, characteristics of the soils being studied, the availability of soil data, time constraints and the financial. Machine learning techniques are trained and tested using these features. According to Devi et al. (2016) also worked on soil nutrients with the motive of agribusiness is the backbone of the Indian economy and the primary source of income for a sizable section of the country's population. According to them, a variety of factors, including soil, climate, rainfall, composts, and pesticides, affect harvest yield.

Rahman et al. (2018) have studied various soil types with macro and micronutrients for agriculture. They studied that crop growth depends on the soil type because soil and crop compatibility is important in agriculture. Each soil may have different soil properties and nutrients based on which crop growth will depend. To understand the varieties of crops that grow better in particular soil types, farmers need to know the traits, nutrients, and qualities of various soil types. They have used machine learning techniques that are beneficial for prediction.

Some of the important research gaps addressed in this research are Limited Temporal Modeling: Several state-of-the-art frameworks do not focus on learning long- or short-term dependencies of soil nutrient data, which are essential for accurate long-term predictions. Challenges of Feature Extraction: Traditional models usually struggle to extract meaningful features from raw sensor data resulting in lower performance. Limitations of Traditional ML Approach: The traditional machine learning approaches that have been applied for soil nutrients prediction are at best limited in their complexities and may not sufficiently approximate non-linear relationship that might exist in the data.

The overall structure of the paper is as follows: The first section provides an overview of agriculture and soil nutrients, outlining their significance, and the goals of the proposed research. The second section discusses the literature on the identification of soil nutrients and the subsequent steps that were done. A detailed discussion of the proposed model and its mathematical model together with its design is given in the third section. Complete implementation and outcomes with different parameters are given in the fourth section. The paper's final part followed by references provides the summary of the whole process of identifying and categorizing the soil nutrients.

3 Proposed Methodology

This section discusses the detailed proposed methodology with the mathematical model using deep learning mechanisms. The proposed framework is named as one dimensional Convolutional Neural Network with Long Short Term Memory for soil nutrients prediction and type estimation. This approach leverages the strengths of the CNN model in extracting features and LSTMs in handling temporal dependencies. The proposed methodology is divided into multiple stages like data collection, preprocessing, feature selection, classification and estimation. Figure 2 shows the overall proposed framework design.

3.1 Preparation of the data

Data preparation is a very crucial stage of any research; it helps to work on the actual need of the real-time problems. The proposed model considered the soil data of a particular region to train the model to predict the soil nutrients. The Karnataka state nutrient information, which is accessed via online GIS servers, is obtained from the India Soil Health System. This is managed by the Indian Council of Agricultural Research and it is used to train the proposed framework. The district-level data from Karnataka state provided the soil nutrient elements used in the training. The data is preprocessed before being incorporated into the model. The entire dataset is divided into training data of about 80% and testing of about 20%. The dataset is having the macro nutrients for good crop production like Calcium, Nitrogen, Phosphorous, Sulphur, Boron, Zinc, Manganese, Potassium, and Iron as C, N, P, S, B, Zn, Mn, K, Fe.

3.2 Data preprocessing

This step is also essential, because it ensures that the proposed 1D-CNN-LSTM framework receives soil nutrient input data of high quality and validity. Data Normalization, Removing Outliers and Missing Value Treatment: The following techniques were applied for these tasks: Mean imputation was used to deal with missing values in the dataset. Missing values in numerical features like soil nutrients (Nitrogen, Phosphorus, Potassium) and pH were filled with their respective mean values. This method maintains the overall distribution of the data while maintaining continuity. We normalize the dataset using Z-score normalization and min-max normalization to ensure that all features contribute equally to the model. Feature selection methods were used in order to eliminate redundancy and enhance computational efficiency. Discrete values with few observations were also cleaned and preprocessed in order to remove non-informative features with low variance or high correlation with other features. Thereafter, data cannot be used for training which prevents overfitting and improves model generalization capability.

Through optimal prediction of soil nutrients detection need for the crop cultivation based on the nutrients available. The proposed Enhanced convolutional neural network based model for soil nutrient composition prediction contributes to achieving food crop yield. In order to forecast the ideal soil nutrient composition using enhanced convolutional neural network utilized the significant parameters of various macro nutrients along with soil pH. One dimensional convolutional neural networks are used as a solution that have rapidly attained the required ideal performance levels in a range of applications. One notable advantage of one dimensional Convolutional Neural Networks is their compatibility and easy configuration, as they are only capable of performing one dimensional convolutions. This makes the transition to on-demand, cost-effective hardware implementation possible. Given the usefulness of one dimensional convolutional neural networks for temporal data analysis, we have decided to build the soil nutrient prediction process on the optimized one dimensional Convolutional Neural Network, which flattens the layer level enhancement of traditional two dimensional convolutional neural networks.

One-dimensional convolutional neural network has a special network structure that alternates between the convolution and the pooling layers. It contains a fully linked hidden layer in contrast to typical neural network. One input layer, three fully connected layers (C1, C2, C3), five convolutional layers (L1, L2, L3, L4, L5), and an output layer O. Although it has been demonstrated that these enhancements are more effective in particular applications that deal with one-dimensional temporal data. This one-dimensional Convolutional Neural Network is intended to be an improved version of a two-dimensional convolutional neural network. The input layer of the one-dimensional convolutional neural network remains an inactive layer and accepts one-dimensional temporal data. The output layer consists of a fully linked layer with the same number of neurons as classes. As seen in the image, there are five one-dimensional convolutional neural network layers in a row with a one-dimensional filter kernel size of five and a subsampling feature of four.

In this case, the hidden layer kth neuron of the convolutional neural network performs a series of convolutions first, after which the sum is sent via a subsampling operation to cover the activation function f. The main difference between a two-dimensional Convolutional Neural Network and a one-dimensional convolutional neural network is basically this. The one dimensional array replaces the two dimensional matrix in both the kernel map and the feature map. Feature extraction and classification are therefore merged into a single process that may be improved to increase the effectiveness of the classification.

To be more precise, the Convolutional Neural

Network layers process the raw one-dimensional datasets further and start learning to extract the properties that can help the fully connected layers do classification. The one-dimensional Convolutional Neural Network process is advantageous because it involves minimal computational complexity. This is because the order of one-dimensional convolutional neurons, that essentially performs the linear weighted summing of two one dimensional arrays. It can be used efficiently for both forward and backward transmission operations in parallel.

Equation 1 depicts the one dimensional Forward propagation process in each Convolutional Neural Network layer, while Equation 2 and 3 summarizes the back propagation process and defines it as follows:

$$V_i^L = b_i^L + \sum_{x=1}^{M} C1D(W_i^{L-1}, S_i^{L-1}) \quad (1)$$

where,

V_i^L — Output feature (or activation value) of the *i-th neuron/filter* in layer L before applying the activation function;

b_i^L —Bias term for the *i-th filter/neuron* in layer L;

M —Number of input channels (or the number of feature maps) coming from the previous layer (layer L^{-1});

$C1D(.)$— Represents a 1-Dimensional Convolution operation**;**

W_i^{L-1}—Convolution kernel/weight corresponding to the *x-th input channel* from the previous layer;

S_i^{L-1} —Input feature map (signal) from the *x-th channel* of the previous layer.

These values can be used to compute the weights and biases and the learning factor, ε, once the weight and bias have been computed.

$$W_i^{L-1}(k+1) = W_i^{L-1}(k) - \varepsilon \frac{\delta E}{\delta W_i^{L-1}} \quad (2)$$

$$b_i^L(k+1) = b_i^L(k) - \varepsilon \frac{\delta E}{\delta W_i^L} \quad (3)$$

Finally, flattened feature vector obtained from the 1D CNN i.e. X = {F1, F2,……..,Fi}. Here, F1, F2,…,Fi are the max pooled values for the filters.

The LSTM has adequate control over which data should be retained and which should be deleted at every training iteration. At the i-th time step, the LSTM receives the output vector of the previous hidden state h_{i-1} and the vector representation of the i-th word in the sentence v_i. The LSTM procedure begins with deciding what data should be removed from the cell state S_{i-1}. The forget gate g_i depicted in Equation 4 is where this step is determined.

$$g_i = \sigma w_{gi} v_i + \sigma w_{gh} h_{i-1} + \sigma b_g \quad (4)$$

Choosing the data that will be kept in the cell state S_i is the next step. This step consists of two steps: first, as indicated in Equation 5, compute the value that needs to calculate through the input gate p_i, and then, as indicated in Equation 6, generate a new vector representing the candidate cell state value S_i.

$$p_i = \sigma w_{pi} v_i + \sigma w_{ph} h_{i-1} + \sigma b_p \quad (5)$$

$$\widetilde{S_i} = tanh\, tanh\, w_{si} v_i + tanh\, tanh\, w_{sh} h_{i-1} + tanh\, tanh\, b_s \quad (6)$$

Equation 7 illustrates how the forget gate g_i , input gate, and candidate vector $\widetilde{S_i}$ are used to update the cell state $S_{i-1}$to a new cell state S_i.

$$S_i = g_i\, S_{i-1} + v_i\, \widetilde{S_i} \quad (7)$$

Using the modified cell state, the output is determined in the last step. This process takes place at the output gate o_i, as indicated by Equation 8. The hidden state $h_i$is then found by applying the vector o_i to the cell state S_i, as shown in Equation 9. The following time step will make use of the vector h_i.

$$o_i = \sigma w_{oi} v_i + \sigma w_{oh} h_{i-1} + \sigma b_o \quad (8)$$

$$h_i = o_i * tanh\, tanh\, S_t \quad (9)$$

Here $W_{\{g_i, g_h, p_i, p_h, S_i, S_h, o_i, o_h\}}$ shows the weight matrix of the output and $b_{\{g,p,S,o\}}$ is a bias vector

The integrated model first extracts feature from the input sequence using a 1D CNN, and then feeds those features into an LSTM network to identify temporal relationships. The data input is represented s given in Equation 10. Here S represents number of samples and L represents the length of the samples $Y\epsilon\, R^{SL}$.

Applying the 1D CNN for extracting the features.

$$X = \max(\sum_{x=0}^{M-1} SW_{i,k}\, Y_{i+k} + b_k((Y))) \quad (10)$$

Here, $X\epsilon\, R^{SL'f}$ S is the number of samples, L represents the length of the samples, and f is the number of features. Then the LSTM is applied to the output of 1DCNN i.e. X as given in Equation 11.

$$Z = o_i * tanh\, S_t(X) \quad (11)$$

In this case, the Dense layer maps the LSTM output to the ultimate soil nutrient level estimate $Z \epsilon R^S$.

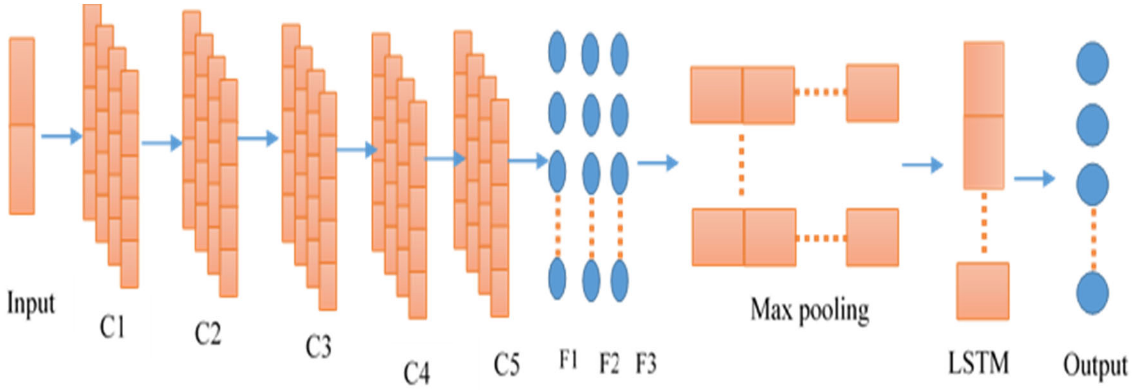

Figure 2 Proposed 1DCNN-LSTM design diagram

Figure 2 shows the overall design diagram of the proposed model. In which, one-dimensional Convolutional Neural Networks will receive the preprocessed data as input. Convolutional Neural Network first processes a data representation in order to identify the important features from the data. After that, the Convolutional Neural Network output is fed into the long short term memory model, which uses the data's order to construct a new representation. Moreover, the long short term memory model output is mapped into two categorization classes using a fully linked layer. Finally, the soil nutrients are classified based on the features mapping with hyper parameters tuning in the proposed model. The trained deep learning model provides accurate predictions by leveraging the IoT sensor data, and integrates the IoT sensor system with the machine learning framework, thus resulting in an intelligent prediction model in terms of accuracy and reliability which focuses on precision agriculture.

4 Results Discussion

This section discusses the implementation of the proposed hybrid model for the soil nutrients detection from the particular region data. The proposed model implementation shows the forward and back-propagation processes in concealed one dimensional Convolutional Neural Network layers. Utilizing benchmarked Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), 1D-Convolutional Neural Network (1DCNN), Long Short Term Memory (LSTM), and Artificial Neural Network (ANN) schemes along with the Enhanced one dimensional Convolutional Neural Network-LSTM for assessing for a specific geographic area of utilizing nutrient information from the dataset. The parameters like Recall, Precision, Accuracy, and F1-measure along with the error measures R^2, MAD, RMSE, and MAE are computed. The graphical comparison is made between the proposed enhanced Convolutional Neural Network-LSTM and the baseline models.

The proposed framework is tested using the Google Colab platform and the datasets are stored on Google Drive. The Karnataka State nutrient datasets are sourced from online GIS servers and linked to the Indian Council of Agricultural Research's soil health system are used in the training and testing process of the proposed framework. This website is essential for giving district-by-district soil-related parameters for every state in the nation. Initially, the pre-processing of the data is done before it is added to the model to remove the unwanted data, outliers, and noise. Next, the dataset is divided into 20% testing and 80% training segments. Subsequently, the pre-processed data is introduced into proposed model evaluation procedure whereby the model performance measure and error metrics are assessed. Finding the hyperparameters to regulate the learning process and determine the values of model parameters is accomplished by fine-tuning the hyperparameters on the test set. Such a cutting-edge deep learning technique is intended to predict the nutrient content of the soil for the crop's sustainable growth.

4.1 Preprocessing of the data

Processing data is a second, very important stage of any prediction or classification that gathers and cleans the unprocessed data. Some data preprocessing is required at this stage to avoid data inaccuracy, inconsistency, and incompleteness. Data cleansing and windowing are two ways of preprocessing techniques used in the suggested strategy. Partial data handling and data conversion are important for further processing to deliver well-formatted data. Pre-processing is started with handling missing data M and the data normalization Nz will be performed. The machine learning classifier accuracy in predicting outcomes is affected by the presence of missing or null data. As shown in Equation 12, the missing values are replaced by using the mean value rather than deleting or removing it.

$$M(x) = \left\{ \left(\frac{\sum_1^i M_i}{n}\right), if\ i = \frac{null}{missing} \right.$$

$$M_i,\quad otherwise \qquad (12)$$

The missing value imputation via the mean method ensuring the generation of continuous data necessary for algorithm training while mitigating the introduction of outliers. To prevent overfitting, inefficient performance, and prolonged computations resulting from redundant features. Then, the feature selection algorithms are employed to reduce the feature set and enhance computational efficiency. Various feature selection techniques are utilized to identify optimal feature vectors (Swaminathan et al., 2023).

Following feature extraction, the calculated features are standardized to mean and unit variance, thereby compressing the features within specific ranges. This standardization process, particularly beneficial for features with dynamic range, contributes to improved classification accuracy in the soil nutrients detection system. In particular, the Z-score normalization approach is used in the proposed method for normalizing the dataset. Equation 13 is used to divide the mean value of feature vectors by the standard deviation following subtraction from each individual signal data point.

$$Z = \frac{i-\underline{i}}{\sigma} \qquad (13)$$

When features are rescaled, standardization or Z-score normalization is used to achieve a standard normal distribution with a zero mean and unit variance. As seen in Equation 14, standardization (S) also lessens the skewness of the data distribution:

$$S(i) = \frac{i-\underline{i}}{\propto} \qquad (14)$$

Here y represents the instances in the n-dimensional of the feature vector, shown in Equation 15.

$$i \in V^x \qquad (15)$$

$$\underline{i} \in V^x \qquad (16)$$

$$\propto \in V^x \qquad (17)$$

Above Equations 16 and 17, calculate the standard deviation and attribute mean value. Then the data can be split into fixed-length sequences. This splitting helps the data be properly divided for testing, training, and validation of the proposed hybrid model.

The data should be split into three categories to ensure representative samples among various soil nutrients. To find the best number of convolutional layers, we tried multiple sets of layers from 3 to 7. We evaluated performance on accuracy, accuracy, precision, recall, and F1-score metrics. The five layer model yields good accuracy, precision, recall, and F1-score as shown below.

4.2 Performance measures

Accuracy: The overall classification accuracy of the model is calculated to show the efficacy of the model. It shows the correct prediction of the labels against the actual predictions of the labels. Figure 3 presents a comparative analysis of the proposed model performance with different existing models.

Precision: It is the ratio of correctly detected positive predictions to the entire number of positive predictions. It represents the overall precision of the proposed framework with true positive and false positive values. Figure 4 provides a comparative representation of the suggested model's precision.

Recall: The ratio of actually predicted true positive values to the total sum of true predicted positive and false negative values. It is the recall value for the proposed framework compared to the existing methods is shown in Figure 5.

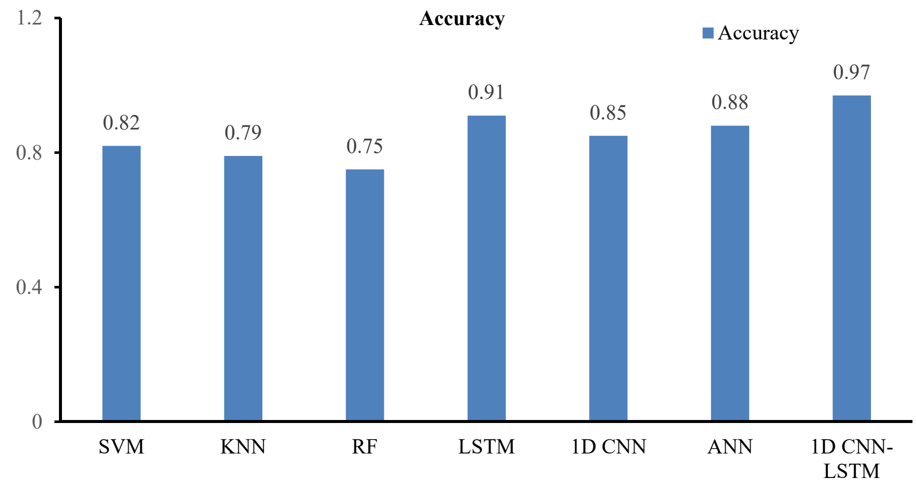

Figure 3 Accuracy comparison of the existing methods with the proposed framework

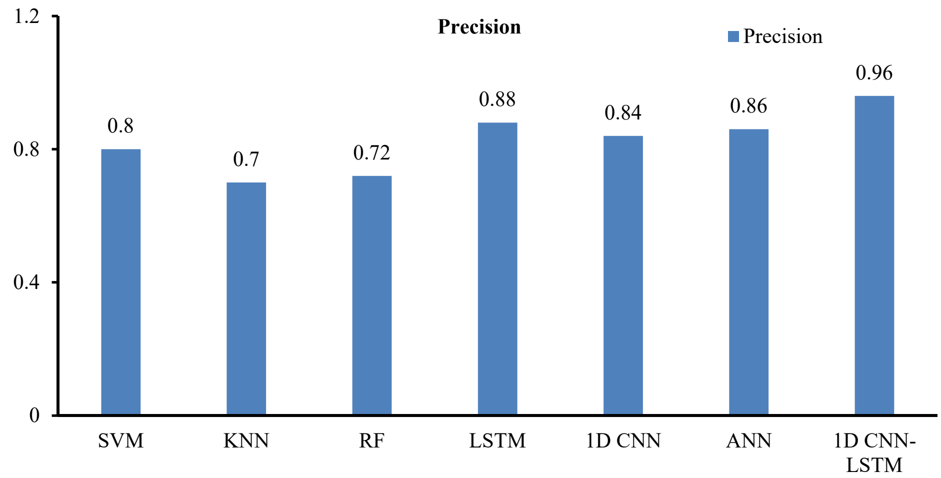

Figure 4 Precision comparison of the existing methods with the proposed framework.

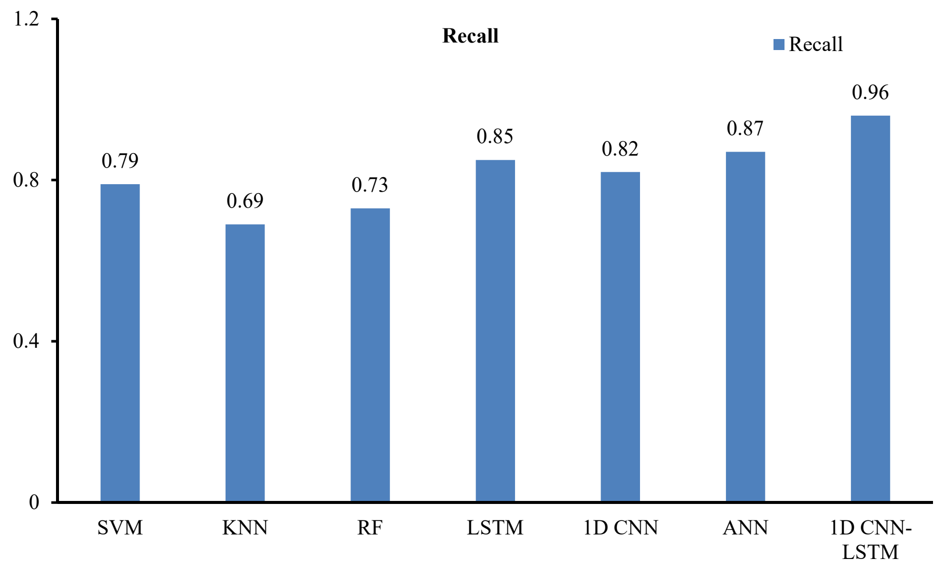

Figure 5 Recall comparison of the existing methods with the proposed framework

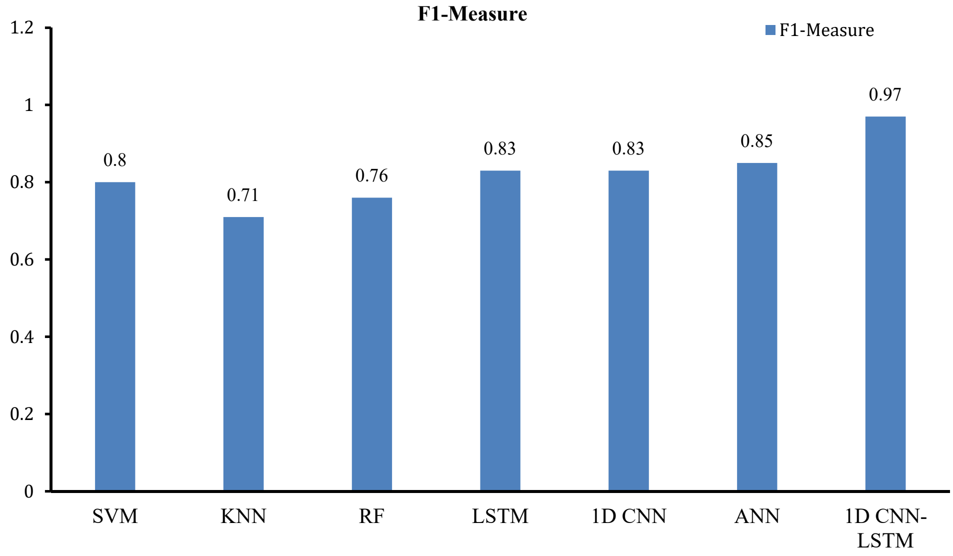

Figure 6 F1-Measure comparison of the existing methods with the proposed framework

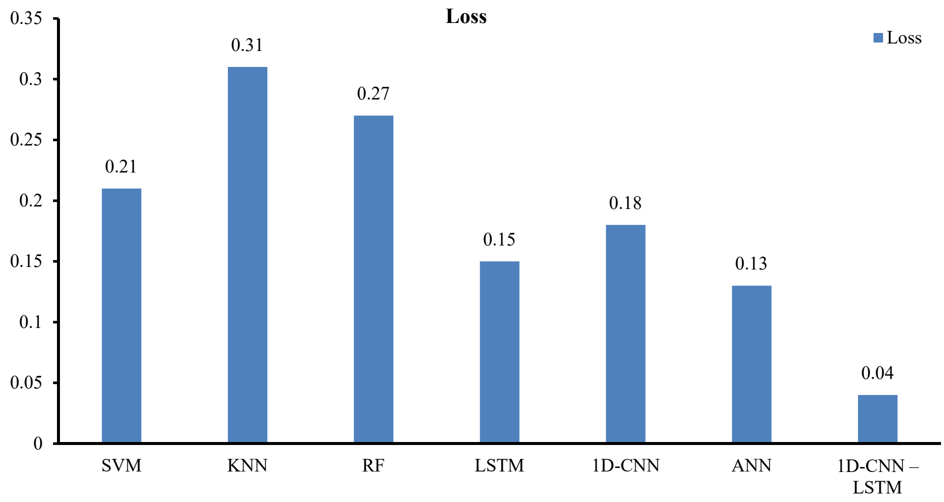

Figure 7 Loss comparison of the existing methods with the proposed framework.

Table 1 The overall performance metrics and related representation of the proposed framework

Model	Accuracy	Precision	Recall	F1-Measure	Loss
SVM	0.82	0.80	0.79	0.80	0.21
KNN	0.79	0.70	0.69	0.71	0.31
RF	0.75	0.72	0.73	0.76	0.27
LSTM	0.91	0.88	0.85	0.83	0.15
1D-Convolutional Neural Network	0.85	0.84	0.82	0.83	0.18
ANN	0.88	0.86	0.87	0.85	0.13
1D-Convolutional Neural Network –LSTM	0.97	0.96	0.96	0.97	0.04

F1-Measure: The proposed framework performance against the baseline model is shown in Figure 6. The F1 measure is the harmonic mean of precision and recall. It provides the model fitness to the real world implementation.

Loss: The proposed framework that has been suggested produces a very low loss value, which enhances the accuracy rate. Figure 7 illustrates the loss comparison between the proposed framework and the existing mechanisms. The loss is the miss predicted labels for the data to the total number of labels predicted.

Table 1 shows the overall performance metrics and related representation of SVM, KNN, RF, LSTM, 1D-Convolutional Neural Network, ANN and 1D-Convolutional Neural Network- LSTM.

4.3 Error metric Information

RMSE: The Root of the Mean of the Squared Errors is the square root of the difference, measured over the total number of observations, between the actual values associated with the research variables of soil nutrients and the anticipated values of the utilized model.

MAE: The mean of absolute values represents the discrepancy between the actual values related to the research variable of soil nutrient as determined over the total number of observations and the anticipated values of the used model.

MAD: The mean absolute deviation of a dataset is the average distance between each data point and the mean. It helps us determine the degree of variability in a dataset. Table 2 presents the calculated error metrics for the proposed framework in comparison to all baseline models. The corresponding graphical representation, as presented in

Figure 8, illustrates the proposed framework optimal performance in terms of achieving low error metrics, and minimal loss during the prediction process.

Table 2 The calculated error metrics of the proposed framework in comparison to all baseline models

Model	RMSE	MAE	MAD
SVM	0.70	0.73	0.69
KNN	0.79	0.60	0.59
RF	0.75	0.62	0.63
LSTM	0.41	0.68	0.65
1D-Convolutional Neural Network	0.65	0.54	0.62
ANN	0.58	0.66	0.67
1D-Convolutional Neural Network -LSTM	0.38	0.21	0.26

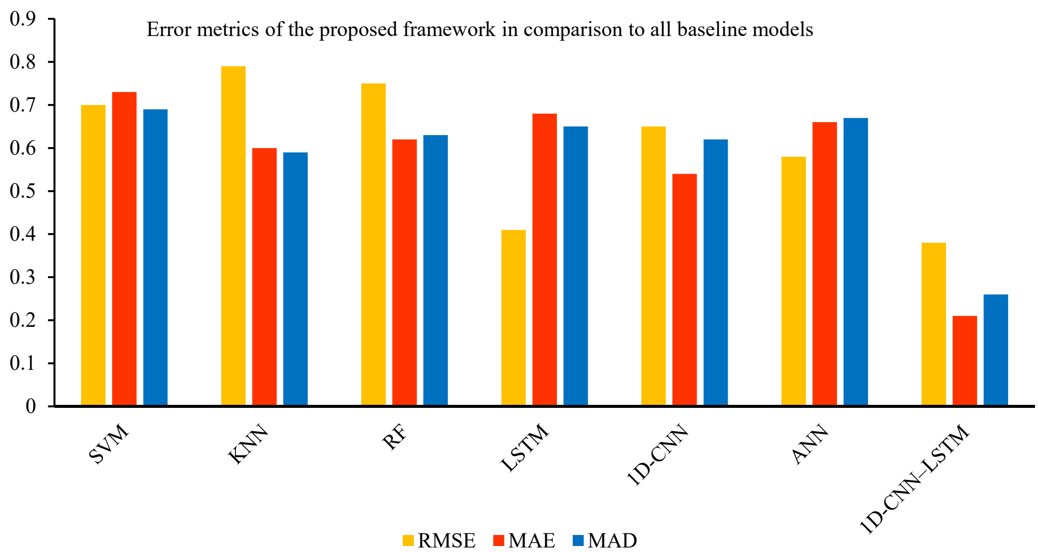

Figure 8 The proposed framework optimal performance in terms of achieving low error metrics

The Karnataka state soil nutrient dataset is used to evaluate the prediction efficacy of the suggested framework. Nitrogen (N), Phosphorous (P), and Potassium (K), the three main macronutrients of soil, have been estimated with the highest prediction accuracy rates using the Low, Medium, and High categorization sets for the proposed model. The results of the forecast show these outcomes quite clearly. Various categorization labels are built using the nutrient dataset values, which precisely represent the soil nutrient composition of various nutrients in the state of Karnataka. The similar benchmarked techniques and the proposed framework go through a three-step experimental procedure. Although the suggested 1D-CNN-LSTM model was trained and tested on soil data specific to Karnataka, India, the principles of design and architecture support implementation over different soil compositions in different geographical positions.

The proposed framework is first evaluated for performance metrics like accuracy, recall, F1-score, and precision as well as error metrics like MAD, RMSE, and MAE, with its benchmarked comparison models. In terms of mean accuracy and precision, different figures compare the performance of the benchmarked SVM, KNN, RF, LSTM, 1D convolutional neural network, and ANN models with the proposed framework. The proposed framework demonstrated a significantly improved overall accuracy, less prediction loss and high precision. The proposed framework therefore has the potential to outperform the baseline models SVM, KNN, RF, LSTM, 1D-convolutional neural network, and ANN. When extending the model to other regions or contexts, there are several limitations to take into account. High-quality soil nutrient data is necessary for the model to work well. For areas with little to no data to collect, the model would underperform. For instance, you might have sandy soils, clay soils, or volcanic soils and soil properties can vary quite a bit across the region. This means that the same model might need heavy retraining or fine-tuning to work on such variations.

5 Conclusion

Soil nutrients plays a very important role in the agricultural crop production. As farming grows more intensive and demand rises the quality of the soil may decrease due to the usage of pesticides. One cannot rely on good soil with insufficient nutrient levels. This will reduce the crop yield. Precise information on the distribution of soil nutrients required for precision farming is provided by smart soil prediction. Automated soil prediction systems are primarily driven by deep learning and machine learning techniques. This paper offers a framework for identifying soil nutrients in sensor-gathered data and forecasting soil quality using hybrid deep learning. The classifiers will receive the soil data gathered by the sensors and use it to categorize the various soil nutrients. Subsequently, it will be classified into several categories by binary classifiers. This is intended to show how much more appropriate it is for any type of agricultural land to evaluate the state of the soil and detect any nutrient deficits.

Benchmark methods such as K-Nearest Neighbors, Support Vector Machine, Long Short Term Memory, Random Forest, Artificial Neural Network, and One-Dimension Convolutional Neural Network schemes are compared with the proposed enhanced Convolutional Neural Network-LSTM framework. According to the proposed model experimentation results, this model has achieved an accuracy of 98%. The future enhancement can be done for predicting the crop suitable with the available soil nutrients ratio in the soil. Further can be also done the soil nutrient prediction using the image dataset with hybrid deep learning techniques.

References

Ahmad, S., A. Kalra, and H. Stephen. 2010. Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, 33(1): 69-80.

Aouichaoui, A. R. N., R. Al, J. Abildskov, and G. Sin. 2021. Comparison of group-contribution and machine learning-based property prediction models with uncertainty quantification. *Computer Aided Chemical Engineering*, 50: 755-760.

Arunesh, K., and V. Rajeswari. 2017. Agricultural soil lime status analysis using data mining classification techniques. *International Journal of Advanced Technology in Engineering and Science*, 5(2): 27-35.

Bhargavi, P., and S. Jyothi. 2009. Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils. *International Journal of Computer Science and Network Security*, 9(8): 117-122.

Chartrand, G., P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang. 2017. Deep learning: a primer for radiologists. *RadioGraphics*, 37(7): 2113-2131.

Chatterjee, S., N. Dey, and S. Sen. 2020. Soil moisture quantity prediction using optimized neural supported model for sustainable agricultural applications. *Sustainable Computing: Informatics Systems*, 28: 100279.

Chiranjeevi, M. N., and R. B Nadagoudar. 2018. Analysis of Soil Nutrients using Data Mining Techniques. *International Journal of Recent Trends in Engineering and Research*, 4(7): 103-107.

Devi, M. P. K., Anthiyur, U., & Shenbagavadivu, M. S. (2016). *Improved crop yield estimate and soil data analysis using data mining*. International Journal of Modern Computer Science, 4(6).

Gholap, J., A. Ingole, S. Gargade, J. Gohil, and V. Attar. 2012.

Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction. arXiv preprint arXiv:1206.1557.

Golicz, K., S. H. Hallett, R. Sakrabani, and G. Pan. 2019. The potential for using smartphones as portable soil nutrient analyzers on suburban farms in central East China. *Scientific Reports*, 9(1): 16424.

Han, J., K. Mao, T. Xu, J. Guo, Z. Zuo, and C. Gao. 2018. A soil moisture estimation framework based on the CART algorithm and its application in China. *Journal of Hydrology*, 563: 65–75.

Ibukunoluwa Moyin-Jesu, E. 2015. Use of different organic fertilizers on soil fertility improvement, growth and head yield parameters of cabbage (*Brassica oleraceae* L) *International Journal of Recycling of Organic Waste in Agriculture*, 4(4): 291-298.

Indiveri, G., and S. C. Liu. 2015. Memory and information processing in neuromorphic systems. *Proceedings of the IEEE*, 103(8): 1379-1397.

Juhi Reshma, S. R., and D. J. Aravindhar. 2021. Fertilizer Estimation using Deep Learning Approach. *Natural Volatiles & Essential Oils*, 8(4): 5745-5752.

Kim, H. J., K. A. Sudduth, and J. W. Hummel. 2009. Soil macronutrient sensing for precision agriculture. *Journal of Environmental Monitoring*, 11(10): 1810-1824.

Kumar, A., and J. Kaur. 2024. Machine Learning and Deep Learning for Soil Analysis and Classification of Micro and Macro Nutrient Using IOT.[Preprint] 10.21203/rs.3.rs-4016181/v1.

Liu, W., Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi. 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234: 11-26.

Mmbaga, G. W., K. M. Mtei, and P. A. Ndakidemi. 2014. Extrapolations on the use of Rhizobium inoculants supplemented with phosphorus (P) and potassium (K) on growth and nutrition of legumes. *Agricultural Sciences*, 05(12): 1207-1226.

Moonrungsee, N, S. Pencharee, and J. Jakmunee. 2015. Colorimetric analyzer based on mobile phone camera for determination of available phosphorus in soil. *Talanta*, 136(4): 204-209.

Nikam, S. S. 2015. A Comparative study of classification Techniques in Data Mining Algorithms. *Oriental Journal of Computer Science & Technology*, 8(1): 13-19.

Pekel, E. 2020. Estimation of soil moisture using decision tree regression. *Theoretical & Applied Climatology*, 139: 1111–1119.

Prakash, S., and S. S. Sahu. 2020. Soil moisture prediction using shallow neural network. *International Journal of Advanced Research in Engineering and Technology*, 11(6): 426–435.

Prieto, A., B. Prieto, E. M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, and I. Rojas. 2016. Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214: 242-268.

Puno, J. C., E. Sybingco, E. Dadios, I. Valenzuela, and J. Cuello. 2017. Determination of soil nutrients and pH level using image processing and artificial neural network. In *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 1-6. Manila, Philippines, 01-03 December.

Rahman, S. A. Z., K. Chandra Mitra, and S. M. Mohidul Islam. 2018. Soil Classification using Machine Learning Methods and Crop Recommendation Based on Soil Series. In *2018 21st International Conference of Computer and Information Technology (ICCIT)*, 1-4. Dhaka, Bangladesh, 21-23 December.

Ramesh, V., and K. Ramar. 2011. Classification of agricultural land soils: A data mining approach. *Agricultural Journal*, 6(3): 82-86.

Riese, F. M., and S. Keller. 2019. Soil Texture Classification with 1D Convolutional Neural Network based on hyperspectral Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5: 615-621.

Sajindra, H., T. Abekoon, E. M. Wimalasiri, D. Mehta, and U. Rathnayake. 2023. An artificial neural network for predicting groundnut yield using climatic data. *AgriEngineering*, 5(4): 1713-1736.

Sarker, I. H. 2022. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science*, 3(2): 158.

Shahena, S., M. Rajan, V. Chandran, and L. Mathew. 2021. Conventional methods of fertilizer release. In *Controlled Release Fertilizers for Sustainable Agriculture*, eds. F. B. Lewu, T. Volova, S. Thomas, and K. R. Rakhimol, ch. 1, 1-24. New York: Academic Press.

Shi, Z., W. Ji, R. A. Viscarra Rossel, S. Chen, and Y. Zhou. 2015. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library. *European Journal of Soil Science*, 66(4): 679-687.

Swaminathan, B., S. Palani, and S. Vairavasundaram. 2023. Feature fusion based deep neural collaborative filtering model for fertilizer prediction. *Expert Systems with Applications*, 216: 119441.

Tantalaki, N., S. Souravlas, and M. Roumeliotis. 2019. Data-driven decision making in precision agriculture: the rise of big data in agricultural systems. *Journal of Agricultural & Food Information*, 20(4): 344-380.

Viscarra Rossel, R. A., and J. Bouma. 2016 Soil sensing: a new paradigm for agriculture. *Agricultural Systems*, 148: 71-74.

Wang, Z., S. Li, and S. Malhi. 2007. Effects of fertilization and other agronomic measures on nutritional quality of crops. *Journal of Science of Food and Agriculture*, 88(1): 7-23.

Wu, X., M. Bai, Y. Li, T. Du, S. Zhang, Y. Shi, and Y. Liu. 2020. The effect of fertigation on cabbage (*Brassica oleracea* L. var. *capitata*) grown in a greenhouse. *Water*, 12(4): 1076.