INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Morisita-Horn's Feature Matching Pursuit Based Contingency Correlative Recursive Deep Neural Network for Soil Fertility Prediction

R Jayalakshmi[1]*, C Rajagopal[2], K Sridharan[3]

1 Assistant Professor, Department of Computer Science, Sri Vidya Mandir Arts and Science College(Autonomous), Uthangarai, Krishnagiri, Tamil Nadu, India
2 Assistant Professor and Head, Department of Computer Science, Sri Vidya Mandir Arts and Science College (Autonomous), Uthangarai, Krishnagiri, Tamil Nadu, India
3 Assistant Professor, Department of Computer Applications, Sri Vidya Mandir Arts and Science College (Autonomous), Uthangarai, Krishnagiri, Tamil Nadu, India

OPEN ACCESS

Received: 14-11-2024

Accepted: 01-02-2025

Published: 04-03-2025

Citation: Jayalakshmi R, Rajagopal C, Sridharan K (2025) Morisita-Horn's Feature Matching Pursuit Based Contingency Correlative Recursive Deep Neural Network for Soil Fertility Prediction. Indian Journal of Science and Technology 18(9): 658-670. https://doi.org/10.17485/IJST/v18i9.3704

* **Corresponding author**.

jayasvm179@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives: Soil is a non-renewable resource monitored continuously to prevent degradation and obtain precision agriculture. Due to the high cost and low efficiency of conventional field sampling and laboratory testing models, it is hard to achieve large-scale and real-time monitoring of dynamic soil properties. Therefore, a novel technique is introduced for enhanced prediction of soil fertility. **Methods:** Morisita-Horn's Feature Matching Pursuit based Contingency Correlative Recursive Deep Neural Network (MFMP-CCRDNN) is introduced for soil fertility prediction. **Findings:** Experimental assessment is carried out on factors such as prediction accuracy, false-positive rate, processing time, and space complexity. The quantitative assessment results reveal the MFMP-CCRDNN technique attains higher prediction accuracy and lesser false positive rate, time, and space complexity than the state-of-the-art methods. **Novelty:** The Recursive Deep Neural Network is designed with multiple layers such as one input layer, two hidden layers, and one output layer. Here, trained the input features and data are considered from the dataset for performing the prediction process. Initially, Input features are analyzed in the first hidden layer. Preprocessing is carried out to eradicate noisy data. Morisita-Horn's Feature Matching Pursuit is to select the significant features. Then the selected features are sent to the next hidden layer. Here, data classification is performed using a Contingency Correlative Recursive Deep Neural Network based on McNemar's test. Then the classified data are sent to the output layer for accurate Fertility Prediction with minimum time.

Keywords: Soil Fertility Prediction; Pearson Correlative Recursive Deep Neural Network; Morisita-Horn's Feature Matching Pursuit; McNemar's test; Mean square contingency coefficient

1 Introduction

Soil physical and chemical properties play a most important role in estimating soil quality and fertility. Soil is one of the important factors as healthy soils provide healthy yields and precision agriculture. Therefore, predicting the crop yield based on the soil parameters is a prospective research area. Various Deep-learning-based models are broadly used to extract significant soil features for fertility prediction. The development of deep learning approaches effectively analyzes soil features and plant growth characteristics for predicting soil fertility. Here, soil features are considered to be soil pH, temperature, soil moisture, potassium, phosphorus, and organic carbon.

Table 1. Summarization of literature and proposed MFMP-CCRDNN works

Method	Contribution	Merits	Demerits	Novelty of proposed MFMP-CCRDNN
Deep Recurrent Q-Network (DRQN)	Deep Recurrent Q-Network over the Q-Learning reinforcement learning technique [1]	Attains higher prediction accuracy	It failed to construct a more robust model efficiency	Recursive Deep Neural Network used for enhancing the training process
Deep Convolutional Neural Networks (DCNNs)	The soil properties are analyzed to predict the performance of multi-task modeling [2]	Accurate predictions for soil properties	Both prediction accuracy and time complexity analysis were not focused	Prediction accuracy and time complexity were considered
Extreme Learning Machine (ELM)	Various activation functions [3]	Maximum pH classification level	Deep feature analysis was not performed	Recursive Deep Neural Network employed for deeply analyzing the features
Partial Least Squares Regression model	Soil fertility and productivity were predicted through machine learning [4]	Estimate soil's fertility with higher efficiency	The time consumption was not focused	Time complexity was considered
Geographically Weighted Principal Component Analysis (GWPCA)	Soil fertility quality index (SFQI) analysis [5]	An effective method for SFQI assessment in large-scale areas.	Deep learning analysis was not performed	Recursive Deep Neural Network performed for classification
Machine learning techniques	Two activation functions for better prediction [6]	Better classification of soil nutrient parameters	Difficult to handle huge sizes of training samples	The huge size of training samples considered
Cuckoo–Grey wolf-based Correlative Naïve Bayes classifier	Big data classification [7]	Enhanced data classification process	Deep learning model fails to attain higher result	Recursive Deep Neural Network performed to enhance accuracy
A fuzzy correlative naive Bayes (FCNB) classifier	Correlative naive Bayes (CNB) classifier [8]	Improves the classification performance	Prediction performance is difficult	Contingency Correlative Recursive Deep Neural Network used for improving prediction performance
Four machine learning techniques	ELM, ANN, classification, regression trees and group method of data handling [9]	Estimation of soil temperature	It failed to use more samples for analyzing the soil temperature	Soil temperature examined by using Mean square contingency correlation coefficient

Continued on next page

Table 1 continued

A hybrid Fractionally Autoregressive Integrated Moving Average (FAIMA) and networks (FFBPNN)	Daily soil temperature estimation [10]	Soil temperature is accurately predicted with minimum error	The complexity was not minimized	Relevant features selected with less time using Morisita-Horn's Feature Matching Pursuit
Extreme Learning Machine (ELM) algorithm	Performs feature selection process for forecasting the soil status [11]	The performance of ELM is improved for the prediction	It failed to improve efficiency	McNemar's test performed among the testing and training features to enhance efficiency
Different machine learning algorithms	Prediction of Soil organic carbon [12]	Attains higher prediction accuracy	It was not efficient to enhance Soil prediction	Soil prediction improved using Recursive Deep Neural Network
Linear mixed model	Spatial variability and mapping of soil fertility status [13]	Determine soil fertility with higher accuracy	Soil properties cause issues in fertility prediction	Real-time monitoring of dynamic soil properties considered
Residual Maximum Likelihood and Empirical Best Linear Unbiased Predictor (REML-EBLUP)	The spatial prediction of soil pH value [14]	The prediction error is minimized	However, accuracy during fertility prediction was not focused	McNemar's test used to improve accuracy
A regional-scale hyper spectral prediction approach	Prediction of soil organic carbon with spatial distributions [15]	This resulted in higher prediction accuracy	But, time for classified data was not minimized	Important features are selected using Feature Matching Pursuit to minimize time.
An Intelligent Data Prediction (IDP) method	Soil heavy metal prediction [16]	Effectively classifies soli levels	However, it failed to improve the speed and accuracy of model	McNemar's test is used in Recursive Deep Neural Networks to enhance efficiency
The fuzzy control system	Based on fuzzy rules, soil quality and fertility estimation [17]	Increased soil quality and fertility prediction accuracy	The soil quality assessment using the Deep learning model is difficult	Recursive Deep Neural Network introduced for soil quality assessment
Machine learning models	Regression and classification techniques for crop yield prediction with score of soil level [18]	Achieve higher accuracy in soil quality prediction and assessment	It fails to address issues on soil level identification	Soil level identification measured using novel feature selection and classification
Machine Learning Strategy	Soil nutrient prediction Using Spectroscopic Method [19]	Prediction of soil properties in accurate manner	However, predictions were limited	Soil quality computed with Mean square contingency correlation coefficient

Continued on next page

Table 1 continued

Multi-target Stacked Generalization	The soil fertility prediction [20]	Soli quality prediction with minimum error	However, it fails to achieve enhanced prediction accuracy	Prediction accuracy improved with Recursive Deep Neural Network
1D convolution neural networks (1D-CNN)-based soil fertility classification method	1D-CNN-based soil fertility classifier and fertilizer prescription [21]	Higher performance result of soil fertility prediction	However, the deep learning technique failed to apply as a classifier	Recursive Deep Neural Network employed for performing classification
Deep regression model	Predict tea crop yield [22]	Improve accuracy	However, it failed to consider space complexity	Space complexity considered for dimensionality reduction technique
3D-CNN ConvLSTM ViT named convolutional long short-term memory (ConvLSTM), three-dimensional convolutional neural network (3D-CNN), and vision transformer (ViT)	Precise crop yield forecasting [23]	Improving decision-making processes	However, time was not considered	Vital features chosen using Feature Matching Pursuit with less time
Topsoil fertility Exergy appraisal	Measure soil fertility [24]	Increase accuracy	But, time was not minimized	Feature Matching Pursuit is used for selecting essential features with less time
DL and dimensionality reduction (DR) methods	Crop yield prediction [25]	Lower time	However, the huge volume and complexity of agricultural data were major challenges	Huge datasets such as soil health datasets considered
Optimization-based feature selection and filtering approaches	Select vital features [26]	Improve crop production	However, it failed to perform classification	Classification carried out by Recursive Deep Neural Network
Multi-parametric Deep Neural Network (MDNN)	Accurate prediction of crop yield [27]	Enhance crop yield predictive performance	However, the time was not considered	Time focused on Feature Matching Pursuit
Hybrid deep learning framework	Temporal and spatial data extracted [28]	Enhanced accuracy	But, it failed to improve classification performance	Recursive Deep Neural Network developed for increasing classification performance

The major contributions of the proposed MFMP-CCRDNN technique are summarized as follows:

- To predict soil fertility with higher accuracy and minimized prediction time, a novel technique called Morisita-Horn's Feature Matching Pursuit based Contingency Correlative Recursive Deep Neural Network (MFMP-CCRDNN) is introduced based on feature selection and classification process.
- Morisita-Horn's Feature Matching Pursuit is applied for feature selection by the MFMP-CCRDNN technique. The novelty of Morisita-Horn's index is employed in the proposed MFMP-CCRDNN technique to estimate the similarity between

two features and the range from '0' and '1'. '+1' denotes a similar feature for an accurate forecasting process whereas 0' represents no similarity among the two features. The relevant feature is discovered and other features are eradicated from the dataset. In this way, soil fertility prediction time and space complexity are reduced.
- Mean square contingency coefficient is utilized in the MFMP-CCRDNN technique for classification with the novelty of McNemar's test. The similarity between training features and testing features is measured. Based on similarity value, the samples are classified into different classes. Types of soil fertility such as low fertile, very low fertile, medium fertile, high fertile, and very high fertile are predicted. With this, the prediction accuracy is enhanced and the false positive rate is minimized.
- Finally, experimental assessments are carried out with the various methods and it also helps to find that the MFMP-CCRDNN technique outperforms the other related approaches.

The rest of the article is systematized into different sections. Section 2 reviews the related work. The description of the MFMP-CCRDNN technique and system frameworks is detailed in section 3. Section 4 explains the experimental setup, result analysis, and comparison with other state-of-the-art methods presented in Section 5. Finally, the conclusion is presented in section 6.

Research gap

Soil fertility prediction is described as the process of estimating various soil levels for improving a given recent plot and season in meteorological associations. The fertility prediction helps to analyze the soil data, monitoring data, and other modeling methods. By managing soil features level, fertility prediction is accurately performed for maintaining agricultural growth and security globally. The utilization of a deep learning model effectively performs prediction on soil fertility in agriculture with higher accuracy. While comparing with other machine learning models, fertility prediction is one of the most attractive and challenging tasks in the soil agriculture field. However, conventional methods do not consider the importance of soil level and provide issues with soil properties in different agriculture stages. Weather conditions and insect killer usage greatly influence soil degradation. Various methods and technologies have been designed for better prediction of Soil fertility but they often face challenges related to accuracy and time complexity in soil properties. Based on atmospheric various conditions, the classification of sample data in fertility prediction was not sufficient.

2 Methodology

Soil fertility is the main part of soil production. Soil management in agriculture research aims to enhance crop yield and dynamic soil parameters. Various methods were employed to examine the soil prediction. However, the time of sample analysis was higher and more challenging. Therefore, a novel technique is needed to enhance the soil fertility prediction with less time for enhancing crop productivity. Prediction of soil parameter levels is employed to avoid unnecessary expenditure on fertilizers. This prediction is utilized to save time for analyzing the soil health and environmental conditions. In this section, MFMP-CCRDNN is introduced.

Figure 1 illustrates the architecture of the proposed MFMP-PCRDNN technique that consists of three phases namely preprocessing, feature selection and classification. First, the input soil property features $F = f_1, f_2, \ldots, f_n$ acquired and collected from the soil health dataset for prediction. The soil health dataset is collected from https://soilhealth.dac.gov.in/. At first, the preprocessing is employed to remove the noisy data and clean the data. In the second phase, the feature selection process is carried out to improve soil quality analysis for accurate prediction. Thirdly, with the selected feature, classification is performed using Mean square contingency coefficient based classification. Finally, a brief explanation of the process of the proposed MFMP-CCRDNN technique is given below.

2.1 Recursive Deep Neural Networks

The recursive neural network is a kind of deep neural network used for deeply analyzing the feature and performing the classification with the help of multiple layers. A recursive neural network is formed by applying the same set of weights recursively over a structured input, to generate a structured prediction over variable-size input structures. The structural design of the recursive neural network comprises neurons like the nodes that are connected from one layer to another layer to make the entire network.

Figure 2 depicts the block diagram of a schematic structure of a Recursive Deep Neural Network to perform the soil fertility prediction.

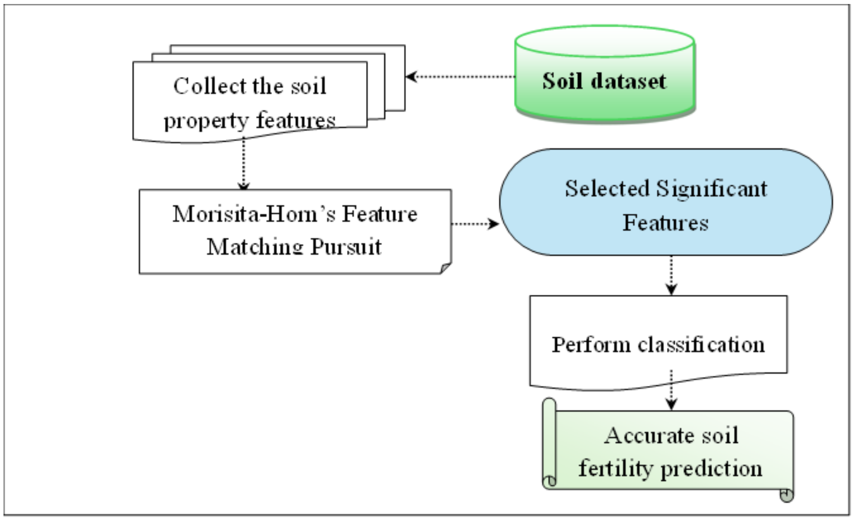

Fig 1. Architecture of the Proposed MFMP-CCRDNN

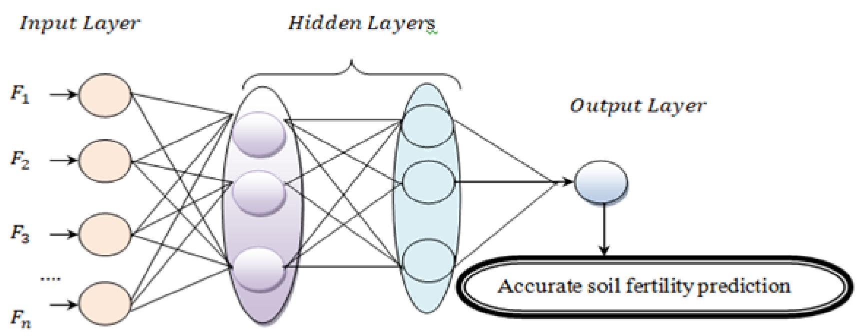

Fig 2. Schematic Structure of Recursive Deep Neural Network

The above network structure includes the input layer and multiple hidden layers. The input layer consists of input of a number of features $F = f_1, f_2, \ldots, f_n$ from dataset. In the first hidden layer, preprocessing and feature selection are carried out. Preprocessing is to eliminate noise data. The input is transformed into the first hidden layer. Followed by, the Morisita-Horn's Feature Matching Pursuit is performed at the first hidden layer. Then the similarity between the two features is estimated by applying a Morisita-Horn's index. Relevant features are identified and irrelevant features are eliminated to reduce the dimensions of the input features. In the second hidden layer, the classification is performed with the selected features. Finally, the similarity value is sent to the output layer where the Mean square contingency coefficient is applied to classify the results.

The activity of the neuron at the input layer is expressed as follows,

$$i(t) = b + \left[\sum_{i=1}^{n} F_i * \beta\ \right] \quad (1)$$

Where, $i(t)$ denotes neuron activity, F_i indicates the number of features, β indicates a weight between the input and hidden layer, $'b'$ indicates a bias that stored the value is +1, F_i indicates features. The input is transferred into the first hidden layer where the preprocessing and feature selection.

- **Preprocessing**

The proposed MFMP-CCRDNN technique is to perform image preprocessing. Different types of noise are included in the input data. This type of noise affects the data illustration and minimizes classification accuracy. Hence, it is necessary to remove the high-frequency noise and provide a better visual interpretation. The preprocessing is vital to clean and transform the raw data. It is used for removing the noisy data and cleaning the data.

- **Morisita -Horn's Feature Matching Pursuit**

The proposed MFMP-CCRDNN technique performs the feature selection using Morisita-Horn's Feature Matching Pursuit. Owing to the huge volume of data, dimensionality reduction is challenging. Feature selection aims to select significant features and eliminate irrelevant features to reduce the time. Feature selection is to minimize the dimensionality of the dataset. In our work, Morisita-Horn's Feature Matching Pursuit is a dimensionality reduction statistical technique used for finding the possible projections in multidimensional data. Matching pursuit is utilized for identifying the best match of multidimensional data. The feature matching pursuit maps the relevant features into lower-dimensional space as given below.

$$f(x) : F_s \to S \tag{2}$$

Where, $f\ (x)$ indicates a projection Matching Pursuit to project the significant features (F_s) related to the subset 'S'. The significant features are identified through Morisita-Horn's index. Morisita-Horn's index is a statistical measure to discover the similarity between two features. The similarity range varies from 0 or 1. Morisita-Horn's index is computed as given below,

$$\vartheta = 2 * \left[\frac{\sum f_i * f_j}{\sum {f_i}^2 + \sum {f_j}^2} \right] \tag{3}$$

From the Equation (3), Morisita-Horn's index 'ϑ' is measured based on the sum of the product of paired score of two features '$\sum f_i * f_j$' and the squared score of the feature f_i is '$\sum {f_i}^2$' and a squared score of feature 'f_j' is $\sum {f_j}^2$'. Morisita-Horn's index provides the output value between '0' and '1'. The index with '+1' indicates a similar feature whereas 0' represents no similarity between the two features.

$$\vartheta = \begin{cases} +1 & \text{Similar features} \\ 0 & \text{Dissimilar features} \end{cases} \tag{4}$$

The value of '$\vartheta = +1$' indicates that the features are similar whereas the value of '0' indicates that the features are not similar. Then the selected features are given as the input to the next hidden layer for data classification. In this way, more similar feature sets are selected for classification to reduce the time complexity.

- **Mean square contingency coefficient based classification**

After the feature selection, classification is performed using the Mean square contingency coefficient in the second hidden layer. The relevant feature set is utilized as the input of the next hidden layer. Classification is carried out to classify the different classes of soil fertility levels such as Low fertile, Very Low fertile, Medium fertile, High fertile, Very High fertile. The proposed technique uses the input as selected relevant features $F = f_1, f_2, \ldots, f_m$ and the number of classes $C = c_1, c_2, \ldots, c_k$. The contingency correlation coefficient is an arithmetical measure of the relationship between two variables. It values from -1 to 1. The mean square contingency correlation coefficient is also called as phi coefficient. Then applying the Mean square contingency correlation coefficient is to measure the correlation between the relevant features and number of classes using McNemar's test. McNemar's test is a statistical test for soil fertility prediction on a matched pair of soil data. For measuring the McNemar test, the data is placed into a 2×2 contingency table, with the cell frequencies equaling the number of pairs. McNemar's test as given below,

$$r_\varphi \to \delta = \left(\frac{(f_m - f_t|^2}{f_m + f_t} \right] \tag{5}$$

From Equation (5), 'δ' denotes McNemar's test results, 'r_φ' represents Mean square contingency correlation coefficient. The mean square contingency correlation coefficient is defined as the mean square of the difference between the relevant features 'f_m ' and the testing feature value 'f_t'. The output of the coefficient returns a value between -1 and +1. Therefore, the hidden layer output is obtained as follows,

$$R\ (t) = [\textstyle\sum_{i=1}^{n} F_i * \beta\ ] + [\beta *\ R\,(t-1)] \tag{6}$$

Where, $R\ (t)$ indicates an output of a hidden layer, $R\ (t-1)$ indicates output from the previous hidden layer and 'β 'indicates the weight of the layers, F_i represents the input feature value. Here, the same set of weights is recursively used over a structured input. The classification output is obtained at the output layer,

$$Y = [\beta *\ R\ (t)] \tag{7}$$

Where 'Y' denotes an output of the recursive deep neural network, β indicates a similar weight between hidden and output, $R\ (t)$ indicates an output of a hidden layer. In this way, the soil fertility level is correctly predicted with higher accuracy based on the correlation coefficient.

The algorithmic process of MFMP-CCRDNN is described as given below:

// Algorithm 1: Morisita-Horn's Feature Matching Pursuit based Contingency Correlative Recursive Deep Neural Network

Input: Soil Dataset '???', Features '?? = $?_1, ?_2, \ldots, ?_?$', Class '?? = $?_1, ?_2, \ldots, ?_?$'
Output: Increase soil fertility prediction accuracy

Begin
Collect the number of features ?? = $?_1, ?_2, \ldots, ?_?$ into **an** input layer
Transform the features into the first hidden layer
For each feature
Measure Morisita-Horn's index '??' ---**hidden layer 1**
If (?? = +1) **then**
Features are said to be a relevant
Select the relevant features
else
Features are said to be an irrelevant
Remove the irrelevant features
End if
For each selected Feature '$?_?$' ---**hidden layer 2**
Initialize number of classes ?? = $?_1, ?_2, \ldots, ?_?$
Measure the Mean square contingency correlation '$?_?$'
if ($?_?$ > +1) **then**
soil is classified as Very High Fertile'
else if ($?_?$ = 1) **then**
soil is classified as 'High Fertile'
else if ($?_?$ = 0) **then**
soil is classified as 'Medium Fertile'
else if ($?_?$ = - 1) **then**
soil is classified as 'Low Fertile'
else if ($?_?$ < - 1') **then**
soil is classified as 'Very Low Fertile's
End if
Repeat the process until minimum error is attained
End for
End

Algorithm 1 given above explains the step-by-step process of soil fertility prediction using a Recursive Deep Neural Network. The number of features is collected from the dataset. After that, the dimensionality reduction technique called Morisita-Horn's Feature Matching Pursuit is applied to select the more relevant features in the first hidden layer to minimize the time as well as space complexity. Followed by, the classification is performed using Mean square contingency correlation. Based on the correlation measure, the soil fertility level is correctly predicted in the second hidden layer. This in turn increases the prediction accuracy.

3 Results and Discussion

The experimental evaluation of proposed MFMP-CCRDNN, existing DRQN[1], DCNN[2], Deep Regression[22], and 3D-CNN ConvLSTM ViT[23] are implemented using Java for soil fertility prediction using soil health dataset collected from https://soilhealth.dac.gov.in/ during the period 2016–2017 for Erode district, Tamil Nadu. This dataset consists of 15 attributes and a total of more than 5000 instances from the soil sample test report. The Sample No, Soil pH value, electrical conductivity (EC), Organic carbon (OC), Nitrogen (N), Phosphorus (P), Potassium (K), Sulphur (S), Zinc (Zn), Iron (Fe), Copper (Cu), Manganese (Mn), Calcium (Ca), Boron (B) and Class (Very high/ High/ Medium/ Low/Very low). Among the 15 features, more relevant features are selected for classification. For the experimental evaluation, 5000 to 50000 numbers of samples are considered as input.

Cross-validation is to calculate the performance of the model's ability to carry out unseen data. The aim of cross-validation is to improve the model generalization. By using cross-validation, the dataset is divided into two sets namely training and testing. More data samples (70%) are used for training and the remaining (30%) are considered for testing. In our work, the **10-fold cross-validation** is employed to estimate the results.

The performance results of MFMP-CCRDNN are compared to existing methods with various metrics such as soil fertility prediction accuracy, false-positive rate, soil fertility prediction time, and space complexity. The different graphical results show the performance results of the proposed and existing methods.

- **Soil fertility prediction accuracy:** It is defined as the number of samples accurately predicted through the classification to the total number of samples. The accuracy rate is measured as follows.

$$SFP_{acc} = \left(\sum_{i=1}^{n} \frac{S_{accpred}}{S_i} * 100 \right) \tag{8}$$

Where, SFP_{acc}s denotes soil fertility prediction accuracy, $S_{accpred}$ denotes the number of samples accurately predicted, the number of samples provided as input 'S_i' for experiments. It is measured in terms of percentage (%).

- **False- positive rate**: It is measured as the number of soil samples that were incorrectly predicted to the total number of samples. This is mathematically estimated as given below.

$$FPR = \left(\sum_{i=1}^{n} \frac{S_{wronglypred}}{S_i} * 100 \right) \tag{9}$$

Where 'FPR' denotes a false positive rate, '$S_{wronglypred}$' denotes samples wrongly predicted, S_i denotes the total number of samples. It is measured in terms of percentage (%).

- **Soil fertility prediction time**: It is defined as the amount of time consumed by the algorithm to perform Soil fertility prediction through classification. It is mathematically expressed as given below.

$$SFP_{time} = \sum_{i=1}^{n} S_i * Time\ [SFP] \tag{10}$$

Where, SFP_{time} denotes a soil fertility prediction time, 'S_i' denotes the number of samples considered for simulation, '$Time\ [SFP]$' denotes a time involved in predicting soil fertility. It is measured in terms of milliseconds (ms).

- **Space complexity**: It is defined as the amount of memory space consumed by the algorithm to perform Soil fertility prediction through classification. It is measured using the following formula,

$$Comp_{Space} = \sum_{i=1}^{n} S_i * Mem\,[SFP] \tag{11}$$

Where, $Comp_{Space}$ denotes a Space complexity, 'S_i' denotes the number of samples considered for simulation, '$Mem\,[SFP]$' denotes a memory involved in predicting soil fertility. It is measured in terms of megabytes (MB).

Table 2. Soil Fertility Prediction Accuracy

Number of samples	Soil fertility prediction accuracy (%)				
	MFMP-CCRDNN	**DRQN**	**DCNN**	**Deep Regression**	**3D-CNN ConvL-STM ViT**
500	92	86.6	88	90	91
1000	91.8	85	87	88	89.5
1500	92	86.66	88.66	89.66	90

Continued on next page

Table 2 continued

2000	91.5	85.5	88	88.75	89.25
2500	91.2	86.4	88.4	89.2	89.8
3000	91.66	84.66	88.33	89.33	90.33
3500	92.28	87.14	89.71	90.85	91.28
4000	91.25	85.25	87	88	90.62
4500	91.33	84.66	88.66	89.77	90.55
5000	93	86	89	89.8	90

Table 2 given above provides the soil fertility prediction accuracy results using MFMP-CCRDNN and existing DRQN[1], DCNN[2], Deep Regression[22], and 3D-CNN ConvLSTM ViT[23]. It is obvious from Table 1 that the MFMP-CCRDNN provides higher soil fertility prediction accuracy than the existing methods. With this, the Soil fertility prediction accuracy using MFMP-CCRDNN was found to be 92%, 86.6%[1], 88%[2], 90%[22], and 91%[23]. Followed by, various results are observed for each method with different counts of input samples. Finally, the performance of MFMP-CCRDNN is compared to existing methods. The comparison results prove that the soil fertility prediction accuracy of MFMP-CCRDNN is considerably increased by 7% when compared to[1], 4% when compared to[2], 3% when compared to[22], and 2% when compared to[23]. The reason behind the improvement was the application of Mean square contingency correlation in a Recursive Deep Neural Network based on McNemar's test. The McNemar's test is performed between the testing and training features and returns the classification of different fertility levels with higher accuracy.

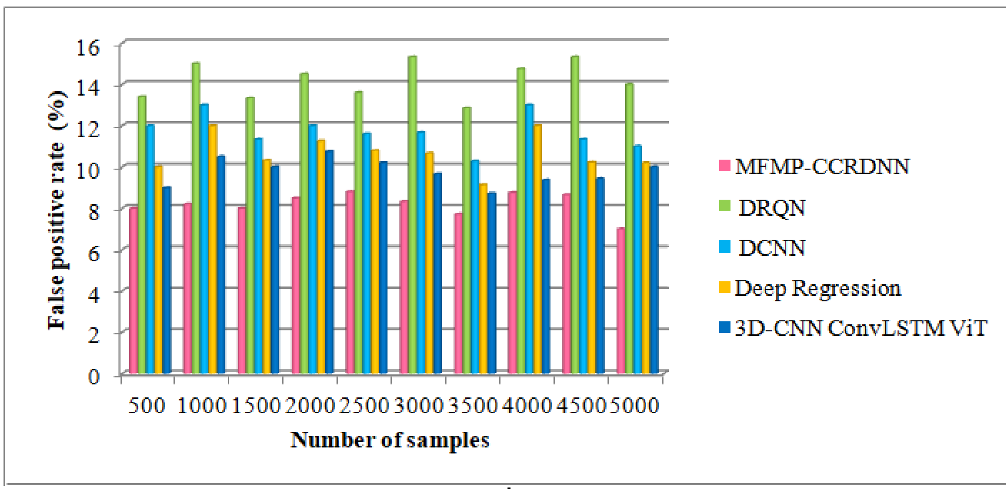

Fig 3. Graphical plot of false-positive rate

Figure 3 illustrates the performance result of the false-positive rate versus a number of samples in the ranges from 500 to 5000. As shown in Figure 4, 'x' axis represents the number of samples and 'y' axis represents the false positive rate. The pink, green, sky blue, yellow, and blue color denotes the MFMP-CCRDNN technique, existing DRQN[1], DCNN[2], Deep Regression[22], and 3D-CNN ConvLSTM ViT[23] . When taking 500 samples for the experimental process, 50 samples are incorrectly predicted using the proposed MFMP-CCRDNN technique. For that reason, the proposed MFMP-CCRDNN technique achieves 8% FPR whereas conventional[1,2,22,23] are observed as 13.4%, 12%, 10% and 9% respectively. In the same way, other runs are calculated in order to investigate the soil fertility prediction performance of the proposed MFMP-CCRDNN technique and the existing two methods. Compared to all three methods, the proposed MFMP-CCRDNN provides better performance in terms of achieving the minimum false positive rate. From this simulation result, the false positive rate using MFMP-CCRDNN is said to be comparatively minimized by 42%, 30%, 23%, and 16% when compared to[1,2,22,23] . The reason behind the minimum false positive rates was owing to the application of McNemar's test results used in the Mean square contingency correlation coefficient with the training features and the testing feature value. Moreover, the performance of Recursive Deep Neural Network effectively analyzes the feature value and minimizes the incorrect classification of soil fertility level.

Table 3. Soil Fertility Prediction Time

Number of samples	Soil fertility prediction time (ms)					
	MFMP-CCRDNN	DRQN	DCNN	Deep Regression	3D-CNN ViT	ConvLSTM
500	27.5	36	34	31.5	30	
1000	32.5	52.5	46.5	42.6	34.6	
1500	38.25	56.25	50.25	48.75	45	
2000	53	69.76	63.1	60.81	58.3	
2500	61.25	81.5	71.37	69.5	65.75	
3000	67.5	85.8	79.65	78.52	69.13	
3500	75.25	89.95	85.92	83.1	78.25	
4000	84.2	99.52	94.2	92.8	88.2	
4500	92.47	110.25	101.02	99.92	97.55	
5000	102.2	115	106.5	111.5	107.5	

The performance results of the soil fertility prediction time of five different techniques namely MFMP-CCRDNN and existing DRQN[1], DCNN[2], Deep Regression[22], and 3D-CNN ConvLSTM ViT[23] are shown in Table 2. The observed results indicate that the MFMP-CCRDNN minimizes the soil fertility prediction time. After obtaining the ten results, the observed time consumption of the MFMP-CCRDNN technique is compared to the results of existing methods. The average of ten results indicates that the proposed MFMP-CCRDNN technique reduces the overall time consumption by 22%, 15%, 13%, and 7% as compared to the[1,2,22,23], respectively. From the obtained result, the soil fertility prediction time of all the methods is slowly increased while increasing the number of input samples for different runs. But comparatively, the prediction time gets reduced using the MFMP-CCRDNN technique. The major reason is to apply the dimensionality reduction technique called Morisita-Horn's Feature Matching Pursuit for selecting the significant features and removing the other features from the dataset for soil fertility level classification to minimize the time consumption.

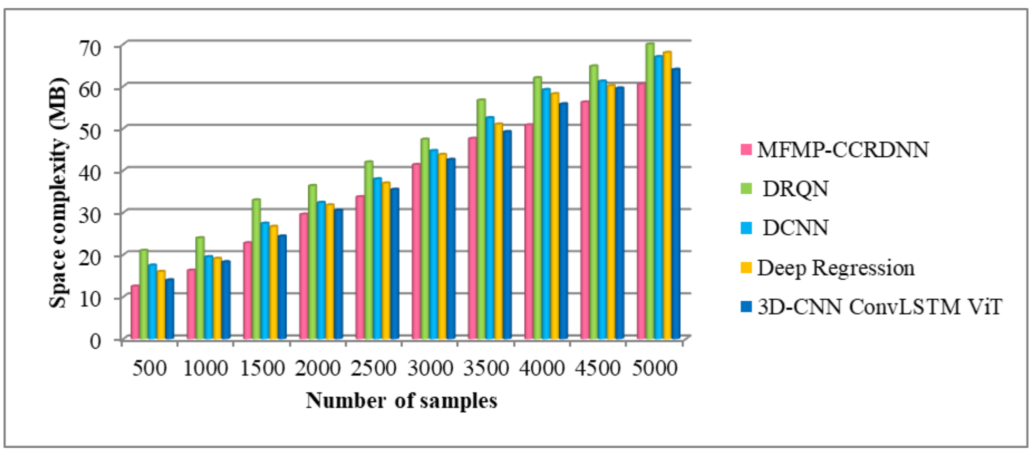

Fig 4. Graphical plot of Space complexity

Finally, Figure 4 shows the space complexity with respect to a number of samples. For simulation purposes, the number of samples is considered ranging from 500 to 5000. In the figure, 'x' axis denotes the number of samples and 'y' axis represents the space complexity of proposed and existing methods. However, with the simulations conducted for 500 samples, the memory consumed for predicting the space complexity was found to be 0.025, 0.042, 0.035, 0.032 and 0.028 using the MFMP-CCRDNN technique, existing[1,2,22,23]. With this, the space complexity using MFMP-CCRDNN was found to be $12.5MB$, $21MB$ using[1], $17.5MB$ using[2], 16using[22] and $14MB$ using[23]. Also, the various results are measured by getting different counts of input samples. From this observed result, the space complexity using MFMP-CCRDNN is said to be comparatively minimized by 22%, 13%, 11%, and 6% when compared to[1,2,22,23] respectively. The reason behind the minimum space complexity was owing to the application of finding the relevant features for classifying the different soil fertility levels. As a result, the proposed MFMP-

CCRDNN technique consumes less memory space for predicting the soil fertility level.

4 Real-Time Applications

Agriculture is a real-time application area in India for crop yield prediction based on soil properties. In agriculture, soil fertility prediction is a significant process for providing superior crop productivity. Since the soil characteristics are diverse from area to area. MFMP-CCRDNN technique is proposed in this research work for precise analysis of soil fertility in an application area of agriculture. The proposed MFMP-CCRDNN technique is used to remove noise and deeply examine soil features in the first hidden layer with preprocessing and Morisita-Horn's Feature Matching Pursuit concepts. Significant features are chosen in less time. Next, second hidden layer employs a classification process. Contingency Correlative Recursive Deep Neural Network classifies data with higher accuracy.

5 Conclusion

An efficient MFMP-CCRDNN technique is introduced to enhance the classification of samples for predicting soil fertility and productivity with minimum time. The main objective of soil fertility prediction is achieved with the application of Morisita-Horn's Feature Matching Pursuit and Mean square contingency coefficient-based classification process. The proposed Deep Neural Network comprises three processes namely preprocessing, feature selection, and classifier process. At first, noisy data is removed from the dataset by using preprocessing. Morisita-Horn's Feature Matching Pursuit is performed to discover the most relevant features from the dataset. After selecting relevant features, a Contingency Correlative Recursive Deep Neural Network based on McNemar's test is applied to effectively classify data points. Here, McNemar's test is carried out to classify data for accurate Fertility Prediction with minimum time. Hence, the MFMP-CCRDNN technique effectively classifies the soil sample data for accurate prediction of soil fertility level. The comprehensive experimental assessments are carried out with respect to a number of samples based on metrics of soil fertility prediction accuracy, false-positive rate, soil fertility prediction time, and space complexity. The results of the proposed technique is conducted and compared with two conventional algorithms. The observed results have confirmed that the proposed MFMP-CCRDNN technique has higher soil fertility prediction accuracy with minimized false-positive rate, soil fertility prediction time, and space complexity than the other state-of-the-art methods. However, the various parameters such as precision, recall, and F-measure are not considered. Also, the proposed MFMP-CCRDNN technique is considered only one dataset. In future work, a novel deep learning method will be used to consider different parameters and datasets for finding the soil fertility level with less time.

References

1) Elavarasan D, Vincent PMD. Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. *IEEE Access*. 2020;8:86886–86901. Available from: https://ieeexplore.ieee.org/document/9086620.
2) Zhong L, Guo X, Xu Z, Ding M. Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. *Geoderma*. 2021;402. Available from: https://doi.org/10.1016/j.geoderma.2021.115366.
3) Suchithra MS, Pai ML. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. *Information Processing in Agriculture*. 2020;7(1):72–82. Available from: https://doi.org/10.1016/j.inpa.2019.05.003.
4) Helfer GA, Barbosa JLV, dos Santos R, da Costa AB. A computational model for soil fertility prediction in ubiquitous agriculture. *Computers and Electronics in Agriculture*. 2020;175. Available from: https://doi.org/10.1016/j.compag.2020.105602.
5) Chen J, Qu M, Zhang J, Xie E, Huang B, Zhao Y. Soil fertility quality assessment based on geographically weighted principal component analysis (GWPCA) in large-scale areas. *CATENA*. 2021;201. Available from: https://doi.org/10.1016/j.catena.2021.105197.
6) Pant J, Pant P, Pant RP, Bhatt A, Pant D, Juyal A. Soil Quality Prediction For Determining Soil Fertility In Bhimtal Block of Uttarakhand (India) Using Machine Learning. *International Journal of Analysis and Applications*. 2021;19(1):91–109. Available from: https://etamaths.com/index.php/ijaa/article/view/2240.
7) Banchhor C, Srinivasu N. Integrating Cuckoo search-Grey wolf optimization and Correlative Naive Bayes classifier with Map Reduce model for big data classification. *Data & Knowledge Engineering*. 2020;127. Available from: https://doi.org/10.1016/j.datak.2019.101788.
8) Banchhor C, Srinivasu N. FCNB: Fuzzy Correlative Naive Bayes Classifier with MapReduce Framework for Big Data Classification. *Journal of Intelligent System*. 2020;29(1):994–1006. Available from: https://www.degruyter.com/document/doi/10.1515/jisys-2018-0020/html?lang=en&srsltid=AfmBOoqstQg6uXgsiX2Ue1Tv-K729PSqeSZCVInr_N29XuzBRdte5mz5.
9) Alizamir M, Kisi O, Ahmed AN, Mert C, Fai CM, Kim S, et al. Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *PLoS ONE*. 2020;15(4). Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7156082/.
10) Mehdizadeh S, Fathian F, Safari MJS, Khosravi A. Developing novel hybrid models for estimation of daily soil temperature at various depths. *Soil and Tillage Research*. 2020;197. Available from: https://doi.org/10.1016/j.still.2019.104513.
11) Pham BT, Nguyen-Thoi T, Ly HB, Nguyen MD, Al-Ansari N, Van-Quan Tran, et al. Extreme Learning Machine Based Prediction of Soil Shear Strength: A Sensitivity Analysis Using Monte Carlo Simulations and Feature Backward Elimination. *Sustainability*. 2020;12(6). Available from: https://doi.org/10.3390/su12062339.

12) John K, Isong IA, Kebonye NM, Ayito EO, Agyeman PC, Afu SM, et al. Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. *Land*. 2020;9(12). Available from: https://doi.org/10.3390/land9120487.
13) Soropa G, Mbisva OM, Nyamangara J, Nyakatawa EZ, Nyapwere N, Lark RM. Spatial variability and mapping of soil fertility status in a high potential smallholder farming area under sub humid conditions in Zimbabwe. *SN Applied Sciences*. 2021;3(396). Available from: https://link.springer.com/article/10.1007/s42452-021-04367-0.
14) Makungwe M, Chabala LM, Chishala BH, Lark RM. Performance of linear mixed models and random forests for spatial prediction of soil pH. *Geoderma*. 2021;397. Available from: https://doi.org/10.1016/j.geoderma.2021.115079.
15) Bao Y, Ustin S, Meng X, Zhang X, Guan H, Qi B, et al. A regional-scale hyperspectral prediction model of soil organic carbon considering geomorphic features. *Geoderma*. 2021;403. Available from: https://doi.org/10.1016/j.geoderma.2021.115263.
16) Chen F, Zhang C, Zhang J, Cao W. IDP: An Intelligent Data Prediction Scheme Based on Big Data and Smart Service for Soil Heavy Metal Content Prediction. *IEEE Access*. 2021;9:32351–32367. Available from: https://ieeexplore.ieee.org/document/9358137.
17) Pant H, Lohani MC, Bhatt A, Pant J, Singh MK. Rule Descriptions for Soil Quality and Soil Fertility Assessment using Fuzzy Control System". *International Journal of Recent Technology and Engineering (IJRTE)*. 2020;8(6):1341–1346. Available from: https://www.ijrte.org/wp-content/uploads/papers/v8i6/F7663038620.pdf.
18) Pant H, Lohani MC, Bhatt A, Pant J, Joshi A. Soil Quality Analysis and Fertility Assessment to Improve the Prediction Accuracy using Machine Learning Approach. *International Journal of Advanced Science and Technology*. 2020;29(3). Available from: http://sersc.org/journals/index.php/IJAST/article/view/27039.
19) ml JT, Chambers O. Machine Learning Strategy for Soil Nutrients Prediction Using Spectroscopic Method. *Sensors*. 2021;21(12). Available from: https://dx.doi.org/10.3390/s21124208.
20) Santana EJ, dos Santos FR, Mastelini SM, Melquiades FL, Barbon S. Improved prediction of soil properties with multi-target stacked generalisation on EDXRF spectra. *Chemometrics and Intelligent Laboratory Systems*. 2021;209. Available from: https://doi.org/10.1016/j.chemolab.2020.104231.
21) M S, D JC, Lingappa M. 1D convolutional neural networks-based soil fertility classification and fertilizer prescription. *Ecological Informatics*. 2023;78. Available from: https://doi.org/10.1016/j.ecoinf.2023.102295.
22) Ramzan Z, Asif HMS, Yousuf I, Shahbaz M. A Multimodal Data Fusion and Deep Neural Networks Based Technique for Tea Yield Estimation in Pakistan Using Satellite Imagery. *IEEE Access*. 2023;11:42578–42594. Available from: https://dx.doi.org/10.1109/access.2023.3271410.
23) Nejad SMM, Abbasi-Moghadam D, Sharifi A. ConvLSTM–ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2024;17:17489–17502. Available from: https://dx.doi.org/10.1109/jstars.2024.3464411.
24) Valero A, Palacino B, Ascaso S, Valero A. Exergy assessment of topsoil fertility. *Ecological Modelling*. 2022;464. Available from: https://dx.doi.org/10.1016/j.ecolmodel.2021.109802.
25) Subramaniam LK, Marimuthu R. Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*. 2024;8. Available from: https://dx.doi.org/10.1016/j.prime.2024.100611.
26) Mehla A, Deora SS, Dalal S. Improving Crop Yield Prediction Models with Optimization-Based Feature Selection and Filtering Approaches. *Indian Journal Of Science And Technology*. 2023;16(47):4512–4524. Available from: https://dx.doi.org/10.17485/ijst/v16i47.1602.
27) Kalaiarasi E, Anbarasi A. Crop yield prediction using multi-parametric deep neural networks. *Indian Journal of Science and Technology*. 2021;14(2):131–140. Available from: https://indjst.org/articles/crop-yield-prediction-using-multi-parametric-deep-neural-networks.
28) Wang L, Chen Z, Liu W, Huang H. A Temporal–Geospatial Deep Learning Framework for Crop Yield Prediction. *Electronics*. 2024;13(21). Available from: https://dx.doi.org/10.3390/electronics13214273.