ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

SOIL CLASSIFICATION USING MACHINE LEARNING FOR CROP SUGGESTION

S. Sakthi Saranya
M.Phil Research Scholar, Department of Information Technology, Bharathiar University
Coimbatore – 641046, saisakthisaranya@gmail.com

Dr. W. Rose Varuna
Assistant Professor, Department of Information Technology, Bharathiar University
Coimbatore – 641046, rosevaruna@buc.edu.in

Abstract: Soil plays a prominent role and the appropriate soil selection for a crop is a significant aspect of agriculture. Progression of agriculture and yield can be increased with the assistance of computational algorithms. *Machine Learning (ML)* provides a pivotal tool for *Decision Support System (DSS)* in soil classification and recommendation systems. Numerous, *ML* algorithms have been deployed in the classification of soil and recommendation of crops. In this research work, an efficient *Gradient Boosted Tree (GBT)* is utilized in feature selection and classification is accomplished using *Feed Forward Neural Network (FFNN)*. This work considers ten different soils and an appropriate crop is recommended for soil. The framework of *GBT* is efficient and robust in handling data with nonlinear relationships. The *FFNN's* scalability, adaptability, and interpretability offer the highest accuracy while classifying complex data. Finally, the *Soil-Crop suitable Matrix (SCM)* is constructed to map the suitable crop source for appropriate soil, whereas the scores are generated for distinct crops for a specific type of soil. The performance of the proposed *Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF)* is compared with existing *state-of-art* techniques, where the *ML-SCCRF* outperforms the existing frameworks.

Keywords: Machine learning, accuracy, soil classification, feature selection, crop recommendation, and matrix.

1. Introduction

Agriculture is the science and art of farming that comprises efforts to harvest crops and nurture livestock. It is the foundation of food security and the backbone of human society. It also makes an essential contribution to income development and poverty reduction in developing countries by providing food, employment etc. Food production is the main reason for agricultural activities such as cultivating plants, and other existing forms of food and these products, which are used to prolong and improve human life [1, 2]. The main tasks encompassed in agricultural statistics are gathering processing, and investigation of statistical data that characterizes the current status, development of agriculture, and the fulfilment of production plans. The processed statistical data is utilised in describing the long-range and annual goals for production in agriculture [3].

Agricultural data has provided farmers with rough information on water cycles, rainfall, fertilizer necessities, etc., Agricultural data can allow farmers to make precise decisions about crop cultivation for improved productivity. It is primarily used to maximize land yields and ensure the best possible use of land. Weather forecast data can be used to provide advice on planting, crop management and harvest schedules. Dimension data supports the farmers in improving the processes with a tremendous amount of information. Additionally, they make decisions that can be adapted to the farm‟s specific needs. The information obtained helps the farmers identify efficiencies that lead to high productivity, profitability, lower input costs, and optimized fertilizer use.

Agriculture plays a significant role in the economic growth of a country. Agriculture offers food and raw ingredients, whereas provides employment opportunities for a large population. Most of the population entirely depends on agriculture as a food source for livelihood [4]. Crop selection is generally made with the farmers‟ experience and knowledge. However, this manual suggestion might not be favourable for considering factors like climate change, drought, lack of monsoon and floods. The cultivation of crops is highly dependent on the weather and geographical outline of the land [5, 6].

Soil nutrient levels vary from region to region, so frequent plantings in the same soil affect the nutrients and plant development. Several crops are destroyed each year owing to a lack of technical knowledge and random weather conditions that are essential in determining yields and profits [7]. Based on the soil and climatic information, the farmers possess the complexity to predict weather and crops. Identifying the best crops for a specified land can support the farmers in deciding on crop cultivation [8, 9].

Agricultural production has not been reliant only on natural factors rather it could also be based on the inputs specified. Farming productivity depends on soil type, type of fertilizers, accessibility of water, crop type, humidity, temperature, weather conditions, etc., These agricultural parameters contain a huge potential to augment or reduce crop productivity. With the implementation of emerging technologies, conventional farming can be replaced by precision farming which increases productivity. It recommends the farmers concerning the best crop to cultivate on the farms for the highest possible yield and profit [10]. Appropriate crop selection must be made based on geographical and climatic conditions to increase productivity. Due to the infinite subdivision of land holdings, the size of the land holdings is reducing day by day.

Farmers cannot make consistent progress in agriculture because rainfall is unpredictable. The automated classification of soil and relevant crops can make considerable progress in agriculture [11]. Crops are classified into three types based on the period of cultivation. The Kharif crops are cultivated in the monsoon season. The Kharif season is usually from June to September. Paddy, ragi, maize, groundnut, and cotton are Kharif crops. Rabi season runs from November to March or April. Rabi crops include green peas, pulses, and wheat. The summer crop grows in the summer season, which starts from March to July. Therefore, summer crops are those that are grown during this time [12]. These crops are grown on irrigated lands and do not depend on the monsoons. Summer crops include green gram and black gram [13].

The kharif crops namely ragi (finger millet), cotton, paddy (rice), groundnut, and maize in distinct variety of soil that is suitable for the development. The paddy necessitates clayey soil with well-drained texture with the retention ability of good amount of water. The ragi can be grown in the

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

sandy loam or red soil with the suitable drainage facility and modest fertility in the farm. Maize is harvested in most fertile natured soil likely sandy loam or loamy and groundnut crop desires well-drained soil called sandy loam with efficient retention of moisture ability [14]. Cotton expects good moisture holding ability of soil and necessitates sandy loam soil with well drainable nature. The crops such as rabi, and wheat prefers appropriate soil. Summer crops such as green gram (mung bean) and black gram (urad bean) thrive in well-drained sandy loam or loamy soils. While these are general guidelines, soil fertility, pH levels, and local conditions should always be considered for optimal crop growth and yield [15].

The yield maximisation in agriculture is attained by the soil classification and crop recommendation system. Significantly, *Machine Learning (ML)* is essential for agricultural *Decision Support Systems (DSS)*. The proposed *Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF)* is encompassed with *Gradient Boosted Tree (GBT)* for feature selection, *Feed Forward Neural Network (FFNN)* for soil classification, and *Soil-Crop suitable Matrix (SCM)* for crop recommendation system. The proposed *ML-SCCRF* optimizes the compatibility of the soil-crop and makes advancements in the practice of agriculture.

The remainder of the article is organized as follows: The overview of agriculture, the necessity of soil classification, and recommendations are detailed in the *Introduction* section. The recent research in soil classification and recommendation systems is elaborated with a research gap in the *Related Works* section, the proposed *Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF),* which is a combination of *Gradient Boosted Tree (GBT)* for feature selection, *Feed Forward Neural Network (FFNN)* for soil classification, and *Soil-Crop suitable Matrix (SCM)* for crop recommendation system is detailed in *Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF)* section, the attained outcome throughout this research framework is depicted with discussion in *Result and Discussion* section, and the article is concluded with a suggestion for future work in *Conclusion and Future Work* section.

2. Related Works

The research in paper [16] highlights the significance of agriculture in the economy of India and plays prominent role in the productivity of crop. It necessitates the accurate forecasting of soil variety for management of weed and optimized selection of crop. This research work presents an advanced technique for deploying *Deep Learning (DL)* and *Machine Learning (ML)* techniques for multiclass classification of soil. *Artificial Intelligence (AI)* approaches namely the selection of *Q-HOG* feature and multi-stacking ensemble framework. By assessing numerous *DL* framework and traditional *ML* frameworks, the research attains effective accuracy rates of *98.96%* in soil classification.

The second paper [17] emphases on emerging a cost-effective soil classification system for countryside farmers. It highlights the importance of texture of soil in the selection of crop and properties involved in transmission of water. The study indents to address the restrictions of conventional soil investigation methods like the hydrometer test that is time-consuming and

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

expensive. By accumulating the samples of soil from West Guwahati, Assam, and deploying image processing approaches, the research seeks to generate an available classification system. Utilizing feature extraction techniques and *Multi-Class Support Vector Machine (MC-SVM)* classification, the proposed technique attains an average accuracy of *91.37%*, offering farmers with a consistent tool for informed selection crop based on characteristics soil.

Suruliandi et al. (2021) [18] made a comparative study of different attribute selection methods for predicting crops using different classifiers by considering soil properties and environmental factors such as season, precipitation, temperature and texture. Khaki et al. (2020) have proposed a *Deep Learning (DL)* structure by applying the *Recurrent Neural Network (RNN)* and *Convolutional Neural Networks (CNNs)* to forecast crop yield according to management practices and environmental data. In the *CNN-RNN* model, *Deep Fully Connected Neural Networks (DFNN)* and *Random Forest(RF)* was utilized for corn and soya bean yield prediction.

Guo et al. (2020) [19] examined eleven climate, environment, and geography combinations for the site-based rice yield prediction by applying *Multi Linear Regression (MLR)* and *ML* techniques such as *Back Propagation (BP)*, *RF*, and *Support Vector Machine (SVM)*. Experimental results of this study have demonstrated that *ML* methods have given more precision than the *MLR* technique. Furthermore, the results showed that *SVM* had attained the maximum precision and accuracy of yield by considering the environment and climatic variables.

Farhat Abbas et al. (2020) [20] have utilized the four ML algorithms as *K-Nearest Neighbors (KNN),Logistic Regression (LR),Elastic Net (EN), Support Vector Regression (SVR)* for the potato (Solanum tuberosum) tuber yield prediction according to the data of soil and crop properties composed through proximal sensing. The modelling methods have been exploited with several statistical parameters for yield prediction estimation.

Bhimanpallewar et al. (2020) [21] discussed *ML* methods for the agriculture advisory system. An agriculture advisory system was developed, which takes soil and environmental parameters as input using the *C4.5 Decision Tree (DT)* algorithm.

Mariappan et al. (2020) [22] presented a model which analyzed various soil parameters and suggested crops for cultivation. It utilizes *KNN* to map soil and crop data and predict suitable crops for cultivation. The results showed that *SVM* had attained the maximum precision and accuracy of yield by considering soil properties. Kalimuthu et al. (2020) proposed a *Bayesian Algorithm* without any boosting technique to recommend suitable crops, assuming soil temperature, humidity, and water content as crucial factors. The developed model achieved better accuracy in forecasting the suitable crop.

Jenskie et al. (2020) [23] have enhanced a *Fuzzy Logic-based Method* in which a suitable amount of fertilizer has been provided to the soil. The soil nutrients and season might directly influence the growth and yield of the crop. Lack of soil nutrient levels might result from planting infection, whereas employing an unnecessary amount of soil fertilizer affects crop enhancement. Four types of fertilizer recommendations were proposed to resolve the issue: Urea, Complete, Muriate of Potash, and Sophos. The grouping and amount of these fertilizers are executed according to the

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

fuzzy rules and input parameters. These kinds of soil fertilizer recommendations have been exploited for rice in congenital light soil.

Ashok (2020) [24] presented a crop recommendation system using holistic regression-based models by employing various ML algorithms using environmental factors and yield prediction for significant crops such as rice, ragi, gram, potato and onion.

The research presented in [16] and [17] primarily focuses on soil classification and crop selection independently. However, there exists a research gap in integrating these two aspects into a unified framework for holistic agricultural decision-making. While [16] emphasizes soil classification accuracy using advanced machine learning techniques, and [17] addresses the need for cost-effective soil classification for farmers, there's an opportunity to combine these efforts into a comprehensive system that offers precise soil classification alongside informed crop recommendations, bridging the gap between soil analysis and crop selection in agricultural decision-making. The proposed work integrates soil classification accuracy with informed crop recommendations, offering a comprehensive framework for agricultural decision-making, and bridging the gap between soil analysis and crop selection.

3. Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF)

This section investigates into the proposed *Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF)*, which integrates *Gradient Boosted Tree (GBT)* for effective feature selection, *Feed Forward Neural Network (FFNN)* for enriched soil classification, and *Soil-Crop suitable Matrix (SCM)* for appropriate crop recommendation system. The proposed methodology for *ML-SCCRF* is depicted in Figure 1.

Pre-Processing of Soil Image

Pre-processing is an crucial stage in image examination intended at refining the image quality before further analysis of image. It includes numerous consecutive steps to enhance the soil images indicating diverse soil types. Initially, the original images are acquired from the publicly available source [25]. The acquired soil images are converted to grayscale to simplify processing and then contrast enhancement is applied to enhance visual quality, followed by the deployment of a median filter to minimize noise and smoothens the soil images. These pre-processing steps are significant for preparing input data and ensuring that subsequent classification models can effectively analyze as well as classify soil types. Each step contributes to refining image quality and it subsequently helps in accurate classification.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

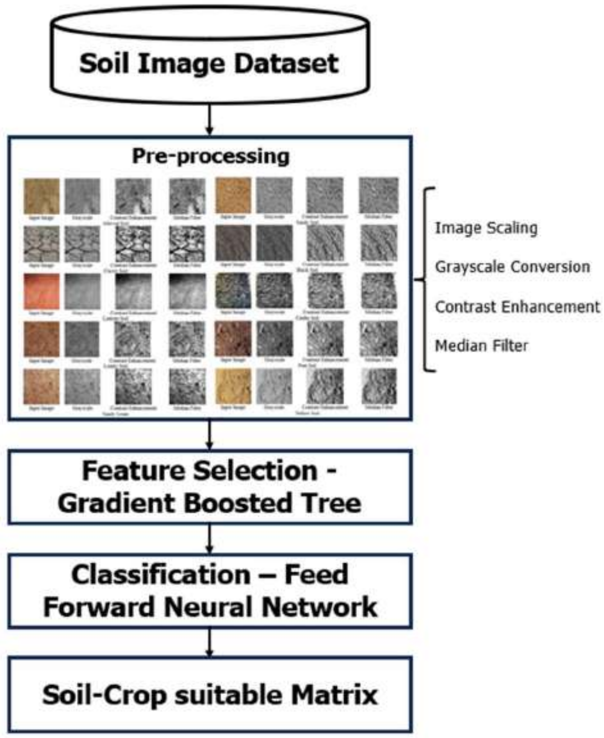


Figure 1. Overall Methodology

Feature Selection using Gradient Boosted Tree

Feature selection involves identifying relevant features from a dataset to improve model performance and reduce complexity. Gradient boosted trees efficient in this task, which captures complex relationships effectively. In soil classification, texture features like the *HSV* histogram highlight color distribution, while color auto correlogram and color moments capture spatial correlations and color similarities. Gabor features analyze texture patterns using wavelets, and discrete wavelet transform decomposes images into frequency bands, enhancing texture characterization. Integrating these diverse features offers a comprehensive understanding of soil textures, facilitating accurate classification and crop recommendation. This approach optimizes model efficiency and ensures informed agricultural decisions based on soil characteristics. Boosting, exemplified by the "*AdaBoost*" method, uplifts weaker learners through sequential re-weighting of training data, refining classifiers iteratively. Misclassified instances are emphasized, guiding subsequent classifiers to focus on challenging samples. *XGBoost*, or *eXtreme Gradient Boosting*, stands as a versatile and efficient algorithm supporting various objectives like regression and classification. It ensures model regularization, preventing overfitting, and supports parallelization for faster training. By strategically pausing tree construction and adjusting sample weights, *XGBoost* optimizes learning, making it a prevalent choice in machine learning applications. The paradigms of the standard gradient boosting are given below,

$$F_m(x) = \sum_{i=1}^{m} \beta_i f_i(x) \text{---------}(1)$$

where the base learner proportion is signified as m and f act as a base learner, β act as a coefficient

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

factor. The weak learner is indicated by $f_i(x)$ and total learners in GBT is $F_m(x)$.

The main objective is to attain an efficient and generic whole model with minimized loss functional value and is given in Equation 2.

$$F_m(x) \arg min \sum_{i=1}^{n} L(y_i F_m(x)) = \beta_m(x) \arg min \sum_{i=1}^{n} L(y_i, F_{m-1}(x) + \beta_m F_m(x)) \text{----(2)}$$

Greedy method is utilized by gradient lifting and initially F is a constant value, where the value progressively increases the performance. The loss function is minimized by the gradient boosting and feasible by assigning negative gradient's loss function in each step. The new learner is given with the Equation 3.

$$\beta_m F_m(x) = \gamma \frac{\partial L(y_i, F_{m-1}(x)}{\partial F_{m-1}(x)} \text{---------(3)}$$

where size is signified by γ, which includes the negative gradient and negative value with β parameter. The Equation 3 eliminates the pseudo residual or gradient is generally indicated as *γ*.

$$R_{im} = \gamma \frac{\partial L(y_i, F_{m-1}(x)}{\partial F_{m-1}(x)} \text{---------(4)}$$

From the Equation 4, the value of *F* is acquired through training samples and the pseudo residual value is acquired. The current learner is trained and determined with the assistance of base learner. In the training space, the appropriate values are determined by *x* and *y*. The value of base learner is determined from Equation 4, minimises the loss function and acquires the value of *γ*.

$$F_m(x) = F_{m-1}(x) + \beta_m f_i(x) \text{---------(5)}$$

The process of regularisation is enriched by the target learning rate *L* in *Xgboost*, which is based on the traditional framework of gradient boosting. The value of *L* is estimated by Equation 6.

$$L = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \text{---------(6)}$$

The first term is a distinguishable convex loss function, which computes the variation among the target and projected values. The objective function Obj and the optimal leaf node of fraction w in XGBoost are identical to w that is given in Equation 7 and 8.

$$objec = -\frac{1}{2} \sum_{j-1}^{T} \frac{G_j}{H_j + \lambda} + \gamma T \text{---------(7)}$$

$$w_i^* = \frac{G_j}{H_j + \lambda} \text{---------(8)}$$

Texture features like the HSV histogram highlight color distribution, while color auto correlogram and color moments capture spatial correlations and color similarities. Gabor features analyze texture patterns using wavelets, and discrete wavelet transform decomposes images into frequency bands, enhancing texture characterization. Diverse machine learning frameworks are appropriate for various tasks. *XGBoost*, based on a tree model, handles image data effectively. *XGBoost* offers advantages including model input data invariance, interpretability, and simplicity of parameter adjustment, making it suitable for low-dimensional data. The procedure involved in feature selection is given in Algorithm 1.

Algorithm 1. Feature Selection using GBT
Procedure GradientBoostedTrees(Input: Training data (X, y), Number of boosting rounds (M)) Initialize F(x) as a constant value For m = 1 to M: Compute the negative gradient of the loss function R_{im} for each training sample Train a base learner $f_{i(x)}$ to minimize the loss function using the negative gradients Compute the optimal coefficient β_m using a line search or other optimization technique Update the ensemble model: $F_{m(x)} = F_{(m-1)}(x) + \beta_m * f_{i(x)}$ Update the residuals: $R_{im} = -\partial L(y_i, F_{(m-1)}(x)) / \partial F_{\_(m-1)}(x)$ End for Return F(x) End Procedure

Classification – Feed Forward Neural Network

The features acquired from the GBT is utilized in classification with Feed Forward Neural Network (FFNN) that is efficient in pattern classification and independent from distribution of class respective of knowledge. The whole network is trained as a batch and it utilises one generic hidden layer framework. The FFNN framework comprises of input, hidden, and output layers that is interconnected by the links comprises of weights. In the training process, weights in the network are updated iteratively in four process that is initiated with output estimation in the hidden layer. The entire process concerns the optimization process and regulate the weights for exact classification using FFNN.

$$k_j = af_H\left[\sum_{i=1}^{N_j} \omega_i(i,j) l_i\right] \quad j = 1,2,3,\dots\dots,N_H \text{---------(9)}$$

where the i^{th} neuron value in the input layer is signified as l_i, j^{th} neuron value in the hidden layer is signified as k_j, and the activation function (sigmoid) is signified as af_H. The activation function, sigmoid estimation is given as in Equation 10.

$$af_H(l) = \frac{1}{1+exp(-l)} \text{---------(10)}$$

The output layer comprises neuron with output information is given in Equation 11.

$$OP_k = af_{op}\left[\sum_{j=1}^{N_H} \omega_2(j,m) y_j\right] \quad m = 1,2,3,\dots\dots,N_o \text{---------(11)}$$

The activation function in output layer is signified as af_o and random value is assigned for every weight, which is modified by the delta based rule and the its relevant to the learning samples. Incidence of loss rate or error in the network is signified as the variation amongst output and

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

target value is given in Equation 12.

$$ER_i = error\left(\sum_{m=n}^{N_o}(OP_k - T_k)\right) \qquad n = 1,2,3, \dots \dots, N_s \text{---------(12)}$$

where the authentic value in k^{th} neuron is signified as T_k, which is identified as soils and sample count as N_s. The FFNNs fitness value is calculated as the average of error rate and it is given in Equation 13.

$$af(\omega) = \sum_{l=1}^{N_s} ER_i \text{---------(13)}$$

The main goal of *FFNN* is to reduce the occurrence of error rate and it impacts the feature vector that is considered as an appropriate target values. The feature vectors attained from *GBT* is trained with the aid of *FFNN*. Diverse soils are categorized with minimal occurrence of error and the categorized soils are mapped with crop.

Soil-Crop suitable Matrix (SCM)

The process involves generating a soil-crop suitability matrix derived from the FFNN-classified soil types. This matrix delineates the compatibility between various soil types and crop species, drawing upon agronomic expertise, research insights, and past crop performance records. Each cell in the matrix symbolizes the suitability score for a particular crop grown in a specific soil type. By structuring the matrix with soil types as rows and crop species as columns, suitability scores are assigned to populate the cells. These scores, grounded in compatibility assessments, illuminate the degree of suitability for each crop within the context of distinct soil types.

Algorithm 2. Soil-Crop suitable Matrix (SCM)

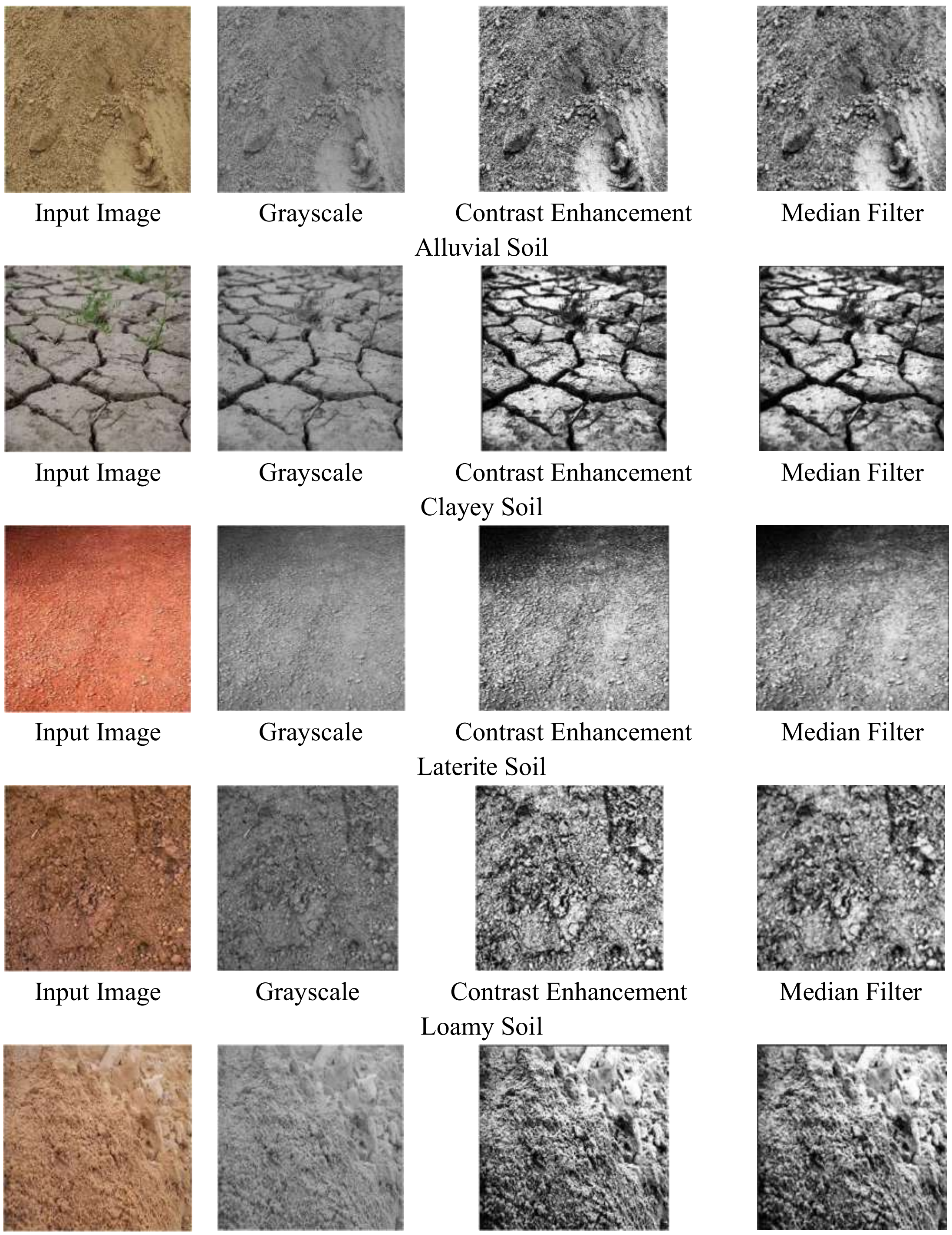
Procedure CreateSoilCropSuitabilityMatrix(Input: Classified soil types, Crop species)
  Initialize an empty matrix M with dimensions (number of soil types) x (number of crop species)
  For each soil type s:
    For each crop species c:
      Determine the suitability score or rating of crop c for soil type s based on agronomic knowledge, research findings, and historical crop performance data
      Assign the suitability score to cell (s, c) in matrix M
  End for
  Return M
End Procedure

4. Result and Discussion

The dataset provided has been compiled with the best knowledge of the Indian Council of Agricultural Research (ICAR) and is available for public use in both public and private endeavors. It serves the purpose of facilitating precision agriculture by enabling informed decision-making for farmers. Augmented from various sources including rainfall, climate, and fertilizer data, the dataset encompasses parameters such as nitrogen (N), phosphorous (P), potassium (K) ratios in

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

soil, temperature, humidity, pH value, and rainfall. These factors are crucial in determining the suitability of a farm for various crops. Additionally, the dataset includes information on electrical conductivity, sulphur, copper, iron, manganese, zinc, boron, and crop types. Furthermore, the dataset includes a collection of soil images categorized into distinct soil types, aiding in soil classification and analysis tasks. It serves as a valuable resource for beginners in machine learning, providing an opportunity to develop predictive models for recommending crop types based on soil parameters [25]. The pre-processing of soil images is given in Figure 2.



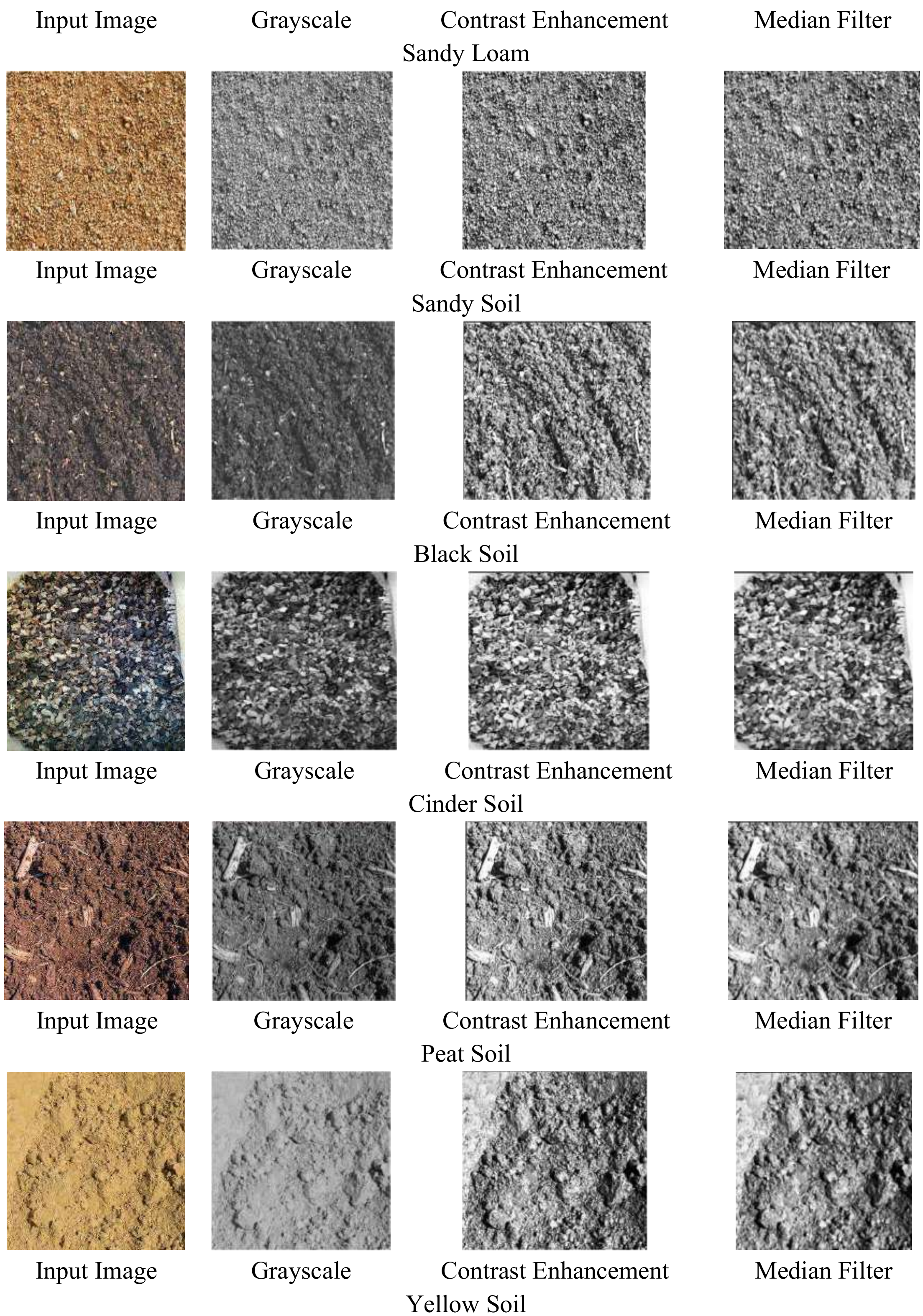


Figure 2. Preprocessing of Soil Images

These pre-processing steps are essential in preparing soil images for further analysis and

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

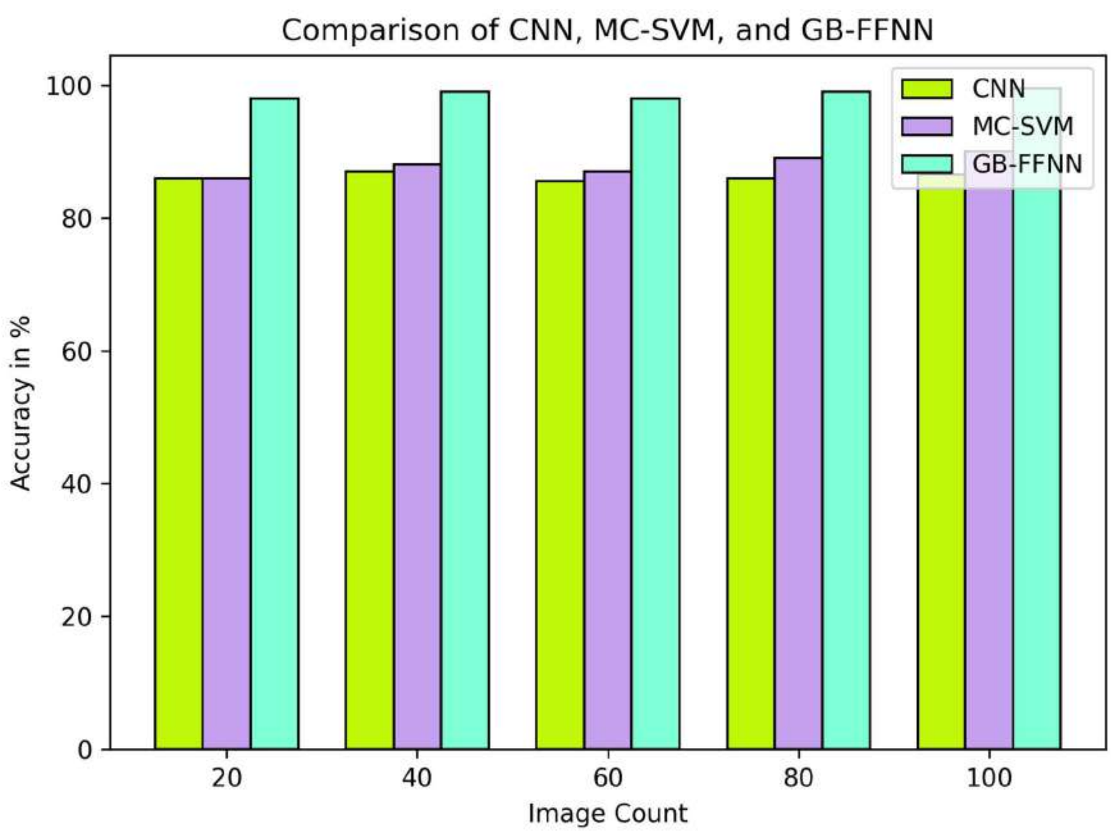
classification tasks, as they help to standardize the images and enhance their features, ultimately facilitating accurate identification and characterization of various soil types. By systematically applying these transformations, researchers can ensure consistency and reliability in the analysis of soil imagery, leading to more robust and insightful findings in the field of soil science and agriculture.

Accuracy: It is the measure of the amount of the exactly categorized data amongst every instance in the entire dataset. Accuracy estimated the count of appropriately forecasted instances to the whole count of instances. The accuracy is estimated using Equation 14. The experimental value is given in Table 1 and depicted in Figure 3.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{---------(14)}$$

Table 1. Comparison of Accuracy

Image Count	CNN	MC-SVM	GB-FFNN
20	86	86	98
40	87	88	99
60	85.5	87	98
80	86	89	99
100	86.5	90	99.5



ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

Figure 3. Comparison of Accuracy

Recall: The ability of the classifier to measure and forecast the appropriate values is determined as *True Positive (TP)* rate. It is estimated as the proportion of appropriately forecasted positive values to the total count of definite positive values. The recall is estimated using Equation 15. The experimental value is given in Table 2 and depicted in Figure 4.

$$Recall = \frac{TP+TN}{TP+FN} \text{---------}(15)$$

Table 2. Comparison of Recall

Image Count	CNN	MC-SVM	GB-FFNN
20	84	93	96.5
40	85	94	97
60	79	89	94
80	84	90.5	95.5
100	85	91	95

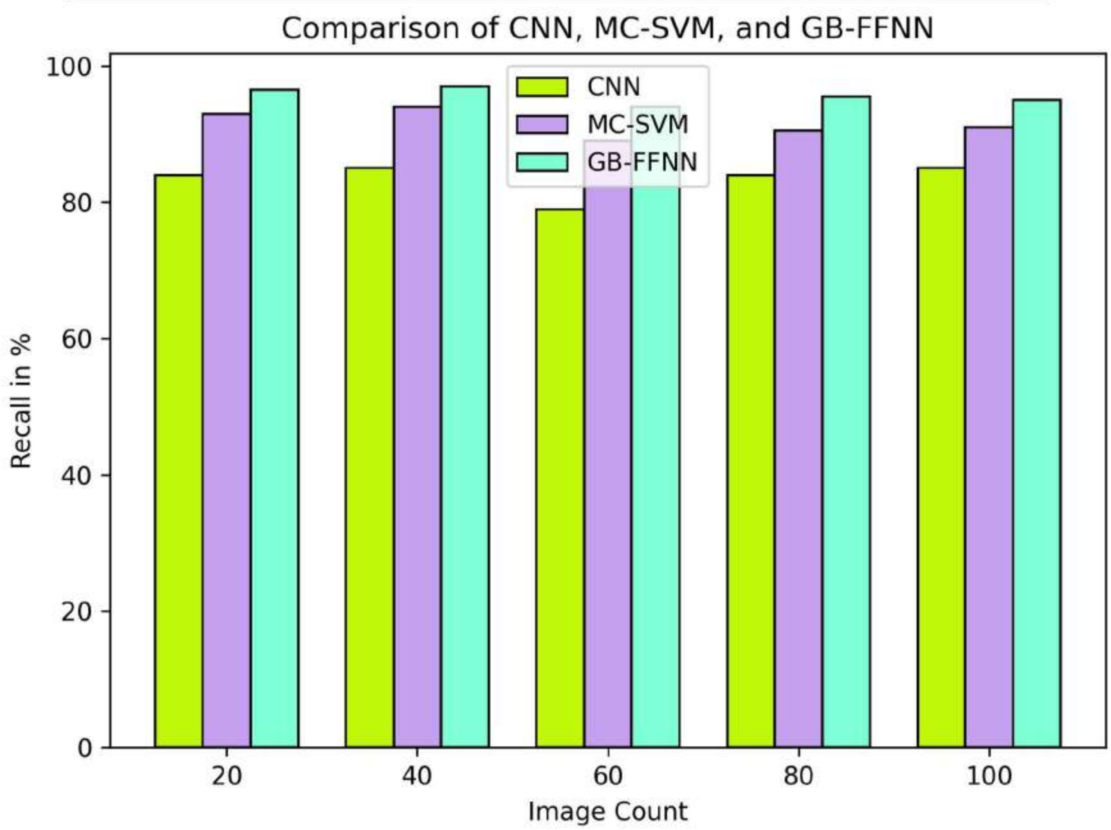


Figure 4. Comparison of Recall

F-measure: It integrates the outcome of recall and precision into one metric that offers a stability amongst the performance metrics. The harmonic mean or F-measure is estimated using Equation 16. The experimental value is given in Table 3 and depicted in Figure 5.

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \text{---------}(16)$$

Table 3. Comparison of F-Measure

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

Image Count	CNN	MC-SVM	GB-FFNN
20	82	93.5	97.5
40	84	94.5	98
60	84	91	94
80	84.5	92.5	95.5
100	86	91	96

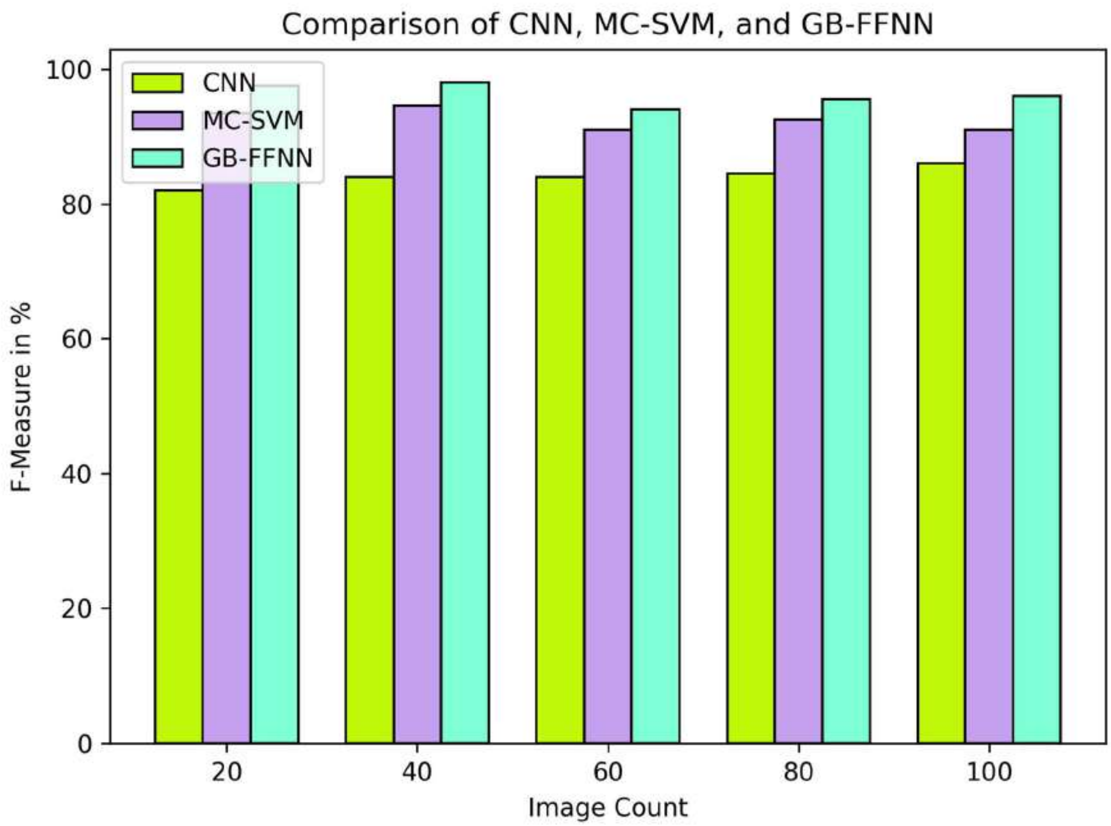


Figure 5. Comparison of F-Measure

Precision estimates the ratio of exactly categorized positive values amongst the all values forecasted as positive by the classifier. It is estimated as the proportion of *TP* to the summation of *TP* and *False Positives (FP)*. The precision is estimated using Equation 17. The experimental value is given in Table 4 and depicted in Figure 6.

$$Precision = \frac{TP}{TP+F} \text{---------(17)}$$

Table 4. Comparison of Precision

Image Count	CNN	MC-SVM	GB-FFNN
20	82	93.5	96.5
40	83	94.5	97
60	84.5	92	95

80	84.5	93.5	96.5
100	87	92	97

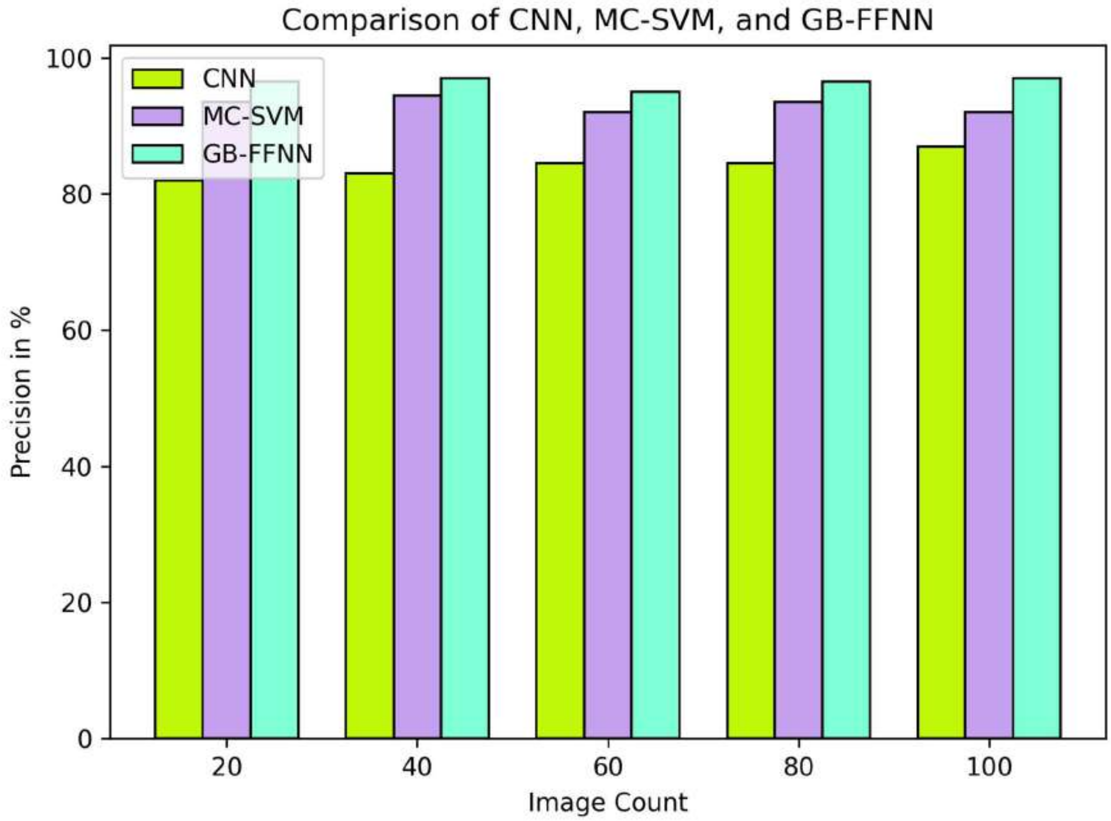


Figure 6. Comparison of Precision

The misclassification in the soil classification is noted as error rate that quantifies the ratio of wrongly classified instances in the soil dataset. This is estimated as the proportion of wrongly forecasted values to the whole count of values. The experimental value is given in Table 5 and depicted in Figure 7.

Table 5. Comparison of Error Rate

Image Count	CNN	MC-SVM	GB-FFNN
20	0.5765	0.5242	0.4462
40	0.5827	0.5213	0.4698
60	0.5928	0.5471	0.4887
80	0.6456	0.5821	0.4902
100	0.6872	0.6252	0.5135

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

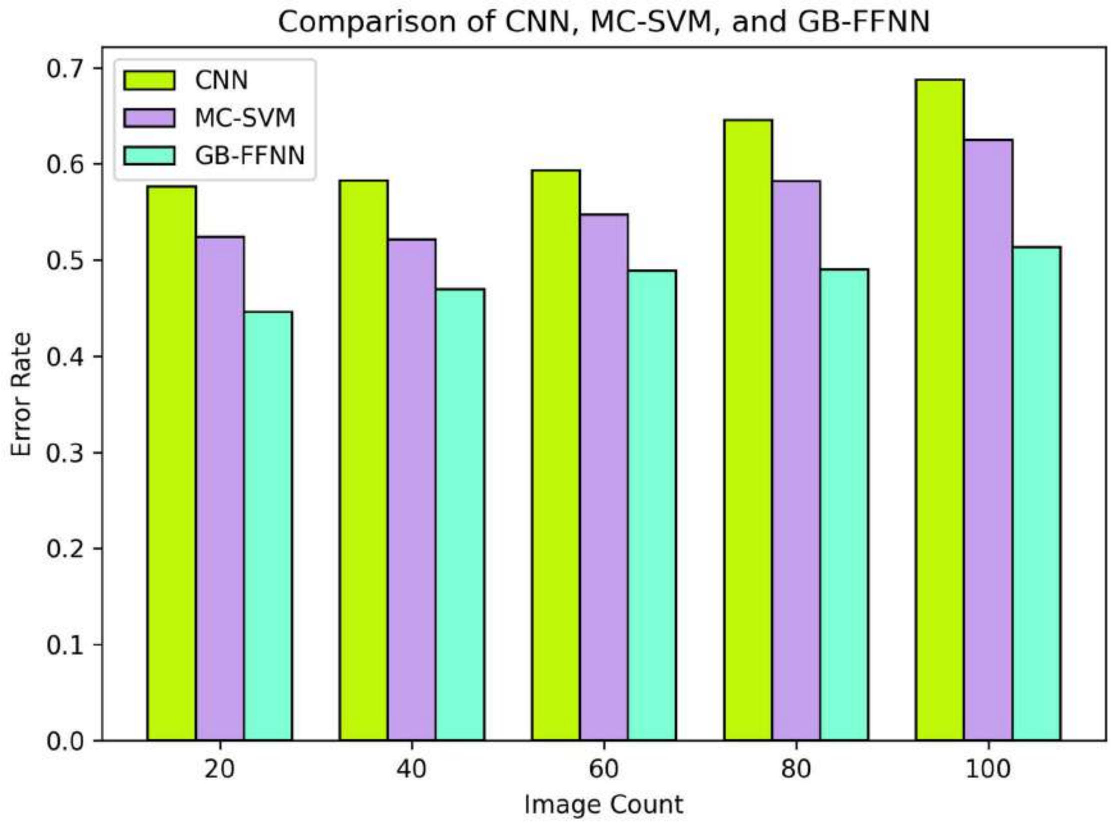


Figure 7. Comparison of Error Rate

The comparative analysis of Accuracy, Recall, F-Measure, and Precision provides valuable insights into the performance of existing techniques - CNN, MC-SVM, and the proposed GB-FFNN - in soil image analysis. Accuracy, as a representation of systematic mistakes and statistical bias, is crucial for assessing the reliability of classification models. The analysis compares the performance of three different classifiers, namely Convolutional Neural Network (CNN), Multi-Class Support Vector Machine (MC-SVM), and Gradient Boosted Feed Forward Neural Network (GB-FFNN), across various accuracy metrics. Accuracy, which measures the proportion of correctly classified instances, indicates that GB-FFNN consistently outperforms CNN and MC-SVM, achieving the highest accuracy across different image counts, reaching up to 99.5% accuracy with 100 images. Recall, evaluating the classifier's ability to identify all relevant instances, also demonstrates GB-FFNN's superiority, with higher recall scores compared to CNN and MC-SVM. Similarly, F-measure, combining precision and recall, highlights GB-FFNN's balanced performance, exhibiting the highest F1-score across different image counts. Precision, indicating the proportion of correctly classified positive instances, showcases GB-FFNN's precision surpassing that of CNN and MC-SVM. Lastly, the error rate, quantifying misclassified instances, reveals that GB-FFNN consistently exhibits the lowest error rate, indicating its superior classification performance compared to CNN and MC-SVM across varying image counts. Overall, these findings underscore the effectiveness of GB-FFNN in accurately classifying instances in the dataset, making it a promising choice for classification tasks.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

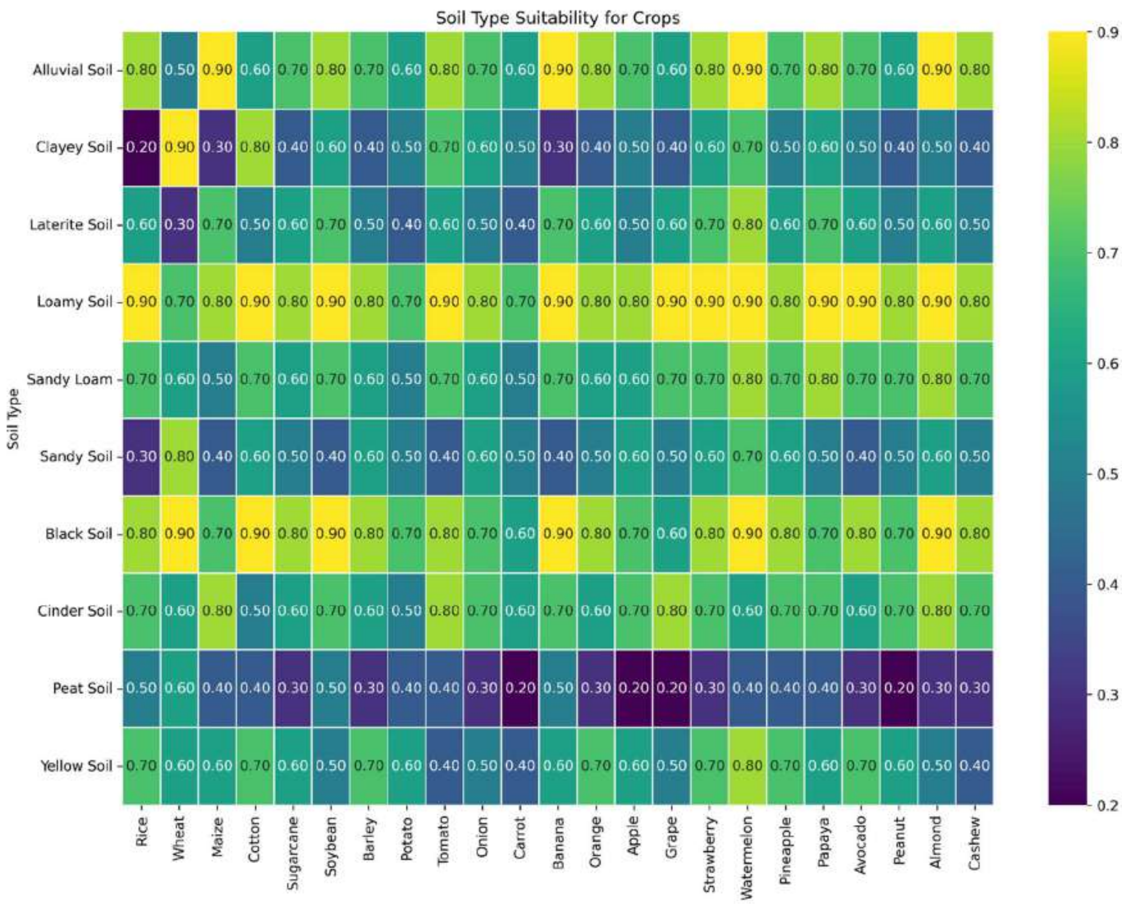


Figure 8. Soil Crop suitability Matrix

The suitability of various crops for different soil types, we first calculated the average suitability scores for each crop across all soil types. For instance, for Alluvial Soil, the average suitability scores for different crops were as follows: Rice (0.8), Wheat (0.5), Maize (0.9), Cotton (0.6), Sugarcane (0.7), Soybean (0.8), Barley (0.7), Potato (0.6), Tomato (0.8), Onion (0.7), Carrot (0.6), Banana (0.9), Orange (0.8), Apple (0.7), Grape (0.6), Strawberry (0.8), Watermelon (0.9), Pineapple (0.7), Papaya (0.8), Avocado (0.7), Peanut (0.6), Almond (0.9), Cashew (0.8). By comparing these scores across soil types, we can determine the crops that are generally more suitable for specific soil types and identify any patterns or trends in crop suitability.

5. Conclusion

Agriculture is a vital sector that contributes considerably to the economy of any country. However, in recent years, the agricultural industry has been gradually diminishing due to economic and environmental conditions and other problems in the farming community. Therefore, various recommendation systems using ML have been implemented to utilize the latest technical insights in promoting the agricultural sector. An efficient recommendation model is developed in this work by considering the factors such as soil, climate and crop rotations which helps the farmers decide on crop cultivation. By cultivating the most effective crops, productivity has been increased without wasting any resources with the help of ML technology. Moreover, choosing suitable crops helps in enhancing economic and environmental sustainability.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

Computational algorithms, particularly Machine Learning (ML), offer significant support through Decision Support Systems (DSS) in this domain. This research leverages the efficiency of Gradient Boosted Trees (GBT) for feature selection and employs Feed Forward Neural Network (FFNN) for soil classification. A Soil-Crop suitable Matrix (SCM) is constructed to map suitable crop choices for specific soil types, with scores generated for each crop-soil combination. The proposed Machine Learning assisted Soil Classification and Crop Recommendation Framework (ML-SCCRF) outperforms existing techniques, highlighting its effectiveness in agricultural decision-making.

In future, the efficiency of soil classification and crop recommendation is enhanced by the *Deep Learning (DL)* models. The *DL* models are efficient in handling the complex and high dimensional data. Diverse soil types across the universe and diverse crop combination as well as rotation can be suggested with the recommender system. The learning of *DL* can be improvised by tuning of hyper-parameter, which is utilized in optimizing accuracy of *DL* framework and utilization of resource is optimized by swarm-based techniques. In future this work can be extended to predict the climatic factors like temperature, rainfall and etc, Implementing Hybridization in data mining techniques and Regression techniques for cultivating and predicting more crops.

Reference

1. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.
2. Akhter, R., & Sofi, S. A. (2022). Precision agriculture using IoT data analytics and machine learning. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5602-5618.
3. Reddy, K. S. P., Roopa, Y. M., LN, K. R., & Nandan, N. S. (2020, July). IoT based smart agriculture using machine learning. In *2020 Second international conference on inventive research in computing applications (ICIRCA)* (pp. 130-134). IEEE.
4. Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, *9*, 4843-4873.
5. Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010.
6. Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2023). Impact of machine learning on management, healthcare and agriculture. *Materials Today: Proceedings*, *80*, 2803-2806.
7. Suchithra, M. S., & Pai, M. L. (2020). Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. *Information processing in Agriculture*, *7*(1), 72-82.
8. Wadoux, A. M. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359.

9. Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, *81*, 401-418.
10. Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., & Scholten, T. (2020). Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sensing*, *12*(14), 2234.
11. Reddy, Y. R. Prediction Of Ground Properties Using Machine Learning Techniques. *International Journal of Current Science (IJCSPUB), ISSN*, *2250*(1770), 12-15.
12. Taghizadeh-Mehrjardi, R., Nabiollahi, K., Rasoli, L., Kerry, R., & Scholten, T. (2020). Land suitability assessment and agricultural production sustainability using machine learning models. *Agronomy*, *10*(4), 573.
13. Srivastava, P., Shukla, A., & Bansal, A. (2021). A comprehensive review on soil classification using deep learning and computer vision techniques. *Multimedia Tools and Applications*, *80*(10), 14887-14914.
14. Saranya, N., & Mythili, A. (2020). Classification of soil and crop suggestion using machine learning techniques. *International Journal of Engineering Research & Technology (IJERT)*, *9*(02), 671-673.
15. Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. *Mathematical and Computer Modelling of Dynamical Systems*, *27*(1), 117-140.
16. Padmapriya, J., & Sasilatha, T. (2023). Deep learning based multi-labelled soil classification and empirical estimation toward sustainable agriculture. *Engineering Applications of Artificial Intelligence*, *119*, 105690.
17. Barman, U., & Choudhury, R. D. (2020). Soil texture classification using multi class support vector machine. *Information processing in agriculture*, *7*(2), 318-332.
18. Thomas, D. (2023, May). A review of crop yield prediction based on Indian agriculture sector using machine learning. In *AIP Conference Proceedings* (Vol. 2773, No. 1). AIP Publishing.
19. Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., ... & Senthilnath, J. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecological Indicators*, *120*, 106935.
20. Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, *10*(7), 1046.
21. Bhimanpallewar, R. N., & Narasingarao, M. R. (2020, January). Alternative approaches of machine learning for agriculture advisory system. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 27-31). IEEE.
22. Mariappan, AK, Madhumitha, C, Nishitha, P & Nivedhitha, S 2020, „Crop recommendation system through soil analysis using classification in Machine Learning‟. International Journal of Advanced Science and Technology, vol. 29, no. 3, pp. 12738-12747.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

23. Jenskie Jerlin, Haban, I, John Carlo VPuno, Argel A Bandala, Robert Kerwin Billones, Elmer P Dadios & Edwin Sybingco 2020, „Soil fertilizer recommendation system using Fuzzy Logic‟, in IEEE Region 10 Conference (Tencon), pp.1171-1176.
24. Ashok Patidar, Chandrawanshi, SK & Neeraj Kumar 2020, „Crop yield forecasting of sorghum (sorghum bicolor l.) by using statistical technique for Tapi and Surat districts of South Gujarat, India‟, International Journal of Current Microbiology and Applied Sciences, vol. 9, no. 8, pp.3979-3985.
25. https://www.kaggle.com/code/prasanshasatpathy/soil-type-image-classification