International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 06, pp. 26-33, June

A Whale Optimization Algorithm Enabled CNN-GRU-Mamba Hybrid Framework for Precise Soil Fertility Classification Using Agricultural Soil Parameters

P. Kavitha[3]

[3]Research Scholar, Department of Electrical and Electronics Engineering,
Government College of Engineering, Tirunelveli, Tamilnadu, India.

kavitha.paulsamy6@gmail.com

ABSTRACT: Soil Fertility (SF) classification is important for sustainable agriculture, efficient nutrient management, and improved crop productivity. In this paper a Whale Optimization Algorithm-enabled Convolutional Neural Network - Gated Recurrent Unit -Mamba (WOA-CNN-GRU-Mamba) hybrid structure is proposed for precise SF classification using agricultural soil parameters. Data pre-processing based on missing-value handling and duplicate removal, data visualization based on outlier detection is used to improve data quality. Relevant soil attributes are then transformed through feature engineering and normalization to provide reliable inputs for model learning. The proposed hybrid model integrates CNN for extracting local feature patterns, GRU for learning sequential dependencies, and Mamba for capturing long-range relationships efficiently. Furthermore, the WOA is used to optimize the model parameters and improve classification performance and convergence. Experimental results is demonstrated in Python software using SDN Intrusion Detection dataset the proposed structure attains an 98% for accuracy, precision, F1-score, as well as recall of 97% ROC-AUC of 100%. The proposed WOA-CNN-GRU-Mamba framework provides an accurate and computationally efficient solution for intelligent SF assessment and support precision agriculture and sustainable soil management.

Keywords: Soil Fertility, Agricultural Soil Parameter, Whale Optimization Algorithm, Convolutional Neural Network, Gated Recurrent Unit, Mamba

1. Introduction

Healthy soil cover is important for maintaining agricultural productivity as well as supporting sustainable ecosystems [1]. However, intensive plowing, improper farming practices, excessive use of chemical fertilizers, and overuse of crop protection chemicals disturb the natural balance of the soil, slowly reducing its fertility and accelerating soil degradation [2-3]. Furthermore, accurate classification of soil assets is important for better land-management decisions as well as improving ecological crop production in exactness agriculture [4]. Soil prepared for planting, depends on various physical and chemical characteristics [5-6]. Identifying these soil constraints supports agriculturalists use properties more capably, develop crop harvests, as well as minimize environmental effects [7]. Researchers have introduced composite measures multiple soil properties into a single indicator to provide a more assessment of SF.

2. Related Work

Several techniques are utilized for precise SF Classification using agricultural soil parameters. In [8] Gunasekaran et al (2025) is proposed Random Forest (RF) algorithm then Multilayer Perceptron and Long Short Term Memory for SF prediction. It effectively captures nonlinear relationships among soil parameters. Nevertheless, requires large number of trees increase memory requirements. Consequently, in [9] Tynchenko et al (2024) have presented Genetic Algorithm (GA) with Deep Neural Network (DNN) is used for multiclass soil-property classification. Nonetheless, performance depends strongly on the selected search ranges and GA situations. Additionally, in [10] Rahman et al (2025) have developed a hybrid Particle Swarm Optimization–Genetic Algorithm (PSO-GA) approach to optimize a Decision Tree (DT) model for crop cultivation forecasting based on soil classification. However, require much iteration. Furthermore, in [11] Selvi et al (2025) have invented Enhanced Spatiotemporal Graph Neural Network (E-STGNN) with a hybrid Particle Swarm Optimization–Red Kite Optimization (PSO-RKO) algorithm to increase SF prediction. Nevertheless, repeated optimization and model evaluation increase convergence time. Moreover, in [12] Sarangi et al (2024) have presented Decision Tree (DT) classifier predicts SF by dividing soil samples into different fertility classes. It identifies soil parameters that strongly influence fertility classification. Nonetheless, large numbers of soil parameters produce unnecessarily complex trees.

2.1 Research Contribution

The proposed CNN-GRU-Mamba, CNN extracts local feature patterns, GRU learns sequential dependencies between soil parameters, and Mamba efficiently captures long-range relationships, provides soil characteristics. WOA optimize CNN-GRU-Mamba parameters, and improving convergence.

3. Proposed Methodology

For precise SF classification using agricultural soil parameters at first data pre-processing based on handling missing values and removing duplicate records to improve data quality. Whereas data visualization based on outlier detection are performed to identify abnormal observations. Then relevant soil parameters are transformed through feature engineering based on normalization to generate consistent and informative inputs. The CNN component extracts local patterns and relationships between soil attributes, whereas the GRU captures sequential dependencies within the feature representations. The Mamba component further learns long-range dependencies efficiently, allowing representation of complex soil characteristics. Lastly, the WOA is used to optimize the important model parameters, improving convergence and classification performance. The block diagram for the proposed structure is presented in figure 1.

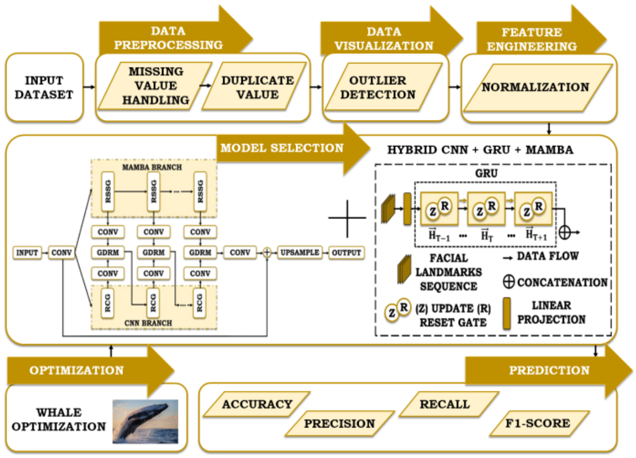

Figure 1: *Block diagram for proposed SF*

3.1 Data Pre-processing

For SF classification the data pre-processing is utilized to increase the quality as well as dependability of agricultural soil parameters. Missing values in constraints like nitrogen (N), phosphorus (P), potassium (K), pH, moisture, as well as organic carbon are identified and appropriately handled to maintain complete information. Duplicate records are also detected and removed to prevent repeated samples from biasing the learning process.

3.2 Feature Engineering

Feature engineering based on normalization is used to find the relevant agricultural soil features, like N, P, K, pH, moisture, organic carbon, as well as electrical conductivity. The selected features are then normalized to a common scale, reducing differences in their numerical ranges

$$X_{norm} = \frac{X-\bar{X}}{X_{max}-X_{min}} \tag{1}$$

Where, X_{norm} signifies normalized feature value, X denote novel feature value, $\bar{X}$ specifies mean of the feature, X_{max}& X_{min} indicates maximum as well as minimum assessment of the feature. The normalized data is fed to proposed CNN-GRU-Mamba.

3.3 Proposed CNN-GRU-Mamba

For accurate SF classification the proposed CNN-GRU-Mamba extracts soil characteristics. CNN extracts local patterns between soil constraints, GRU absorbs sequential dependencies, and Mamba captures long-range relationships efficiently. Figure 2 displays the CNN-GRU-Mamba structure.

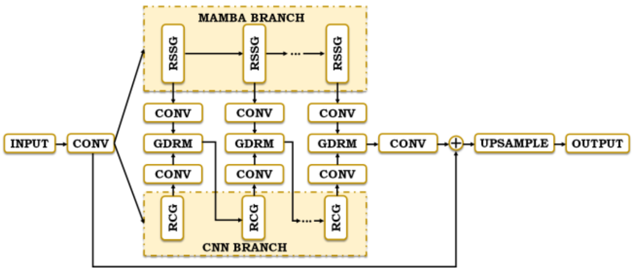

Figure 2: *Structure of CNN-GRU-Mamba*

CNN-GRU-Mamba frameworks, convolutional kernels with different sizes are first applied to the input soil attributes to capture multi-scale local feature patterns.

For input traffic sequence X, convolution is applied as

$$h_i = \sigma(W_c * X_i + b_c) \tag{2}$$

Where, W_c indicates convolution kernel, b_c represents bias and σ denotes the activation function. The refined features are subsequently processed by GRU layers, which learn dependencies among the ordered soil-feature representations

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + s_z) \tag{3}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + s_r) \tag{4}$$

Activation calculation is assumed through

$$\widetilde{h_t} = \tanh(W_h x_t + U_h(r_t \odot h_{t-1} + s_h) \tag{5}$$

Memory at time t is given by

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h_t} \tag{6}$$

Here, h_{t-1} denotes hidden state from the earlier time step, r_t *and* z_t indicates update as well as reset gates, $\odot$ signifies element-wise multiplication, σ denotes sigmoid function, W, U, s indicates trainable weights as well as biases. GRU identify the temporal patterns in soil, like slow variations in packet size, occurrence.

The GRU representation is then supplied to the Mamba module is displayed in fig.3, which uses a state-space mechanism to efficiently model long-range dependencies among soil attributes. Mamba selectively retains important information whereas efficiently propagating relevant feature relationships.

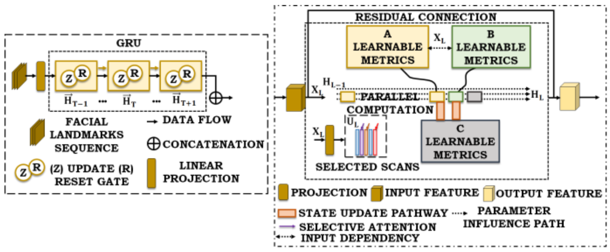

Figure 3: *GRU-Mamba operation*

The Mamba network selectively updates its internal state by

$$h_t = f(h_{t-1}, x_t, \theta) \tag{7}$$

Where, $f(\cdot)$ signifies the selective state-space function, θ denotes the learnable network parameters. The accurate SF is calculated by

$$\hat{y} = W h_t + b \tag{8}$$

Where $\hat{y}$ indicates predicted SF, W denotes output weight matrix, b represent bias vector. Lastly, the CNN-GRU-Mamba representation is distributed over a dense layer to forecast the corresponding SF class, like low, medium, high fertility of SF classification

3.3.1 Whale Optimization Algorithm

The WOA efficiently hunt through a whole solution space by mimicking humpback whale hunting methods, like encircling prey and by spiral-shaped bubble nets.

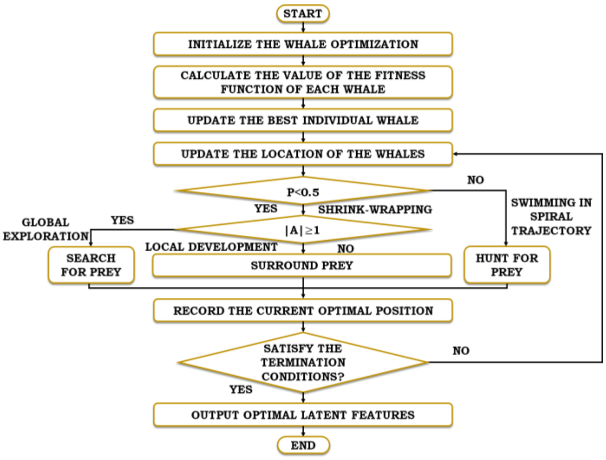

Figure 4: *Flowchart for WOA*

The function of humpback whales is to locate and recognize their prey. Here the whale considers that this is the best way of action as well as instructs other whales to approach the target prey. The circular prey approach is well-defined by

$$H = |J.R^*(k) - R(k)| \tag{9}$$

$$R(k+1) = R^*(k) - I.H \tag{10}$$

Where, k signifies present repetition, R specifies velocity location, and I and J indicates vectorial coefficients.WOA flowchart is shown in fig. 4 shows the three techniques for encircling prey are bubble-net attacking, pointed for prey, as well as enclosing prey: If there are N whales in the populace, then $X_i = \mathrm{i}\{1, 2, \ldots, \mathrm{N}\}$ shows the location vector of the i^{th}whale, a single whale, as well as offers a feasible solution to a given optimization problem.

$$R(k+1) = \begin{Bmatrix} R(k) - I.H & if\ P < 0.5 \\ H.e^{mn}.\cos(2\pi l) + R^*(k) & if\ P \geq 0.5 \end{Bmatrix} \tag{11}$$

Once the prey that is the ideal whale individual have been found, other whales in the population change their postures to approach the prey over a number of cycles. Thus the WOA shows an efficient solution for SF.

4. Result and discussion

For SF classification the proposed CNN-GRU-Mamba is implemented in Python software using Soil Nutrients dataset with 1,500 rows and 19 columns. The dataset columns X are Name, Photoperiod, Temperature, Rainfall, pH, Light_Hours, Light_Intensity, Rh, Nitrogen, Phosphorus, Potassium,Yield, Category_pH, Soil_Type, Season, N_Ratio, P_Ratio, K_Ratio and Y is Fertility.

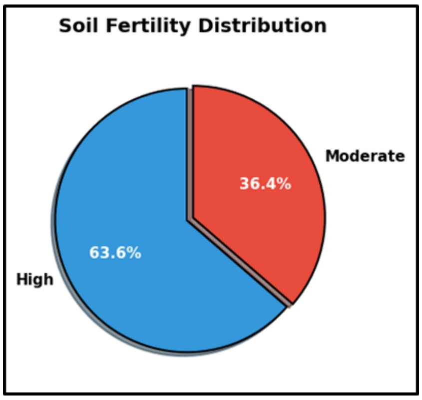

Figure 5: *Soil Fertility distribution*

Figure 5 displays the soil fertility distribution. The high have the SF of 63.6% and moderate have the SF distribution of 36.4% respectively.

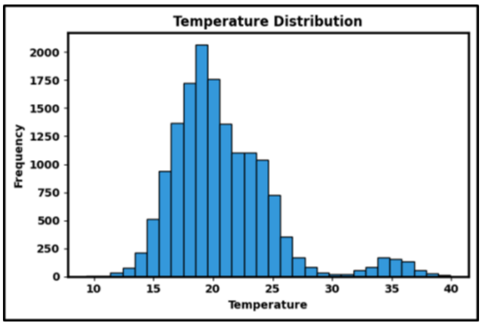

Figure 6: *Temperature distribution*

Figure 6 shows the temperature distribution. It is concentrated mainly around 18–22°C, with the highest frequency near 19°C, whereas a smaller secondary cluster appears around 34–36°C, demonstrating two distinct temperature ranges.

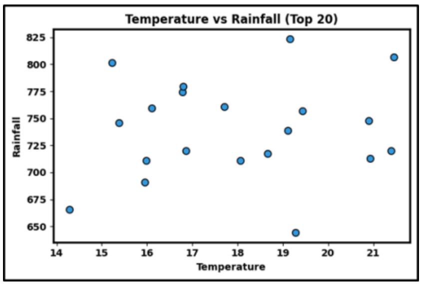

Figure 7: *Temperature vs rainfall (Top 20)*

Figure 7 displays the temperature vs rainfall (top 20). Temperature ranges from 14 to over 21 Rainfall ranges from 650 to 825. The data points are scattered widely across the grid, indicating no strong linear correlation between temperature and rainfall for these top 20 locations.

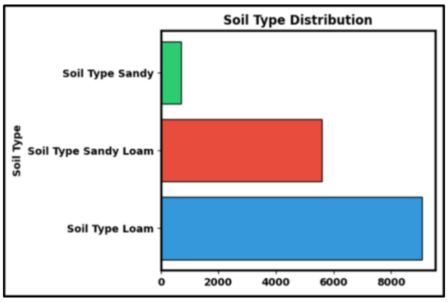

Figure 8: *Soil Type Distribution*

Figure 8 shows the soil type distribution. Soil type sandy have the value of 700 units, soil type sandy loam have the value of 5600 units and soil type loam have the value of 9100 units.

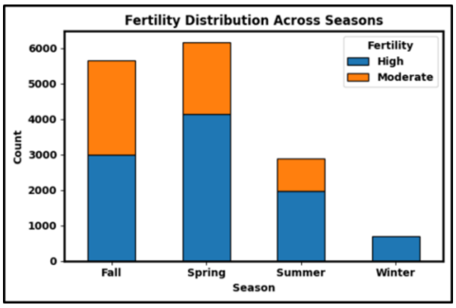

Figure 9: *Fertility Distribution across Seasons*

Figure 9 displays the fertility distribution across seasons. Fall shows the second-highest total count of 5,650, featuring a high fertility count of about 3,000 and a large moderate fertility segment of roughly 2,650.Summer and winter drops the summer totals around 2,900 counts (2,000 high, 900 moderate), whereas winter drops greatly to an overall total of roughly 700 counts, consisting almost entirely of high fertility with negligible moderate data.

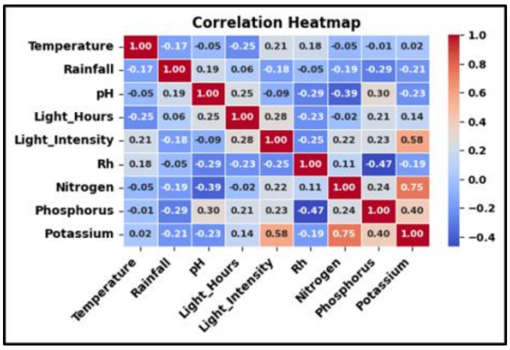

Figure 10: *Correlation Heatmap*

Figure 10 display the correlation heatmap. The statistical relationships between environmental and soil nutrient variables on a scale from -0.4 to 1.0, where red specifies a positive correlation then blue specifies a negative correlation

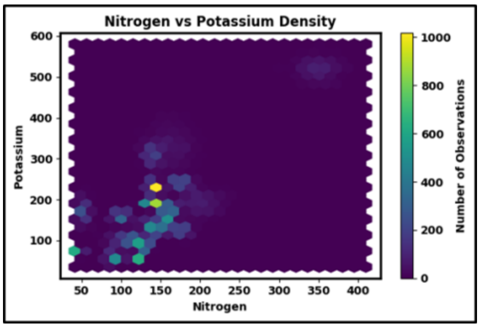

Figure 11: *Nitrogen vs potassium density*

Figure 11 shows the nitrogen vs potassium density. A hexagonal binning plot presenting the density distribution of Potassium levels relative to Nitrogen levels across a dataset. Hexagonal binning is used as an alternative to scatter plots to manage data point overlap.

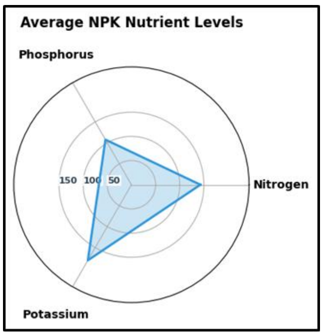

Figure 12: *Average NPK nutrient levels*

Figure 12 displays average NPK nutrient levels. The average nutrient levels are 125 for Nitrogen, 75 for Phosphorus, and 140 for Potassium. Nitrogen point sits exactly halfway among the 100 and 150 grid lines, representing 125. For phosphorus the point sits exactly halfway between the 50 and 100 grid lines, indicating 75.For potassium the point extends just below the 150 grid line, indicating nearly 140.

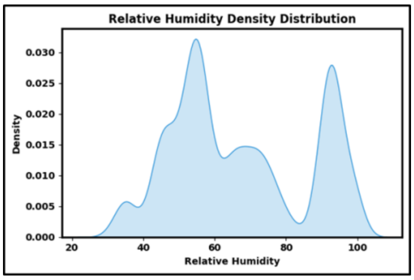

Figure 13: *Relative humidity density distribution*

Figure 13 shows relative humidity density distribution. The relative humidity distribution is multimodal, with major peaks around 55%, density of 0.032and 93% density of 0.028, indicating two dominant humidity ranges in the dataset.

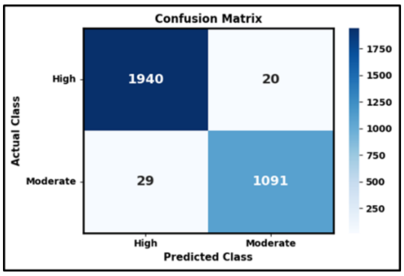

Figure 14: *Confusion matrix*

Figure 14 displays confusion matrix for proposed WOA-CNN-GRU-Mamba. It have the higher value of 1940 and moderate of 1091 respectively.

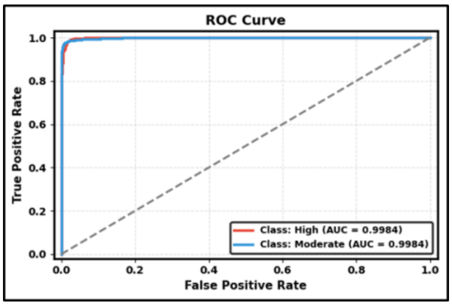

Figure 15: *ROC curve*

Figure 15 displays ROC curve. The class high and moderate have the AUC value of 0.998, when using the proposed WOA-CNN-GRU-Mamba technique.

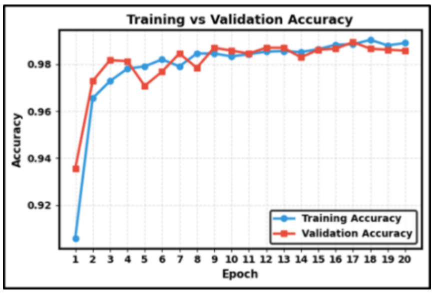

Figure 16: Training vs validation accuracy

Figure 16 shows training vs validation accuracy. It quickly improves from 92.2% and 93.5% at epoch 1 to approximately 98.9% and 98.5% by epoch 20, indicating stable and strong model performance.

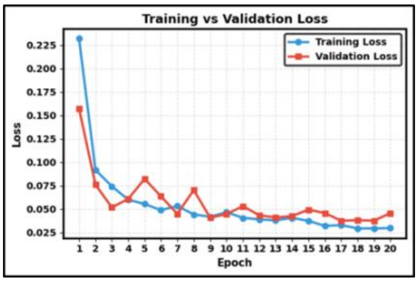

Figure 17: *Training vs validation loss*

Figure 17 displays training vs validation loss. It decrease from 0.23 and 0.15 at epoch 1 to nearly 0.03 and 0.04 by epoch 20, representing stable convergence and effective model learning.

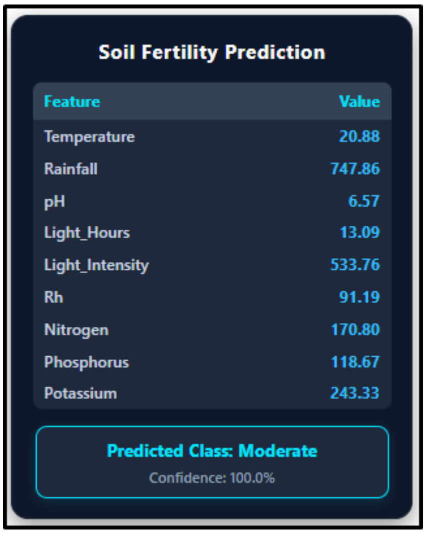

Figure 18: *Soil fertility prediction*

Figure 18 shows SF. The model predicts with 100% confidence based on the input parameters temperature 20.88°C, rainfall 747.86, pH 6.57, light hours 13.09, light intensity 533.76, Rh 91.19, nitrogen 170.80, phosphorus 118.67, and potassium 243.33.

Figure 19: *Soil fertility prediction (crop: kale)*

Figure 19 displays SF prediction (crop: kale [winter/ short day]). The model predicts high SF with 100% confidence for Kale (winter/short day) based on temperature 17.57°C, rainfall 721.20, pH 5.49, light hours 9.54, light intensity 79.39, RH 91.49, nitrogen 198.70, phosphorus 57.13, and potassium 117.21.

3.2.2 Comparison with the Proposed Technique

In the evaluation with proposed technique the proposed method is compared with the existing techniques as presented in table 1. Sarangi et al [12] have the accuracy of 89%, precision of 90%, recall of 88 % , Humeniuk et al [1] have the accuracy of 94%, precision of 70%, recall of 81%, as well as F1-score of 91%, Sharafat et al [5] have the accuracy of 95.9%, precision & recall of 92% as well as F1-score of 95.8%, Gunasekaran, et al [8] have the accuracy, precision, recall as well as F1-score of 97%, and the proposed work have the higher value with recall of 97% then accuracy, recall as well as F1-score of 98% using proposed WOA-CNN-GRU-Mamba for SF classification.

Table 1: Comparison with the proposed Technique

Study	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Sarangi et al [12]	89	90	88	-
Humeniuk et al [1]	94	70	81	91
Sharafat et al [5]	95.9	92	92	95.8
Gunasekaran, et al [8]	97	97	97	97
Proposed Work	98	98	97	98

5. Conclusion

In this paper the proposed WOA-CNN-GRU-Mamba framework provides an effective approach for precise SF classification using agricultural soil parameters. The incorporation of CNN, GRU, and Mamba allows the model to learn local feature patterns, sequential dependencies, and long-range relationships among soil characteristics, whereas WOA optimizes the model parameters to improve convergence and classification performance. The experimental results demonstrate in python software using Soil Nutrients dataset with strong performance, achieving accuracy, precision, and F1-score of 98% recall of 97%, and 100% ROC-AUC. The proposed framework is suitable for intelligent SF assessment, precision agriculture, efficient nutrient management, and sustainable soil management. The future work focus on IoT-based real-time soil sensing for continuous fertility monitoring and prediction.

References

1. Halyna Humeniuk; Dmytro Tymoshchuk; Andrii Sverstiuk, Year: 2026, “Soil quality classification from chemical composition using machine learning methods with SHAP-based explanation”, Environmental Challenges, Vol: 22, pp. 101404.

2. S. Ajith; S. Vijayakumar; N. Elakkiya, Year: 2025, "Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms", Discover Food, Vol: 5, No: 1, pp. 67.

3. Nurullah Acir, Year: 2025, "Predicting soil fertility in semi-arid agro ecosystems using interpretable machine learning models: A sustainable approach for data-sparse regions", Sustainability, Vol: 17, No: 16, pp. 7547.

4. Jiacheng Niu; Huaizhi Tang; Yuanyuan Zhang; Bingbo Gao; Wenjun Ji; Qiangyi Yu; Yuanfang Huang, Year: 2025, "Optimization of crop yield prediction models based on stratified soil properties", Smart Agricultural Technology, Vol: 12, pp. 101466.

5. MD Shaifullah Sharafat; Nilavro Das Kabya; Rahimul Islam Emu; Mehrab Uddin Ahmed; Jakaria Chowdhury Onik; Mohammad Aminul Islam; Riasat Khan, Year: 2025, "An IoT-enabled AI system for real-time crop prediction using soil and weather data in precision agriculture", Smart Agricultural Technology, Vol: 12, pp. 101263.

6. G. Srivarshini; M. Sujithra; B. Dhanalakshmi; K. Selvakumar, Year: 2025, "Smart farming with agri CNN-LSTM fusion: leveraging soil suitability analysis using real-time sensor data", Discover Applied Sciences, Vol: 7, No: 10, pp. 1151.

7. Olusegun Folorunso; Oluwafolake Ojo; Mutiu Busari; Muftau Adebayo; Joshua Adejumobi; Daniel Folorunso; Femi Ayo; Orobosade Alabi; Olusola Olabanjo, Year: 2025, "GeaGrow: a mobile tool for soil nutrient prediction and fertilizer optimization using artificial neural networks", Frontiers in Sustainable Food Systems, Vol: 9, pp. 1533423.

8. Kanimozhi Gunasekaran; Karmel A; Pemmareddy Sreevardhan, Year: 2025, "Real-time soil fertility analysis, crop prediction, and insights using machine learning and deep learning algorithms" Frontiers in Soil Science, Vol: 5, pp. 1652058.

9. Yadviga Tynchenko; Vadim Tynchenko; Vladislav Kukartsev; Tatyana Panfilova; Oksana Kukartseva; Ksenia Degtyareva; Van Nguyen; Ivan Malashin, Year: 2024, "Soil properties classification in sustainable agriculture using genetic algorithm-optimized and deep neural networks", Sustainability, Vol: 16, No: 19, pp. 8598.

10. Fardowsi Rahman; Md Ashikur Rahman Khan; Mahbubul Alam, Year: 2025, "Hybrid PSO-GA Optimization for enhancing decision tree performance in soil classification and crop cultivation prediction", Evolutionary Intelligence, Vol: 18, No: 1, pp. 30.

11. CP Thamil Selvi; M. Manimaraboopathy; M. Jeyalakshmi; J. Narmadha, Year: 2025, "Intelligent Soil Fertility Forecasting Using Enhanced STGNN and Hybrid Swarm-Based Optimization", Results in Engineering, Vol: 27, pp. 106866.

12. Ashutosh Sarangi; Sailesh Kumar Raula; Sohamdev Ghoshal; Swadhin Kumar; Chinta Sai Kumar; Neelamadhab Padhy, Year: 2024, "Enhancing process control in agriculture: Leveraging machine learning for soil fertility assessment", Engineering Proceedings, Vol: 67, No: 1, pp. 31.