Journal of Scientific & Industrial Research
Vol. 85, June 2026, pp. 511-519
DOI:10.56042/jsir.v85i6.6134

Predicting Topsoil Suitability for Healthy Crop Production using a Threshold-Guided Deep Neural Network Model

Eric Yaw Agbezuge & Balakrishnan P*
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, 632 014, India

Received 11 October 2023; revised 28 November 2025; accepted 10 September 2026

Healthy food production for the growing population usually affected by soil nutrient imbalance is a great challenge. This study presents a data-driven framework for predicting topsoil suitability for healthy crop production using physicochemical soil parameters and a Deep Neural Network (DNN) model. Soil samples were categorized using agronomic threshold ranges derived from international standards and national guidelines to group soils as suitable or unsuitable. A fully connected Deep Neural Network was implemented to discover the nonlinear relationships between soil attributes and suitability outcomes. To ensure robustness, the model was evaluated alongside some traditional classifiers using stratified cross-validation and independent testing. The Deep Neural Network achieved competitive performance on noise-free data and demonstrated superior robustness under increasing levels of Gaussian noise. The integration of threshold-based labelling with deep learning provides a practical pre-harvest decision-support tool. The proposed framework contributes to sustainable agriculture by enabling early detection of soil conditions that may compromise crop safety and productivity.

Keywords: Heavy metals, Machine learning, Precision agriculture, Physicochemical parameters, Soil nutrients

Introduction

Among the key challenges of modern societies is ensuring food security for the growing population. Several factors affect food security. The common ones are soil and water quality, post-harvesting loss and climatic conditions. Different research works have given a considerable amount of recognition to this challenge based on the relevance of the threat, sources and their effects in diverse ways. One area that is catching the eyes of most researchers now is the effect of soil pollution on healthy crop production, which adversely affects food security. Agricultural systems rely heavily on soil health, which directly influences plant nutrient availability, crop yield, and the safety of food products. Degraded soil conditions can lead to reduced agricultural productivity and increased vulnerability to contamination by harmful substances.[1–4]

Soil quality is determined by biological, chemical, and physical characteristics. Chemical composition, however, is the crucial characteristic in regulating nutrient availability and plant growth. Soil nutrients can be grouped into macronutrients and micronutrients. Macronutrients are needed in large quantities for crop development. Micronutrients, on the other hand, are needed in low quantities for enzyme activation and metabolic processes in plants.[5] When the concentrations of these elements are in excess, they adversely affect crop quality and human health.

Apart from these nutrients, there exist other naturally occurring metallic elements called heavy metals, which can affect the quality of agricultural soils. Through natural geological processes and anthropogenic activities, these metals can accumulate in soils. Once these metals enter the soil, they may be absorbed by plant roots and subsequently transferred to edible plant tissues, creating potential health hazards for consumers.[6,7]

The presence of potentially toxic metals in agricultural soil is an important environmental and food-safety concern because these elements can enter the food chain through crop uptake. To reduce the associated risks, international and national regulatory institutions have developed soil-quality guidelines that specify acceptable concentrations of metals in soil intended for agricultural use. The World Health Organization (WHO), the United States Environmental Protection Agency (US EPA), and other environmental authorities have consequently established maximum permissible limits for selected metals.[8–13] These limits provide a basis for determining whether soil conditions are suitable for

*Author for Correspondence
E-mail: balakrishnan.p@vit.ac.in

safe agricultural production. However, such assessments are not consistently undertaken prior to cultivation in many farming environments. Consequently, crops may be grown on soils whose chemical characteristics have not been adequately evaluated, potentially creating implications for consumer health and the wider food supply system.[14]

The growing availability of Artificial Intelligence (AI) provides an alternative means of analysing complex environmental and agricultural data. Unlike conventional approaches that may depend heavily on predefined relationships between variables, Machine Learning (ML) and Deep Learning (DL) methods can learn patterns directly from data. This capability has contributed to their increasing use in agricultural applications, including crop productivity estimation, plant health assessment, weed identification, and precision agriculture.[15] Their capacity to represent complicated and nonlinear interactions is particularly useful when agricultural datasets contain numerous variables and complex relationships.

Despite these developments, research on AI applications in agriculture has predominantly concentrated on visual information. Deep learning architectures such as convolutional neural networks (CNNs) and transformer models, for instance, have received considerable attention in tasks involving plant disease recognition, crop surveillance, and other image-based agricultural applications.[16–21] Furthermore, the application of EfficientNetB0 an image-oriented deep learning approach for precision agriculture and vision transformer based plant disease detection, produced an outstanding performance accuracy values of 99.51% and 90.99% respectively.[22,23] The same level of attention however has not been given to the analysis of quantitative soil chemistry measurements. Consequently an important research gap remains concerning the use of deep learning to interpret numerical soil-property data and determine the suitability of agricultural soils. Addressing this gap could provide a data-driven approach for identifying potentially unsuitable soils before crop cultivation and thereby contribute to safer and more informed agricultural decision-making.

This study develops a fully connected Deep Neural Network (DNN)-based predictive framework for assessing the suitability of topsoil for producing healthy crops. The model applies soil physicochemical parameters and agronomic threshold rules from international standards and guidelines[8–13] to classify soils into suitable and unsuitable categories. The study also performs a comparative evaluation with traditional classifiers.

The novelty of this study is not about the architectural complexity of the neural network, but about the integration of agronomic threshold-based labelling, controlled synthetic data generation, and a reproducible learning pipeline designed for soil suitability prediction to assist in pre-harvest decision making. The summary of the contribution of this study is:

1. A threshold-guided soil suitability classification framework using physicochemical parameters based on agronomic standards.
2. A DNN model using structured soil data.
3. A comparative assessment with classical classifiers.
4. A reproducible pipeline incorporating pre processing, augmentation, and validation.

Related Work

The focus of this section is reviewing existing studies relevant to soil contamination and agricultural safety, soil quality evaluation methods and machine learning in agriculture.

Soil Contamination and Agricultural Safety

Soil is contaminated by heavy metals through industrial activities, urban runoff, mining processes, and irrigation with contaminated wastewater.[8,24] Long-term accumulation of these elements can lead to the degradation of soil fertility and contamination of agricultural products. This means the concentration levels need to be monitored and controlled to prevent potential absorption of these toxic metals from the soil by the crops.[14] Studies have shown that plants can absorb heavy metals from soil through the roots to the edible plant tissues, causing health complications in humans feeding on them.[6,25,26]

Several research works have analysed heavy metal concentrations in soil samples using chemical analysis techniques and statistical methods. For example, regression models have been applied to investigate associations between soil properties and metal absorption in crops such as rice and tobacco.[6] Other studies have explored the effects of wastewater irrigation on heavy metal mobility.[27] These studies focused on retrospective analysis rather than predictive modelling. Thus, there remains a need for computational tools capable of predicting soil suitability as part of the pre-harvest activities.

Soil Quality Evaluation Methods

Traditional soil quality assessment methods involve comparing measured soil parameters with established threshold values derived from environmental guidelines. These parameters typically include soil pH, electrical conductivity, and nutrient concentrations.[28]

Threshold-based evaluation approaches are useful for determining whether soil conditions fall within acceptable ranges for agricultural production. However, these methods often require manual interpretation by soil experts and may not be scalable for large-scale monitoring.

Furthermore, soil characteristics interact in complex ways that may not be adequately captured through simple threshold comparisons. This limitation highlights the need for advanced computational approaches capable of modelling multidimensional relationships among soil parameters.

Machine Learning in Agriculture

The use of the latest techniques in machine learning in the field of agriculture has provided great benefits to farmers in minimizing losses due to the critical recommendations and insights about crops, which cover pre-harvest, harvest and post-harvest activities, since these techniques can discover patterns from large datasets. Applications include crop yield forecasting, irrigation optimisation, pest detection, and precision farming decision-support systems.[15,16]

Recent studies have increasingly explored machine learning and deep learning for soil-related prediction and classification tasks. Sumathi *et al.*[17] demonstrated the application of deep learning to soil quality prediction for smart agriculture, highlighting the potential of neural models for learning complex relationships among soil attributes. Even though this study explicitly acknowledges that advanced deep-learning architectures, such as convolutional neural networks (CNNs) and transformers have been widely applied to agricultural problems[22,23] this research distinguishes these image-oriented and large-scale dataset applications from the present study, which uses structured numerical soil physicochemical data. A recent systematic review by Folorunso *et al.*[18] examined machine-learning approaches for predicting soil nutrient properties and reported the growing application of data-driven methods in digital soil assessment. Similarly, El Behairy *et al.*[20] investigated machine-learning approaches for soil quality prediction in dryland environments, while Dengiz *et al.*[21] evaluated classification and regression-based machine-learning algorithms for assessing soil quality in cultivated lands. These studies, including[29–31] demonstrate the increasing use of data-driven approaches for soil assessment; however, their objectives and modelling settings differ from the present study, which integrates agronomic threshold rules with a DNN to classify topsoil as suitable or unsuitable for crop production

Research Gap

Although recent studies have demonstrated the potential of machine learning and deep learning for soil-quality prediction, soil-property estimation, and soil classification, relatively limited attention has been given to integrating agronomic threshold rules directly into a binary learning framework for pre-harvest soil suitability assessment. Existing studies generally focus on predicting individual soil properties, estimating soil quality indicators, or classifying soil types rather than determining whether a soil sample satisfies a set of agronomic suitability conditions for crop production

Therefore, there is a clear need for predictive frameworks capable of integrating soil chemistry data with regulatory standards to support early decision-making in agricultural planning. This study handles this gap by developing a DNN-based classification model capable of predicting soil suitability based on physicochemical soil parameters.

Materials and Methods

This study devises soil suitability assessment as a supervised binary classification problem based on agronomic threshold rules. The methodology blends structured dataset construction, deterministic preprocessing, meticulous synthetic data generation, noise sensitivity analysis, and deep learning model development to ensure reproducibility, robustness, and validity.

The process pipeline involved in implementing a Deep Neural Network Framework for predicting soil suitability for healthy crop production is presented in Fig. 1.

Dataset Construction and Preprocessing

The dataset used in this research consists of soil samples characterized by physicochemical parameters commonly used in agronomic evaluation. Each soil sample is represented by an eleven-dimensional feature vector containing heavy metal, macro- and micronutrient measurements.

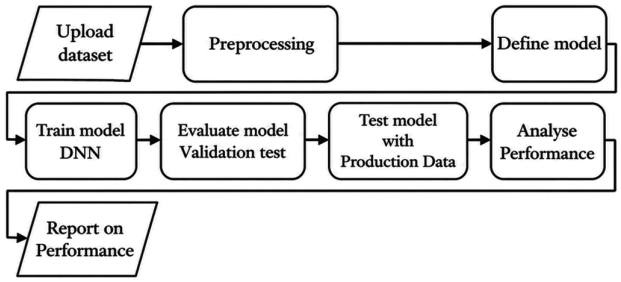


Fig. 1 — Deep neural network-based healthy crop production processes

The dataset containing a total of 1542 record scan be expressed mathematically as

$$D = \{(x_i, y_i)\}_{i=1}^{N} \quad \ldots (1)$$

where,

- $N = 1542,$ denotes the total number of samples
- $x_i \in \mathbb{R}^{11}$represents the feature vector
- $y_i \in \{0,1\}$denotes the suitability class label

Each feature vector is defined as

$$x = [N, P, K, pH, EC, S, Cu, Fe, Mn, Zn, B] \quad \ldots (2)$$

where, each variable represents a soil nutrient or physicochemical property. All features are measured using standard agronomic units. The elements also assume their usual chemical symbols. Before model development, the dataset undergoes several preprocessing steps to ensure numerical stability and reproducibility. Class distribution was approximately balanced after preprocessing.

Soil suitability labels were generated based on agronomic threshold ranges derived from international soil quality guidelines. A soil sample is classified as suitable when all parameters fall within acceptable ranges. This is the positive class category with a class label of 1.

$$y_i = \begin{cases} 1, & if\ x_{ij} \in [L_j, U_j],\ \forall j \\ 0 & Otherwise \end{cases} \quad \ldots (3)$$

where, L_j= lower threshold and U_j= upper threshold

Features scaling was performed using z-score standardization. Given that μ represents the mean of the dataset and σ denotes the standard deviation, the z-score is given by the formula

$$z = \frac{x-\mu}{\sigma} \quad \ldots (4)$$

The dataset was split into training and testing in the ratio 80:20, respectively, using stratified sampling to ensure that the class distribution remains consistent across training and testing subsets. Model validation was conducted using stratified five-fold cross-validation. The training ensured that each fold is used once for validation, and the remaining folds are used for training. All numerical values are rounded to two decimal places. The dataset is randomly shuffled using a fixed random seed of 42 to eliminate ordering bias.

To enhance the representation of edge cases and unsuitable soil conditions, synthetic samples are generated using controlled deviations from agronomic thresholds. Synthetic data are generated by introducing deviations beyond threshold limits:

$$x' = x + \delta \quad \ldots (5)$$

where, δ is a controlled perturbation factor with lower deviation factor, upper deviation factor, number of violated features and random seed parameters. The seed value for all random tasks was set to 42.

Generated samples were categorised based on threshold violations, in compliance with the deterministic classification rule. To illustrate real-world dataset scenarios, Gaussian noise was introduced to the dataset as

$$x_{noise} = x + \mathcal{N}(0, \sigma^2) \quad \ldots (6)$$

Noise levels considered were 0%, 5%, 10% and 20%, which respectively represent the original dataset, low noise, moderate noise and high noise. This enables evaluation of model robustness under uncertainty.

Model Development

The primary classification model evaluated in this study was a fully connected DNN. Three established machine-learning classifiers—Random Forest, Support Vector Machine, and XGBoost—were implemented as baseline models for comparative evaluation

The models were implemented using the Python scientific computing ecosystem, including Scikit-learn, TensorFlow, and XGBoost libraries. GridSearch CV with five-fold stratified cross-validation and F1-score as the selection criterion was used to identify the best hyperparameter configuration for each baseline model. The resulting best estimators were subsequently evaluated using the five-fold stratified cross-validation procedure. The hyperparameter settings of the baseline models are summarized in Table 1.

Deep Neural Network Architecture

A fully connected DNN was designed to model the nonlinear relationships between soil parameters and soil suitability classes. The DNN architecture is illustrated in Fig. 2.

From Fig. 2, the network consists of an input layer with 11 neurons due to the number of soil features, followed by two hidden layers whose neurons are respectively 64 and 32 with the Rectified Linear Unit (ReLU) activation function. The decreasing neuron counts for the hidden layers enable hierarchical feature learning. Dropout chances of 30% and 20%, respectively, were applied in the two hidden layers. The output layer has one neuron with a sigmoid activation function to provide binary classification.

The DNN used the fixed Adam optimizer configuration provided by Keras, trained for 50 epochs, with the specified architecture and dropout rates. The DNN was evaluated using the same five-fold stratified cross-validation scheme defined for the comparative experiments. The DNN architectural configuration for reproducibility is presented in Table 2.

The model prediction is well-defined as $\hat{y} = \sigma(W_3 f(W_2 f(W_1 X + b_1) + b_2) + b_3)$, where, σ

Table 1 — Hyperparameter space for the baseline models

Model	Hyperparameters searched	Values
Random Forest	n_estimators	100, 200
	max_depth	5, 10, None
XGBoost	n_estimators	100, 200
	max_depth	3, 5
	learning_rate	0.01, 0.1
SVM	C	0.1, 1, 10
	kernel	RBF, linear
Selection criterion	F1-score	5-fold CV

Table 2 — Deep neural network architectural configuration

Parameter	Deep neural network configuration
Input dimension	11 features
Hidden layer 1	64 neurons
Hidden-layer activation	ReLU
Dropout 1	0.30
Hidden layer 2	32 neurons
Hidden-layer activation	ReLU
Dropout 2	0.20
Output layer	1 neuron
Output activation	Sigmoid
Optimizer	Adam
Loss function	Binary cross-entropy
Training epochs	50
Classification threshold	0.50
Feature scaling	Standard Scaler within each fold
Cross-validation	5-fold StratifiedK Fold
Shuffle	Yes
Random state	42
Performance metrics	Accuracy, Precision, Recall, F1-score, ROC-AUC
DNN hyperparameter optimization	Not performed

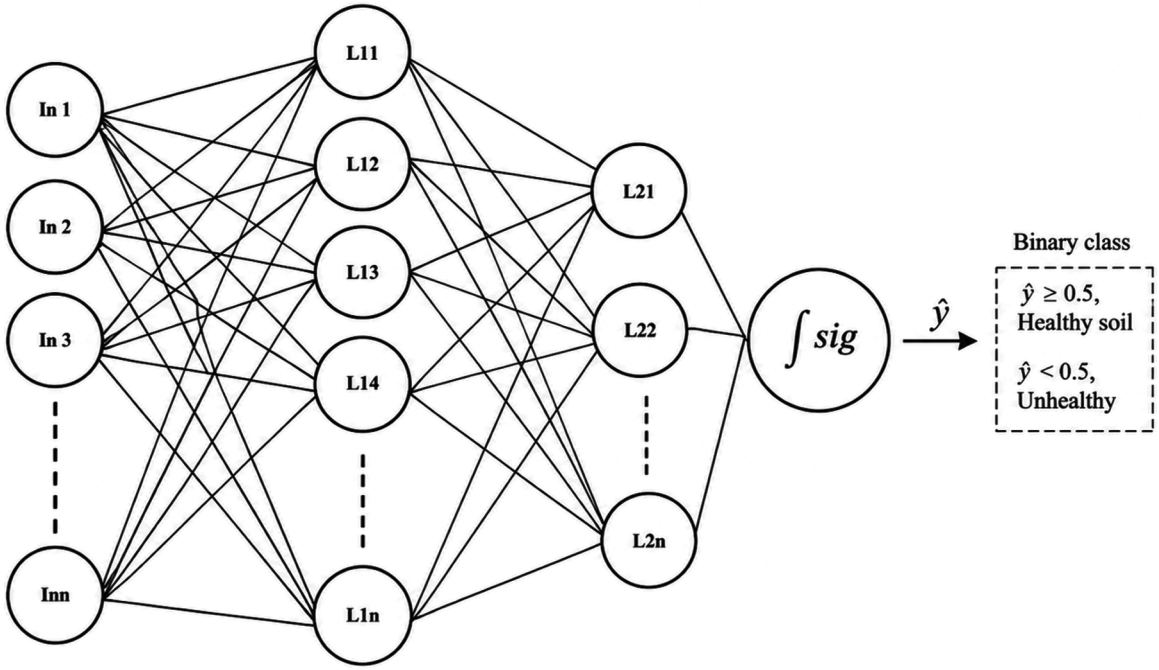


Fig. 2 — Deep Neural Network architecture for soil suitability prediction

represent the sigmoid function and f expressed as an activation function of the products of the weights (W), the input values (X) and the biases (b). The loss function is a binary cross-entropy calculated by the formula:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[\, y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad \ldots (7)$$

Performance Metrics

The performance of the DNN and the baseline models was assessed using accuracy, precision, recall, and F1-score. Using different metrics ensures a complete view of how the models perform. Accuracy represents correctly categorized samples:

$$Accuracy = \frac{(True\ Positive\ (TP) + True\ Negatives\ (TN))}{(TP + TN + FP + FN)} \quad \ldots (8)$$

Precision measures the proportion of correctly predicted suitable samples among all samples predicted as suitable:

$$Precision = \frac{TP}{TP+FP} \quad \ldots (9)$$

Recall, also known as sensitivity, represents the proportion of correctly predicted suitable samples among all actual suitable samples:

$$Recall = \frac{TP}{TP+FN} \quad \ldots (10)$$

The F1-score provides a harmonic mean of precision and recall:

$$F_1 Score = \frac{2\times (Precision * Recall)}{Precision + Recall} \quad \ldots (11)$$

These metrics provide insight into the reliability of the proposed model compared with the baseline traditional classifiers for soil suitability assessment. ROC-AUC was also computed to measure how the DNN model distinguishes between classes across different classification thresholds.

Results

To assess the robustness of the models under unpredictable levels of uncertainty, Gaussian noise was introduced into the dataset at four intensities: 0% (baseline), 5% (low), 10% (moderate), and 20% (high). The results of the baseline data are given in Table 3.

From Table 3, the models achieve very high performance on noise-free data. This reveals their strong ability to learn deterministic relationships from clean soil data instead of discovering patterns. The results when 5% Gaussian noise perturbation was done are shown in Table 4.

Table 3 — Performance of 0% Gaussian noise

Model	Accuracy	Precision	Recall	F1-score
DNN	99.71	99.65	99.78	99.71
RF	99.85	99.80	99.88	99.84
SVM	99.62	99.58	99.65	99.61
XGBoost	99.88	99.84	99.90	99.87

Table 4 — Performance of 5% Gaussian noise

Model	Accuracy	Precision	Recall	F1-score
DNN	99.12	98.95	99.28	99.11
RF	98.64	98.40	98.90	98.64
SVM	97.85	97.60	98.10	97.85
XGBoost	98.92	89.70	99.05	98.87

Table 5 — Performance of 10% Gaussian noise

Model	Accuracy	Precision	Recall	F1-score
DNN	98.43	98.10	98.72	98.41
RF	97.10	96.80	97.45	97.12
SVM	95.95	95.60	96.30	95.94
XGBoost	97.65	97.30	97.95	97.62

Table 6 — Performance of 20% Gaussian noise

Model	Accuracy	Precision	Recall	F1-score
DNN	97.65	97.20	98.05	97.62
RF	95.60	95.10	96.20	95.64
SVM	94.20	93.80	94.70	94.25
XGBoost	96.30	95.90	96.75	96.32

From Table 4, it can be seen that there is a marginal decline in the models' performance, indicating that the models are resistant to minor noise introduction. The DNN recall of 99.28% remains high, which is essential for avoiding false-negative predictions. The results of 10% Gaussian noise perturbation are given in Table 5.

From Table 5, the models' performance is seen to be degraded. However, the DNN model shows better generalisation compared to traditional models. The results of 20% Gaussian noise perturbation are given in Table 6.

From Table 6, it can be seen that the traditional classifiers have generalization limitations as more controlled noise is introduced. This provides a realistic assessment of the models. The proposed DNN model is more robust compared to baseline methods. The performance degradation trend of accuracy in percentage is given in Table 7

Table 7 — Performance degradation trend (Accuracy %)

Model	0%	5%	10%	20%
DNN	99.71	99.12	98.43	97.65
RF	99.85	98.64	97.10	95.60
SVM	99.62	97.85	95.95	94.20
XGBoost	99.88	98.92	97.65	96.30

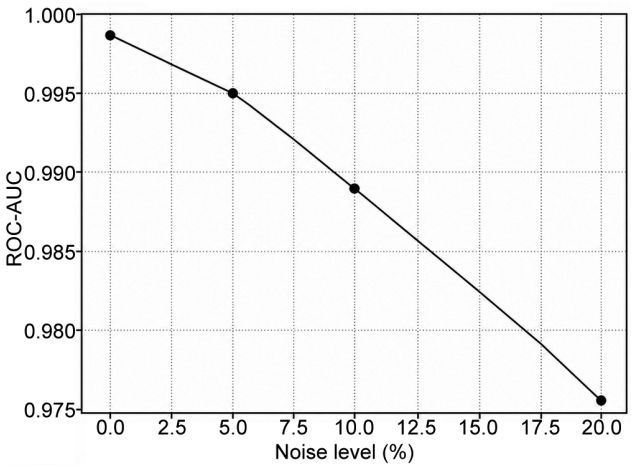


Fig. 3 — The deep neural network ROC-AUC values with Gaussian noise levels

	Predicted positive	Predicted negative
Actual positive	True positive (TP) 144	False negative (FN) 3
Actual negative	False positive (FP) 4	True negative (TN) 157

Fig. 4 — Confusion matrix of deep neural network model at 20% gaussian noise perturbation

From Table 7, with a 0% noise perturbation, XGBoost has the highest accuracy of 99.8% but drops to the second position when 5% Gaussian noise is introduced. DNN, however, remains the model with the highest accuracy when 5% and more Gaussian noise is introduced, proving its resistance to noise. The ROC-AUC values plotted against the Gaussian Noise levels can be seen in Fig. 3.

From Fig. 3, the DNN AUC values reduce gradually as the Gaussian noise levels increase from 0.999 at 0% noise to 0.975 at 20%, confirming the strong discriminative capability of the proposed DNN model under realistic data uncertainty conditions. The confusion of the final DNN model at 20% Gaussian noise perturbation is illustrated in Fig. 4.

The confusion matrix when the final DNN model was tested using the testing dataset is presented in Fig. 4. The DNN produces only 3 false negative result confirming its ability to minimise misclassification of false negatives.

Discussion

The pollution of agricultural produce by contaminated soil, due to excessive concentration levels of heavy metals, has been a critical challenge that results in food loss globally and also leads to economic loss to farmers. The aim of this research is to find a solution to this challenge through a proposed DNN model for the prediction of topsoil suitability for healthy crop production. This developed DNN model was compared with three traditional classifiers.

One of the challenges of this work was to obtain dataset for the training of the DNN. Using international standards and national guidelines regarding allowable thresholds of soil nutrients and heavy metals, a dataset consisting of 1542 records was generated to train the DNN and the baseline models. This dataset followed a deterministic rule, which is not practical in real-world scenarios. A controlled Gaussian noise of 5%, 10% and 20% were introduced into the dataset.

The results demonstrate that the DNN is especially effective at uncovering the complex, non-linear relationships within soil properties. When the DNN model was compared to the traditional classifiers used as baseline models in this study, some important trends were noticed. In the noise-free dataset, the models performed exceptionally well. As noise at varying levels was introduced, the DNN maintained higher levels of accuracy and reliability than traditional models, proving its suitability for practical use. The DNN model consistently produced a high recall rate, indicating that it is good at identifying suitable soil, making it an ideal choice for high-stakes decisions.

Conclusions

Most systems designed by researchers for agricultural activities have focused much on harvesting and post-harvesting practices, with few concentrating on the pre-harvesting, monitoring and evaluation stage, which is critical in determining the quality of the end product of agricultural enterprise. In this research, we proposed a model for predicting topsoil suitability for agronomy based on a Deep Neural Network (DNN) for healthy crop production. The proposed model yielded a performance accuracy rate of 97.65% at 20% Gaussian noise perturbation. This study used a dataset based on established rules, which might not capture every single quirk of nature. Future work will need to test the DNN model on a

much larger scale in actual fields to see how it is affected by changing weather, different soil textures, and microbial activity. Future research should also explore integration with real-time soil sensors, geospatial soil datasets, and explainable AI techniques to enhance model transparency and applicability.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1 Abdulrahaman S, Population growth and food security in Nigeria (2010 - 2012), *ncAJBMR* , **1(3)** (2013) 41–53.

2 Smith V H & Glauber J W, Trade, policy, and food security, *Agric Econ*, **51(1)** (2019) 159–171, doi:10.1111/agec.12547.

3 Adetiloye K A, Agricultural financing in Nigeria: an assessment of the agricultural credit guarantee scheme fund (ACGSF) for food security in Nigeria (1978-2006), *J Econ*, **3(1)** (2012) 39–48, doi:10.1080/09765239.2012.11884951.

4 Montanarella L, Pennock D J, McKenzie N, Badraoui M, Chude V, Baptista I, Mamo T, Yemefack M, Aulakh M S, Yagi K, Hong SY, Vijarnsorn P, Zhang G L, Arrouays D, Black H, Krasilnikov P, Sobocká J, Alegre J, Henriquez C R, Mendonça-Santos M L, Taboada M, Espinosa-Victoria D, AlShankiti A, Alavi Panah S K, Elsheikh E A E M, Hempel J, Arbestain M C, Nachtergaele F & Vargas R, World's soils are under threat, *Soil*, **2(1)** (2016) 79–82, doi:10.5194/soil-2-79-2016.

5 Fageria N K, Filho M P B, Moreira A & Guimarães C M, Foliar fertilization of crop plants, *J Plant Nutr*, **32(6)** (2009) 1044–1064, doi:10.1080/01904160902872826.

6 Golia EE, Dimirkou A & Mitsios I K, Heavy-metal concentration in tobacco leaves in relation to their available soil fractions, *Commun Soil Sci Plant Anal*, **40(1-6)** (2009) 106–120, doi:10.1080/00103620802623570.

7 Zhang M K, Liu Z Y & Wang H, Use of single extraction methods to predict bioavailability of heavy metals in polluted soils to rice, *Commun Soil Sci Plant Anal*, **41(7)** (2010) 820–831, doi:10.1080/00103621003592341.

8 Vodyanitskii Y N, Standards for the contents of heavy metals in soils of some states, *Ann Agrar Sci*, **14(3)** (2016) 257–263, doi:10.1016/j.aasci.2016.08.011.

9 Kinuthia G K, Ngure V, Beti D, Lugalia R, Wangila A & Kamau L, Levels of heavy metals in wastewater and soil samples from open drainage channels in Nairobi, Kenya: community health implication, *Sci Rep*, **10(1)** (2020) 8434, doi:10.1038/s41598-020-65359-5.

10 Osmani M, Bani A & Hoxha B, Heavy metals and Ni phytoextraction in the metallurgical area soils in Elbasan, *AJAS*, **14(4)** (2015) 414–419.

11 Ogundele D T, Adio A A & Oludele O E, Heavy metal concentrations in plants and soil along heavy traffic roads in north central Nigeria, *J Environ Anal Toxicol*, **5(6)** (2015) 1–5, dx.doi:10.4172/2161-0525.1000334.

12 McKeague J A & Wolynetz M S, Background levels of minor elements in some Canadian soils, *Geoderma*, **24(4)** (1980) 299–307, doi:10.1016/0016-7061(80)90057-9.

13 He Z, Shentu J, Yang X, Baligar VC, Zhang T & Stoffella P J, Heavy metal contamination of soils: sources, indicators, and assessment, *J Environ Indic*, **9** (2015) 17–18.

14 Karanja N, Njenga M, Prain G, Kang'ethe E, Kironchi G, Githuku C, Kinyari P & Mutua G K, Assessment of environmental and public health hazards in wastewater used for urban agriculture in Nairobi, Kenya, *Trop Subtrop Agroecosystems*, **12(1)** (2010) 85–97.

15 Meshram V, Patil K, Meshram V, Hanchate D & Ramkteke S D, Machine learning in agriculture domain: a state-of-art survey, *Artif Intell Life Sci*, **1** (2021) 100010, doi:10.1016/j.ailsci.2021.100010.

16 Lin J, Chen X, Pan R, Cao T, Cai J, Chen Y, Peng X, Cernava T & Zhang X, GrapeNet: A lightweight convolutional neural network model for identification of grape leaf diseases, *Agriculture*, **12(6)** (2022) 887, doi:10.3390/agriculture12060887.

17 Sumathi P, Karthikeyan V V, Kavitha M S & Karthik S, Improved soil quality prediction model using deep learning for smart agriculture systems, *Comput Syst Sci Eng*, **45(2)** (2023) 1545–1559, doi:10.32604/csse.2023.027580.

18 Folorunso O, Ojo O, Busari M, Adebayo M, Joshua A, Folorunso D, Ugwunna CO, Olabanjo O & Olabanjo O, Exploring machine learning models for soil nutrient properties prediction: A systematic review, *Big Data Cogn Comput*, **7(2)** (2023) 113, doi:10.3390/bdcc7020113.

19 Pandian J A, Kanchanadevi K, Kumar V D, Jasińska E, Goňo R, Leonowicz Z & Jasiński M, A five convolutional layer deep convolutional neural network for plant leaf disease detection, *Electronics*, **11(8)** (2022) 1266, doi:10.3390/electronics11081266.

20 El Behairy R A, El Arwash H M, El Baroudy A A, Ibrahim M M, Mohamed E S, Rebouh N Y & Shokr M S, An accurate approach for predicting soil quality based on machine learning in drylands, *Agriculture*, **14(4)** (2024) 627, doi:10.3390/agriculture14040627.

21 Dengiz O, Alaboz P, Saygın F, Adem K & Yüksek E, Evaluation of soil quality of cultivated lands with classification and regression-based machine learning algorithms optimization under humid environmental condition, *Adv Space Res*, **74(11)** (2024) 5514–5529, doi:10.1016/j.asr.2024.08.048.

22 Subhan M, Rana N, Qamar F, Jeribi F, Alhazmi AH, Tahir A & Shahzadi R, Crop disease detection using EfficientNetB0 deep learning approach for precision agriculture, *Discov Comput*, **29(1)** (2026) 25, doi:10.1007/s10791-025-09881-y.

23 Barman U, Sarma P, Rahman M, Deka V, Lahkar S, Sharma V & Saikia M J, Vit-SmartAgri: vision transformer and smartphone-based plant disease detection for smart agriculture, *Agronomy*, **14(2)** (2024) 327, doi:10.3390/agronomy14020327.

24 Aydin ME, Aydin S, Beduk F, Tor A, Tekinay A, Kolb M & Bahadir M, Effects of long-term irrigation with untreated municipal wastewater on soil properties and crop quality, *Environ Sci Pollut Res*, **22(23)** (2015) 19203–19212, doi:10.1007/s11356-015-5123-1.

25 Grigg J, Environmental toxins; their impact on children's health, *Arch Dis Child*, **89(3)** (2004) 244–250, doi:10.1136/adc.2002.022202.

26 Fischer A B, Georgieva R, Nikolova V, Halkova J, Bainova A, Hristeva V, Penkov D & Alandjiisk D, Health

risk for children from lead and cadmium near a non-ferrous smelter in Bulgaria, *Int J Hyg Environ Health*, **206(1)** (2003) 25–38, doi:10.1078/1438-4639-00194.

27 Christou A, Eliadou E, Michael C, Hapeshi E & Fatta-Kassinos D, Assessment of long-term wastewater irrigation impacts on the soil geochemical properties and the bioaccumulation of heavy metals to the agricultural products, *Environ Monit Assess*, **186(8)** (2014) 4857–4870, doi:10.1007/s10661-014-3743-4.

28 Firdous S, Begum S & Yasmin A, Assessment of soil quality parameters using multivariate analysis in the Rawal Lake watershed, *Environ Monit Assess*, **188(9)** (2016), doi:10.1007/s10661-016-5527-5.

29 Shafi U, Mumtaz R, Iqbal N, Zaidi S M H, Zaidi S A R, Hussain I & Mahmood Z, A multi-modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning, *IEEE Access*, **8** (2020) 112708–112724, doi:10.1109/ACCESS.2020.3002948.

30 Paudel D, Boogaard H, De Wit A, Janssen S, Osinga S, Pylianidis C & Athanasiadis I N, Machine learning for large-scale crop yield forecasting, *Agric Syst*, **187** (2021) 103016, doi:10.1016/j.agsy.2020.103016.

31 Amani M A & Marinello F, A deep learning-based model to reduce costs and increase productivity in the case of small datasets: a case study in cotton cultivation, *Agriculture*, **12(2)** (2022) 267, doi:10.3390/agriculture12020267.