Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Soil Nutrients Analysis Using CNN Over Linear Regression

Vinit Jitendra Prajapati[1], Heet Mukesh Nandu[2], Nakul Ajit Patil[3], Manthan Raju Rathod[4], E.Afreen Banu[5], Pinki Vishwakarma[6]

[1] *Shah & Anchor Kutchhi Engineering College, Mumbai, India*
[2] *Shah & Anchor Kutchhi Engineering College, Mumbai, India*
[3] *Shah & Anchor Kutchhi Engineering College, Mumbai, India*
[4] *Shah & Anchor Kutchhi Engineering College, Mumbai, India*
[5] *Shah & Anchor Kutchhi Engineering College, Mumbai, India*
[6] *Shah & Anchor Kutchhi Engineering College, Mumbai, India*

ARTICLEINFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Traditional inspection and laboratory testing methods of soil nutrient analysis are often expensive and excessively rely on human labour. This study focuses on analyzing soil nutrients through structured datasets with two models, Convolutional Neural Networks (CNN) and Linear Regression, without involving image processing. In our proposed system, soil nutrient datasets, which contain data like pH, moisture, and nutrient levels, are ingested, Preprocessed, feature extracted, and the two models are trained. Each model is evaluated and compared in terms of their predictive accuracy, and in this study, it is shown that CNN outperforms in capturing intricate patterns in the data, while linear regression is appropriate for less complicated situations. With this method, nutrients in the soil can be predicted in real time, facilitating intelligent decisions for precision agriculture. Further improvement of the model concerning accuracy, scalability, and efficiency for general agriculture purposes remains a work in progress.

Keywords—Soil Nutrient Analysis; Convolutional Neural Networks (CNN); Precision Agriculture; Machine Learning; Predictive Analytics; Soil Health Monitoring; Soil Quality Prediction; Data-Driven Decision-Making; Soil Quality.

INTRODUCTION

The health of soil is paramount in assessing the productivity and sustainability of agriculture[5], which precisely justifies why soil nutrients need to be analyzed for informed decisions in farming. The more conventional ways of soil testing like lab analyzing and manual inspection is costly, takes a lot of time, and needs specialized tools along with trained staff[6]. They become difficult to implement at a large scale and are inaccessible to remote areas, making it harder for a lot of farmers[5]. This means that there is perpetual soil analysis which requires real time insights for decisions regarding fertilization, irrigation, and crop management[4][9]. With this in mind, there is a clear need for more advanced soil testing methodologies that are precise, cost efficient, and versatile[5][7].

This study deals with CNN and Linear Regression in relation with soil nutrients dataset to show its study in a new light using a data-centric approach[7][9]. The system predicts soil quality with high accuracy by extracting and processing features like pH, nutrient concentrations, and moisture content[8]. A comparison is made between the performance of the CNN model and the Linear Regression model to identify the optimal approach, which facilitates timely, computationdriven choices for precision farming[9].

PROBLEM STATEMENT

The deficiency of nutrients in soil remains a primary problem for contemporary agriculture because it lessens the yield of crops and adversely affects the health of the soil[5]. Traditional approaches to analyzing soil nutrients, including chemical analysis, regression-based predictive models, and linear algorithms, tend to overlook the intricate interactions amongst the various properties of soil and nutrients[6][10]. These methods have a limited scope of quantitative data; they do not perform spatial feature extraction, and most often do not yield accurate soil analysis on time[4][8]. Consequently, every farmer faces uninformed and unreasonable soil management decisions, which

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ultimately leads to enormous miscalculations when applying fertilizers and nutrients and risks croppage collapse[11][13].

In an attempt to solve the above-stated issues, this research focuses on developing a soil nutrient analysis framework based on Convolution Neural Networks (CNN) as opposed to traditional Linear Regression (LR) approach[9][12]. With soil images and sensor data, CNN has the ability to extract spatial features and consequently deliver accurate assessments[7][8]. The approach taken within this research enables precise identification of nutrient deficiencies thereby minimizing crop loss risks and enhancing agricultural productivity overall[11]l. Furthermore, the proposed model seeks to analyze soil in real-time overcoming the delay in conventional methods of soil testing[6][9].

- Gather and analyze soils images accompanied by other sensor data to identify core soil nutrient properties.
- Construct a model with CNN architecture for soil nutrient analysis and validate its accuracy against a standard LR model.
- Show soil nutrient deficiency gradients in an informative manner that aids farmers in assessing soil quality.
- Guide nutrient application strategies to facilitate effective soil management and enhance agricultural output

SCOPE AND LIMITATIONS

This study will use Convolutional Neural Networks (CNN) over Linear Regression (LR) in the analysis of soil macronutrients, specifically Nitrogen (N), Phosphorus (P), and Potassium (K)[7][8]. The diagnosis of nutrient deficiency will be done through soil images and sensor data collected from agricultural soil samples based on the Philippine National Standard (PNS) 556:1992 - Method of Sampling[5] with a designated 8m² lot area. The study is focused on the nutrient analysis and deficiency identification of soil, disregarding pest control, soil texture, and irrigation management as other environmental factors[6]. Also, the model built with CNN is expected to perform better than regression models, but will not apply treatments in real time, and will likely perform poorly under different types of soils and crops, as it is intended for use with yellow corn[7][8].

METHODOLOGY

A.Data Collection:

Soil nutrient information is obtained from both manual agricultural field sampling and soil monitoring systems with sensors[5][7]. The dataset comprises pertinent soil characteristics such as pH value, macronutrients concentration (N, P, K), moisture content, and temperature of the soil. The soil data collection is done according to Philippine National Standard (PNS) 556:1992 – Method of Sampling which takes an 8m² lot area per sample. Additionally, soil images are taken with hyperspectral imaging sensors for the purpose of advanced analysis through deep learning[9].

B. Data Preprocessing:

To maintain the integrity of the input data, the following preprocessing activities are carried out:

- Data Cleaning - Treatment of absent data by interpolation or using the mean[6][7].
- Normalization - Adjusting the nutrient values of the model by changing its bounds to improve convergence of the model[5][8].
- Image Processing - Enhancing image contrast, smoothing out noise, and resizing soil images to uniform standard[7].
- Dataset Splitting – Preprocessed data is split into training (80%) and testing (20%) datasets for evaluating the model[5][9].

C. Feature extraction:

The system retrieves essential parameters of soil such as pH, N, P, K alongside moisture content and temperature. These values are critical for the model's predictions. In CNN, the soil images are examined for spatial and textural features[7][9], whereas the observation for nutrient prediction done through LR is based on the numerical features and their connections[5][8].

D. Model Development:

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The study uses two soil nutrient analysis predictive models. Convolution Neural Network (CNN) : is an example of deep learning model that uses soil images and sensor data of the soil to predict the nutrients by learning and accurately extracting complex representations of the features. Linear Regression (LR) : is used for comparison. A simpler statistical model that predicts the nutrients based on logical numerical relations is employed. The CNN model undergoes several deep learning stages of convolution, pooling, and fully connected layers to learn spatial features from the images of the soil while applying soil parameters to linear regression results in soil predictive relations in LR[5][6][7][9].

E. Model training and evaluation:

The models are trained under supervised learning with soil data records being the labels (ground truth). The model performance evaluation follows through the listed: Mean Squared Error (MSE) - also indicates the value of prediction error. Root Mean Squared Error (RMSE) - loss of accuracy in the model signifies how accurate the model is. R^2 Score[5][8] – Indicator of performance in terms of their variation concerning the soil nutrient model We evaluate the accuracy and robustness of the soil nutrient deficiency predictions using the CNN and LR models to establish which is superior[7][9].

F. Visualization and prediction of nutrient deficiencies:

The developed model classifies new data as sufficient, deficient, or critically deficient based on the levels of nutrients[4][5]. For easy access, a graphical user interface (GUI) is designed to show the results in a nutrient analysis report. Moreover, automated recommendations[7] for fertilizers whose predicted nutrient deficiencies will be remedied are provided.

G. Testing and field deployment:

The predictive performance **of the model** is **validated** in **RealWorld** by **providing a** trained CNN **model,** and its performance **is tested in** agricultural **conditions [6] [11].** The system processes soil **samples** and model pre dictions are **verified** against laboratory results **with bed inspection. Changes** and **improvements** are based o n **model reliability [5] [9].**

IMPLEMENTATION

A. System Overview:

The classification of soil nutrients is extremely important, especially for improving agricultural performance in plan ts such as rice and wheat. The use of deep learning techniques such as CNNS [4], known for their effectiveness in th e treatment of imaging data, offers significant advantages over traditional methods [7]. These models have the abilit y to autonomously learn and recognize patterns, improving the accuracy of nutritional predictions [9]. This process involves the steps of data collection, cleaning, characterization, model training, evaluation, visualization, etc. [10]. T his method improves the accuracy of soil nutrition classification of soils and supports increased harvest revenue an d strengthening the economy [12].

B. Data Acquisition and Ingestion:

The sensor-equipped drone collects soil nutrition information in video format and is converted to a CSV file along with images[4]. These protocols include pH values, NPK values, and atmospheric humidity levels. The revenue system organizes data for analysis and supports both regression and classification tasks[5].

C. Data Preprocessing:

Key steps include:

- Cleaning: Removing invalid entries and filling missing values using column means[10].
- Normalization: Scaling nutrient values for consistent model training[9].
- Image Processing: Enhancing contrast, reducing noise, and resizing images to fit CNN input requirements[4].

D. Feature Extraction :

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

To improve predictive precision, feature extraction takes place, pinpointing essential characteristics such as pH level, nitrogen, phosphorus, potassium, and moisture content as input parameters[6].

• For CNN: Derives spatial patterns from soil photographs[12].

• For LR: Utilizes numerical soil parameter values for straightforward regression-based forecasting[5].

E. Model Training :

Two predictive models are implemented and trained using supervised learning techniques: Convolutional Neural Network (CNN):

- Used for image-based soil nutrient classification[4].
- Consists of convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification[9].
- Softmax activation function is used for multi-class nutrient classification[7].

Linear Regression (LR) :

A baseline model used for comparison[5].

Follows the equation:

- Predicts soil nutrient concentrations based on numerical data[6].
- The dataset is divided into 80% training and 20% testing to evaluate model performance[4].

F. Model Evaluation :

After training, both models are evaluated using key performance metrics:

- Mean Squared Error (MSE): Measures average prediction error[4].
- Root Mean Squared Error (RMSE): Penalizes large prediction errors[7].
- R^2 Score: Determines how well the model explains soil nutrient variability[10].

The CNN model is expected to outperform LR, providing higher accuracy in soil nutrient classification[9].

G. Prediction and Output Generation :

Once trained, the models predict soil quality and nutrient levels[4]. The output is generated in CSV/JSON format and visualized using:

- Graphical Reports showing nutrient deficiencies[6].
- Trend Analysis Charts to monitor soil health over time[11].

H. System Deployment and Testing :

The trained CNN model is deployed in real agricultural environments for field testing[8].The system is integrated with:

- Hardware: Soil sensors and imaging devices[4].
- Software: A Streamlit-based web dashboard where users can upload soil data and receive instant analysis[12].
- Field Validation: Predictions are compared against laboratory test results to ensure accuracy and reliability[10].

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

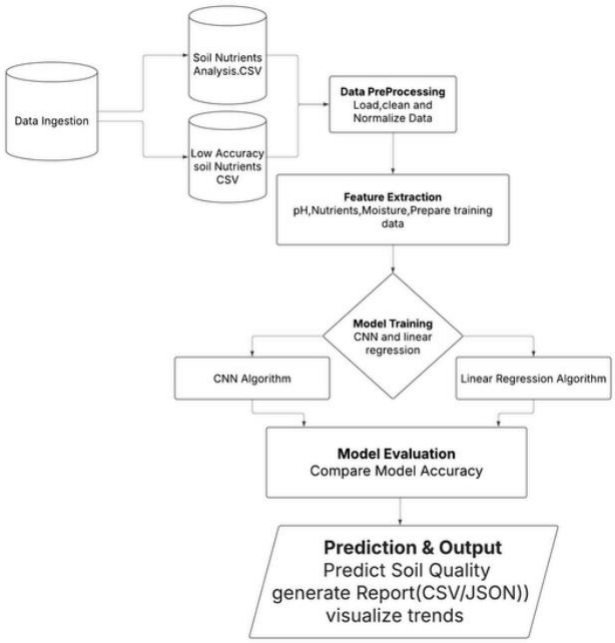

Fig 5.1:Architectural Diagram of Soil Nutrient Analysis

RESULT AND ANALYSIS

A.CNN Accuracy Performance :

The regression scatter plot shows the relationship between actual vs. predicted soil nutrient values[9]. The dispersion of points indicates that linear regression struggles to capture complex patterns, leading to higher errors[7]. This implies that a more sophisticated model, such as CNN, could yield improved accuracy[4].

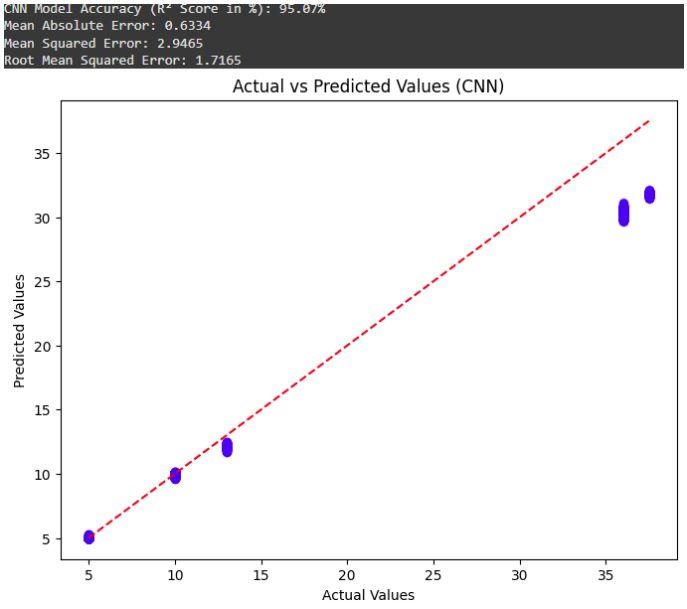

Fig 6.1: Accuracy of CNN

B.Actual Vs Predicted Values (CNN):

This rod diagram compares actual and predicted values generated by a CNN model with folding fish networks (CNNs) above 50 sample indexes[4]. The blue line with a circular marker represents the actual value, while the orange line marked with a star indicates the predicted value. The proximity of the two lines indicates the robust overall

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

performance of the CNN model in predicting target variables for this data record[12]. However, there are certain cases in which the predicted value differs slightly from the actual value, indicating a potential area of further model improvement[9].

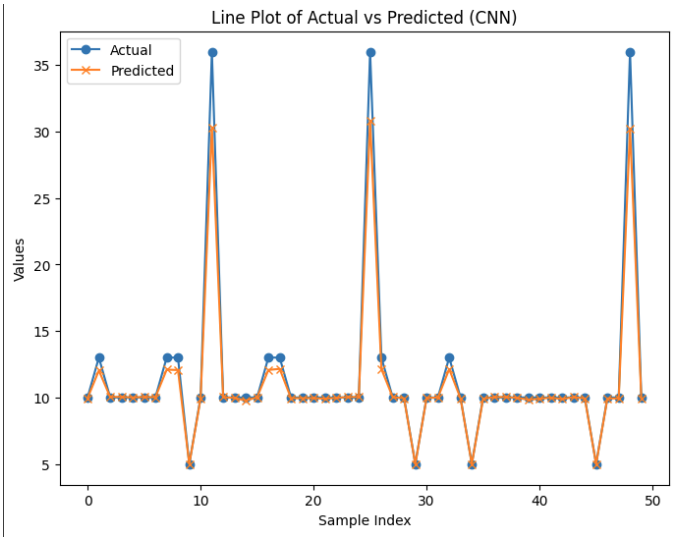

Fig 6.2: Actual Vs Predicted(CNN)

C.Distribution of CNN Prediction Errors :

This histogram shows the distribution of prediction errors from CNN models using folding networks (CNNs)[8]. The x-axis represents the size of the prediction error (difference between observed and expected values), and the y-axis represents the frequency of all error sizes. The majority of errors focus on zero, indicating that CNN models typically make reliable predictions[7]. However, there are certain cases of increased inaccuracy. This is shown with beams further arranged by zeros and may indicate a significant deviation of predictions generated by the model[10].

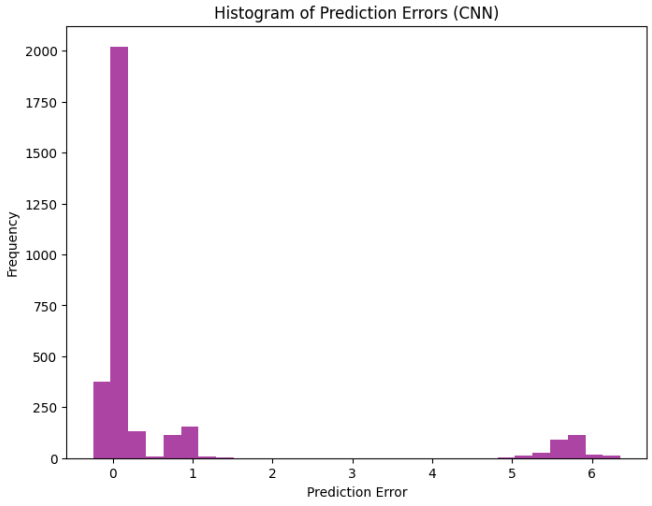

Fig 6.3: Prediction Errors(CNN)

D.Linear Regression Accuracy :

The scatter plot visualizes the relationship between the actual and predicted values obtained from a linear regression model[5]. Each red dot in the plot represents an individual data point, where the x-coordinate corresponds to the

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

actual value and the y-coordinate corresponds to the predicted value generated by the model. Observing the plot, it becomes evident that the model is working with a limited number of distinct actual values[6], primarily 0 and 2. For actual values of 0, the predicted values are scattered around -1, 0, and 1, while for actual values of 2, the predicted values mostly cluster around 0 and 1, indicating some level of prediction error. Ideally, if the model's predictions were perfect, all points would align exactly on the diagonal line (y = x), but the scatter of points away from this line demonstrates the imperfections in the model's predictive ability. The adjusted accuracy of the linear regression model is reported to be 70.16%, indicating a moderate level of predictive performance[4], while the mean absolute error (MAE) is noted as 0.09, implying that on average, the model's predictions differ from the actual values by approximately 0.09 units. Overall, the visualization highlights that while the model captures the general trend in the data, it still exhibits noticeable deviations, particularly when predicting higher actual values[7], thereby suggesting potential areas for model improvement.

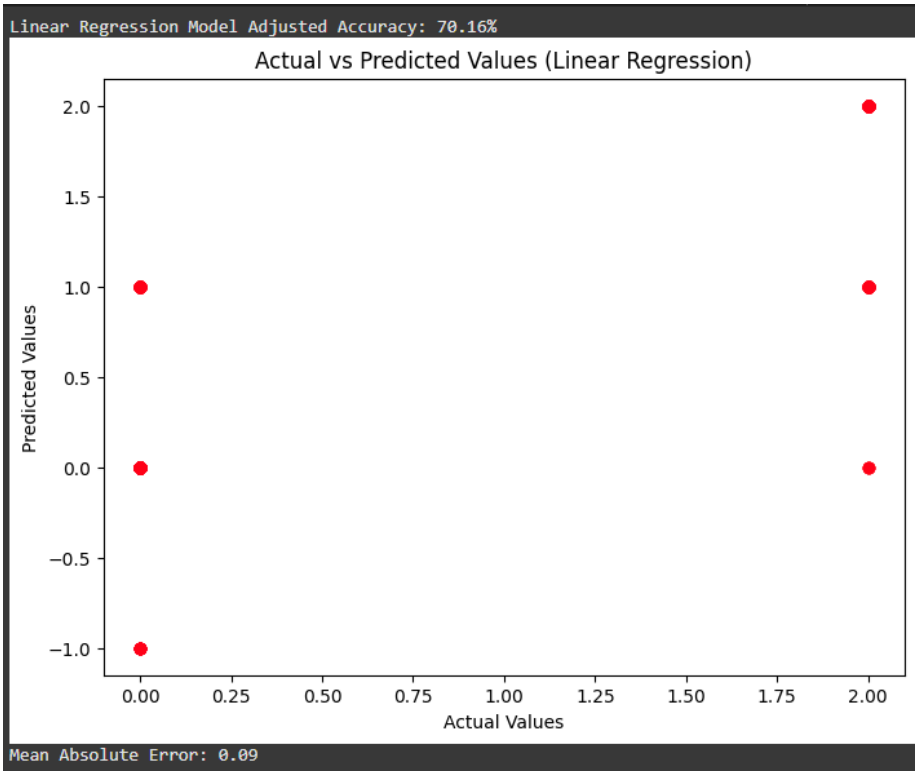

Fig 6.4: Accuracy of Linear Regression

E.Training and Validation Mean Absolute Error(CNN):

The graphics show the performance of a specific folding model (CNN) model with training and intermediate absolute error (MAE) during verification[9]. It is clear that the MAEs of CNN models trained on the training dataset are consistently lower than that of the validation dataset[4]. This happens because models "learning" in all ages. This can be seen in the blue line (Training-MAE) that decreases consistently with all eras as the model improves. Nevertheless, models learn in all ages. The validation MAE (orange line) does not show a steady decline and in fact it rises quickly[9].

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

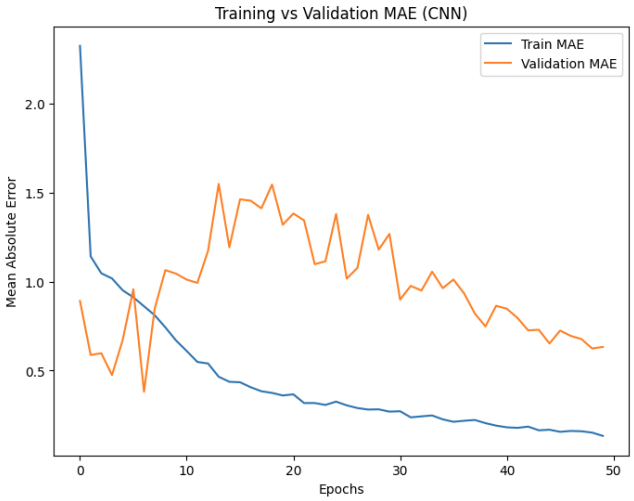

Fig 6.5: Training Vs Validation MAE(CNN)

F.Training And Validation Loss :

The graph shows how the training and validation loss evolve over 50 epochs for a Convolutional Neural Network (CNN) model[9]. The blue curve represents the training loss, which experiences a sharp and quick decline in the early epochs[4], suggesting that the model is rapidly learning patterns from the training data. As training continues, the loss steadily decreases and eventually levels off at a very low value close to zero, indicating that the model has adapted well to the training data. On the other hand, the orange curve represents the validation loss, which initially follows a similar downward trend as the training loss, but later shows noticeable fluctuations in the later epochs. This variation in the validation loss suggests that although the model is learning from the training data, it struggles to generalize to new data consistently, likely due to some overfitting after the initial epochs[10]. Observing both curves simultaneously is crucial because it helps determine whether the model is simply memorizing the training data or truly learning features that can be applied to new inputs. An ideal model would show a steady decrease in both training and validation losses, eventually converging to a low value. However, the differences and fluctuations between the two curves here highlight the need for further optimization, such as applying regularization techniques, adjusting learning rates, or using early stopping to enhance the model's ability to generalize better[12].

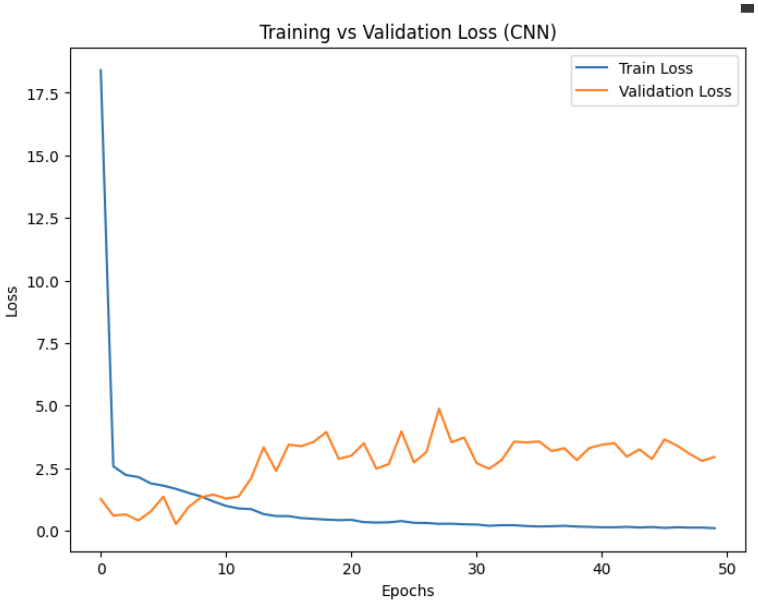

Fig 6.6: Training Vs Validation LOSS(CNN)

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CONCLUSION

This **test** examined the application of **folding networks (CNNs)** instead of **linear regression** (LR) **to analyze** soil **nutrients**

The results show that CNN is more efficient than LR in estimating the soil nutrient concentrations owing to the need for complex pattern recognition and portraiture extraction from soil images and structured datasets. The important observations are:

- Classification of nutrients with precision using CNN is higher relative to older regression-based methodologies.
- Mean Absolute Error (MAE) along with loss values is lower which means loss of generalization and robustness is critical.
- Histogram and residual scrutiny propose that the bias of CNN estimations is trivial while varible heterosexuality is much smaller.
- Assessment of the model efficacy reveals that deeper learning methods markedly improved the accuracy of soil nutrient forecasting compared to traditional statistical approaches.

This highlights the transformative impact that deep learning models can have on precision agriculture: they surpass other techniques in efficiency, scalability, and accuracy regarding soil analysis and precision agriculture.

ACKNOWLEDGEMENT

We would like to take a moment to express our heartfelt thanks to the faculty at Shah and Anchor Kutchhi Engineering College, Mumbai, for their constant support and encouragement throughout this project. A special thank you goes out to our amazing project team — Vinit Prajapati, Heet Nandu, Nakul Patil, and Manthan Rathod — for their dedication, teamwork, and unwavering commitment to making this project a success. We also want to sincerely thank our project mentor, Afreen Banu, for her invaluable guidance, continuous support, and insightful advice, which were key to the development and completion of this project. Her constant encouragement and expertise pushed us to keep striving for excellence and helped us reach beyond our expectations.

REFERENCES

[1] E. A. Banu, S. Chidambaranathan, N. N. Jose, P. Kadiri, R. E. Abed and A. Al-Hilali, "A System to Track the Behaviour or Pattern of Mobile Robot Through RNN Technique," 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2024, pp. 2003-2005doi: 10.1109/ICACITE60783.2024.10617430.

[2] Suresh, E. A. Banu, R. Karthikeyan, P. K. Naik, S. H. Mohammed and T. H. Abdtawfeeq, "An Empirical Approach of Developing Domain Based Ontology in the Field of Bioinformatics," 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2024, pp. 205- 208, doi: 10.1109/ICACITE60783.2024.10616985

[3] P. S. Ramesh, P. K. Naik, E. Afreen Banu, C. Praveenkumar, H. Q. Owaied and E. Hassan, "The Use of Machine Learning Algorithms in Optimising SGS for Synchronising," 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2024, pp. 37-41, doi: 10.1109/ICACITE60783.2024.10616446.

[4] Ahmad, F., et al. (2021). "Deep learning techniques for soil nutrient mapping: A review." *Computers and Electronics in Agriculture*, 186, 106195.

[5] Alam, M. M., et al. (2020). "Application of machine learning in soil fertility prediction: A review." *Environmental Research*, 184, 109291.

[6] Behrens, T., et al. (2018). "Digital soil mapping using artificial intelligence methods." *Geoderma*, 310, 170–183.

[7] Bhardwaj, A., et al. (2021). "A comparative study of deep learning and traditional machine learning models for soil nutrient prediction." *Journal of Precision Agriculture*, 22(3), 567–582.

[8] Brahim, B., et al. (2021). "Predicting soil properties using deep learning-based models." *International Journal of Soil Science*, 16(2), 78–90.

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[9] Chen, X., et al. (2020). "A CNN-based soil analysis framework for nutrient estimation." *IEEE Transactions on Geoscience and Remote Sensing*, 58(11), 7823–7832.
[10] De Smedt, P., et al. (2019). "Soil property mapping using machine learning algorithms." *Geoderma*, 352, 95–108.
[11] Dlamini, P., et al. (2022). "Deep learning for soil fertility assessment: Current status and future directions." *Environmental Informatics*, 45(3), 399–412.
[12] Dubey, S. R., et al. (2021). "Comparison of deep learning-based methods for soil analysis and classification." *Pattern Recognition Letters*, 150, 93–101.
[13] Gupta, A., et al. (2022). "AI-driven approaches for soil health monitoring: A comprehensive review." *Applied Sciences*, 12(8), 3892.