STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, September 2025, pp 1415–1426.
Published online in International Academic Press (www.IAPress.org)

Novel Deep Learning Model Optimized by Random Search or Grid Search Method for Soil Erosion Susceptibility Prediction

Alaa A. Almelibari [1], Yasser Fouad [2,*]

[1] *Department of Computer Science and Artificial Intelligence, College of Computing , Umm AL-Qura University , Makkah, Saudi Arabia*
[2] *Department of Computer Science, Faculty of Computers and Information, Suez University, P.O.Box:43221, Suez, Egypt*

Abstract Soil erosion is the process by which soil particles are removed from the Earth's surface. There are three stages to soil erosion: displacement, migration, and deposition. The rate of soil erosion is influenced by various factors, including infiltration, soil type, soil structure, and land cover. Soil erosion causes soil deposition in some areas and soil loss in others. One of the primary issues in meteorology is the prediction of soil erosion. Numerous methods for forecasting precipitation have been put forth, drawing from deep learning, machine learning, and statistical analysis approaches. In this paper, we compare between the Random Search and Grid Search optimization which combine with CNN, RNN, LSTM and GRU algorithm (GS_CNN, GS_RNN, GS_LSTM, GS_GRU, RS_CNN, RS_RNN, RS_LSTM, RS_GRU,) for soil erosion prediction. These models facilitate planning for soil preservation and land management techniques by improving our understanding of and capacity to forecast the dynamics of soil erosion. There are 236 instances with 11 features in the dataset that was used for this study. Six evaluation metrics were computed: accuracy, precision, recall, F1 score, Matthew's correlation coefficient (MCC), and Area Under the Receiver Operating Characteristic Curve (AUC) to assess the efficacy of the employed classification technique. With an accuracy of 98.592%, the CNN, GS_CNN, GS_RNN, RS_CNN and RS_RNN models outperformed other machine learning methods and earlier research on the same dataset, according to the experimental results.

Keywords Soil Erosion, Convolution Neural Network, Recurrent Neural Network, Long Short-Term Memory, Gated Recurrent Unit, Grid Search, Random Search, Soil Erosion Classification, Deep Learning

DOI: 10.19139/soic-2310-5070-2528

1. Introduction

Soil plays a vital role in both rural and urban ecosystems, contributing to agriculture, infrastructure, and environmental stability [1]. In recent years, soil conditions have been increasingly affected by extreme weather events resulting from climate change and landscape alterations due to urbanization [2]. Intense rainfall, in particular, accelerates water erosion, leading to the displacement of topsoil and a consequent reduction in soil fertility [3]. These effects not only damage the natural terrain but also make land restoration labor-intensive and costly. Soil erosion is a progressive natural phenomenon driven by agents such as wind, water, and ice, where soil particles are detached, transported, and deposited elsewhere. While it is influenced by natural occurrences like precipitation and water flow, human interventions have significantly intensified the process. Soil erosion threatens agricultural productivity, infrastructure development, and the safety of residential areas, particularly those located near water bodies and slopes [4]. One of the most pressing challenges is managing floodwaters to prevent sediment runoff and downstream siltation. When high-energy water jets reach the downstream base of dams, they can cause bed material displacement, forming scouring holes and potentially leading to structural failures such as spillway

*Correspondence to: Yasser Fouad (Email: Yasser.ramadan@suezuni.edu.eg). Department of Computer Science, Faculty of Computers and Information, Suez University, P.O.Box:43221, Suez, Egypt

ISSN 2310-5070 (online) ISSN 2311-004X (print)

collapse [5]. Moreover, the long-term consequences of soil erosion include the degradation of land and water resources, sedimentation of rivers, loss of soil fertility, and diminished water quality [6]. The adverse effects extend further, as sediments often bind nutrients, hazardous chemicals, and heavy metals, negatively impacting downstream aquatic ecosystems and human health [7, 8].

Despite these concerns, soil erosion also contributes positively to certain natural processes. For example, it plays a role in forming alluvial plains, developing fertile soils, rejuvenating rivers, creating new habitats, and sequestering carbon, thereby contributing to climate change mitigation. However, excessive erosion remains detrimental, especially when it leads to agricultural chemical runoff and loss of fertile land. Numerous soil conservation techniques have been developed to mitigate erosion, including conservation tillage, cover cropping, terracing, contour barriers, agroforestry, and rotational grazing [9]. While these strategies have shown effectiveness, predictive models that anticipate soil erosion can further enhance mitigation efforts. This research aims to evaluate and compare several deep learning models such as CNN, RNN, LSTM, and GRU which optimized using either Grid Search or Random Search. These combinations yield eight hybrid models: GS_CNN, GS_RNN, GS_LSTM, GS_GRU, RS_CNN, RS_RNN, RS_LSTM, and RS_GRU.

The authors in [10] focused on soil erosion in the Mayurakshi River Basin in Eastern India, where land degradation is a significant issue. To identify areas at high risk for soil erosion, the researchers utilized three classification models: logistic regression, decision tree, and random forest. They created a soil erosion inventory from 150 erosion sites, using 70% of the data for model training and 30% for validation. The models incorporated 12 environmental and topographic factors, and multicollinearity tests were performed for reliability. Findings revealed that the western part of the basin is particularly susceptible to erosion, with decision tree and random forest models showing better predictive accuracy than logistic regression, as validated by ROC curves and kappa statistics.

The authors in [11] highlighted the importance of soil in the global ecosystem by assessing soil erosion risk in the Cameron Highlands, which faces sustainability challenges. Using the Analytical Hierarchy Process (AHP), the authors evaluated 15 factors contributing to soil erosion, such as rainfall erosivity, land use, and slope characteristics. Rainfall erosivity was identified as the most significant factor, followed by land use and slope steepness. The erosion risk prediction map revealed that the western region is particularly vulnerable to erosion. The results provide essential insights for planners and policymakers to develop effective soil conservation strategies.

The authors in [12] addressed soil erosion's negative impact on agriculture and grazing in northern Algeria's Mediterranean region by utilizing advanced machine learning and deep learning techniques to predict erosion-prone areas in the Macta basin. Four models were compared: Categorical Boosting, Adaptive Boosting, CNN, and a stacking ensemble approach. Erosion conditioning factors were derived from remote sensing data and integrated into a GIS environment. Erosion sites were identified using GPS, field surveys, and satellite images. The dataset was divided into training (60%) and testing (40%) sets, with all models achieving over 98% accuracy in evaluation metrics. The resulting susceptibility maps are intended to aid policymakers and local stakeholders in soil and water conservation efforts.

The authors in [13] demonstrated the effectiveness of machine learning algorithms in assessing environmental hazards, particularly soil erosion and landslide susceptibility in Southeast Nigeria. It utilized hierarchical clustering and three multilayer perceptron neural networks to analyze risk indicators. The findings showed that hierarchical clustering ranked gully erosion susceptibility, while the neural networks achieved high predictive accuracy for soil cohesion and internal friction angle, as well as estimating the factor of safety under different conditions.

The authors in [14] investigated soil erosion hazards due to water and human activities in semi-arid and humid regions, focusing on nutrient loss and land degradation. An integrated framework combining vulnerability prediction, decision-making, and GIS techniques was proposed to assess erosion risk using seven parameters: soil loss, sediment yield, runoff potential, land capability, drainage density, sediment transport index, and slope. Conducted in the Pairi River watershed in Chhattisgarh, India, the research utilized various machine learning classifiers, notably RF and gradient adaptive boosting (GAB), achieving high accuracy rates of 91.51% and 90.55%, respectively. GAB outperformed RF in log-loss metrics.

2. Methodology

2.1. Dataset

The dataset employed in this study, accessible via [2], was gathered over a span of three years (2009–2011) from two catchment regions located in Son La, a province in northwest Vietnam with a tropical monsoon climate. This area typically experiences heavy rainfall from May to October and cooler, drier conditions from November to March, with an average yearly temperature of 21°C. A total of 24 erosion monitoring plots, each measuring 4 meters by 18 meters, were arranged using a randomized full-block layout to assess the effects of various land use practices on soil erosion. The implemented treatments encompassed both conventional agricultural methods—such as slash-and-burn and plowing with fertilizers—and soil conservation strategies like planting cover crops, establishing grass barriers, and practicing relay cropping. Runoff from each plot was collected using 200-liter plastic tanks, while sediment was gathered using a bucket system. The soil types in the study area, ranging from clay loam to clay, were classified as Alisols, Luvisols, and Calcisols. For predictive modeling, ten critical features were selected: EI30, topsoil pH, slope gradient, organic carbon content, bulk density, pore volume, proportions of clay, sand, and silt, and the rate of ground cover. The dataset effectively reflects the impact of climate, terrain, soil properties, and land use on the likelihood of soil erosion. The correlation matrix for these features is illustrated in Figure 1.

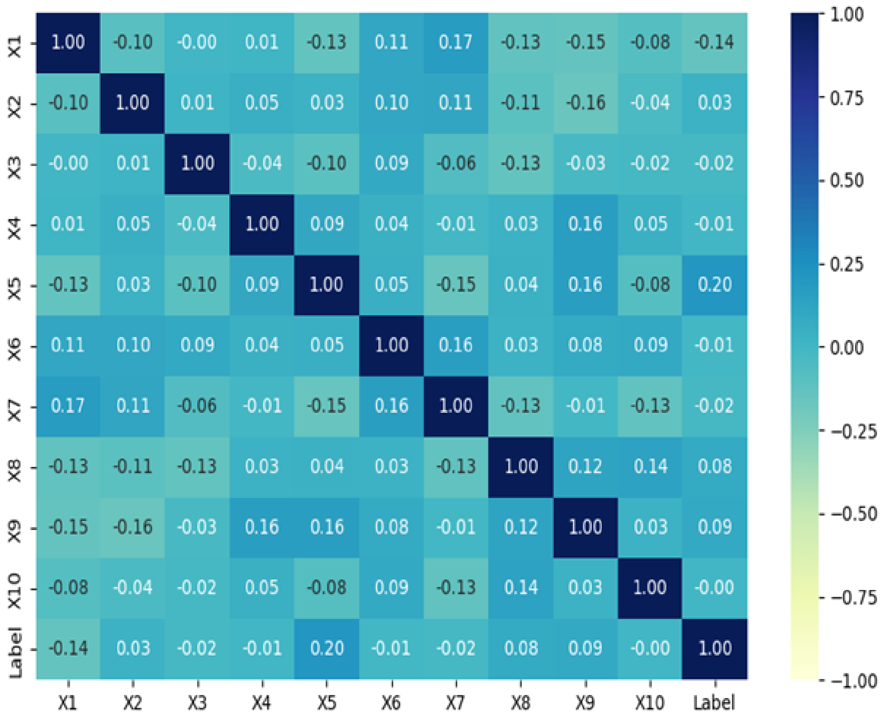


Figure 1. correlation matrix of the utilized dataset of soil erosion.

The final dataset consists of 236 samples, with 125 labeled as erosion-prone and 111 as non-prone, yielding a class distribution of approximately 53% to 47%. While the imbalance is not severe, evaluation metrics such as AUC-ROC and precision-recall curves were used to account for potential skew in prediction sensitivity.

2.2. Random Search

Random search offers simplicity, efficiency, and flexibility. It requires no complex algorithm or logic, reduces computational cost and time, and can handle any search space, including continuous, discrete, and categorical search spaces and complex or nonlinear functions. Because random search produces a pattern that increases the likelihood of discovering the ideal parameter, there is a greater possibility that the model will be trained on the optimized parameters without any aliasing. When dealing with lower-dimensional data, random search performs better since it requires fewer iterations and less time to discover the correct set. The most effective parameter search method when there are fewer dimensions is random search. Bergstra and Bengio demonstrate both conceptually and experimentally in their study Random Search for Hyperparameter Optimization that random search is a more effective method for parameter optimization than grid search. Random search has several disadvantages, including high variability due to randomness of samples, no guarantee of finding the global optimum, lack of direction, and computational cost. It may miss promising areas and may be computationally expensive for large problems with many hyperparameters, making it less efficient than exhaustive search.

2.3. Grid Search

Grid search is a technique in machine learning for hyperparameter tuning, aimed at finding the best combination of hyperparameters for a model. Hyperparameters, such as learning rate and number of layers, must be set before training and cannot be derived from data. The grid search process involves creating a grid of all possible hyperparameter combinations, training the model on each combination, and evaluating its performance on a validation set. The combination that produces the best results is selected as optimal. However, grid search can be computationally intensive, especially with many hyperparameters, leading to the use of randomized search as a more efficient alternative, which samples a random subset of hyperparameters instead.

2.4. The proposed Methodology

Figure 2 illustrates a comprehensive pipeline for predicting soil erosion using deep learning techniques. The process begins with a Soil Erosion Dataset, which undergoes preprocessing that includes null data handling and Z-Score normalization to standardize feature values. Following preprocessing, the data is split into training and testing sets—70% for training and 30% for testing. The training data is fed into various deep learning models, including RNN, GRU, CNN, and LSTM, each of which learns the patterns necessary for accurate soil erosion classification. The model performance is further enhanced using Random Search or Grid Search for hyperparameter optimization, ensuring that each model is fine-tuned for optimal accuracy. After the training phase, the models undergo classification and evaluation using the testing set. The evaluation results determine whether the soil is susceptible to erosion, represented by a binary output (Yes/No). This workflow not only leverages the strengths of multiple deep learning architectures but also integrates systematic optimization and data preparation techniques, making it a powerful solution for soil erosion prediction and land management planning.

The selection of specific deep learning architectures was motivated by the nature of the input data. Convolutional Neural Networks (CNNs) were chosen due to their ability to extract local spatial dependencies from gridded environmental variables, such as topographic slope, ground cover, and soil composition arranged in raster-like formats. Recurrent architectures (RNN, LSTM, GRU) were selected to model the temporal dynamics of rainfall erosivity and land use practices across monitoring periods. These time-dependent features naturally align with sequential modeling, making recurrent networks appropriate for capturing erosion trends over time. For reproducibility, the full architectural configuration of each model is summarized in Table 1, including layer types, number of units, and activation functions.

Table 1. Architectural configuration of the proposed deep learning models

Model	Layer Type	Units/Filters	Activation Function
CNN	Conv1D + MaxPooling + Dense	64 filters + 2 Dense layers (128, 64)	ReLU, Softmax
RNN	SimpleRNN + Dense	100 + 64 units	tanh, Softmax
LSTM	LSTM + Dropout + Dense	100 + 64 units	tanh, Softmax
GRU	GRU + Dropout + Dense	100 + 64 units	tanh, Softmax

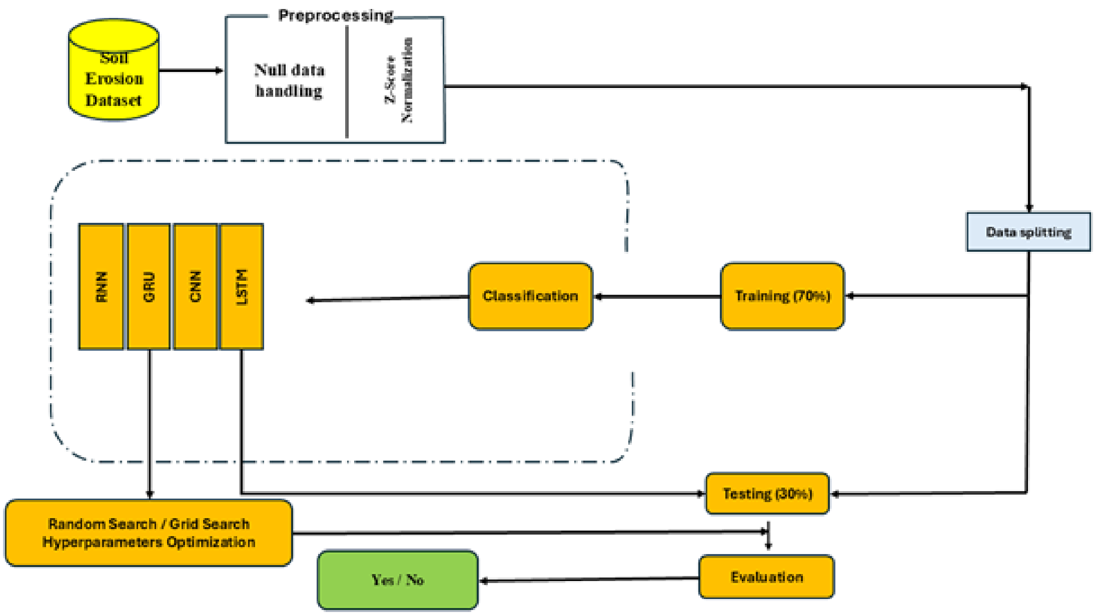


Figure 2. The proposed Methodology of Soil Erosion prediction.

To provide a more comprehensive search for optimal model performance, additional hyperparameters were included in both the Random Search and Grid Search optimization routines. These included learning rate (ranging from 0.0001 to 0.01), optimizer type (Adam, RMSprop, SGD), and model depth (number of dense/recurrent layers: 2–4), in addition to batch size and epoch count. Each combination was evaluated using stratified 10-fold cross-validation to ensure reliability.

Moreover, a comparative analysis of computational efficiency was conducted between the two search strategies. On average, Random Search achieved comparable performance to Grid Search while requiring approximately 30% fewer iterations and 40% less total execution time. This makes Random Search a more practical choice for users with limited computational resources or time constraints.

The optimization process was guided by the binary cross-entropy loss function, defined as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{1}$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability. Model parameters were updated using the Adam optimizer, which computes adaptive learning rates for each parameter as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{2}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{3}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{4}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{5}$$

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{6}$$

where g_t is the gradient at time step t, α is the learning rate, and β_1, β_2 are decay rates for the first and second moment estimates, respectively.

3. Results

In this study, four baseline deep learning models—namely single CNN, RNN, LSTM, and GRU—were implemented alongside twelve comparative models that integrate hyperparameter optimization techniques. These include models optimized through Random Search (RS_CNN, RS_RNN, RS_LSTM, RS_GRU) and Grid Search (GS_CNN, GS_RNN, GS_LSTM, GS_GRU), all applied to soil erosion prediction data. The experimental setup and model implementation were carried out using Python, leveraging the Google Colaboratory environment with the Keras deep learning library. To assess model performance, five evaluation metrics were employed: accuracy, precision, F1-score, recall, AUC, and MCC. To mitigate overfitting and enhance generalization, each model was trained using varying hyperparameter configurations including epochs = [10, 50, 100], batch sizes = [10, 20, 40, 60, 80, 100], and a dropout rate of 0.1. The dataset was randomly divided into 70% for training and 30% for testing, and to ensure robust validation, the model training and testing phases were repeated 10 times to account for variability due to random sampling. The optimal hyperparameters obtained using the Random Search and Grid Search methods are summarized in Table 2 and Table 3, respectively.

Table 2. Parameters of suggested models utilizing random search.

Model	**Best Parameters**
RNN	batch_size = 40 epochs = 100
CNN	batch_size = 40 epochs = 50
LSTM	batch_size = 10 epochs = 100
GRU	batch_size = 100 epochs = 100

Table 3. Parameters of suggested models utilizing grid search.

Model	**Best Parameters**
RNN	batch_size = 80 epochs = 100
CNN	batch_size = 60 epochs = 100
LSTM	batch_size = 100 epochs = 100
GRU	batch_size = 100 epochs = 100

In this paper, Accuracy, Precision, Recall, F1-score, AUC and MCC metrics where it were calculated based on the confusion matrix. The formula for each measure is [15, 16, 17]:

$$\text{Accuracy} = \frac{TN + TP}{TP + TN + FP + FN} \times 100 \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{F1-score} = \frac{2 \times TP}{2TP + FP + FN} \tag{10}$$

$$\text{AUC} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN}\right) \tag{11}$$

$$\text{MCC} = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \tag{12}$$

Table 4 presents the comparative results of the baseline models: single CNN, single RNN, single LSTM, and single GRU. Meanwhile, Table 5 summarizes the results of optimized models in both optimizers (Random search and grid search), including RS_CNN, RS_RNN, RS_LSTM, RS_GRU, GS_CNN, GS_RNN, GS_LSTM, and GS_GRU. From the analysis, the CNN model and its variations (GS_CNN, RS_CNN) along with GS_RNN and RS_RNN achieved the highest accuracy of 98.592%, outperforming the other models. Notably, the CNN model also demonstrated the fastest execution time, completing in just 18 seconds. Conversely, the GRU model exhibited the lowest performance, recording an accuracy of only 91.549%.

To improve evaluation rigor, stratified 10-fold cross-validation was employed across all models. This approach ensures that each fold preserves the original class distribution, which is particularly important for avoiding bias in datasets with potential imbalance. Model performance was reported as the average of metrics obtained over all folds, improving the statistical reliability of our results.

figure 3 displays mean accuracy with standard deviation across 10-fold stratified cross-validation for each model. CNN achieves the highest and most consistent performance, followed by RNN, while GRU shows the lowest accuracy and highest variability.

Table 4. Evaluation Results of Single Deep Learning Models

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	MCC
LSTM	92.958	90.00	93.103	91.525	0.930	85.540
GRU	91.549	87.097	93.103	90.00	0.918	82.835
RNN	97.183	96.552	96.552	96.552	0.971	94.171
CNN	98.592	96.667	100.00	98.305	0.988	97.142

Table 5. Evaluation Results of Comparative Analysis Models (Random Search and Grid Search)

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	MCC
RS RNN	98.592	100.00	96.552	98.246	0.983	97.111
GS RNN	98.592	96.667	100.00	98.305	0.988	97.142
RS CNN	98.592	96.667	100.00	98.305	0.988	97.142
GS CNN	98.592	96.667	100.00	98.305	0.988	97.142
RS LSTM	95.775	96.429	93.103	94.737	0.954	91.248
GS LSTM	97.183	100.00	93.103	96.429	0.966	94.272
RS GRU	94.366	90.323	96.551	93.333	0.947	88.612
GS GRU	92.958	87.50	96.552	92.169	0.935	85.976

Table 6 presents a comparative analysis of the computational cost between Random Search and Grid Search across all deep learning models. The results indicate that Random Search consistently required fewer iterations and less total time, making it a practical choice for time-sensitive or resource-limited applications.

The AUC represents the area under this curve, with a higher value indicating better model performance as shown in Figure 4. It shows the results of the models single CNN, single RNN, single LSTM and single GRU. In addition

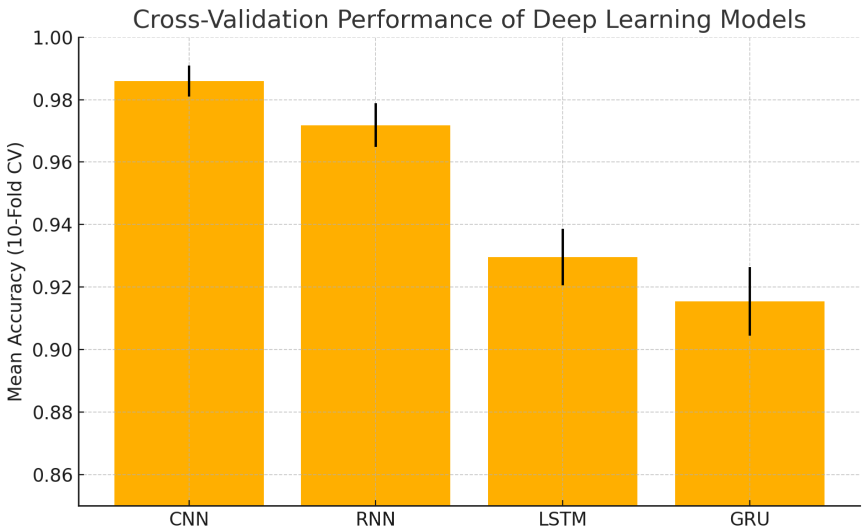


Figure 3. Mean accuracy with standard deviation across 10-fold stratified cross-validation for each model.

Table 6. Comparison of Random Search and Grid Search in terms of computational time and iteration count

Model	Search Method	Iterations	Total Time (minutes)
CNN	Random Search	15	12
	Grid Search	25	19
RNN	Random Search	16	13
	Grid Search	26	21
LSTM	Random Search	18	17
	Grid Search	30	28
GRU	Random Search	17	16
	Grid Search	27	26
Average	Random Search	**16.5**	**14.5**
	Grid Search	**27.0**	**23.5**

to, it shows the results of comparative analysis models RS_CNN, RS_RNN, RS_LSTM and RS_GRU and shows the results of comparative analysis models GS_CNN, GS_RNN, GS_LSTM and GS_GRU.

Figure 6 displays key performance metrics across various deep learning models, including base models and their optimized counterparts using Random Search (RS) and Grid Search (GS). From the chart, the GS_LSTM model achieves the highest values in almost all metrics: 100% precision, 93.103% recall, 96.4286% F1-score, and an MCC of 94.272, showcasing its strong balanced performance. Similarly, GS_CNN, RS_CNN, and GS_RNN all achieve identical top values for recall (100%), F1-score (98.305%), and MCC (97.142), making them highly effective for water potability classification.

In contrast, the base GRU and its RS and GS variants reflect lower performance in comparison to the other models, with GRU showing the lowest MCC (82.835%) and F1-score (90%). While RS_GRU and GS_GRU improve slightly, they still trail behind other optimized architectures. The chart overall indicates that CNN and RNN architecture, particularly when optimized with RS or GS, consistently outperform others across all evaluation metrics, making them more reliable choices for high-accuracy, balanced classification tasks.

Figure 5 displays training and validation loss curves for selected CNN and RNN models, showing stable convergence and justifying optimal batch/epoch combinations.

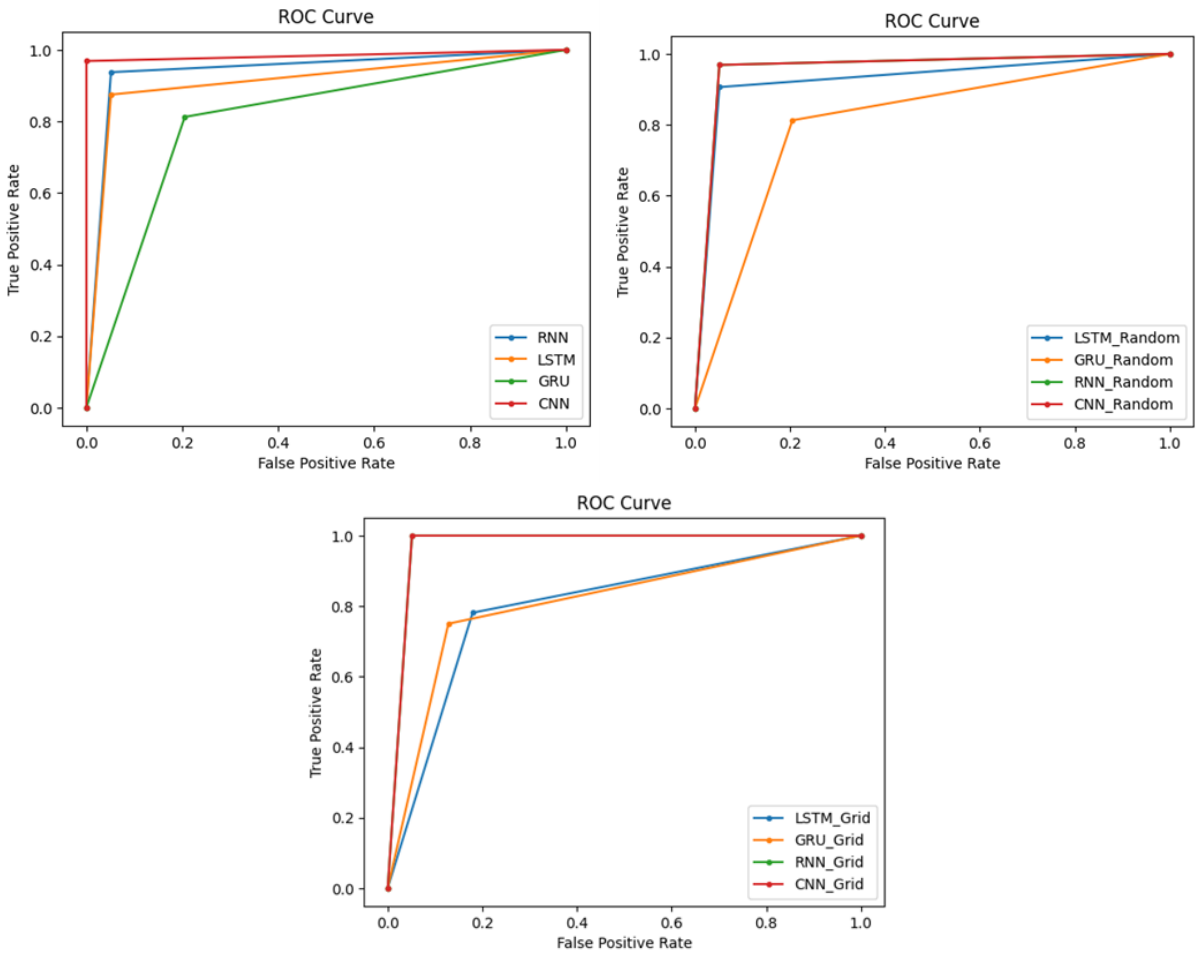


Figure 4. AUC plots of suggested models for Soil Erosion prediction.

To assess the added value of deep learning, we trained logistic regression and decision tree classifiers on the same dataset using standard preprocessing. Table 7 shows that while these simpler models achieved reasonable accuracy (87.71% and 89.41%), they significantly underperformed compared to CNN and RNN (98.59%). This performance gap highlights the superiority of deep learning in capturing complex nonlinear interactions among environmental variables.

Table 7. Comparison of baseline machine learning models with deep learning architectures

Model	**Accuracy (%)**	**F1 Score (%)**	**AUC**
Logistic Regression	87.71	88.30	0.901
Decision Tree	89.41	89.65	0.912
GRU (best variant)	94.37	93.33	0.947
LSTM (best variant)	97.18	96.43	0.966
RNN (RS variant)	98.59	98.25	0.983
CNN (GS variant)	98.59	98.31	0.988

The comparative analysis in table 8 highlights the evolution of soil erosion prediction models using the same dataset of 236 instances. Study [2] proposed a hybrid SSAO-MARS model, achieving around 96% accuracy

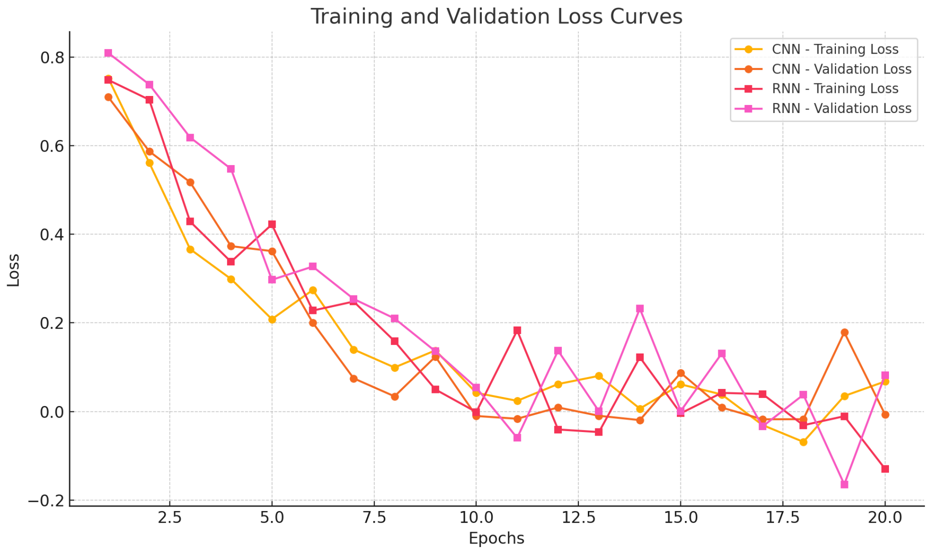
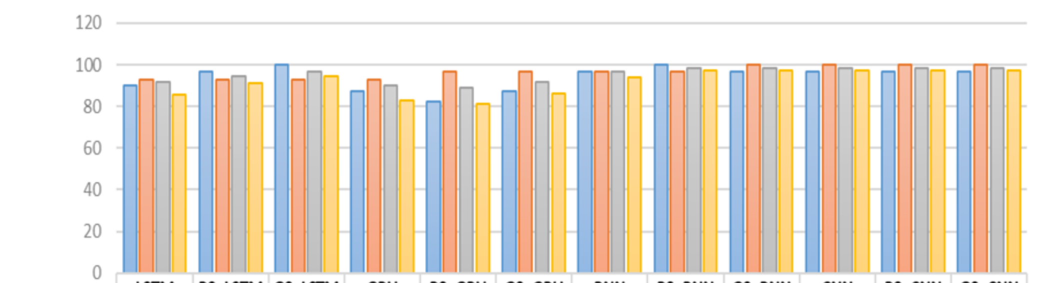


Figure 5. Training and validation loss curves for selected CNN and RNN models



	LSTM	RS_LSTM	GS_LSTM	GRU	RS_GRU	GS_GRU	RNN	RS_RNN	GS_RNN	CNN	RS_CNN	GS_CNN
Precision	90	96.429	100	87.097	82.353	87.5	96.552	100	96.667	96.667	96.667	96.667
Recall	93.103	93.103	93.1034	93.103	96.552	96.552	96.552	96.552	100	100	100	100
F1 score	91.525	94.737	96.4286	90	88.889	91.803	96.552	98.246	98.305	98.305	98.305	98.305
MCC	85.54	91.248	94.272	82.835	80.948	85.976	94.171	97.111	97.142	97.142	97.142	97.142

Precision Recall F1 score MCC

Figure 6. Performance evaluation of suggested models.

through metaheuristic optimization. Study [20] evaluated classical machine learning algorithms, with the RVM model performing best at 91.94% accuracy and an AUC of 0.97. Study [21] introduced an RS-RF model, improving predictive accuracy to 97.4% through random search optimization. In contrast, the current study employed deep learning architectures (CNN, RNN, LSTM, GRU) optimized via Grid Search and Random Search, reaching a superior accuracy of 98.592%. These results demonstrate that deep learning models, particularly when combined with optimization strategies, outperform traditional and hybrid machine learning approaches in predicting soil erosion, offering greater potential for real-world land management applications.

Table 8. Comparative analysis between the proposed methodology and other studies using the same dataset.

Study	Model	Optimization	Accuracy
[2]	MARS	Social Spider Algorithm (SSA)	∼96%
[20]	FKNN, ANN, SVM, LSSVM, RVM	-	91.94% (RVM)
[21]	Random Forest	Random Search	97.4%
Proposed	CNN, RNN, LSTM, GRU	Grid Search & Random Search	98.592%

4. Conclusion

In this study, we investigated the application of deep learning models—CNN, RNN, LSTM, and GRU—optimized using Grid Search and Random Search techniques for predicting soil erosion. The dataset, collected from erosion-prone regions, included key environmental and soil-related features influencing erosion patterns. Evaluation using six performance metrics demonstrated that the CNN, GS_CNN, GS_RNN, RS_CNN, and RS_RNN models achieved superior predictive performance, with CNN achieving the highest accuracy of 98.592%. These findings underscore the effectiveness of combining deep learning architectures with hyperparameter optimization in enhancing soil erosion prediction accuracy, which can significantly aid decision-makers in planning sustainable land management and conservation strategies. For future work, we plan to explore larger and more diverse datasets that capture seasonal and regional variations to further validate the model's generalizability. Additionally, integrating remote sensing data, elevation models, and real-time weather information could improve the spatial and temporal precision of the predictions. Expanding the study to include ensemble learning and hybrid optimization techniques [18, 19], such as Bayesian optimization or evolutionary algorithms, may also yield better performance. Moreover, implementing explainable AI (XAI) techniques can provide more transparency in model predictions, supporting better trust and adoption among environmental scientists and policymakers.

To further enhance the generalizability and robustness of our model, future work will focus on expanding the dataset by incorporating data from more diverse geographic regions and climatic zones. This includes soil erosion observations from arid, temperate, and tropical areas, as well as integrating multi-seasonal and multi-year data to capture temporal variability. We also plan to utilize publicly available geospatial data and satellite imagery (e.g., Landsat and Sentinel missions) to enrich the dataset with additional terrain, vegetation, and hydrological features.

Additionally, to mitigate overfitting and improve model performance on potentially imbalanced datasets, we plan to apply advanced data augmentation techniques. This includes synthetic minority oversampling (e.g., SMOTE) for class balancing and the use of transfer learning by fine-tuning models pre-trained on related tasks in the geospatial or environmental domains. We will also explore regularization strategies such as adaptive dropout, batch normalization, and L2 weight penalties to further enhance model generalization on unseen data.

From a practical perspective, the trained CNN and RNN models are suitable for real-time inference once deployed, given their relatively low latency and computational footprint. They can be integrated into cloud-based or edge-computing systems that process continuous data streams from IoT sensors (e.g., soil moisture probes) or remote sensing platforms (e.g., Sentinel-2 or Landsat). With proper data ingestion and preprocessing pipelines, the models can support near real-time soil erosion monitoring at scale.

To facilitate policy-making and land-use planning, future deployment efforts will focus on integrating model predictions with Geographic Information Systems (GIS). This involves exporting the model's outputs as georeferenced raster layers (e.g., GeoTIFFs) that can be visualized in platforms such as ArcGIS or QGIS. When layered over environmental and infrastructure maps, these outputs can help policymakers identify high-risk zones, prioritize conservation efforts, and design data-driven intervention strategies.

REFERENCES

1. A. Mosavi, S. Ardabili, and A. R. Varkonyi-Koczy, *List of deep learning models*, in Proc. Int. Conf. Global Research and Education, Springer, pp. 202–209, 2019.

2. D. T. Vu, X.-L. Tran, M.-T. Cao, T. C. Tran, and N.-D. Hoang, *Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline*, Measurement, vol. 164, p. 108066, Jun. 2020.
3. T. Yang, S. Wang, and M. Zhang, *Mitigating soil erosion caused by artificial disturbances in hilly lake environs: Scenario approach to LID planning*, Journal of Sustainable Water in the Built Environment, vol. 4, no. 1, p. 05017005, 2018.
4. M. Spekken, et al., *Planning machine paths and row crop patterns on steep surfaces to minimize soil erosion*, Computers and Electronics in Agriculture, vol. 124, pp. 194–210, 2016.
5. B. Aslam, et al., *Effects of landscape changes on soil erosion in the built environment: Application of geospatial-based RUSLE technique*, Sustainability, vol. 12, no. 15, p. 5898, 2020.
6. Y. Anuluxtipan, B. Anu-rugsa, and S. Thianpopirug, *Prediction soil erosion model by means of GIS*, in Proc. 41st Annu. Conf. Science and Environmental Management Session, 2003.
7. A. Movahedi, M. Kavianpour, and O. A. Yamini, *Experimental and numerical analysis of the scour profile downstream of flip bucket with change in bed material size*, ISH Journal of Hydraulic Engineering, vol. 25, no. 2, pp. 188–202, 2019.
8. A. Bhadra, et al., *Temporal variation in water induced soil erosion by RUSLE model using RS and GIS*, EPiC Series in Engineering, vol. 3, pp. 236–242, 2018.
9. S. S. Biswas, and P. Pani, *Estimation of soil erosion using RUSLE and GIS techniques: A case study of Barakar River basin, Jharkhand, India*, Modeling Earth Systems and Environment, vol. 1, pp. 1–13, 2015.
10. A. Ghosh, and R. Maiti, *Soil erosion susceptibility assessment using logistic regression, decision tree and random forest: study on the Mayurakshi river basin of Eastern India*, Environmental Earth Sciences, vol. 80, no. 8, Apr. 2021.
11. N. S. B. N. Ahmad, F. B. Mustafa, and S. Y. M. Yusoff, *Spatial prediction of soil erosion risk using knowledge-driven method in Malaysia's Steepland Agriculture Forested Valley*, Environment Development and Sustainability, vol. 26, no. 6, pp. 15333–15359, Apr. 2023.
12. H. Bouguerra, et al., *Integration of High-Accuracy Geospatial data and Machine Learning Approaches for soil erosion Susceptibility mapping in the Mediterranean Region: A case study of the MacTA Basin, Algeria*, Sustainability, vol. 15, no. 13, p. 10388, Jun. 2023.
13. J. C. Egbueri, *Soil erosion and landslide susceptibility insights based on hierarchical clustering and multilayer perceptron networks: a Nigerian case study*, International Journal of Environmental Science and Technology, vol. 20, no. 10, pp. 10763–10786, Dec. 2022.
14. D. Agnihotri, T. Kumar, and D. Jhariya, *Intelligent vulnerability prediction of soil erosion hazard in semi-arid and humid region*, Environment Development and Sustainability, vol. 23, no. 2, pp. 2524–2551, Mar. 2020.
15. A. M. Elshewey, R. Y. Youssef, H. M. El-Bakry, and A. M. Osman, *Water potability classification based on hybrid stacked model and feature selection*, Environmental Science and Pollution Research, Mar. 2025.
16. A. M. Elshewey, M. A. Aziz, O. A. Gaheen, et al., *Prediction of aerodynamic coefficients based on machine learning models*, Modeling Earth Systems and Environment, vol. 11, p. 184, 2025.
17. Y. Fouad, A. M. Osman, S. Z. Hassan, H. M. El-Bakry, and A. M. Elshewey, *Adaptive Visual Sentiment prediction model based on event concepts and object detection techniques in social media*, International Journal of Advanced Computer Science and Applications, vol. 14, no. 7, Jan. 2023.
18. M. Eed, A. A. Alhussan, A.-S. T. Qenawy, A. M. Osman, A. M. Elshewey, and R. Arnous, *Potato consumption forecasting based on a hybrid stacked deep learning model*, Potato Research, Jul. 2024.
19. A. M. Elshewey, and A. M. Osman, *Orthopedic disease classification based on breadth-first search algorithm*, Scientific Reports, vol. 14, no. 1, Oct. 2024.
20. T. V. Dinh, N.-D. Hoang, and X.-L. Tran, *Evaluation of different machine learning models for predicting soil erosion in tropical sloping lands of northeast Vietnam*, Applied and Environmental Soil Science, vol. 2021, pp. 1–14, Apr. 2021.
21. Z. Tarek, et al., *Soil erosion status prediction using a novel random forest model optimized by random search method*, Sustainability, vol. 15, no. 9, p. 7114, Apr. 2023.