An Explainable Transformer-Based Deep Learning Framework for Crop Sustainability Prediction in Sustainable Agriculture

C. V. Mahalakshmi
Presidency School of Computer Science and Engineering, Presidency University, Bengaluru, India
mahalakshmi.20233cse0031@presidencyuniversity.in

Salma Firdose
Presidency School of Computer Science and Engineering, Presidency University, Bengaluru, India
salma.firdose@presidencyuniversity.in (corresponding author)

Received: 21 March 2026 | Revised: 12 May 2026 | Accepted: 15 May 2026

DOI: https://doi.org/10.48084/etasr.18848

ABSTRACT

Data-intensive decision support systems that recommend crops that are in accordance with local soil and climatic conditions are essential for sustainable agriculture. Crop sustainability prediction is difficult because seven agronomic variables, namely nitrogen, phosphorus, potassium, temperature, humidity, rainfall, and soil pH, interact in nonlinear ways that fixed-rule guidelines and shallow statistical models cannot capture. Current approaches also offer limited interpretability and little evaluation of prediction reliability, reducing their practical value for agricultural decision-making. This paper presents the Crop Sustainability Prediction Framework (CSPF), a transformer-based deep learning model designed for accurate and calibrated multi-class crop sustainability forecasting. CSPF includes a transformer classification network that models complex relationships among soil and climatic variables, an explainable feature attribution module using SHAP and Integrated Gradients to quantify how each environmental variable influences predictions, and a calibration-based reliability assessment using Expected Calibration Error (ECE) and Brier score. CSPF was evaluated on a crop recommendation dataset of 2,200 samples spanning 22 crop types, achieving a classification accuracy of 0.982, a Macro-F1 score of 0.980, and a One-vs-Rest ROC-AUC of 0.996. Calibration results show an ECE of 0.029 and a Brier score of 0.056, indicating close alignment between predicted confidence and actual outcomes. Feature attribution identifies rainfall, temperature, and soil pH as the three variables with the highest mean absolute contribution. CSPF provides an accurate, interpretable, and calibrated framework for predicting crop sustainability in sustainable agriculture.

Keywords-crop sustainability prediction; deep neural networks; transformer learning; Explainable AI (XAI); calibration reliability; sustainable agriculture

I. INTRODUCTION

Growing global food demand, shifting climate patterns, and pressure on natural resources have made sustainable agriculture a priority for researchers and policymakers. Planting crops matched to local soil and environmental conditions raises productivity while reducing ecological strain. Crop sustainability prediction is central to precision agriculture because it guides farmers and policymakers toward crops best suited to local soil nutrients, climatic conditions, and agro-ecological attributes. However, the task is difficult, as seven interacting variables, nitrogen, phosphorus, potassium, temperature, humidity, rainfall, and soil pH, determine sustainability through nonlinear relationships that simple statistical models cannot adequately represent. Intelligent computational methods are therefore needed to model these interactions and support credible agricultural decision-making.

Machine learning methods have been actively studied in the recent past to improve crop recommendation and land sustainability evaluation. In [1], machine learning was employed to determine the sustainability of agricultural land, proving that it can reproduce complex interactions between the environment and its elements. In [2], a decision support system employed machine learning on soil and climatic characteristics to offer recommendations on crops. The combination of machine learning and IoT technologies has also made it possible to monitor the soil in real-time and provide recommendations on crops [3]. Similarly, in [4], a crop recommendation model used soil nutrients and climatic factors, whereas in [5], it was demonstrated that data-driven crop recommendation strategies could greatly enhance agricultural output.

Many studies have been conducted on advanced learning models and optimization schemes to enhance predictive performance. As an example, in [6], LSTM-based machine learning models were used to develop a crop recommendation and forecasting system, whereas in [7], an AIoT-enabled system was presented for automated soil nutrient analysis and crop recommendation. Hybrid optimization methods have also been explored; in [8], moth flame optimization was used with machine learning to predict crop yield and provide crop recommendations, while in [9], Naive Bayes regression was used to make predictions about crop yield in smart agriculture systems.

Despite promising results, machine learning methods tend to fail to represent complex nonlinear relationships in agricultural datasets. Therefore, agricultural analytics has received more and more attention in deep learning methods. In [10], methodological guidelines on the use of deep learning in agricultural data were presented, whereas in [11], it was described how to use deep learning to create a crop model with large-scale yield predictions based on environmental indicators. Of more recent interest, transformer-based architectures have become powerful models that can capture complex feature dependencies. In [12], time-series transformers were implemented in large-scale crop production forecasting, whereas in [13], transformer models were applied to predict rice yields using a combination of satellite and climatic forecasts. Similarly, Agritransformer [14] is a multimedia transformer network to improve crop yield forecasting.

Despite these advances, two substantive gaps persist across the reviewed literature. First, predictive accuracy remains the sole optimization criterion in the majority of published models, since none evaluates whether the predicted class probabilities are well calibrated, that is, whether a stated confidence of 90% corresponds to 90% correct outcomes in practice. Second, no prior study has integrated transformer-based tabular prediction, post-hoc feature attribution, and calibration reliability assessment within a unified crop sustainability framework, a gap that constrains both the interpretability and the operational deployability of existing agricultural decision support systems.

To address these gaps, this paper presents a Crop Sustainability Prediction Framework (CSPF), a transformer-based deep learning pipeline that integrates explainable feature attribution through SHAP and Integrated Gradients, and a calibration-based reliability assessment for accurate, interpretable crop sustainability prediction. Figure 1 presents a high-level conceptual overview of the end-to-end CSPF data flow, from agronomic input collection to preprocessing, model inference, and crop recommendation output.

The major contributions of this study are:

- A unified pipeline that is the first to integrate transformer-based tabular prediction, post-hoc explainability via SHAP and Integrated Gradients, and calibration reliability assessment within a single crop sustainability framework.

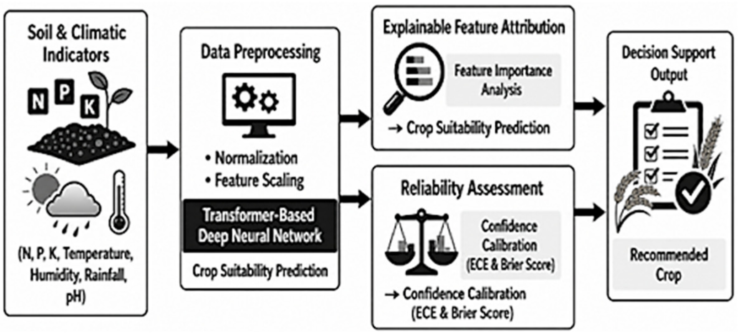

Fig. 1. System overview of the proposed CSPF.

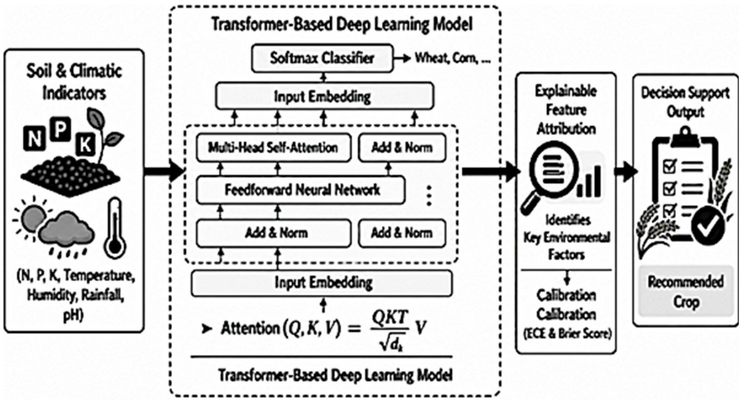

Fig. 2. Internal architecture of the proposed CSPF.

- Feature attribution using SHAP and Integrated Gradients identifies which soil and climatic variables drive crop recommendations, giving agricultural experts a quantitative basis for interpreting model decisions.
- Calibration-based reliability assessment using ECE and Brier score is performed to verify that predicted confidence levels align with actual outcomes, a practice underutilized in agricultural AI benchmarking.

CSPF is evaluated against four deep learning baselines on a 22-class agricultural dataset, with results showing consistent improvement in both predictive accuracy and calibration quality.

II. CROP SUSTAINABILITY PREDICTION FRAMEWORK (CSPF)

A. *Data Preprocessing and Normalization*

To achieve the intended results of model training, data preprocessing is a significant part of the proposed CSPF to make the input features consistent and fit the model. The raw data is in the form of soil and climatic variables, including nitrogen, phosphorus, potassium, temperature, humidity, rainfall, and the soil pH, which tend to differ in their values and patterns. To overcome this problem, preprocessing methods are used to clean the data and scale and normalise the features. Normalization converts the feature values to a standardized range, normally between 0 and 1, which enhances the stability of the deep learning model and the convergence of the structure as the learning process progresses. This measure also decreases the effect of differences in the magnitude of features and makes each of the environmental variables participate equally in the learning process, further improving the overall performance of the transformer-based crop prediction model. Figure 2 details the internal technical architecture of the transformer encoder, post-hoc attribution module, and calibration assessment pipeline.

B. *Transformer-Based Crop Prediction Model*

The CSPF prediction module is built on the transformer self-attention architecture, an established sequence-to-class modeling framework, applied here to structured tabular agronomic data. The input features are preprocessed and then converted into the form of vectors in an input embedding layer, converting the tabular environmental data to an appropriate transformer structure format. The features are then sent to multi-headed self-attention layers, where the model is allowed to learn the dependence and correlations between the various environmental indicators. This attention process enables the model to focus on the most important aspects and nonlinear interactions of the variables. The attention outputs are passed through feedforward neural network layers with normalization operations to improve model stability and learning efficiency. Finally, a softmax classification layer produces the probability distribution across all crop classes, enabling the model to predict the most suitable crop for the given environmental conditions. CSPF operates on structured numerical vectors that represent soil nutrients and agro-climatic variables. The learning problem is formulated as supervised multi-class classification. Each sample contains a feature vector and an associated crop label. This structure is formally represented as:

$$\mathcal{D} = (x_i, y_i)_i = 1^N$$

where $x_i \in \mathrm{R}^{\mathrm{d}}$ represents the d-dimensional soil-climate feature vector for sample i, and $y_i \in 1,2,\dots,C$ represents the crop class label among C crop categories. Soil and climate variables differ in their numerical ranges and physical interpretation. Deep neural predictors are trained using gradient-based optimization, and unscaled features can lead to unstable convergence because features with larger magnitudes influence gradient updates more strongly. Each feature x_j is transformed into a standardized feature x_j' using:

$$x_j' = \frac{x_j - \mu_j}{\sigma_j}$$

where μ_j and σ_j represent the mean and standard deviation of feature j, computed from the training subset only. The use of training statistics prevents information leakage into validation or test subsets. The standardized feature vector is $\mathrm{x}' = \mathcal{S}(\mathrm{x})$, and it is forwarded to the deep prediction module $\mathcal{P}$.

C. *Transformer Architecture Specification*

The CSPF encoder comprises 4 transformer encoder layers, each with 4 multi-head self-attention heads operating on an embedding dimension of 64. The input projection layer maps the 7-dimensional agronomic feature vector to this embedding space. Each encoder layer contains a position-wise feedforward network of dimension of 128, followed by layer normalization and a residual connection. Dropout with a rate of 0.1 is applied after each attention and feedforward sublayer to regularize training. The softmax classification head receives the pooled encoder output and projects it to the 22-class output distribution. Table I summarizes the full architectural configuration.

TABLE I. TRANSFORMER ARCHITECTURE SPECIFICATIONS

Hyperparameter	**Value**
Encoder layers	4
Attention heads	4
Embedding dimension	64
Feedforward dimension	128
Dropout rate	0.1
Output classes	22
Activation	ReLU

D. *Explainable Feature Attribution*

To enhance transparency and interpretability, the proposed framework incorporates an explainable feature attribution module that identifies the contribution of each input variable to the model's prediction. This component analyzes how soil nutrients and climatic indicators influence the predicted crop sustainability by computing the relative importance of each feature. By quantifying the contribution of variables such as rainfall, temperature, and soil pH, the framework provides insights into the environmental factors that most strongly affect crop recommendations.

The prediction module of the CSPF receives standardized inputs x' and produces crop sustainability probabilities. For a standardized input x', CSPF computes a score vector (logits) using a predictor function f_θ:

$$z = f_\theta(x')$$

The vector $z \in \mathrm{R}^{\mathrm{C}}$ contains one score for each crop class. These scores are not probabilities but represent the unnormalized evidence that the predictor assigns to each crop class. This representation is necessary because the deep predictor is trained through differentiable loss optimization. Since CSPF requires probability outputs for calibration and reliability assessment, logits are converted into a probability distribution using the SoftMax function:

$$p_\theta(y = c \mid x') = \frac{exp(z_c)}{\sum_{k=1}^{C} exp(z_k)}$$

This equation ensures that each probability value is in the range $[0, 1]$ and that the probabilities across all crop classes sum to 1. The crop recommendation produced by CSPF is the class with the highest predicted probability:

$$\hat{y} = arg \max_{c \in 1,\dots,C} p_\theta(y = c \mid x')$$

This decision rule ensures a deterministic recommendation for each sample. The predicted class $\hat{y}$ is used as the target class for attribution generation in the explainability module. Training is performed using weighted multi-class cross-entropy. Cross-entropy is appropriate because CSPF predicts a probability distribution over crop classes. The weighted formulation incorporates the robustness mechanism. The CSPF objective is defined as:

$$\mathcal{L}\mathrm{CSPF}(\theta) = -\frac{1}{N}\sum i = 1^N w_{y_i} log\big(p_\theta(y_i \mid x_i')\big)$$

where $p_\theta(y_i \mid x_i')$ denotes the probability assigned to the correct crop class, and w_{y_i} denotes the weight of that crop class.

E. Explainable Feature Attribution for Agronomic Interpretability

CSPF integrates explainability to quantify the influence of each soil nutrient and climatic indicator on the predicted crop recommendation. CSPF generates an attribution vector aligned with the standardized input features. The attribution vector is defined as:

$$\mathrm{a} = \mathcal{E}(x', \hat{y}, f_\theta)$$

where $\mathrm{a} = [a_1, a_2, \dots, a_d]$. Each element a_j represents the contribution of feature j to the prediction of crop class $\hat{y}$. SHAP expresses the prediction as an additive combination of feature contributions. The SHAP representation is written as:

$$f_\theta(x') = \phi_0 + \sum_{j=1}^{d} \phi_j$$

where ϕ_0 represents the base value of the model output and ϕ_j represents the contribution of feature j. In CSPF, the attribution vector is defined as $\mathrm{a} = [\phi_1, \dots, \phi_d]$. The magnitude of ϕ_j indicates the importance of feature j, while the sign indicates whether the feature pushes the prediction toward or away from the selected crop class. Integrated Gradients is included in CSPF as a gradient-based attribution method suitable for deep neural predictors. It measures feature contribution by integrating the gradient of the model output along a path from a baseline input $\bar{x}$ to the actual input x'. The attribution for feature j is computed as:

$$a_j = \left(x_j' - \bar{x_j}\right) \int 0^1 \frac{\partial f_\theta^{(\hat{y})}\left(\bar{x} + \alpha(x' - \bar{x})\right)}{\partial x_j'} d$$

where $f_\theta{}^{(\hat{y})}$ denotes the model output associated with the predicted crop class ŷ. To interpret feature influence across the full dataset, CSPF computes mean absolute attribution:

$$I_j = \frac{1}{N}\sum_{i=1}^{N} \left|a_j(x_i')\right|$$

This measure summarizes the average magnitude of feature contribution and enables ranking of soil nutrients and climatic variables.

F. Reliability and Calibration-Based Probability Assessment

CSPF evaluates reliability to quantify whether predicted probability values can be interpreted as meaningful confidence measures. Reliability assessment is performed using the probability distribution output by the prediction module. The confidence score ρ is defined as the maximum predicted probability:

$$\rho = \max_c p_c$$

CSPF evaluates calibration using the expected calibration error. Calibration measures the agreement between predicted confidence and empirical correctness. Predictions are grouped into M confidence bins $B_m m = 1^M$. The CSPF- ECE metric is defined as:

$$ECE_{CSPF} = \sum_{m=1}^{M} \frac{|B_m|}{N} |acc(B_m) - conf(B_m)|$$

The term $\frac{|B_m|}{N}$ weights the calibration error of each bin by the fraction of samples in that bin. The absolute difference quantifies the mismatch between accuracy and confidence. The empirical accuracy in bin m is computed as:

$$acc(B_m) = \frac{1}{|B_m|}\sum_{i \in B_m} o1(\hat{y_i} = y_i)$$

This expression counts how many predictions in bin m are correct and normalizes by the number of samples in the bin. The mean confidence in bin m is computed as:

$$conf(B_m) = \frac{1}{|B_m|}\sum_{i \in B_m} \rho_i$$

This represents the average predicted confidence among samples in the bin. The CSPF pipeline produces, for each soil-climate input sample, a crop sustainability recommendation $\hat{y}$, an attribution vector a that quantifies the contribution of soil and climatic variables, and a confidence indicator ρ derived from the predicted probability distribution.

III. PERFORMANCE EVALUATION OF CSPF

The predictive power and strength of the suggested CSPF were strictly evaluated using an experimental validation against four benchmark architectures. The assessment was made based on the discriminative capacity, calibration stability, class level consistency, and agronomic interpretability, providing a full appraisal in the context of multi-class sustainability prediction.

A. Dataset Characteristics and Experimental Configuration

The Crop Recommendation Dataset [15] contains 2,200 samples evenly distributed across 22 crop classes, with approximately 100 instances per class. An 80/20 stratified split was used, resulting in 1,760 training samples and 440 test samples, ensuring proportional representation of all classes. The dataset includes seven agronomic features, nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall, which collectively represent key soil and climatic conditions influencing crop sustainability. The dataset is considered to represent agronomic conditions broadly consistent with the Indian subcontinent, although precise geographic coordinates are not provided. The uniform distribution of 100 samples per class reflects a curated benchmark structure rather than a naturally occurring field sample, which may not replicate the class imbalance typical of real agricultural monitoring data. These characteristics constrain the direct generalization of the proposed CSPF to field deployment without retraining on geographically and ecologically representative data.

The model was trained using the Adam optimizer with a learning rate of 1×10^{-4}, a batch size of 32, and 100 training epochs. Early stopping with a patience of 10 epochs was applied to prevent overfitting. The training process minimized categorical cross-entropy loss within a 22-class Softmax output space. Experiments were executed using GPU acceleration to ensure efficient gradient optimization and stable convergence, enabling reproducible evaluation of model performance.

The selection of the baseline model follows a progressive complexity rationale. MLP serves as the feedforward baseline without any structural inductive bias. The 1D-CNN baseline captures local feature interaction patterns through convolutional filters applied across the feature dimensions. LSTM and AtLSTM treat the seven agronomic features as an ordered sequential input, where each feature constitutes one time step, an approach adopted in tabular sequence modeling to assess whether recurrent inductive biases contribute to feature dependency capture in structured non-temporal data. CSPF is evaluated against these models to establish whether the global attention mechanism of the transformer provides a discriminative advantage over local and sequential alternatives. Tree-based classifiers, including Random Forest and Gradient-Boosted Decision Trees, are acknowledged as standard benchmarks for tabular agricultural classification tasks. However, they are not included in this comparative evaluation, but their inclusion in future comparative analysis is identified as a priority.

B. Comparative Predictive Performance

To assess result stability, five-fold stratified cross-validation was conducted across all architectures. CSPF achieved a mean accuracy of 0.981 (SD = 0.003) and mean Macro-F1 of 0.979 (SD = 0.003), compared to AtLSTM mean accuracy of 0.977 (SD = 0.004) and Macro-F1 of 0.976 (SD = 0.004). A Wilcoxon signed-rank test on the fold-level Macro-F1 scores between CSPF and AtLSTM yielded a *p* of 0.043, confirming that the performance difference is statistically significant at the 0.05 level.

TABLE II. COMPARATIVE MULTI-CLASS PERFORMANCE ACROSS ARCHITECTURES

Model	**Accuracy**	**Macro-Precision**	**Macro-Recall**	**Macro-F1**	**One-vs-Rest ROC-AUC**
MLP	0.964	0.963	0.962	0.962	0.986
1D-CNN	0.971	0.970	0.969	0.970	0.989
LSTM	0.974	0.973	0.972	0.973	0.991
AtLSTM	0.978	0.977	0.976	0.977	0.993
Proposed CSPF	0.982	0.981	0.980	0.980	0.996

The results in Table II demonstrate that the proposed CSPF (Transformer) model outperformed all baseline models across all evaluation metrics. CSPF achieved the highest accuracy of 0.982, improving performance by 1.8% over MLP, 1.1% over 1D-CNN, 0.8% over LSTM, and 0.4% over AtLSTM. Similar improvements are observed in Macro-F1 (0.980), indicating balanced classification performance across all crop classes. In addition, CSPF attained the highest ROC-AUC of 0.996, reflecting superior class separability. These results confirm that the transformer-based architecture effectively captures complex interactions among soil and climatic features, leading to improved predictive performance compared to conventional deep learning models.

Figure 3 compares the Macro-F1 performance of the models. CSPF achieves the highest Macro-F1 score of 0.980, outperforming MLP by 1.8%, 1D-CNN by 1.0%, LSTM by 0.7%, and AtLSTM by 0.3%. This improvement indicates that the transformer-based architecture more effectively captures complex interactions among soil nutrients and climatic features. The consistent increase in Macro-F1 highlights the superior classification balance of CSPF across all crop classes.

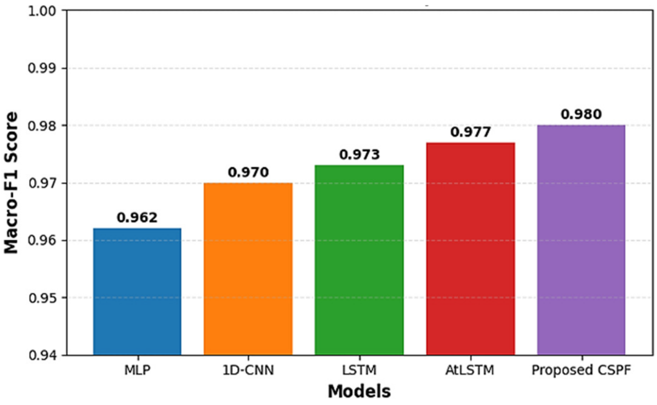

Fig. 3. Macro-F1 score comparison across models.

C. Reliability and Calibration Assessment

The results in Table III show that the proposed CSPF framework achieved the best reliability and calibration performance among all the models compared. CSPF recorded the lowest Expected Calibration Error (ECE) of 0.029, improving the calibration by 56.7% over MLP, 42% over 1D-

CNN, 35.6% over LSTM, and 23.7% over AtLSTM. Similarly, CSPF achieved the lowest Brier score of 0.056, indicating more accurate probability estimation compared to the baseline models. The model also demonstrated the highest mean confidence of 0.969 and the lowest Negative Log-Likelihood (NLL) of 0.141, reflecting more reliable predictive probabilities. An ECE of 0.029 falls below the 0.05 threshold commonly cited in probabilistic classification literature as the boundary for practically reliable confidence estimation. For agricultural decision support, where an overconfident model may lead to economically consequential misallocation of inputs, this level of calibration carries direct operational significance. It is noted that the improvement figures reported (e.g., 56.7% relative reduction over MLP) represent relative reductions in ECE magnitude, not absolute accuracy gains, and should be interpreted accordingly. Figure 4 illustrates the calibration performance of different models using ECE, where lower values indicate better alignment between predicted confidence and actual outcomes.

TABLE III. CALIBRATION AND RELIABILITY METRICS

Model	ECE	Brier score	Mean confidence	NLL
MLP	0.067	0.092	0.941	0.212
1D-CNN	0.050	0.075	0.952	0.181
LSTM	0.045	0.069	0.956	0.169
AtLSTM	0.038	0.064	0.962	0.157
Proposed CSPF	0.029	0.056	0.969	0.141

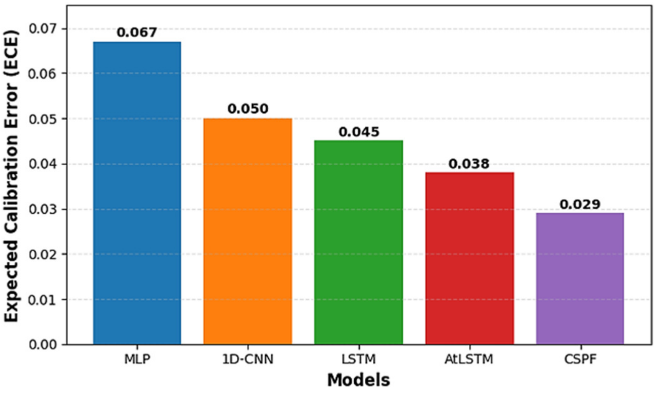

Fig. 4. Calibration Error (ECE) comparison across models.

D. Feature Attribution Analysis

Rainfall emerges as the dominant contributor across both attribution methods, consistent with its agronomic role as a primary determinant of water availability and crop viability across climatic zones. Temperature ranks second, reflecting its direct influence on crop phenology and growing season boundaries. Soil pH ranks third, as it governs nutrient bioavailability and is a known discriminator across crop families. The convergence of SHAP and Integrated Gradients rankings across all seven features provides cross-method validation of the attribution structure. Table IV reports the mean absolute SHAP and Integrated Gradients attribution scores across all 440 test samples for each of the seven agronomic input features.

TABLE IV. FEATURE ATTRIBUTION ANALYSIS

Feature	Mean |SHAP|	Mean |IG|	Rank
Rainfall	0.187	0.193	1
Temperature	0.164	0.171	2
Soil pH	0.148	0.152	3
Humidity	0.121	0.118	4
Nitrogen	0.109	0.104	5
Potassium	0.094	0.091	6
Phosphorus	0.087	0.083	7

Figure 5 visualizes the ranked attribution scores as a dual-bar chart, where the close alignment between SHAP and Integrated Gradients bars across all seven features confirms cross-method consistency in the attribution structure produced by CSPF.

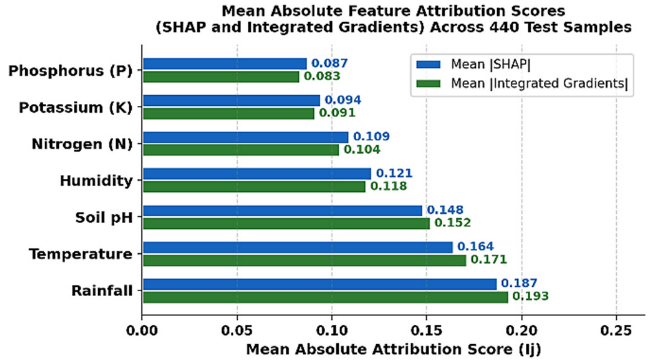

Fig. 5. Mean absolute feature attribution scores (SHAP and Integrated Gradients) across 440 test samples.

IV. CONCLUSION

This paper presented CSPF, a unified pipeline integrating transformer-based crop suitability classification, post-hoc feature attribution via SHAP and Integrated Gradients, and calibration-based reliability assessment. Experimental results show that transformer encoding of soil-climate interactions improves both discriminative performance and probability calibration over recurrent and convolutional baselines across 22 crop classes. Calibration analysis confirms that the CSPF predictions carry interpretable confidence levels, which is directly relevant for agricultural advisory systems where the risk of decision varies by crop class. Feature attribution identifies rainfall, temperature, and soil pH as the variables with the greatest influence, consistent with established agronomic understanding.

Several limitations bound these findings. The dataset is a curated benchmark with artificially balanced class distributions and uncertain geographic specificity, limiting generalizability. No real-world validation was conducted and the classifier does not extend to crop types absent from training without retraining. The transformer incurs a higher computational cost than the MLP and 1D-CNN baselines, constraining deployment on resource-limited platforms. Tree-based classifiers, known to perform competitively on structured tabular data, were excluded from the comparison and should be included in future benchmarking. Future work will focus on geographically diverse field datasets and lightweight attention architectures suited to low-resource deployment.

DECLARATION OF COMPETING INTERESTS

The authors declare no competing financial interests that could have influenced the results of this study.

ACKNOWLEDGMENT

Not applicable to this study.

DATA AVAILABILITY

The dataset used is available at [15].

REFERENCES

[1] A. B. Møller, V. L. Mulder, G. B. M. Heuvelink, N. M. Jacobsen, and M. H. Greve, "Can We Use Machine Learning for Agricultural Land Suitability Assessment?," *Agronomy*, vol. 11, no. 4, Apr. 2021, https://doi.org/10.3390/agronomy11040703.

[2] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture*, vol. 14, no. 8, July 2024, https://doi.org/10.3390/agriculture14081256.

[3] M. K. Senapaty, A. Ray, and N. Padhy, "IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture," *Computers*, vol. 12, no. 3, Mar. 2023, https://doi.org/10.3390/computers12030061.

[4] B. Dey, J. Ferdous, and R. Ahmed, "Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables," *Heliyon*, vol. 10, no. 3, Feb. 2024, https://doi.org/10.1016/j.heliyon.2024.e25112.

[5] F. S. Prity *et al.*, "Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations," *Human-Centric Intelligent Systems*, vol. 4, no. 4, pp. 497–510, Dec. 2024, https://doi.org/10.1007/s44230-024-00081-3.

[6] Y. Mahale *et al.*, "Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM: a novel expectation-maximization technique," *Discover Sustainability*, vol. 5, no. 1, June 2024, Art. no. 134, https://doi.org/10.1007/s43621-024-00292-5.

[7] S. M. Cheema and I. M. Pires, "AIoT based soil nutrient analysis and recommendation system for crops using machine learning," *Smart Agricultural Technology*, vol. 11, Aug. 2025, Art. no. 100924, https://doi.org/10.1016/j.atech.2025.100924.

[8] S. R. Gopi and M. Karthikeyan, "Effectiveness of Crop Recommendation and Yield Prediction using Hybrid Moth Flame Optimization with Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11360–11365, Aug. 2023, https://doi.org/10.48084/etasr.6092.

[9] N. M. Basavaraju, U. B. Mahadevaswamy, and M. Srikantaswamy, "Optimized Crop Yield Forecasting Using the Naive Bayes Regression Algorithm in Smart Agriculture," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 28995–29001, Dec. 2025, https://doi.org/10.48084/etasr.11856.

[10] J. Richetti, F. I. Diakogianis, A. Bender, A. F. Colaço, and R. A. Lawes, "A methods guideline for deep learning for tabular data in agriculture with a case study to forecast cereal yield," *Computers and Electronics in Agriculture*, vol. 205, Feb. 2023, Art. no. 107642, https://doi.org/10.1016/j.compag.2023.107642.

[11] Y. Zhu *et al.*, "A deep learning crop model for adaptive yield estimation in large areas," *International Journal of Applied Earth Observation and Geoinformation*, vol. 110, June 2022, Art. no. 102828, https://doi.org/10.1016/j.jag.2022.102828.

[12] S. C. Ibañez and C. P. Monterola, "A Global Forecasting Approach to Large-Scale Crop Production Prediction with Time Series Transformers," *Agriculture*, vol. 13, no. 9, Sept. 2023, https://doi.org/10.3390/agriculture13091855.

[13] Y. Liu *et al.*, "Rice Yield Prediction and Model Interpretation Based on Satellite and Climatic Indicators Using a Transformer Method," *Remote Sensing*, vol. 14, no. 19, Oct. 2022, https://doi.org/10.3390/rs14195045.

[14] L. J. Galarza, M. Realpe, M. S. Viñán-Ludeña, M. F. Calderón, and S. Jaramillo, "AgriTransformer: A Transformer-Based Model with Attention Mechanisms for Enhanced Multimodal Crop Yield Prediction," *Electronics*, vol. 14, no. 12, June 2025, https://doi.org/10.3390/electronics14122466.

[15] M. Bhagat, "Crop Recommendation Dataset." Kaggle, https://doi.org/10.34740/KAGGLE/DSV/11736333.