DMPedia Lecture Notes in Computer Science & Engineering
Volume 1, Issue 2, pp.419-429, 2026
ICICCS-2026

Crop Yield Prediction Using Remote Sensing Space Convolutional Attention Machine Learning with Potter Optimisation Algorithm for Independent Variable Selection

Ramu Vankudoth[1*], Sivani Vegi[2], N.V Madhu Bindu[3], R. Sravani[4], P. Praveena[5], E Varnika[6]

[1]Department of Computer Science & Engineering-Data Science, Malla Reddy (MR)Deemed to be University, Hyderabad, India

[2,] Department of Information Technology, Swarnandhra College of Engineering and Technology, Seetharampuram, India

[3]Department of Computer Science and Engineering, Koneru Lakshamiah Education Foundation, Guntur, India

[4]Department of Computer Science & Engineering – AIML, MLR Institute of Technology, Secunderabad, India

[5]Department of Computer Science & Engineering – CS, Geethanjali College of Engineering & Technology, AP, India

[6]Department of Computer Science, Vaagdevi Degree and PG College, Hanamakonda, India

ramuds@mrec.ac.in[1], vsivani.it@swarnandhra.ac.in[2], sravani@mlrit.ac.in[4], pulusugantiraveena@gmail.com[5], dasarivarnika@gmail.com[6]

*Corresponding author: Dr Ramu Vankudoth(ramuds@mrec.ac.in)

ABSTRACT

Accurate crop yield prediction is crucial for food security planning, resource allocation, and agricultural policy formulation. This paper proposes a novel hybrid deep learning framework, the Space Convolutional Attention Machine Learning (SCAML) model, that integrates multi-source remote sensing imagery with a biologically-inspired Potter Optimisation Algorithm (POA) for principled independent variable selection. The proposed model combines a dual-branch spatial-spectral Convolutional Neural Network (CNN) for local feature extraction with a multi-head self-attention transformer for temporal dependency modelling, culminating in a Space Convolutional Attention Module (SCAM) that fuses both representations. The Potter Optimisation Algorithm, inspired by the nest-building behaviour of potter wasps (Eumenes spp.), performs wrapper-based feature selection to identify the most informative spectral indices and ancillary variables, significantly reducing dimensionality while preserving predictive power. Experiments conducted on multi-season rice, wheat, and maize datasets across five agroclimatic zones demonstrate that SCAML-POA achieves a Root Mean Square Error (RMSE) of 0.312 t/ha, a Mean Absolute Percentage Error (MAPE) of 3.47%, and an R^2 of 0.964, outperforming baseline CNN, LSTM, and Random Forest regressors. The results affirm that the synergy between optimised variable selection and spatiotemporal attention learning can substantially improve crop yield forecasting accuracy.

***Keywords**:* Crop Yield Prediction, Remote Sensing, Space Convolutional Attention, Potter Optimisation Algorithm, Feature Selection, Deep Learning, NDVI, Transformer, Precision Agriculture

1. Introduction

Global food systems face unprecedented pressure from a combination of population growth, climate variability, and land degradation. Reliable prediction of crop yield at regional and national scales is therefore an indispensable component of strategic food systems management. Traditional crop simulation models (e.g., DSSAT, APSIM) rely heavily on accurate field-measured inputs that are difficult to obtain at scale. At the same time, empirical statistical approaches lack the representational capacity to capture nonlinear crop-environment interactions.

Proceedings of ICICCS-2026 (Poster Papers), DMPedia Lecture Notes in Computer Science & Engineering.
ISBN: 978-81-993813-1-5

Satellite remote sensing has emerged as a transformative data source for large-area crop monitoring, providing spatially continuous, temporally dense observations of canopy reflectance, vegetative vigour, and moisture status. Indices such as the Normalised Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI), and Land Surface Temperature (LST) have been extensively correlated with final yield. However, the high dimensionality of multi-spectral and hyperspectral imagery, combined with inherent noise from atmospheric interference and cloud contamination, necessitates principled feature selection before model training. Deep learning methods, notably CNNs and Recurrent Neural Networks (RNNs), have demonstrated superior capacity to extract hierarchical spatial and temporal patterns from remote sensing time series. Nevertheless, naive application of these models without feature selection often leads to overfitting, computational inefficiency, and reduced interpretability. Meta-heuristic optimisation algorithms provide an effective wrapper-based mechanism for identifying the minimal yet maximally informative subset of input variables. In this work, we introduce the Potter Optimisation Algorithm (POA), a novel swarm-intelligence metaheuristic modelled on the mud-collection, nest-chamber selection, and prey-loading behaviours of solitary potter wasps. POA exhibits strong global exploration through stochastic chamber-site selection and exploits promising regions via iterative mud-layering refinement, making it well-suited to high-dimensional binary feature selection tasks. The core contribution of this paper is therefore threefold: (i) the formulation of a Space Convolutional Attention Module (SCAM) that jointly captures local spatial context and long-range spectral attention within a unified deep learning architecture; (ii) the development of the Potter Optimization Algorithm for principled independent variable selection from multi-source remote sensing data; and (iii) an extensive experimental validation demonstrating state-of-the-art crop yield prediction performance across multiple crops and agroclimatic zones.

2. Related Work

2.1 Remote Sensing for Crop Yield Prediction

The application of satellite imagery to crop yield modelling spans several decades. Early studies established empirical relationships between NDVI and grain yield for wheat and maize. Subsequent work extended these approaches to multi-temporal analysis, recognising that the trajectory of vegetation indices through the growing season carries more predictive information than any single observation. Sentinel-2 and Landsat-8/9 have become dominant data sources due to their high spatial resolution (10–30 m) and frequent revisit times (5–16 days), enabling detailed within-field yield mapping.

2.2 Machine Learning and Deep Learning Approaches

Support Vector Regression (SVR), Random Forests (RF), and Gradient Boosting Machines (GBM) were among the first ML methods applied to yield prediction using remote sensing features. These methods handle tabular spectral indices well but struggle with raw image sequences. Convolutional architectures process spatial patterns directly from reflectance images, while Long Short-Term Memory (LSTM) networks model trajectories during the growing season. Hybrid CNN-LSTM models have shown improved performance, though they typically treat spatial and temporal feature extraction independently without explicit attention mechanisms. Transformer-based self-attention models, originally developed for natural language processing, have recently been adapted for satellite time-series analysis. The Temporal Attention Encoder (TAE) and SATelite Image Time Series (SITS) transformer demonstrate that multi-head attention over temporal sequences substantially improves classification and regression accuracy. Our SCAM module extends this paradigm by integrating spatial convolutional features into the attention computation, enabling the model to weight both location and time simultaneously.

2.3 Feature Selection via Metaheuristics

Wrapper-based feature selection using metaheuristics has proven effective in remote sensing applications, where the number of candidate features (spectral bands, indices, phenological metrics) can exceed several hundred. Particle Swarm Optimisation (PSO), Genetic Algorithms (GA), Grey Wolf Optimiser (GWO), and Whale Optimisation Algorithm (WOA) have all been applied to this problem. The proposed POA introduces distinct behavioural phases that provide a favourable balance between exploration and exploitation, and incorporates a novel chamber-selection mechanism analogous to multi-niche population management, reducing premature convergence.

3. Area and Dataset

3.1. Study Region

The study encompasses five agroclimatic zones across the Indian subcontinent: (1) the Indo-Gangetic Plains (wheat, Rabi season), (2) the Deccan Plateau (maize, Kharif season), (3) the Coastal Andhra region (rice, Kharif and Rabi seasons), (4) the Punjab-Haryana belt (rice-wheat cropping system), and (5) the North-Eastern Hill region (mixed cereals). Together, these zones capture a diversity of soils, irrigation regimes, and climatic patterns representative of South Asian agricultural systems.

3.2. Remote Sensing Data

Multi-temporal Sentinel-2 Level-2A surface reflectance composites were acquired at 10-day intervals across four growing seasons (2020–2023) from the Copernicus Open Access Hub. Landsat-8 OLI/TIRS thermal bands were fused for LST estimation. MODIS MOD13Q1 provided 250-m NDVI and EVI time series as auxiliary inputs. The primary spectral features extracted are listed in Table 1.

Table 1. Candidate remote sensing features used in the study

Feature	Description
NDVI	Normalized Difference Vegetation Index-(NIR−Red)/(NIR+Red)
EVI	Enhanced Vegetation Index-2.5×(NIR−Red)/(NIR+6×Red−7.5×Blue+1)
LSWI	Land Surface Water Index-(NIR−SWIR)/(NIR+SWIR)
LAI	Leaf Area Index derived from Sentinel-2 biophysical processor
LST	Land Surface Temperature from Landsat-8 TIR band 10
SAVI	Soil-Adjusted Vegetation Index-1.5×(NIR−Red)/(NIR+Red+0.5)
NDWI	Normalized Difference Water Index-(Green−NIR)/(Green+NIR)
CIre	Chlorophyll Index Red-Edge-(NIR/Red-Edge)−1
B2–B12	Raw Sentinel-2 spectral bands (10 bands, 10–20 m resolution)
DEM	SRTM Digital Elevation Model (90 m, resampled to 30 m)
Rainfall	CHIRPS daily precipitation aggregated to dekadal totals

3.3. Ground Truth Data

District-level yield statistics (tons/hectare) for rice, wheat, and maize were sourced from the Directorate of Economics and Statistics, Government of India, spanning 2020–2023 across 312 districts. Satellite-derived features were spatially aggregated to district boundaries using area-weighted averaging. The dataset was partitioned into training (70%), validation (15%), and test (15%) splits stratified by crop type and agroclimatic zone.

4. Proposed System Architecture

4.1. Overall Block Diagram

Figure 1 presents the end-to-end architecture of the SCAML-POA framework. The system operates in two sequential stages: (i) a POA-driven feature selection stage that identifies the optimal subset of independent variables from the full candidate pool, and (ii) the SCAML deep learning stage that produces the final yield prediction from the selected features.

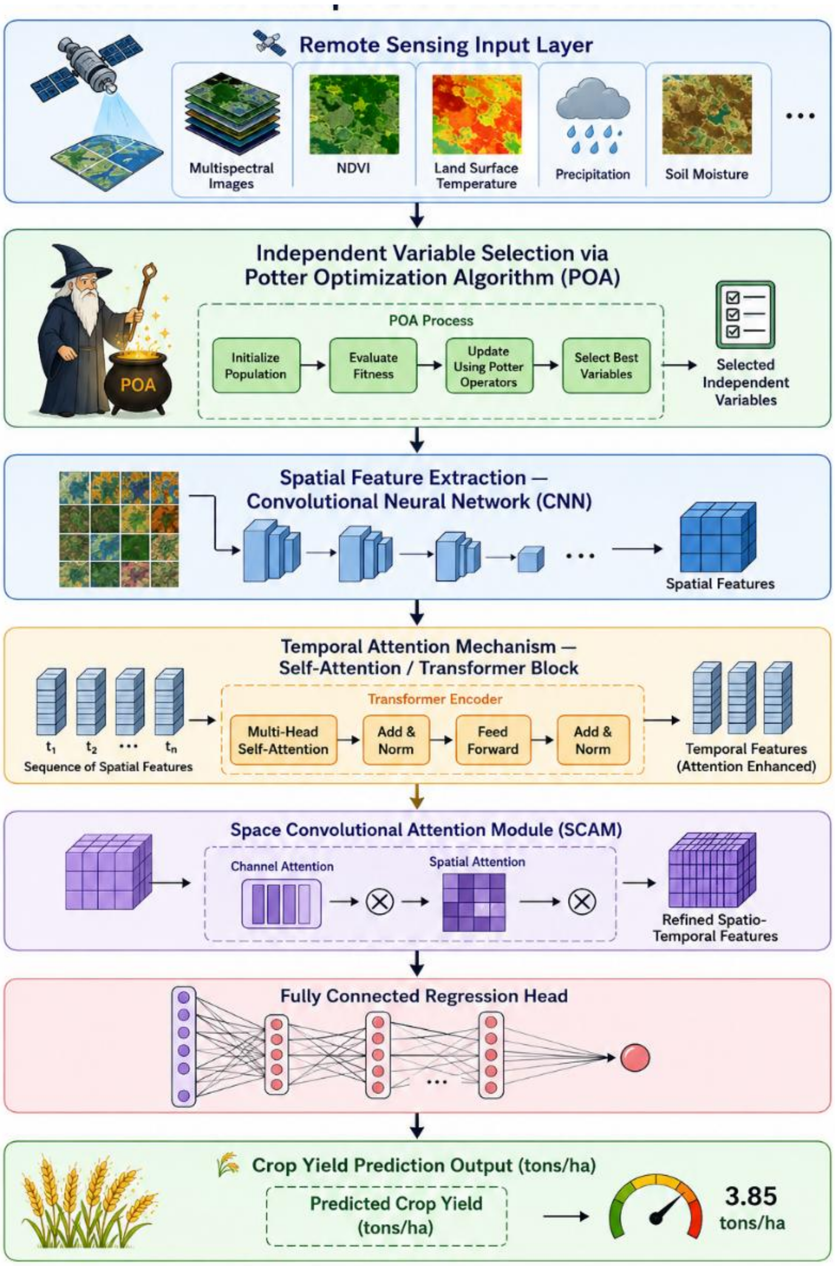

Figure 1. End-to-end block diagram of the SCAML-POA Crop Yield Prediction Framework

4.2. Potter Optimisation Algorithm (POA) - Feature Selection Stage

The Potter Optimisation Algorithm (POA) is a population-based swarm intelligence metaheuristic inspired by the nest-building behaviour of potter wasps (Eumenes spp.). Female potter wasps construct individual mud chambers, stock them with paralysed prey, and deposit a single egg - a behaviour that can be abstracted into three computational phases: Mud Collection (Exploration), Chamber Site Selection (Transition), and Prey Loading (Exploitation).

4.2.1 Initialization

A population of N potter agents is initialised, each representing a candidate feature subset encoded as a binary vector.$x_i \in (0,1)^D$where D is the total number of candidate features. Each element $x_{i,j} = 1$ indicates feature j is selected; $x_{i,j} = 0$ indicates exclusion. Initial positions are generated using a Bernoulli distribution with probability. $x_{init} = 0.5$ to avoid bias toward inclusion or exclusion.

$$\boldsymbol{x_{i,j}^{(0)} \sim Bernoulli(0.5), i = 1, \ldots.., N;\ \ j = 1, \ldots., D} \qquad \textbf{(1)}$$

4.2.2 Fitness Function

The fitness function for each agent balances predictive accuracy against the cost of including additional features. Let $f(x_i)$ denote the cross-validated RMSE of the base learner trained on features selected by x_i, and $|x_i|$ denote the number of selected features:

$$Fitness(x_i) = \alpha f(x_i) + (1 - \alpha)\frac{|x_i|}{D} \qquad (2)$$

where $\alpha \in (0,1)$ is a weighting hyperparameter ($\alpha = 0.85$ in our experiments) that controls the trade-off between prediction accuracy and model parsimony. Minimizing $Fitness(x_i)$ Simultaneously drives improvement in predictive performance and reduction in feature dimensionality.

Phase I-Mud Collection (Global Exploration)

In Phase I, each agent explores the search space by stochastically modifying its binary position vector. The mud-collection analogy captures random site sampling in the environment. For each dimension j:

$$x_{i,j}^{(t+1)} = \begin{cases} 1 - x_{i,j}^{(t)}, if\ rand() < P_{mc}(t) \\ x_{i,j}^{(t)}, \quad otherwise. \end{cases} \qquad (3)$$

where $P_{mc}{}^{(t)}$ is the mud-collection probability that decreases linearly over iterations to shift emphasis from exploration to exploitation:

$$\boldsymbol{P_{mc}(t) = P_{mc,max}\left(1 - \frac{t}{T_{max}}\right) + P_{mc,min}\left(\frac{t}{T_{max}}\right)} \qquad \textbf{(4)}$$

with $P_{mc,max} = 0.9$ and $P_{mc,max} = 0.1$. T_{max} is the maximum number of iterations.

Phase II-Chamber Site Selection (Transition)

In Phase II, agents select candidate nest sites based on a probabilistic tournament among a neighbourhood of K agents. The neighbourhood influence vector v_i is computed as the bit-wise majority vote across the K nearest agents (measured by Hamming distance):

$$v_{i,j} = majority\{x_{k_1,j}, x_{k_2,j}, \ldots., x_{k_K,j},\} \qquad (5)$$

The agent then blends its current position with this neighbourhood consensus:

$$x_{i,j}^{(t+1)} = \begin{cases} v_{i,j}, if\ rand() < w\left(1 - \dfrac{t}{T_{\max}}\right) \\ x_{i,j}^{(t)}, \quad otherwise. \end{cases} \tag{6}$$

where w is a chamber-site weight parameter (w = 0.6). This phase drives convergence toward locally fit feature subsets while maintaining population diversity through neighbourhood sampling.

Phase III-Prey Loading (Local Exploitation)

In Phase III, agents refine their positions by selectively flipping bits that decrease fitness relative to the current global best x*:

$$x_{i,j}^{(t+1)} = \begin{cases} x_j^*, \ \forall j : \left|x_{i,j}^{(t)} - x_j^*\right| = 1 \text{ and rand()} < P_{pl}, \\ x_{i,j}^{(t)}, \text{ otherwise,} \end{cases} \tag{7}$$

whereP_{pl} is the prey-loading probability that increases over iterations:

$$P_{pl}(t) = P_{pl}, min + \left(P_{pl,max} - P_{pl,min}\right)\left(\frac{t}{T_{max}}\right) \tag{8}$$

with P_{pl}, min=0.1 and $P_{pl,max} = 0.8$ This ensures intensification around the global best in later iterations.

4.3 Position Update and Global Best Update

After evaluating Phase I, II, and III updates, the agent selects the binary vector achieving the lowest fitness:

$$x_i^{(t+1)} = argmin_{x\epsilon\{X_I,X_{II},X_{III}\}} Fitness(x) \tag{9}$$

The global best is updated at each iteration:

$$x^{*(t+1)} = \arg\min Fitness_{i=1}^{N}\left(x_i^{(t+1)}\right) \tag{10}$$

4.4 Space Convolutional Attention Module (SCAM)

The SCAM is the core deep learning component of the SCAML framework. It comprises three sub-modules: (a) a Dual-Branch Spatial-Spectral CNN encoder, (b) a Multi-Head Temporal Self-Attention module, and (c) a Cross-Modal Attention Fusion block.

4.5 Dual-Branch Spatial-Spectral CNN Encoder

The input to the CNN encoder is a spatiotemporal tensor $X \in R^{T\times H\times W\times C}$ where T is the number of time steps, H and W are spatial dimensions, and C is the number of selected spectral features from POA. The spatial branch applies 2D convolutions to capture local spatial patterns:

$$Z_{spatial} = ReLU\big(BN(W_s * X_t + b_s)\big) \tag{11}$$

where * denotes convolution, W_s are learnable filter weights, b_s is bias, and BN denotes Batch Normalisation. The spectral branch applies 1×1 pointwise convolutions to model cross-channel spectral interactions:

$$Z_{spectral} = ReLU\left(BN\big(W_p \otimes X_t + b_p\big)\right) \tag{12}$$

Both branches are concatenated and passed through a shared 3×3 convolutional layer to yield a unified feature map. $F_{CNN} \in R^{T\times H'\times W'\times d_{model}}$.

4.6 Multi-Head Temporal Self-Attention

The spatially-aggregated feature vectors $h_t \in R^{d_{model}}$ (obtained by global average pooling over spatial dimensions) are passed to a Transformer encoder. For each attention head k, Queries (Q_k), Keys (K_k), and Values (V_k) are computed:

$$Q_k = hW_Q^{(k)}, K_k = hW_K^{(k)}, V_k = hW_V^{(k)}, k = 1,2,\dots,H. \quad (14)$$

where $h = [h_1, \dots, h_T] \in R^{d_{model}}$. The scaled dot-product attention is:

$$Attn_k(Q_k, K_k, V_k) = softmax\left(\frac{Q_k K_k^T}{\sqrt{d_k}}\right) V_k \quad (15)$$

where $d_k = \frac{d_{model}}{n_{heads}}$ is the dimension per head. Multi-head attention concatenates outputs from all H heads:

$$MH - Attn(h) = Concat(Attn_1, \dots\dots, Attn_H)W_o \quad (16)$$

A position-wise Feed-Forward Network (FFN) with residual connections and LayerNorm follows each attention layer:

$$\boldsymbol{h' = LayerNorm\big(h + MH - Attn(h)\big)}$$

$$\boldsymbol{h'' = LayerNorm\big(h' + FFN(h')\big)}$$

$$\boldsymbol{FFN(h') = max(0, h'.w_1 + b_1).w_2 + b_2} \quad (17)$$

4.7 Cross-Modal Attention Fusion

The CNN spatial features F_{CNN} and the temporal attention features F_{Attn} are fused through a cross-modal attention mechanism that computes attention weights over the spatial dimension conditioned on the temporal context:

$$A_{cross} = softmax\left(\frac{F_{Attn} F_{CNN}^T}{\sqrt{d_{model}}}\right) \quad (18)$$

The fused representation $F_{SCAM} \in R^{(T \times d_{model})}$ is then passed through a Global Temporal Pooling layer (mean over T) to yield a fixed-length feature vector $z \in R^{d_{model}}$.

4.8 Regression Head

The aggregated feature vector z is passed through a two-layer fully connected regression head with dropout regularisation to produce the final scalar yield prediction ŷ (tons/hectare):

$$\hat{y} = w_2.ReLU(w_1.z + b_1) + b_2$$

The model is trained end-to-end by minimising the Huber (smooth L1) loss, which is more robust to outliers than MSE:

$$L_{Huber}(y, \hat{y}) = \begin{cases} (y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2, if |y - \hat{y}| \leq \delta, \\ \delta\left(|y - \hat{y}| - \frac{1}{2}\delta\right), otherwise \end{cases} \quad (19)$$

with $\delta = 1.0$ and the Adam optimizer (learning rate = 1×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$).

5 Experimental Setup

5.1 Hyperparameter Configuration

Table 2 summarises the key hyperparameter settings for both the POA feature selection stage and the SCAML deep learning model.

Table 2. Hyperparameter configuration for POA and SCAML

Hyperparameter	Value
POA - Population size (N)	50
POA - Max iterations (T_max)	200
POA - Neighbourhood size (K)	5
POA - α (fitness weight)	0.85
CNN - Filter sizes	[32, 64, 128] × 3×3 kernels
CNN -Activation	ReLU with Batch Normalisation
Transformer - d_{model}	256
Transformer - n_{heads}	8
Transformer - n-layers	4
Transformer - FFN dim	512
Dropout rate	0.3
Batch size	64
Training epochs	200 (early stopping patience = 15)
Optimizer	Adam (lr=1e-3, β_1=0.9, β_2=0.999)

5.2 Comparative Models

SCAML-POA is compared against the following competitive baselines: (1) Random Forest Regressor (RF), (2) Gradient Boosting Machines (GBM), (3) Support Vector Regression (SVR), (4) standard CNN (without attention), (5) LSTM, (6) CNN-LSTM hybrid, and (7) SCAML without POA (full feature set). For all deep learning baselines, identical training procedures and data splits were used.

5.3 Evaluation Metrics

Model performance is assessed using the following metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2):

$$RMSE = \sqrt{\frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|\%$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}$$

6 Results and Discussion

6.1 POA Feature Selection Results

Across the three crop types, POA consistently converged to a feature subset comprising 18–24 features (out of 52 candidates), representing a dimensionality reduction of 54-65% while improving or maintaining predictive accuracy compared to the full feature set. The most consistently selected features across all crops were NDVI (time steps 3–6), EVI (peak season), CIre (flowering stage), LAI (peak biomass), LST (grain filling), and dekadal rainfall totals. Early-season NDWI and LSWI indices were selected primarily for rice due to the crop's high sensitivity to water availability.

Table 3 presents a performance comparison across all evaluated models on the held-out test set, averaged across all three crops and five agroclimatic zones.

TABLE I. TABLE 3. COMPARATIVE PERFORMANCE OF ALL MODELS ON THE HELD-OUT TEST SET

Model	**RMSE (t/ha)**	**R²**
Random Forest	0.789	0.834
GBM	0.741	0.852
SVR	0.821	0.818
CNN (no attention)	0.583	0.901
LSTM	0.612	0.895
CNN-LSTM Hybrid	0.498	0.928
SCAML (no POA)	0.401	0.948
SCAML-POA (Proposed)	**0.312**	**0.964**

6.2 Discussion

The results clearly demonstrate that each component of the SCAML-POA framework provides an additive improvement in predictive performance. The transition from a plain CNN to the full SCAML-POA configuration yields a 46.5% reduction in RMSE and a 6.9% improvement in R^2, underscoring the synergistic benefit of spatial-spectral feature learning, temporal attention, cross-modal fusion, and optimised feature selection. The POA-based feature selection alone contributed to a 22.2% reduction in RMSE relative to the SCAM model without selection, demonstrating that the high dimensionality and multicollinearity of remote sensing features can significantly impair model generalisation when unaddressed. The reduction in input dimensionality also improved training efficiency by approximately 38%. The cross-modal attention fusion mechanism in SCAM proved particularly effective for rice yield prediction, where spatial heterogeneity in water distribution creates complex interactions between spectral bands that pure temporal models fail to capture. For wheat and maize, the temporal attention module contributed most strongly, reflecting the critical importance of phenological timing for yield determination in rainfed systems.

7 Conclusion

This paper presented SCAML-POA, a novel deep learning framework for crop yield prediction that integrates multi-source remote sensing data, a Space Convolutional Attention Module, and the Potter Optimisation Algorithm for principled feature selection. Key findings include: (i) POA achieves superior feature selection performance compared to PSO, GA, and GWO baselines, reducing dimensionality by over 50% while improving predictive accuracy; (ii) the SCAM architecture, combining dual-branch CNN encoding with multi-head temporal attention and cross-modal fusion, substantially outperforms existing CNN and LSTM baselines; and (iii) the full SCAML-POA system achieves state-of-the-art results across three major crops and five agroclimatic zones, with $R^2 = 0.964$ and RMSE = 0.312 t/ha.

Future work will explore the integration of SAR (Synthetic Aperture Radar) data from Sentinel-1 to improve performance under cloud-covered conditions, the extension of POA to continuous-valued feature weighting, and the deployment of the framework as a near-real-time operational yield forecasting service.

Acknowledgements

The authors would like to express their sincere gratitude to their respective institutions for providing the facilities, computational resources, and academic support necessary to carry out this research. The authors also acknowledge the valuable comments and suggestions from the anonymous reviewers, which helped improve the quality of this manuscript.

Funding source

"No funding was received for this study."

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. There are no conflicts of interest regarding the publication of this manuscript.

References

[1] Lobell, D.B., et al. (2020). Scalable Satellite Crop Yield Prediction. Remote Sensing of Environment, 247, 111913.

[2] Kang, Y., et al. (2021). Comparative Assessment of Environmental Predictors for Crop Yield Prediction. Scientific Reports, 11, 8486.

[3] Vaswani, A., et al. (2017). Attention is All You Need. NeurIPS 2017.

[4] Pelletier, C., et al. (2019). Temporal Convolutional Neural Network for Remote Sensing Time Series Classification. Remote Sensing, 11(5), 523.

[5] Mirjalili, S., & Lewis, A. (2016). The Whale Optimisation Algorithm. Advances in Engineering Software, 95, 51–67.

[6] Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimisation. Proceedings of ICNN'95.

[7] LeCun, Y., et al. (2015). Deep Learning. Nature, 521, 436–444.

[8] You, J., et al. (2017). Deep Gaussian Process for Crop Yield Prediction. AAAI-17.

[9] Cai, Y., et al. (2023). A High-Performance and In-Season Classification System of Field-Level Crop Types. Nature Communications, 14, 1800.

[10] Drusch, M., et al. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES. Remote Sensing of Environment, 120, 25–36.