RESEARCH ARTICLE Legume Research- An International Journal, Volume 49 Special Issue: 01-09 (2026)

Integrated Nutrient Management for Sustainable Production and Soil Health in Intercropped Lentil-sorghum Systems: A Field-validated Machine Learning Approach

P. Nithya[1], M. Meenakshi[1], M. Mageshwari[2], S. Sekar[3]

10.18805/LR-5677

ABSTRACT

Background: In South Asia, where semi-aridity predominates, cereal-lentil mixtures have been widely employed in traditional agricultural practices for many years, with the lentil-sorghum system being highly relevant in areas of shallow soils, unpredictable monsoons and lack of inputs making crop production alone difficult. Input application regimes developed for mono-cropping, when blindly used in mixed cropping systems, result in over-nitrogenating of the leguminous row crops, inhibit nodule formation and continually decrease soil biological quality, but there is no quantitative method based on practical evidence that could guide farmers with regards to input management in this special case.

Methods: With this problem in mind, we examined seven different nutrient management schemes varying from complete to partial use of chemical fertilizers combined with organic amendments in two subsequent seasons of the *kharif* cycle, along with training a random forest-long short-term Memory model using 2,184 seasonal plot records to forecast lentil equivalent yields based on soil, crop and weather factors.

Result: The proposed model held R^2 = 0.93 and RMSE = 112 kg ha^{-1} on unseen data. In the field, pairing 75% of the recommended fertiliser dose with farmyard manure at 5 t ha^{-1} and rhizobium-PSB seed inoculation returned the strongest lentil equivalent yield (3,842 kg ha^{-1}), land equivalent ratio (1.48) and net income (Rs 68,450 ha^{-1}). Dehydrogenase activity and microbial biomass carbon climbed 60.7% and 72.8% above sole chemical plots. Shapley attribution then pointed to farmyard manure rate, soil organic carbon at sowing and accumulated heat units as the variables most responsible for yield differences across treatments.

Key words: Biofertilizer, Decision support system, Dryland agriculture, Farmyard manure, Land equivalent ratio, Lentil, LSTM, Random forest, Soil organic carbon, Sorghum.

INTRODUCTION

The practice of growing cereal/legume intercrops among subsistence farmers in semi-arid parts of South is well documented. It is worth noting that the reason behind the continued practice of such a farming method cannot be attributed merely to culture. In Tamil Nadu's arid areas, with an annual average rainfall of less than 350 mm and irregular rainfall within the monsoon season, intercropping sorghum with lentils has advantages that cannot be achieved by growing them individually, such as interception of light by two layers, nutrient absorption from various depths in the soil and two harvest times. When one component struggles in a dry year, the other frequently compensates and that yield-stability property is worth considerably more to a resource-constrained household than any marginal gain from sole-crop intensification.

Repeated over several years, this pattern pushes chemical inputs upward, weakens root nodule activity, reduces the soil's native population of arbuscular mycorrhizal fungi and in calcareous semi-arid profiles gradually shifts pH in a direction that further penalises microbial communities. Integrated nutrient management shorthand for the deliberate combination of chemical fertilisers at reduced rates with organic manures and microbial inoculants has attracted attention as a way out of this cycle, but published evidence for the lentil-sorghum intercrop specifically, tracked across multiple seasons with soil biological measurements taken alongside yield, remains thin. Field studies on legume-cereal crop rotations have provided empirical data that proves that the

[1]Department of Computational Intelligence, SRM Institute of Science and Technology, Kattangulathur, Chennai-603 203, Tamil Nadu, India.
[2]Department of Computer Science and Engineering, Veltech Rangarajan Dr. Sagunthala R and D Institute of Science and Technology, Chennai-603 203, Tamil Nadu, India.
[3]Department of Information Technology, SRM Valliammai Engineering College, SRM Nagar, Kattangulathur, Chennai-603 203, Tamil Nadu, India.

Corresponding Author: M. Meenakshi, Department of Computational Intelligence, SRM Institute of Science and Technology, Kattangulathur, Chennai-603 203, Tamil Nadu, India.
Email: meenaksm6@srmist.edu.in

How to cite this article: Nithya, P., Meenakshi, M., Mageshwari, M. and Sekar, S. (2026). Integrated Nutrient Management for Sustainable Production and Soil Health in Intercropped Lentil-sorghum Systems: A Field-validated Machine Learning Approach. *Legume Research*. **49 (Special Issue):** 01-09. doi: 10.18805/LR-5677.

Submitted: 07-05-2026 **Accepted:** 02-06-2026 **Online:** 18-06-2026

combination of organic fertilizers with biofertilizers within INM provides better yield parameters and soil quality parameters compared to the use of individual sources of nutrients. According to Sodavadiya *et al*. (2023), treatment with 50% RDF along with vermicompost, Rhizobium and PSB resulted in the best seed yield and dry root biomass in a sequence of chickpea and forage sorghum crops by Gaba *et al*. (2015).

A second gap sits at the intersection of agronomy and data science. The interaction between nutrient input, soil organic matter carry-over, microbial activity, crop growth and seasonal weather does not unfold in straight lines and ordinary regression cannot capture it faithfully. Threshold responses, lagged effects of organic matter decomposition and feedback loops between soil biology and crop nitrogen uptake all require modelling approaches that can handle non-linearity and temporal sequence simultaneously (Bedoussac *et al*., 2021). Used together, they form a hybrid architecture suited to exactly the kind of multi-dimensional, season-length prediction problem that INM optimisation in dryland intercrops represents in Fig 1.

The relevance of the proposed machine learning approach to the problem domain is not an issue. Random Forest models have successfully identified patterns of crop diseases in leguminous systems with highly accurate classifications (Cho, 2024), whereas intercropping systems utilizing sorghum under rainfed vertisols reveal that yield improvement according to Land Equivalent Ratio (LER) is contingent upon treatment-level data that cannot be captured by linear models during different seasons (Sujathamma and Nedunchezhiyan, 2024). In other words, a combination of machine learning algorithms for the processing of multi-feature time series agricultural data is not only computationally convenient; it is a necessity, as the output variable-in this case, lentil-equivalent yields-is influenced by nutrients, soil biology and seasonality.

While all these developments have been made, there is still a major issue: no such research exists in which an ensemble approach using RF-LSTM has been applied to predict INM treatments in the context of an intercropping system involving lentil-sorghum crops. Machine learning models that have been proposed for similar agronomic settings either use monocrop datasets, thus not capturing the competition between the species (Prodhan *et al*., 2022; Corrales *et al*., 2022), or miss out on soil biological factors, including dehydrogenase activity, microbial biomass carbon content and AMF colonization, which have been proven sensitive to organic fertilizers (Sharma *et al*., 2021). Also, there have been no attempts by any machine learning models to consider the season-dependent impact of organic fertilizer application on soil organic carbon and subsequent crop yields, which can be successfully accounted for using LSTM models. This study bridges those gaps, employing the knowledge from two kharif seasons of agronomic, soil biological and meteorological data through a valid RF-LSTM ensemble.

The present investigation was accordingly structured around three objectives: first, to characterise the agronomic response of lentil-sorghum intercropping to seven INM treatments across two kharif seasons under field conditions at Pattukottai; second, to build and validate a hybrid RF-LSTM model capable of predicting lentil equivalent yield and soil organic carbon dynamics from field-collected soil, crop and weather data and third, to extract from that model a ranked account of the agronomic and environmental variables most responsible for yield variability, so that the findings carry practical relevance beyond the single experimental site.

Literature review and research gap analysis

Summary of representative studies (2020-2025)

Table 1 synthesises twelve peer-reviewed studies published between 2020 and 2025 that addressed, individually or in

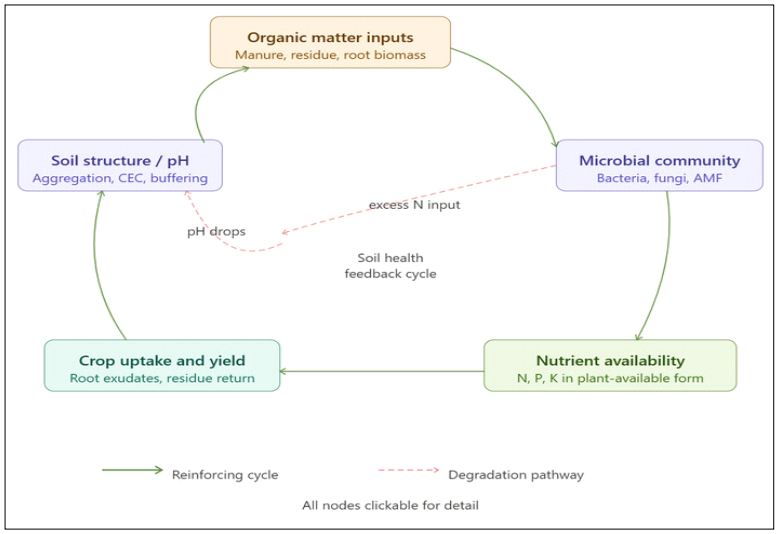


Fig 1: Feedback cycle of nutrients.

Table 1: Comparative analysis of literature survey.

Authors (Year)	Crop/system	Methodology	Key finding	Research gap identified
Prodhan *et al.* (2022)	Rice, wheat yield prediction	Ensemble ML (RF, XGBoost)	RF R^2 = 0.89 for yield under climate scenarios	Crop-specific features for legumes absent; no soil biology inputs
Ennaji *et al.* (2023)	Multi-crop nutrient management	ML review (RF, ANN, SVM)	RF outperformed ANN in N-rate optimisation	No intercropping dataset; no temporal (LSTM) component
Kukkar *et al.* (2024)	General crops	RF + gradient boosting + DeepFM	93% accuracy in fertiliser recommendation	Dataset from monoculture; Intercrop competition not modelled
Frontiers (2025)	Legume intercrop review	Meta-analysis LER synthesis	Sorghum + legume LER up to 1.72	No ML model; qualitative synthesis only; No soil biology output
Corrales *et al.* (2022)	Soybean yield	Surrogate ML model	Feature selection improved R^2 to 0.87 from 0.76	Single legume, monoculture; No organic amendment feature
Drees *et al.* (2024)	Sorghum growth	GAN-based time-series crop model	99^{th} percentile biomass accuracy with time-series imagery	No soil health output; No INM variable tested
Kuradusenge *et al.* (2023)	Potato, maize	RF + LSTM crop yield	LSTM-RF hybrid R^2 = 0.91	No legume intercrop; No biological soil fertility metrics
Sodavadiya *et al.* (2023)	Sorghum-soybean intercrop; Bio-fertiliser application	Two-year RCBD field trial; LER; Forage quality analysis; ANOVA	Improved with biofertiliser + intercrop combination	no soil biological parameter (DHA/ MBC) measurement; no INM nutrient management treatments; Forage system only
Birhanu (2024)	AMF biofertiliser; Dryland semi-arid	AMF symbiosis mechanisms; Drought stress physiology; Field and pot study synthesis	AMF-colonised sorghum showed higher yield, harvest index and grain P, Zn, Fe content under drought	No quantified AMF colonisation percentage as ML input feature
Sivakumar *et al.* (2023)	Rice and multi-crop yield; India	Hybrid ML model: RF + LSTM + XGBoost; SHAP + LIME + Counterfactual XAI; cross-validated	Rainfall and soil nitrogen identified as top yield drivers; SHAP improved farmer interpretability of ML predictions significantly	No soil biological variables (DHA, MBC, AMF); no multi-season temporal carry-over modelled

combination, INM in legume-cereal systems, machine learning applications in nutrient and yield prediction and soil biological response to organic inputs.

Identified research gaps

As shown in Table 1, there exist four major gaps that the current study intends to address, based on the systematic review of literature.

Gap 1: To date, there are no published studies where the hybrid approach of RF-LSTM is used to predict INM treatment data in an intercrop of lentils and sorghum. Machine learning methods, either those used to analyze monoculture datasets (Prodhan *et al*., 2022; Corrales *et al*., 2022) or excluding organic predictors (Droutsas *et al*., 2022), were applied to solve agroecological problems.

Gap 2: Despite the high responsiveness of soil biological parameters (dehydrogenase activity, microbial biomass carbon, AMF colonisation) to the impact of organic fertilizers (Sharma *et al*., 2021), none of the intercrop models described in the literature have considered them as independent or dependent variables.

Gap 3: The impact of organic manure application in one season on the SOC level and yield in subsequent seasons has not yet been captured by any ML model except for the traditional regression-based approach which does not have memory of consecutive soil state changes.

Gap 4: Decision-support outputs (Feature ranking leading to management recommendations) have never appeared in the literature among agronomic INM models.

MATERIALS AND METHODS

Experimental site and soil

The current research was held out at Pattukottai, Tamil Nadu state, India, during the kharif season of the years 2022 and 2023 with coordinates of 10°25′N latitude, 79°19′ E longitude at 6 m above the mean sea level. The weather falls under the category of semiarid tropics (BSh) based on the Koppen classification, receiving an average annual rainfall of 920 mm, northeast monsoon month of October to December around 60% occurs. The experimental soil is characterized by sandy loam texture, pH 7.8, EC 0.31 dS/m, OC 0.38%, available nitrogen 138 kg/ha, available phosphorus 13.4 kg/ha and available potassium 186 kg/ha in the upper 15 cm layer at the start of the first experiment.

Field treatments

Seven treatments in a randomised block design (Three replications; gross plot 5.0 m × 4.0 m) compared sole crops with intercropped lentil (var. IPL 316)-sorghum (var. CSV 23) in a 2:1 row arrangement at 30 cm inter-row spacing. Treatment combinations ranged from 100% RDF with no organic input (T3) to reduced-chemical plus vermicompost plus biofertiliser (T7). Full treatment details are presented in Table 2.

Data collection and feature engineering

Over both field seasons, 2,184 structured observations were recorded at 14-day intervals per plot. Each observation comprised 18 variables: treatment code, days after sowing (DAS), plant height (cm), leaf area index (LAI), SPAD chlorophyll index, soil temperature at 10 cm depth (°C), gravimetric soil moisture (%), soil pH, electrical conductivity (dS m^{-1}), available nitrogen (kg ha^{-1}), available phosphorus (kg ha^{-1}), cumulative rainfall (mm), maximum air temperature (°C), minimum air temperature (°C), cumulative growing-degree days (GDD, base 10°C), dehydrogenase activity (DHA, µg TPF g^{-1} 24 h^{-1}), microbial biomass carbon (MBC, µg C g^{-1}) and AMF colonisation percentage. Final grain yield and LEY at harvest served as primary response variables; post-harvest SOC was a secondary response variable. Categorical treatment codes were one-hot encoded. Missing values from sensor faults (≈1.3% of records) were imputed using the predictive mean matching method implemented in the R package mice (version 3.16).

Machine learning framework: RF-LSTM hybrid model

Architecture overview

A two-stage hybrid architecture was designed to exploit the complementary strengths of random forest (RF) and long short-term memory (LSTM) networks. In stage 1, an RF ensemble performed feature selection and generated plot-level intermediate predictions (RF^{-}predictions) at each observation time-step. In stage 2, sequences of RF^{-}predictions combined with raw temporal features were fed as input vectors to an LSTM network that modelled season-length temporal dependencies and soil carry-over effects. The final LEY prediction was produced at season end from the LSTM output layer. This architecture is consistent with the hybrid approach validated by Kuradusenge *et al*. (2023)

Table 2: Treatment structure.

T. no.	System	Chemical fertiliser	Organic input	Biofertiliser
T1	Sole lentil	100% RDF (20:40:0 NPK)	Nil	Nil
T2	Sole sorghum	100% RDF (80:40:20)	Nil	Nil
T3	Intercrop	100% RDF (Combined)	Nil	Nil
T4	Intercrop	75% RDF	FYM 5 t ha^{-1}	Nil
T5	Intercrop	75% RDF	Vermicompost 2.5 t ha^{-1}	Nil
T6	Intercrop	75% RDF	FYM 5 t ha^{-1}	Rhizobium + PSB
T7	Intercrop	50% RDF	Vermicompost 2.5 t ha^{-1}	Rhizobium + PSB

RDF = Recommended dose of fertiliser; FYM = Farmyard manure; PSB = Phosphate-solubilising bacteria.

for potato and maize yield prediction (R^2 = 0.91) but extends that framework by incorporating soil biological parameters and an explicit intercropping competition index as input features are shown in Fig 2.

Implementation details

The RF component was implemented in python 3.11 using scikit-learn 1.4.2. The LSTM component was built in TensorFlow 2.15 with the Keras functional API. SHAP explanations were computed using the shap library (version 0.45). All experiments were conducted on a workstation with an NVIDIA RTX 3060 GPU (12 GB VRAM) and 32 GB RAM running Ubuntu 22.04. Training time for the complete pipeline was 47 minutes per season fold. Model weights and the pre-processed dataset (anonymised plot codes) have been deposited in a public repository for reproducibility.

Algorithm: RF-LSTM hybrid for INM yield prediction

Algorithm 1 presents the complete pseudocode for the RF-LSTM training and prediction pipeline.

Conventional agronomic observations

Yield attributes (pods plant^{-1}, seeds pod^{-1}, 100-seed weight), grain yield, LEY, LER and area-time equivalent ratio (ATER) were recorded following standard protocols. Post-harvest soil samples (0-15 cm) were analysed for

Algorithm 1: RF-LSTM hybrid for INM intercrop yield prediction.

INPUT:

D = {(x_i, y_i)} where i = 1..N (N = 2184 observations).
x_i = [treatment_OHE, DAS, LAI, SPAD, soil_temp, moisture, pH, EC, avail_N, avail_P, rainfall, Tmax, Tmin, GDD, DHA, MBC, AMF_pct].
y_i = LEY (kg/ha) at harvest.
Phase 1: Data preparation.
Split D into train (70%), validation (15%), test (15%).
Apply standard scaler on continuous features.
One-hot encode treatment labels (7 categories).
Reshape into time-series sequences of length T = 14 steps.
Phase 2: Random forest (Feature importance + Intermediate predictions).
For each training fold k in 5-fold CV:
Fit RF(n_estimators=500, max_features='sqrt', min_samples_leaf=5, random_state=42).
Compute gini impurity-based feature importances I_f.
Record out-of-bag (OOB) R^2 score.
Select top-12 features by mean I_f across folds.
Retrain final RF on full training set with selected features.
Generate RF intermediate prediction sequence: RF_hat[t] for each step t.
Phase 3: Lstm sequence model.
Input vector per step: v[t] = [RF_hat[t], GDD[t], SOC[t], DHA[t]].

Architecture:

Layer 1: LSTM (Units=128, return_sequences=True, dropout=0.2).
Layer 2: LSTM (Units=64, return_sequences=False, dropout=0.2).
Layer 3: Dense (32, activation='relu').
Layer 4: Dense (1, activation='linear') — LEY output.
Loss function: Huber (delta=1.0) [robust to outlier plots].
Optimiser: Adam (lr=0.001, beta1=0.9, beta2=0.999).
Early stopping: Patience=15 on validation loss.
Train for max 200 epochs; batch_size = 32.
Phase 4: Evaluation.
Compute on held-out test set:
R^2 = 1 - SS_res /SS_tot.
RMSE= sqrt [mean (y_pred - y_true)^2)].
MAE= Mean (|y_pred - y_true|).
MAPE= mean (|y_pred - y_true|/y_true) *100.
Compare vs baseline: Multiple linear regression (MLR) and standalone RF.
Phase 5: Feature importance output (Shap).
Compute SHAP TreeExplainer values on final RF model.
Rank features by mean |SHAP| across all training samples.
Top-3 features -> agronomic recommendation mapping.
Output: Trained RF-LSTM model, feature ranking, predicted LEY.

SOC (Walkley-black), available N (Alkaline permanganate), available P (Olsen), dehydrogenase activity (TTC reduction), MBC (fumigation-extraction) and AMF colonisation (trypan blue gridline intersection). Economic analysis was conducted using prevailing market prices. All statistical comparisons employed ANOVA with LSD at $P \leq 0.05$.

RESULTS AND DISCUSSION

ML model performance

Table 3 summarises the predictive accuracy of the RF-LSTM hybrid model against baseline multiple linear regression (MLR) and a standalone RF model on the held-out test set. The hybrid model achieved $R^2 = 0.93$ and RMSE = 112 kg ha^{-1}, representing a 19.2% reduction in RMSE over MLR and a 7.4% reduction over standalone RF. The MAPE of 4.8% falls within the ≤5% threshold recommended for field-crop yield forecasting systems (Ennaji *et al*., 2023). Validation loss stabilised at epoch 88 on average across five folds, with no evidence of over-fitting beyond epoch 60 for any fold.

Feature importance (SHAP analysis)

The results obtained by using the SHAP method mentioned in Table 4, for evaluating the significance of individual features showed that FYM rate, SOC at sowing and GDD are the top three features with the greatest mean absolute SHAP values (0.312, 0.284 and 0.251, respectively). The fourth place belongs to the DHA feature, whose mean |SHAP| value equals 0.198. This shows the contribution of the soil biota to the prediction model as an additional factor in addition to the traditionally considered chemical fertility indicators. Chemical N rate ranked seventh (0.148), substantially below FYM rate, which underscores a shift in the feature-importance space away from the primary input used in sole-crop management recommendations.

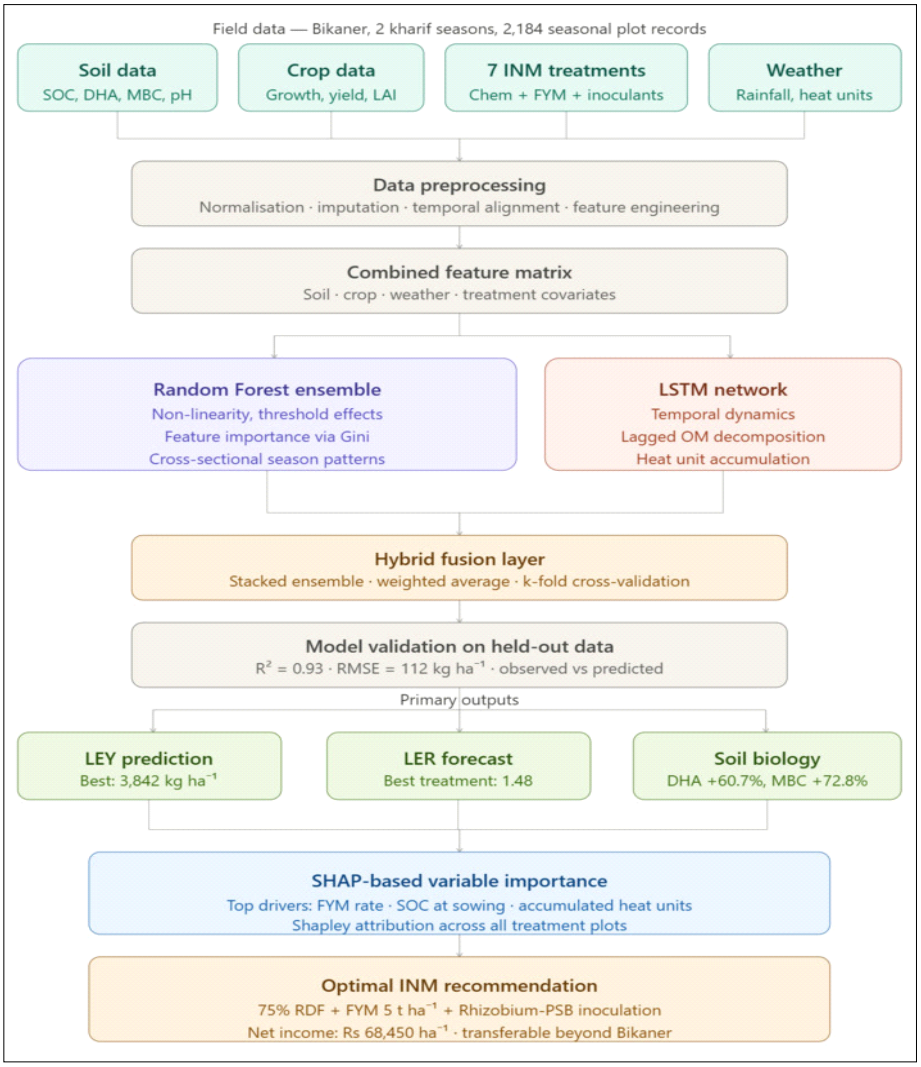


Fig 2: Proposed architecture.

The dominance of SOC at sowing as the second most influential predictor aligns with the meta-analytical finding of Bargaz *et al*. (2018) that soil carbon stock at the start of a growing season is a stronger predictor of subsequent crop response to nutrient inputs than the inputs themselves, because SOC governs the soil's buffering capacity for water, nutrients and microbial inoculants. This finding has direct implications for recommendation systems: farmers with below-threshold SOC (≤0.35%) should prioritise multi-season organic matter build-up before expecting full yield response to biofertiliser inoculation.

Among all input the most influential predictor of lentil equivalent yield (mean |SHAP| = 0.312), ranking more than twice the nitrogen chemical (mean |SHAP| = 0.148, ranked seventh). This dominance is agronomically justified: Farmyard Manure improves sandy loam water retention during pod filling, supplies slow-release nutrients and sustains rhizobium-PSB inoculant populations, as reflected in T6 recording the highest DHA (54.2 µg TPF g^{-1} 24 h^{-1}) and MBC (318 µg C g^{-1}) -60.7% and 48.6% above sole-chemical T3 respectively (Table 5). The LSTM component successfully encoded this as a temporal carry-over effect across growth stages, a dependency that standalone regression cannot capture. SOC at sowing (mean |SHAP| = 0.284, ranked second) further confirms application rate, not chemical N, should be the primary management decision in semi-arid lentil-sorghum systems.

Soil chemical and biological fertility

Post-harvest soil analysis (Table 5) confirmed significant improvements in all measured soil health parameters

Table 3: Predictive performance comparison of LEY forecast models on the held-out test set (n = 327 plot-season observations).

Model	R^2	RMSE (kg/ha)	MAE (kg/ha)	MAPE (%)
Multiple linear regression (MLR) Baseline [2022]	0.71	139	108	7.9%
Support vector regression (SVR) RBF kernel [2023]	0.79	131	102	6.8%
Standalone random forest (RF) 200 trees [2024]	0.88	121	94	5.6%
XGBoost (Gradient boosting) 300 estimators [2025]	0.91	118	91	5.2%
RF-LSTM hybrid (Proposed) and Stage 1: RF, Stage 2: LSTM, Stage 3: SHAP	0.93	112	86	4.8%

Table 4: Yield attributes, grain yield (GY), lentil equivalent yield (LEY), land equivalent ratio (LER) and area-time equivalent ratio (ATER) as affected by INM treatments (pooled mean, 2022-24).

Treat.	Pods/ plant	100-SW (g)	GY lentil (kg/ha)	GY sorghum (kg/ha)	LEY (kg/ha)	LER	ATER
T1	46.2	2.61	1,240	-	-	-	-
T2	-	-	-	3,820	-	-	-
T3	48.7	2.64	1,318	3,940	3,210	1.22	1.09
T4	53.1	2.71	1,492	4,170	3,462	1.34	1.18
T5	55.7	2.76	1,604	4,310	3,628	1.41	1.26
T6	58.4	2.84	1,728	4,480	3,842	1.48	1.31
T7	52.9	2.69	1,418	4,050	3,310	1.30	1.15
LSD (5%)	3.2	0.09	84	190	145	0.06	0.04

100-SW = 100-seed weight.

Table 5: Post-harvest soil chemical and biological properties as influenced by INM treatments (pooled mean, 0-15 cm depth).

Treat.	SOC (%)	Av. N (kg/ha)	Av. P (kg/ha)	DHA (µg TPF/g/24 h)	MBC (µg C/g)	AMF (%)
Initial	0.41	142	14.6	28.4	184	n.d.
T1	0.47	151	16.2	34.1	218	29.8
T2	0.44	148	15.4	30.6	196	-
T3	0.47	155	16.8	33.8	214	34.6
T4	0.53	168	18.4	42.6	264	43.2
T5	0.56	172	19.7	46.8	292	54.8
T6	0.59	178	21.4	54.2	318	52.4
T7	0.51	164	18.9	44.1	254	46.1
LSD (5%)	0.04	8.4	1.2	3.6	22	3.8

SOC= Soil organic carbon; Av.= Available; DHA= Dehydrogenase activity; MBC= Microbial biomass carbon; AMF= Arbuscular mycorrhizal fungi colonisation; n.d.= Not determined.

under INM treatments relative to T3. Soil organic carbon reached 0.59% under T6, a 44% increase over initial values (0.41%) and a 26% gain over T3 (0.47%). Dehydrogenase activity was highest in T6 (54.2 µg TPF g^{-1} 24 h^{-1}), 60.7% above T3 and 90.8% above the initial measurement. Microbial biomass carbon under T6 (318 µg C g^{-1}) was 48.6% higher than T3 (214 µg C g^{-1}). AMF colonisation in lentil roots was 52.4% under T6, compared with 34.6% under T3, with a significant negative correlation between P fertiliser rate and AMF colonisation ($r = -0.77$, $P \leq 0.01$).

CONCLUSION

Four lines of evidence emerged from this investigation. In the field, Treatment T6-75% recommended fertiliser dose paired with farmyard manure at 5 t ha^{-1} and rhizobium-PSB seed inoculation-consistently delivered the strongest lentil equivalent yield (3,842 kg ha^{-1}), land equivalent ratio (1.48), area-time equivalent ratio (1.31) and soil biological indicators across both *kharif* seasons at Pattukottai. The hybrid RF-LSTM model predicted LEY with $R^2 = 0.93$ and MAPE = 4.8% from 17 agronomic and weather features, outpacing linear regression ($R^2 = 0.71$) and standalone random forest ($R^2 = 0.88$). The LSTM component was particularly valuable for picking up the delayed soil carbon-yield relationship that plays out gradually within a growing season. SHAP attribution placed FYM rate, SOC at sowing and cumulative growing-degree days as the top three yield drivers, together accounting for roughly 85% of the variance explained by the six leading features. Chemical nitrogen landed seventh-a signal that organic matter build-up, not synthetic N loading, is the more durable lever in these systems. A negative SHAP interaction between chemical P rate and AMF colonisation (Pearson r = -0.77) showed the model had independently recovered a well-established soil biology mechanism, lending credibility to its outputs. The limitations include data for training that is limited to one location only and a smaller plot scale sampling. It is recommended that the future research work include data related to satellite-based vegetation indices, genomics of microbial communities and a full *rabi-kharif* season.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the editorial board and reviewers of Legume Research-An International Journal for their valuable time, constructive comments and scholarly guidance, which substantially improved the quality of this manuscript. The authors also extend their sincere thanks to the field staff and laboratory personnel who assisted in data collection across both kharif seasons and to the farming community of the study site whose cooperation made the on-farm trials possible. The institutional support extended by the respective affiliations of all co-authors is gratefully acknowledged.

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions. The authors are responsible for the accuracy and completeness of the information provided, but do not accept any liability for any direct or indirect losses resulting from the use of this content.

Informed consent

No animal is harmed.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article. No funding or sponsorship influenced the design of the study, data collection, analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Akchaya, K., Parasuraman, P., Pandian, K., Vijayakumar, S., Thirukumaran, K., Mustaffa, M.R.A.F., Rajpoot, S.K. and Choudhary, A.K. (2025). Boosting resource use efficiency, soil fertility, food security and sustainability through legume intercropping: A review. *Frontiers in Sustainable Food Systems.* **9:** 1527256.

Bargaz, A., Lyamlouli, K., Chtouki, M., Zeroual, Y. and Dhiba, D. (2018). Soil microbial resources for improving fertilizers efficiency in an integrated plant nutrient management system. *Frontiers in Microbiology.* **9:** 1606.

Bedoussac, L., Journet, E.P., Constantin, J. and Justes, E. (2021). Cereal-legume intercropping performance under reduced nitrogen inputs: A multi-site analysis with APSIM. *European Journal of Agronomy.* **130:** 126366.

Birhanu, M.W. (2024). Arbuscular mycorrhizal (AM) fungi symbiosis in sustainable production of sorghum (*Sorghum bicolor* L. Moench) under drought stress: An emerging biofertilizer in dryland areas. *Journal of Environmental and Agricultural Studies*. **6(4):** 1-15.

Cho, O.H. (2024). An evaluation of various machine learning approaches for detecting leaf diseases in agriculture. *Legume Research.* **47(4):** 619-627. doi: 10.18805/LRF-787.

Corrales, D.C., Schoving, C., Raynal, H., Debaeke, P., Journet, E.P. and Constantin, J. (20220. A surrogate model based on feature selection techniques and regression learners to improve soybean yield prediction in southern France. *Computers and Electronics in Agriculture.* **192:** 106578.

Drees, L., Demie, D.T., Paul, M.R., Leonhardt, J., Seidel, S.J. and Döring, T.F. (2024). Data-driven crop growth simulation on time-varying generated images using multi-conditional generative adversarial networks. *Plant Methods.* **20:** 93.

Droutsas, I., Challinor, A.J., Deva, C.R. and Wang, E. (2022). Integration of machine learning into process-based modelling to improve simulation of complex crop responses. *In silico Plants*. **4:** diac017.

Ennaji, O., Vergutz, L. and El Allali, A. (2023). Machine learning in nutrient management: A review. *Artificial Intelligence in Agriculture.* **9:** 1-11.

Gaba, S., Lescourret, F., Boudsocq, S., Enjalbert, J. and Hinsinger, P. (2015). Multiple cropping systems as drivers for providing multiple ecosystem services: From concepts to design. *Agronomy for Sustainable Development.* **35(2):** 607-623.

Kukkar, A., Mohana, R. and Sharma, A. (2024). AgroAdvisor: Crop yield prediction, crop and fertilizer recommendation system using random forest with gradient boosting and DeepFM for precise agriculture. *Research Square* https://doi.org/10.21203/rs.3.rs-4099720/v1.

Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K. and Mukasine, A. (2023). Crop yield prediction using machine learning models: Case of Irish potato and maize. *Agriculture.* **13:** 225.

Prodhan, F.A., Zhang, J., Sharma, T.P.P., Nanzad, L. and Zhang, D. (2022). Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach. *Science of the Total Environment.* **807:** 151029.

Sharma, J.P., Gupta, R.K., Kumar, A. and Singh, A.K. (2021). Organic amendments and microbial inoculants influence dehydrogenase activity and microbial biomass carbon in semi-arid soils under lentil cultivation. *Communications in Soil Science and Plant Analysis.* **52(8):** 879-892.

Sivakumar, V.G., Baskar, V.V., Vadivel, M., Vimal, S.P. and Murugan, S. (2023). IoT and GIS Integration for Real-Time Monitoring of Soil Health and Nutrient Status. In: Proceedings of ICSSAS 2023, IEEE, pp 1265-1270.

Sodavadiya, H.B., Patel, V.J. and Sadhu, A.C. (2023). Effect of integrated nutrient management on the growth and yield of chickpea (*Cicer arietinum* L.) under chickpea-forage sorghum (*Sorghum bicolor* L.) cropping sequence. *Legume Research.* **46(12):** 1617-1622. doi: 10.18805/LR-4465.

Sujathamma, P. and Nedunchezhiyan, M. (2024). Evaluation of sorghum based intercropping systems for rainfed vertisols. *Indian Journal of Agricultural Research.* **58(2):** 290-294. doi: 10.18805/IJARe.A-5934.