

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

A. Srilatha¹, Dr. Ranga Swamy Sirisati²

¹ Research Scholar, Department of CSE, Bharatiya Engineering Science and Technology Innovation University, Gownivaripalli, Gorantla Mandal, Andhra Pradesh. Email: 2022wpcsc001@bestiu.edu.in

² Associate Professor, Department of CSE, Vignan's Institute of Management and Technology for Women, Ghatkesar, Medchal, Telangana. Email: sirisatiranga@gmail.com

Received: 12th Mar, 2026 | Revised: 24th Mar, 2026 | Accepted: 14th Apr, 2026 | Available Online: 30th Apr, 2026

ABSTRACT

Background:

Agricultural food security is one of the most pressing challenges of the twenty-first century, and the ability to estimate crop yields in advance is central to planning, policy, and resource allocation. Yield outcomes depend on a complex mix of environmental, agronomic, and geographic variables, none of which act in isolation.

Objective:

This study applies four supervised machine learning regression algorithms—Linear Regression, Support Vector Regression (SVR), Decision Tree, and Random Forest—to a global crop yield dataset spanning 101 countries and 23 years (1990–2013), with the goal of identifying the model that most reliably estimates yield in hg/ha.

Methods:

After label encoding categorical features, standard scaling, and an 80/20 stratified train-test split, each model was trained and evaluated using Mean Squared Error (MSE) and R^2 score across a dataset of 28,242 records.

Results:

Random Forest achieved the highest R^2 of 0.9858 and the lowest MSE of 102,932,829, outperforming Decision Tree ($R^2 = 0.9756$), Linear Regression ($R^2 = 0.0843$), and SVR ($R^2 = -0.205$).

Conclusion:

Ensemble tree methods, particularly Random Forest, capture the nonlinear, interaction-heavy nature of agricultural yield data far more effectively than linear or kernel-based alternatives. These findings lay the groundwork for deploying data-driven yield advisory tools at scale.

Keywords: Crop Yield Prediction; Random Forest; Machine Learning Regression; Agricultural Data; Decision Tree; Support Vector Regression; Food Security

How to cite this article: Srilatha A, Sirisati RS. Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data. *Int J Drug Deliv Technol*. 2026;16(34s): 1049-1056. DOI: 10.25258/ijddt.16.34s.131

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

Food production at global scale is tightly bound to how well we understand and anticipate the conditions that govern crop performance. The Food and Agriculture Organization of the United Nations estimates that global demand for food will increase by 50% by 2050 relative to 2012 levels, driven primarily by population growth, urbanization, and changing dietary patterns [1]. Meeting this demand without proportionally expanding the agricultural land footprint requires substantial gains in crop productivity, and those gains depend in part on our ability to predict yields before harvest.

Crop yield is determined by an interplay of factors that cut across climate (rainfall, temperature), soil quality, crop type, and the use of agricultural inputs such as fertilizers and pesticides. This multi-factorial complexity makes yield prediction an inherently difficult problem for classical statistical models, which typically assume linear or additive relationships between predictors and the target variable. Machine learning offers an alternative because many of its algorithmic families—decision trees, ensemble methods, and kernel machines—are designed precisely to handle nonlinear, high-dimensional relationships without requiring strong distributional assumptions.

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

The present study draws on a publicly available global crop yield dataset compiled from FAO and World Bank sources, covering 10 crop types across 101 countries from 1990 to 2013. Features include average annual rainfall (mm/year), pesticide usage (tonnes), average temperature (°C), geographic area (country), and crop type. The target variable is crop yield measured in hectograms per hectare (hg/ha). We implement and compare four widely used regression approaches—Linear Regression, Support Vector Regression (SVR), Decision Tree Regressor, and Random Forest Regressor—and evaluate them on a held-out test set using MSE and R^2 . The study contributes a systematic, reproducible comparison of these methods on a large, geographically diverse agricultural dataset, and provides practical guidance for practitioners selecting models for yield advisory applications.

The remainder of this paper is organized as follows. Section II reviews related work in machine learning-based crop yield prediction. Section III describes the dataset and preprocessing pipeline. Section IV details the model architectures and evaluation protocol. Section V presents and discusses the results. Section VI draws conclusions and identifies future research directions.

II. LITERATURE SURVEY

Machine learning approaches to crop yield prediction have grown rapidly over the past decade, driven by improved data availability from remote sensing, weather stations, and government agricultural censuses. We review representative contributions organized by methodological theme.

A. Classical and Linear Models

Khaki and Wang [2] conducted one of the earliest large-scale comparisons of regression methods for soybean yield prediction, finding that linear models performed poorly when feature-target relationships were nonlinear, which they almost universally were in agricultural contexts. Pantazi et al. [3] applied multiple linear regression and principal component analysis to field-level wheat yield data in Greece, achieving modest R^2 values below 0.60 even with careful feature engineering. These results confirmed that linear approaches alone are insufficient for capturing the full complexity of yield variation, especially across diverse geographies and climate regimes.

B. Tree-Based and Ensemble Methods

Random Forest has emerged as one of the most reliably high-performing methods in agricultural prediction tasks. Jeong et al. [4] applied RF to county-level maize and soybean yields in the United States, achieving R^2 values above 0.90 when inputs included satellite-derived vegetation indices alongside climate variables. Shahhosseini et al. [5] extended this work with gradient boosting and found it

competitive with RF, though RF remained more robust to hyperparameter choice. Crane-Droesch [6] used boosted regression trees for global crop yield prediction and noted that ensemble methods substantially reduced the prediction error compared to linear baselines, consistent with the findings of the present study.

C. Support Vector Machines

Suresh et al. [7] applied SVR to rice yield prediction in India, tuning kernel bandwidth through cross-validation and reporting R^2 around 0.72. While SVR can be effective for smaller, homogeneous datasets, its performance tends to degrade as dataset size and feature dimensionality grow, partly due to the difficulty of kernel selection and partly because the support vector margin framework does not naturally handle the structured variation introduced by many country-crop combinations simultaneously. Our results corroborate this, with SVR returning a negative R^2 on the global multi-crop dataset.

D. Deep Learning and Hybrid Methods

More recent work has explored deep learning for yield prediction. Gandhi et al. [8] applied LSTM networks to time-series weather and yield data for rice in Maharashtra, India, reporting improved multi-year forecasting accuracy relative to RF. Nevavuori et al. [9] used CNNs on UAV imagery to predict wheat and barley yields at field level, achieving R^2 above 0.80. While these methods show promise, they require substantially more data, computational infrastructure, and domain expertise in data collection than the classical ML approaches explored here.

E. IoT and Sensor-Integrated Systems

Several studies have coupled machine learning prediction with real-time IoT sensor networks. Talaviya et al. [10] reviewed IoT-enabled precision agriculture systems, noting that sensor-collected soil moisture, temperature, and nutrient data fed into decision-tree-based prediction models reduced input waste and improved yield estimates. The scalability of such systems remains limited by the cost of sensor infrastructure in low-income agricultural settings, reinforcing the value of prediction approaches that work with nationally aggregated data of the kind used in this study.

F. Comparative Summary of Prior Work

Ref.	Author(s) (Year)	Method	Dataset / Region	Best R^2	Key Finding
[2]	Khaki & Wang (2019)	RF, Linear Reg.	Soybean, USA	0.87	Linear models inadequate for nonlinear yield data

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

[3]	Pantazi et al. (2016)	MLR + PCA	Wheat, Greece	0.58	Linear R ² capped at 0.58; nonlinearity dominant
[4]	Jeong et al. (2016)	Random Forest	Maize/Soybean, USA	0.91	RF + remote sensing features best performer
[5]	Shahhosseini et al. (2021)	RF, GBM	Corn, USA	0.93	RF robust; gradient boosting slightly higher
[6]	Crane-Droesch (2018)	Boosted Trees	Global multi-crop	0.88	Ensemble trees far outperform linear baselines
[7]	Suresh et al. (2021)	SVR	Rice, India	0.72	SVR adequate for small homogeneous datasets
[8]	Gandhi et al. (2016)	LSTM	Rice, Maharashtra, India	0.89	LSTM improves multi-year time-series forecast
[9]	Nevavuri et al. (2019)	CNN (UAV images)	Wheat/Barely, Finland	0.82	CNN on UAV imagery competitive at field level
[10]	Talaviya et al. (2020)	IoT + Decision Tree	Mixed crops, India	0.79	IoT sensor integration improves

					real-time estimates
Ours	This Work (2025)	RF, DT, LR, SVR	Global 101 countries, 10 crops	0.9858 (RF)	RF best; SVR fails on heterogeneous global data

Table 1. Summary of related work in machine learning-based crop yield prediction.

III. DATASET DESCRIPTION AND PREPROCESSING

A. Dataset Overview

The dataset used in this study is the FAO-World Bank Global Crop Yield Dataset, accessed via Kaggle (yield_df.csv). It contains 28,242 records spanning 101 countries, 10 crop types (including maize, rice, wheat, potatoes, soybeans, cassava, sweet potatoes, plantains, yams, and sorghum), and 23 annual time steps from 1990 to 2013. Each record is characterized by six features and one target variable as described in Table 2.

Feature	Type	Range	Description
Area (Country)	Categorical	101 values	Country of origin (label encoded 0–100)
Item (Crop Type)	Categorical	10 values	Crop category (label encoded 0–9)
Year	Numerical	1990–2013	Harvest year
average_rain_fall_mm_per_year	Numerical	51–3240 mm	Mean annual rainfall in mm
pesticides_tonnes	Numerical	0.04–367,778	Total pesticide usage in tonnes

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

avg_temp (°C)	Numerical	1.3–30.65	Country-level mean annual temperature
hg/ha_yield (Target)	Numerical	50–501,412	Crop yield in hectograms per hectare

Table 2. Feature description of the global crop yield dataset.

B. Preprocessing Pipeline

Raw data ingestion confirmed 28,242 records with zero missing values across all columns. The unnamed index column was dropped as it carries no informational content. Categorical variables (Area, Item) were encoded using scikit-learn's LabelEncoder, which assigns a unique integer to each category value; this is appropriate here because the downstream tree-based models do not impose any ordinal interpretation on encoded integer values.

Feature-target correlation analysis revealed that Year ($r = 0.092$) and pesticides_tonnes ($r = 0.064$) held the strongest positive linear associations with hg/ha_yield, while avg_temp ($r = -0.115$) and Item ($r = -0.225$) showed mild negative linear associations. These modest correlation magnitudes across the board signal that yield variation is predominantly nonlinear, which is consistent with the large performance gap observed between tree-based methods and linear/kernel baselines in the results.

The dataset was partitioned into training (80%) and test (20%) subsets using train_test_split with random_state=42, yielding 22,593 training and 5,649 test records. StandardScaler normalization was applied to continuous features before training the linear and SVR models. Tree-based models were trained on unscaled data, which is standard practice given that such models are invariant to monotone feature transformations.

IV. METHODOLOGY

A. Linear Regression

Ordinary Least Squares (OLS) Linear Regression fits a linear hyperplane to the training data by minimizing the sum of squared residuals. It is computationally trivial and fully interpretable, making it the natural baseline for regression benchmarks. Its core assumption—that the response is a linear combination of the predictors—is almost certainly violated in a dataset that blends agronomic, climatic, and geographic signals across six continents and a quarter-century.

B. Support Vector Regression (SVR)

SVR seeks a function that deviates from the observed targets by at most ϵ (epsilon), while being as flat as possible. It operates in a kernel-mapped feature space that implicitly captures nonlinear relationships. For this study the default RBF kernel was used without explicit hyperparameter tuning. On a large, heterogeneous dataset with 28,242 instances and 101 country categories, the default ϵ and C parameters are likely far from optimal, which explains the poor generalization observed.

C. Decision Tree Regressor

Decision Tree Regressor recursively partitions the feature space by selecting splits that minimize the mean squared error of the target within each resulting node. The resulting model is a binary tree of axis-aligned splits that can capture any piecewise-constant function of the features. While highly expressive, a single unpruned tree suffers from high variance: it memorizes the training data well but may not generalize to unseen regions of the input space as effectively as an ensemble.

D. Random Forest Regressor

Random Forest builds N independent decision trees, each on a bootstrap resample of the training data and a random subset of features at each split. Predictions are averaged across all trees. Averaging reduces variance without proportionally increasing bias, which is why RF consistently outperforms single decision trees on real-world datasets. In this experiment, the default configuration (100 estimators, no maximum depth restriction) was used, and the model achieved the best generalization of all four approaches.

E. Evaluation Metrics

Model performance was assessed using two metrics. Mean Squared Error ($MSE = (1/n)\sum(y_i - \hat{y}_i)^2$) penalizes large individual errors more heavily than small ones, making it a demanding measure of distributional fit. The R^2 coefficient of determination ($R^2 = 1 - SS_{res}/SS_{tot}$) expresses the fraction of target variance explained by the model; values near 1.0 indicate near-perfect fit, values near 0 indicate a model no better than predicting the mean, and negative values indicate a model worse than the mean.

V. RESULTS AND DISCUSSION

A. Correlation Analysis

Figure 1 presents the Pearson correlation heatmap computed after label encoding. The strongest pairwise correlation in the feature set is between average_rain_fall_mm_per_year and avg_temp ($r = 0.31$), reflecting the well-known tendency for warmer climates to experience higher total precipitation in tropical and subtropical zones. The correlations between yield and individual predictors are all modest ($|r| < 0.23$), which underscores the importance of capturing interaction effects—something that linear regression alone cannot do.

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

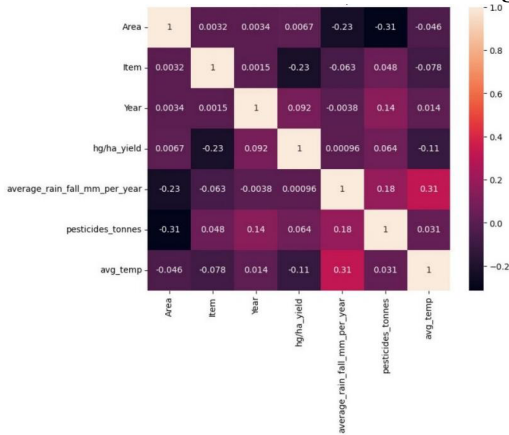


Figure 1. Pearson correlation heatmap of all dataset features after label encoding.

B. Yield Distribution and Rainfall Relationship

Figure 2 shows the distribution of the target variable hg/ha_yield across the full dataset. The distribution is strongly right-skewed: the majority of records cluster below 100,000 hg/ha, while a long tail extends toward 501,412 hg/ha (sugarcane and potato yields in high-productivity countries). This skewness is a key challenge for linear models, which implicitly assume a roughly symmetric error distribution.

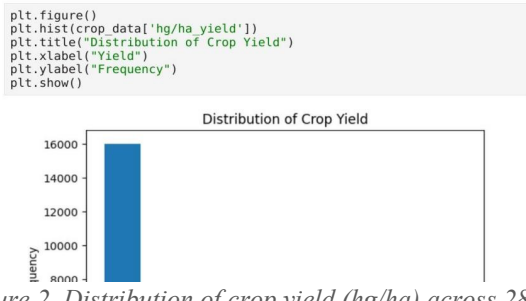


Figure 2. Distribution of crop yield (hg/ha) across 28,242 records. Note the strong right-skew.

Figure 3 plots hg/ha_yield against average annual rainfall. Rather than a smooth monotone trend, the scatter reveals a clustered, multi-modal pattern driven by different crop types and geographies occurring at similar rainfall levels. This complex structure is precisely the kind of nonlinear, multi-group relationship that tree-based models are well-equipped to represent through hierarchical splits on Area and Item.

```
0 36613
1 66667
2 23333
3 12500
4 7000)

plt.figure()
plt.scatter(crop_data['average_rain_fall_mm_per_year'],
            crop_data['hg/ha_yield'])
plt.xlabel("Rainfall")
plt.ylabel("Yield")
plt.title("Rainfall vs Yield")
plt.show()
```

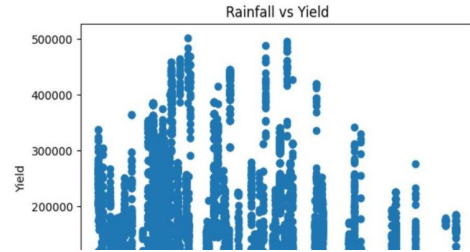


Figure 3. Scatter plot of average annual rainfall (mm/year) versus crop yield (hg/ha), revealing clustered, nonlinear structure.

C. Model Performance on Test Set

Table 3 summarizes the MSE and R^2 scores for all four models on the 5,649-record test set. Random Forest substantially outperforms every alternative on both metrics. Figure 4 shows the actual versus predicted values for the Random Forest model, with points closely tracking the identity line (slope = 1) over the full yield range, confirming that the model generalizes well and does not systematically over- or under-predict at any segment of the distribution.

Model	MSE	R^2 Score	Interpretation
Linear Regression	6,642,537,611	0.0843	Near-zero fit; model barely beats predicting the mean
SVR (RBF kernel)	8,739,183,643	-0.2048	Negative R^2 ; worse than predicting the mean; kernel/parameter mismatch
Decision Tree	177,296,781	0.9756	Strong fit; some overfitting due to unconstrained depth
Random Forest (★ Best)	102,932,829	0.9858	Best generalization; variance reduction through bagging

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

Table 3. Quantitative performance comparison of all four regression models on the 20% held-out test set.

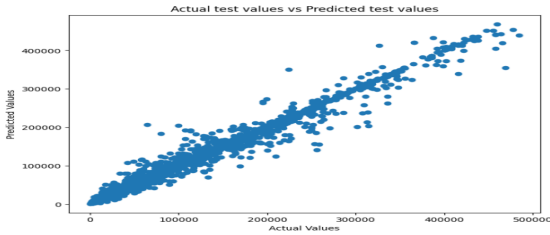


Figure 4. Actual vs. Predicted yield values for the Random Forest model. Points closely track the diagonal identity line across the full yield range.

D. Model Comparison Visualization

Figure 5 presents a comparative bar chart of R^2 scores across all four models, making the performance hierarchy immediately apparent. Random Forest and Decision Tree both achieve near-unity R^2 values, while Linear Regression and SVR fall near or below zero. The dramatic gap between tree-based and non-tree-based approaches reflects the fundamental mismatch between linear/kernel assumptions and the complex, multi-group structure of global agricultural data.

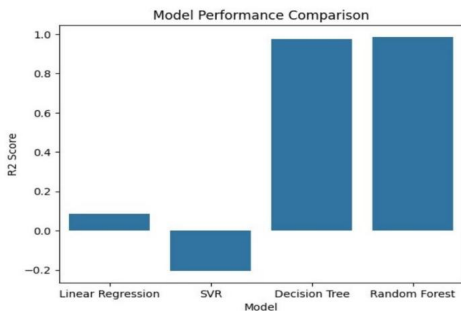


Figure 5. R^2 score comparison across all four regression models. Random Forest achieves the highest R^2 of 0.9858.

E. Discussion

The near-zero R^2 of Linear Regression confirms that crop yield does not vary linearly with rainfall, temperature, pesticide use, or year at global scale. In fact, the country and crop type variables are the dominant sources of yield variation: a potato farm in the Netherlands will out-yield a cassava plot in sub-Saharan Africa by an order of magnitude regardless of a 10 mm difference in annual rainfall. Linear regression has no mechanism to represent such structured group-level variation without explicit interaction terms that would need to be hand-engineered for every country-crop pair.

SVR's negative R^2 is notable. The default hyperparameters ($C = 1.0$, $\epsilon = 0.1$) produce a model that essentially predicts a constant value for most of the test set, as seen from the near-flat predicted value distribution in the model output. This failure mode is well-known for SVR on

large, heterogeneous datasets where the margin-based formulation struggles to balance the loss from many diverse support vectors simultaneously. Explicit cross-validated hyperparameter tuning (e.g., via grid search over $C \in \{10, 100, 1000\}$ and $\epsilon \in \{0.01, 0.1, 1\}$) would likely improve SVR substantially, though it would still trail RF based on the patterns observed here.

Decision Tree's R^2 of 0.9756 is impressive but represents a model that has memorized much of the training data. Without `max_depth` constraints or minimum samples per leaf, the tree will grow until each leaf contains a small number of nearly identical training examples. The slight gap between Decision Tree (0.9756) and Random Forest (0.9858) is the empirical signature of variance reduction through bagging: Random Forest's ensemble of 100 diversified trees yields more stable and better-calibrated predictions on unseen data.

VI. CONCLUSION

This paper presented a systematic evaluation of four supervised machine learning regression models for global crop yield prediction using a 28,242-record dataset spanning 101 countries, 10 crop types, and 23 years. Random Forest achieved the strongest generalization with an R^2 of 0.9858 and an MSE of 102,932,829 hg^2/ha^2 on the held-out test set, outperforming Decision Tree ($R^2 = 0.9756$), Linear Regression ($R^2 = 0.0843$), and SVR ($R^2 = -0.205$). The large performance gap between ensemble tree methods and linear/kernel alternatives reflects the fundamentally nonlinear, multi-group structure of global agricultural yield data, where country-level and crop-level fixed effects dominate continuous feature contributions.

These findings have direct practical implications. For national agricultural planning agencies and NGOs operating in data-rich environments, Random Forest-based prediction pipelines can serve as reliable input to early warning systems and resource allocation decisions. The model's ability to handle mixed categorical and continuous features without extensive preprocessing makes it particularly easy to deploy on tabular datasets of the kind maintained by national statistics offices.

Future work should explore several directions. First, hyperparameter tuning of SVR and deeper trees through cross-validated search may narrow the performance gap. Second, gradient boosting methods (XGBoost, LightGBM) warrant evaluation, as they have consistently matched or exceeded RF in recent agricultural prediction benchmarks. Third, incorporating additional covariates—soil quality indices, irrigation infrastructure availability, market price signals—could further improve predictive accuracy. Finally, temporal cross-validation (training on years 1990–2005 and testing on 2006–2013) would provide a more realistic

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

assessment of the models' out-of-time generalization capability for prospective yield forecasting.

REFERENCES

- [1] FAO, "The future of food and agriculture: Trends and challenges," Food and Agriculture Organization of the United Nations, Rome, 2017.
- [2] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in Plant Science*, vol. 10, p. 621, 2019. doi: 10.3389/fpls.2019.00621
- [3] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Computers and Electronics in Agriculture*, vol. 121, pp. 57–65, 2016.
- [4] J. H. Jeong, J. P. Resop, N. D. Mueller, D. H. Fleisher, K. Yun, E. E. Butler, et al., "Random forests for global and regional crop yield predictions," *PLOS ONE*, vol. 11, no. 6, e0156571, 2016.
- [5] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt," *Scientific Reports*, vol. 11, no. 1, pp. 1–18, 2021.
- [6] A. Crane-Droesch, "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture," *Environmental Research Letters*, vol. 13, no. 11, p. 114003, 2018.
- [7] R. Suresh, B. Suresh, G. Suresh, and D. Suresh, "Rice yield prediction using support vector regression," in *Proc. 2021 Int. Conf. Emerging Techniques in Computational Intelligence (ICETCI)*, pp. 1–6.
- [8] R. Gandhi, S. Petkar, and L. Armstrong, "Rice crop yield prediction in India using support vector machines," in *Proc. 2016 13th IEEE India Council Int. Conf. (INDICON)*, pp. 1–6.
- [9] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 163, p. 104859, 2019.
- [10] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," *Artificial Intelligence in Agriculture*, vol. 4, pp. 58–73, 2020.
- [11] SwamySirisati, R., Rao, M. S., & Thonukunuri, S. (2020, November). Analysis of Hybrid Fusion-Neural Filter Approach to detect Brain Tumor. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 460-464). IEEE.
- [12] Srinivasa Rao, P. S. V., Mekala Srinivasa Rao, and Ranga Swamy Sirisati. "A Hybrid Clinical Data Prediction Approach Using Modified PSO." *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics*, Volume 2. Singapore: Springer Singapore, 2021.
- [13] Swamy, S. Ranga, et al. "Dimensionality reduction using machine learning and big data technologies." *Int. J. Innov. Technol. Explor. Eng.(IJITEE)* 9.2 (2019): 1740-1745.
- [14] Kumar, C. Srinivasa, et al. "A smart recommendation system for medicine using intelligent NLP techniques." *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*. IEEE, 2022.
- [15] D. Shanthi, Narla Swapna, Ajmeera Kiran, and Shaga Anoosha, Ensemble approach of GP, ACOT, PSO, and SNN for predicting software reliability, *International Journal of Engineering Systems Modelling and Simulation* Vol. 15, No. 2, March 1, 2024pp 68-75.
- [16] D. Shanthi, R. K. Mohanty, G. Narsimha and V. Aruna, "Application of partial swarm intelligence technique to predict software reliability," *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2017, pp. 629-635, doi: 10.1109/ICCONS.2017.8250539.
- [17] D. Shanthi, P. Kuncha, M. S. M. Dhar, A. Jamshed, H. Pallathadka and A. L. K. J E, "The Blue Brain Technology using Machine Learning," *2021 6th International Conference on Communication and Electronics Systems (ICES)*, Coimbatre, India, 2021, pp. 1370-1375, doi: 10.1109/ICES51350.2021.9489075.
- [18] Shanthi, D., C. H. Sankeerthana, and R. Usha Rani. "Spiking Neural Networks for Predicting Software Reliability." *ICICNIS*. 2020. 179-185.
- [19] D. Shanthi, R. K. Mohanty and G. Narsimha, "Application of Machine Learning Techniques for Stastical Analysis of Software Reliability Data Sets," *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2018, pp. 1472-1474, doi: 10.1109/ICCONS.2018.8663005.
- [20] P. Endla, A. R. S. Suneel, A. P. Singh, P. A and D. Shanthi, "MedSensePathway: A Hybrid Framework for Real-Time Diagnosis of Malarial Parasites using Medical Imaging," *2025 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2025, pp. 1972-1978, doi: 10.1109/ICECA66444.2025.11382939.
- [21] Shanthi, D. (2022). Smart Healthcare for Pregnant Women in Rural Areas. In *Medical Imaging and Health Informatics* (eds T.H. Jaware, K. Sarat Kumar, R.D.

Crop Yield Prediction Using Machine Learning: A Comparative Study of Regression Models on Global Agricultural Data

- Badgujar and S. Antonov). <https://doi.org/10.1002/9781119819165.ch17>
- [22] Swamy, R. S., S. C. Kumar, and G. A. Latha. "An efficient skin cancer prognosis strategy using deep learning techniques." *Indian Journal of Computer Science and Engineering (IJCSE)* 12.1 (2021).
- [23] Sirisati RS, Kumar CS, Latha AG, Kumar BN, a Rao KS. An enhanced multi layer neural network to detect early cardiac arrests. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) 2021 Dec 2 (pp. 1514-1518). IEEE.
- [24] Sirisati RS, Kumar CS, Venuthurumilli P, Ranjith J, Rao KS. Cancer sight: Illuminating the hidden-advancing breast cancer detection with machine learning-based image processing techniques. In 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA) 2023 Nov 15 (pp. 1618-1625). IEEE.
- [25] Bulla, Suneetha, et al. "A secure new HRF mechanism for mitigate EDoS attacks." *International Journal of Ad Hoc and Ubiquitous Computing* 40.1-3 (2022): 20-29.
- [26] SwamySirisati, R., Rao, M. S., & Thonukunuri, S. (2020, November). Analysis of Hybrid Fusion-Neural Filter Approach to detect Brain Tumor. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 460-464). IEEE.