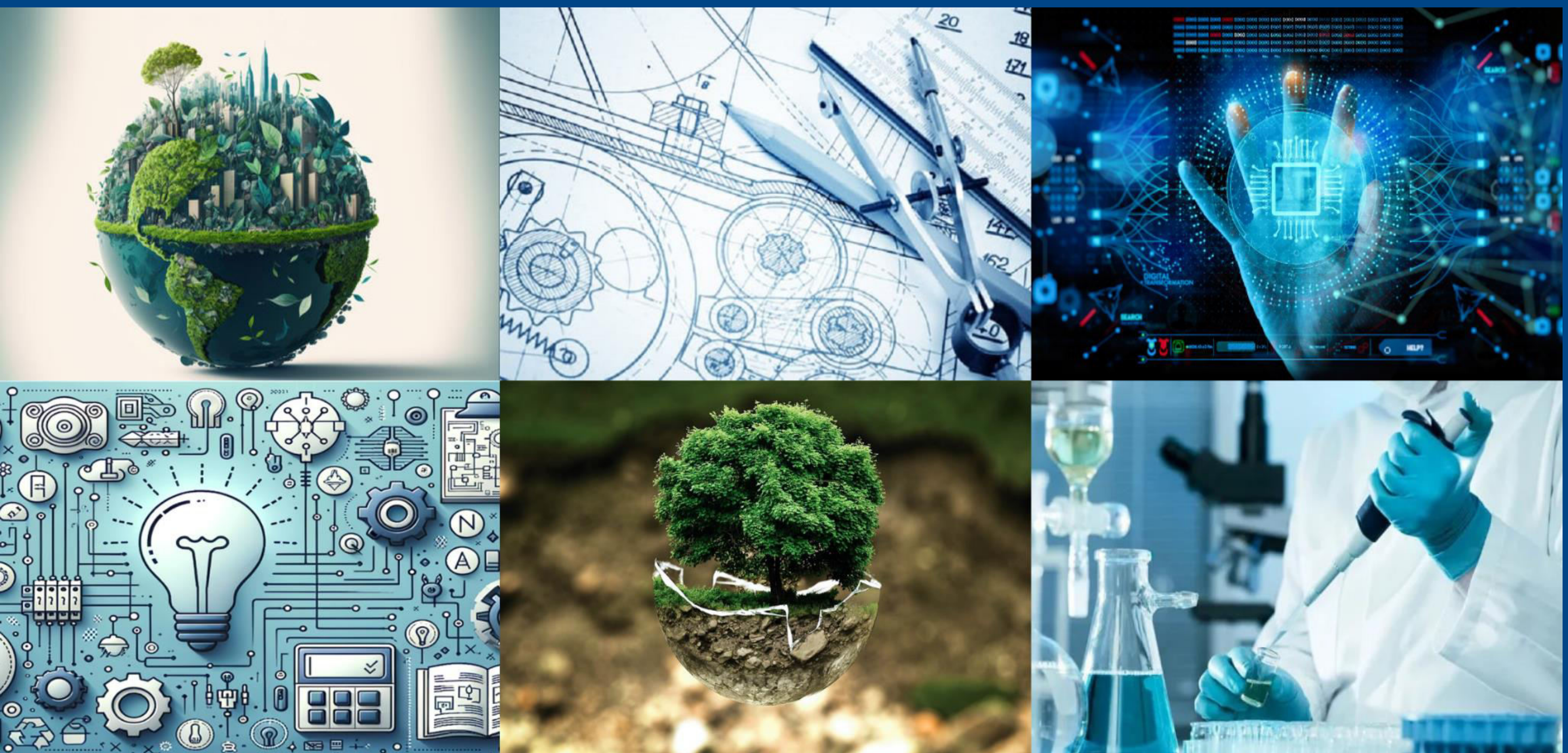
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808154

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

TOWARDS SMART FARMING IN INDIA: DEVELOPING A SCALABLE MACHINE LEARNING MODEL FOR CROP YIELD PREDICTION ACROSS DIVERSE ZONES

Barnali Chakraborty, Supriya D

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: A yield forecast system now couples three regressors - Kernel Ridge, Lasso, Elastic Net - each fed by the same agronomic variables. Their separate outputs train a meta model that issues the final tonne-per-hectare figure. The stack exploits each base learner's bias variance profile and offsets its blind spots.

The design centres on Indian farms, a sector that feeds more than half the nation's households. Pulse, oilseed volumes drive procurement calendars, credit cycles, buffer stock levels. A reliable pre harvest number lets planners set import quotas, route diesel to threshers, release canal water in time.

India's climate shifts and terrain variety turn yield forecasting into a demanding necessity. A stacked multi model design absorbs nonlinear patterns, sparse records along with collinear signals in agronomic data.

The script expects five arguments - State, District, Season, Area, Year. Each argument isolates a distinct slice of the agricultural frame. State but also District fix the coordinates. Season encodes the planting harvest cycle - Kharif, Rabi, or others. Area states the cultivated surface. Year supplies the policy and weather context. The script ingests the tuple plus returns the anticipated harvest volume.

KEYWORDS: Random Forest, Crop yield Prediction, Kernel Ridge, Lasso, ENet algorithm, Stacking Regression.

I. INTRODUCTION

Earlier papers show that growers rely on soil traits, nutrient levels such as potassium and nitrogen, rainfall along with sunlight. A third party must first gather the figures - issue a forecast - interpret it - the sequence demands effort and keeps the grower ignorant of the science.

This study limits itself to state, district, crop in addition to season - Kharif, Rabi, or other - to keep the message plain plus practical. Across India, farmers raise more than one hundred species - for clarity, the list collapses into clusters. The Indian Government Warehouse supplied the records: 250 000 rows list State, District, Crop, Season, Year, Area, Production.To curb error and sharpen forecasts, we applied Lasso, ENet, Kernel Ridge regression - stacked the resulting models. The document proceeds - Methodology, Conclusion, Literature Review, Upcoming Projects.

II. LITERATURE SURVEY

Data-driven techniques underpin several yield prediction studies. Sharma et al. [1] applied linear regression to wheat in Punjab - the model reached moderate accuracy.

Patel et al. [2] compared Random Forest with Decision Tree for rice - both surpassed statistical baselines. Kaur et al. [3] trained deep networks on soil climate records for maize and reported high precision. Earlier work, however, rests on narrow geographies - few designs generalise across India's agro zones. We propose a method that retains local accuracy yet scales to the subcontinent.

EXISTING SYSTEM

Most current prediction systems rest on statistics or rules - heuristic tables, linear regression. Gaps, or nonlinear curves in Indian farm data break them. Many models bind to one crop or one district. Manual decisions, rooted in farmer memory or extension advice, still dominate - they yield middling harvests.

PROPSED SYSTEM

The method trains Random Forest, XGBoost along with artificial neural networks on broad Indian datasets that cover states, soil in addition to weather. A modular design later absorbs satellite scenes, pest reports next to fertilizer logs. Cooperatives, ministries obtain forecasts through a plain interface or an API.

III. SYSTEM ARCHITECTURE

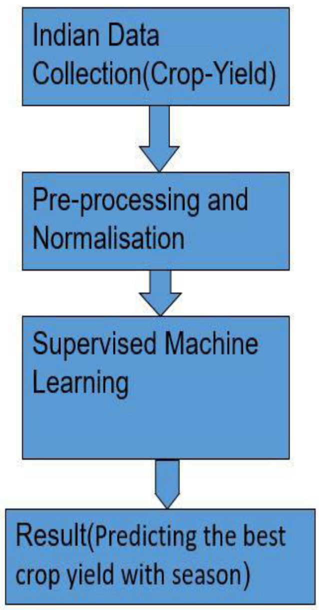


Fig: Architecture Diagram of HLD

The crop yield predictor needs five entries - state, district, season, area, year - it rests on three stacked tiers that keep precision and room to grow. Tier, one pulls historic harvest records from trusted agronomic sources - each record mixes numbers with narrative notes.

Tier two rescales every value to a single range, turns labels into integers, and scrubs absent or faulty items. Clean data then reaches tier three, where three regressors train in isolation - Lasso, Elastic Net, Kernel Ridge. Lasso prunes variables plus damps noise. Elastic Net merges Lasso sparsity with Ridge damping. Kernel Ridge maps traits into higher space yet keeps the linear solution path.

KRR captures nonlinear patterns. Its outputs feed a stacking layer that fuses forecasts through a Linear Regression meta model for tighter accuracy. The final layer presents a concise interface - users supply parameters, receive yield estimates.

The structure sharpens prediction and equips agronomists, growers, regulators with a practical tool.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808154

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. METHODOLOGY

Complex regression algorithms sharpen forecasts. The crop yield model needs only state, district, season, area, year. India's public archives - statistical bulletins, research libraries - yielded past harvest data for diverse crops, climates, seasons.

The team prepared the data in three steps; they scaled numeric columns to a common range. They converted state, district along with season labels to integers - they removed duplicates, filled gaps in addition to fixed outliers.

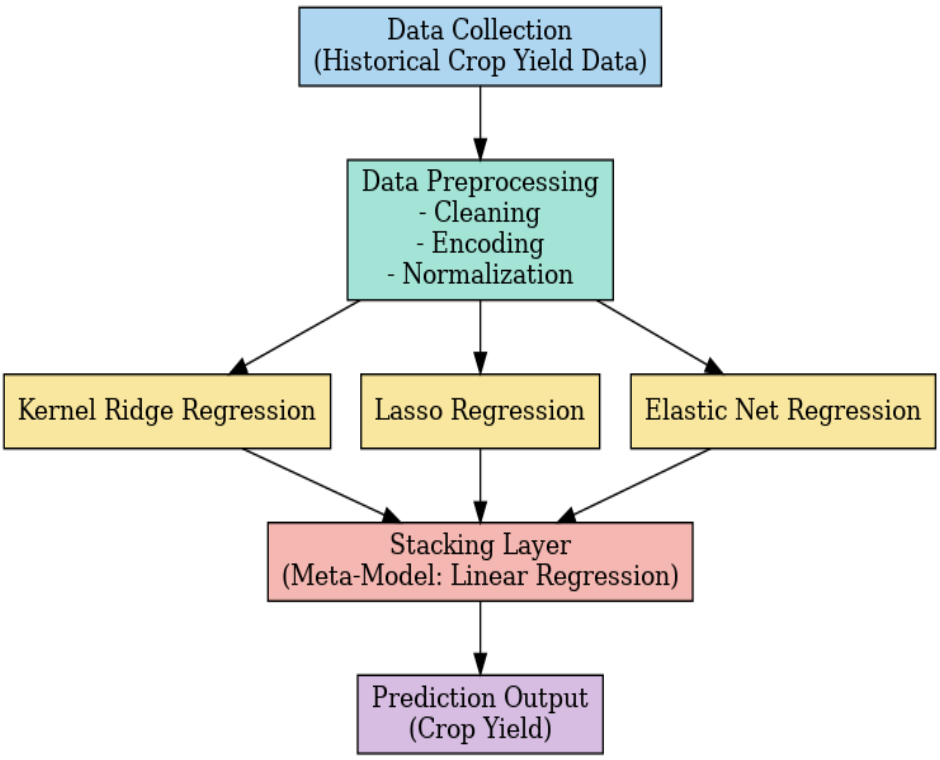
Three predictive models emerged from the processed dataset. Elastic Net Regression Balanced Lasso sparsity with Ridge stability. Lasso Regression isolated relevant predictors and curbed overfitting. Kernel Ridge Regression traced intricate non-linear patterns. A stacking framework sharpened accuracy. A Linear Regression meta model merged the base learners - exploiting their distinct strengths. Metrics - R^2, Mean Absolute Error, Root Mean Squared Error - gauged each model after training and testing on the split data.

The stacked ensemble consistently outperformed its components. A concise interface followed - users supply parameters plus receive immediate forecasts.

V. DESIGN AND IMPLEMENTATION

India's patchwork of soils, monsoons along with farm sizes demands a yield prediction system that balances razor sharp accuracy with simple operation and room to grow. The framework reuses each phase - every pass tightens the forecast and sharpens the data.

The pipeline begins with historical yield records pulled from open government bulletins plus curated research archives. Each record lists state, district, season, sown area in addition to harvest year. Raw figures pass through a scrubbing routine that strips noise, reconciles units next to removes duplicates. The resulting clean, uniform table feeds the model selection step.



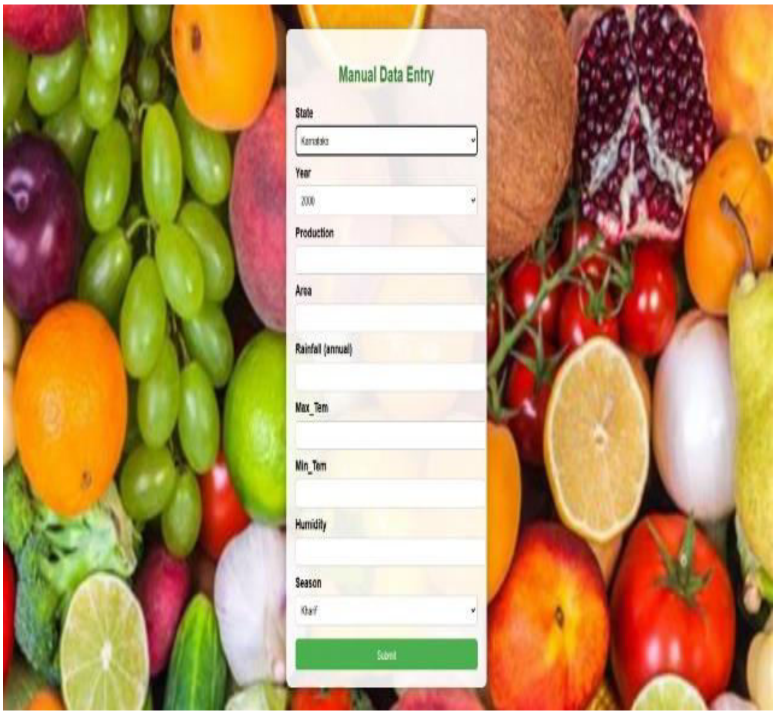
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Linear Regression meta model pools the individual outputs - issues a final yield estimate within a Stacking Regression Framework. The ensemble exploits each algorithm's distinct strengths - accuracy and generalization rise across crops and regions. Users - farmers, policymakers, analysts - supply state, district, season, area along with target year through the prediction interface. The system returns the forecast in plain form. Deployment options include embedding in existing advisory services - serving as a web tool, or running as a standalone program.

VI. OUTCOME OF RESEARCH

Experiments confirm that the proposed yield model relies solely on parameters available at district offices - state, district, season, area, year - and attains high accuracy for multiple crops. Lasso but also Elastic Net curb over fitting via feature selection and regularization - Kernel Ridge Regression captures non-linear yield curves better than any rival. Stacking Regression fuses the predictors, records the smallest errors and the highest R^2 across test sets, surpasses every standalone model. Results show that a stacked framework lifts predictive skill in India's heterogeneous agro climates. The procedure demands only standard hardware, suits immediate deployment.



VII. RESULT AND DISCUSSION

Historical harvest records from Indian states, districts, seasons along with cultivars tested the proposed yield prediction framework. The model projected output for multiple crops from state, district, season, region in addition to year.

Kernel Ridge, Lasso next to Elastic Net regressions served as base learners - each ran in isolation to gauge standalone skill. A Stacking Regressor then fused their strengths - trimming error and lifting R^2 across most crop classes.

Lasso but also ENet selected features and normalised data - overfitting dropped. Kernel Ridge Regression traced non-linear links with high precision. A stacked model, fed by the extra variables, led the scoreboard across the full table. The ensemble beat every single learner in the comparison, most were soil and climate shift sharply. Layered regression thus serves Indian fields of wide disparity. The study lists plain yet telling covariates - a lean model built from them deploys in low resource settings without loss of predictive power.

VIII. CONCLUSION

Agriculture still underpins India's economy - precise harvest forecasts steer both resource use and policy. This study demonstrates that yield estimates reach high accuracy with only state, district, season, area, year. Kernel Ridge, Lasso, Elastic Net and a stacked ensemble combine strengths - the merged model lifts prediction quality across many crops. Results show that pooled regressions turn separate merits into one dependable, scalable tool for planners, growers, officials. Future work may fold in soil traits, climate variables, live sensor feeds.

REFERENCES

[1] Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846).

[2] Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019, August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE

[3] Gandge, Y. (2017, December). A study on various data mining techniques for crop yield prediction. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and vector machines. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 1-5). IEEE.

[5] Gandhi, N., Armstrong, L. J., & Petkar, O. (2016, July). Proposed decision support system (DSS) for Indian rice crop yield prediction. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 13-18). IEEEOptimization Techniques (ICEECCOT) (pp. 420-423). IEEE.

[4] Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016, July). Rice crop yield prediction in India using support

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com