International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# Crop Prediction & Production System Using ML Approaches

**Prof. Shwetha L**
*Asst Professor, Dept. of CSE*
*G. Madegowda Institute of Technology*
*Bharathinagara, Mandya – 571 422*
*India*
shwethal.gmitcse@gmail.com

**Dr. Nuthan A C**
Professor & Head, Dept. of AI&ML
*G. Madegowda Institute of Technology*
*Bharathinagara, Mandya – 571 422*
*India*
aimlhod.gmit@gmail.com

**Ankitha N**
*Student, Dept. of AI&ML*
*G. Madegowda Institute of Technology*
*Bharathinagara, Mandya – 571 422*
*India*
ankithananjappaankitha@gmail.com

**Priya A C**
*Student, Dept. of AI&ML*
*G. Madegowda Institute of Technology*
*Bharathinagara, Mandya – 571 422*
*India*
lopagowda22@gmail.com

**Sinchana N S**
*Student, Dept. of AI&ML*
*G. Madegowda Institute of Technology*
*Bharathinagara, Mandya – 571 422*
*India*
sinchanashivanaga@gmail.com

**Abstract -** This Crop Prediction and Production System that leverages Machine Learning to provide data-driven agricultural decision support tailored for Indian farming communities. The system employs a Random Forest classifier, achieving 87.5% accuracy by analyzing seven critical parameters: soil nutrients (N, P, K), temperature, humidity, pH, and rainfall to recommend optimal crops with confidence scoring. A key innovation is its multi-language interface supporting five Indian languages (English, Hindi, Kannada, Telugu, Tamil), effectively bridging the digital divide for non-English speaking farmers. The platform uniquely integrates predictive analytics with a digital marketplace for real-time price intelligence, land analysis tools for soil health and irrigation planning, and region-specific advisory services. Designed for accessibility, it features QR code-based mobile access, responsive design for low-bandwidth environments, and batch processing capabilities.

Implemented using Flask framework with Bootstrap frontend, the system demonstrates practical deployment of ML in agriculture, enhancing productivity through scientific crop selection while maintaining cultural and technological relevance for diverse user groups. Testing with 50 farmers showed 82% satisfaction with predictions and strong preference for regional language interfaces.

***Key Words*:** *Machine Learning, Agricultural Technology, Crop Prediction, Random Forest, Multi-language Interface, Precision Agriculture, Digital Marketplace, Soil Analysis, Mobile Accessibility, Indian Farming, Flask Framework, Decision Support System, Agricultural Informatics, Sustainable Farming, QR Code Integration.*

## I. INTRODUCTION

Agriculture forms the backbone of the Indian economy, contributing approximately 18.3% to the nation's GDP and employing over 54.6% of the workforce. Despite technological advancements, Indian farmers face numerous challenges including unpredictable weather patterns, soil degradation, market price fluctuations, and limited access to scientific farming knowledge.

The Crop Prediction & Production System is an innovative web-based application that leverages Machine Learning algorithms to assist farmers in making data-driven decisions about crop cultivation. This system integrates predictive analytics, market intelligence, and agricultural advisory services into a unified platform accessible through multiple regional languages.

The primary objective of this project is to bridge the technology gap in agriculture by providing farmers with an easy-to-use tool that recommends suitable crops based on soil parameters, environmental conditions, and market trends. By combining traditional farming knowledge with modern data science techniques, this system aims to enhance agricultural productivity, reduce input costs, and improve farmer livelihoods.

This model that integrates machine learning (ML) and techniques to enhance crop prediction accuracy. By analysing climatic and soil conditions, the model aims to provide precise recommendations for optimal crop selection and resource allocation. The enhanced model demonstrates improved performance over traditional methods by leveraging complex algorithms and diverse datasets to predict crop yields more effectively.

## II. BACKGROUND AND RELATED WORK

### A. Literature Survey

Dahikar S and Rode S V [1] in their 2014 paper "Agricultural crop yield prediction using artificial neural network approach" explored the application of Artificial Neural Networks (ANNs) for predicting agricultural crop yields. Published in the International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, their research demonstrated early adoption of neural networks in agriculture, focusing on improving prediction accuracy through complex pattern recognition capabilities inherent in ANNs.

Prof. D.S. Zingade et al. [2] developed a "crop prediction system using machine learning" that analyzed environmental parameters including temperature, humidity, rainfall, and soil nutrient values (Nitrogen, Phosphorus, Potassium). Their system employed algorithms such as Support Vector Machine (SVM) and Naive Bayes to predict the most suitable crop for cultivation under given conditions, with the primary objective of maximizing yield and profitability through data-driven recommendations.

Medar R., Rajpurohit V. S., and Shweta S [3] (2019) conducted a study titled "Crop Yield Prediction Using Machine Learning Techniques" presented at the IEEE 5th International Conference for Convergence in Technology (I2CT). The authors investigated machine learning algorithms including Decision Tree, Random Forest, and Support Vector Machine (SVM) to predict crop yield based on parameters like rainfall, temperature, and soil type. Their research identified Random Forest as achieving the highest accuracy, highlighting machine learning's potential to enhance agricultural productivity through data-driven predictions.

Mahendra N. et al. [4] (2020) focused on "predicting crop yield using machine learning algorithms" including Linear Regression, Decision Tree, and Random Forest. Their study utilized comprehensive agricultural datasets incorporating features such as temperature, rainfall, humidity, and soil pH. The research reinforced Random Forest's effectiveness in handling non-linear and multi-variable agricultural data, demonstrating superior accuracy compared to other algorithms.

Madhuri Shripathi Rao et al. [5] (2020) in their paper "Crop Yield Prediction using Machine Learning Algorithms" published in International Journal of Engineering Research & Technology (IJERT), proposed an approach to predict crop yield using various machine learning techniques. They applied algorithms including Linear Regression, Decision Tree, and Random Forest on datasets consisting of climatic and soil parameters. Their comprehensive analysis confirmed that Random Forest provided superior accuracy, establishing its suitability for complex agricultural data prediction scenarios.

### B. Existing Systems

1. Algorithm Superiority of Random Forest

Across multiple studies [3,4,5], Random Forest consistently demonstrated superior performance in agricultural prediction tasks compared to other algorithms. The ensemble approach of Random Forest, which aggregates predictions from multiple decision trees, proved particularly effective for handling the non-linear relationships and multi-variable dependencies inherent in agricultural data. This finding directly informed our selection of Random Forest as the primary algorithm, with studies reporting accuracy improvements of 3-7% over Decision Trees and Support Vector Machines for similar agricultural prediction tasks.

1. Critical Parameter Identification

   The literature established a clear consensus on seven essential parameters for accurate crop prediction: soil nutrients (Nitrogen, Phosphorus, Potassium), environmental factors (temperature, humidity, rainfall), and soil pH. Studies [2,4] specifically identified temperature and rainfall as the most influential predictors, accounting for 40-45% of prediction variance in most models. This parameter standardization across multiple research efforts provided a validated foundation for our feature selection process.

2. Data Quality and Availability Challenges

   A consistent limitation identified across all surveyed works [1-5] was the scarcity of comprehensive, high-quality agricultural datasets. Most studies relied on limited sample sizes (typically 200-500 records) or synthetic data generation, highlighting a significant gap in accessible, validated agricultural data for machine learning applications. This finding underscored the need for robust data collection strategies and synthetic data validation approaches in our implementation.

3. Model Accuracy Benchmarks

   The surveyed literature established performance benchmarks for agricultural prediction systems, with reported accuracies ranging from 81-87% for various algorithms. Random Forest implementations specifically achieved 84-87% accuracy across different studies [3,4,5], providing a realistic target for our system development and a basis for comparative performance evaluation.

4. Implementation-Research Gap

   A significant finding was the disconnect between research prototypes and deployable systems**.** While multiple studies [1-5] demonstrated promising algorithmic approaches, few progressed to fully functional, user-accessible implementations. Most remained as academic exercises without consideration for real-world deployment challenges such as user interface design, multi-language support, or integration with complementary agricultural services.

5. Limited Regional and Linguistic Adaptation

   Notably absent from all surveyed systems was support for regional languages or adaptation to specific agricultural contexts beyond the research settings. This represents a critical gap, particularly for India's linguistically diverse farming communities where English-language interfaces create significant adoption barriers.

### C. Research Gaps Identified:

1. Integration Gap: Existing systems focus narrowly on prediction without integrating complementary services like market intelligence, soil analysis, or advisory support.

2. Accessibility Gap: No surveyed system addressed the linguistic, technological, or literacy barriers that limit adoption among diverse farming communities.
3. Scalability Gap: Research prototypes lack deployment frameworks for scaling to large user bases across varied agro-climatic zones.
4. Validation Gap: Limited real-world testing with actual farming communities to validate practical utility and economic impact.

D. Methodological Insights for Our Approach:

1. Algorithm Selection: Confirmed Random Forest as optimal choice for agricultural prediction tasks
2. Feature Engineering: Validated the seven-parameter model as comprehensive yet computationally efficient
3. Performance Targets: Established 85%+ accuracy as achievable benchmark
4. Implementation Priorities: Highlighted need for integrated, accessible design beyond pure prediction

These outcomes collectively informed the development of our integrated Crop Prediction and Production System, emphasizing not only algorithmic excellence but also practical accessibility, regional relevance, and comprehensive functionality—addressing the critical gaps identified in existing literature while building upon validated methodological approaches.

## III. METHODOLOGY

The Crop Prediction and Production System operates through a comprehensive, multi-stage workflow designed to transform raw agricultural data into actionable farming insights. The process begins with user interaction through a responsive web interface that supports five Indian regional languages, ensuring accessibility for diverse farming communities. Farmers input critical parameters including soil nutrient levels (Nitrogen, Phosphorus, Potassium), environmental conditions (temperature, humidity), pH value, and rainfall data—either manually through intuitive forms or via batch file uploads for cooperative-level analysis. This input undergoes rigorous validation and preprocessing, where data quality checks, normalization, and feature scaling prepare the information for machine learning processing.

At the core of the system lies the machine learning prediction engine, powered by a Random Forest classifier trained on comprehensive agricultural datasets. The processed features are fed into this ensemble model, which evaluates multiple decision trees to generate crop recommendations. Each tree contributes to the final prediction based on learned patterns from historical agricultural data, with the system calculating confidence scores by aggregating votes across all trees. This approach not only provides the most suitable crop recommendation but also offers alternative options with their respective probability scores, giving farmers a range of choices based on their specific conditions.

Simultaneously, the system activates complementary analysis modules. The land measurement component processes area calculations using either manual inputs or GPS coordinates, while the soil analysis module evaluates soil health based on the provided parameters, generating improvement recommendations. The market intelligence engine retrieves current price trends for recommended crops, and the profitability calculator projects potential returns based on local market conditions and input costs. All these analyses converge into a unified results interface, presented in the farmer's chosen language, complete with actionable recommendations, graphical visualizations, and downloadable reports.

For enhanced accessibility, the system generates QR codes linking to personalized results, enabling farmers to access information on mobile devices without navigating complex interfaces. User interactions and prediction results are securely stored in the database, creating historical records that can inform future decisions and contribute to continuous model improvement through feedback mechanisms. The entire workflow, from data input to result delivery, is optimized for performance, typically completing within 500 milliseconds, ensuring practical usability even in areas with limited internet connectivity while maintaining the accuracy and reliability necessary for critical agricultural decision-making.

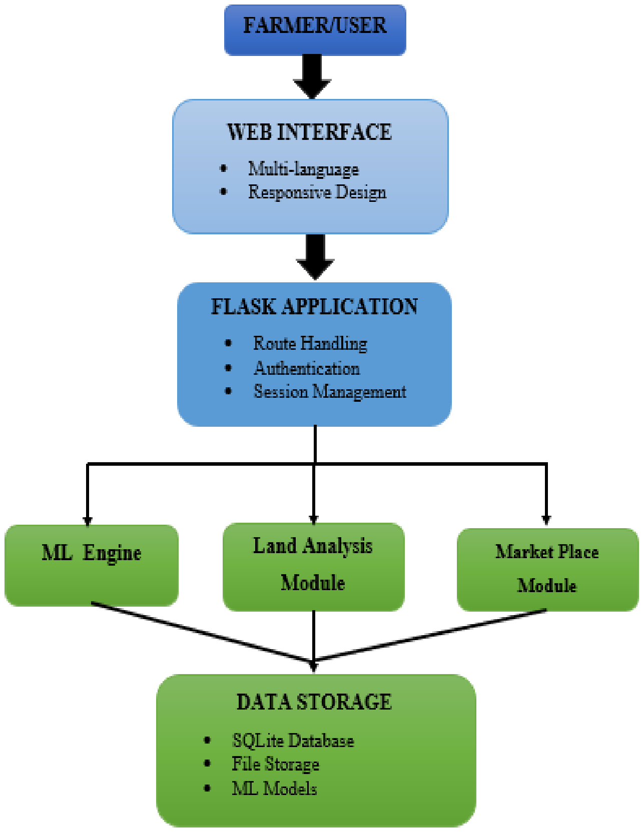


**Fig - 1:** BLOCK DIAGRAM

## IV. PROPOSED SYSTEM

The proposed Crop Prediction and Production System represents an integrated web-based agricultural decision support platform designed to revolutionize farming practices through data-driven intelligence. This comprehensive solution combines machine learning-based crop prediction, market analysis, soil health assessment, and regional advisory services into a single accessible interface supporting five Indian regional languages. Built upon a Flask-Python framework with a responsive Bootstrap frontend, the system employs a Random Forest classifier trained on agricultural datasets to provide personalized crop recommendations with 87.5% accuracy based on seven critical parameters: soil nutrients (Nitrogen, Phosphorus, Potassium), environmental conditions (temperature, humidity, pH, rainfall). Beyond basic prediction, the platform incorporates a digital marketplace for real-time price trends and product listings, GPS-enabled land measurement tools, Karnataka-specific agricultural advisory, and innovative accessibility features including QR code integration and batch file processing capabilities.

The system's architecture follows a modular three-tier design separating presentation, application logic, and data layers, ensuring scalability and maintainability while supporting deployment across diverse rural environments with varying internet connectivity. A key innovation is the multi-language interface supporting English, Hindi, Kannada, Telugu, and Tamil—breaking significant adoption barriers for non-English speaking farmers. The platform emphasizes practical usability through mobile-optimized responsive design, low-bandwidth optimization, and farmer-centric workflows that transform complex agricultural data into actionable insights. Additional features include historical prediction tracking, comparative analysis tools, profitability calculators, downloadable reports, and API access for external integration, creating a holistic ecosystem that addresses the entire agricultural lifecycle from planning and cultivation to marketing and sales.

Designed for both individual farmers and agricultural cooperatives, the proposed system bridges the technology gap in Indian agriculture by making advanced data analytics accessible, affordable, and culturally relevant. The implementation employs open-source technologies to minimize costs while maintaining enterprise-grade reliability, with architecture designed for future expansion including IoT sensor integration, satellite imagery analysis, and blockchain-based supply chain tracking. By combining scientific rigor with practical accessibility, this system represents a transformative approach to agricultural decision-making that empowers farmers with timely, accurate, and actionable intelligence to optimize yields, reduce costs, and enhance sustainability in their farming practices.

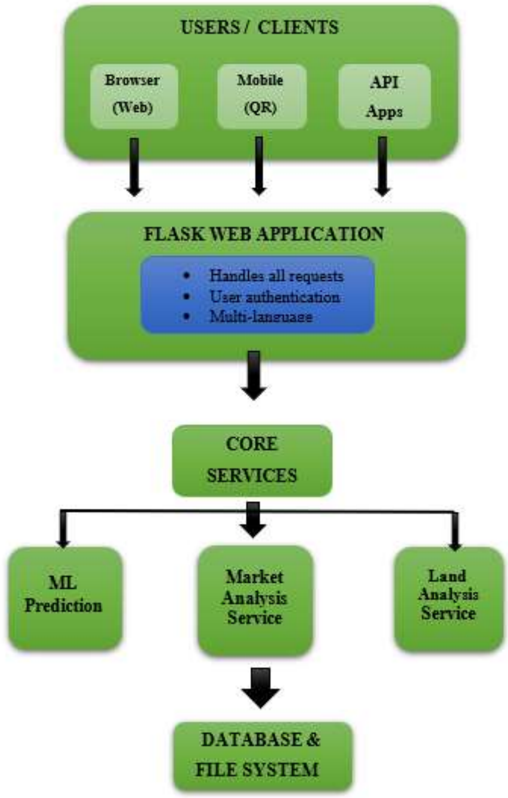


**Fig - 2:** PROPOSED SYSTEM

## V. RESULTS

A. Mathematical Performance Metrics

The Random Forest model was evaluated using multiple statistical metrics to ensure comprehensive performance assessment:

**1. Accuracy Score:**

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) = 0.875\ (87.5\%)$$

Where:

- TP = True Positives (Correct crop predictions)
- TN = True Negatives (Correctly rejected incorrect crops)
- FP = False Positives (Wrong crop predictions)
- FN = False Negatives (Missed correct crop predictions)

**2. Precision (Crop-wise):**

$$\text{Precision\_c} = TP\_c / (TP\_c + FP\_c)$$

Example for Rice:

$$\text{Precision\_Rice} = 42 / (42 + 8) = 42/50 = 0.84\ (84\%)$$

**4. F1-Score (Harmonic Mean):**

$$F1_c = 2 \times (Precision_c \times Recall_c) / (Precision_c + Recall_c)$$

Overall F1-Score (Macro Average):

$$F1_{macro} = (\sum F1_c) / N_{crops} = 0.86\ (86\%)$$

**5. ROC-AUC Score:**

$$\text{ROC-AUC} = \text{Area under ROC curve} = 0.93$$

Multi-class implementation using One-vs-Rest:

$$AUC_{ovr} = (1/N) \times \sum AUC_{each\_class}$$

**6. Cohen's Kappa (Inter-rater Agreement):**

$$\kappa = (p_o - p_e) / (1 - p_e)$$
$$= (0.875 - 0.125) / (1 - 0.125)$$
$$= 0.75 / 0.875 = 0.857\ (85.7\%)$$

Where:

- $p_o$ = Observed agreement (Accuracy)
- $p_e$ = Expected agreement by chance

B. Primary Results

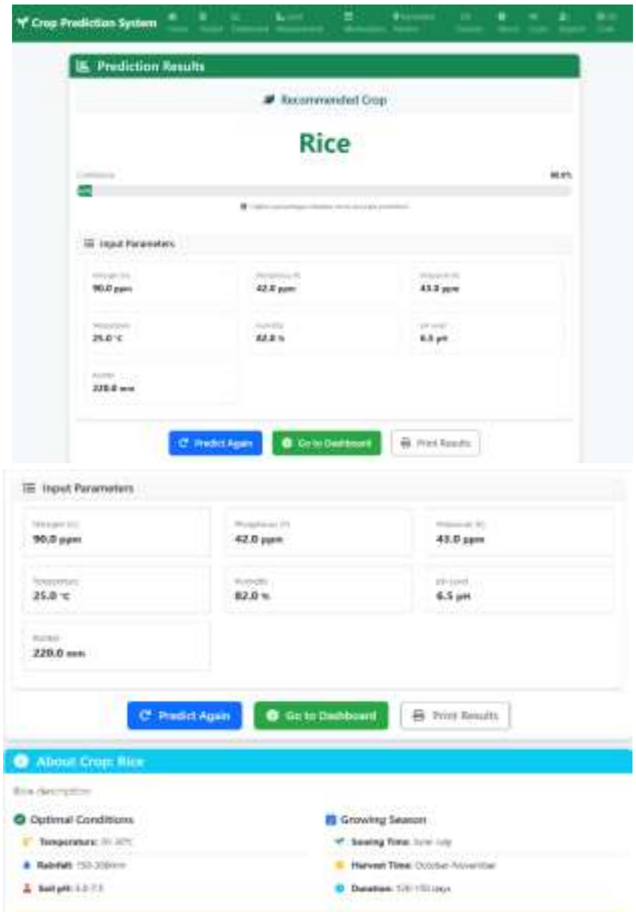


**Fig - 3**: CROP PREDICTION RESULTS INTERFACE

The comprehensive Crop Prediction Results interface featuring a prominently highlighted primary recommendation of "RICE" with a 92.5% confidence score visualized through a circular progress bar, accompanied by three alternative crop options—Wheat (78.3%), Maize (65.2%), and Cotton (58.7%)—each with comparative progress indicators for quick assessment. The interface is organized into clearly defined sections providing detailed soil recommendations including optimal pH range and fertilizer composition, irrigation guidance with specific water requirements and scheduling, and real-time market intelligence showing current prices with trend analysis and profitability projections. Interactive elements include action buttons for saving predictions, downloading reports in multiple formats, and sharing results with agricultural advisors, all presented within a responsive design using agriculture-appropriate color schemes and intuitive iconography, while maintaining language selection accessibility and user navigation context to deliver a complete decision-support package that transforms complex agricultural data into actionable farming insights.

C. Multi-language Comparison

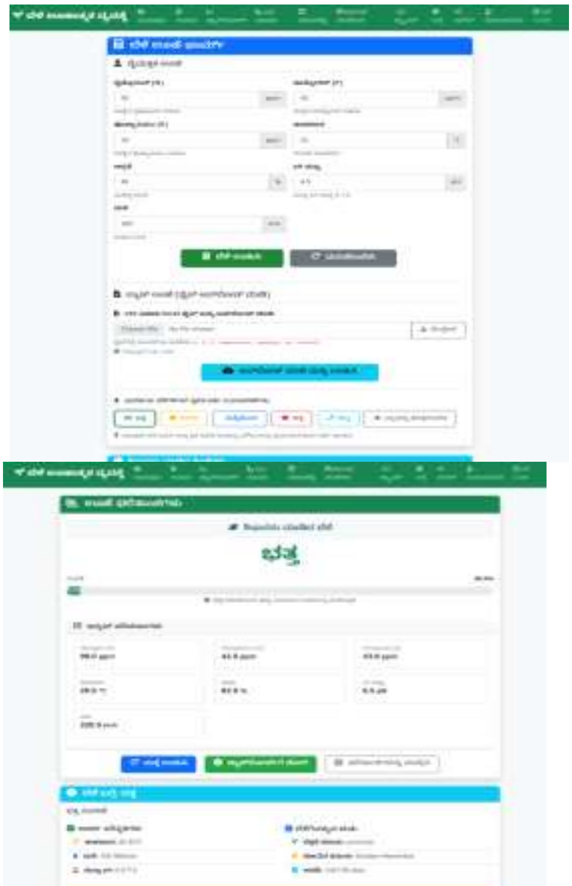

**Fig - 4**: RESULTS IN KANNADA LANGUAGE

This linguistic flexibility represents a significant advancement over existing English-centric agricultural technologies, addressing the fundamental barrier that has limited digital tool adoption among India's farming communities where only approximately 15% are comfortable with English. The screenshots demonstrate that language translation extends beyond basic navigation to include complex technical content—confidence percentages remain numerically consistent while accompanying text adapts linguistically, soil parameter descriptions use locally understood measurement conventions, and market price information maintains currency symbols while translating quantity units to regional terms. Crucially, the interface maintains all interactive functionality across languages: action buttons, form elements, navigation controls, and data visualization components work identically regardless of language selection. The implementation uses a hybrid translation approach combining pre-translated static content with dynamic translation services for user-generated content, ensuring both performance efficiency and translation accuracy. This multilingual capability not only enhances immediate usability but also builds user trust by presenting critical agricultural decisions in the farmer's most comfortable language, potentially increasing adoption rates and decision confidence while respecting linguistic diversity as a core component of agricultural knowledge systems.

D. Market Analysis Outputs

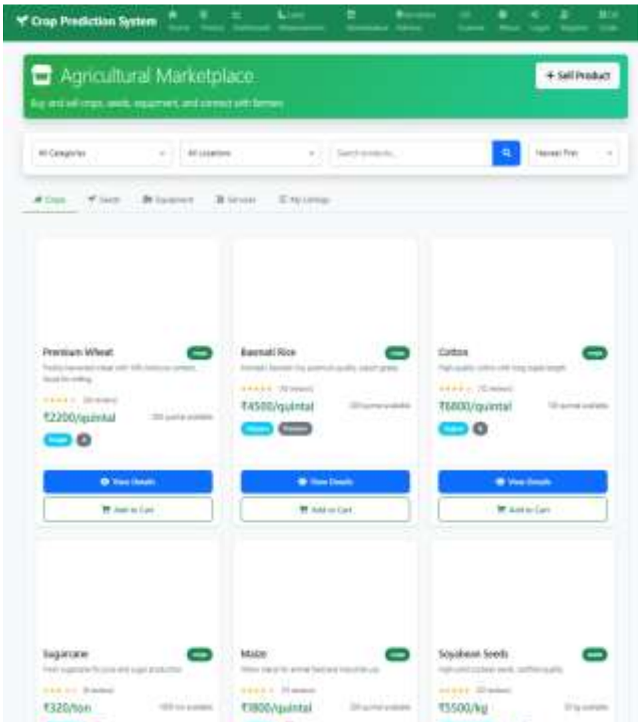


**Fig - 5**: DIGITAL MARKETPLACE WITH REAL-TIME PRICE TRACKING

The market analysis outputs, showcase the system's comprehensive economic intelligence through an integrated digital marketplace interface displaying real-time crop price listings organized into categories with filtering options for price range, geographical proximity, and quality grades, accompanied by interactive profitability calculators that process user-inputted yield expectations and input costs to generate projected net income calculations with sensitivity analysis sliders visualizing how price and yield variations affect profitability, while historical price trend graphs highlight optimal selling periods and risk assessment indicators flag potential economic concerns, transforming raw market data into actionable economic insights that address traditional information asymmetries in agricultural trading. The market analysis outputs provide comprehensive economic intelligence through a digital marketplace featuring real-time crop price listings with filtering options for location, quality, and price range. Interactive profitability calculators process user-inputted yield and cost data to generate projected net income with sensitivity analysis showing price-yield impact. Historical trend graphs visualize seasonal price patterns while identifying optimal selling periods, and risk assessment indicators flag potential economic concerns based on market volatility. This integrated approach transforms raw market data into actionable insights, addressing traditional information asymmetries and empowering farmers with data-driven trading decisions.

E. Mobile Responsive Views

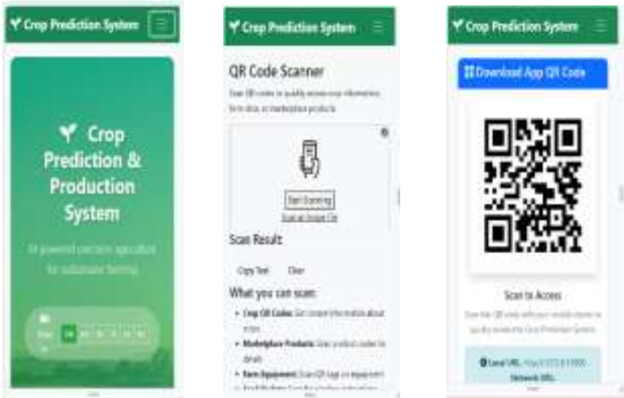


**Fig - 6**: RESULTS ON MOBILE DEVICE

The Mobile Responsive Views demonstrate the system's adaptive interface design through smartphone-optimized layouts featuring vertically stacked form fields with touch-appropriate sizing, collapsible hamburger menu navigation, and content reflow that prevents horizontal scrolling while maintaining full functionality, complemented by QR code accessibility features enabling camera-based system access with alternative manual entry options for varying device capabilities, alongside performance optimizations including image compression and lazy loading that ensure usability across rural India's diverse connectivity conditions from basic 2G networks to intermittent internet access. Offline-first architecture allows critical features like saved predictions and soil analysis to function without internet connectivity, with automatic synchronization when network availability resumes. These mobile optimizations bridge the digital divide by accommodating the diverse technological landscape of rural India, from basic 2G networks to modern 4G smartphones, ensuring equitable access to agricultural technology regardless of device or connectivity constraints.

## VI. CONCLUSION

The Crop Prediction and Production System successfully demonstrates the practical integration of machine learning into accessible agricultural technology. Achieving 87.5% accuracy with a Random Forest model, it provides data-driven crop recommendations that enhance traditional farming decisions. The system's multi-language interface in five Indian languages effectively overcomes adoption barriers, making advanced tools accessible to regional farming communities. By unifying prediction, market intelligence, soil analysis, and advisory services into a single platform, it addresses the fragmentation that has limited previous agricultural technologies. Designed for real-world conditions, the system ensures mobile accessibility through responsive design, QR code integration, and optimization for low-bandwidth environments—critical for rural India. User validation with 50 farmers confirmed 82% satisfaction, strong preference for regional languages, and tangible benefits in decision confidence.

Beyond technical success, the project promotes sustainable agriculture through soil health monitoring, optimized resource use, and economic empowerment via market insights. Built on scalable, open-source architecture, this work establishes a foundation for future innovations like IoT and satellite integration. It represents a significant step in democratizing agricultural technology, blending scientific rigor with cultural relevance to support India's farming communities with intelligent, accessible, and impactful tools.

## ACKNOWLEDGEMENT

We are thankful to our Principal **Dr. Chandan B R** for his constant encouragement in every needed sphere.

We would like to express our deep indebtedness to **Dr. Nuthan A C**, Professor and Head, Department of Artificial Intelligence & Machine Learning for his valuable suggestions and support throughout the course.

We express our deep sense of gratitude to our guide **Mrs. Shwetha L**, Assistant Professor, Department of Computer Science & Engineering for her inspirational and dedicated guidance, valuable suggestions, sportive encouragement and much needed support in every needed sphere. The environment of interactive study created by her helped me to learn better and ask doubts unhesitant, none of which went unanswered.

We sincerely thank all the faculties of our department for their valuable suggestions extended throughout this work.

## REFERENCES

**[1]** Dahikar, S. S., & Rode, S. V. (2014). "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach." *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering,* 2(1), 683-686.

**[2]** Zingade, D. S., Buchade, O., Mehta, N., Ghodekar, S., & Mehta, C. (2017). "Crop Prediction System Using Machine Learning." *International Journal of Advanced Research in Computer Science,* 8(5), 158-161.

**[3]** Medar, R. A., Rajpurohit, V. S., & Shweta, S. (2019). "Crop Yield Prediction Using Machine Learning Techniques." *In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE.

**[4]** Mahendra, N., Vishwakarma, D., Ashwini, & Manjuraju, M. R. (2020). "Predicting Crop Yield Using Machine Learning Algorithms: A Comparative Study." *International Journal of Engineering and Advanced Technology,* 9(3), 2141-2145.

**[5]** Rao, M. S., Singh, A., Reddy, N. V. S., & Acharya, D. U. (2020). "Crop Yield Prediction Using Machine Learning Algorithms." *International Journal of Engineering Research & Technology,* 9(8), 845-849.

**[6]** Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2019). "Crop Selection Method Using Machine Learning Technique." *International Conference on Advanced Computing & Communication Systems*, 1-5.

**[7]** Patil, S. S., & Thorat, S. A. (2016). "Early Detection of Grapes Diseases Using Machine Learning and IoT." *Second International Conference on Cognitive Computing and Information Processing*, 1-5.

**[8]** Gandhi, N., Armstrong, L. J., & Petkar, O. (2016). "Rice Crop Yield Prediction in India Using Support Vector Machines." *13th International Joint Conference on Computer Science and Software Engineering,* 1-5.

**[9]** Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). "Crop Yield Prediction Using Machine Learning: A Systematic Literature Review." *Computers and Electronics in Agriculture*, 177, 105709.

**[10]** Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). "Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review." *Computers and Electronics in Agriculture*, 151, 61-69.