International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

AI-Powered Agriculture Chatbot for Smart Farming Assistance

Chepuri Sailaja [1], Mr. M. Chiranjeevi [2]

[1] M. Tech Scholar, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology (Autonomous), Anakapalli District, Narsipatnam (R.D.), Andhra Pradesh, India

Email: sailchepuri@gmail.com

[2]Associate Professor, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology (Autonomous), Anakapalli District, Narsipatnam (R.D.), Andhra Pradesh, India

Email: chiru.mf@gmail.com

***Abstract*—** Agriculture plays a vital role in the economic development of many countries, especially in regions where farming is the primary source of livelihood. However, farmers often face challenges such as unpredictable weather conditions, lack of awareness about suitable crops, pest infestations, and limited access to expert advice. This paper presents an intelligent agriculture chatbot integrated with a crop prediction system using Convolutional Neural Networks (CNN), Random Forest, and Support Vector Machine (SVM) classifiers. The chatbot component employs Natural Language Processing (NLP) with TF-IDF vectorisation and cosine similarity to provide real-time conversational assistance on crop selection, fertiliser usage, pest control, irrigation, and market guidance. The CNN-based crop prediction module analyses environmental and soil parameters—including soil type, climate, season, and geographic location—to recommend the most suitable crop for cultivation. Comparative evaluation of three models reveals that CNN achieves 95% accuracy, Random Forest achieves 90%, and SVM achieves 80% on the crop prediction task. The hybrid system—combining deep learning prediction with NLP-driven chatbot interaction—provides a comprehensive, accessible, and scalable smart farming platform that bridges the technology gap for rural farmers.

Keywords: Agriculture chatbot, crop prediction, Convolutional Neural Network, Natural Language Processing, TF-IDF, smart farming, machine learning, precision agriculture, soil analysis, sustainable agriculture.

1. INTRODUCTION

Agriculture is undergoing a profound technological transformation driven by the convergence of artificial intelligence, machine learning, and data analytics. In many developing economies, including India, farming remains the primary livelihood for a significant portion of the population. Yet farmers in rural areas routinely face challenges that reduce productivity and increase risk: unpredictable weather patterns, soil degradation, pest outbreaks, and critically, the absence of timely access to agricultural expertise. Traditional knowledge transfer mechanisms—government extension services, agricultural universities, local experts—are insufficient in scale and timeliness to meet the information demands of millions of smallholder farmers.

Machine learning models have been widely deployed for crop prediction, disease detection,

and yield estimation. However, traditional algorithms such as Decision Trees and Random Forests, while useful, struggle to capture complex non-linear relationships in high-dimensional agricultural datasets. Deep learning, and specifically Convolutional Neural Networks (CNNs), has demonstrated superior pattern recognition capability in structured and unstructured data alike, enabling more accurate feature extraction and classification.

Complementing prediction capability with interactive accessibility is the agriculture chatbot—a conversational AI system enabling farmers to query the system in natural language without requiring technical knowledge. By integrating NLP-based query understanding with CNN-based crop prediction, the proposed AgriAI system creates a unified smart farming platform that delivers both data-driven crop recommendations and real-time advisory support through a single user interface.

This paper describes the design, implementation, and evaluation of such a hybrid intelligent system. The platform is implemented using Python and the Flask web framework, with models trained on agricultural datasets from Kaggle. Three classifiers—CNN, Random Forest, and SVM—are comparatively evaluated on crop prediction accuracy, and the chatbot is trained using TF-IDF vectorisation on a dataset of 20,000+ agricultural inquiry-response pairs.

1.1 Motivation

The motivation for this research is rooted in the persistent digital divide between advanced AI-driven agricultural tools and the rural farming communities who need them most. Existing precision agriculture systems are typically designed for technically proficient users or large commercial farms equipped with IoT infrastructure. Smallholder farmers in India—who cultivate the majority of the country's agricultural land—lack access to personalised, real-time agronomic guidance. The development of a voice-and-text accessible chatbot integrated with an AI crop prediction engine addresses this gap directly. Furthermore, CNN's demonstrated superiority over traditional classifiers in feature extraction from agricultural data provides the technical motivation for its deployment as the core prediction engine.

1.2 Problem Statement

Despite the availability of crop recommendation systems and agricultural advisory platforms, the following critical gaps remain unaddressed: (1) Traditional ML models fail to capture complex non-linear environmental-crop relationships, reducing prediction accuracy; (2) Existing chatbot systems lack integration with predictive ML engines, providing only static FAQ-style responses; (3) Current systems do not support natural language interaction for non-technical rural users; (4) Agricultural recommendation tools are rarely accessible via mobile or web platforms in Tier 2/Tier 3 regions; (5) No unified platform currently combines real-time conversational AI with personalized, region-specific, season-aware crop recommendations using deep learning. This study addresses these five gaps through a fully integrated, deployable AgriAI system.

1.3 Objectives

The objectives of this study are:

- To develop a CNN-based crop prediction model trained on regional soil, climate, and season parameters achieving measurably higher accuracy than traditional ML classifiers.
- To build an NLP-driven chatbot trained on 20,000+ agricultural inquiry-response pairs capable of answering queries on crop selection, pest control, fertiliser usage, irrigation, and weather guidance.
- To integrate the chatbot and prediction modules into a unified web platform using Flask, supporting real-time user interaction and on-demand crop recommendation.
- To comparatively evaluate CNN, Random Forest, and SVM classifiers on accuracy, F1-score, and recall metrics for the crop prediction task.
- To design a scalable, deployable system architecture extensible to mobile platforms, IoT sensor integration, and multilingual support.

2. LITERATURE REVIEW

[1] Ramesh and Vydeki (2020) proposed an optimised deep neural network combined with the Jaya optimisation algorithm for recognition and classification of paddy leaf diseases, achieving high accuracy in multi-class plant

disease classification. Their work demonstrates that deep learning models with optimisation heuristics outperform conventional classifiers on agricultural image datasets [1].

[2] Kamilaris and Prenafeta-Boldú (2018) conducted a comprehensive survey of deep learning applications in agriculture, encompassing crop yield prediction, plant disease detection, weed identification, and livestock monitoring. Their analysis confirms CNN as the dominant architecture for agricultural image-based classification tasks and motivates its application in the crop recommendation domain [2].

[3] Too et al. (2019) comparatively evaluated fine-tuned deep learning models including VGG, ResNet, and DenseNet for plant disease identification, concluding that transfer learning significantly improves classification accuracy with limited training data—a finding applicable to agricultural datasets where labelled samples are scarce [3].

[4] Chen et al. (2018) developed a deep learning-based detection system for rice plant diseases using image analysis, achieving detection accuracy exceeding 90%, and demonstrating that CNN-based systems can provide practical disease detection support comparable to expert diagnosis [4].

[5] Ferentinos (2018) demonstrated that CNN models trained on the PlantVillage dataset achieve classification accuracy of up to 99.53% for plant disease detection, establishing deep learning as the state-of-the-art approach for image-based agricultural diagnostics [5].

[6] Tiwari, Shukla, and Tiwari (2020) surveyed machine learning approaches for crop yield prediction, evaluating Random Forest, SVM, and neural network models on agricultural datasets, confirming ensemble methods and deep learning models as superior predictors compared to linear statistical approaches [8].

[7] Devlin et al. (2019) introduced BERT, a transformer-based pre-trained language model that achieves state-of-the-art performance across NLP tasks. Their architecture informs the chatbot development direction for future enhancement, where agricultural query understanding can be improved through fine-tuned BERT embeddings replacing TF-IDF representations [17].

[8] Jeong et al. (2016) applied Random Forest classifiers for regional and global crop yield prediction, confirming their effectiveness as interpretable baseline models for crop recommendation systems when combined with soil and climate feature engineering [10].

[9] LeCun, Bengio, and Hinton (2015) established the theoretical foundations of deep learning, demonstrating that hierarchical feature learning through multiple neural network layers enables performance surpassing human-level accuracy in pattern recognition tasks—validating CNN's application in complex agricultural prediction scenarios [13].

[10] Singh et al. (2018) reviewed deep learning approaches for plant stress phenotyping, identifying CNN and LSTM architectures as particularly effective for temporal agricultural data modelling. Their findings motivate the inclusion of hybrid CNN-RNN architectures in the proposed system's future enhancement roadmap [11].

3. EXISTING SYSTEM

Existing crop prediction and agricultural advisory systems primarily employ three categories of approaches, each with distinct technical capabilities and operational limitations.

Traditional ML-Based Crop Prediction: Systems employing Decision Trees, Random Forest, SVM, and K-Nearest Neighbors (KNN) analyse structured soil and climate inputs to recommend crops. These models are trained on manually engineered feature sets comprising parameters such as N-P-K ratios, rainfall, temperature, and humidity. While functional for standard scenarios, traditional ML algorithms are limited in their ability to model complex non-linear interactions between environmental parameters and crop suitability. They require manual feature selection, are susceptible to overfitting on high-dimensional datasets, and their performance degrades when generalising to new geographic regions or crop varieties not represented in training data.

Static Rule-Based Advisory Systems: Many agricultural extension services provide digitised decision trees or condition-action rule sets for crop and fertiliser recommendations. These systems offer no adaptability—they cannot learn from new data, detect novel pest

patterns, or adjust to climate variability. Their advice is bounded by expert-defined rule sets, which quickly become outdated as agricultural conditions change.

Basic Chatbot Implementations: Early agricultural chatbots rely on keyword matching and pattern-based response retrieval without genuine NLP understanding. These systems cannot handle paraphrased queries, contextual follow-up questions, or domain-specific vocabulary variations. Critically, they operate in isolation from predictive ML models, providing only static FAQ-style responses without access to data-driven crop recommendations. The absence of conversational intelligence limits their utility for diverse farmer queries.

The collective limitations of existing systems include: inability to capture complex non-linear agricultural data relationships; absence of real-time adaptive learning; disconnected advisory and prediction capabilities; lack of user-friendly natural language interfaces; and limited accessibility for non-technical rural users operating on mobile devices with intermittent connectivity.

4. PROPOSED SYSTEM

The proposed AgriAI system is a hybrid intelligent platform integrating a CNN-based crop prediction engine with an NLP-driven agricultural chatbot, deployed through a Flask web application with a responsive HTML/CSS frontend. The system architecture comprises five functional layers working in a coordinated pipeline.

4.1 Data Collection and Dataset

The crop prediction module uses the Kaggle crop prediction dataset (crop_prediction_data.csv), comprising records with four categorical features: place (Indian geographic region), soil type, climate condition, and season—mapped to crop labels. The chatbot training dataset (chatbot_data.csv) contains 20,000+ agricultural inquiry-response pairs covering pest control, fertiliser recommendations, weather guidance, irrigation methods, and crop-specific advice. Data is loaded, validated for completeness, and passed to the preprocessing pipeline.

4.2 Preprocessing and Feature Engineering

Categorical features (place, soil, climate, season, crop) are encoded using scikit-learn's LabelEncoder. Feature vectors are standardised using StandardScaler to zero mean and unit variance before model training. For the chatbot, text data undergoes TF-IDF (Term Frequency-Inverse Document Frequency) vectorisation with English stop-word removal. Outlier detection and duplicate removal are applied to both datasets to ensure training data quality. The prediction dataset is split 80%/20% into training and test sets using stratified random sampling.

4.3 Model Training and Architecture

CNN Model: A 1D-Convolutional Neural Network is implemented using TensorFlow/Keras. The architecture comprises: Conv1D layer (64 filters, kernel size 2, ReLU activation), Flatten layer, Dense hidden layer (64 neurons, ReLU), Dropout layer (rate 0.2 for regularisation), and output Dense layer (softmax, neurons = number of crop classes). The model is compiled with the Adam optimiser, sparse categorical cross-entropy loss, and trained for 5 epochs with batch size 32. Trained model is serialised as cnn_model.h5.

Random Forest Model: An ensemble of 100 decision trees (RandomForestClassifier, n_estimators=100) is trained on the scaled feature matrix. The model provides high interpretability through feature importance ranking and serves as the performance baseline. Serialised as rf_model.pkl.

SVM Model: A Support Vector Classifier with RBF kernel and probability calibration (SVC, probability=True) is trained on a 10,000-sample subset for computational efficiency. Serialised as svm_model.pkl.

Chatbot Model: TF-IDF vectorisation transforms all chatbot training questions into a TF-IDF matrix. At inference time, user query vectors are compared against this matrix using cosine similarity; the response corresponding to the highest-similarity question (threshold > 0.2) is returned. The vectoriser and TF-IDF matrix are serialised as chatbot_model.pkl.

4.4 System Integration and Deployment

All four model artefacts are loaded at startup by the Flask application (app.py). The system exposes RESTful API endpoints: /ask (POST)

for chatbot queries, /predict (POST) for crop prediction, /dashboard (GET) for performance analytics. The frontend is built with HTML5, CSS3, and Chart.js, providing three user-facing modules: Agriculture ChatBot interface, Crop Prediction Engine (dropdown-based input), and Performance Dashboard (accuracy bar chart, F1-score line chart, recall horizontal bar chart). User authentication (login/signup) is also supported.

4.5 System Requirements

Category	Specification
Processor	Intel Core i5 / Pentium IV 2.4 GHz or above
RAM	Minimum 4 GB (8 GB recommended for CNN training)
Storage	80 GB HDD / SSD
Operating System	Windows 10 / Ubuntu 20.04 or above
Frontend	HTML5, CSS3, JavaScript, Chart.js
Backend Framework	Python 3.8+, Flask
ML/DL Libraries	TensorFlow 2.x, Keras, Scikit-learn, Pandas, NumPy
Database	MS SQL / SQLite (user management)
Cloud Deployment	AWS / Google Cloud / Azure (optional)

TABLE I. HARDWARE AND SOFTWARE REQUIREMENTS

4.6 Advantages of the Proposed System

- CNN achieves 95% crop prediction accuracy—5–15% above traditional ML baselines—through automatic hierarchical feature extraction from agricultural parameters.
- NLP chatbot with TF-IDF and cosine similarity enables natural language query handling across 20,000+ agricultural topics without keyword-matching constraints.
- Unified Flask platform integrates prediction and chatbot in one deployable web application accessible on desktop and mobile browsers.
- Scalable architecture supports future integration with IoT sensor feeds, weather APIs, satellite imagery, and voice-based input for semi-literate farmer users.
- Multilingual extension and offline-capable mobile deployment are architecturally planned as near-term enhancements.

5. RESULTS AND DISCUSSIONS

5.1 Model Performance Comparison

Table II presents the comparative evaluation results for the three crop prediction classifiers across accuracy, F1-score, and recall metrics, evaluated on the 20% held-out test set.

Model	Accuracy (%)	F1-Score	Recall
Convolutional Neural Network (CNN)	95.00%	0.94	0.93
Random Forest Classifier	90.00%	0.89	0.88
Support Vector Machine (SVM)	80.00%	0.79	0.78

TABLE II. CROP PREDICTION MODEL PERFORMANCE COMPARISON

5.2 Dataset Statistics

Table III summarises key characteristics of the training datasets used for both system components.

Dataset Component	Details
Chatbot Training Records	20,000+ inquiry-response pairs
Crop Prediction Records	50,000 crop data entries (estimated)
Geographic Coverage	Multiple Indian regions (states/districts)
Unique Crop Classes	Multiple varieties across Kharif/Rabi/Zaid seasons
Input Features (Prediction)	Place, Soil Type, Climate, Season
Chatbot Topics Covered	Pest control, fertiliser, weather, irrigation, markets

TABLE III. DATASET OVERVIEW

5.3 Test Case Validation

Table IV presents system-level test case results covering user registration, admin authentication, and core functionality validation.

S.No	Test Case	Input	Expected Result	Actual Result	Status
1	User Registration	All fields filled	Registration success	Registration successful	PASS

S.No	Test Case	Input	Expected Result	Actual Result	Status
2	User Registration	Missing mandatory field	Registration rejected	Registration unsuccessful	PASS
3	Admin Login	Valid credentials	Admin dashboard loaded	Admin home page opened	PASS
4	Chatbot Query	"How to treat fungus in wheat?"	Relevant advisory response	Correct response returned	PASS
5	Crop Prediction	Place, Soil, Climate, Season	Crop recommendation	CNN prediction displayed	PASS
6	Low-similarity Query	Unrecognised query text	Fallback message returned	Fallback response shown	PASS

TABLE IV. SYSTEM TEST CASE RESULTS

5.4 Discussion

The CNN model outperforms both Random Forest (90%) and SVM (80%) on crop prediction accuracy. This performance advantage is attributable to CNN's ability to learn hierarchical feature representations directly from encoded input vectors through convolutional operations—capturing non-linear interactions between soil type, climate, and season that linear or tree-based models approximate less precisely. The Dropout layer (rate 0.2) and Adam optimiser effectively mitigate overfitting while maintaining training efficiency across 5 epochs.

The Random Forest model (90% accuracy) provides a strong ensemble baseline and offers interpretable feature importance rankings, revealing that geographic location (place) and soil type are the highest-importance predictors of crop suitability—a finding consistent with domain knowledge. The SVM (80%), while computationally efficient for small-scale real-time predictions, shows reduced accuracy due to the non-linear complexity of the feature-crop relationship that the RBF kernel partially but insufficiently captures.

The chatbot demonstrates strong coverage across the 20,000+ training topics, with cosine similarity thresholding (0.2) effectively balancing recall against inappropriate response delivery for unrecognised queries. Queries involving pesticide management, seasonal crop selection, and irrigation scheduling are handled with high response relevance. The primary limitation identified is sensitivity to highly paraphrased queries that diverge substantially from training question phrasing—a gap addressable through BERT-based embedding replacement in future iterations.

The integrated Flask platform successfully demonstrates end-to-end functionality: a farmer selects their geographic region, soil type, climate, and current season via dropdown menus; the CNN model returns a crop recommendation in under two seconds; and the chatbot simultaneously handles follow-up questions about cultivation practices for the recommended crop. The real-time analytics dashboard transparently displays all three model accuracy metrics, fostering user trust through model explainability.

6. CONCLUSION

This paper presented AgriAI—an intelligent smart farming platform integrating a CNN-based crop prediction engine with an NLP-driven agricultural chatbot. The system was developed using Python, TensorFlow/Keras, scikit-learn, and Flask, trained on Kaggle agricultural datasets, and evaluated across three ML/DL classifiers.

The CNN model achieved 95% crop prediction accuracy—the highest among all evaluated classifiers—demonstrating deep learning's superiority over traditional algorithms in capturing complex environmental-to-crop mappings. The TF-IDF chatbot effectively handled 20,000+ agricultural inquiry categories, providing real-time conversational guidance on pest management, fertiliser selection, irrigation, and market conditions through a user-friendly web interface.

The unified platform successfully bridges the expertise gap for rural farmers, delivering both data-driven precision recommendations and natural language advisory support without requiring technical knowledge. The system's scalable architecture supports future extension with IoT sensor integration, real-time weather API feeds, voice-based multilingual input, image-based disease detection using CNN plant pathology models, and BERT-enhanced chatbot understanding for improved paraphrase handling.

AgriAI establishes that hybrid AI systems combining deep learning prediction with conversational AI can meaningfully improve

agricultural decision-making, productivity, and economic outcomes for smallholder farmers—demonstrating that AI-driven solutions can bridge the technology-agriculture divide in developing economies.

REFERENCES

[1] S. Ramesh and D. Vydeki, "Recognition and classification of paddy leaf diseases using optimized deep neural network with Jaya algorithm," Information Processing in Agriculture, vol. 7, no. 2, pp. 249–260, 2020.

[2] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[3] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[4] J. Chen, D. Zhang, A. Nanehkaran, and D. Li, "Detection of rice plant diseases based on deep learning," Journal of Agricultural Science, vol. 10, no. 5, pp. 42–50, 2018.

[5] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[6] S. Sladojevic et al., "Deep neural networks based recognition of plant diseases by leaf image classification," Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801.

[7] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[8] P. Tiwari, S. Shukla, and R. Tiwari, "Crop yield prediction using machine learning approaches," International Journal of Engineering Research & Technology, vol. 9, no. 3, pp. 506–510, 2020.

[9] R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques," International Journal of Computer Science and Information Security, vol. 14, no. 5, pp. 452–458, 2016.

[10] J. H. Jeong, J. P. Resop, N. D. Mueller, et al., "Random forests for global and regional crop yield predictions," PLOS ONE, vol. 11, no. 6, 2016.

[11] A. K. Singh, B. Ganapathysubramanian, S. Sarkar, and A. Singh, "Deep learning for plant stress phenotyping: Trends and future perspectives," Trends in Plant Science, vol. 23, no. 10, pp. 883–898, 2018.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NIPS, 2012, pp. 1097–1105.

[15] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016. [Online]. Available: https://www.tensorflow.org/

[16] F. Chollet, "Keras: Deep learning library for Python," 2015. [Online]. Available: https://keras.io/

[17] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019.

[18] A. Vaswani et al., "Attention is all you need," in Proc. NIPS, 2017.

[19] S. Young, M. Gašić, B. Thomson, and J. D. Williams, "POMDP-based statistical spoken dialogue systems: A review," Proceedings of the IEEE, vol. 101, no. 5, pp. 1160–1179, 2013.

[20] R. Yan, Y. Song, and H. Wu, "Learning to respond with deep neural networks for retrieval-based human-computer conversation system," in Proc. SIGIR, 2016.

AUTHORS



Chepuri Sailaja, Pursuing MTech,
Department of Computer Science and Engineering,
Avanthi Institute of Engineering and Technology,
Makavarapalem, Narsipatnam, Anakapalle-531113
Andhra Pradesh, India. Her Main Research interest areas include Artificial Intelligence, Machine learning and Agentic AI, Large Language Models, Cloud Computing.



M.CHIRANJEEVI is working as an Associate Professor in the Department of Computer Science and Engineering at Avanthi Institute of Engineering & Technology, Makavarapalem,AP,India. His received the B.Tech degree in Computer
Science and Engineering from JNTU, Hyderabad and received the M.Tech degree in
Computer Science and Technology
from Berhampur University.
He has a total of 16 years of Teaching Experience. His research interests include
Data Warehousing and Data Mining, Cloud Computing,Network Security and RDBMS.He has Published more than 15 papers in various national and internationals journals.