ENHANCING AGRICULTURAL DECISION-MAKING WITH COMBINED LANGUAGE MODELS

N. Anand Babu[1], Harshavardhan Kaduduri[2], Jogu Harsha Vardhan[3],

Machkuri Sai Prasuna[4]

Email: anandbabu.csegnit@gniindia.org, kaduduriharshavardhan@gmail.com, joguharshavardhan11@gmail.com, saiprasuna42@gmail.com

Assistant Professor, Department of CSE[1]
Students, Department of CSE[2,3,4]
Guru Nanak Institute of Technology, Hyderabad, Telangana, India

Abstract: Agricultural decision-making depends on timely and accurate knowledge, yet many farmers lack access to expert guidance, leading to reduced productivity and crop losses. This paper presents a lightweight and scalable agricultural question-answering system that utilizes DeBERTa-v3 sentence embeddings for semantic representation and FAISS-based similarity search for efficient answer retrieval. Unlike generative approaches, the system retrieves responses from a verified knowledge base, ensuring factual accuracy and eliminating hallucinated outputs. The model is optimized for CPU-only environments, making it suitable for deployment in resource-constrained settings. Experimental results demonstrate improved retrieval precision, faster inference, and higher reliability compared to conventional methods. The proposed system provides an effective AI-driven advisory solution to enhance accessibility and support sustainable agricultural practices.

Keywords: Agricultural Question Answering, DeBERTa-v3, Sentence Embeddings, FAISS Indexing, Semantic Retrieval, Sustainable Farming

I.INTRODUCTION

Agriculture plays a crucial role in ensuring food security, economic stability, and rural livelihoods. Farmers often face challenges such as crop diseases, pest infestations, nutrient deficiencies, and limited awareness of government schemes, making timely and accurate advisory support essential. With the increasing use of digital platforms, large volumes of farmer queries are generated through call centres, mobile applications, and online portals. However, delivering reliable, consistent, and domain-specific responses remains difficult due to the complexity and diversity of agricultural problems.

Recent advancements in natural language processing and machine learning have enabled automated agricultural question-answering systems. Although generative large language models show strong language understanding capabilities, they often suffer from issues such as hallucination, inconsistency, and high computational cost in real-world applications. To overcome these challenges, this work proposes a retrieval-based agricultural question-answering system that uses semantic similarity to provide accurate, expert-validated responses, ensuring reliability, efficiency, and practical usability for farmers.

II. LITERATURE SURVEY

J. Singh, A. Verma & P. Kumar (2024) This study analyses the evolution of agricultural question-answering systems from traditional rule-based and keyword-based methods to advanced transformer-based approaches. It highlights that earlier systems lacked contextual understanding and struggled with domain-specific agricultural terminology. Although modern models improve semantic understanding, they often suffer from hallucination and data limitations. The authors recommend retrieval-based systems using curated knowledge bases to ensure accurate, reliable, and scalable solutions for farmers.

L. Chen, B. Patil & M. Sharma (2025) This research introduces a deep learning–based retrieval system that encodes farmer queries and expert knowledge into semantic embeddings. The system retrieves the most relevant answers from a structured database, ensuring consistency and correctness. It outperforms keyword-based methods and works efficiently on CPU-only systems, making it suitable for real-world agricultural applications.

A. Reddy, S. Sharma & T. Gupta (2024) This study evaluates lightweight AI models for agricultural advisory in resource-limited environments. It highlights that large generative models are computationally expensive and unsuitable for rural deployment. Instead, embedding-based retrieval systems provide faster responses, lower memory usage, and better energy efficiency. The authors conclude that lightweight models combined with semantic retrieval techniques are ideal for building scalable and reliable agricultural advisory systems.

M. Zhang, L. Wang & Y. Zhou (2024) This paper explores the use of transformer-based models to improve semantic understanding in agricultural systems. It emphasizes the importance of domain-specific fine-tuning, as general models struggle with agricultural terminology. The study shows that advanced embeddings significantly improve accuracy in tasks like crop disease detection and fertilizer recommendation. It concludes that models like DeBERTa enhance contextual understanding and retrieval performance in agricultural QA systems.

III. EXISTING SYSTEM

The existing agricultural question-answering systems are primarily based on transformer models such as BERT, Agricultural-BERT, and LLaMA. These systems use generative and ensemble approaches to understand user queries and produce responses. They are capable of handling complex natural language inputs and extracting contextual meaning, which improves query understanding compared to traditional methods. However, these systems rely heavily on large-scale models that require high computational power, memory, and often GPU support. The generative nature of these models can lead to hallucinated or vague responses that are not always factually accurate. Additionally, they involve longer training and inference times, making real-time response generation difficult. Due to these limitations, existing systems are less efficient, harder to deploy, and not well-suited for rural or low-resource environments where fast, reliable, and cost-effective solutions are needed.

Existing System Disadvantages

- High computational cost and requires powerful hardware (GPU, high memory)
- Generates hallucinated or inaccurate responses
- Difficult to deploy in rural or low-resource environments
- Increased system complexity and maintenance

Proposed System

The proposed system is a retrieval-based agricultural question-answering model that uses DeBERTa-v3 sentence embeddings and FAISS similarity search. Instead of generating answers, it retrieves the most relevant and expert-validated responses from a structured knowledge base. This approach ensures factual accuracy, reduces computational cost, and provides fast and reliable results. The system is lightweight, scalable, and designed to work efficiently in CPU-only and low-resource environments.

Proposed System Advantages:

- Eliminates hallucinated and inaccurate responses
- Provides fast and reliable answer retrieval
- Low computational cost and memory usage
- Works efficiently on low-resource systems (CPU-only)

IV. SYSTEM ARCHITECTURE

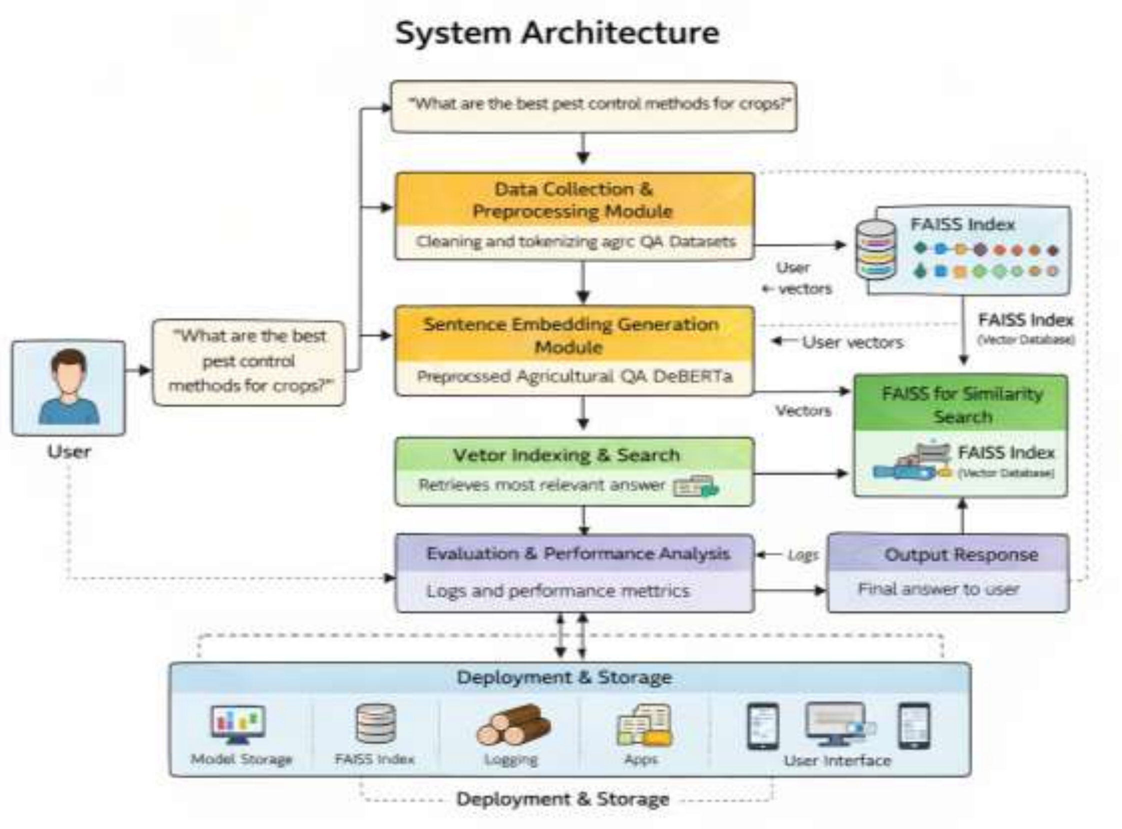

Figure 1 : SYSTEM ARCHITECTURE

The system architecture consists of several components working together to deliver accurate agricultural answers. The User Interface Layer allows users to enter queries, which are cleaned and tokenized by the Text Preprocessing Module. The processed query is then converted into semantic embeddings using the DeBERTa-v3 Encoder. These embeddings are matched against stored knowledge using the FAISS Similarity Engine, which quickly identifies the closest answers from the Answer Database containing pre-encoded agricultural information. Finally, the Result Module displays the top and most relevant answer to the user, ensuring fast, factual, and reliable advisory support.

Methodology
Modules Name:

- Data Collection and Preprocessing Module
- Sentence Embedding Generation Module
- Similarity Search and Retrieval Module
- Answer Selection Module
- Evaluation and Performance Analysis Module

- Model Storage and Deployment Module

1. Data Collection and Preprocessing Module: This module is responsible for loading the agricultural question–answer datasets, including training, validation, and testing data. The data is cleaned by removing duplicates, handling missing values, and normalizing text through lowercasing and trimming. Proper preprocessing ensures that the input data is consistent and suitable for semantic representation, which directly impacts retrieval accuracy.

2. Sentence Embedding Generation Module: In this module, farmer questions are converted into dense numerical vectors using a pretrained sentence transformer model. These embeddings capture the semantic meaning of the text rather than relying on exact keyword matching. By representing queries in a high-dimensional embedding space, the system can effectively identify semantically similar questions even when wording differs.

3. Similarity Search and Retrieval Module: This module performs fast similarity search using a vector indexing mechanism based on cosine similarity. A FAISS index is constructed using the embeddings of training questions. When a new query is received, its embedding is compared against the indexed vectors to retrieve the most similar historical question. This enables efficient and accurate retrieval even for large datasets.

4. Answer Selection Module: Once the most relevant historical question is identified, this module retrieves its corresponding expert-validated answer. Since the answers are sourced from previously verified advisory data, the system ensures factual correctness and consistency. This module eliminates the risk of hallucination by avoiding answer generation.

5. Evaluation and Performance Analysis Module: This module evaluates the effectiveness of the system using retrieval accuracy metrics based on similarity thresholds. Both training and testing datasets are used to assess system performance. Additionally, qualitative analysis is performed by comparing retrieved answers with ground truth responses to ensure agricultural relevance and correctness.

6. Model Storage and Deployment Module: The final module handles saving the trained sentence embedding model, FAISS index, and processed datasets for future use. This enables easy loading and deployment of the system in real-time applications such as web portals or mobile advisory platforms. The modular design allows seamless updates when new agricultural data becomes available.

V. IMPLEMENTATION

The proposed system is implemented using a retrieval-based approach that combines Natural Language Processing techniques with efficient similarity search. The implementation begins with data collection and preprocessing, where agricultural question–answer datasets are cleaned, normalized, and prepared for further processing. Next, sentence embeddings are generated using the DeBERTa-v3 model, which converts user queries and stored questions into dense vector representations capturing semantic meaning. These embeddings are indexed using FAISS, enabling fast and efficient similarity search. When a user submits a query, it is pre-processed and converted into an embedding, which is then compared with stored vectors in the FAISS index to find the most relevant match. The corresponding expert-validated answer is retrieved from the knowledge base and displayed to the user. The system is developed using Python with libraries such as PyTorch, Transformers, and FAISS, and integrated with a Flask-based web interface for user interaction. It is optimized to run on CPU-only environments, ensuring low computational cost and easy deployment in real-world agricultural settings.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Algorithm Used

Existing Algorithm

The existing system primarily uses transformer-based models such as BERT and LLaMA for processing agricultural queries. Initially, the input query is pre-processed by cleaning and tokenizing the text. The processed query is then passed through a transformer model, which converts it into contextual embeddings and understands the semantic meaning. In generative-based systems, the model predicts and generates a response based on learned patterns from large datasets. In some cases, ensemble techniques are used to improve accuracy by combining multiple models. The generated output is then returned as the final response to the user. Although this approach provides good language understanding, it involves heavy computation, longer processing time, and may produce inconsistent or hallucinated answers due to its generative nature.

Proposed Algorithm

The proposed system follows a retrieval-based approach for answering agricultural queries. Initially, the input query is pre-processed by cleaning and normalizing the text. The processed query is then converted into a semantic vector using the DeBERTa-v3 embedding model. All stored agricultural questions are also pre-converted into embeddings and indexed using FAISS for efficient similarity search. When a user submits a query, its embedding is compared with the indexed vectors to find the most similar question based on cosine similarity. Once the closest match is identified, the corresponding expert-validated answer is retrieved from the knowledge base and presented to the user. This approach avoids generating new responses and ensures accurate, consistent, and fast results

VI. EXPERIMENTAL RESULTS

Landing Page

Figure 2 : Landing Page

The landing page of the agricultural advisory system provides a brief overview of the project and acts as the entry point for users. It includes a navigation bar with options like home, register, and login. The page displays the system title and an abstract explaining how it helps farmers by providing accurate answers using embedding models and FAISS-based retrieval. Overall, it is designed to be simple, informative, and user-friendly.

Prediction Result Page

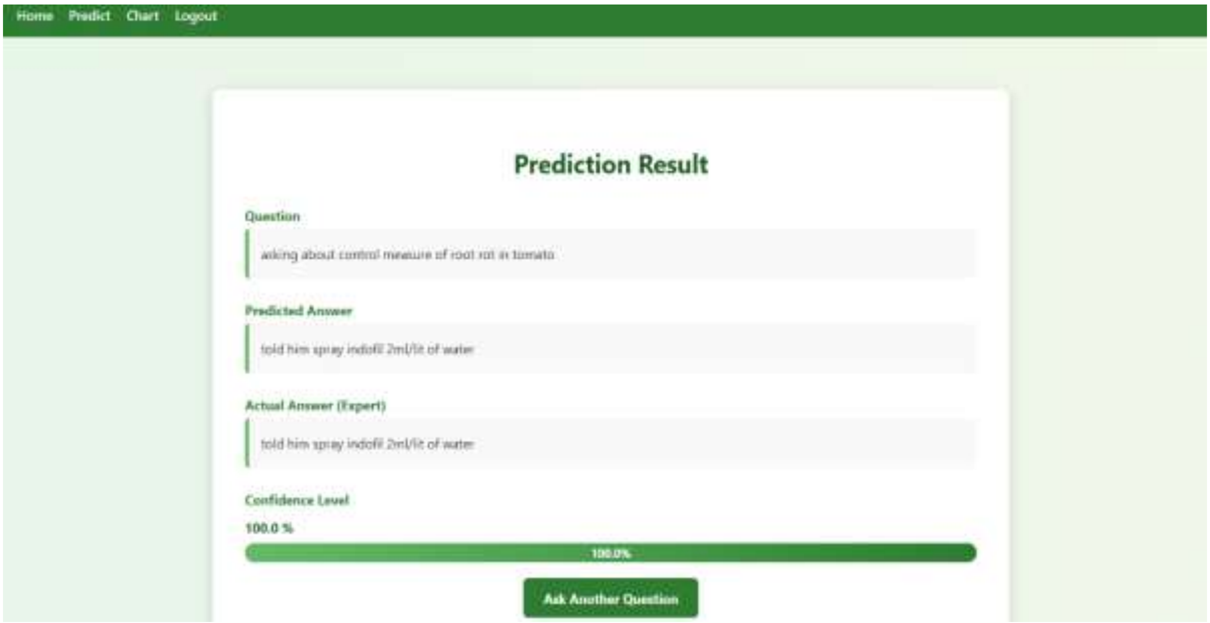

Figure 3 :Prediction Result Page

This screen represents the result module of the Agricultural Advisory System. It displays the outcome of the user's query after processing. The page shows the entered question, the predicted answer generated by the system, and the actual expert-validated answer for comparison. This helps in evaluating the accuracy and reliability of the system. Additionally, a confidence level is provided to indicate the certainty of the prediction. The interface is designed to present information clearly and effectively, allowing users to understand the results easily. An option is also available to ask another question, enabling continuous interaction with the system.

Performance Evaluation Page

Figure 4 : Performance Evaluation Page

This screen represents the performance evaluation module of the Agricultural Advisory System. It displays the overall accuracy metrics of the system, including training accuracy and testing accuracy. The training accuracy indicates how well the model has learned from the training dataset, while the testing accuracy reflects its performance on unseen data. These metrics demonstrate the effectiveness and reliability of the proposed system. The interface presents the results in a clear and concise manner, allowing users to easily understand the system's performance.

VII. CONCLUSION

The proposed agricultural question-answering system successfully demonstrates an efficient and reliable approach to support farmers with accurate advisory information. By adopting a retrieval-based methodology using DeBERTa-v3 embeddings and FAISS similarity search, the system overcomes the limitations of traditional generative models such as hallucination, high computational cost, and inconsistency. The system ensures fast, accurate, and context-aware responses by retrieving expert-validated answers from a structured knowledge base. Its lightweight design and ability to run on CPU-only environments make it suitable for real-world deployment, especially in rural and resource-constrained areas. Overall, the project provides a scalable, cost-effective, and practical solution for enhancing agricultural decision-making, contributing to improved productivity and better access to reliable farming knowledge.

VIII. FUTURE ENHANCEMENT

The proposed retrieval-based agricultural question-answering system can be enhanced by adding multilingual support to handle regional languages, improving accessibility for farmers. It can be extended with region-specific, soil-specific, and crop-specific recommendations for more personalized advisory. A hybrid approach combining retrieval with controlled generative models can provide detailed yet accurate responses. Integrating real-time data such as weather, soil conditions, pest alerts, and market trends will further improve decision-making. Additional features like voice-based interaction, continuous knowledge updates, mobile application deployment, and integration with government platforms can expand usability, making the system a comprehensive and user-friendly agricultural advisory solution.

REFERENCES

[1] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi,C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, and X. Xie, ‘‘A survey on evaluation of large language models,’’ ACM Trans.Intell. Syst. Technol., vol. 15, no. 3, pp. 1–45, Mar. 2024.

[2] J. Howard and S. Ruder, ‘‘Universal language model fine-tuning for textclassification,’’ 2018, arXiv:1801.06146.

[3] Y. Bengio, R. Ducharme, and P. Vincent, ‘‘A neural probabilistic language model,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 13, Mar. 2000,pp. 1–11.

[4] K. Tan, ‘‘Large language models for crop yield prediction,’’ Tech. Rep.,2024.

[5] Y. Luo, X. Cai, J. Qi, D. Guo, and W. Che, ‘‘FPGA—Accelerated CNN for real-time plant disease identification,’’ Comput. Electron. Agricult.,vol. 207, May 2023, Art. no. 107715.

[6] Y. Xu, J. Lu, and J. Zhang, ‘‘Bridging the gap between different vocabularies for LLM ensemble,’’ 2024, arXiv:2404.09492.

[7] T. Wang, X. Xu, C. Wang, Z. Li, and D. Li, ‘‘From smart farming towards unmanned farms: A new mode of agricultural production,’’ Agriculture, vol. 11, no. 2, p. 145, Feb. 2021.

[8] J.Li,M.Xu,L.Xiang,D.Chen,W.Zhuang,X.Yin,andZ.Li,‘‘Foundation models in smart agriculture: Basics, opportunities, and challenges,’’Comput. Electron. Agricult., vol. 222, Jul. 2024, Art. no. 109032.

[9] K. Pan, X. Zhang, and L. Chen, ‘‘Research on the training and applicationmethods of a lightweight agricultural domain-specific large languagemodel supporting Mandarin Chinese and Uyghur,’’ Appl. Sci., vol. 14,no. 13, p. 5764, Jul. 2024.

[10] J. Mahapatra and U. Garain, ‘‘Impact of model size on fine-tuned LLMperformance in data-to-text generation: A state-of-the-art investigation,’’2024, arXiv:2407.14088.