Genome Publications
https://doi.org/10.61096/978-81-990998-5-2_2

Chapter 2

Smart AgroAssist: An Instant Advisory Platform for Supporting Farmers Across Regions

Sheela S Maharajpet
Department of MCA, Acharya Institute of Technology, Bangalore – 560107, India
Vishwapriya R
Department of MCA, Acharya Institute of Technology, Bangalore – 560107, India
Krutika D R
Department of MCA, Acharya Institute of Technology, Bangalore – 560107, India.

ABSTRACT: Recent advancements in information science have significant potential to support sustainable agriculture in India. In this context, we present a framework for a text-based query-response generation system aimed at providing timely assistance to farmers across the country. A key challenge in developing such a system lies in building a robust knowledge base capable of addressing plant protection-related queries from a highly diverse farming population. To address this, we utilize eight years of call-log data from a nationwide agricultural helpline to create a comprehensive knowledge repository. Furthermore, we develop three response-retrieval models that incorporate approximate matching and spatial search capabilities to effectively interpret user queries and retrieve relevant responses. To evaluate the system's effectiveness, we compile a diverse question set comprising 755 queries related to 151 different crops cultivated in India. The performance of the proposed models is assessed using three key metrics: Accuracy Percentage, Crop-weighted Performance Score, and Average Response Retrieval Time. Experimental analysis demonstrates that the proposed framework, AgriResponse, is well-suited for real-world deployment, with each retrieval model offering unique strengths in different application contexts.

Keywords: AgroAssist, query-response system, plant protection, Agricultural, farmer assistance, spatial search, crop-based queries.

Citation: Sheela S Maharajpet, Vishwapriya R, Krutika D R. Smart AgroAssist: An Instant Advisory Platform for Supporting Farmers Across Regions. Machine Learning in Research and Practice: A Multidisciplinary Perspective. Genome Publications. 2025; Pp9-19.
https://doi.org/10.61096/978-81-990998-5-2_2

1. INTRODUCTION

Agriculture remains a vital sector globally, providing the majority of food and raw materials while significantly contributing to national economic growth. However, with the global population rising steadily, agricultural systems are under pressure to meet increasing food demands[1]. This challenge emphasizes the need for advanced technological solutions to help farmers optimize the use of available resources[1]. In this context, the rapid growth of Information and Communication Technology (ICT) infrastructure plays a crucial role in facilitating knowledge sharing and support for the farming community[1]. As mobile phone usage expands among farmers, so does the need for accessible agricultural support through helplines and assistance centers[2].

In India, where nearly 150 million individuals are engaged in farming, the government has consistently taken steps to support this vast population[2]. A notable initiative is the launch of the Kisan Call Center (KCC), a telephonic support system established in 2004[2]. This service, available in regional languages via the toll-free number "1800-180-1551", connects farmers with agricultural advisors to resolve their queries[2]. Over the years, research has shown that KCC has positively impacted farmers' productivity and economic well-being[3],[4]. At KCC, when a farmer places a call with a query, the operator attempts to provide an immediate solution[2]. If the issue is beyond the operator's expertise, the call is escalated to an agricultural expert[2]. However, delays may occur when specialists are unavailable, affecting the timely resolution of farmers' concerns[6]. To address this gap, the present work proposes a text-based query-response generation system that acts as a virtual assistant to replicate the role of KCC operators and agricultural experts[9]. This system is designed to handle plant protection-related queries from farmers across the country efficiently[9].

Agriculture remains the backbone of India's economy, with a vast majority of its population dependent on farming for their livelihood[1]. However, despite rapid advancements in information and communication technologies, a significant information gap still persists between agricultural experts and grassroots-level farmers particularly when it comes to timely solutions for plant protection problems[6]. Farmers often encounter challenges such as pest attacks, crop diseases, and inappropriate chemical usage, which demand immediate attention[9]. Traditional advisory mechanisms like local extension services and helpline centers, though helpful, are often overwhelmed by the volume of inquiries and hindered by language, accessibility, and regional knowledge limitations[6]. To bridge this gap, the integration of intelligent information systems into agriculture has become increasingly vital[9]. In this context, AgriResponse is proposed as a real-time, text-based agricultural query-response system specifically tailored to answer plant-protection-related queries from Indian farmers[9]. Unlike conventional systems that are either rule-based or limited to localized datasets, AgriResponse leverages eight years of call-log records from the Kisan Call Centers (KCCs), creating a rich and diverse knowledge base representative of national agricultural diversity[9].

One of the primary challenges in building such a system lies in handling the multilingual, unstructured, and error-prone queries typically submitted by farmers[9]. To overcome this, AgriResponse employs three robust response-retrieval models, each incorporating approximate string matching, semantic analysis, and spatial search capabilities to effectively interpret and respond to a wide variety of queries[9],[14]. The system is designed to operate not just at a district or state level, but across national agricultural landscapes, ensuring scalability and inclusivity[9]. By offering accurate, timely, and region-specific solutions, AgriResponse has the potential to significantly improve decision-making at the farm level, reduce crop losses, and enhance productivity[9]. The system also serves as a valuable support tool for call center operators and agricultural advisors, allowing for efficient triaging of queries and reducing dependence on real-time human expertise[9]. This paper details the architecture, methodology, evaluation, and practical implications of the AgriResponse system, presenting it as a novel solution for advancing digital agriculture in India[9].

II. LITERATURE REVIEW

T. Wang and Z. Wang, this research an artificial oasis is a desert region that can be cultivated and inhabited by humans under human supervision[1]. Although it offers a way to increase urbanization and population in arid areas, agricultural reclamation in artificial oases has an adverse effect on the environment[1]. GOAL In the mostly arid northwest Chinese province of Xinjiang, agricultural production is centered in piedmont oases[1]. As a result of its agricultural growth, Xinjiang is now China's largest producer of grapes and cotton. This research aims to comprehend. [1]

The Government of India launched the Kisan Call Center (KCC) in 2004 as part of the Kisaan Knowledge Management System (KKMS) with the goal of giving farmers nationwide real-time, region-specific agricultural advice[2]. Farmers can get help with crop management, pest control, weather forecasts, market prices, livestock care, and government programs by speaking with trained agricultural experts, known as Farm Tele-Advisors (FTAs), in their local language over a toll-free number[2]. Every call is recorded in the KKMS database for analysis and service enhancement by the system, which runs through several regional centers[2]. The KCC increases the availability of expert knowledge by utilizing technology and multilingual assistance, enabling farmers to make well-informed decisions and raising overall agricultural productivity. [2]

P. Jaisridhar, introduces sincerely to the feet of my family Goddess Angalamman, Lord Hayagreevar, and Guru Vedhathri Maharishi, whose blessings gave me the strength and courage to complete my thesis successfully. I express my heartfelt gratitude to my guide, whose constant support, encouragement, and belief in my abilities guided me throughout this journey. I also thank the advisory committee, faculty members, and staff of the Dairy Extension Division for their valuable guidance and assistance. Special thanks to my colleagues and friends whose companionship and advice were truly meaningful. Lastly, I appreciate the library resources, which greatly supported my research and thesis writing.[3]

S. Kavitha and A. Nallusamy, The socioeconomic traits of Kisan Call Center (KCC) beneficiaries and non-beneficiaries in Mahbubnagar district, Telangana[4], are investigated in this study. According to information gathered from farmers, the majority of the beneficiaries were young, had higher levels of education, double-cropped, used canal irrigation, and kept in close communication with extension services[4]. Despite having less farming experience, they also demonstrated a positive perception of mobile use in farming[4]. The majority of non-beneficiaries were middle-aged, had lower levels of education, practiced single cropping, had little interaction with extension agents, and used mobile phones less frequently in agriculture[4]. The study recommends creating an interactive mobile app in regional languages and raising awareness of KCC through the media. [4]

K. Chachra et al. The study investigates Kisan Call Centers (KCC) function and influence in resolving issues in the Indian agriculture industry[5]. Since agriculture accounts for a sizable portion of GDP and exports, the establishment of KCCs on January 21, 2004, sought to improve agricultural extension by facilitating direct, reciprocal communication between experts and farmers[5]. KCCs, which have 14 locations throughout India, offer farmers ongoing, specialized phone support in their native tongues[5]. The implementation and efficacy of this initiative in enhancing farming outcomes and knowledge transfer are examined in this chapter.[5]

An overview of the Kisan Call Center (KCC) initiative, which was started to close the communication gap between farmers and agricultural experts, is given by the Department of Agriculture and Farmers Welfare (2020)[6]. Through a toll-free helpline that is available in several local languages, the KCC seeks to provide farmers with timely, dependable, and region-specific advice[6]. This service improves the efficacy of agricultural practices nationwide and facilitates well-informed decision-making.[6]

P. Ajawan, P. Desai, and V. Desai, In order to improve Kisan Call Centers (KCC) responsiveness, particularly in times of crisis like the COVID-19 pandemic[7], the study introduces Smart Sampark, a virtual conversational system. Traditional KCCs depend on human experts to answer farmers' questions, but in times of emergency when experts might not be available, the system has limitations[7]. This gap is filled by Smart Sampark, which uses cosine similarity and Natural Language Processing (NLP) to automatically answer farmer questions based on past KCC data[7]. After 36 months of testing on data from the Belagavi district, the system's accuracy was 86%, and it could be further enhanced with larger datasets.[7]

The Open Government Data (OGD) Platform India, a centralized portal created to make it simple to access datasets made available by different government departments, is offered by the National Informatics Centre (2021)[8]. By making government data publicly accessible to the general public, researchers, and developers, this platform seeks to advance transparency, innovation, and data-driven decision-making[8]. It provides agricultural and rural development datasets that can be used to improve digital solutions and policymaking, making it an essential resource for programs like Kisan Call Centers.[8]

S. Godara, In order to support sustainable agriculture, the study presents AgriResponse, an online query-response system that offers prompt, precise responses to questions from Indian farmers[9]. With an emphasis on plant protection, the system uses eight years' worth of call log records from national farmer helplines to create an extensive knowledge base[9]. It uses spatial search and approximate matching to integrate three response-retrieval models[9]. Accuracy, crop-weighted performance, and response time are used to assess the system, which is validated against a question bank of 755 queries covering 151 crops. The outcomes show how useful and flexible AgriResponse is in a variety of agricultural support situations.[9]

D. G. Bobrow et al. The study presents GUS (Genial Understander System), one of the first experimental dialog systems created to investigate artificial intelligence's ability to understand language. Within a limited domain, GUS facilitates goal-oriented English conversations with users by assuming the role of a travel agent helping to arrange a basic round-trip ticket[10]. GUS achieves realistic interactions without requiring full language complexity or extensive world knowledge by limiting the conversation to a particular context[10]. The paper highlights GUS's role in influencing future research directions in AI and natural language processing, presenting it as a fundamental step in the development of dialog systems.[10]

III. EXISTING SYSTEM

Earlier question-answering (QA) systems built using linguistic approaches typically begin by processing user queries through steps like tokenization, part-of-speech tagging, and syntactic parsing[10]. These systems then interpret the meaning of the question to fetch relevant answers[10]. However, such methods are limited in scope because they rely on structured databases that can only handle questions related to specific, predefined topics[10]. Initial development of these systems began in the 1960s, utilizing natural language processing techniques to convert questions into a standard format for querying database[12]s. In the 1970s, dialogue-based models emerged, which also depended on structured data sources[13].

Over time, systems began to evolve by integrating web-based content with internal knowledge bases[11]. This hybrid approach allowed for broader question coverage, including both commonly asked and more random queries[11]. Subsequent models utilized online text data and developed heuristic techniques to store relevant information locally for future use. In the early 2010s, systems incorporating semantic information started gaining attention[14]. These used semantic roles, lexical databases like WordNet, and detailed linguistic analysis to improve accuracy[14]. For instance, some models extracted semantic relationships from domain-specific literature, while others pulled information from large-scale online platforms like Wikipedia[14]. Techniques such as named entity recognition, syntactic analysis, and ontology-based keyword extraction were commonly applied in these systems[14]. Parallel to linguistic models, statistical methods also gained popularity. These approaches relied on analysing various features of the questions to estimate likely answers[11],[15]. By training models on annotated datasets, these systems could categorize and retrieve answers based on similarity measures[15]. Algorithms such as support vector machines, Naive Bayes, and k-nearest neighbours were frequently used. A notable example from the early 2000s used a maximum entropy model and word frequency features for classification[15]. Other systems developed during this time adopted similar classification strategies, including chunking questions into phrases using statistical models and employing n-gram statistics to rank potential answers[15].

Machine learning also played a key role in enhancing QA systems. For example, ranking models were trained on feature sets to answer "why" questions, while Bayesian techniques were applied to classify questions using ontology-based structures[15]. Hierarchical classifiers were used to extract accurate answers from community QA platforms[15]. Later models combined traditional information retrieval methods with newer techniques like TF-IDF scoring and tree distance algorithms[14]. Some integrated external knowledge sources with multi-level tag classification systems to improve semantic relevance. Neural network models also began to emerge, especially in specialized areas like medical question answering, where hybrid methods were applied. More recently, with the rise of deep learning, large-scale language models have been adapted for QA systems[14]. These models were fine-tuned on specific domains, such as health-related datasets during the COVID-19 pandemic, to improve relevance and accuracy. Additional enhancements included mechanisms for evaluating question similarity, simulating human-like reasoning to determine if a given question could be adequately answered[14].

Limitations of Existing System

Despite these advancements, existing systems face several limitations. A significant issue is the imbalance in data distribution: many crop types in the knowledge base have very few entries, while a small number of popular crops dominate the dataset. This skew leads to inaccuracies, as incorrect responses are often associated with less-represented crops[9]. Furthermore, response retrieval times (RRT) also reflect this imbalance. Queries related to underrepresented crops are usually resolved quickly, whereas questions about more common crops often take longer to process, highlighting inefficiencies in the current systems[9].

IV. PROPOSED SYSTEM

This study presents the development of AgriResponse, an intelligent query-response system designed to assist farmers by providing answers to plant protection-related queries in text format. The system is not only intended for direct use by farmers but can also support helpline operators by offering additional insights or alternatives, especially in the absence of subject-matter experts. One of the main challenges in designing such a solution lies in building a comprehensive knowledge base capable of addressing a wide range of agricultural queries from across the country. Another key challenge is constructing a response-retrieval mechanism that can handle spelling variations, language inconsistencies, and optimize search speed without compromising accuracy9]. To address the first issue, historical query-call records from the Kisan Call Center (KCC) data servers and the Open Government Data (OGD) Platform of India have been used[2],[8]. These datasets, spanning the past eight years, are publicly available and include rich details such as farmer location, the exact questions asked, and the responses provided by the helpline experts[8],[9].

Advantages of Proposed System

Innovative Query Response Framework: The AgriResponse system offers a fresh approach to answering plant protection-related questions posed by farmers, addressing a long-standing need for scalable and intelligent agricultural support[9].

Spelling Error Resilience: The model includes error-tolerant features that recognize and process misspelled words in both queries and stored responses, increasing its reliability in real-world use cases[9].

Multiple Answer Retrieval: In cases where multiple valid solutions exist often due to regional agricultural practices the system is capable of presenting several appropriate responses to a single query.

Scalability Across Regions: Unlike previous models which are often limited to local or district-level applications, AgriResponse is designed for nationwide deployment, ensuring broader impact and usability[9].

Open-Access Resources: To support transparency and encourage collaborative innovation, the knowledge base, model code, and question datasets developed through this research are made publicly available[9].

Multilingual Consideration: The system takes into account the diverse linguistic nature of Indian farmers and the KCC records, accommodating mixed-language inputs and addressing translation inconsistencies for improved performance[9].

Performance Evaluation Contribution: AgriResponse also addresses a notable gap in existing literature by introducing reliable metrics for assessing QA model performance, contributing to standardization and benchmarking in agricultural AI systems[9],[11],[14].

System Architecture

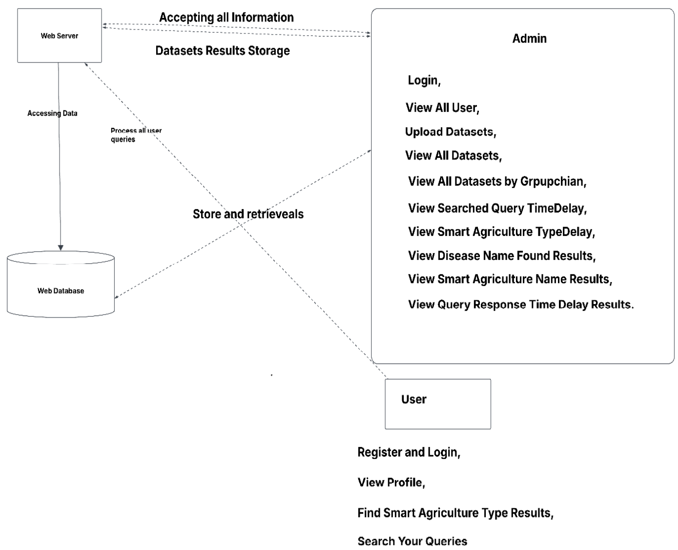

Fig 1. System Architecture

V. METHODOLOGY

The development of the AgriResponse system follows a structured and systematic methodology to ensure the accurate, scalable, and efficient retrieval of agricultural information in response to plant-protection-related queries. The methodology is composed of five major phases: data acquisition, preprocessing, knowledge base construction, response retrieval modelling, and performance evaluation[9].

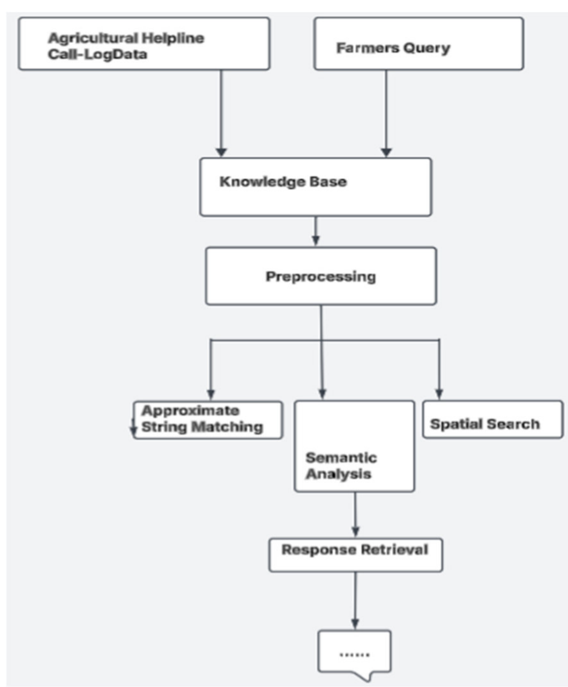

Fig 2: Methodology flow

Data Acquisition

The AgriResponse system is built on a foundation of large-scale, real-world data collected from the Kisan Call Centres (KCCs) operating across India[2],[9]. Over eight years of historical query logs were compiled, which include detailed information such as the farmer's question, the name of the crop, the expert's response, the location of the caller (district and state), and the timestamp[9]. To enhance the diversity and depth of the dataset, additional agricultural data was integrated from the Open Government Data (OGD) Platform India[8]. These datasets offer structured and validated agricultural information, which strengthens the coverage and credibility of the resulting knowledge base[8],[9].

Data Preprocessing

Given the unstructured and multilingual nature of the collected data, a robust preprocessing pipeline was necessary to normalize and standardize the inputs[9]. The first step involved language normalization, where queries expressed in various Indian languages were transliterated and translated into a consistent textual format[9]. This was followed by cleaning operations that removed special characters, noise, and irrelevant metadata from the text. Spelling corrections were applied using edit-distance-based algorithms, such as Levenstein distance, to accommodate common typographical errors[9],[14]. Semantic tagging was then performed to enrich the data with labels like crop name, disease category, and region, enabling more precise and context-aware search functionality in subsequent stages[9],[14].

Knowledge Base Construction

Once pre-processed, the refined dataset was used to construct a structured and indexed knowledge base[9]. Each record in the knowledge base contains the standardized user query, corresponding expert response, crop-related tags, geographic metadata, and semantic annotations[9]. The indexing mechanisms support fast and flexible lookups based on multiple parameters such as crop name, query keywords, and user location. The knowledge base is designed to be scalable and adaptable, allowing for updates as new data becomes available and supporting robust query matching under varying linguistic and semantic conditions[9].

Response Retrieval Models

To handle diverse user inputs, the system implements a tri-model response retrieval strategy[9]. The first model is the Exact Matching Model, which performs direct keyword-based lookups in the knowledge base. It is most effective for well-formed queries that match existing entries closely. The second model, the Approximate String Matching Model, addresses the issue of noisy or loosely structured queries by using text similarity measures such as n-gram overlap and cosine similarity, thereby tolerating spelling errors and word-order variations[9],[14]. The third and most advanced model is the Spatial-Semantic Matching Model, which integrates semantic analysis techniques like TF-IDF or Word2Vec with geographic filtering based on the user's district or region[9],[14],[15]. This model ensures contextually appropriate answers, even for ambiguous or partially correct inputs, making the system more adaptive and user-friendly[9],[14].

Query Processing Workflow

The operational pipeline begins when a user submits a free-text agricultural query via the system interface[9]. The input is first normalized and pre-processed using the methods described earlier. Based on the nature of the query, the system then selects the most appropriate response retrieval model[9]. The selected model retrieves the top-k relevant results from the knowledge base, ranking them by relevance and accuracy[9]. The final answers are presented to the user, potentially with multiple valid responses if applicable, thereby accommodating regional or methodological variations in plant protection[9].

Evaluation Metrics

The effectiveness and efficiency of the AgriResponse system are measured using three key performance metrics[11],[14]. Accuracy Percentage quantifies the proportion of correct answers retrieved compared to expert validations[11]. The Crop-Weighted Performance Score (CWPS) is designed to account for the imbalance in query distribution across different crops, ensuring that performance is not biased towards more commonly discussed crops[11]. Lastly, Average Response Retrieval Time (RRT) measures the time taken to generate a response, reflecting the system's real-time usability[11]. Experiments conducted on a curated testbed of 755 queries across 151 crops demonstrate that AgriResponse, particularly when using all three retrieval models, performs with high accuracy and low latency, validating the robustness of the methodology[9].

VI. RESULTS

The AgriResponse system was evaluated using a dataset of 755 agricultural queries covering 151 crop types[9]. Three retrieval models Exact Matching, Approximate String Matching, and Spatial-Semantic Matching were tested based on Accuracy, Crop-Weighted Performance Score (CWPS), and Average Response Retrieval Time (RRT)[9],[11].

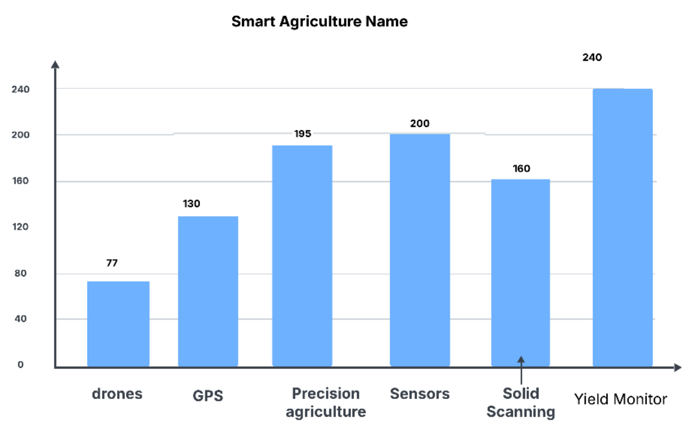

Fig 3: View Graph of all agriculture name

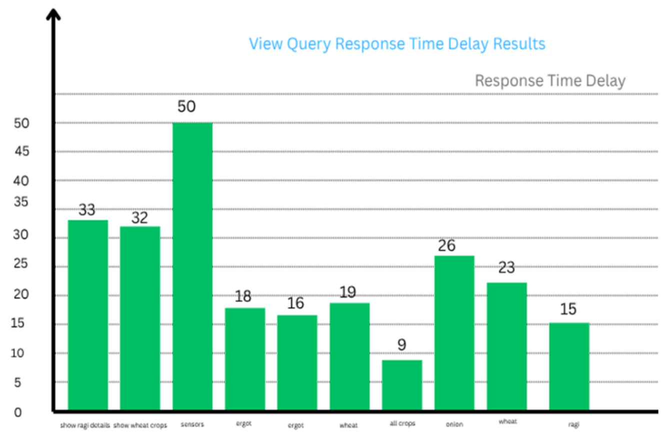

Fig 4: View Query Response Time Delay Results

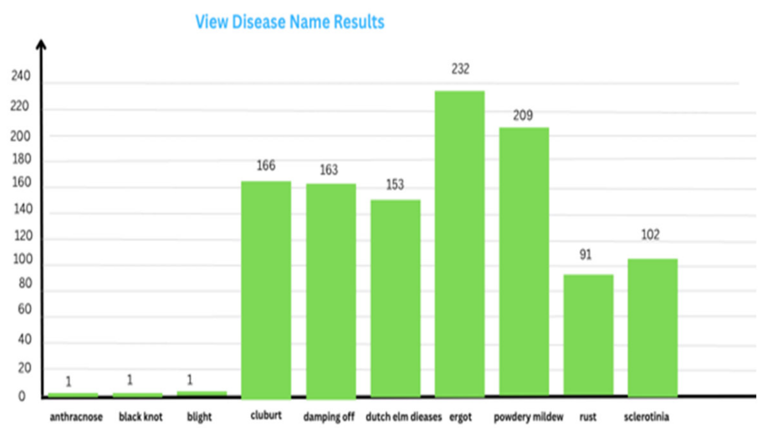

Fig 5: View Disease Name Results

The Exact Matching Model showed limited accuracy, particularly with informal or misspelled queries[9]. The Approximate Matching Model improved performance by handling textual variations[9]. The Spatial-Semantic Model achieved the highest accuracy and CWPS by incorporating both linguistic similarity and regional relevance[9],[14].

In terms of efficiency, all models maintained an RRT under 1.5 seconds, with the Exact Matching Model being the fastest[9]. Overall, the Spatial-Semantic Model provided the best balance between accuracy and response time, proving suitable for real-world agricultural advisory support[9].

VII. CONCLUSION

The growing reliance on ICT in agriculture highlights the urgent need for intelligent support systems to address farmers' queries promptly. Expert unavailability and delays in advice from helplines can severely affect both farmers' livelihoods and the economy. To address this, we introduced AgriResponse, a real-time, text-based framework aimed at delivering plant protection solutions to Indian farmers. It also serves as a support tool for helpline operators seeking second opinions.

Key challenges included building a diverse knowledge base covering various crops and designing efficient query-response models. To overcome these, historical call-log data from helpline centers were used to construct the knowledge base, and three response retrieval models RRM1, RRM2, and RRM3 were developed and tested on 755 queries across 151 crops using metrics like Accuracy Percentage (AP), Crop-Weighted Performance Score (CWPS), and Average Response Retrieval Time (ARRT). Among the models, RRM3 offered the best overall balance, making it the most suitable for practical use. Future improvements will focus on integrating external knowledge sources and advanced string-matching with machine learning to enhance system accuracy and responsiveness.

REFERENCES

[1] T. Wang, Z. Wang, L. Guo, J. Zhang, W. Li, H. He, R. Zong, D. Wang, Z. Jia, and Y. Wen, "Experiences and challenges of agricultural development in an artificial oasis: A review," Agricult. Syst., vol. 193, 2021, Art. no. 103220.

[2] Kisaan Knowledge Management System Government of India. (2021). Kisan Call Center. [Online].

[3] P. Jaisridhar, "Impact of Kisan call center on technological adoption among dairy farmers of Tamilnadu," Ph.D. thesis, NDRI, Karnal, Karnal, India, 2013.

[4] S. Kavitha and A. Nallusamy, "A study on socio economics characteristics of Kisan call centre beneficiaries and non-beneficiaries in Mahbubnagar district of telangana," J. Pharmacognosy Phytochem., vol. 8, no. 3, pp. 4660–4663, 2019.

[5] K. Chachra, G. Seelam, H. Singh, M. Sarkar, A. Jain, and A. Jain, "The impact of Kisan call centers on the farming sector," in Environmental and Agricultural Informatics: Concepts, Methodologies, Tools, and Applications. Hershey, PA, USA : IGI Global, 2020, pp. 66–78.

[6] Department of Agriculture and Farmers Welface. (2020). Kisan Call Centre. [Online].

[7] P. Ajawan, P. Desai, and V. Desai, "Smart Sampark–An approach towards building a responsive system for kisan call center," in Proc. IEEE Bengaluru Humanitarian Technol. Conf. (B-HTC), Oct. 2020, pp. 1–5.

[8] National Informatics Centre. (2021). Open Government Data Platform India. [Online].

[9] S. Godara. (2021). Agriresponse Online Repository. [Online]. Available: https://github.com/Samarth-Godara/AgriResponse

[10] D. G. Bobrow, R. M. Kaplan, M. Kay, D. A. Norman, H. Thompson, and T. Winograd, "GUS, a frame-driven dialog system," Artif. Intell., vol. 8, no. 2, pp. 155–173, 1977.

[11] S. K. Dwivedi and V. Singh, "Research and reviews in question answering system," Proc. Technol., vol. 10, pp. 417–424, Jan. 2013.

[12] B. F. Green, A. K. Wolf, C. Chomsky, and K. Laughery, "Baseball: An automatic question-answerer," in Proc. Papers Presented Western Joint IRE-AIEE-ACM Comput. Conf. IRE-AIEE-ACM (Western), 1961, pp. 219–224.

[13] W. A. Woods, "Progress in natural language understanding: An application to lunar geology," in Proc. June 4–8, Nat. Comput. Conf. Expo. (AFIPS), 1973, pp. 441–450.

[14] S. Godara. (Oct. 2021). Agriresponse: A Query-Response Generation Model for Indian Farmers. [Online]. Available: https://www.codeocean.com

[15] P. Clark, J. Thompson, and B. Porter, A knowledge-based approach to question-answering, in Proc. AAAI, vol. 99, 1999, pp. 43–51.