



A Small Language Model Framework for Intelligent Agricultural Knowledge Retrieval and Context-Aware Decision Support

Mohammed Danish Kolhar¹, Abdul Hannan R. Dalal², Yuvraj Lahoti^{3*}

¹ Department of Computer Engineering, Vishwakarma University, Pune, Maharashtra, India

² Department of Artificial Intelligence, Vishwakarma University, Pune, Maharashtra, India

³ Department of Commerce & Management, Vishwakarma University, Pune, Maharashtra, India

*Corresponding Author: yuvraj.lahoti@vupune.ac.in ,
dalalabdulhannan@gmail.com

Abstract.

Artificial intelligence is transforming modern agriculture by enabling intelligent advisory services, efficient resource utilization, and data-driven decision-making. However, many language-model-based agricultural systems rely on computationally intensive architectures that are difficult to deploy in resource-constrained farming environments. This paper proposes a Small Language Model (SLM) framework for intelligent agricultural knowledge retrieval and context-aware decision support. The proposed framework integrates natural language query understanding, semantic knowledge retrieval, contextual relevance analysis, and recommendation generation within a lightweight architecture designed for efficient deployment. The agricultural knowledge retrieval layer utilizes domain-specific repositories containing crop, disease, irrigation, fertilizer, and weather information, while a context-awareness mechanism incorporates crop type, environmental conditions, seasonal factors, and symptom descriptions to improve recommendation relevance. An evaluation framework based on retrieval performance, computational efficiency, and deployment-oriented metrics is also presented to facilitate future implementation and assessment. The proposed architecture emphasizes resource efficiency, edge deployment capability, and scalability, making it suitable for agricultural environments with limited computational infrastructure and network connectivity. The framework provides a structured foundation for developing intelligent agricultural advisory systems that combine efficient knowledge retrieval with context-aware decision support for modern smart farming applications.

Keywords: Small Language Models (SLMs), Agricultural Knowledge Retrieval, Decision Support Systems, Smart Farming, Precision Agriculture, Context-Aware Computing

1 Introduction

One of the most important sectors in the world today is agriculture. It generates the food consumed by man, stabilizes economies, and enhances rural well-being. However, challenges are arising daily for farmers and other stakeholders in the agricultural sector. Increasing food demands, changing climatic patterns, and resource constraints necessitate decision making in crops, management of diseases, water, nutrients, resources so that farming can maintain its productivity and sustainability. With rapid growth of digital technology, there are more knowledge, data, and tools readily available to farmers in support of their decisions and transformations of the ways of farming, providing timely and correct information to the right people in relation to specific and real-world decisions in varied agricultural circumstances.

1.1 Background and Motivation

The field of agriculture is rapidly evolving with the integration of digital tools, data-driven approaches, and smart farming practices. Frameworks such as Agriculture 4.0 and smart farming have been proposed to address complex issues such as food security, optimal resource utilization, environmental sustainability and production demands. These developments bring sensing, data analysis, automation, and intelligent decision-making to farming practices by enabling farmers to monitor crops and manage their resources more efficiently to improve yield [14-17]. The growing trend towards precision agriculture and site-specific management practices which optimize crop performance, decrease input costs, and lessen the environmental footprint have fueled the transformation of agriculture [15,16,24]. The emergence and importance of digital advisory platforms are growing, as they enable real-time information and decision support for farmers [19,21,25]. Furthermore, the advancements in Artificial Intelligence have enhanced the capabilities of these systems in interpreting and leveraging agricultural data and responding effectively to farmers [19,21,25]. Natural Language Processing and language-model based techniques are presenting new opportunities to provide advice and knowledge through conversational and intuitive interfaces so as to enable more accessible acquisition of domain-specific information [1-6]. Overall, all these trends emphasize the growing need for intelligent systems that can collect agricultural data and provide decision support to farmers in agriculture.

1.2 AI-Based Agricultural Decision Support

With increasing farmer access to digital technology and agricultural data, smart decision support systems are increasingly developing. These are designed to integrate agronomic information into timely, useful decision guidance and decision recommendations at every level of crop production. Applications for crop recommendation, diagnosis of disease, planned irrigation and fertilizer management exist, among others, to improve agricultural productivity, reduce waste, and contribute to a sustainable agricultural system. IT, data analysis, and data-driven processes allow for a better understanding of fields and more efficient and immediate agricultural management [17,19,21,24].

Recently, decision support systems have become more sophisticated in collecting real-time data, conducting predictive analytics and providing automated decisions. Smart Irrigation Solutions help with wise water use, disease identification systems detect crop illness earlier than humans might notice, and fertilizer advisories deliver targeted nutrient management. These modern information systems give site-specific suggestions helping to inform and enable better agriculture at the right time [19,21,22,25,26].

1.3 Language Models in Agriculture

Language models provide new ways for the access to and association of the agricultural information and its users. Many researchers applied LLMs for the dissemination, advice, decision support and generation of agricultural specialized knowledge [1–6]. Contrary to structured conventional information systems, LLM can process natural language queries and offer context-adapted answers that provide an experience much more natural to interact with agricultural systems. The ability of language models to answer a wide range of agricultural queries and extract domain-specific knowledge efficiently has recently attracted considerable attention for enhancing agricultural extension services, farm management, and knowledge-sharing systems [1–6]. Additionally, the combination of multimodality to the language models helps to make better decision making by incorporating vision and textual knowledge. This has been seen for the plant disease identification, crop monitoring, advisory services and intelligent farming assistants [2, 3, 8, 11]. The ag chatbot or the advice language models are also promising for the use in developing countries with less technical users. However, the above models need much computer power which is a constraint for resource poor farms [1, 4, 5, 12].

1.4 Small Language Models

However, the advent of Small Language Models (SLMs) has led to an alternative to the large-scale approach, which requires much less computational effort to achieve language understanding and generation capabilities. The objective of this is to make a model that has less parameters and optimized architectures that can be deployed in environments that are constrained in their processing power, memory and connectivity. An interest in SLMs has been recently documented in a number of new research efforts for their capability to perform effectively for specialized applications where they do not require as much infrastructure as larger models [1,7]. One difference between SLMs and LLMs is related to their deployment and resource usage. If high performance computing infrastructure and cloud services is required, then LLM can be deployed on these, but if it is not required then SLM can be used on edge devices and lower latency local systems with less costs and requirements [1]. This is especially significant for agricultural settings where access to the internet and high-end computers might not always be dependable. Moreover, SLMs offer the potential to be efficient in resource use, deployed at the edge, and offer local inference, making them a promising basis for the development of easy-to-use, scalable knowledge retrieval and decision support solutions in agriculture, information and communication, and healthcare [1,7].

1.5 Research Gaps

Existing research has extensively explored the application of intelligent technologies in agriculture, including precision farming, crop monitoring, disease diagnosis, irrigation management, and digital advisory systems [14-25]. More recently, language-model-based approaches have demonstrated considerable potential for agricultural information access, knowledge dissemination, and decision support through natural language interactions [1-6]. These trends signal an increased interest in the adoption of advanced computational technologies for enhancing agricultural production and gaining access to domain-specific information. However, while there are several emerging studies focusing on large-scale language models and computationally demanding applications of agricultural intelligence, very few have looked at lightweight solutions that will work well in resource-constrained settings. Current research also lacks a comprehensive framework to link the retrieval of agricultural knowledge to context-aware decision support through the use of Small Language Models. There has been little exploration into creating an architecture that integrates crop-specific information, environmental conditions and seasonal variables, in addition to a user query, into a lightweight and deployable model. Therefore, there is a need for a framework that leverages the efficiency of Small Language Models to provide intelligent agricultural knowledge retrieval and context-aware recommendations while maintaining suitability for low-resource and edge-computing environments. Table 1 summarizes the characteristics, strengths, and limitations of representative AI-based, LLM-based, and SLM-based technologies reported in the agricultural domain, highlighting the need for lightweight and context-aware agricultural intelligence frameworks [1-7,15-17,19,21,24,25].

Table 1. Comparative Analysis of Existing AI, LLM, and SLM Technologies in Agriculture

Ref.	Technology Category	Primary Application Domain	Key Strengths	Identified Limitations
[15]	Deep Learning	Crop monitoring, yield prediction, disease detection	High predictive capability for complex agricultural data	Requires large datasets and computational resources
[16]	Machine Learning	Precision agriculture and farm analytics	Effective for classification and prediction tasks	Limited contextual understanding and adaptability

Ref.	Technology Category	Primary Application Domain	Key Strengths	Identified Limitations
[17]	Big Data Analytics	Smart farming and data-driven agriculture	Supports large-scale agricultural data management	Dependence on data quality and infrastructure
[19]	AI-Based Systems	Irrigation, pesticide, and herbicide optimization	Improves resource utilization and operational efficiency	Limited integration of contextual knowledge
[21]	Decision Support Systems	Agriculture 4.0 advisory and planning	Supports informed agricultural decision-making	Often relies on pre-defined rules and structured inputs
[22]	IoT-Based Smart Agriculture	Real-time field monitoring and management	Enables continuous data collection and monitoring	Requires reliable connectivity and sensor infrastructure
[24]	Smart Farming Platforms	Crop data management and farm operations	Facilitates digital transformation of agriculture	Challenges in handling diverse agricultural contexts
[25]	Machine Learning Frameworks	Precision agriculture applications	Supports automated analysis and recommendations	Limited natural language interaction capabilities
[1]	Small Language Models	Agricultural advisory and edge-based applications	Lightweight deployment and lower computational requirements	Limited domain-specific agricultural implementations
[2]	Multimodal Language Models	Agricultural knowledge analysis and interpretation	Ability to process multiple information modalities	High computational complexity

Ref.	Technology Category	Primary Application Domain	Key Strengths	Identified Limitations
[3]	Large Language Models	Agricultural knowledge generation and advisory support	Advanced language understanding and information retrieval	Resource-intensive deployment requirements
[4]	Large Language Models	Agricultural information access and decision support	Natural language interaction and knowledge assistance	Risk of unreliable or contextually inappropriate outputs
[5]	Large Language Models	Agricultural workforce and advisory systems	Enhanced accessibility to agricultural information	Dependence on significant computational resources
[7]	Small Language Models	Domain-specific language processing	Efficient inference and reduced infrastructure requirements	Limited exploration in agricultural decision support frameworks

There is already a significant body of research on agricultural intelligence systems, but few studies have focused on integrating SLM into agricultural knowledge retrieval and context-aware decision support in a lightweight and deployable system. While numerous studies have been conducted on agricultural intelligence systems, there is limited research that deals with integrating SLM into agricultural knowledge retrieval and context-aware decision support, either light and deployable. This gap is a driver to create the proposed framework. The evolution of agricultural technologies is shown in Figure 1, from the rule-based systems to more intelligent systems based on language models that can support more advanced agricultural decision-making.

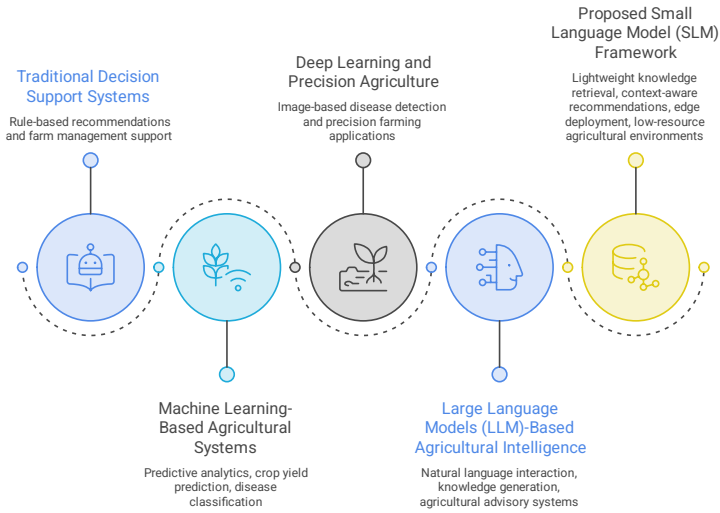


Figure 1. Evolution of agricultural intelligence systems from traditional decision support systems to the proposed Small Language Model (SLM)-based framework for intelligent agricultural knowledge retrieval and context-aware decision support.

2 Proposed Framework

2.1 Framework Overview

The proposed framework is designed to enable the easy retrieval and support of knowledge to assist selection in agriculture by introducing a lightweight Small Language Model (SLM) architecture. The framework aims to help the users gain agricultural information relevant to them by extracting the information from natural language queries and integrate it with appropriate knowledge resources. The framework aims to facilitate agricultural decision-making within a structured and efficient framework through the combination of language understanding, knowledge retrieval, and contextual analysis. The design focuses on computational efficiency, scalability, and applicability, catering to scenarios where high-end computing infrastructure is not easily accessible. There are seven major modules in the system: User query Interface, Text Preprocessing Layer, SLM Understanding Engine, Agricultural Knowledge Retrieval layer, Context awareness engine, Decision support generator and Advisory output layer. It starts whenever a user types in an ag related query related to growing crops, symptoms of a disease, how to water plants, how wet soil or nutrients affect plants. The query submitted is passed through the Text Preprocessing Layer, with the objective of eliminating irrelevant symbols and normalizing text to have a uniform representation for input strings prior to the semantic analysis. The input query goes through the Text Preprocessing Layer with the aim of discarding irrelevant symbols, normalizing the text

and returning the representation of the input strings for the consideration of semantic syntax. The processed query is then passed on to the SLM Understanding Engine that tries to "understand" the intent of the query and extracts relevant agricultural concepts to be further processed. When the understanding of a query is established, the Agricultural Knowledge Retrieval Layer is used for searching in domain specific repositories that hold crop information, descriptions of disease, fertilizer recommendations, guidelines for irrigation and weather-related knowledge. The retrieved information is then transformed into a data-ready format and analysed by the Context Awareness Engine, which, using these pieces of information like the presence of a crop, environmental condition, seasonal characteristics, soil condition and descriptions of symptoms, is able to progressively increase the relevance of the retrieved knowledge. Based on this analysis the Decision Support Generator tells which recommendations and responses to consider according to the study area. The resulting products are presented to users in a context-specific manner in the Advisory Output Layer providing forward-looking advice and recommendations to take action on agricultural issues. The proposed SLM-based agricultural advisory framework overall process is represented in Figure 2 showing what modules interact each other for generating the knowledge and decision-making support in context. The proposed framework assumes the deployment of a lightweight transformer-based Small Language Model containing approximately 1–3 billion parameters, making it suitable for execution on edge devices and resource-constrained computing environments. The SLM accepts natural language agricultural queries as input and generates semantic representations that are subsequently used for agricultural knowledge retrieval. Rather than relying on computationally intensive fine-tuning, the framework is designed to support retrieval-augmented prompting using domain-specific agricultural knowledge repositories. This design enables efficient query understanding while reducing computational overhead and maintaining flexibility for future implementation using lightweight open-source SLM architectures.

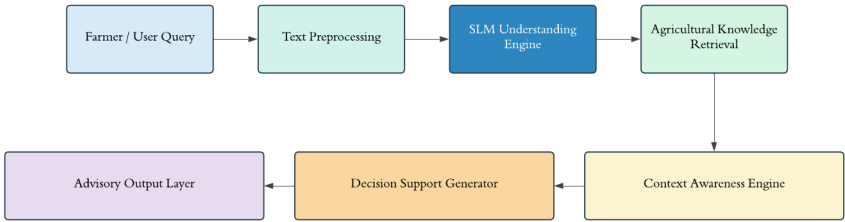


Figure 2. Proposed SLM Framework Architecture

2.2 Agricultural Knowledge Retrieval Layer

The Agricultural Knowledge Retrieval Layer is the major information access layer of the proposed framework. This layer will be responsible to extract relevant agricultural knowledge to match user query before its contextual interpretation and recommendation creation. Once query has been analysed by SLM Understanding Engine and semantic elements of the query has been extracted the entities and meaning associated

to that agricultural concept are passed to the agricultural knowledge retrieval layer for information retrieval. The system's performance would strongly rely upon the retrieval performance in this module of getting relevant knowledge from domain knowledge. The retrieval layer will access a number of Agricultural knowledge repositories which comprises crop cultivating details, details regarding disease, fertilizer details, water managing, weather related advisory etc which act as the knowledgebase for this framework and used for developing recommendations based upon agricultural knowledge. Once user queries, information retrieval system would look into these repositories and retrieve knowledge entries showing higher relevance toward the queries. Before retrieval, agricultural documents are transformed into semantic vector representations and indexed using an embedding-based indexing mechanism. This enables efficient semantic search by organizing agricultural knowledge according to vector similarity rather than exact keyword matching, thereby improving the retrieval of contextually relevant information. Retrieved knowledge is subsequently validated through relevance ranking and consistency with the corresponding agricultural knowledge repository before being forwarded to the Context Awareness Engine for further processing. The measure used to identify the level of relevancy between a user's query and the stored agricultural knowledge entries is Cosine Similarity, originally defined in the Vector Space Model by Salton et al. [35]. Cosine Similarity is an information retrieval measure widely used as it calculates the angular likeness between text represented as vectors without taking in account the documents' length. Thanks to its computationally effective nature and ability to performs well on semantic similarity detection, it is also used in our system in order to identify the relevant agricultural knowledge records to the user's query. The relevance is then calculated through Equation (1). During retrieval, the semantic embedding generated from the user query is compared with the embeddings of indexed agricultural knowledge entries. Cosine Similarity is computed between these vector representations, and knowledge entries with the highest similarity scores are selected as candidate responses for subsequent contextual evaluation.

$$Sim(Q, K) = \frac{Q \cdot K}{|Q| \times |K|} \quad (1)$$

where Q represents the user query vector, K denotes the agricultural knowledge vector, and $Sim(Q, K)$ indicates the similarity score. Higher similarity values correspond to greater semantic relevance and are prioritized for subsequent contextual evaluation. The framework utilizes the Cosine Similarity measure proposed by Salton et al. [35] to identify relevant knowledge entries. Retrieved knowledge is subsequently verified through domain consistency and relevance validation. Candidate entries are checked for consistency with the identified crop, agricultural context, and associated knowledge repository before being forwarded to the Context Awareness Engine. This verification step reduces the likelihood of irrelevant or contradictory recommendations and improves the reliability of the generated advisory responses.

2.3 Context Awareness Engine

Though the Agricultural Knowledge Retrieval Layer has been responsible to retrieve appropriate information from various specialized databases, the retrieved knowledge

may not be equally applicable to all agricultural contexts. Decisions in agriculture are usually context-dependent; for example, certain recommended advice to manage crop yield is depending on the characteristics of the crop grown, prevailing environment conditions, the season etc., and the symptoms observed. So, the contextual Awareness Engine is introduced in the framework to check applicability of the retrieved knowledge before recommendation is generated. The Contextual Awareness Engine serves to ensure the relevance of advisory feedback to the contextualized scenarios. It takes in several contextual information pieces that include crop type, soil type, weather, season, reported symptoms. These parameters help differentiate between seemingly identical agricultural contexts which require entirely different recommendations, e.g., requirements of irrigation may differ based on the crop type and the season, while advice about crop disease management is dependent on the symptoms reported and the current environment conditions. A Context Relevance Score is introduced in the framework, in order to rank retrieved knowledge with respect to the relevance with the context. This score is a combination of three main context information's, namely Crop Relevance, Symptom Relevance and Weather Relevance. It is formulated to describe the input of each context into the ranking of recommendation as described by Equation (2).

$$Score = \alpha C + \beta S + \gamma W \quad (2)$$

where C denotes crop relevance, S represents symptom relevance, and W indicates weather relevance. The coefficients α , β , and γ are weighting factors assigned to each component based on its significance. Higher context relevance scores indicate a stronger match to the user's conditions and are therefore prioritized during recommendation ranking. This framework-specific scoring formulation systematically incorporates agricultural context to improve the relevance of advisory recommendations. In the proposed framework, crop relevance is determined by evaluating the correspondence between the identified crop entity in the user query and the retrieved agricultural knowledge. Symptom relevance is estimated by matching reported symptoms with disease descriptions contained in the knowledge repository, whereas weather relevance is derived from the consistency between environmental conditions and crop-specific cultivation requirements. The weighting coefficients α , β , and γ are normalized such that $\alpha + \beta + \gamma = 1$ and may be assigned using expert knowledge or optimized during future experimental implementation based on application-specific requirements.

2.4 Decision Support Generation

The Decision Support Generator is the last analytical part of the proposed framework and its responsibly transforming retrieved knowledge with contextual insights into actionable agricultural recommendations. The irrelevant information is identified by the Agricultural Knowledge Retrieval Layer and the contextual appropriateness of the identified information is assessed by the Context Awareness Engine, before they are combined by the Decision Support Generator, which generates advisory responses according to the user's query. The goal of this part is to ensure that the final recommendations are relevant both to the retrieved knowledge, and specific to cases with agricultural

contexts corresponding to the user's request. The knowledge retrieved from the retrieval phase and then contextualized in the contextual evaluation phase is used for the creation of the recommendation. The framework values relevance of the input information to the domain and semantic relevance to the context of agriculture, thus giving priority to pertinent information to the user's agricultural environment. The method allows the system to minimize the chances of serving general recommendations and offer guidance based upon attributes of the crop, environmental factors, seasonal factors, and field symptoms observed. The resulting recommendations are then presented in an easily interpreted and easily implemented way. Based on the type of query submitted to the Decision Support Generator, the output of advisory may vary. This includes crop management advice, disease identification suggestions, irrigation advice, fertilizer options and general agriculture advice. This generated response is presented to the end user via the Advisory Output Layer which sits between the framework and the user. The proposed framework seeks to enhance access to the relevant agricultural information in various farming situations and to facilitate informed agricultural decision making by embedding the knowledge retrieval process within the analysis of context and generation of recommendations. The generated recommendations may further be prioritized according to the calculated context relevance score, enabling the framework to present the most suitable advisory response before alternative recommendations. This ranking mechanism supports more consistent and context-sensitive agricultural decision support.

3 Methodology

3.1 Query Processing Workflow

The proposed framework employs a sequential methodology to change the user's agricultural question to a context-aware suggestion by using techniques such as language comprehension, knowledge retrieval, contextual analysis, and decision generation. The main features of the workflow are relevant agricultural information being identified, evaluated and presented in a structured manner. Its design incorporates several processing levels to support more relevant and applicable advisory response while keeping computational efficiency. The whole cycle is initiated by sending an agriculture-related query by the User Query Interface. The query might involve questions such as growing a specific crop, the disease symptoms, what they should irrigate, how would they apply fertilizers, how their soil condition is or various other farming-related topics. The text is then fed into the Text Preprocessing Layer for applying some normalization or cleaning operations to enhance the process of analysis of the query. The processed query is then sent to the SLM Understanding Engine which helps in recognizing the semantic intent of the user query and extracting the relevant agriculture entities to be retrieved in knowledge base. After understanding the query, the Agricultural Knowledge Retrieval Layer is activated in order to fetch the required information based on highest semantic relevance to the user query from an agriculture specific repository.

The knowledge retrieved are then filtered by Context Awareness Engine which uses auxiliary information like crop type, weather, season, soil type and the description of symptoms. The filtered content allows to prioritize the information based on its relevance to the user's farming environment. The subsequent information entries are then taken up by the Decision Support Generator for creating recommendations and advice. Finally, the results are put through to Advisory Output Layer that transforms the fetched knowledge into actual farming advice for the user. Figure 3 illustrates the overall workflow of the Agricultural Query Processing.

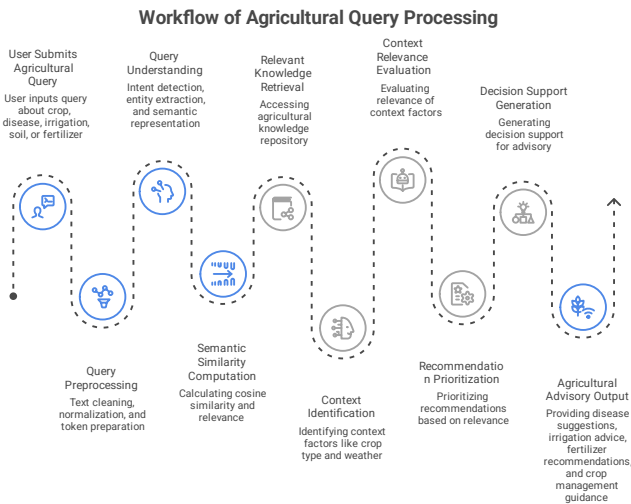


Figure 3. Workflow of Agricultural Query Processing

3.2 Knowledge Sources

Information retrieval and information recommendation generation can be achieved effectively with a comprehensive agricultural knowledge source. In the system, several domain-specific knowledge bases are considered such as crop management, crop health, nutrient requirements, irrigation practices, environment, etc, to answer various agricultural questions. Information sources used in information retrieval and information for agriculture decision support are the combination of those knowledge bases. The Crop Data provides information concerning the characteristics, growth regulation, growing requirement and crop management. Disease Database is built successfully by representing different symptoms, reasons, precautions and cure relating to various crops. Fertilizer Knowledge Base stores guidelines on nutrient management, nutrient suggestion on different applications and nutrient needs of different crops. Irrigation Knowledge Base also stores information about irrigation scheduling, water requirements and needed practices to achieve efficient irrigation. Additionally, the Weather Information Repository (WIR) is also a component which represents environment information and integrates with the recommender can be developed to provide context aware recommendation. Multiple knowledge sources are integrated so as to allow the framework to access information from various knowledge areas in agriculture, lending

itself to greater diversity and relevance of the recommendations produced. The framework proposes using structured agricultural repositories to ensure that information is consistent, domain specific and can be used for support services in agriculture.

3.3 Model Selection Strategy

Choosing the right language model is among the key issues in building agricultural decision support systems. As powerful as LLM-driven skills are in comprehending natural speech or creating knowledge, there are often significant computational power, memory and access to sophisticated computing systems needed to turn the LLM into an actionable generator. These features may pose problems when deployed in an agricultural setting, where hardware is limited, Internet connectivity is potentially sparse and costs are more significant. The framework proposed uses a Small Language Model approach because of the suitability for light-weight and resource-efficient deployment. Generally, the Small Language Models consume fewer computational power and yet retain the capacity of processing and understanding domain specific queries. The smaller model size reduces memory requirements, improves inference speed, and enhances deployment on edge devices and local systems. These characteristics make SLMs well-suited for agricultural advisory services, particularly in environments that require rapid response times and cost-effective operation. Apart from the efficiency of the computation, SLMs can be deployed locally to a greater extent and agricultural knowledge retrieval can be achieved without using a lot of cloud resources. It is especially important in low connectivity and rural areas where there are no guarantees of a continuous connection to an outside service. Therefore, the approach of SLM-based architecture addresses the goal of designing an easily accessible, scalable and context-aware framework for agricultural decision support.

As shown in Table 2, SLMs require fewer computational resources, lower memory consumption, and reduced deployment costs compared to LLMs. Their faster inference and suitability for edge and offline environments make them particularly appropriate for agricultural applications operating under resource constraints. Consequently, the proposed framework adopts an SLM-based architecture to support efficient knowledge retrieval and context-aware decision support while maintaining deployment feasibility and scalability.

Table 2. Comparison Between Large Language Models and Small Language Models for Agricultural Deployment

Feature	Large Language Models (LLMs)	Small Language Models (SLMs)
Model Size	High	Low
Computational Requirement	High	Moderate to Low
Memory Consumption	High	Low

Feature	Large Language Models (LLMs)	Small Language Models (SLMs)
Inference Speed	Moderate	Faster
Deployment Cost	High	Lower
Edge Device Deployment	Limited	More Suitable
Offline Operation	Challenging	More Feasible
Infrastructure Dependency	Significant	Reduced
Suitability for Resource-Constrained Environments	Limited	High
Applicability to Local Agricultural Advisory Systems	Moderate	High

4 Evaluation Framework

4.1 Evaluation Metrics

The success of the proposed framework can be evaluated with plots of retrieval performance, recommendation quality as well as computational efficiency measures. Because multiple components of this framework are involved in agricultural knowledge retrieval, agricultural analysis of knowledge context and generation of knowledge suggestions, multiple criteria must be adopted to evaluate the whole system. The performance of the model is measured using these indices, providing a formal approach for evaluating the accuracy, relevance, effectiveness and applicability of the model to agricultural consulting field. Some traditional classification metrics like Accuracy, Precision, Recalls and F1-Score could be employed to gauge how correctly the system retrieve relevant knowledge and output the recommendations. Precision refers to what percentage of the recommendations is relevant (irrelevant) and Recall is the percentage of the relevant agricultural knowledge retrieve for a given query. F1-Score is a metric that combines both precision and recall in a single measure and position it at a midpoint to evaluation efficiency of retrieval. Beyond the retrieval quality metrics, other operational metrics like inference time, memory consumption and response latency may be adopted to assess the efficiency of the framework from a computational perspective. Moreover, measures relevant to deployment are very relevant for the suitability evaluation in the field of agriculture. Using factors like the utilization of resources, deployment cost, scalability and operational feasibility on low-resource systems, help with the

analysis of the proposed framework's viability. These metrics are in line with the main requirement in building up a lightweight DSS for agriculture that can be used efficiently in resource-poor and limited connectivity environments. The metrics that are included in the proposed framework are summarized in Table 3.

Table 3. Evaluation Metrics for the Proposed Framework

Metric	Description	Purpose
Accuracy	Proportion of correct predictions or recommendations	Measures overall framework correctness
Precision	Ratio of relevant recommendations among retrieved recommendations	Evaluates recommendation relevance
Recall	Ratio of correctly identified relevant recommendations	Measures retrieval completeness
F1-Score	Harmonic mean of Precision and Recall	Provides balanced performance evaluation
Inference Time	Time required to generate recommendations	Measures computational efficiency
Response Latency	Time between query submission and output generation	Evaluates user experience
Memory Consumption	Amount of memory utilized during execution	Assesses resource efficiency
Deployment Cost	Estimated computational and infrastructure requirements	Evaluates deployment feasibility
Scalability	Ability to maintain performance with increasing workload	Assesses practical applicability

The proposed evaluation framework with both performance and deployment-oriented metrics are given in Table 3. In terms of the evaluation according to metrics, some of them are used to measure the quality of knowledge retrieval/recommendation generation, such as accuracy, precision, recall and F1-Score, and some are used to measure the operational efficiency of the framework, such as inference time, memory consumption, deployment cost and scalability. These supplementary metrics allow the system to be fully evaluated in terms of both accuracy of the recommendations and ensuring that they are also implemented. The proposed evaluating framework makes use of the F1-Score metric for evaluating the effectiveness of retrieval. F1-Score is a very common

measure of performance used in many information retrieval and classification problems as it takes into account both precision and recall. Recall and Precision are what is assessed in terms of the completeness of retrieval and meaning of retrieved recommendations, respectively, and F1-Score is a compromise, showing how relevant the retrieved recommendations. This is used especially when evaluating agricultural knowledge retrieval systems where there are two types of relevance to the recommendation (relevance to the individual requester and relevance to the overall recommendation), and two types of retrieval completeness (retrieval of the relevance to the individual requester and retrieval of the overall relevance). The F1-score is computed using equation (3).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

where *Precision* represents the proportion of relevant recommendations among all retrieved recommendations, and *Recall* denotes the proportion of relevant agricultural knowledge successfully retrieved by the framework. Higher F1-Score values indicate a better balance between retrieval accuracy and retrieval completeness, thereby reflecting the overall effectiveness of the proposed framework.

4.2 Expected Comparative Performance

The proposed framework aims to create a form of agricultural advice system that is efficient for our end-users and alternative to the computationally expensive language-based advice approach. Large Language Models are powerful tools for language comprehension, but may have strict demands for computing resources, infrastructure, and costs in the agricultural context. Unlike the conventional technique that targets to scale the model size, the proposed Small Language Model framework emphasis on resources-efficiency, high responsiveness, and appropriateness of the technique to contexts of limited computing resources and connectivity. By virtue of the inherent features of Small Language Models (SLMs) and the comparative results derived in Table 2, it is expected that the proposed framework contributes on query response, resource efficiency, real-scale deployment, and operation efficiency. With confined computing computations, SLMs achieve query efficiently, lower resource-demand, while still providing useful agricultural answers and context-based recommendations by their features on fast response, lower costs and field-deployment in an appropriate way for the farming application. An agricultural knowledge-retrieval and evaluating module can be added to improve the appropriateness of the response to the context. Combining semantic matching and context score evaluation in a joint model would make it possible for responses to be contextually-relevant and informative, considering various factors such as time, season and user. Efficient and context-aware recommendations provide an excellent foundation for adaptive agricultural advisory services adaptable to a range of agricultural environments. Finally, conceptually compares the proposed framework of using SLM with current LLM-based methods in Figure 4.

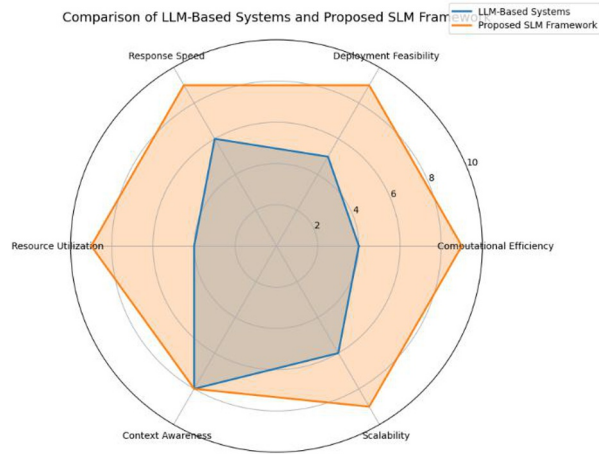


Figure 4. Conceptual comparison between conventional LLM-based agricultural systems and the proposed SLM framework across key deployment and performance dimensions.

The performance of the proposed framework is compared with the traditional LLM-based agricultural systems as shown in Figure 4. The proposed framework is expected to offer improved efficiency with less computational load, more deployment feasibility and speed, and better resource use when compared to both approaches in providing access to agricultural knowledge and decision support capabilities. The characteristics of the system are in line with developing a lightweight, ease-to-use agriculture advisory system which could be useful in diverse range of agriculture settings, including those of resource-poor regions.

5 Results and Discussions

The described system is expected to acquire the farming details more efficiently and offer some tips based on the situation, always demanding minimal effort. A Small Language Model integrates with domain-specific knowledge resources like Agricultural Knowledge Repositories to offer users relevant information and advisory services in an efficient manner, without the need for large-scale, power-hungry infrastructure. A semantic query understanding, knowledge retrieval and contextual evaluation unification is expected to give better relevance for the recommendations, both in relation to the user intent in a query, and with respect to the agricultural conditions related to the query. The lightweight architecture will be more likely to be deployed in agricultural applications which have limited resources, including computing power, connectivity and budgets. The proposed framework will be expected to have better inference efficiency, memory efficiency and deployment flexibility than the language model-based systems.

Such properties can help pave the way for agricultural advisory solutions that are suitable for low resource and edge computing settings and can satisfy reasonable expectation of recommendation quality. Recommendations taking into account contextual factors (e.g. crop type, weather events, seasonal information, soil properties etc.) and containing descriptions of the symptoms are expected to be more applicable. In addition to semantic matching the framework incorporates contextual evaluation to increase the relevance of the advisory products in the context and support agricultural decision making. It may find use in optimized crop management, disease diagnosis, irrigation planning, nutrient management and so on. From practical perspective, the proposed framework can be used as a basis to develop an ecological and functional accessible Agricultural Advisory System providing support for farmers and agricultural practitioners and Extension Services. Delivering agricultural information in a user-friendly, structured manner through light-weight language processing, knowledge retrieval and contextual reasoning. Thus, the framework can aid in the dissemination of intelligent agricultural support systems, and at the same time address the challenges that are normally encountered in large scale deployment of language-model based systems.

6 Challenges and Limitations

It is important that the benefits of the proposed framework/framing be acknowledged, but there are a number of issues and limitations to consider. Quality, completeness and credibility of agricultural knowledge repositories used to retrieve knowledge play a major role in making the framework efficient. Inaccurate, outdated or incomplete knowledge sources, can affect the quality of the knowledge sources retrieved, the relevance of the retrieved information, and subsequently the quality of the recommended information. Thus, it is imperative to keep agricultural information bases accurate and up-to-date for proper functioning of a framework. One of the disadvantages is agricultural studies' regional diversity of agricultural conditions. There may be significant variations between different properties in terms of farm management, disease prevalence, soil condition, and weather and resource levels. Thus, general knowledge recommendations given by knowledge repositories cannot always be adapted to local agronomic needs. Further mechanism of localization might be required in order to increase the degree of the suitability of the recommendations throughout different agricultural settings. Vague, partial and ill-structured user queries will also be a challenge in the framework. In some cases, the user's intent may have to be identified correctly because of the use of informal language, or different terms (agricultural related issues). Not only that, but the presented context-aware mechanism considers the crop, symptom and weather information, but there are other context factors specific to a farm's management practices, market conditions, pest pressures, etc., which are not explicitly addressed by the presented mechanism that influence decisions made for farming. Lastly, while Small Language Models have benefits related to computation and deployment, they typically have fewer parameters and lack in reasoning ability than their larger counterparts. In general, lightweight models tend to be cumbersome to operate for highly intricate

questions in the domain of agriculture that demand extensive reasoning or knowledge to give an answer, which cannot be done. These constraints are a motivation for future improvements, both in the amount of knowledge and context covered, and in the ability for localization and decision-making capacity.

7 Future Scope

The proposed framework provides a foundation for lightweight agricultural knowledge retrieval and context-aware decision support; however, several opportunities exist for further enhancement. One important direction is the incorporation of multilingual language support, enabling farmers to submit agricultural queries in regional and local languages. Such functionality would improve accessibility, promote wider adoption among diverse farming communities, and reduce language barriers in agricultural advisory services. Another promising extension is the integration of voice-based interaction, allowing users to communicate with the advisory system through speech. This would particularly benefit farmers with limited digital literacy and simplify access to agricultural recommendations. In addition, mobile deployment could further enhance accessibility by enabling the framework to operate efficiently on smartphones and other portable devices. Future work may also focus on integrating the framework with Internet of Things (IoT) devices, environmental sensors, and weather monitoring systems. Real-time field data can strengthen contextual analysis and improve the accuracy of recommendations related to irrigation scheduling, disease prevention, nutrient management, and crop planning. Furthermore, personalized recommendation mechanisms that consider farm-specific characteristics, cultivation history, and resource availability could improve decision support. Another potential research direction involves deploying the framework in edge-computing environments to enable localized processing while reducing dependence on continuous Internet connectivity. Collectively, these enhancements can improve the effectiveness, scalability, and practical applicability of the proposed framework while supporting the wider adoption of intelligent agricultural advisory systems.

8 Conclusion

The efficient generation of access to relevant knowledge and informed decisions are becoming more important as the agricultural system changes in terms of production, environment and resource management. Language-model based technologies have been advanced over the last few years to bring more capabilities to field experimental farming, however, their operation can be limited by computational load, reliance on infrastructure and costs. These challenges underscore the importance of light-weight, low-cost solutions that can provide specific information on agriculture in a variety of agricultural settings. The authors in this paper presented a Small Language Model Framework for Intelligent Agricultural Knowledge Retrieval and Context-Aware Decision Support. The framework has been integrated into an efficient

system design composed of computationally efficient building blocks like Natural Language query Understanding, Agriculture Knowledge retrieval, context-based evaluation, and recommendation generation. The system design focuses on important parameters including crop cultural information, environmental conditions, symptoms and weather information for effective, context-aware recommendations. The system architecture described here incorporates important features such as efficiency in computation, deploy-ability, flexibility, and compatibility to the computational resources available in an agricultural context. The system architecture proposed here forms a basic building block for developing lightweight advisory systems in agriculture which combines the aspect of knowledge retrieval with context aware decision support. It may allow the users to access agricultural knowledge, and receive recommendations at an actionable context using the semantic matching and contextual evaluation approach in the proposed system. Beside from the existing concept and techniques in smart farming, it would be more efficient by incorporating features like real time data access, natural language interface in other languages and context specific advice services.

References

1. Peláez, C. A., Solano, A., Corchado, J. M., & De la Prieta, F. (2025). Design of a GenAI UX layer with small language models for edge computing in smart agriculture. *Array*, 100632. <https://doi.org/10.1016/j.array.2025.100632>
2. Haghighat, M., Saleh, A., & Azghadi, M. R. (2025). Multimodal language models in agriculture: A tutorial and survey. *Information Fusion*, 104042. <https://doi.org/10.1016/j.inffus.2025.104042>
3. Li, H., Wu, H., Li, Q., & Zhao, C. (2025). A review on enhancing agricultural intelligence with large language models. *Artificial Intelligence in Agriculture*, 15(4), 671-685. <https://doi.org/10.1016/j.aiaa.2025.05.006>
4. De Clercq, D., Nehring, E., Mayne, H., & Mahdi, A. (2024). Large language models can help boost food production, but be mindful of their risks. *Frontiers in Artificial Intelligence*, 7, 1326153. <https://doi.org/10.3389/frai.2024.1326153>
5. Marinoudi, V., Benos, L., Villa, C. C., Kateris, D., Berruto, R., Pearson, S., ... & Bochtis, D. (2024). Large language models impact on agricultural workforce dynamics: Opportunity or risk?. *Smart Agricultural Technology*, 9, 100677. <https://doi.org/10.1016/j.atech.2024.100677>
6. Shaikh, T. A., Rasool, T., Venington, K., & Yaseen, S. M. (2025). The role of large language models in agriculture: harvesting the future with LLM intelligence. *Progress in Artificial Intelligence*, 14(2), 117-164. <https://doi.org/10.1007/s13748-024-00359-4>
7. Corradini, F., Leonesi, M., & Piangerelli, M. (2025). State of the art and future directions of small language models: a systematic review. *Big Data and Cognitive Computing*, 9(7), 189. <https://doi.org/10.3390/bdcc9070189>
8. Zhu, H., Qin, S., Su, M., Lin, C., Li, A., & Gao, J. (2025). Harnessing large vision and language models in agriculture: A review. *Frontiers in Plant Science*, 16, 1579355. <https://doi.org/10.3389/fpls.2025.1579355>
9. Wu, P., & Zhong, Y. (2025). Artificial intelligence in sustainable agriculture: Towards a socio-technical roadmap. *Smart Agricultural Technology*, 101578. <https://doi.org/10.1016/j.atech.2025.101578>

10. Reitano, M., Segovia, M. S., & Nayga Jr, R. M. (2025). A systematic review on the impact of Artificial Intelligence in the agri-food supply chain. *Food Policy*, 137, 102983. <https://doi.org/10.1016/j.foodpol.2025.102983>
11. Liu, J., & Wang, X. (2025). Multimodal large models for smart agriculture: Frameworks and applications. *Expert Systems with Applications*, 130394. <https://doi.org/10.1016/j.eswa.2025.130394>
12. Ge, C., Zhang, G., Wang, Y., Shao, D., Song, X., & Wang, Z. (2025). Research Status and Development Trends of Artificial Intelligence in Smart Agriculture. *Agriculture*, 15(21), 2247. <https://doi.org/10.3390/agriculture15212247>
13. Chan, Y. C., Fan, Y. C., Li, H. C., & Yang, M. D. (2026). SMART: Structured multimodal agricultural retrieval-augmented transformer for plant disease diagnosis. *Computers and Electronics in Agriculture*, 250, 111882. <https://doi.org/10.1016/j.compag.2026.111882>
14. Klerkx, L., Jakku, E., & Labarthe, P. (2019). A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. *NJAS: wageningen journal of life sciences*, 90(1), 1-16. <https://doi.org/10.1016/j.njas.2019.100315>
15. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147, 70-90. <https://doi.org/10.1016/j.compag.2018.02.016>
16. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
17. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming-a review. *Agricultural systems*, 153, 69-80. <https://doi.org/10.1016/j.agsy.2017.01.023>
18. Benos, L., Tsaopoulos, D., & Bochtis, D. (2020). A review on ergonomics in agriculture. part II: Mechanized operations. *Applied Sciences*, 10(10), 3484. <https://doi.org/10.3390/app10103484>
19. Talaviya, T., Shah, D., Patel, N., Yagnik, H., & Shah, M. (2020). Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artificial intelligence in agriculture*, 4, 58-73. <https://doi.org/10.1016/j.aiia.2020.04.002>
20. Chlingaryan, A., Sukkariéh, 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. <https://doi.org/10.1016/j.compag.2018.05.012>
21. Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 170, 105256. <https://doi.org/10.1016/j.compag.2020.105256>
22. Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2019). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE access*, 7, 129551-129583. <https://doi.org/10.1109/ACCESS.2019.2932609>
23. El Bilali, H., Dambo, L., Nanema, J., Tietiambo, S. R. F., Dan Guimbo, I., & Nanema, R. K. (2023). Impacts of the COVID-19 pandemic on agri-food systems in West Africa. *Sustainability*, 15(13), 10643. <https://doi.org/10.3390/su151310643>
24. Saiz-Rubio, V., & Rovira-Más, F. (2020). From smart farming towards agriculture 5.0: A review on crop data management. *Agronomy*, 10(2), 207. <https://doi.org/10.3390/agronomy10020207>
25. Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE access*, 9, 4843-4873. <https://doi.org/10.1109/ACCESS.2020.3048415>

26. Fountas, S., Carli, G., Sørensen, C. G., Tsiropoulos, Z., Cavalaris, C., Vatsanidou, A., ... & Tisserye, B. A. (2015). Farm management information systems: Current situation and future perspectives. *Computers and electronics in agriculture*, 115, 40-50. <https://doi.org/10.1016/j.compag.2015.05.011>
27. Lowenberg-DeBoer, J., & Erickson, B. (2019). Setting the record straight on precision agriculture adoption. *Agronomy journal*, 111(4), 1552-1569. <https://doi.org/10.2134/agronj2018.12.0779>
28. Tzounis, A., Katsoulas, N., Bartzanas, T., & Kittas, C. (2017). Internet of Things in agriculture, recent advances and future challenges. *Biosystems engineering*, 164, 31-48. <https://doi.org/10.1016/j.biosystemseng.2017.09.007>
29. Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture*, 13(6), 693-712. <https://doi.org/10.1007/s11119-012-9274-5>
30. Verdouw, C., Tekinerdogan, B., Beulens, A., & Wolfert, S. (2021). Digital twins in smart farming. *Agricultural Systems*, 189, 103046. <https://doi.org/10.1016/j.agry.2020.103046>
31. Aubert, B. A., Schroeder, A., & Grimaudo, J. (2012). IT as enabler of sustainable farming: An empirical analysis of farmers' adoption decision of precision agriculture technology. *Decision support systems*, 54(1), 510-520. <https://doi.org/10.1016/j.dss.2012.07.002>
32. Kamble, S. S., Gunasekaran, A., & Gawankar, S. A. (2020). Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. *International journal of production economics*, 219, 179-194. <https://doi.org/10.1016/j.ijpe.2019.05.022>
33. Balafoutis, A., Beck, B., Fountas, S., Vangeyte, J., Van der Wal, T., Soto, I., ... & Eory, V. (2017). Precision agriculture technologies positively contributing to GHG emissions mitigation, farm productivity and economics. *Sustainability*, 9(8), 1339. <https://doi.org/10.3390/su9081339>
34. Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering*, 114(4), 358-371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>
35. Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. <https://doi.org/10.1145/361219.361220>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

