Frontiers in Sustainable Food Systems

TYPE Original Research
PUBLISHED 09 January 2025
DOI 10.3389/fsufs.2024.1445795

Check for updates

OPEN ACCESS

EDITED BY
Fernando Casanoves,
Centro Agronomico Tropical De Investigacion Y Ensenanza Catie, Costa Rica

REVIEWED BY
Athina Koutouleas,
The University of Copenhagen, Denmark
Didier Snoeck,
CIRAD—UMR AbSys, France

*CORRESPONDENCE
Edinson David León Chilito
✉ leoc@unicauca.edu.co

RECEIVED 08 June 2024
ACCEPTED 06 December 2024
PUBLISHED 09 January 2025

CITATION
León Chilito ED, Casanova Olaya JF, Corrales JC and Figueroa C (2025) Sustainability-driven fertilizer recommender system for coffee crops using case-based reasoning approach.
Front. Sustain. Food Syst. 8:1445795.
doi: 10.3389/fsufs.2024.1445795

Sustainability-driven fertilizer recommender system for coffee crops using case-based reasoning approach

Edinson David León Chilito[1]*, Juan Fernando Casanova Olaya[1,2], Juan Carlos Corrales[1] and Cristhian Figueroa[1]

[1]Telematics Engineering Group, Faculty of Electronics and Telecommunications Engineering, Universidad del Cauca, Popayán, Cauca, Colombia, [2]Ecotecma S.A.S, Popayán, Cauca, Colombia

Introduction: In recent years, the increased demand for food has prompted farmers to increase production to support economic expansion. However, the excessive use of mineral fertilizers poses a significant threat to the sustainability of food systems. In Colombia, coffee cultivation plays a fundamental role in the economy, thus creating a recognized demand to elevate its production while minimizing its environmental impact sustainably.

Methodology: The study follows the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining) developing of a fertilizer recommender system (FRS) for coffee crops. This process includes business understanding, where the key factors influencing coffee production were identified; data understanding and preparation, where agroclimatic data and expert knowledge were collected and processed; modeling, which involved building a case-based reasoning (CBR) system to recommend fertilizer doses and frequencies, and evaluation, where expert feedback was gathered to assess the system's performance. The CBR system integrates soil, crop, and climate variables to provide tailored recommendations for nitrogen, phosphorus, and potassium applications.

Results: The results revealed that the FRS was deemed acceptable for application in the region, with expert evaluations rating the recommendations based on their experience and knowledge. Additionally, valuable feedback was provided to facilitate future enhancements to the system.

Discussion: Based on expert feedback and system performance, the proposed FRS meets the minimum requirements for deployment in real crops, serving as a valuable tool for small-scale farmers. Future work will expand the case base and refine recommender algorithms to improve accuracy and usability.

KEYWORDS

crop management, knowledge base farming, environmental sustainability, expert system, smart farming

1 Introduction

Agricultural production contributes to Colombia's economic growth and development, contributing 8.3% to the national gross domestic product (World Bank, 2022). However, agriculture substantially impacts the environment by producing food, fuel, and fibers to meet human needs (Boregowda et al., 2022). It is a leading cause of chemical and organic pollution to surface water and groundwater resources (Drechsel et al., 2023). It contributes to the release of greenhouse gasses (GHG) emissions (N2O), contributing to climate change (Rodríguez-Espinosa et al., 2023).

Colombia's four major export products are coffee, flowers, bananas, and sugar. Coffee, in particular, substantially impacts the country's economic growth, with 884,000 ha cultivated and 540,000 families relying on coffee production (Vélez-Vallejo, 2022; León-Burgos et al., 2022). The coffee industry is particularly affected by climate change, which threatens cultivable land and compels farmers to seek higher altitudes for optimal growing conditions for coffee crops (Bilen et al., 2023).

In this respect, coffee producers increasingly prioritize sustainability, which involves extra costs impacting farm profitability. Improving fertilizer application is one action that can contribute to sustainability and the reduction of climate change. Fertilizer misuse, particularly overapplication, significantly affects the economy and the environment (Martín Alonso et al., 2016). According to Lenka et al. (2016) and Sainju (2017), this is because crops typically utilize only 50–60% of the applied fertilizer, releasing a residual portion into the environment through natural processes such as leaching, denitrification, surface runoff, and soil erosion.

Analyzing agroclimatic and crop data is essential to provide accurate recommendations on coffee crop fertilizers and other agricultural inputs. This helps to improve crop efficiency and maintain environmental responsibility. Unfortunately, many small Colombian farmers do not have access to technological tools for data collection due to a lack of knowledge or education in using them (Chaves, 2016). Therefore, coffee farmers require better access to data about their crops, as there is a need for more data throughout the region (Sylvester et al., 2020). In this context, it is necessary to propose research studies that address fertilizer recommendations in scenarios with limited data, relying on scientific literature and expert knowledge (Howland et al., 2015).

This paper proposes a fertilizer recommender system (FRS) to address the environmental impact of fertilizer application in coffee crops in Colombia. The FRS was developed using a case-based reasoning (CBR) approach, a problem-solving methodology that uses past experiences or "cases" to inform new decisions. CBR operates by retrieving the most similar previous cases from a case base, reusing their solutions, revising them if necessary, and retaining the latest solutions for future use (Kolodner, 1992). In this context, the system integrates expert knowledge and agroclimatic data collected from local coffee growers and governmental institutions. The proposed FRS recommends the amount of nitrogen, phosphorus, and potassium fertilizer to be applied to a coffee crop, considering the balance between agricultural production and environmental preservation.

2 Related work

In recent years, the integration of intelligent systems in agriculture has gained significant momentum, with increasingly advanced systems combining expert knowledge with data. To explain these works, three main groups were classified based on the data sources used to generate recommendations or perform decision-support systems.

The first category includes works that collect data through sensors, leveraging real-time data from environmental sensors to monitor and adjust agricultural practices. The second category includes systems that rely on historical data, using past agricultural performance and weather trends to make predictions or provide recommendations. Finally, the third category encompasses systems that integrate expert knowledge, using the perspectives of farming professionals to guide decisions in crop management, fertilization, and pest control.

In the first category, the primary data collection method is sensor-based data gathering, where data from both the soil and the crop plant are collected. These data are subsequently analyzed and used to generate recommendations. For example, Kumar et al. (2019) developed a system that utilizes data from the soil, color sensors, and chemical processes to detect potential nutrient levels in the soil to provide fertilizer recommendations to small farmers in India in crops such as wheat, barley, corn, and sugar cane among others. Other studies, such as Wickramasinghe et al. (2019) and McFadden et al. (2018), employ machine learning algorithms like Support Vector Machines (SVM) and Bayesian models to analyze the data sensors to predict the necessary fertilizer quantities. In these works, the authors use previous information from the farmer that they combine with data from sensors that measure soil fertility to improve the estimation of agricultural production. The systems are developed and tested in small areas of crops where we can find corn, peanuts, beans, bananas, tomatoes, and sugarcane, among others. Among such works, Sujithra et al. (2019) developed a classification model where the input parameters consist of soil variables (where NPK, pH, temperature, and humidity stand out) collected by wireless sensors. The system experiments with J48, SVM, and k-means decision tree algorithms to select the most suitable classifier. The results indicated that the J48 algorithm better classified NPK availability in soil than the others, so it was chosen to make a more accurate classification. Subsequently, the data they collected in the field was taken as test data and compared with the trained data that had already entered the system to suggest fertilizers for cases with macronutrient deficiency in the soil.

The study by Qin et al. (2018) proposed a content-based RS for predicting the optimal nitrogen rate for corn crops in the USA. For this, they captured data from weather stations, sensors, and soil profile samples, then tested some ML algorithms such as Linear Regression (LR), Ridge Regression (RR), Most Minor Absolute Shrinkage and Selection Operator (LASSO), and Gradient Boosting Regression Trees (GBRT). To evaluate their results, they used R2, MAE, and RMSE. The ridge regression algorithm presented the best performance with 70% success in the evaluation. The research by Islam et al. (2020) enables the determination of nitrogen demand in plants using a dataset of 6,000 rice leaf images. These images were classified using a Convolutional Neural Network (CNN) and a decision tree to determine the necessary amount of nitrogen that farmers should apply.

The second category comprises studies about FRS based on the SVM algorithm to analyze historical data from governmental institutions. Suchithra and Pai (2018) created an FRS that generates fertilizer type and quantity recommendations in this category. Their system leverages historical records of soil and crop variables spanning multiple years. Another study in this category is Jiang et al. (2020), which used historical data such as applied nitrogen rates, crop yield, location, rainfall, temperature, pH, soil organic

10.3389/fsufs.2024.1445795

carbon, phosphorus, and potassium, from 31 experimental plots of 60 m^2 over 5 years located in the central corn producing region of northwest China, to construct a quadratic model. Puntel et al. (2019) employed some regression algorithms using data from 54 agroclimatic variables (like pH, organic matter, elevation, depth, previous crop yield, soil moisture, soil nitrate, precipitation, air temperature, and among others) which are obtained from historical databases and meteorological stations in the region, to recommend to farmers the optimal rate of nitrogen to apply. Vieira Fontoura et al. (2017) implemented an RS focused only on the nitrogen nutrient, designed for wheat and barley crops in Parana, Brazil. The proposed RS is based on content with data from 70 field experiments carried out between 2007 and 2012. These historical data correspond to Organic Matter (OM) values, pH, P, K, applied N rates, and the yield of the crops planted in that period. Thus, the system tries to obtain the maximum economic efficiency of N application rates in crops through data correlation analysis.

The second category also included studies where data could be more present, complete, or partially collected. These studies often rely on expert information to support their investigations. Ren and Lu (2012) and Zhang et al. (2011) proposed a DSS to recommend the most suitable fertilizers for specific crops, employing historical knowledge databases encompassing data from soil, crops, fertilization, and previous yields. Hossain and Siddique (2020) propose to address the problem of intensive input use in Bangladesh through the Soil Resources Development Institute's Online Fertilizer Recommender System (OFRS), which uses a national database to generate specific fertilizer recommendations. Cholissodin et al. (2016) developed a knowledge-based RS using experimental fertilizer data, employing algorithms such as Artificial Neural Networks (ANN) and Bayesian Improved Particle Swarm Optimization (BIPSO) to determine the required fertilizer dosage. Finally, a knowledge-based RS was also identified that utilizes fuzzy logic to recommend NPK fertilizer dosages, as presented by Sumaryanti et al. (2019).

The third category included studies that provide precise recommendations tailored to the specific needs of farmers and crops using expert knowledge represented as ontologies and agricultural data from sensors and image analysis. The drawback of works in this category, like Acuña (2019) and Chougule et al. (2019), which developed an RS that was fed with historical data and expert information from government databases, which was converted and stored in ontologies, subsequently the data were analyzed and studied with two machine learning algorithms: which were the grouping of k -means (k -means clustering) and random forest. The data that made up the knowledge base was a history of the last three years of the NPK content in the soil, the types of crops that grew in that soil, the climatic conditions, and what the production of those crops was like. Finally, with this history, the RS recommends to farmers based on the region, NPK content in the soil, and crop type, stating that the system's performance is highly accurate and that they expect it to achieve the goal of improving agricultural production in that country area.

While significant progress has been made in integrating sensor data, historical data, and expert knowledge for agricultural recommendations, several challenges remain. This study was conducted in the Cauca Department, a southwestern Colombia region, characterized by its diverse altitudes and predominantly Coffea arabica cultivation. The region's latitude influences the number of seasons, and coffee is often grown under shade trees rather than in full sun. One of the primary challenges in Cauca is the need for more data, as many farmers need access to advanced technological tools, which limits the integration of sensor and historical data. Future research should focus on developing scalable and adaptable systems that can leverage global data sets and expert knowledge to provide universally applicable recommendations, especially in technologically constrained regions.

3 Materials and methods

3.1 Phase 0: methods

Before developing the FRS, we conducted a Systematic Literature Review (SLR) following Kitchenham guidelines Kitchenham and Charters (2007) to identify relevant works in the field, as explained in the previous section. The objective was to identify agricultural smart systems and the most common and significant agroclimatic variables considered in these systems.

The SLR process began with an initial search in Scopus and Web of Science databases, focusing on Recommender Systems (RS), Prediction Systems, Decision Support Systems (DSS), and Expert Systems within the agricultural domain. After applying exclusion criteria to omit non-relevant works such as secondary sources, non-English publications, and non-agricultural studies, we identified 102 articles that met our inclusion criteria. These articles were further analyzed based on their geographic focus, with notable contributions from countries such as India (36), Indonesia (10), China (8), and the US (7).

In addition, we analyzed the agroclimatic variables used in these systems. The variables were categorized, and a bar chart (Figure 1) was generated to illustrate the frequency of these variables in the reviewed studies. The most frequently used variables were temperature (36 articles), pH levels (28 articles), disease incidence (26 articles), and soil moisture (16 articles). Other significant variables included soil NPK content (12 articles), crop yield (11 articles), crop type (11 articles), and pests (10 articles).

This analysis is visualized in Figure 2, which shows the distribution of the most used agroclimatic variables in smart systems within agriculture. The frequency of these variables highlights the diversity of factors that must be considered in agricultural decision-making systems, emphasizing the complexity of integrating environmental and crop-specific data.

The results of this SLR provided insights into the state of the art in smart agriculture systems and established a basis for the design of the system proposed in this work. This review also highlighted the importance of incorporating expert knowledge and agroclimatic data to improve the accuracy and relevance of recommendations.

Following this review, we adopted the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology to develop the FRS. The CRISP-DM phases that guided our development include business understanding, data understanding, modeling, and evaluation. The system was created in the Cauca Department, located in southwestern Colombia (with approximate

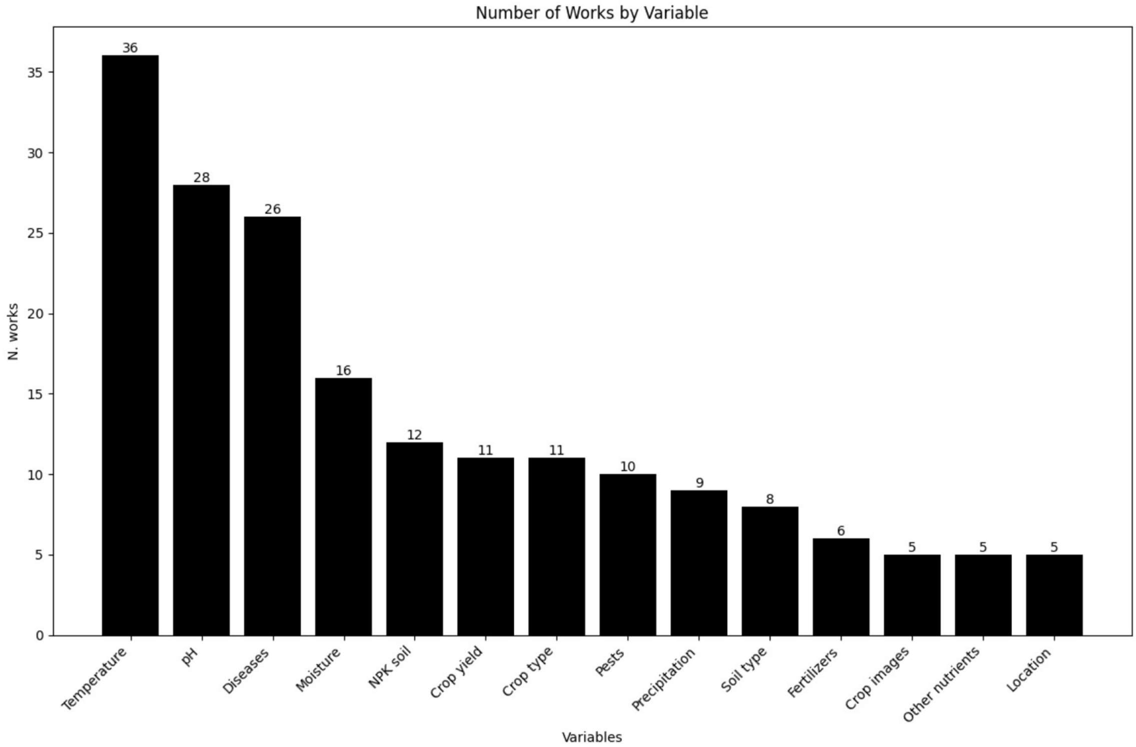

FIGURE 1
Input variables in the works.

coordinates of 2○30' N latitude and 76○30' W longitude). Cauca is characterized by a diverse range of altitudes (from 1,000 to 2,800 m above sea level), which influences its climate and, consequently, its agricultural practices. Coffee cultivation in Cauca primarily involves coffea arabica grown under shaded trees, although some coffee is grown in full sun. The region's latitude leads to distinct rainy and dry seasons, directly affecting fertilization practices. The phases of the CRISP-DM methodology applied in this work are detailed in the following sections.

3.2 Phase 1: study area

Initially, a business understanding was conducted to identify the critical factors in the development and growth of a crop. Like any other plant species, coffee cultivation requires essential elements for its development. Three of these elements, known as organic constituents, are freely available in the environment: carbon, hydrogen, and oxygen. According to Sadeghian (2008) they are obtained from water and the atmosphere, representing 95% of the plant's weight. The remaining 5% is found in the soil and is known as minerals, which are classified as macronutrients such as nitrogen, phosphorus, potassium, calcium, magnesium, and sulfur, as well as micronutrients like iron, manganese, copper, boron, chlorine, molybdenum, and nickel.

It is essential to mention that macronutrients are required in larger quantities and are further categorized into primary macronutrients (N, P, and K) and secondary macronutrients (S, Ca, and Mg). However, while N, P, and K are applied annually as fertilizers, Ca and Mg are considered soil amendments. These amendments are typically used in a single dose, usually before planting, to correct soil pH. As Sadeghian Khalajabadi (2017) explained, although primary macronutrients are applied more frequently, the doses of Ca and Mg are calculated to have a long-term effect. Additionally, due to changes and climatic phenomena that occur over time, various natural processes can lead to the loss of these nutrients in the soil.

- Leaching: According to Sadeghian Khalajabadi et al. (2015), leaching is the displacement of nutrients and subtances below the crop's root zone toward water bodies due to excessive moisture in the soil.
- According to Valdivielso (2020), surface runoff is precipitation that flows over the soil surface under the influence of gravity without infiltrating into the soil.
- Erosion: the wearing away of the Earth's surface due to various natural events such as rainfall, sunlight, and natural disasters, as well as causes generated by human improper use of soil resources.
- Denitrification: due to the presence of a large number of microorganisms that use nitrite and nitrate instead of oxygen, the production of gaseous forms of nitrogen occurs, including nitrous oxide, which is one of the leading air pollutants, by Sadeghian Khalajabadi et al. (2015).

The fertilization process addresses the nutrient deficiency in the soil, which provides the necessary supplements for the

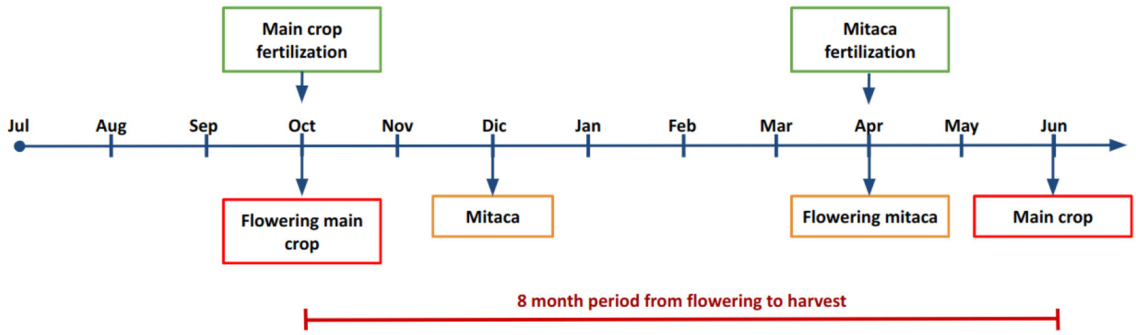

FIGURE 2
Schedule of coffee cultivation in Cauca (illustrated with experts).

plant to be productive and prevent nutrient losses. The fertilizer dose is typically calculated based on the crop's specific nutrient requirements, considering factors such as soil nutrient content, plant nutrient uptake, and expected yield. Soil testing is often used to determine the current levels of critical nutrients like N, P, and K, and recommendations are made to apply enough fertilizer to meet the plant's needs without over-fertilizing. Pozas (2008) states that one crucial factor to consider in this process is the stage of the crop. There are two stages in coffee crops: vegetative (juvenile) and production (adult). This study is centered around the production stage of coffee, which starts with the first crop harvest. It's worth noting that the timing and frequency of fertilizer application are just as important as the quantity applied. Hence, this initial phase also identifies the most appropriate periods for fertilizing a coffee crop during its production stage.

It is essential to recognize that crop fertilization needs to be timed correctly. It is crucial to determine the time from flowering to harvest for coffee crops. Figure 2 illustrates a coffee-growing period, known as a "coffee year." This figure shows two flowering periods and two harvest periods throughout the year.

In the Cauca department, the coffee year typically starts in July and ends in June of the following year. The flowering of the coffee crop usually occurs between September and November, with 8 months until the harvest or production of the crop. Thus, it is essential to fertilize the coffee during this period. However, coffee also has a second flowering, which results in a smaller harvest known locally as "Mitaca." The Mitaca typically represents 40–50% of the main harvest. Consequently, there are two periods of fertilization in coffee throughout the year to account for the main harvest and the Mitaca.

3.3 Phase 2: determination of variables

A study was conducted to collect data on the agroclimatic factors that affect crop fertilization or are considered significant from the domain perspective. The goal was to determine which variables would be addressed in the system. Afterward, these variables were gathered using a wireless sensor network, information provided by coffee farmers, and a service for extracting historical meteorological data from weather stations.

TABLE 1 Agroclimatic variables studied.

Variable type	Variable	Unit of measurement
Crop	Planting density	No. of plants per hectare (plan/ha)
Crop	Shade coverage	Percentage (%)
Crop	Flowering date	Date
Climate	Rainfall	millimeters of rain
Soil	N level	mg/kg
Soil	P level	mg/kg
Soil	K level	mg/kg
Soil	Moisture	Percentage (%)
Soil	pH	pH level

Determination of variables: This work involved collaborating with various agricultural experts affiliated with ECOTECMA SAS. Then, a knowledge base was constructed with their guidance. This involved a systematic compilation process encompassing reviewing reports, books, summaries, yearbooks, and bulletins sourced from CENICAFÉ. This knowledge base was constructed to determine the relevant soil, climate, and coffee crop variables deemed crucial for comprehensive study and analysis. Table 1 shows these variables.

Table 1 displays the fundamental variables for studying coffee crops. These variables are defined as follows:

- Planting density: This factor depends on the type of coffee planted by the coffee farmer, with the most common being Robusta coffee. A low planting density is considered when values are below 5,000 trees per hectare, and a medium or average density falls between 5,000 and 6,000 trees (approximate values for Robusta coffee), and a high density is considered when the number of trees per hectare exceeds 6,000, Arcila et al. (2007) noted.
- Shade coverage: This variable refers to the shade the coffee crop receives per hectare. It is measured in percentage, and low values are considered when below 30%, medium values range from 30 to 60%, and high values are above 60%. It is important to note that for proper development in coffee

crops, water availability, and sunlight exposure must be controlled to develop coffee crops properly. Effective shade management in the crop contributes to maintaining soil fertility, nutrient recycling, and erosion reduction (essential during dry seasons), supported by the findings of Farfán and Mestre (2004).
- The flowering date refers to the date when the crop flowers. Knowing this date is essential because, according to Sadeghian (2008), the fertilization process should be initiated on this day.
- Rainfall precipitation measures the accumulated rainfall in a specific region during a day. By studying these accumulations over a certain period, the state of the climatic season can be identified. Therefore, based on the collected expert information, it was determined to establish an analysis of historical meteorological data to classify the climatic season in a specific period (dry, normal, or rainy), as mentioned in Gast et al. (2013).
- Soil nitrogen: Nitrogen levels in coffee crops can range between 0 and 225 mg/kg. The appropriate nitrogen range is between 51 mg/kg for optimal crop development and 87 mg/kg. If nitrogen levels fall below 51 mg/kg, it leads to a nutritional deficiency. This shortage can adversely affect chlorophyll, essential for photosynthesis, thereby hindering the healthy growth of the plant. On the other hand, nitrogen levels exceeding 87 mg/kg suggest an overabundance or misuse of this nutrient. This represents waste from an economic standpoint for the farmer and carries the risk of causing environmental pollution. This information and the next were taken from Sadeghian Khalajabadi (2017).
- Soil phosphorus: The possible range of values in the soil is between 0 and 80 mg/kg, with suitable values for coffee crops falling between 10 and 20 mg/kg. If the values are below 10 mg/kg, the plant may exhibit uneven yellowing in older leaves, accompanied by reddish spots, and in severe cases, defoliation. If the value exceeds 20 mg/kg, it is considered a high soil phosphorus value, which can lead to the blocking of boron absorption in plants.
- Soil potassium: Its values range from 0 to 546 mg/kg. The appropriate values for coffee crops are between 78 and 156 mg/kg. A potassium value below 78 mg/kg reduces fruit size and leaf defoliation. If the potassium value exceeds 156 mg/kg, block in the absorption of micronutrients in plants.
- pH: Sadeghian Khalajabadi (2016) observes that it is measured on a scale of 0–14. The appropriate pH value for coffee crops should be between 5.5 and 6.5. If the soil pH is below 5, it is considered acidic soil, which affects the growth of plant roots and hinders the proper absorption of nutrients. It should be noted that acidic soils also block the absorption of potassium and nitrogen while promoting the absorption of manganese at levels that can be toxic to crops. If the pH is above 6.5, it leads to a blockage in the absorption of phosphorus, iron, zinc, and copper, resulting in a lower availability of these nutrients.

After the variables to be included in the development of this work were determined, the system architecture was designed, as shown in Figure 3. The collection and processing of data in the CBR system was done from three primary data sources that are integrated to generate the system recommendations:

IoT sensors in the field: A sensor network was deployed on a small farm in Piendamó, Cauca, to collect real-time critical soil data. The sensors include the 7-in-1 NPK, humidity, temperature, salinity, and electrical conductivity (EC). This sensor, with an NPK measurement range of 0 to 1,999 mg/kg and an accuracy of ±2%, was essential for measuring nutrient levels in the soil. The sensor sends its data to a collector via RS485, a communication protocol that ensures data transmission. The collector, in turn, transmits the data to a LORA gateway that uses RF communication at frequencies of 915–960 MHz with a range of up to 2 km. Finally, the data is sent via GPRS/GSM to a central server where it is stored and processed.

Features of the IoT devices used include:

- 7-in-1 NPK sensor: measures nitrogen, phosphorus, potassium, humidity, temperature and electrical conductivity with high accuracy and a resolution of 1 mg/kg for NPK.
- Collector: responsible for collecting and transmitting the data from the sensors, using the MODBUS-RTU protocol and LORA modulation with a range of up to 2 km.
- LORA gateway: device that connects the sensor data to the central server using network technologies such as GPRS and GSM.

Historical data comes from weather stations and public databases that record precipitation, temperature, and other agroclimatic conditions over time. This data is captured and sent to a processing server via HTTP. The server stores historical and real-time sensor data; for later use in the CBR. Besides automatically collecting data, the system feeds specialist information. This knowledge base contains rules and recommendations drawn from previous research and consultations with field professionals, which enriches the recommendations generated. Expert knowledge is integrated into the CBR system, which allows fertilization recommendations to be adjusted based on the specific farm context and soil conditions.

The system architecture therefore, combines three data sources: real-time sensors, historical databases, and expert knowledge. The central server processes all this data, allowing the CBR system to create new cases and make recommendations for crop fertilization. This integration of multiple sources of information makes it possible to improve the accuracy of recommendations and adapt them to the specific needs of each agricultural situation.

3.4 Phase 3: system construction

In this phase, an exploratory analysis of the variables determined in the previous section was conducted, identifying the numerical and categorical variables. Subsequently, the structure of each case in the CBR system was established, defining the data that constitute the problem and the solution for each case.

Case structure:

- Problem: This represents the part of the case that describes the situation that needs to be resolved or for which a solution needs to be found. It is represented by data or information that describes the need or problem. For the proposed CBR system, the problem consists of the variables determined in

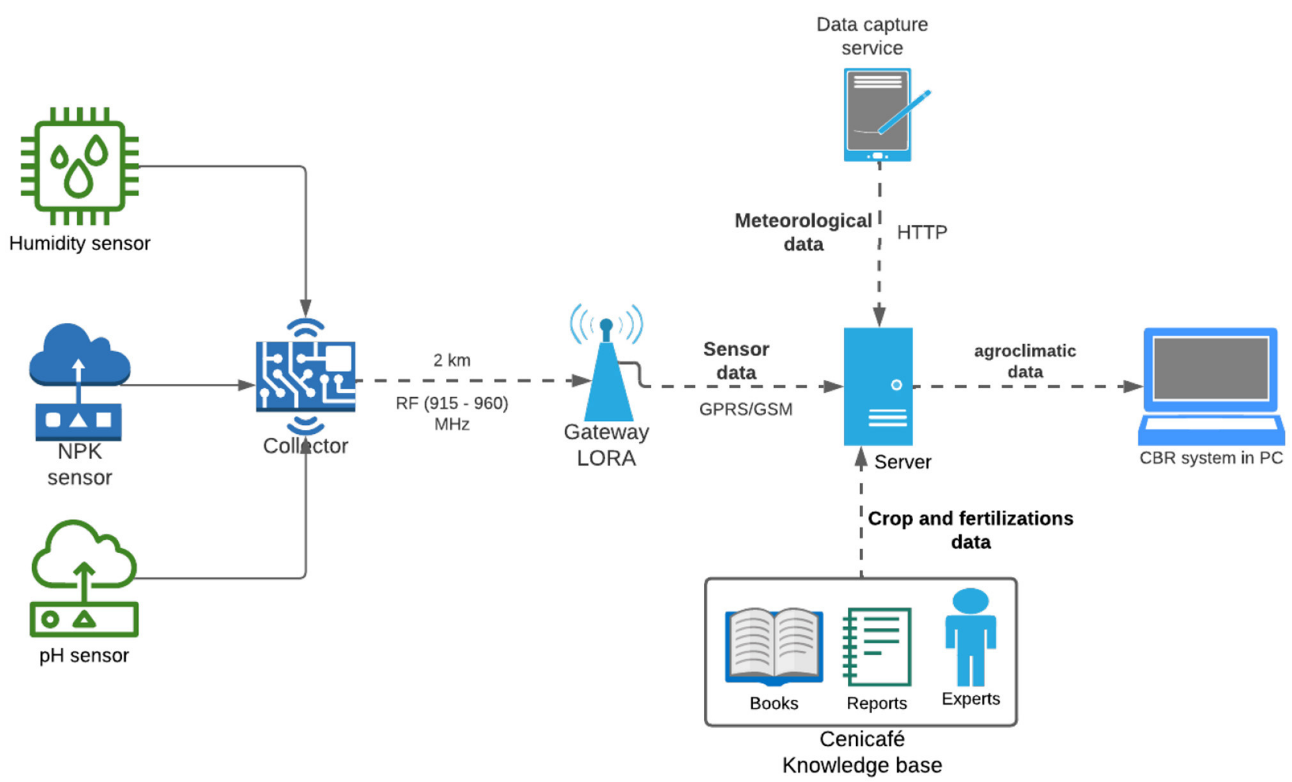

FIGURE 3
System architecture.

the previous section, except for the flowering date and rainfall precipitation, which are used to classify the climatic season that occurred 4 months before the flowering date.

- Solution: The solution was constructed using the system's output data, which corresponds to the fertilizer quantity rates that the system will recommend to the coffee farmer based on the data comprising the case problem. Three recommend fertilizer rates are provided due to the three most essential macronutrients identified in the first phase of CRISP-DM (N, P, and K).

3.4.1 Classification of climatic season

It is necessary to organize the climate seasons from 2006 to the established date to classify the climatic season based on the date of the last crop flowering. These seasons were identified based on historical records of dry or rainy periods each year obtained from the National Weather Service (2023). Using the classification provided by the NWS, eight years were identified as usual, five years as rainy, and three years as dry from 2006 to 2021.

With these historical classifications, historical data was analyzed to define a new classification of the climatic season for the year in which the last flowering date in the crop occurred. The mentioned classification was based on information suggested by experts. It involved obtaining the accumulated precipitation (PP) for the last 4 months before the flowering date and the historical accumulated PP for those same months in the years classified as standard within the 2006–2021 period. The obtained values are compared, and the season is categorized based on the comparison results. Suppose the accumulated PP in the year of the flowering date exceeds the historical average PP by more than 35%. In that case, the climatic season at that flowering date is classified as rainy. Conversely, the season is dry if the comparison shows that the accumulated PP is lower by 25% than the historical average. The season is classified as usual if these conditions still need to be met.

Therefore, by classifying a climatic season at the beginning of the fertilization period in the crop, the system can define the frequency of fertilizer application it recommends. According to expert information, applying fertilizer three times per period (at 2, 4, and 6 months after the flowering date) is appropriate if the season is classified as usual or rainy. On the other hand, if the season is classified as dry, applying fertilizer twice per period (at 3 and 6 months after the flowering date) is recommended.

3.4.2 Case base

The case base is the core of any CBR system, as it directly depends on the existence of a case base to perform all the steps in a CBR cycle, as Sànchez-Marrè (2001) mentioned. A case can be constructed by human experts or past experiences of the system. Cases can be represented in various ways, such as rules, logical formulas, frames, and database records (Shang, 2005). In this work, the case base was constructed, considering the knowledge base explained in the data collection process and determining variables for the system. Subsequently, experts validated this case base.

Different possible combinations of problem data were considered, and the most suitable solutions were determined based on expert knowledge. In each case, the variables of climatic season, humidity, pH, and NPK in the soil represent the problem of a case, and the fertilizer rate data represent the solution. For example, in a case where there is a rainy season, normal pH, low NPK, and high humidity, it may indicate unfavorable soil conditions for the crop, which would require a high NPK fertilizer rate. In this way, 300 problems are constructed, representing different situations that can occur in a coffee crop, and expert solutions are assigned to each of these problems, resulting in a case base of 300 cases.

3.5 Phase 4: modeling

In this phase, the model of the developed CBR system was implemented using Python. The CBR system follows the classic retrieval, reuse, revision, and retention steps described by Kolodner (1992), also known as the 4 R's of CBR. The first two steps in our implementation–retrieving the most similar cases and determining a new solution—were the focus. The data manipulation and case creation were done using Python libraries such as Pandas and Numpy, which facilitated the processing of historical climatic data and expert knowledge. These tools were used to build the case base for the system. The code was developed and tested in Python Notebooks, using the Google Colab platform to effectively run and visualize the results. This environment allowed for interactive data analysis and collab development, making refining the CBR model based on the input data easier.

- Retrieval: Given a new problem, the CBR system uses an algorithm to retrieve the cases that are most similar to it. Miller (2019) used the K-Nearest Neighbors (KNN) algorithm, which can be used for classification or regression tasks. This algorithm uses a factor k, which indicates the number of nearest neighbors (similar cases) to consider making a prediction.

  A comparison was made between each case data point using a similarity measure to find the most similar cases (nearest neighbors) given a new input. According to Gabel (2010), the choice of similarity measures in a CBR system depend on the composition of the problem variables. The Hamming distance or the Simple Matching Coefficient (SMC) is commonly used if the values are discrete. The cosine similarity measure is used if the data is symbolic or character strings. Similarly, the Manhattan, Euclidean, and Mahalanobis distances are used for real-valued numeric data. In this work, the Euclidean distance was chosen due to the numerical nature of the problem data and the advantage of feature weighting that this similarity measure possesses.

  Next, by calculating the distance between each data point of the two cases, similarities are obtained for each problem variable. It means that two problems can be similar from the perspective of a particular variable but entirely different when viewed from another variable. Therefore, assigning a weight to each variable was considered to strengthen the similarity in some variables rather than others. It is done based on the impact variables have on the crop, as not all variables affect it in the same proportion. For example, two problems that only differ in the shade of the crop cannot have the same similarity as two problems that only differ in the climatic season. In other words, the season has a more significant impact in considering those two problems differently. In summary of this step in CBR, a new entry is entered into the system, and the k most similar cases are obtained based on the Euclidean distance calculated between the incoming problem and each case stored in the case base.
- Reuse: With the identification of the k most similar cases, the solutions of those cases were reviewed to determine or adapt a new solution for the new problem. In CBR, Gabel (2010) observes different methods to find a new solution based on rules, conditions, formulas, expert guidance, and constraints. Additionally, one of the methods to determine a new solution is the regression technique, which was used to predict the fertilizer rates for the three studied macronutrients. Therefore, the average fertilizer rates from the most similar cases found were used to determine the fertilizer rates for the new problem.
- Recommendations: As explained earlier, the first three variables of a case's solution indicate the rates (low, medium, high, or very high) of NPK fertilizer to apply based on the general soil condition. After obtaining the predictions, these rates are converted into fertilizer amounts expressed in kilograms per hectare (kg/ha) per year. The recommended fertilizer rates were obtained through consultations with experts and studies from Cenicafé according to Sadeghian (2008) and FNC (2013). These studies determined the maximum amounts of N, P, and K fertilizer required for a coffee crop in Colombia, based on a case with high density and low shade.

  This study determined the fertilizer quantities for various soil conditions that can occur in a case. Additionally, these fertilizer quantities are scaled according to the crop planting density. Recommending a medium rate, for example, in a density of 3,000 plants per hectare, is different from making the same recommendation in a density of 7,000 plants per hectare. Therefore, it is essential to define three high rates for low, medium, and high planting densities and three medium rates for low, medium, and high densities in a periodical way. Table 2 presents the determined quantities for each studied nutrient, which domain experts validated.

3.6 Phase 5: evaluation

For this phase, the evaluation of the RS depends directly on the perspective of the application domain, in this case, agriculture; therefore, it is imperative to validate the findings in consultation with experts in the domain. This validation process would help alleviate the uncertainty of the developed model for future applications. Six case studies (randomly chosen from the case base) were addressed, covering information related to crops, soils, and climatic conditions. Table 3 shows the information for each case and the respective recommendation provided by the system, considering the units of measurement of each variable.

TABLE 2 Amount of fertilizer to recommend according to the crop's soil condition and planting density.

Nutrient	Fertilizer rate	Fertilizer quantity (kg/ha per year)		
		High density	Medium density	Low density
Nitrogen	Low	150	180	210
	Medium	180	210	240
	High	210	240	270
	Very high	240	270	300
Phosphorus	Low	30	36	42
	Medium	36	42	48
	High	42	48	54
	Very high	48	54	60
Potassium	Low	135	165	195
	Medium	165	195	225
	High	195	225	255
	Very high	225	255	295

These six cases were obtained from data collected by sensors on several farms in the Department of Cauca and information shared by Ecotecma experts in their investigations. In developing our RS, we have reached a level of technological maturity 3, indicating that our project has passed the theoretical phase and created a functional prototype, according to National Aeronautics and Space Administration (2017). This prototype has been designed to demonstrate the viability of the underlying RS concept; using the data and scenarios provided in the case above. While this prototype is functional, it is essential to note that it is designed for proof of concept and initial experimentation, rather than for large-scale implementation or commercial use. Future work will focus on advancing this prototype to higher TRL levels and improving the CBR KB, aiming to create a scalable recommender system that can be effectively adapted to various agricultural conditions and requirements.

The system evaluation involved the collaboration of six domain experts whose credentials and specific areas of experience provide significant validity to their assessments. These include:

- Expert 1: A Biologist with a Ph.D. in Environmental Sciences and postdoctoral experience in soil management in coffee agroecosystems. Over 7 years of experience in science, technology, and innovation in Colombia's agricultural sector and natural resource conservation.
- Expert 2: Agricultural Engineer, Master in Agroecology, PhD in Environment and Society, professor at the University of Cauca, and coordinator of the Agroecology and Territory component of the Center for Innovation and Social Appropriation of Coffee Growing.
- Expert 3: Agricultural Engineer. Agronomist at a coffee development company in the region.
- Expert 4: Agronomist Engineer. Agronomic Advisor and Researcher for the Cauca Soils Project at the Government of Cauca.
- Expert 5: Environmental engineer. Coordinator in environmental management activities in Cauca. Doctoral student in Telematics Engineering in the research field of agriculture.
- Expert 6: Business administrator and farmer.

The RS was evaluated using a structured survey. The process consisted of open recommendations for the first three cases and multiple-choice evaluations for the final three cases.

In the first part of the survey, experts were provided with information on three cases detailing crop, soil, and climate conditions. Based on their expert knowledge, they were asked to recommend the amounts of N, P, and K fertilizers to be applied per hectare per year. The system's recommendations were compared with these open responses, and subsequently, for the following three cases, the experts rated the system's suggestions using a modified Likert Scale (LS) (1 = Inadequate, 2 = Unsatisfactory, 3 = Acceptable, 4 = Effective, and 5 = Optimal).

Table 4 shows the recommendations given by the experts for the first three cases. As well as the recommendations generated by the RS. Likewise, Table 5 shows the expert's ratings for the following three cases, where each one made a general rating for the three recommendations (N, P and K) generated by the RS, based on the given options.

To analyze the similarity between the CBR system's recommendations and those provided by the experts, the Pearson correlation coefficient (PCC) was calculated for each case and nutrient. This coefficient measures the strength of the linear relationship between the two variables, in this case, the fertilization recommendations by the experts and those generated by the system. The coefficient results are shown at the end of Table 4, where the variability between the values is observed.

N shows a moderate correlation in the three cases, with values ranging between 0.578 and 0.665. A high correlation was observed for P in Case 1 (0.894). Still, in Case 2 the correlation was negative (−0.178), indicating significant discrepancies between the system's recommendations and those of the experts in that context. Finally, very high correlations were found for K, with values close to 1 in all three cases, suggesting a solid alignment.

To analyze the consistency and reliability of the expert's assessments, we calculated the Intraclass Correlation Coefficient (ICCa) using Python and the Pingouin library. The ICCa assesses agreement by analyzing the variance between the rater's ratings relative to the total variance. An ICCa value close to 1 indicates high consistency or agreement between raters, while a value close to 0 or negative reflects a lack of significant agreement. This metric is beneficial when assessing the reliability of assessments provided by different people using a similar scale.

For the first three cases, three coefficients were calculated based on the three recommendations given by each expert for each case. The results, divided by nutrient (N, P, and K), are presented below:

- ICCa for N: The ICCa for the expert's N recommendations was 0.266, indicating low agreement between experts. This

TABLE 3 System recommendations based on case information.

Variable	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Planting density (No. of plants per ha)	5,500	4,500	5,700	5,800	6,700	7,800
Shade coverage (%)	30	30	38	20	50	30
Climatic season	Rainy	Dry	Rainy	Rainy	Normal	Dry
Moisture soil (%)	50	30	75	60	35	15
N level soil (mg/kg)	40	60	45	60	95	55
P level soil (mg/kg)	12	6	8	42	80	60
K level soil (mg/kg)	196	100	125	216	160	93
pH soil	5.2	4.5	5	6.1	5.1	6.9
N fertilizer recommendation (kg per ha per year)	214.29	188.57	235.51	222.86	227.14	270
P fertilizer recommendation (kg per ha per year)	47.14	42	50.57	41.14	44.57	54
K fertilizer recommendation (kg per ha per year)	203.57	173.57	225	177.86	220.71	242.15

TABLE 4 Expert recommendations for cases 1, 2, and 3.

Subject	Case 1			Case 2			Case 3		
	Recommendations per hectare per year								
	N	P	K	N	P	K	N	P	K
Expert 1	275.75	47.5	-	246	42.5	-	246	42.5	-
Expert 2	285	38	171	221	51	221	238	51	221
Expert 3	250	45	230	245	443	225	260	51	240
Expert 4	289.5	15.05	180	275	25.07	224.1	298.5	25.07	211.65
Expert 5	280	99	297	229	81	243	171	102	307
Expert 6	266	38	171	221	51	221	238	51	221
CBR system	231.42	44	177.86	217.14	44.57	202.14	244.28	51.42	216.43
PCC	0.578	0.894	0.987	0.665	-0.178	0.939	0.612	0.431	0.883

TABLE 5 Expert assessment for cases 4, 5, and 6.

Subject	Case 4		Case 5		Case 6	
	LS	Value	LS	Value	LS	Value
Expert 1	Acceptable	3	Unsatisfactory	2	Unsatisfactory	2
Expert 2	Acceptable	3	Acceptable	3	Acceptable	3
Expert 3	Inadequate	1	Inadequate	1	Unsatisfactory	2
Expert 4	Acceptable	3	Effective	4	Effective	4
Expert 5	Unsatisfactory	2	Inadequate	1	Unsatisfactory	2
Expert 6	Effective	4	Effective	4	Effective	4

suggests a significant discrepancy in how each expert assessed the nitrogen recommendations.

- ICCa for P: The ICCa for P was very low (−0.003), indicating no consistent agreement between experts. This lack of agreement highlights the challenges in determining phosphorus levels, which are often more context-specific and sensitive to local soil chemistry.
- ICCa for K: The ICCa for potassium was also negative (−0.17), again indicating poor agreement. This could be due to expert's different approaches to addressing potassium levels under various conditions.
- The average ICCa for the three nutrients combined was also low, reflecting the overall discrepancy in expert opinions. These findings indicate that while experts provided valuable information, their recommendations often needed to be more consistent, likely due to the high variability and complexity of coffee fertilization practices. These discrepancies underline the need to incorporate more local

environmental data into the FRS to better align with expert knowledge.

Furthermore, for the final three cases, we calculated the ICCa for these categorical ratings, indicating poor agreement between experts (−0.03). The lack of consensus may indicate the subjective nature of these assessments, where factors such as individual experience, specific field conditions, and interpretation of the data provided may generate such discrepancy in the formation of expert opinions. The Notebooks used for the calculation of the ICCa, as well as the surveys and the expert's responses, are included in the supplementary data attached to this work, allowing the reproduction and verification of the results obtained.

Finally, by averaging the PCC values between the system's recommendations and those provided by the experts for the first three cases (1, 2, and 3), an average value of 0.646 was obtained. This result indicates a moderate agreement between the system's recommendations and the experts. On the other hand, for Cases 4, 5, and 6, in which the experts evaluated the system's recommendations using a Likert scale, the average of the ratings was 2.66, which falls between the Regular and Acceptable categories, with a slight inclination toward the Acceptable category. This result suggests that, although the system provides recommendations that are mostly seen as adequate, there are still areas for improvement. The PCC analysis and the use of LS reinforce the idea that the CBR system has good potential for fertilizer recommendation in coffee crops but still requires adjustments and greater incorporation of local contextual data to achieve greater alignment with the recommendations. These results provide a solid foundation for continued development and refinement of the system to improve its accuracy and applicability in coffee agriculture.

4 Discussion

The evaluation of the CBR system revealed essential insights into its performance in providing fertilization recommendations for coffee crops. Despite the system showing moderate to high correlation for some nutrients—particularly potassium—discrepancies between the system's outputs and expert recommendations highlight areas where improvement is needed. For instance, nitrogen and phosphorus showed variability across different cases, with a negative correlation for phosphorus in one of the evaluated cases. These findings suggest that the promising system still requires refinement to better capture the nuances of fertilization practices, particularly in regions like Cauca, where environmental factors and soil characteristics vary significantly.

The ICCa further highlighted the inconsistencies among expert evaluations, particularly for nitrogen and phosphorus. Low agreement among experts, as reflected by negative or near-zero ICCa values, suggests that the complexity of coffee fertilization may lead to divergent opinions depending on individual experience and local knowledge. This is consistent with previous studies showing similar challenges in developing uniform fertilizer recommendations across diverse agricultural contexts. For instance, Kumar et al. (2019) and Suchithra and Pai (2018) emphasize integrating local environmental data, such as soil pH, organic matter content, and crop-specific conditions, into intelligent systems to improve recommendation accuracy.

These results reinforce the need for further development of the CBR system. Incorporating additional variables such as organic matter, crop age, the nutrients exported by the future harvest, and more detailed local soil and climate data could improve the alignment between the system's recommendations and expert opinions. Similar improvements have been suggested in works like Wickramasinghe et al. (2019), where sensor-based data collection has been shown to enhance the precision of FRS.

Considering these findings, future iterations of the system should focus on increasing its adaptability to different environmental conditions. This could involve integrating real-time sensor data and expanding the knowledge base with region-specific agricultural data, similar to the approaches seen in McFadden et al. (2018) and Ren and Lu (2012). By doing so, the system can move toward providing more contextually relevant recommendations that better align with expert knowledge while maintaining flexibility across various regions and agricultural practices.

5 Conclusions

Recommender systems have provided multiple insights into various crops, allowing farmers to improve production, mitigate risks such as diseases and pests, improve decision-making in various agricultural practices, and even reduce associated environmental impacts. However, their implementation in real-world environments must be enhanced by technological limitations in capturing the data necessary for these systems to function effectively, especially in the Colombian region. In this sense, this research proposed an RS that, unlike existing works, is based on expert knowledge obtained through interviews with domain experts and scientific research from Colombian private institutions related to coffee cultivation. The approach's recommendations are based solely on current crop status and climatic conditions; rather than historical soil information or crop production records. Consequently, the implemented system successfully addressed the problem of the scarcity of data needed to generate recommendations. It was demonstrated through evaluation that the results obtained were close to the expert's suggestions, but there were many corrections regarding more information to be analyzed. To increase the case base and the agroclimatic variables studied. Considering that the tests carried out reached the laboratory level, it is an initial prototype that can receive many improvements in the future.

Regarding implementing the system, it was determined that leveraging expert knowledge in agriculture is essential for crops with limited data availability, especially when it is challenging to access historical data on crucial parameters such as climate, crop management, and soil. The CBR, which has proven effective in other application domains, demonstrated in this study that its application in agriculture, together with the participation and collaboration of experts, can contribute to supporting the sustainability of small farmers.

Furthermore, the knowledge base built can be important for future research in coffee cultivation, as it establishes a mechanism to automatically identify relevant variables in coffee cultivation by analyzing the importance and meaning of soil, crop, and climate variables. This allows us to determine which variables are more appropriate or have greater weight than others.

6 Future works

Considering the research opportunities that arise with the development of this research project, the following future work is proposed.

The current FRS provides suggestions solely for the quantity and frequency of fertilizer application. However, an essential aspect of fertilization management that remains to be explored is the type of fertilizer. Future work could focus on refining the recommendations by considering the types of fertilizers available in the market, which vary in cost and nutrient composition. In this sense, a key enhancement would be the ability of the system to recommend specific fertilizer formulations based on the crop's nutrient requirements at different growth stages. For example, based on the results of this study, the IoT system could recommend an initial application of a complete NPK fertilizer that covers the phosphorus requirement at flowering, followed by a second application of urea + KCl to meet the nitrogen and potassium needs during the fruit growth and filling stages.

In addition, developing a dashboard interface for farmers and domain experts to provide more detailed information beyond fertilizer quantities is proposed as a next step. This interface could offer early alerts related to fertilization, including analysis of seasonal changes and their impact on nutrient absorption and loss. It could also integrate soil condition monitoring, helping farmers optimize fertilizer application timing and effectiveness.

Although this study demonstrated that the system's recommendations were in line with those provided by experts, more research is needed to assess the long-term impact of these recommendations on coffee production. Future evaluations should be conducted over multiple years, tracking fertilization events, coffee yield, and quality. To that end, the recommender system could be complemented with automated data collection modules that monitor fertilization events and production outcomes, allowing feedback to be incorporated into the CBR system and improving its accuracy over time.

Finally, it is essential to mention that while this system was developed for coffee cultivation, the architecture and approach can be adapted for other crops. This system could be extended by modifying the case base to consider crop-specific characteristics to optimize fertilization practices for various agricultural contexts.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

EL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JFCO: Conceptualization, Resources, Supervision, Validation, Visualization, Writing – review & editing. JCC: Conceptualization, Funding acquisition, Methodology, Project administration, Software, Supervision, Writing – review & editing. CF: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is within the framework of the research project: Incremento de la Oferta de Prototipos Tecnológicos en Estado Pre-Comercial Derivados de Resultados de I+D Para el Fortalecimiento del Sector Agropecuario en el Departamento del Cauca (Grant Number: BPIN 2020000100098), funded by the Sistema General de Regalías (SGR) de Colombia under the Convocatoria de Formación de Capital Humano de Alto Nivel Para Promover el Liderazgo Social y comunitario y Afrontar el 4RI en el Departamento del Cauca (Grant Number: BPIN 202100010029). Additionally, this work also is funded by the Telematics Engineering PhD program from the University of Cauca.

Acknowledgments

The authors are grateful to the Telematics Engineering Group (GIT) of the University of Cauca, and the Ministry of Science Technology and Innovation of Colombia (Minciencias) for the Ph. D. support granted to Edinson David Leon Chilito.

Conflict of interest

JFCO was employed by Ecotecma S.A.S.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Acuña, D. G. H. (2019). *Sistema de recomendación de libros basado en ontologías asociadas a tesauros: El caso de la biblioteca de la facultad de ingeniería de sistemas e informática de la universidad nacional mayor de san marcos* (Bachelor's thesis). Universidad Nacional Mayor de San Marcos.

Arcila P., J., Farfán V., F., Moreno B., A., Salazar G., L. F., and Hincapié, G. E. (2007). "Densidad de siembra y productividad de los cafetales," in *Sistemas de producción de café en Colombia, Volume 1 of* Fitotecnia, Prácticas de cultivo, Caficultura, 1 Edn. (Chinchiná, Caldas: Editorial Blanecolor Ltda.), 131–144.

Bilen, C., El Chami, D., Mereu, V., Trabucco, A., Marras, S., and Spano, D. (2023). A systematic review on the impacts of climate change on coffee agrosystems. *Plants* 12:10102. doi: 10.3390/plants12010102

Boregowda, N., Jogigowda, S. C., Bhavya, G., Sunilkumar, C. R., Geetha, N., Udikeri, S. S., et al. (2022). Recent advances in nanoremediation: carving sustainable solution to clean-up polluted agriculture soils. *Environ. Pollut*. 297:118728. doi: 10.1016/j.envpol.2021.118728

Chaves, P. A. (2016). Grado de aceptación e implementación de las tic por campesinos agrícolas con vocación cafetera en el municipio del tambo-cauca (Master's thesis). Universidad Nacional Abierta y a Distancia UNAD, Popayán, Colombia.

Cholissodin, I., Dewi, C., and Eunike Endariahna Surbakti (2016). "Integrated ANN and bidirectional improved PSO for optimization of fertilizer dose on Palawija plants," in *2nd International Conference on Science in Information Technology (ICSITech)* (Balikpapan: IEEE), 193–197.

Chougule, A., Jha, V. K., and Mukhopadhyay, D. (2019). "Crop suitability and fertilizers recommendation using data mining techniques," in *Progress in Advanced Computing and Intelligent Engineering, Vol. 714*, eds. C. R. Panigrahi, A. K. Pujari, S. Misra, B. Pati, and K.-C. Li (Singapore: Springer Singapore), 205–213.

Drechsel, P., Marjani Zadeh, S., and Salcedo, F. (2023). *Water Quality in Agriculture: Risks and Risk mitigation*. Rome: FAO.

Farfán V. F., and Mestre M. A. (2004). Manejo del sombrío y fertilización del café en la zona central colombiana. *Cenicafé*, 330:1–8. doi: 10.38141/10779/0330

FNC (2013). *Manual del cafetero colombiano, Vol. 2, 1 Edn*. Chinchiná, Caldas: LEGIS.

Gabel, T. (2010). *Problem Solving By Case-Based Reasoning*. Friburg: Machine Learning Lab—Friburg University.

Gast, F., Benavides M. P., Sanz U. J. R., Herrera P. J. C., Ramírez Builes, V. H., Cristancho A. M., et a;. (2013). "Gestión del riesgo agroclimático. Vulnerabilidad y capacidad de adaptación del sistema de producción de café," in *Investigación y tecnología para la sostenibilidad en la agricultura, Vol. 1 of* Manual Del Cafetero Colombiano (Chinchiná: LEGIS), 91–114.

Hossain, M., and Siddique, M. N. A. (2020). Online fertilizer recommendation system (OFRS): a step towards precision agriculture and optimized fertilizer usage by smallholder farmers in Bangladesh. *Eur. J. Environ. Earth Sci.* 1, 1–9. doi: 10.24018/ejgeo.2020.1.4.47

Howland, F. C., Muñoz, L. A., Staiger, S., Cock, J., and Alvarez, S. (2015). Data sharing and use of ICTS in agriculture: working with small farmer groups in Colombia. *Knowl. Manag. Dev. J.* 11, 44–63.

Islam, T., Uddin, R., Rizan, R., Tusher, Y., Hossain, M. A., and Galib, S. (2020). Nitrogen fertilizer recommendation for paddies through automating the leaf color chart (LCC). *Int. J. Adv. Comput. Sci. Appl.* 11, 745–752. doi: 10.14569/IJACSA.2020.0110891

Jiang, W., Xing, Y., Wang, X., Liu, X., and Cui, Z. (2020). Developing a sustainable management strategy for quantitative estimation of optimum nitrogen fertilizer recommendation rates for maize in Northeast China. *Sustainability* 12:2607. doi: 10.3390/su12072607

Kitchenham, B. A., and Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001*. Keele: Keele University and Durham University Joint Report.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artif. Intell. Rev*. 6, 3–34.

Kumar, A., Kumar, A., De, A., Shekhar, S., and Singh, R. K. (2019). IoT based farming recommendation system using soil nutrient and environmental condition detection. *Int. J. Innov. Technol. Explor. Eng.* 8, 3055–3060. doi: 10.35940/ijitee.K2335.0981119

Lenka, S., S, R., Coumar, V., Dotaniya, M., and Saha, J. (2016). *Impact of Fertilizers Use on Environmental Quality*. (Kalyani: Bidhan Chandra Krishi Viswavidyalaya), 11.

León-Burgos, A. F., Unigarro, C., and Balaguera-López, H. E. (2022). Can prolonged conditions of water deficit alter photosynthetic performance and water relations of coffee plants in central-west Colombia? *South Afr. J. Bot.* 149, 366–375. doi: 10.1016/j.sajb.2022.06.034

Martín Alonso, G. M., Pérez Díaz, A., Espinosa, R. R., González, C. B., Viñals Núñez, R., and Nualles, M. V. (2016). Comparison of two models of response to nitrogen doses in corn and coffee. *Cult. Trop*. 37, 155–164.

McFadden, B. R., Brorsen, B. W., and Raun, W. R. (2018). Nitrogen fertilizer recommendations based on plant sensing and Bayesian updating. *Precision Agri*. 19, 79–92. doi: 10.1007/s11119-017-9499-4

Miller, M. (2019). *The Basics: KNN for Classification and Regression*. Towards Data Science.

National Aeronautics and Space Administration (2017). *Technology Readiness Level Definitions*.

National Weather Service (2023). *NOAA's Climate Prediction Center*. College Park, MD.

Pozas, J. S. (2008). Fertilidad del suelo y nutrición del café en Colombia—Guia práctica. *Cenicafé* 32:45.

Puntel, L. A., Pagani, A., and Archontoulis, S. V. (2019). Development of a nitrogen recommendation tool for corn considering static and dynamic variables. *Eur. J. Agron*. 105, 189–199. doi: 10.1016/j.eja.2019.01.003

Qin, Z., Myers, D. B., Ransom, C. J., Kitchen, N. R., Liang, S., Camberato, J. J., et al. (2018). Application of machine learning methodologies for predicting corn economic optimal nitrogen rate. *Agron. J.* 110, 2596–2607. doi: 10.2134/agronj2018.03.0222

Ren, Z., and Lu, X. (2012). "Design of fertilization recommendation knowledge base and appllication," in *First International Conference on Agro- Geoinformatics (Agro-Geoinformatics)* (Shanghai: IEEE), 203–207.

Rodríguez-Espinosa, T., Papamichael, I., Voukkali, I., Gimeno, A. P., Candel, M. B. A., Navarro-Pedreño, J., et al. (2023). Nitrogen management in farming systems under the use of agricultural wastes and circular economy. *Sci. Tot. Environ*. 876:162666. doi: 10.1016/j.scitotenv.2023.162666

Sadeghian Khalajabadi, S. (2016). La acidez del suelo, una limitante común para la producción de café. *Cenicafé* 466, 1–12.

Sadeghian Khalajabadi, S. (2017). Síntomas visuales de deficiencias nutricionales en café. *Cenicafé* 478:12. doi: 10.38141/10779/0478

Sadeghian Khalajabadi, S., Osorio, H. G., and Suárez, E. A. (2015). Lixiviación de nutrientes en suelos de zona cafetera. *Prácticas que ayudan a reducirla*. 40, 1–36.

Sadeghian, S. (2008). Fertilidad del suelo y nutrición del café en Colombia : Guía práctica. *Boletín Técnico Cenicafé* 32, 1–44.

Sainju, U. M. (2017). Determination of nitrogen balance in agroecosystems. *MethodsX* 4, 199–208. doi: 10.1016/j.mex.2017.06.001

Sànchez-Marrè, M. (2001). *Principles of Case-Based Reasoning*.

Shang, Y. (2005). "5—expert systems," in *The Electrical Engineering Handbook*, ed. W.-K. Chen (Burlington: Academic Press), 367–377.

Suchithra, M. S., and Pai, M. L. (2018). "Improving the performance of sigmoid Kernels in multiclass SVM using optimization techniques for agricultural fertilizer recommendation system," in *Soft Computing Systems, Vol. 837*, eds. I. Zelinka, R. Senkerik, G. Panda, and P. S. Lekshmi Kanthan (Singapore: Springer), 857–868.

Sujithra, T., Thanjaivadivel, M., and Durai, S. (2019). Fertilizer recommendation system for coconut cultivation. *Int. J. Civ. Eng. Technol*. 8, 813–819.

Sumaryanti, L., Lamalewa, L., and Istanto, T. (2019). Implementation of fuzzy multiple criteria decision making for recommendation paddy fertilizer. *Int. J. Mech. Eng. Technol.* 10, 236–243.

Sylvester, J., Valencia, J., Verchot, L. V., Chirinda, N., Romero Sanchez, M. A., Quintero, M., et al. (2020). A rapid approach for informing the prioritization of degraded agricultural lands for ecological recovery: a case study for Colombia. *J. Nat. Conserv*. 58:125921. doi: 10.1016/j.jnc.2020.125921

Valdivielso, A. (2020). *Qué es la escorrentía*. iAgua.

Vélez-Vallejo, R. (2022). *Informe del gerente—90 congreso nacional de cafeteros. Technical Report*. Bogotá: Federación Nacional de Cafeteros de Colombia.

Vieira Fontoura, S. M., Viero, F., Moraes, R. P. D., and Bayer, C. (2017). Nitrogen fertilization of no-tillage winter cereals in the South-Central Region of Paraná, Brazil. *Revista Brasileira de Ciência do Solo* 41, 1–16. doi: 10.1590/18069657rbcs20170009

Wickramasinghe, C., Lakshitha, P., Hemapriya, H., Jayakody, A., and Ranasinghe, P. (2019). "Smart crop and fertilizer prediction system," in *2019 International Conference on Advancements in Computing (ICAC)* (Malabe: IEEE), 487–492.

World Bank (2022). *Agriculture (% of GDP)—Colombia | Data*. World Bank Group.

Zhang, H., Zhang, L., Ren, Y., Zhang, J., Xu, X., Ma, X., et al. (2011). "Design and implementation of crop recommendation fertilization decision system based on WEBGIS at village scale," in *Computer and Computing Technologies in Agriculture IV, volume 345*, eds. D. Li, Y. Liu, and Y. Chen (Berlin, Heidelberg: Springer), 357–364.