Precision farming using autonomous data analysis cycles for integrated cotton management

Raul Toscano-Miranda[a,c], Jose Aguilar[b,c,e], Manuel Caro[a], Anibal Trebilcok [d], Mauricio Toro[c]

[a]Department of Educational Informatics, Universidad de Córdoba, Colombia
[b]CEMISID, Universidad de Los Andes, Mérida, Venezuela
[c]GICOMP, Universidad EAFIT, Medellín, Colombia
[d]Department of Agronomic Engineering, Universidad de Córdoba, Colombia
[e]IMDEA Networks Institute, Leganés, Madrid, Spain

Abstract

Precision farming (PF) allows the efficient use of resources such as water, and fertilizers, among others; as well, it helps to analyze the behavior of insect pests, in order to increase production and decrease the cost of crop management. This paper introduces an innovative approach to integrated cotton management, involving the implementation of an Autonomous Cycle of Data Analysis Tasks (ACODAT). The proposed autonomous cycle is composed of a classification task of the population of pests (boll weevil) (based on eXtreme Gradient Boosting-XGBoost), a diagnosis-prediction task of cotton yield (based on a fuzzy system), and a prescription task of strategies for the adequate management of the crop (based on genetic algorithms). The proposed system can evaluate several variables according to the conditions of the crop, and recommend the best strategy for increasing the cotton yield. In particular, the classification task has an accuracy of 88%, the diagnosis/prediction task obtained an accuracy of 98%, and the genetic algorithm recommends the best strategy for the context analyzed. Focused on integrated cotton management, our system offers flexibility and adaptability, which facilitates the incorporation of new tasks.

Keywords: Precision Farming, Artificial Intelligence, Data Analysis. Autonomous Systems, Integrated Cotton Management.

1. Introduction

Precision Farming (PF) involves technologies for data collection, data analysis, and decision-making (Say et al., 2018). Data collection technologies, such as sensors, are used to understand the environment (Cui et al., 2022). Data processing technologies use data models for interpretation tasks (Kong et al., 2019). Decision-making technologies also use data models and actuators for planning tasks and changing the environment (Singh & Sharma, 2022).

On the other hand, there is a need to improve cotton production (Ghaffar et al., 2020) and PF technologies can help with this task (Coulibaly et al., 2022). According to Ghaffar et al. (2020), there is a great challenge in the management of cotton cultivation in which factors such as proper management of nutrients, pests, diseases, irrigation, etc. play an important role. In this paper, a PF approach based on autonomous data analysis cycles for integrated cotton management has been used.

1.1. Related works

Several studies have investigated integrated management approaches based on PF. For example, Tribouillois et al. (2022) built an integrated model for crop and water management to optimize irrigation. They used a combination of techniques to reduce water usage while also diversifying the types of crops grown in irrigated watersheds. Hajimirzajan et al., (2021) proposed a large-scale crop planning, which involves a comprehensive strategic framework that employs a decision support system to determine the sustainable use of water, as well as optimal crop selection, timing, and cultivation practices. Aggarwal and colleagues (2022) developed a geospatial analysis system to preserve land fertility, optimize agricultural revenue, and minimize agricultural pollution and water consumption. The system allows land use planning with rotating crops. Wu et al. (2020) developed a model for integrated nutrient management that included four factors: chemical fertilizers, domestic livestock manure, large-scale livestock manure, and cultivated area. The authors found that there is a need to improve integrated nutrient management, expand livestock manure, and control cultivated areas of certain crops.

Diagnostic tasks in agriculture have helped the early diagnosis of crop diseases. For example, Masood et al. (2020) used a Convolutional Neural Network CNN) model to diagnose rice crop regions affected by the disease. The results showed that their proposal outperforms the standard CNN model in terms of recall, precision, F1, and accuracy score. Suleiman (2019) developed an expert system that can identify and diagnose safflower diseases like Cercospora leaf spot, powdery mildew, head rot, and wilt, among others. The expert system provides information on the symptoms, propagation, and survival of each disease. Several studies have focused on the classification tasks of agricultural pests. For example, the identification of Helicoverpa armigera by Kandalkar et al. (2014) involved image segmentation using a saliency map, feature extraction via the discrete wavelet transform, and pest classification through the use of a back-propagation neural-network.

In prediction tasks of crop yield, there are some studies such as the following. Maskey et al. (2019) investigated the correlation between weather parameters and strawberry yield. They used principal component regression, single-layer neural network, and random forest to forecast yield, analyzing various weather conditions. Ali et al. (2018) suggested a hybrid genetic programming model with an integrated Markov Chain Monte Carlo, utilizing climate data such as humidity, rainfall, and temperature. Similarly, Lobell and colleagues (2013) utilized non-linear regression to predict maize yield and demonstrated a notable negative response to temperatures exceeding 30 ºC, and a better response to seasonal rainy seasons.

On the other hand, various technologies have been developed to aid cotton farmers in making decisions about irrigation, fertilization, pest control, and other practices. One of these tools is the use of expert systems, which are computer programs that perform at the level of human experts. For example, COMAX is an expert system that acts as an expert in cotton crop management and determines the best strategy for irrigating, applying fertilizer, and applying defoliants and cotton boll openers (Lemmon, 1986).

Another tool employed is crop simulation modeling, which is a system that simulates the growth and development of cotton plants under different environmental and management conditions. Cotton crop simulation models are mathematical models that can be used to predict the growth, development, and yield of cotton crops. Hearn (1994) developed a cotton crop simulation model, which was validated against six data sets from agronomic experiments. The model demonstrated sensitivity to climatic and agronomic variables, such as irrigation regime, nitrogen fertilizer rate, and sowing date. The Cropping System Model (CSM)-CROPGRO-Cotton model is another cotton crop simulation model that was developed by Pathak et al. (2012). The CSM-CROPGRO-Cotton model is more complex than Hearn's model, and it requires many parameters and inputs. However, the CSM-CROPGRO-Cotton model can be used to predict a wider range of cotton crop traits, such as leaf area index, leaf weight, stem weight, and boll weight.

A third technology is the use of decision support systems, which are computer programs that help users make choices among alternatives based on their values and preferences. Jones & Barnes (2000) proposed a decision support system that allows users to express individual or corporate values and preferences; considers the degree of imprecision associated with each input; reduces several levels of complex information into a single chart; and allows examination of trade-off between alternatives and interests. This decision support system also uses remote sensing data to describe spatial variability in terms that can be related to a crop model, making the decision-making approach feasible for PF applications. The crop model provides information that can be used by the decision support system, and the remote sensing data is used to fine tune the calibration of the crop model, maximizing the accuracy of its results.

Some of the above articles propose expert systems, cotton crop simulation models and decision support systems. Others propose specific diagnostic and prediction tasks, for example, of crop behavior. That work shows that it is possible to develop systems that can help farmers make better decisions about crop management, which will lead to better yields and profits. Our ACODAT system integrates multiple tasks for analyzing the behavior of cotton cultivation in order to make recommendations; to our knowledge, it is the first work with these characteristics.

1.2. Our contribution

To the best of our knowledge, there are no studies that a) implement an autonomous system using ACODAT for integrated cotton management; b) prescribe strategies for the adequate management of the crop; c) integrate different models of knowledge (classification, diagnosis/prediction and prescription) for the management of crops (Toscano-Miranda et al., 2022a; Toscano-Miranda, et al., 2022b); d) concurrently employ several types of variables (fertilizers, climate, the behavior of pests, etc.); and e) and utilize uncertainty models for the prediction/diagnosis of crop yields. These gaps constitute the focus of our study.

Particularly, in this paper, we focus on PF based on the Autonomous Cycle of Data Analysis Tasks (ACODAT) for integrated cotton management. We use ACODAT, which has two advantages. First, ACODAT allows automating the entire process, the phases of monitoring, analysis and decision making. Second, it does so from the process data. According to Sanchez et al. (2016), ACODAT makes use of diverse succeeding data analysis tasks interacting with one another to obtain the necessary knowledge to introduce process improvements. ACODAT has been utilized in various fields, including telecommunications, smart cities, industry 4.0, education, and medicine, as evidenced by different works (Aguilar et al., 2008; Aguilar et al., 2020a; Morales et al., 2019; Sánchez et al., 2020). Morales et al. (2019) focused on the telecommunications sector, where they developed an ACODAT to manage the quality of service in Internet of Things (IoT) platforms, utilizing classification and clustering tasks. It has been employed in smart cities for the purpose of regulating and monitoring heating, ventilation, and air conditioning systems (Aguilar et al., 2020a). The efficiency of production processes in Industry 4.0 has been enhanced through the use of ACODAT. For instance, Sánchez et al. (2020) introduced an architecture that resolves the issues of heterogeneity and actor integration in manufacturing processes. The outcomes demonstrated that ACODAT facilitated interaction among actors such as things, data, people, and services, resulting in the definition of a self-optimization and self-configuration plan. In the educational domain, ACODAT has been implemented to identify learning styles in smart classrooms, demonstrating its usefulness. Monsalve, et al. (2020) utilized ACODAT to study social network and web data, creating knowledge models about students to facilitate ongoing monitoring of their learning process. The findings underscored ACODAT's capacity to generate practical knowledge that can improve the learning experience, particularly in smart classrooms. Finally, the ACODAT approach has been used in the domain of medicine for clinical disease management (Hoyos et al., 2022).

This work aims to define an ACODAT for integrated cotton management. The contributions of this work are the following:

- The definition and implementation of an autonomous system based on ACODAT for integrated cotton management;
- A task of classifying the pest population (boll weevil) according to the level of attack on the cotton crop, based on the work (Toscano-Miranda et al., 2022a);
- An adaptive model for the management of uncertainty based on a fuzzy system (FS) for the prediction/ diagnosis of cotton yield;
- The simultaneous use of information on fertilizers and crop status, climatic variables and level of pest attack, for pest monitoring and control, which improves the prediction/diagnosis yield;
- A prescription task for the generation of strategies for the adequate management of the crop based on the previous tasks of the autonomous cycle.

The paper is structured in the following manner: Section 2 presents the theoretical framework of this paper. Section 3 outlines our integrated cotton management approach based on PF using ACODAT. Section 4 presents a case study to evaluate our proposal, and Section 5 describes the results. Finally, Section 6 shows the conclusions and highlights some of the future directions of this work.

2. Theoretical framework

This section presents concepts about PF for integrated production management, ACODAT, and the Methodology for Data Analytics based oN Organizational characterization through a user-centered design (MIDANO).

2.1. PF for integrated production management

PF aims to reduce costs, increase yield, using the right resources, being friendly to the environment. According to Gandonou (2005), PF is a set of technologies that help the farmer manage the agricultural process. In addition, it aids in production risk management (e.g., through the variable nutrient application), and reduces water consumption (e.g., through drip irrigation).

Say et al. (2018) grouped the PF technologies into three: a) Data collection technologies (e.g., soil sampling and mapping, yield monitoring and remote sensing); b) Data analysis technologies (e.g., geographic information system, economic analysis and modelling); c) and decision-making technologies (e.g., variable rate application, agricultural robots). Next, some examples:

a) Data collection technologies: These technologies detect insects and diseases in crops using field sensors, and remote sensors (Khattab et al., 2019; Lemmon, 1986; Toscano-Miranda, 2022b). In addition, using images for the same tasks (Alves et al., 2020; Caldeira et al., 2021).
b) Data analysis technologies: for predicting the behavior of insects (Hudgins et al., 2017; Toscano-Miranda et al., 2022a), crop growth (Pathak et al., 2012), and crop yield (Maskey et al. (2019), expert systems for decision-making about diseases in crops (Mansour & Abu-Naser, 2019), etc.
c) Decision-making technologies: Automated crop management and treatment using PF (Vulpi et al., 2022), such as irrigation control using robots (Agostini et al., 2017), and spray control for insects or diseases (Song et al., 2017). For this, it is useful the unmanned vehicles in rural farm areas (Mammarella et al., 2021; Saha et al., 2022), geospatial analysis to decision support (Aggarwal et al., 2022), proper use of fertilizers (Stevens et al., 1996), crop management (Hearn, 1994; Jones & Barnes, 2000), etc.

Our work integrates data collection, data analysis and decision-making technologies in an ACODAT.

2.2. ACODAT

Due to the significant increase in data generation, the development of new tools is essential to extract valuable knowledge. ACODAT is useful for this and is based on the autonomic computing paradigm. ACODAT involves a series of interconnected data analysis tasks that must be carried out in conjunction to achieve a desired objective within a given system or context. The tasks perform distinct roles within the cycle and interact with one another (Aguilar et al., 2018; Sanchez et al., 2016; Terán et al., 2017): they observe the process, analyze, and interpret events, and make appropriate decisions. The responsibility of observation tasks is to gather information and data about the environment or system, while analysis tasks interpret and diagnose the system using this data. Knowledge models are constructed to understand the cycle's behavior. Decision-making tasks, on the other hand, are responsible for improving the process by carrying out activities.

The autonomic computing paradigm is oriented to define autonomic characteristics in systems based on a smart control loop, known as MAPE+K (Monitor, Analyze, Plan, Execute, and Knowledge) (Aguilar et al., 2018; Sterritt et al., 2005). An ACODAT collects, filters, and processes data of the supervised problem (the letter M is for this monitoring task). Also, it analyzes/interprets complex situations and predicts forthcoming situations (the letter A is for this analyzing task). Additionally, it establishes the actions that must be carried out/scheduled to reach the system objectives (the letter P is for this planning task) and defines mechanisms to execute the plan (the letter E is for this last task). Because of this, the autonomous cycle requires managing a large amount of information. The letter K corresponds to the knowledge models (e.g., classification, diagnostic, prediction, and prescription models) within the autonomous cycle. The design of the autonomous cycle must include all these aspects to achieve the objectives that give a solution to the problem.

2.3. MIDANO

MIDANO is a methodology that allows gaining a deeper understanding of the data, which relies on organizational characterization as a key component to develop ACODATs (Aguilar et al., 2020b). Fig. 1 shows the three primary phases of MIDANO. The initial phase seeks to familiarize with the organization to define the goal of the data analysis system. The focus of this stage is to recognize and frame the solution to a problem, from the viewpoint of developing data analysis-based applications. Also, it defines the ACODAT for the solution of the problem. The responsibility of Phase 2 is to prepare and treat the data, following the ETL paradigm (Extraction, Transformation, Loading). Its primary goal is to produce high-quality data that can be used to build knowledge models and specify the multidimensional data model of ACODAT. In Phase 3, data analysis tasks are implemented to generate various knowledge models such as descriptive, predictive, classification, and prescriptive (Aguilar et al., 2020b).

Problem characterization and ACODAT definition were accomplished during the first phase of our work using MIDANO. The second phase, which involved data preparation and treatment, was incorporated into the ACODAT to enable real-time processing of data, and increase the autonomy of the process. Additionally, this phase identified the required data sources for ACODAT development. Our work provides a detailed explanation of how each MIDANO phase was applied to cotton crop management.

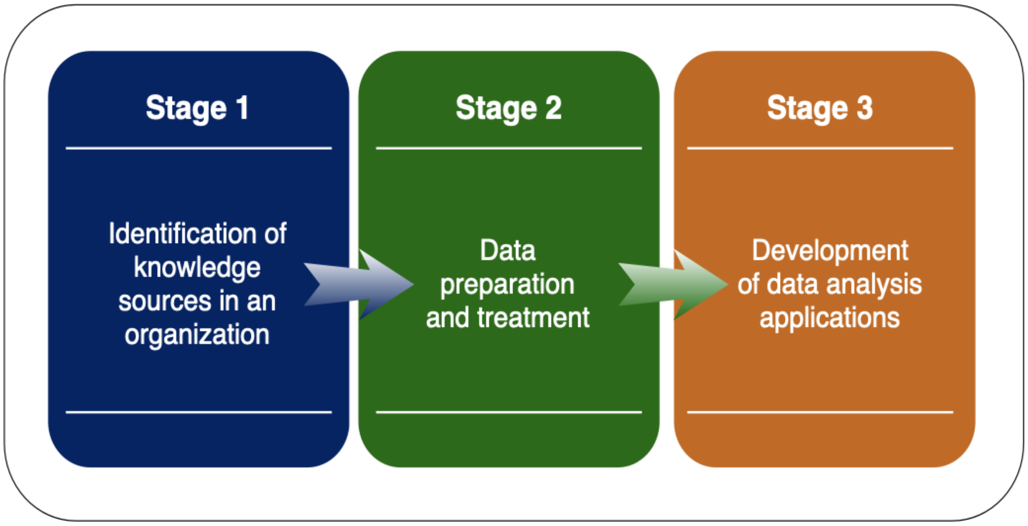


Fig. 1. MIDANO Methodology for Data Analysis from Organizational Characterization. Adapted from (Aguilar et al., 2020b).

3. ACODAT for the integrated management of production processes

This section outlines the process of creating an ACODAT for managing cotton crops. The tasks involved in ACODAT are described in detail. This section provides a general overview of the aspects necessary to implement our approach, which can be applied to other crops and pests. The specific variables are discussed in the Case study section, in which the application and validation of the proposed approach is demonstrated.

3.1. Characterization of the management of cotton crop

The main goal of cotton cultivation is to produce its valuable fiber (Trebilcok, 2020). There are several factors that influence production performance. For this reason, integrated crop management with the help of technologies seeks to improve yields with sustainable management and reduced environmental impacts (Abbas et al., 2020; Ghaffar et al., 2020). For example, if fertilizers are not applied in adequate quantities, then plant growth and development will be slowed, which will lead to lower yields (Ali et al., 2018; Ahmed, et al., 2020a; Ahmed, et al., 2020b). Cotton cultivation requires adequate nutrition, and its demand depends on various factors such as the stage of cultivation, genotype, and environment (Trebilcok, 2020). The water supply and the sowing date also affect yields and the overall growth of the plant (Ali et al., 2018). Regarding insect pests, it is recommended to control all types of cotton insect pests through integrated pest management techniques (Anees & Shad, 2020). Cotton production is also more vulnerable to climate change, which can have a negative impact on yields (Ahmad et al., 2020).

Thus, there is a great challenge in the management of cotton cultivation in which factors such as proper management of nutrients, pests, diseases, irrigation, etc. play an important role (Ghaffar et al.; 2020). In this paper, we focus on a PF using ACODAT for integrated cotton management. Integrated cotton management includes several factors that, when used in a mixed manner, help to make better planning and decision-making. These factors are related to the right management of fertilizers, insect pests, diseases, irrigation, weeds, etc. (Ghaffar et al., 2020). In this study, we included information related to fertilizer management, insect pests, irrigation, climate data and crop stages. These factors are related and were considered for planning and decision-making to assist the farmer in integrated cotton crop management.

3.2. MIDANO Application

We use the MIDANO methodology to design our ACODAT. Inside of our ACODAT are included data preparation and treatment data tasks.

3.2.1 ACODAT specification.

Fig. 2 shows our ACODAT approach for this purpose. ACODAT consists of a trilogy of steps that are linked together through a network of tasks to assist decision-making in cotton crop management. The first step, monitoring, is made up of two tasks: verifying and correcting data. The second step, analytics, involves classifying the population of boll weevils according to climate data and diagnosis/prediction of the cotton yield. The final step, decision-making, involves prescribing the best management strategy for cotton crops.

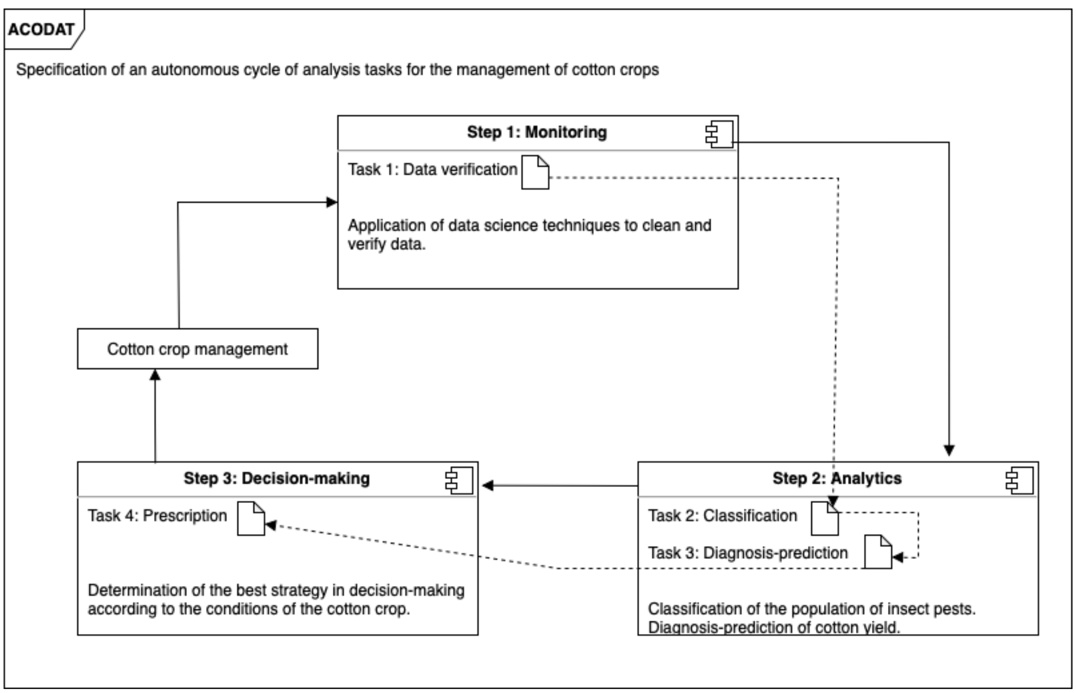


Fig. 2. ACODAT architecture for cotton crop management.

The techniques employed in the data analysis tasks belong to diverse domains of artificial intelligence (AI), including XGBoost (Chen & Guestrin, 2016; Toscano-Miranda, et al., 2022a), fuzzy systems (Cerrada et al., 2005), and genetic algorithms (GA) (Eiben et al. 1999). Therefore, the monitoring, analysis, and decision-making functionalities provided by ACODAT-based self-monitoring are as follows:

- Monitoring tasks: This process includes Task 1 to capture data, clean it and prepare it for the following tasks. In addition, relevant characteristics are extracted and preprocessed, and information about the behavior of insect pests is obtained. The selected features are used in the following steps.
- Analysis tasks: A set of tasks (tasks 2 and 3) to understand, interpret, and predict/diagnose what is happening in the cotton growing process.
- Decision-making tasks: This process includes Task 4 to prescribe the best strategy in the integrated management of cotton crops.

The complete cycle includes four integrated tasks, which communicate with each other and pass information from the first to the last. Each task used different techniques to achieve the objectives. Table 1 shows the interrelation between tasks, data sources and used techniques. The data used are historical data from the study region. The following subsections explain in detail each task in the autonomous cycle.

Table 1
Description of the ACODAT's tasks for integrated cotton management.

Role	**Task name**	**Characteristics of the task**				
		Description	**Data source**	**Analysis type**	**Technique**	**Knowledge model**
Monitoring	Data verification	Verification of data (data processing) and correction of errors	Datasets of monitoring of insects, and Climate data. Both sources are historical data from the study region.	Description	Verification Oversampling / Statistical analysis	Descriptive
Analysis	Classification	Classification of boll weevil population by climate data	Previous task	Classification/ Predictive	XGBoost	Predictive
	Diagnosis/prediction	Diagnosis/ prediction of cotton yield	Previous task, Dataset of cotton production	Diagnosis/ Predictive	Fuzzy logic	Diagnosis/ predictive
Decision-making	Prescription	Determination of the best strategy for the management of cotton crop	Previous task	Optimization	Genetic algorithm	Prescriptive

3.2.2. Monitoring tasks

Task 1 - Verification and data processing

Data Verification was designed as Task 1. This task includes a statistical analysis to evaluate the quality of the data. The modeling results are heavily influenced by the quality of the data. Thus, initially, our ACODAT identifies and fixes any potential data errors. Also, since missing data is common in this type of data, the dataset is purged of rows with missing data. Finally, an oversampling technique was used to balance the classes in the dataset.

In summary, the procedure for this task involves the subsequent actions: 1) extract the structured database about the insect pests, 2) verify if there are errors in the data, 3) delete rows with missing data, 4) Balance the dataset, where the number of samples from the minority class (the class with fewer examples) is increased by creating synthetic examples using the oversampling technique of (Gosain & Sardana, 2017). Fig. 3 shows the steps in this task, while Table 1 lists its main features.

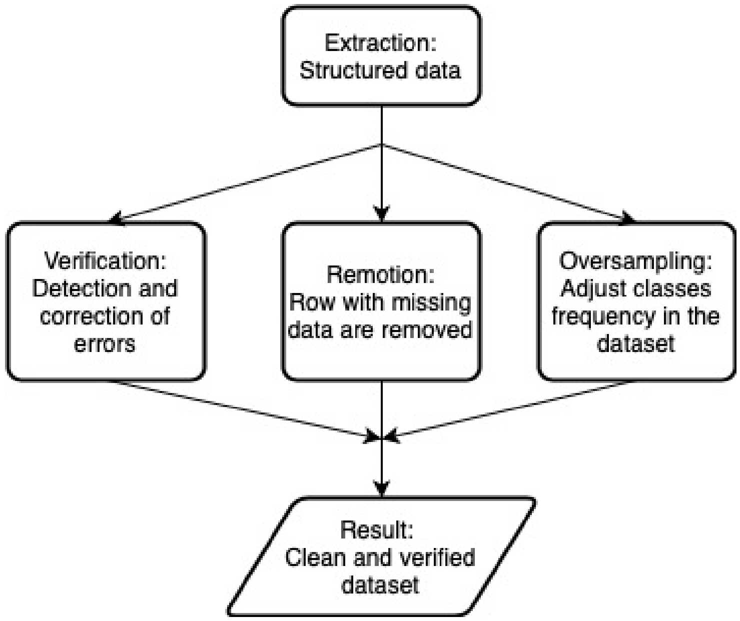


Fig. 3. Activities or sub-tasks related to task 1 (data verification and correction).

3.2.3. Analysis tasks

There are two analysis tasks, one of classification and another of diagnosis/prediction. The following is the description of each task:

Task 2 - Classification of the insect pest population:

The classification techniques are employed in this task to establish the population level of the insect pest. Thus, the classification technique determines the population level of the insect pest, for which it uses specific climatic variables for each city. The XGBoost technique, which has demonstrated the highest accuracy in prior studies (Toscano-Miranda et al., 2022a), was utilized. The main features of this task are detailed in Table 1.

Task 3 - Diagnosis/prediction of crop yield:

After the classification task, we develop the diagnosis/prediction task. This task uses a fuzzy model to diagnose/predict the cotton yield. We used expert opinions to build/define the fuzzy variables, their membership functions, and the fuzzy rules. The process involved in this task is illustrated in Fig. 4. The FS uses input variables that are passed to the fuzzification process. The inference engine uses the rule base and then the defuzzification process is performed to give a crisp output, which is the diagnosis/prediction of cotton yield.

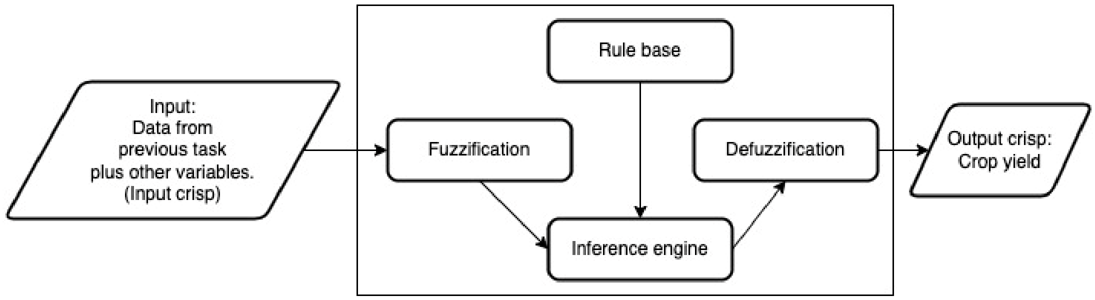


Fig. 4. Steps related to task 3 (Diagnosis/prediction).

Table 2 provides a summary of the input variables, including their descriptions, ranges, fuzzy sets and units of measure. Among these input variables, the attack level of red and black boll weevils is processed and categorized in Task 2 based on the count of boll weevils: Low (0 to 4), Medium (5 to 20), and High (greater than 20). The variable "Crop stage" indicates the phase of the crop in the year, providing insights into the ongoing activities during that phase. In Task 1, boll weevil catches, and climatic data are consolidated into a unified dataset. The variable "Fertilizer" denotes the quantity of fertilizer utilized.

Table 2
Summary of the input variables.

Input variable	Description	Fuzzy sets	Range	Units of measure
Attack level of the red boll weevil	Population of the red boll weevil in the cotton crop.	Low, Medium, and High	[0, 150]	Integer
Attack level of the black boll weevil	Population of the black boll weevil in the cotton crop.	Low, Medium, and High	[0, 200]	Integer
Crop stage	Crop stage in the year.	Vegetative, Flowering, Fruiting, Harvesting, Destruction of soca, and Closing	[0, 12]	Integer
Rainfall	Amount of rain that falls during the day.	Low, Medium, and High	[0, 17]	mm
Fertilizer	Amount of fertilizer used in the crop.	Low, Medium, and High	[0, 18]	Integer (Packages)
Pheromone traps	Number of traps used in the crop.	Absent, Adequate	[0, 1]	Integer
Boll-weevil killing tube	Number of tubes used in the crop.	Absent, Adequate	[0, 1]	Integer

Note: In this study, fertilizer application has been analyzed in a general way without specifying the type of fertilizer (for example, nitrogen, phosphorus or potassium) because this data was not reported in the datasets, only the amount that had been used

This task uses fuzzy sets with membership functions Gaussian, triangular, and trapezoidal. The triangular function is used for the categorical variables, and the trapezoidal/Gaussian functions are used for the rests. Finally, 13 membership functions were defined for the input and output variables. Fig. 5 shows an example with a trapezoidal membership function.

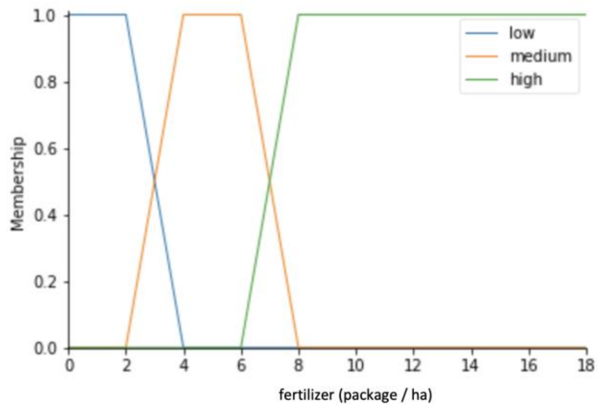


Fig. 5 Example of a trapezoidal membership function.

The experts' answers were also utilized to define the fuzzy rules. The rules are defined as IF-THEN. The antecedents are the input variables and the consequent is the crop yield. Table 3 presents two examples of the rules. For example, Rule number 1 is: IF the red attack level is High AND the black attack level is High AND the crop stage is Vegetative AND the rainfall is High AND the fertilizer is Low AND the pheromone trap is Absent AND the boll-weevil killing tube is Absent THEN the crop yield is Low. Rule 2 defines a different combination in the antecedent, and as a result, the crop yield is Medium. Thirty-eight rules were defined for the system.

Table 3
Rule structure (Example of two of them).

Rule				**If**				**Then**
	Red attack level	Black attack level	Crop stage	Rainfall	Fertilizer	Pheromone trap	Boll-weevil killing tube	Crop yield
1	High	High	Vegetative	High	Low	Absent	Absent	Low
2	High	High	Flowering	Low	High			Medium

3.2.4. Decision-making tasks

Task 4 – Prescribing of strategies for crop management

For decision-making, it was implemented a prescription task. The task was performed with a GA to determine the most efficient strategy for solving the problem. Experts' recommendations in crop management and marketing were identified as the starting point for this task. The crop management prescriptions in this task are based on expert opinion and compiled into a list. The GA optimizes the most efficient strategy for a specific scenario based on the previous task's findings. Table 1 outlines the task's characteristics. Thus, we use expert opinion to build a set of activities for each strategy. One strategy can be shaped by a combination of 13 activities. Specifically, our GA is based on the next procedure:

Algorithm 1: Training procedure of the Genetic Algorithm (GA)

Input: Data from the previous task, synthetic dataset
Output: Strategy recommended according to the best individual

1. Initialize the population
2. Evaluate the population
3. While (stopping condition not satisfied):
    (a) Select the population
    (b) Crossover the population
    (c) Mutate the population
    (d) Evaluate the population
    (e) Update the population
4. Return the best individual in the population

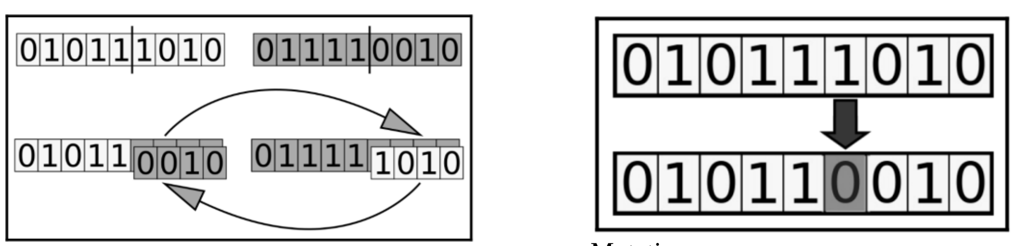
In this task, the result is the prescription of a strategy defined by a set of activities. Thus, an individual in a population is a strategy defined by a binary chain where each bit represents a gene (i.e., an activity). For example:

0	0	1	1	1	1	0	1	1	1	1	0	0

Fig. 6 An individual (prescription).

Therefore, the whole chromosome (individual) is a possible prescription. An activity should be used when a 1 appears, and not when it is 0. Thus, the population is a collection of candidate prescriptions for the context analyzed in cotton cultivation.

The following steps were taken to find the best strategy (see algorithm 1): 1) Initialize population: creating randomly a set of binary chromosomes that depict distinct solutions (possible prescriptions). 2) Evaluate: calculation of the fitness of each chromosome using the fitness function presented in the next paragraph, 3) Generating new individuals through genetic operators: In this stage, the chromosomes of the two fittest parents are selected, to which the crossover and mutation operators are applied (see Fig. 7). 4). Return the best individual in the population (i.e., the best strategy).



Crossover

Mutation

Fig. 7. Example of crossover and mutation processes in a GA.

The fitness function must evaluate possible solutions formulated on the analysis of the crop, in the before process (Task 2) and the context of the crop given as input in this task (Task 3). The fitness function returns values from 0 to 301. This output determines the best recommendation, being 0 as very adequate and the highest value as not adequate. The fitness function evaluates the context of the crop and the activities to be included in the recommendation. If the chromosome includes inappropriate activities, then those activities are penalized. For example, if the recommendation/prescription includes the activities "Conduct soil analysis" and "Apply the necessary amounts of fertilizer according to the soil analysis and the agronomist's recommendations" at the fruiting stage of the crop, this should be penalized. Experts in the management of cotton cultivation do not recommend this because at this stage of cultivation costs increase and it is not necessary. All equations were constructed based on the opinion of cotton crop management experts. The equations are:

$$C1 = A_9 * 100 + A_{10} * 100 + A_{11} * 100 + A_{12} * 100$$

Where C1 is the constraint 1, and 100 is a value that represents the penalization. The previous equation penalizes (i.e., it gives a higher value) in case the prescription includes the following activities in the flowering and fruiting stage: a) put pheromone traps (A_9), b) move the pheromone traps (A_{10}), c) put boll-weevils killing tube (A_{11}), and c) move the boll-weevils killing tube (A_{12}). This penalization is due to these activities are not recommended at these two stages and increasing the costs.

The next equation penalizes in case the prescription includes the following activities in the fruiting stage: soil analysis (A_7) and applying fertilizers (A_8). These activities are economically unfeasible at this stage and increase the costs.

$$C2 = A_7 * 100 + A_8 * 100$$

4. Case study

This section presents the experimental context and the instantiation of ACODAT in a case study for integrated cotton crop management using datasets from a region of Colombia. We test ACODAT to create a regional monitoring system. In this case study, we demonstrate how the ACODAT tasks are executed on particular datasets.

4.1. Context

We identified the data sources according to the MIDANO methodology (Aguilar et al., 2020b). To identify the appropriate sources of knowledge, we engaged with experts in cotton cultivation for this case. For our purpose, we used the next data sources: 1) Network of boll weevil (Anthonomus grandis) monitoring of the Colombian Agricultural Institute (ICA in Spanish), 2) Pheromone traps utilized in each cotton crop deployed by the owners, 3) Climate data from the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM in Spanish) for each site where cotton data were reported, 4) Farm reports of management practices for each field in the study, 5) Crop yields in the area according to the Colombian Cotton Confederation (CONALGODON in Spanish). The reported data is the seed cotton yields, which comprise both cotton seed and fiber. The yield observations were between 2016 and 2021. The study areas consist of the cities in the province of Córdoba, Colombia, where cotton is cultivated. Cotton is one of the main agricultural products in this region, covering about 6,000 hectares of land in 2022.

Our ACODAT was validated using cotton crops from different areas of Córdoba, Colombia, specifically, the cities comprising the Sinú Valley (High Sinú, Middle Sinú, and

Low Sinú) (Trebilcok, 2020), located at ~8°55'33.6"N, 75°48'16.5"W. The data used for this implementation correspond to the Network of boll weevil monitoring operationalized by the ICA and climate data from the IDEAM. These geospatial aspects describe the physical and climatic characteristics of four cities in the department of Córdoba, Colombia: Montería, Lorica, Cereté, and Ciénaga de Oro. These cities are in the lower basin of the Sinú River, which is one of the main water sources and economic activities in the region. The Sinú River flows through Montería, Lorica, and Cereté, providing them with water, fish, and transportation. The four cities have a low altitude above sea level, ranging from 7 m in Lorica to 18 m in Montería. This means that they are close to the Caribbean Sea and have a flat or slightly undulating topography. The low altitude also influences the climate of these cities, which is warm tropical, with high temperatures and humidity throughout the year. The average temperatures in these cities are between 27.3°C and 27.8°C, with little variation among them (Palencia et al., 2006). According to Palencia et al. (2006), the rainfalls increase from north to south. The soils of these cities have heterogeneous chemical characteristics, with acidic and basic soils. The soils of these cities are suitable for agricultural activities, such as rice, corn, and cotton cultivation. Data from the cities of Córdoba: Montería, Cereté, Lorica, and Ciénaga de Oro (from 2016 to 2021) were used for the experiments. We chose these regions because they are cultivated with cotton and have the records of the pheromone traps. For example, Fig. 8 shows the distribution of pheromone traps in Cereté. It can be seen that the most distant measurement between traps is 5.6 km. The pheromone traps are at strategic places close to the cotton crops.

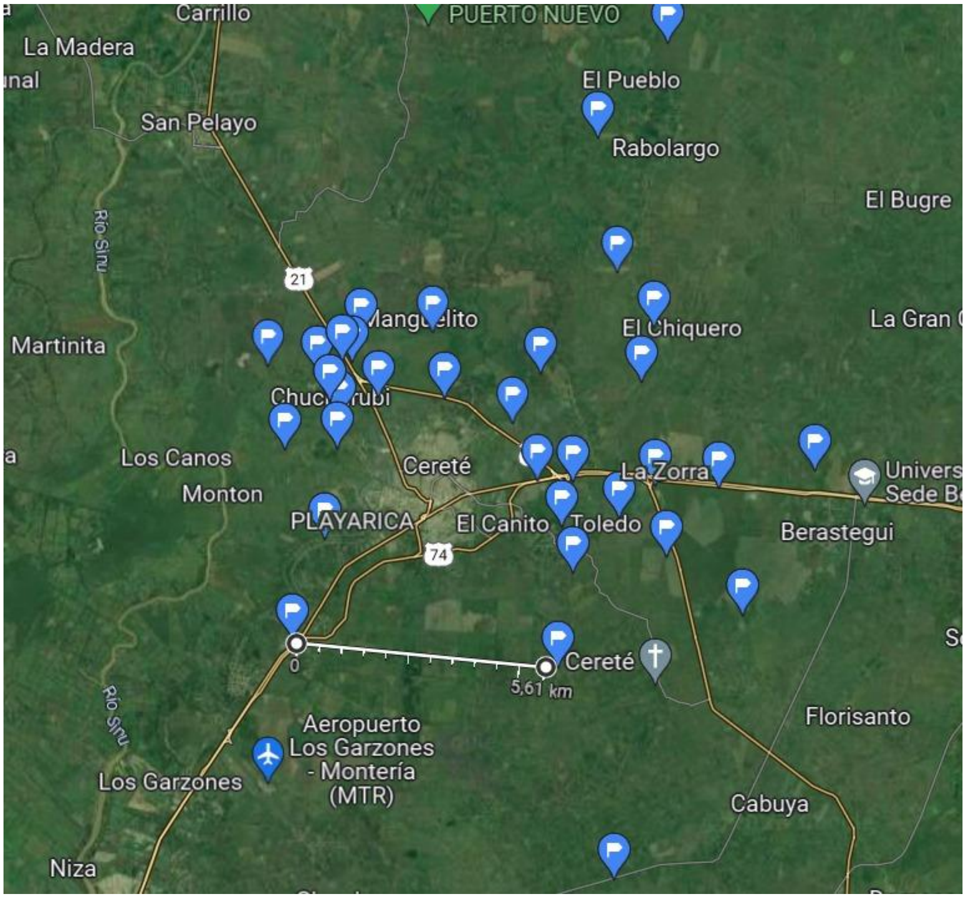


Fig. 8. Distribution of pheromone traps in Cereté.

We collected data from 374 pheromone traps, which attract and capture insects. The dataset consisted of 13,585 samples, each containing the number of boll weevils captured in a trap on a specific date. ICA engineers routinely monitor red and black boll weevil populations using conventional pheromone traps, conducting inspections every 15 days. Engineers record the boll weevil counts manually and enter the data into information system databases. We excluded 11 of the 15 variables in the dataset, such as trap code and GPS name, as they did not provide valuable information. Finally, six variables corresponding to the climatic data and related to the number of boll weevils were selected. Table 4 shows each of the variables, a brief description, and the task where it was used. The climate dataset was merged with the boll weevil capture dataset. The datasets were combined using dates and cities as common identifiers. In addition, the variables related to the stage of cultivation and fertilization were extracted from expert sources, ICA and CONALGODON.

On the other hand, to identify the outliers in the different datasets, the classic Tukey test was used, which refers to a value as an outlier if it is greater than 1.5 times the value of the interquartile range (difference between the first quartile (Q1) and the third quartile). (Q3)). On the other hand, since the outlier values can distort the results of the analysis, their causes were analyzed. They were excluded if they were the result of a data-taking error, but they were left if when analyzing the process they represented anomalous situations (determined by experts).

Table 4
Variables and their descriptions, used in cotton crop management.

Variable	Description	Units of measure	Task	Data source
Red boll weevils	The red boll weevils are the youngest. Quantity of captures of boll weevils.	Integer	1, 2	ICA
Black boll weevils	The black boll weevils are the ones that can procreate. Quantity of captures of boll weevils.	Integer	1, 2	ICA
Rainfall	Amount of rain that falls during the day.	mm	1, 2, 3, 4	IDEAM
Humidity	Hourly relative humidity (average of the day).	%	1, 2	IDEAM
Temperature	Maximum daily temperature, measured in degrees Celsius.	˚C	1, 2	IDEAM
City	City with records of boll-weevil attacks.		1, 2	ICA
Attack level of the red boll weevils	Low, medium, or high level as a result of the previous task.	Integer	2, 3, 4	Task 2
Attack level of the black boll weevils	Low, medium, or high level as a result of the previous task.	Integer	2, 3, 4	Task 2
Crop stage	Growth stage of cotton cultivation.	Integer	3, 4	ICA
Fertilizer	Amount of fertilizers used during growth stages.	Integer (Packages)	3, 4	CONALGODON, experts
Pheromone traps	The use of conventional pheromone traps in the cotton crop.	Integer	4	ICA
Boll-weevil killing tube	The use of boll-weevil killing tube in the cotton crop.	Integer	4	CONALGODON

4.2 Instantiation of ACODAT

4.2.1 Verification and data processing task

In the verification and data processing task, data about the boll-weevil captures were extracted. The dataset contained outlier data in the captures of the boll weevil, temperature, humidity, and rainfall.

To ensure the reliability of the analysis, data points identified as significant outliers in the boll weevil capture data were excluded from the dataset. For example, values of 1200 catches (in 15 days) of boll weevil were considered outliers. Considering the regional climate conditions, specific thresholds were established for the variables of humidity, temperature, and rainfall. Humidity values within the range of 68% to 90% were considered appropriate for inclusion in the analysis, as they represented the relevant range of moisture levels in the region. Similarly, temperature values above 28 ºC and below 50 ºC were considered to encompass the typical temperature range of the area under investigation. In the case of rainfall, values ranging from 0 mm to less than 18 mm were selected as they represented the relevant spectrum of precipitation levels within the region. By defining these specific thresholds, we aimed to focus the analysis on the climatic conditions most pertinent to the study, ensuring the inclusion of meaningful data points.

In some periods of the year, the cities of Cereté, Lorica, and Montería experienced missing data in the climatic variables, including rainfall, temperature, and humidity. To ensure the integrity of the analysis and minimize potential biases caused by missing values, missing data processing was performed using a deletion method based on Mckinney (2010). Under this method, any individual in the dataset with missing data for any variable included in the analysis was excluded from further analysis. By removing individuals with missing data, we aimed to retain complete cases and maintain the reliability and validity of the analysis. This approach enabled a more robust examination of the available variables and their relationships, ensuring that only complete and reliable data were considered in our analysis. Additionally, we employed the Synthetic Minority Oversampling Technique (SMOTE) (Gosain & Sardana, 2017) to even out the classes, given the low occurrence of categories of the boll weevil. Thus, for this first task, data were verified, corrected and balanced.

4.2.2 Classification task

The classification task used XGBoost as the classification technique to determine the population level of the boll weevil. In a previous work (Toscano-Miranda et al., 2022a), this is the best technique for this task among Random Forrest, Support Vector Machine and Backpropagation Neural Networks. XGBoost gave an accuracy of 88%.

This task classified the attack level according to the boll-weevil population on the three labels of the dataset. The labels were low, medium, and high. The input for this task was a dataset that had been cleaned and validated in the previous task. The dataset was divided into 80% for training and validation, and 20% for testing. XGBoost was configured in different ways and 10-fold cross-validation was performed to determine the most optimal combination of hyperparameters. The hyperparameter settings for XGBoost are shown in Table 5.

Table 5
Configuration of the hyperparameters of the XGBoost algorithm used to build the five models.

Algorithm	Best hyperparameters
XGBoost	Mtry = 1 Minimum n = 39 Tree depth = 13 Learn rate = 0.0459 Loss reduction = 0.0189 Sample size = 0.973

4.2.3 Crop yield diagnosis/prediction task

The analysis of cotton production involved the use of a soft computing method that incorporated the knowledge of experts. To perform the analysis, the system considered seven input variables, which were listed in Table 2. These variables included the level of attack from black and red boll weevils, the crop stage, the amount of rainfall, the amount of fertilizer applied, the use of pheromone traps, and the use of boll-weevil killing tubes. By considering these variables, the soft computing approach was able to generate insights into the factors that affect cotton production. This information could be used to improve the management practices of cotton farms and to increase the efficiency and profitability of cotton production. Four of these variables were reused of the previous task, including the classification of the boll-weevil population. As a result of this task, the diagnosis/prediction of cotton yield was obtained. To assess its robustness and adaptability, the system was subjected to tests using various agricultural scenarios. Each scenario defines a different combination of the variables that describe the current situation (characteristics) in the region described by the experts, which are inputs for the predictive and/or prescriptive tasks. Some of these variables are the attack level of the red/black boll weevil, crop stage, rainfall, pheromone traps and boll-weevil killing tube. The scenarios allow the evaluation of the quality of the FS predictions and the strategies generated by the prescription tasks. The knowledge provided by experts was utilized to create the fuzzy rules (see Table 3). The FS was designed with a standard fuzzy Mamdani system that integrates 38 if-then rules. To determine the yield of the crop based on the inferred inputs and rules, the defuzzification process utilized the centroid method, which is also known as the center of gravity (CoG) (Cerrada et al., 2005). This process results in a single crisp value that represents the output of the fuzzy system. As an example, Fig. 9 illustrates the outcome of defuzzification for a given set of inputs using the rules presented in Table 3. The predicted result was medium, with a yield of 2.88 tons/ha.

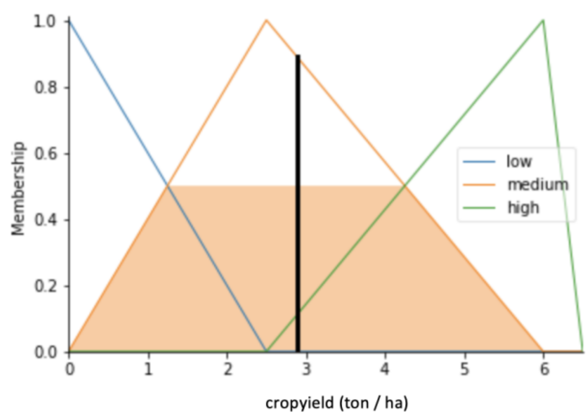


Fig. 9. Examples of defuzzification of the output variable (crop yield with 2.88 tons/ha).

To evaluate the performance of the FS, two measures were utilized as outlined in Table 6. Firstly, the *Coefficient of Determination* (R^2) was used to determine the proportion of the variance in the response variable that can be explained by independent variables. Secondly, the *Mean Squared Error* (MSE) in (tons/ha) was used to determine the difference between predicted and expert values. The R^2 score ranges between 0 and 1, and its high score represents a good result for the FS. On the other hand, the MSE should have a value lower or close to 0 for it to be considered good. These metrics were obtained by comparing the outputs of the FS with the ratings made by domain experts. The FS utilizes fuzzy reasoning, which activates fuzzy rules based on crisp input values such as fertilizer, crop stage, rainfall, pheromone trap data, black attack level, red attack level, and boll-weevil killing tube readings. These crisp values are first converted into fuzzy values and then processed to generate a fuzzy output, which is then converted into a crisp output. This crisp output is the prediction, which is used to calculate metrics such as R^2 and MSE. For the evaluation, 9 scenarios were defined testing more than 50,000 entries. More details can be found in the Results section.

Table 6
Comparison of estimated with observed yields.

R^2	MSE (ton per ha)
0.9374	0.0661

Findings indicate that the FS is capable of producing outputs that correspond with the evaluations of experts, thereby facilitating farmers in choosing the most effective cotton crop management practices to achieve optimal yield under specific circumstances.

4.2.4 Prescriptive task

The task of prescribing helps decision-making regarding the planning and management of cotton cultivation. The aim of this task was to establish the most effective strategy to manage cotton crops according to the context analyzed. It employs a series of prescriptions for the management of cotton crops according to experts in cotton cultivation, management, and marketing. Considering the results of the previous task (i.e., diagnosis/prediction of cotton yield), the GA optimizes the best strategy for a given scenario (it is an input). We use expert opinion to build a set of activities for each strategy. One strategy can be shaped by a combination of 13 activities. The activities considered in our case study are:

0. The cotton crop should be monitored more frequently.
1. The area where the boll weevil was found should be marked, according to the last inspection.
2. Cotton plant bolls that have fallen to the ground should be picked up daily.
3. The bolls affected by the boll weevil should be collected to prevent further feeding and propagation of the boll weevils.
4. The previously demarcated area should be fumigated.
5. Excessive rain must be evacuated using adequate drainage channels.
6. Implement an irrigation system.
7. Conduct soil analysis.
8. Apply the necessary amounts of fertilizer according to the soil analysis and the agronomist's recommendations.
9. Pheromone traps must be placed.
10. Move the pheromone traps frequently (use traps in the area recommended by the engineer and according to monitoring).
11. Place boll-weevil killing tube.
12. Frequently move the kill tubes (use tubes in the area recommended by the engineer and according to the monitoring).

Those recommended activities that are sought to be prescribed have been specifically defined for the study context. According to Trebilcok (2020), Colombia employs various agricultural strategies to manage cotton crops from an entomological perspective. When the boll weevil infects the crop, then specific activities are implemented. This involves distinguishing between two scenarios: when the boll weevil invades the crop in large numbers, or when it appears in isolated foci. In the case of a mass invasion, where the weevils spread and establish themselves extensively throughout the lot, the most effective solution is to closely monitor the crop from day one until day 40, when fruiting begins. During this period, a comprehensive application of insecticide is conducted to eliminate the boll weevils before they have a chance to oviposit. As reproductive structures are not yet present, they cannot serve as a host for the boll weevil's eggs.

Alternatively, if the boll weevils appear in separate foci within the crop (one or multiple foci, depending on the crop area), the agronomist identifies and marks the locations during crop monitoring. By demarcating these foci, the agronomist signals to the farm administrator the presence of boll weevil infestation in those specific areas. Subsequently, the agronomist advises the farm manager to apply insecticide and collect the reproductive structures. Typically, one or two insecticide applications are carried out consecutively, with a time gap of one or two days between them. The objective is to suppress or minimize boll weevil colonization of the crop. During the colonization process, the boll weevils may have caused damage to the reproductive structures through feeding or oviposition. To address this, personnel (one, two, or three individuals, depending on the size of the infestation focus) are assigned to collect the structures. The structures open their bracts within 48 hours and start falling to the ground. The staff can either pick them up from the ground or remove them from the plant before they naturally fall. Damaged structures exhibiting symptomatic open bracts can be easily detached from the plant. This unique strategy ensures a nearly absolute reduction in boll weevil colonization. Staff pick them up from the ground or take them from the plant without waiting for them to fall to the ground. Damaged structures are known for their open square symptomatology and can therefore be torn from the plant. This is a very special strategy to make an almost absolute reduction in the colonization of the boll weevil.

Particularly, this task uses eight variables. The level attack of red and black boll weevil is the result of the classification in Task 2, rainfall is defined from the classification in Task 2, crop yield is the result of the diagnosis/prediction in Task 3, and finally, this task considers also the next variables: the crop stage, the pheromone traps, the boll-weevil killing tube, and the fertilizer. In this task, the result is the prescription of a set of activities (they form a strategy). The GA uses the fitness function that minimizes the cost defined in the previous section. In particular, the fitness function minimizes costs in the proper use of the irrigation system, pheromone traps, boll-weevil killing tubes and fertilizer. When the farmer applies this best/optimal strategy then increases the yield of cotton.

The crossover probabilities were set to 0.9 and mutation to 0.1. Previous research has indicated that the probability values used here have been successful in producing optimal results on comparable problems (Eiben et al., 1999; Hassanat et al., 2019). The crossover operator divides two chosen parents' chromosomes at a random point, resulting in two initial and two final gene subsets. These final subsets are then exchanged, generating two new chromosomes. The mutation operator randomly modifies each offspring's genes on a chromosome level.

5. Results

5.1 Results of Task 1 - Verification and data processing

The boll weevil population was categorized based on data ranges, with the low, medium, and high groups being defined as 0 to 4, 5 to 20, and greater than 20, respectively. These intervals were determined by the ICA. The distribution of the attack-level classes was uneven and required the utilization of SMOTE (Gosain & Sardana, 2017), as well as data standardization. Nonetheless, SMOTE was not used with Ciénaga de Oro and Montería due to their limited number of high-class red boll weevils.

Table 7

Distribution of classes for boll weevil in the Córdoba region.

Class	Red boll weevils	Black boll weevils
Low (0 to 4)	6,456	4,701
Medium (5 to 20)	304	1,244
High (> 20)	83	808

5.2 Results of Task 2 - Classification of the boll-weevil population

XGBoost was selected because (1) it is the technique that has shown good performance in this context (Toscano-Miranda, 2022a), and (2) according to the literature review (Toscano-Miranda, et al., 2022b), it is the most frequent technique among structured data classification techniques. The model for classification was evaluated independently for black and red boll weevils. Three weather features - temperature, humidity and rainfall - were tested in the experiments.

XGBoost achieved an 82% accuracy rate in detecting red boll weevils, the highest among the models tested, but its ability to predict black boll weevils was constantly below 60% (see Table 8).

Table 8

Outcomes of the classification model of black and red boll weevils using rainfall, humidity, and temperature.

Boll weevils	*Accuracy*		*F1-Score*	
	Training	***Test***	***Training***	***Test***
Reds	0.82	0.82	0.82	0.82
Blacks	0.60	0.60	0.59	0.59

Additionally, experiments were performed that solely used rainfall to encompass the entire Córdoba department as well as its cities. The results indicated that the accuracy of the model was lower when using just one attribute rather than all three (see Tables 9 and 10).

Table 9

Results of the model of classification using the XGBoost algorithm and rainfall.

City	Red boll weevil				Black boll weevil			
	Accuracy		*F1-Score*		*Accuracy*		*F1-Score*	
	Training	*Test*	*Training*	*Test*	*Training*	*Test*	*Training*	*Test*
Córdoba	0.75	0.74	0.75	0.73	0.57	0.56	0.57	0.55
Cereté	0.67	0.65	0.67	0.65	0.52	0.49	0.52	0.49
Lorica	0.78	0.73	0.78	0.73	0.60	0.56	0.60	0.56
Ciénaga	FoO	FoO	FoO	FoO	0.69	0.64	0.69	0.64
Monteria	FoO	FoO	FoO	FoO	0.82	0.70	0.82	0.70

Abbreviation: FoO= Fail on oversample.

Feature selection using the ranking of features provided by Random Forest determined that temperature was the main feature. Then, new trials were executed solely using it (see Table 10). The performance of the red boll weevils' algorithm was improved in general for Córdoba using feature selection, resulting in an increase in Accuracy and F1 scores on the training dataset, from 82% (three features) to 83% (temperature only). However, not all cities obtained good results. For this reason, new tests were carried out including the three features as described later in this section.

Table 10

Outcomes of the classification model of red and black boll weevils using temperature.

Boll weevils	*Accuracy*		*F1-Score*	
	Training	*Test*	*Training*	*Test*
Reds	0.83	0.79	0.83	0.79
Blacks	0.62	0.59	0.62	0.59

XGBoost was applied to the data, using three features for each city, as detailed in Table 11. The results showed that Lorica, Cereté and Ciénaga de Oro had better accuracy with black boll weevils, while Lorica and Cereté had better accuracy with red boll weevils. However, when a model was trained using data from all locations in Córdoba, including the samples from Ciénaga de Oro, Cereté, and Lorica, the accuracy for both black and red boll weevils was found to be lower. This decrease in accuracy could potentially be attributed to the unsuccessful oversampling technique applied in Ciénaga de Oro with data of red boll weevils, where the number of captures was predominantly in the low class. This skewed data distribution may have resulted in a biased model. That is, in Ciénaga de Oro, there were few captures of boll weevils; therefore, the categorization in the Medium and High classes was not sufficient to perform oversampling effectively. Specifically, the Low class had 946 records, the Medium class had 36 records, and the High class had only 3 records. This limited representation of the Medium and High classes in Ciénaga de Oro significantly impacted the oversampling process, as the dataset lacked a robust distribution across all classes. Finally, it should be noted that Montería, another city included in the study, had limited available features, with only maximum temperature and rainfall being recorded.

Table 11

Classification model with XGBoost using temperature, rainfall, and humidity. The experiment included four cities of Córdoba.

City	Red boll weevil				Black boll weevil			
	Accuracy		*F1-Score*		*Accuracy*		*F1-Score*	
	Train	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*
*Córdoba	0.82	0.82	0.82	0.82	0.60	0.60	0.59	0.59
Cereté	0.78	0.77	0.78	0.77	0.57	0.52	0.57	0.52
Lorica	0.88	0.88	0.88	0.88	0.66	0.58	0.66	0.58
Ciénaga de Oro	FoO	FoO	FoO	FoO	0.71	0.69	0.71	0.69
Montería	NH	NH	NH	NH	NH	NH	NH	NH

*Córdoba (included Cereté, Lorica, and Ciénaga de Oro). Abbreviation: NH = No humidity. FoO = Fail on oversample.

The experiment was carried out after considering the results of previous experiments, and the models with the highest accuracy, Montería for black boll weevils and Lorica for red boll weevils were used in this test. The purpose of the experiment was to assess whether the best model for one city could result in better classification results for other cities. The models were tested across all other cities to estimate their accuracy levels, and unfortunately, the results showed a decrease in accuracy levels. Specifically, Cereté's accuracy levels dropped from 52% to 29% for black boll weevils and from 77% to 48% for red boll weevils. In other words, the models that worked best for Lorica and Montería did not perform as well in Cereté.

5.3 Results of Task 3 - Diagnosis/prediction of crop yield

This section describes the experiments and scenarios carried out to evaluate the FS. After the FS generated outputs, the results were compared to the crop yield information provided by experts.

Determination of the optimal membership functions for each scenario

Experts were asked to provide specific values for low, medium, and high scales of certain variables through a survey. Each value corresponds to a number on the scale, and the mean and standard deviation were calculated for each value (Table 12).

Table 12

Survey Results: Experts' Assessments.

Variable	Low		Medium		High	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**
Attack level of the red boll weevil	3	1.41	16.66	2.35	25	4.08
Attack level of the black boll weevil	2.66	1.69	15	4.08	25	7.07
Rainfall	2.66	0.47	6	0.81	12.33	1.69
Fertilizer	1.66	0.94	5	2.16	10.33	2.35
Crop yield	1.16	0.23	2.33	0.23	3.83	0.23

Abbreviation: Std= standard deviation

The study used different membership functions for variables such as rainfall, black boll weevil attack level, red boll weevil attack level, fertilizer, and crop yield. These functions included triangular/trapezoidal or Gaussian combinations, while other variables like crop stage, pheromone trap, and boll-weevil elimination tube only had triangular/trapezoidal membership functions. Overall, 32 possibilities were generated for each scenario, leading to a total of 288 combinations (9 scenarios x 32 possibilities). The scenarios are defined by seven input variables that describe the cotton growing context, which determine the expected yield of the crop. Particularly, the input variables are: the different stages of the crop (Vegetable, Flowering or Fruiting), the attack levels of the red and black weevil (low, medium or high), rainfall levels (low, medium or high), fertilizer levels (low, medium or high), the presence of pheromone traps (absent or adequate), and the boll-weevil killing tube presence (absent or adequate). Finally, the expected yield level is defined as low, medium or high (the description of the different scenarios is in Appendix A). The mean and standard deviation were used to create the Gaussian shape in the membership function. The best combination of membership functions was chosen for each scenario, with Table 13 showing the best performance. In some cases, triangular/trapezoidal trends were observed (e.g., scenarios 1 and 6), while in others, Gaussian trends were observed (e.g., scenario 9). The FS results were generally consistent with expert opinion, as shown in the last two columns.

Table 13
Evaluation of the best combination of membership functions.

Scenario	Membership Function								FS	Mean Expert
	Input							Output		
1	T	T	T	G	T	T	T	T	1.24	1.36
2	T	T	T	G	T	T	T	T	1.24	1.63
3	G	G	T	T	G	T	T	T	2.82	2.66
4	G	G	T	T	G	T	T	G	3.83	4
5	G	G	T	T	G	T	T	G	3.83	4
6	T	T	T	T	T	T	T	T	2.88	2.76
7	G	G	T	T	G	T	T	G	1.66	1.50
8	G	G	T	T	G	T	T	G	3.83	4
9	G	G	T	G	G	T	T	G	1.92	1.83

The input variables are fertilizer, crop stage, rainfall, pheromone trap, black attack level, red attack level, and boll-weevil killing tube. The output variable is crop yield. T = triangular / trapezoidal membership function; G = Gaussian membership function

Evaluation of the estimation capabilities of our FS

To further elaborate, the purpose of the test was to evaluate the accuracy and effectiveness of the fuzzy system in predicting crop yield values across various scenarios. The best models, which included formats of the membership functions, were chosen for each scenario, and were used in the test. The test involved considering different values of the input variables that described each scenario, which amounted to more than 50,000 entries that represent the different values of the different input variables of each scenario (representing different observations in different seasons of the year). The fuzzy system generated results

(FS outputs) for each input value, which were then compared to the crop yield established by experts. In order to compare the results with the crop yield established by the experts, the responses from each scenario were averaged to obtain a single crop yield value per scenario. This average value was then compared to the crop yield for each scenario defined by the experts. By comparing the crop yield values predicted by the FS with those established by the experts, the difference between the two was evaluated. Overall, the test was carried out to determine if the FS was consistent in predicting crop yield values that were comparable to those established by experts. This information could then be used to improve the accuracy of crop yield predictions and ultimately assist in decision-making related to crop production.

To assess the effectiveness of our FS, we employed a duo of measures for evaluating its performance. First, we used R^2 (0.9374), and second, the MSE (0.0661) (see Table 6). We can see that the results are very good.

5.4 Results of Task 4 – Prescribing with strategies for crop management

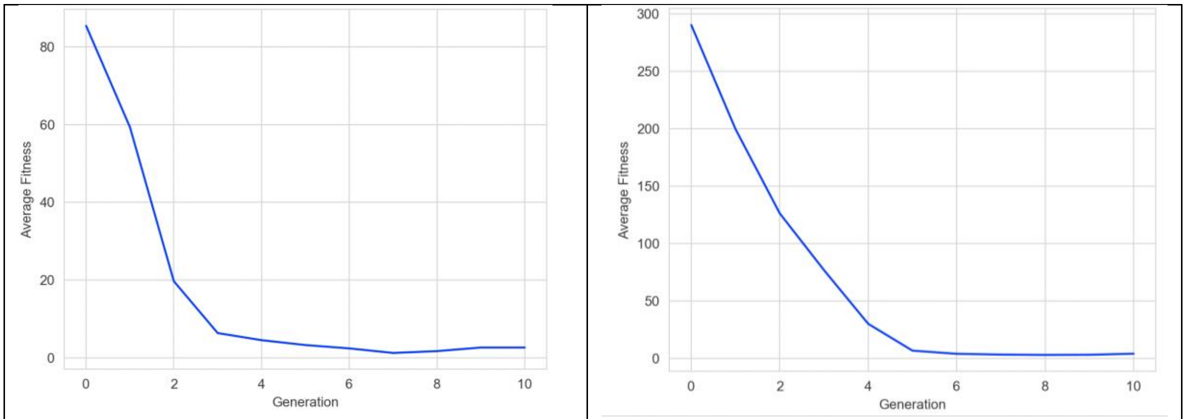
This section shows the results of ACODAT for integrated cotton crop management. For this, real data from cities in the region of Córdoba-Colombia were used. We used different scenarios to validate the experiments. Some scenarios with specific characteristics and others mixed scenarios from the former. In this paper, we present both scenarios to show the application of the autonomous cycle until reaching prescription. Table 14 summarizes the scenarios described in this section. Scenario 1 had a medium level crop yield diagnosis/prediction and Scenario 2 had a low level. According to these levels, a prescription is needed to improve crop yield.

Table 14
Summary of the scenarios.

Scenario	A	B	Crop stage	Rainfall	Fertilizer	C	D	Crop yield
1	Low	Low	Vegetative	High	Medium	Adequate	Adequate	Medium
2	Medium	Medium	Fruiting	Low	NU	NU	NU	Low

Abbreviations: A = Attack level of red boll weevils, B = Attack level of black boll weevils, C = Pheromone trap, D = Boll-weevil killing tube, NU = The farmer did not use this item.

Fig. 10 shows the results using the GA for the scenarios in Table 14. In some scenarios, convergence to optimal prescribing is faster than in others). For example, Fig. 10a. shows a convergence in seven generations, compared to Fig. 10b which shows a convergence in eight generations. The scenarios were tested several times, Fig. 10 shows the average of the generation in which the fitness function reaches the optimal strategy. In these experiments, the average time to complete a generation was 1.35 seconds on a MacBook Pro with a 2.4 GHz quad-core Intel Core i5 processor and 8 GB of 2133 MHz RAM. Fig. 10a begins with values up to 80 and finds the best prescription in generation number 7. Fig. 10b begins with values up to 250 and finds the best prescription in generation number 8. The value in the y-axis indicates the values average of the fitness function. The values higher indicated that the individual was penalized. The values closer to zero are appropriate because is an optimization problem of minimizing the costs.



A. Scenario 1 B. Mixed Scenario

Fig. 10. Results of the minimization of the fitness function (with 10 generations).

5.5 General Discussion of Prescriptive Analysis

In the diagnostic/prescribing task, only cases where the crop yield is low or medium are invoked. Therefore, Table 15 shows poor-performance scenarios. In the "Type" column, *Isolated* refers to a scenario where the yield is only of a type; and *Mixing* refers to a scenario where the yield can be of different types (e.g., low or medium). All prescription results were 100% correct with all activities included in the strategy, and in this sense, the error rate was 0. The generation number needed to reach the prescription was different from scenario to scenario.

Table 15
Example scenarios and their results.

Scenario	The best prescription	No. generations	Error	Crop yield	Type
1	100%	7	0	Medium	Mixing
2	100%	7	0	Low	Isolated
3	100%	7	0	Medium	Isolated
4	100%	8	0	Medium	Isolated
5	100%	8	0	Low	Isolated
6	100%	8	0	Low	Isolated
7	100%	7	0	Low	Isolated
8	100%	8	0	Low	Mixing
9	100%	7	0	Low	Mixing

Now, we took two examples to show the results of the prescription in real conditions. The analysis of scenario 1 indicates a medium level of cotton crop yield and scenario 2 a low level.

Scenario 1:

The characteristics of this scenario are: first, it begins with the classification task of the boll-weevil population: The classification task received input values of temperature, humidity and rainfall of the cultivated area and classified the attack level of the boll weevil as: *low attack level of red boll weevils, low attack level of black boll weevils*.

Second, the diagnosis/prediction task of crop yield received as input values the results of the previous task: a low attack level of red boll weevils and a low attack level of black boll weevils. Additionally, the crop was in the vegetative stage, the rainfall was high (17 mm). Also, at this stage, the farmer used 5 packages of fertilizer (medium), used pheromone traps, and a boll-weevil killing tube. As a result of this task, the diagnosis/prediction of the crop yield was medium (2.88 ton/ha), see Fig. 9.

Third, the prescription task for management crop received as input values the results of the previous task (see Fig. 11): a) a low attack level of red boll weevils, b) a low attack level of black boll weevils, c) a stage of the crop in vegetative, d) a high rainfall (17 mm). Also, at this stage, the farmer e) used five packages of fertilizer (medium), f) used pheromone traps, g) used a boll-weevil killing tube, and mainly, and h) the crop yield was diagnosed as medium. Therefore, according to the medium crop yield, ACODAT should generate a prescription with the best strategy. ACODAT then generates the best strategy as a recommendation to increase the cotton yield to achieve a high level. In this sense, the final prescription is the following chromosome:

1	1	1	1	1	1	0	1	1	0	0	1	1

Fig. 11 Best individual for the first scenario.

Each gene corresponds to an activity. If there is a 0 the activity is not recommended and if there is a 1 the activity is recommended. Table 16 shows the details of each gene on the previous chromosome.

Table 16
Activity configurations of the best recommendation.

Position on chromosome	Gene	Activity
1	1	The cotton crop should be monitored more frequently.
2	1	The area where the boll weevils were found should be marked, according to the last inspection.
3	1	The cotton buds (squares) of the cotton plants that have fallen to the ground must be collected daily.
4	1	The bolls of the cotton plants that have been affected by the boll weevil must be collected to prevent the boll weevil from feeding and spreading.
5	1	The previously demarcated area should be fumigated.
6	1	Excessive rain must be evacuated using adequate drainage channels.
7	0	The irrigation system should NOT be implemented.
8	1	Soil analysis should be performed.
9	1	The necessary amounts of fertilizer should be applied according to soil analysis and agronomist recommendations.
10	0	Pheromone traps must NOT be placed.
11	0	DO NOT move the pheromone traps frequently.
12	1	Boll-weevil killing tubes should be installed.

13	1	Boll-weevil killing tubes should be moved frequently.

This result is correct because the prescription found an optimal strategy, minimizing costs and using activities that improve crop yield. This prescription outcome is an optimal strategy because it defines a set of activities that improve crop yield with minimal cost. The fitness function minimizes costs by proposing the correct use of the irrigation system, pheromone traps, boll weevil killing tubes and fertilizers. When the farmer follows this best/optimal strategy, it increases the yield of cotton. For example, if the farmer uses both pheromone traps and boll-weevil killing tubes at the same time, she/he will unnecessarily increase costs. Therefore, the system recommends using one of them (the boll-weevil killing tubes). The prescription points out that the farmer a) should monitor the cotton crop more, b) should mark the area where the boll weevils were found, according to the last inspection, c) should collect daily the cotton buds (squares) of the cotton plants that have fallen to the ground, d) should collect the bolls from cotton plants that have been affected by the boll weevil and thus prevent further feeding and spread of the boll weevils, e) should fumigate the previously demarcated area, f) should evacuate the excessive rain with draining channels, g) should perform a soil analysis, h) should apply the right amount of fertilizer according to soil analysis and agronomist recommendations, i) should install boll-weevil killing tubes, and j) should move frequently the boll-weevil killing tubes. Activities a), b), c), d), and e) should be performed because monitoring and control activities are needed to quickly eradicate the boll weevil. Activity j) included boll-weevil killing tubes and excluded pheromone traps (i.e., the farmer should not use these activities simultaneously because it increases the cost and it is not necessary). In brief, the prescriptive model gives an accurate suggestion regarding the expert opinion on cotton cultivation.

Scenario 2:

The characteristics of this scenario are: first, it begins with the classification task of the boll-weevil population: The classification task received input values of temperature, humidity and rainfall of the cultivated area and classified the attack level of the boll weevil as*: medium attack level of red boll weevils, medium attack level of black boll weevils*.

Second, the diagnosis/prediction task of crop yield received as input values the results of the previous task: a medium attack level of red boll weevils and a medium attack level of black boll weevils. Additionally, the crop was in the fruiting stage, the rainfall was low (2 mm). Also, at this stage, the farmer did not use fertilizer, pheromone traps, and a boll-weevil killing tube. As a result of this task, the diagnosis/prediction of the crop yield was low (1.23 tons/ha), see Fig. 12.

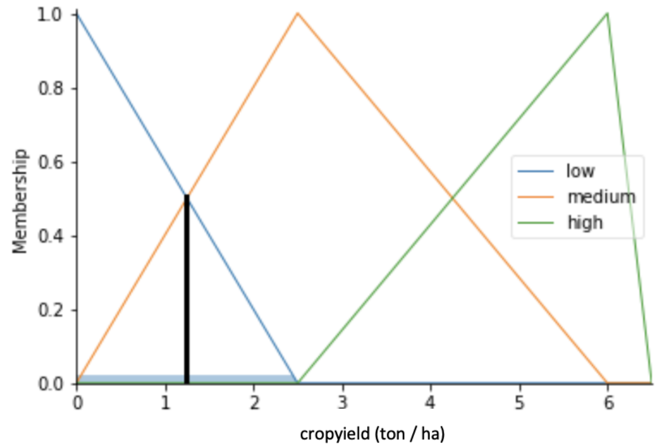


Fig. 12. Defuzzification of the output variable (crop yield with 1.23 tons/ha).

Third, the prescription task for management crop received as input values the results of previous task (see Fig. 13): a) a medium attack level of red boll weevils, b) a medium attack level of black boll weevils, c) a stage of the crop in fruiting, d) a low rainfall (2 mm), e) at this stage the farmer did not use fertilizer, f) nor pheromone traps, g) no tube kills weevils, and mainly, h) the crop yield was diagnosed as low. Therefore, and according to the low crop yield, ACODAT then generates the best strategy as a recommendation to increase the cotton yield to achieve a high level. In this sense, the final prescription is the following:

0	0	1	1	1	0	1	0	0	0	0	0	0

Fig. 13. Best individual for the second scenario.

Table 17 shows the details of each gene on the previous chromosome.

Table 17
Activity configurations of the best recommendation.

Position on chromosome	Gene	Activity
1	0	The cotton crop should NOT be monitored more frequently.
2	0	The area where the boll weevils were found should NOT be marked, according to the last inspection.
3	1	The cotton buds (squares) of the cotton plants that have fallen to the ground must be collected daily.
4	1	The bolls affected by the boll weevil should be collected to prevent further feeding and propagation of the boll weevils.
5	1	The previously demarcated area should be fumigated.
6	0	Excessive rain must NOT be evacuated using adequate drainage channels.
7	1	An irrigation system should be implemented.
8	0	Soil analysis should NOT be performed.
9	0	Fertilizer should NOT be applied.
10	0	Pheromone traps should NOT be placed.
11	0	Pheromone traps should NOT be moved frequently.
12	0	Boll-weevil killing tubes should NOT be placed.
13	0	Boll-weevil killing tubes should NOT be moved frequently.

This result is correct because the prescription found an optimal strategy, minimizing costs and using activities that improve crop yield. For example, in this scenario, the system recommends not doing the soil analysis or applying fertilizer in the fruiting stage because it is not cost-effective and should have been done in earlier stages. The prescription points out that the farmer should a) pick up daily the cotton buds (squares) of the cotton plants that have fallen to the ground, b) collect the bolls affected by the boll weevil to prevent further feeding and propagation of the boll weevils, c) fumigate the previously demarcated area, and d) increase water irrigation with an irrigation system. It should be noted that fumigation is recommended considering the previous demarcation, i.e., as the crop is in the fruiting stage, actions in previous stages should have included demarcation. The system prescribes in real time based on the crop's stage, and in this case, the crop is in the fruiting stage. Therefore, the system should have already recommended this activity (as seen in the previous scenario). In addition, since the crop is in the fruiting stage, the prescription did not include crop analysis activities, fertilizer application, use of pheromone traps, or use of boll-kill weevil tubes, because they are economically unviable at this stage of cultivation. In brief, the prescriptive model gives an accurate suggestion regarding the expert opinion on cotton cultivation.

5.6 General discussion

Our proposal monitored the data and processed it to generate statistical analyses on the behavior of insect pests on cotton crops. A set of variables and expert opinions were considered to diagnose/predict cotton yield. Finally, we use the data processed above to prescribe the best strategy for integrated cotton crop management.

The classification task of the boll-weevil population was performed using XGBoost with 88% of accuracy using climate data. The results of the diagnosis/prediction of cotton yield showed that can a) manage the uncertainty from the variables of the context or the model, b) manage the knowledge of the experts to adapt the model, and c) use concurrently variables of the climate, of the pests, crops, and fertilizers. The results of the prescription task showed that using GA allows determining the optimal strategy according to the context. The system enables assessing the crop conditions in real-time at any stage of its development and provides timely recommendations to improve its performance. For this purpose, we conducted discrete evaluations on different dates or stages of the crop and compared them with expert opinions. Overall, these results show that the integrated use of data collection, data processing and decision-making technologies are useful in PF for cotton crop management.

5.7 Comparison with previous works

This study defines an ACODAT for integrated cotton management. The tasks have been validated by experts with good results in classification, diagnosis/prediction, and prescription tasks. We introduce a set of qualitative criteria in this section to compare our work with other related works. These criteria are:

Criterion 1 - Uncertainty model: whether they proposed uncertainty models for diagnosis/prediction.
Criterion 2 - Integrate management: whether they consider the integrated management of the crop.
Criterion 3 - Production: whether they considered improving the production of the crops.
Criterion 4 - Autonomous systems (AS) that include among other tasks, classification, diagnosis/prediction, and prescription tasks to improve the production.
Criterion 5 - Simultaneous use of Climatic, pests, Fertilizers, and Crop variables (CLFCT).

According to the above criteria, Table 18 shows the comparison with the related works. The existing papers did not meet all the requirements. All the criteria we consider in our work are important because working together allows the operation of a robust system with autonomous tasks for integrated cotton crop management.

Table 18

Comparison with other works.

Work	Uncertainty model	Integrate management	Production	AS	CLFCT
Tribouillois et al. (2022)		✓	✓		
Aggarwal et al. (2022)		✓	✓		
Wu et al. (2020)		✓	✓		
Hajimirzajan et al., (2021)		✓	✓		
This work	✓	✓	✓	✓	✓

Abbreviation: CLFCT= Simultaneous use of Climatic, of pests, of Fertilizer, and of Crop variables. Production = Whether the study considered improving crop production. AS= Autonomous systems that include classification, diagnosis/prediction, and prescription tasks.

Some studies related used integrated management. For example, Tribouillois et al. (2022) built an integrated modeling of crop and water management to optimize irrigation. Hajimirzajan et al., (2021) defined a large-scale crop planning, which involves a comprehensive strategic framework that employs a decision support system to determine the sustainable use of water, as well as optimal crop selection, timing, and cultivation practices. Aggarwal et al. (2022) developed a system of geospatial analysis to preserve land fertility, optimize agricultural revenue, and minimize agricultural pollution and water consumption. Wu et al. (2020) developed a model for integrated nutrient management. It should be noted that the previous authors used integrated crop management because they considered different variables to have a broad management of the analyzed context. But no one of them uses different data analysis tasks, with different variables, and an autonomous cycle to integrate them, which our work does. They also do not consider knowledge obtained from expert recommendations to fit the model.

As previously discussed, our approach is the initial one to combine these criteria and propose an integrated cotton management approach using an ACODAT, which can be developed further with multi-agent systems (Aguilar et al., 2015; Terán et al., 2017). The purpose of integrating the multi-agent systems paradigm is to make the system more adaptable, extendable, and autonomous, as described by Aguilar et al. (2018).

6. Conclusions

This study aimed to develop a system of PF using an ACODAT for the integrated management of cotton. The cycle used tasks of data processing, classification/prediction of cotton yield, and prescribing strategies for integrated cotton management. In the autonomous cycle, each task communicates with the next and passes processed information. Also, each task has its own AI techniques and the integration of all of them produces strategies according to the context of the crop. The combined use of data analysis tasks in one cycle provided notable advantages compared to isolated techniques. To our knowledge, this is the first work to use an autonomous architecture to support integrated cotton management.

We consider some limitations in this work. First, for the diagnosis/prediction of cotton yield, the fertilizer variable only included the amount used. Secondly, for the diagnosis/prediction of cotton yield, we used only the behavior of the boll weevil. Future work should be aimed at improving the diagnosis/prediction model including more variables (e.g., specific fertilizers), and including the behavior of other insect pests and diseases. Third, this proposal did not include pheromone traps with real-time data updating in the case study. This would be an improvement that can be incorporated into the system to have real-time feedback. Fourth, we believe that other validation processes for ACODAT should be studied to evaluate its recommendations at critical stages of cotton growth. Therefore, in future work, we will use cross-validation to evaluate the performance of the ACODAT system at specific stages of the cotton growth cycle over the years. In addition, we have planned to integrate this work with an autonomous cognitive architecture for agriculture. Our approach involves defining a meta-learning task, which will enable us to create models of weevil behavior specific to different regions. To achieve this, we will utilize the transfer learning paradigm, which involves transferring knowledge gained from one task to another related task. By doing so, we hope to improve the accuracy and efficiency of the system's predictions and provide valuable insights to farmers and other stakeholders in the agricultural sector.

Finally, future work should also explore which variables could be calculated in a determinist manner through known mathematical definitions, such as those existing in the literature to determine the yield of cotton based on rainfall. In addition, our ACODAT should be tested in cotton crop simulators such as CropGRO-Cotton, with the respective adaptations to exploit all their variables, such as the estimates of the impact of temperature and nitrogen levels that this simulator provides. As a final point, the models we develop for weevil behavior will be integrated with our cognitive architecture, which is based on the multi-agent systems paradigm. Our decision to use this approach is rooted in the fact that agent theory has already established many effective modeling capabilities and implementations, which can be leveraged to improve the accuracy and efficiency of our models.

Acknowledgments

We thank the Colombian Agricultural Institute (ICA), Colombian Cotton Confederation (CONALGODON) and Colombian Institute of Hydrology, Meteorology and Environmental Studies (IDEAM). We thank Universidad EAFIT and Universidad de Córdoba for their support in this study.

Funding

This work was supported by the Colombian Science, Technology, and Innovation Fund (FCTeI) of the General Royalty System (SGR); Universidad EAFIT; and Universidad de Córdoba.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Ethical Approval

Not applicable

References

Abbas, G., Fatima, Z., Tariq, M., Ahmed, M., Nasim, W., Rasul, G., Ahmad, S., & others. (2020). Applications of Crop Modeling in Cotton Production. In *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 429–445). Springer. https://doi.org/10.1007/978-981-15-1472-2_11

Aggarwal, S., Srinivas, R., Puppala, H., & Magner, J. (2022). Integrated decision support for promoting crop rotation based sustainable agricultural management using geoinformatics and stochastic optimization. *Computers and Electronics in Agriculture*, *200*. https://doi.org/10.1016/j.compag.2022.107213

Agostini, A., Alenyà, G., Fischbach, A., Scharr, H., Wörgötter, F., & Torras, C. (2017). A cognitive architecture for automatic gardening. *Computers and Electronics in Agriculture*, *138*, 69–79. https://doi.org/10.1016/j.compag.2017.04.015

Aguilar, J., Bessembel, I., Cerrada, M., Hidrobo, F., & Narciso, F. (2008). Una Metodología para el Modelado de Sistemas de Ingeniería Orientado a Agentes. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, *12*(38),39-60.

Aguilar J., Jerez M., Exposito E., Villemur T. (2015) CARMiCLOC: Context Awareness Middleware in Cloud Computing, *Latin American Computing Conference (CLEI)*, 2015, https:// doi.org/10.1109/CLEI.2015.7360013.

Aguilar J., Garcia G. (2018) An Adaptive Intelligent Management System of Advertising for Social Networks: A Case Study of Facebook, *IEEE Transactions on Computational Social Systems*, *5*(1), 20-32, https://doi.org/10.1109/TCSS.2017.2759188

Aguilar, J., Ardila, D., Avendaño, A., Macias, F., White, C., Gomez-Pulido, J., de Mesa, J. G., & Garces-Jimenez, A. (2020). An autonomic cycle of data analysis tasks for the supervision of HVAC systems of smart building. *Energies*, *13*(12). https://doi.org/10.3390/en13123103

Aguilar, J., Salazar C., Velasco H., Monsalve-Pulido J, Montoya E. (2020) Comparison and valuation of Different Methods for the Feature Extraction from Educational Contents *Computation 8*(2). https://doi.org/10.3390/computation8020030

Ahmad, I., Ghaffar, A., Haider, G., Ahmad, A., Ahmad, B., Tariq, M., Nasim, W., Rasul, G., Fahad, S., Ahmad, S., & others. (2020). Climate Resilient Cotton Production System: A Case Study in Pakistan. In *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 447–484). Springer. https://doi.org/10.1007/978-981-15-1472-2_11

Ahmed, N., Ali, M. A., Danish, S., Chaudhry, U. K., Hussain, S., Hassan, W., Ahmad, F., & Ali, N. (2020). Role of Macronutrients in Cotton Production. In *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 81–104). Springer. https://doi.org/10.1007/978-981-15-1472-2_11

Ahmed, N., Ali, M. A., Hussain, S., Hassan, W., Ahmad, F., & Danish, S. (2020). Essential Micronutrients for Cotton Production. In *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 105–117). Springer. https://doi.org/10.1007/978-981-15-1472-2_11

Ali, M., Deo, R. C., Downs, N. J., & Maraseni, T. (2018). Cotton yield prediction with Markov Chain Monte Carlo-based simulation model integrated with genetic programing algorithm: A new hybrid copula-driven approach. *Agricultural and Forest Meteorology*, *263*, 428–448. https://doi.org/10.1016/j.agrformet.2018.09.002

Alves, A. N., Souza, W. S. R., & Borges, D. L. (2020). Cotton pests classification in field-based images using deep residual networks. *Computers and Electronics in Agriculture*, *174*(April), 105488. https://doi.org/10.1016/j.compag.2020.105488

Anees, M., & Shad, S. A. (2020). Insect pests of cotton and their management. In *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 177–212). Springer. https://doi.org/10.1007/978-981-15-1472-2_11

Caldeira, R. F., Santiago, W. E., & Teruel, B. (2021). Identification of cotton leaf lesions using deep learning techniques. *Sensors*, *21*(9), 3169. https://doi.org/10.3390/s21093169

Cerrada, M., Aguilar, J., Colina, E., & Titli, A. (2005). Dynamical membership functions: An approach for adaptive fuzzy modelling. *Fuzzy Sets and Systems*, *152*(3), 513–533. https://doi.org/10.1016/j.fss.2004.10.004

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., & Traore, D. (2022). Deep learning for precision agriculture: A bibliometric analysis. In *Intelligent Systems with Applications* (Vol. 16). Elsevier B.V. https://doi.org/10.1016/j.iswa.2022.200102

Cui, M., Qian, J., & Cui, L. (2022). Developing precision agriculture through creating information processing capability in rural China. *Journal of Rural Studies*, *92*, 237–252. https://doi.org/10.1016/j.jrurstud.2022.04.002

Eiben, Á. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, *3*(2), 124–141. https://doi.org/10.1109/4235.771166

Ghaffar, A., ur Rahman, M., Ali, H. R., Haider, G., Ahmad, S., Fahad, S., & Ahmad, S. (2020). Modern concepts and techniques for better cotton production. In *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 589–628). Springer. https://doi.org/10.1007/978-981-15-1472-2_11

Gandonou, J.-M. A. (2005). *Essays on precision agriculture technology adoption and risk management*. University of Kentucky.

Gosain, A., & Sardana, S. (2017). Handling class imbalance problem using oversampling techniques: A review. *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017*, *2017-Janua*, 79–85. https://doi.org/10.1109/ICACCI.2017.8125820

Hajimirzajan, A., Vahdat, M., Sadegheih, A., Shadkam, E., & Bilali, H. El. (2021). An integrated strategic framework for large-scale crop planning: sustainable climate-smart crop planning and agri-food supply chain management. *Sustainable Production and Consumption*, *26*, 709–732. https://doi.org/10.1016/j.spc.2020.12.016

Hassanat, A., Almohammadi, K., Alkafaween, E., Abunawas, E., Hammouri, A., & Prasath, V. B. S. (2019). Choosing mutation and crossover ratios for genetic algorithms-a review with a new dynamic approach. *Information*, *10*(12). https://doi.org/10.3390/info10120390

Hearn, A. B. (1994). OZCOT: A simulation model for cotton crop management. *Agricultural Systems*, *44*(3), 257–299. https://doi.org/10.1016/0308-521X(94)90223-3

Hoyos, W., Aguilar, J. & Toro, M. (2022) A clinical decision-support system for dengue based on fuzzy cognitive maps. *Health Care Manag Sci 25*, 666–681. https://doi.org/10.1007/s10729-022-09611-6

Hudgins, E. J., Liebhold, A. M., & Leung, B. (2017). Predicting the spread of all invasive forest pests in the United States. *Ecology Letters*, *20*(4), 426–435. https://doi.org/10.1111/ele.12741

Jones, D., & Barnes, E. M. (2000). Fuzzy composite programming to combine remote sensing and crop models for decision support in precision crop management. *Agricultural Systems*, *65*(3), 137–158. https://doi.org/10.1016/S0308-521X(00)00026-3

Kandalkar, G., Deorankar, A. v, & Chatur, P. N. (2014). Classification of Agricultural Pests Using DWT and Back Propagation Neural Networks. *International Journal of Computer Science and Information Technologies (IJCSIT)*, *5*(3), 4034–4037. www.ijcsit.com

Khattab, A., Habib, S. E. D., Ismail, H., Zayan, S., Fahmy, Y., & Khairy, M. M. (2019). An IoT-based cognitive monitoring system for early plant disease forecast. *Computers and Electronics in Agriculture*, *166*(September), 105028. https://doi.org/10.1016/j.compag.2019.105028

Kong, Q., Kuriyan, K., Shah, N., & Guo, M. (2019). Development of a responsive optimisation framework for decision-making in precision agriculture. *Computers and Chemical Engineering*, *131*. https://doi.org/10.1016/j.compchemeng.2019.106585

Lemmon, H. (1986). Comax: An expert system for cotton crop management. *Science*, *233*(4759), 29–33. https://doi.org/10.1126/science.233.4759.29

Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, *3*(5), 497–501. https://doi.org/10.1038/nclimate1832

Mammarella, M., Comba, L., Biglia, A., Dabbene, F., & Gay, P. (2021). Cooperation of unmanned systems for agricultural applications: A theoretical framework. *Biosystems Engineering*. https://doi.org/10.1016/j.biosystemseng.2021.11.008

Mansour, A. I., & Abu-Naser, S. (2019). Expert System for the Diagnosis of Wheat Diseases. *International Journal of Academic Information Systems Research (IJAISR)*, *3*(4), 19–26. www.ijeais.org/ijaisr

Maskey, M. L., Pathak, T. B., & Dara, S. K. (2019). Weather based strawberry yield forecasts at field scale using statistical and machine learning models. *Atmosphere*, *10*(7). https://doi.org/10.3390/atmos10070378

Masood, M. H., Saim, H., Taj, M., & Awais, M. M. (2020). *Early Disease Diagnosis for Rice Crop*. 1–5. http://arxiv.org/abs/2004.04775

Mckinney, W. (2010). *Data Structures for Statistical Computing in Python*.

Monsalve, J., Aguilar, J., Montoya, W., Salazar., C. (2024), "Autonomous recommender system architecture for virtual learning environments", *Applied Computing and Informatics*. *20*(1/2), 69-88. https://doi.org/10.1016/j.aci.2020.03.001

Morales, L., Ouedraogo, C. A., Aguilar, J., Chassot, C., Medjiah, S., & Drira, K. (2019). Experimental comparison of the diagnostic capabilities of classification and clustering algorithms for the QoS management in an autonomic IoT platform. *Service Oriented Computing and Applications*, *13*(3), 199–219. https://doi.org/10.1007/s11761-019-00266-w

Palencia, G., Mercado, T., & Combatt, E. (2006). *Estudio agroclimático del departamento de Córdoba*. Universidad de Córdoba, https://www.researchgate.net/publication/333356934_Estudio_Agroclimatico_del_Departamento_de_Cordoba.

Pathak, T. B., Jones, J. W., Fraisse, C. W., Wright, D., & Hoogenboom, G. (2012). Uncertainty analysis and parameter estimation for the CSM-CROPGRO-cotton model. *Agronomy Journal*, *104*(5), 1363–1373. https://doi.org/10.2134/agronj2011.0349

Saha, S., Morita, T., Ospina, R., & Noguchi, N. (2022). A Vision-based Navigation System for an Agricultural Autonomous Tractor. *IFAC-PapersOnLine*, *55*(32), 48–53. https://doi.org/10.1016/j.ifacol.2022.11.113

Sanchez, M., Aguilar, J., Cordero, J., Valdiviezo-Diaz, P., Barba-Guaman, L., & Chamba-Eras, L. (2016). Cloud computing in Smart Educational environments: application in Learning Analytics as service. *Advances in Intelligent Systems and Computing*, *445*. https://doi.org/10.1007/978-3-319-31232-3

Sánchez, M., Exposito, E., & Aguilar, J. (2020). Implementing self-* autonomic properties in self-coordinated manufacturing processes for the Industry 4.0 context. *Computers in Industry*, *121*. https://doi.org/10.1016/j.compind.2020.103247

Say, S. M., Keskin, M., Sehri, M., & Sekerli, Y. E. (2018) Adoption of precision agriculture technologies in developed and developing countries. *The Online Journal of Science and Technology*, *8*(1), 7–15.

Stevens, W. E., Varco, J. J., & Johnson, J. R. (1996). Evaluating cotton nitrogen dynamics in the GOSSYM simulation model. *Agronomy Journal*, *88*(5), 127–132.

Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*. https://doi.org/10.1016/j.compeleceng.2022

Song, X., Yang, C., Wu, M., Zhao, C., Yang, G., Hoffmann, W. C., & Huang, W. (2017). Evaluation of Sentinel-2A satellite imagery for mapping cotton root rot. *Remote Sensing*, *9*(9), 1–17. https://doi.org/10.3390/rs9090906

Sterritt, R., Parashar, M., Tianfield, H., & Unland, R. (2005). A concise introduction to autonomic computing. *Advanced Engineering Informatics*, *19*(3), 181–187. https://doi.org/10.1016/j.aei.2005.05.012

Suleiman, F. M. (2019) Development of A Rule based System for Safflower Disease Diagnosis. *International Journal of Academic Engineering Research (IJAER)*, *3*(8), 1–9.

Terán, J., Aguilar, J., & Cerrada, M. (2017) Integration in industrial automation based on multi-agent systems using cultural algorithms for optimizing the coordination mechanisms. *Computers in Industry*, *91*, 11–23. https://doi.org/10.1016/j.compind.2017.05.002

Toscano-Miranda, R., Hoyos, W., Caro, M., Aguilar, J., Trebilcok, A., & Toro, M (2022). A Classification Model of Cotton Boll-Weevil Population. *XVLIII Latin American Computer Conference (CLEI)*, https://doi.org/10.1109/CLEI56649.2022.9959893

Toscano-Miranda, R., Toro, M., Aguilar, J., Caro, M., Marulanda, A., & Trebilcok, A. (2022). Artificial-intelligence and sensing techniques for the management of insect pests and diseases in cotton: a systematic literature review. *The Journal of Agricultural Science*, *160*, 16–31. https://doi.org/10.1017/ S002185962200017X

Trebilcok, A. (2020). El cultivo del algodonero en Córdoba: Comportamiento de la variedad Nu OPAL (Gossypium hirsutum L.) bajo diferentes arreglos espaciales, *Revista Temas Agrarios 15(2),* 66 – 74.

Tribouillois, H., Constantin, J., Murgue, C., Villerd, J., & Therond, O. (2022). Integrated modeling of crop and water management at the watershed scale: Optimizing irrigation and modifying crop succession. *European Journal of Agronomy*, *140*. https://doi.org/10.1016/j.eja.2022.126592

Vulpi, F., Marani, R., Petitti, A., Reina, G., & Milella, A. (2022). An RGB-D multi-view perspective for autonomous agricultural robots. *Computers and Electronics in Agriculture*, *202*. https://doi.org/10.1016/j.compag.2022.107419

Wu, H., Yang, T., Liu, X., Li, H., Gao, L., Yang, J., Li, X., Zhang, L., & Jiang, S. (2020). Towards an integrated nutrient management in crop species to improve nitrogen and phosphorus use efficiencies of Chaohu Watershed. *Journal of Cleaner Production*, *272*. https://doi.org/10.1016/j.jclepro.2020.122765

Appendix A

			Input					Output
Scenario	**Red attack level**	**Black attack level**	**Crop stage**	**Rainfall**	**Fertilizer**	**Pheromone trap**	**Boll-weevil killing tube**	**Expected Crop yield**
1	150	200	vegetative 0.5	17	1	0	0	low
2	15	15	vegetative 0.5	17	1	0	0	low
3	15	15	vegetative 0.5	2	5	1	1	medium
4	0	0	vegetative 0.5	6	13	1	1	high
5	0	0	flowering 1.5	6	13	NU	NU	high
6	15	15	flowering 1.5	2	5	NU	NU	medium
7	15	15	flowering 1.5	2	1	NU	NU	low
8	0	0	fruiting 3	6	NU	NU	NU	high
9	15	15	fruiting 3	2	NU	NU	NU	low

Abbreviation: NU = Not used