

A fuzzy classification system to analyze the yield of cotton production

Raul Toscano-Miranda^{a,c}, Jose Aguilar^{b,c,e,f}, Manuel Caro^a, Anibal Trebilcok^d,
Mauricio Toro^c

^aDepartment of Educational Informatics, Universidad de Córdoba, Colombia

^bCEMISID, Universidad de Los Andes, Mérida, Venezuela

^cGIDITIC, Universidad Eafit, Medellín, Colombia

^dDepartment of Agronomic Engineering, Universidad de Córdoba, Colombia

^eUniversidad de Alcalá, Departamento de Automática, Spain

^fIMDEA Networks Institute, Leganés, Madrid, Spain

Abstract:

CONTEXT

Properly managing the cultivation of cotton is essential because it directly impacts the amount of cotton that is produced.

OBJECTIVE

The aim of this work is the proposal of a fuzzy classification system for diagnosis-prediction tasks of the cotton crop yield.

METHODS

We used a soft computing method to handle/describe experts' knowledge. Seven input variables (attack level of the red boll weevil, attack level of the black boll weevil, crop stage, rainfall, fertilizer, pheromone traps, and boll-weevil killing tube) were considered in the

system to analyze the cotton production.

RESULTS AND CONCLUSIONS

System tests were carried out on different agricultural scenarios, to determine their robustness and adaptability. According to the results, the fuzzy system has the capability to generate outputs that correspond with the experts' evaluations, which can be used to help farmers select the best practices in cotton crop management, in order to obtain the best yield in a specific context.

SIGNIFICANCE

The developed models enhance our capacity to predict crop yields based on climate data, the soil and pest behaviors, a valuable indicator for decision-making and overall sustainability.

Keywords: Fuzzy System; Classification System; Cotton; Yield; Diagnostic model; Predictive model

1. Introduction

Precision agriculture faces significant difficulties in accurately forecasting and diagnosing crop yield, making it one of the most demanding tasks. Currently, extensive research in agriculture is being carried out to enhance crop yield predictions and diagnoses, with the use of machine learning (ML) algorithms (Lobell et al., 2013; Obsie et al., 2020; Wang et al., 2018). The significance of predicting/diagnosing cotton yield in agricultural areas lies in its potential to enhance cotton crop management practices. Several techniques of Artificial Intelligence (AI) have been used to face the challenge to predict/diagnose crop yield. We focus on *fuzzy classification systems* (FCS). The use of FCS, a soft computing technique that replicates the way humans reason and make decisions, has become prevalent in modeling complex systems. To obtain FCS, experts are consulted to determine the input

variables of the system and define the cause-and-effect relationship among the variables using the 'if-then' rules (Cerrada et al., 2005).

FCS has been used in many different scientific fields for modeling and decision-making. For example, recently in transportation, for predicting the speed limits on Brazilian highways (Lanzaro & Andrade, 2022); in medicine, for diagnosing child anemia (Boadh et al., 2022); for predicting chronic kidney diseases (Hamedan et al., 2020); in business, for recommendation about consumer preference (Mandal et al., 2021); in social science, for improving students' learning performances (Hwang et al., 2020). On the other hand, expert systems have been proposed in agriculture for decision-making and decision-support tasks (Toscano-Miranda, Toro, et al., 2022). More specifically, expert systems have been developed and applied in different fields of agriculture to give advice and make management decisions (Mansour & Abu-Naser, 2019; Mendes et al., 2019; Salman & Abu-Naser, 2019).

On the other hand, for crop yield prediction, some studies have demonstrated the validity of using variables such as fertilizers (Mourhir et al., 2017; Prabakaran et al., 2018), and climatic data (Holzman & Rivas, 2016; Maskey et al., 2019; Obsie et al., 2020). Also, the variables rainfall, humidity, and temperature are useful to predict the attack level of the boll weevil (Toscano-Miranda, Hoyos, et al., 2022). However, so far there are no studies that integrate the above variables with the level of pest attack, crop stages, and techniques for pest monitoring and control. In this work, we included the variables pheromone traps and boll-weevil killing tube that are used to monitor and control the boll weevil, and the variable fertilizer that helps with soil conditions. All the previous variables are related to the right management at a specific crop stage, in this way, these variables together help to the diagnosis/prediction of crop yield.

This work aims to apply a soft computing technique, specifically FCS, to build a diagnosis/predictive model useful to infer the cotton yield in the smart agriculture context. The contributions of this work are the following:

- An adaptive model based on an FCS for the diagnosis/prediction of cotton yield, validated by experts, with a very low error rate;
- An approach for the management of uncertainty based on the fuzzy variables of the context;

- An integration scheme with a previous work (Toscano-Miranda, Hoyos, et al., 2022) that determines the classification of the population of pests (boll weevil) as the level attack ate the cotton crop, used as an input variable in this investigation;
- The simultaneous utilization of climatic variables, level attack of pests, information on fertilizers and crop stage, and techniques for the monitoring and control of pests. Our results show that the use simultaneous of these variables leads to good yield prediction/diagnosis.

The rest of the paper is organized as follows: Section 2 surveys the works related to crop yield prediction/diagnosis using AI. Section 3 outlines the development of the FCS for analyzing cotton yield. Section 4 presents the results of the proposed system, and Section 5 concludes this work by highlighting some of the future directions.

2. Related works

This section explores research that is relevant to the purpose of this study, namely, the use of AI for predicting/diagnosing crop yields. Papageorgiou et al. (2013) proposed a yield prediction in apples using a knowledge-based approach with a dynamic influence graph of *Fuzzy Cognitive Maps* (FCMs). The authors used main soil factors such as soil texture (clay and sand content), soil electrical conductivity, potassium, phosphorus, organic matter, calcium, and zinc. They classified apple yield using an efficient FCM learning algorithm, the non-linear Hebbian learning, and compared it with the conventional FCM tool and benchmark machine learning algorithms.

Maskey et al. (2019) investigated the correlation among various weather parameters related to strawberry yield at the field level, and evaluated yield forecasts using the predictive principal component regression (PPCR) and two ML techniques: (a) a single-layer neural network (NN) and (b) a generic random forest (RF). They used eight attributes: two wetness counts, two wetness minutes, the ambient, canopy and soil temperatures, and the volumetric moisture. Correlation analysis showed that all parameters were significantly correlated with strawberry yield, and provided the potential to develop weekly yield forecasting models. In general, the ML technique showed better skills in predicting strawberry yields when

compared to the principal component regression. More specifically, the NN provided the most skills in forecasting strawberry yield.

The climate data also were used by Ali et al. (2018) and Lobell et al. (2013). Ali et al. (2018) proposed a hybrid genetic programming model integrated with the Markov Chain Monte Carlo (MCMC) based Copula technique. The climate data included rainfall, mean monthly temperature, and mean monthly relative humidity, for the years 1981 to 2013. In a similar way, Lobell et al. (2013) used non-linear regression for the prediction of maize yield. The authors showed that there was a strong negative yield response to temperatures above 30 °C, and a relatively weak response to the seasonal rainfall.

In addition to climatic data, Obsie et al. (2020) used data generated by the Wild Blueberry Pollination Model to predict the best yield. Obsie and colleagues used Multiple linear regression (MLR), boosted decision trees (BDT), RF, and extreme gradient boosting (XGBoost). They found that clone size, honeybee, bumblebee, Andrena bee species, Osmia bee species, maximum upper-temperature ranges, and the number of days with precipitation, were the best predictor variables. The results showed that the XGBoost outperformed other algorithms in all measures of model performance for predicting the yield of wild blueberries.

On the other hand, Ranjan & Parida (2019) employed Sentinel-based both optical (Sentinel-2B) and SAR (Sentinel-1A) sensors data for paddy acreage mapping. The authors used RF classification technique for the paddy acreage mapping. A simple linear regression yield model was developed for predicting yields. Wang et al. (2018) also used remote sensing data. They applied Deep Transfer Learning for predicting soybean crop yields. The transfer learning was applied from Argentina's data to Brazil's data because Brazil had a smaller amount of data. They used the imagery of the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, which was processed with the long short-term memory (LSTM) recurrent NN.

Thus, there is a lot of interest in this area, as has been pointed out by (Klumpenburg et al., 2020). In the scope of this work, which seeks to develop an FCS for crop yield prediction/diagnosis, there are not many works. The closest work is that of Papageorgiou et al. (2011), which used FCM with variables related to the soil. However, our proposal shows the possibility of integrating other contextual variables that have not been included so far for

crop yield prediction/diagnosis. In addition, for the yield management of cotton crops, no papers were found.

Therefore, to our knowledge, there are no studies a) that simultaneously perform a diagnosis/prediction of cotton crop yields; b) that integrate different models of knowledge to characterize the context (in this work, our FCS with the work (Toscano-Miranda, Hoyos, et al., 2022)); c) that simultaneously use different types of variables (climate, behavior of pests, fertilizers, among others); d) that use adaptive models for the diagnosis/prediction of crop yields. In this way, our study focuses on these gaps.

3. Process for the Development of the FCS

This section presents the process to build the FCS for the diagnostic/prediction of cotton yield. For the development of the FCS was used the methodology proposed by (Buchanan, 1983) (see Fig. 1). This methodology has been previously used for implementing expert systems (Brüangel et al., 2019; Ele et al., 2014; Grüger et al., 2022).

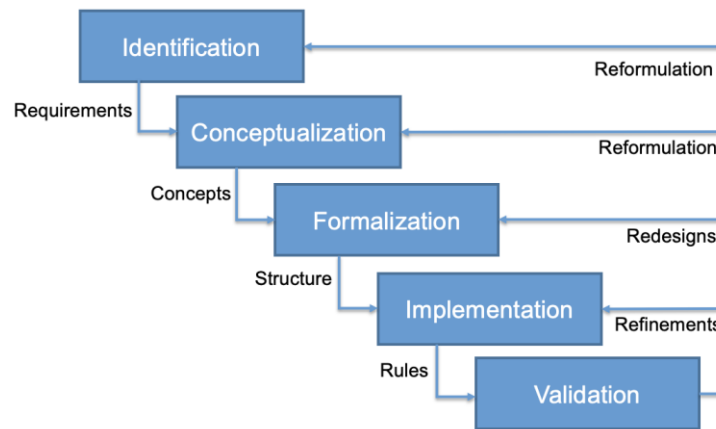


Fig. 1 Adaptation of Buchanan's model

First, let us introduce the concept of fuzzy logic, as it is used in this work. Fuzzy logic is one of the most useful AI techniques (Navinkumar et al., 2020), used to build FCS with the skill to simultaneously handle numerical data and linguistic knowledge (Mendel, 1995).

To build an FCS, it is necessary to define a rule set that represents the business logic. These rules are written in the next format: IF $\langle x \rangle$ THEN $\langle y \rangle$. With these rules, in our case, we can determine the conditions of the cotton crop to estimate the possible crop yield. We can carry out a diagnosis using the conditions of the cotton crops, and the estimation of the yield of the crop would be the prediction of its behavior. To define the fuzzy rules, it is necessary to define the fuzzy variables. Each fuzzy variable is defined by its fuzzy sets, with the boundaries given by the specific context.

Using these key concepts, the five stages of the methodology of C(Buchanan, 1983) in our case are: 1) Identification: In this stage, the characteristics of the cotton yield problem are determined, and the input/output variables were identified for the prediction/diagnosis tasks (this is described in detail in the Input Variable Analysis subsection of the next section). 2) Conceptualization: The fuzzy sets for each input/output variable are established (this is described in detail in the Membership Features subsection of the next section). 3) Formalization: The fuzzy rules for the prediction/diagnosis tasks are designed (this is described in detail in the Fuzzy Rules subsection). 4) Implementation: The results of the previous stages (conceptualizations and designs) are computationally implemented (in our case, we have used the scikit-fuzzy library of Python. This is described in the Defuzzification subsection). 5) Validation: Finally, at this stage, the FCS (its rules) are validated against expert opinion (this is described in detail in the Results section).

4. Case study

4.1 Contextualization

Cotton crops located in various regions of Córdoba, Colombia, were considered as the case study to validate our FCS. These regions involve cities that make up the Sinú Valley (High Sinú, Middle Sinú and Bajo Sinú) (Trebilcock, 2020). This area is located at $\sim 8^{\circ}55'33.6''\text{N}$, $75^{\circ}48'16.5''\text{W}$.

Independently of any other circumstance, the water supply associated with the planting date is determining to a high degree its effect on yields and the general behavior of the plant. However, the Sinú area has a very extended planting date due to different

circumstances, being the main crop rotation of the area with the cultivation of corn, which directly and indirectly is causing an effect on the population dynamics of pests, mainly for the boll-weevil, which makes it pertinent to study the collateral effects that this situation implies. The Sinú Valley area is characterized by a high relative humidity sustained during almost every month of the year, a situation that, together with the high temperature, planting date, rainfall, and population density, creates the conditions for the development of pathogens (Trebilcok, 2020).

This work uses the results of the previous work of (Toscano-Miranda, Hoyos, et al., 2022), which gives the classification of the population of pests (boll weevil) with the level attack at the cotton crop. The predictor variables are temperature, humidity, and rainfall. The prediction (i.e., the population of red and black boll weevils) is used as input variables of the FCS. In addition, in the current investigation, we use another 5 input variables: crop stage, rainfall, fertilizer, pheromone traps, and boll-weevil killing tube. In total, are 7 input variables to estimate the output variable (crop yield). Also, for the regions that make up the area to be planted, the space of time is assumed to be more than 60 days (Trebilcok, 2020).

The crop stages included in the FCS were vegetative, flowering, fruiting, harvesting, destruction of soca, and closing), and the date of planting the cotton in the study region is around August-September. There are several insect pests in the cotton crops, however, for this case study, we used data from boll weevil because of the attack severity in this region (Trebilcok, 2020).

Next, experts were surveyed to gather their viewpoints on various factors linked to the input variables and cotton yield. The information of the survey was used for the creation of the membership functions and fuzzy rules (see survey design subsection). Finally, the system's effectiveness was assessed using multiple performance metrics.

4.2 Analysis of input variables

A group of professionals was requested to assess the significance of multiple factors for the diagnosis/prediction of the cotton yield. Their evaluations were documented on a Likert Scale with five categories that range from weak influence to strong influence. So, the FCS was programmed with seven selected variables as inputs. These variables include the

attack level of the red boll weevil, the attack level of the black boll weevil, crop stage, rainfall, fertilizer, pheromone traps, and the boll-weevil killing tube.

For each input variable, the experts defined numerical ranges for the membership functions. The selection of these variables was based on experience in cotton crop management. The information for these variables was obtained by 1) A network of monitoring of boll weevils implemented by the *Colombian Agricultural Institute* (ICA), 2) Pheromone traps used in each cotton crop implemented by the owner, 3) Records of climate data from the *Institute of Hydrology, Meteorology and Environmental Studies* (IDEAM), 4) Records of climate data obtained in each cotton crop implemented by the owner, 5) Records of the management of each cotton crop implemented by the owner. 6) Crop yields in the area according to the *Colombian Cotton Confederation* (CONALGODON). The characterization of each input variable is as follows:

- Attack level of the red boll weevil: Population of the red boll weevil in the cotton crop. This data is processed and is the output of the previous work of (Toscano-Miranda, Toro, et al., 2022), which used an XGBoost algorithm for the classification of the population of the boll weevil into low (0 to 4), medium (5 to 20), and high (> 20). For this, they used climate data (temperature, humidity, and rainfall). The presence of the boll weevil must be monitored and controlled so that the attack level is zero.
- Attack level of the black boll weevil: It is similar to the red boll weevil, but in this case with blacks.
- Crop stage: Crop stage in the year. The cotton crop has the next stages: vegetative, flowering, fruiting, and harvesting. Then, it is necessary the destruction of soca, and during the closed season, another rotation crop is grown (in the context of the case study, it is usually corn).
- Rainfall: Amount of rain. Three linguistic variables were established: Low, Medium, and High. The desirable is Medium.
- Fertilizer: Amount of fertilizer used in the crop. Nitrogen, phosphorus, and potassium (NPK) fertilizers are used. For this case study, the measure used was bulks applied per hectare.
- Pheromone traps: Number of traps used in the crop. The traps are proportional to the area

of the lot monitored by an agronomist. The ICA has a minimum size of 1 trap per 250 hectares.

- Boll-weevil killing tube: Number of tubes used in the crop. The tubes are proportional to the area of the lot monitored by an agronomist.

4.3 Survey design

A set of questions was designed to collect expert assessments of the state of the crops and the probable cotton yield per hectare. The expert analyzed several scenarios about the crop yield of cotton. Scenarios are a combination of data (i.e., input variables) describing a real context and are used to validate the FCS. The scenarios were presented to the experts, and they adjusted the scenarios using linguistic variables. Each input variable has between two and six linguistic variables. Table 1 displays an overview of the input variables, their fuzzy sets, descriptions, and ranges. The next section contains a comprehensive explanation of the variables and ranges. To classify crop yields, experts in cotton cultivation were asked to sort them into low, medium, or high fuzzy sets, and specify a numeric value between 0 and 6 ton/ha, as is shown in Appendix I.

Table 1

Summary of the input variables.

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

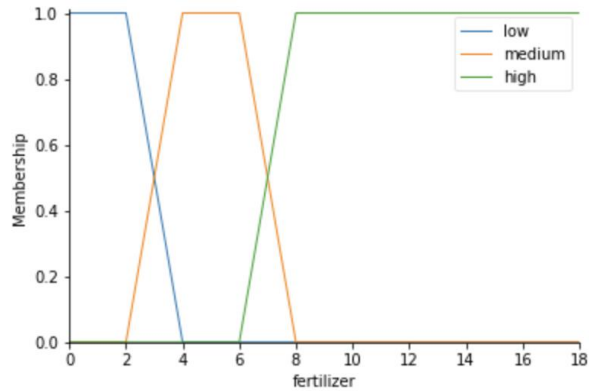
Nine scenarios were designed, which are defined by specific values for the input/output variables. Each variable was defined by a fuzzy value (e.g., low, medium, and high). Thus, each scenario has a different combination of situations for a crop yield waited (i.e., low, medium, or high). For example, one of the scenarios is: 15 red boll weevils (medium attack level), 15 black boll weevils (medium attack level), 0.5 in crop stage (vegetative), 17 mm of rainfall (high), 1 bulk of fertilizer (low), absent pheromone trap (0), and absent boll-weevil killing tube (0). This scenario triggers a specific set of rules from our FCS.

4.4 Membership functions

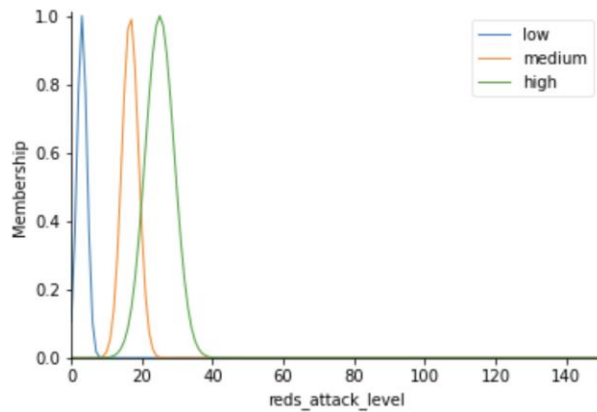
Each variable's fuzzy set was defined using triangular, trapezoidal, and Gaussian membership functions. The triangular function is used for the categorical variables, and the trapezoidal functions are used for the rests. Incorporating Gaussian functions into fuzzy systems enhances the ability to capture uncertainty and nonlinearity.

Linguistic variables, their corresponding fuzzy sets, and membership functions were

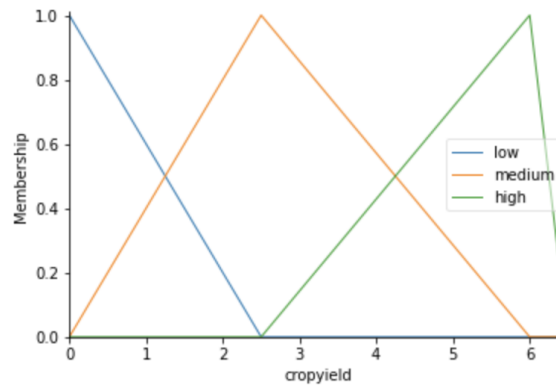
constructed based on the responses from experts. The membership functions were generated for all seven input variables and the output variable, using the same principles as for the input variables. The same considerations for the input variables were used for the output variable. Fig. 2 shows examples of different membership functions for several of the fuzzy variables of our problem.



A) Trapezoidal Member Function for the Fertilizer variable.



B) Gaussian Member Function for the reds_attack_level variable



C) Triangular Member Function for the Crop Yield variable

Fig. 2 Examples of membership functions for different variables of our model.

Finally, 13 membership functions were designed for the input and output variables (see Table 2). For the crop stage, pheromone trap, and boll-weevil killing tube variables, the Gaussian membership function was not used. This design decision is due to different reasons. In the case of the crop stage variable, the planting dates are established by the ICA, and therefore, this question was not asked of the experts. In the case of the pheromone trap and the boll-weevil killing tube, the options are absent or present. Thus, these three variables were not included in the survey, whence they do not have a mean and standard deviation to generate the Gaussian membership function.

Table 2
Membership functions for each fuzzy variable.

Variable	Membership functions	
Red attack level	Trapezoidal	Gaussian
Black attack level	Trapezoidal	Gaussian
Crop stage	Trapezoidal	
Rainfall	Trapezoidal	Gaussian
Fertilizer	Trapezoidal	Gaussian
Pheromone trap	Triangular	
Boll-weevil killing tube	Triangular	
Crop yield	Triangular	Gaussian

Particularly, the membership functions were defined using the viewpoints of experts obtained from the questionnaire. One example of the fuzzy membership functions for the fertilizer variable, an antecedent input, is illustrated in Fig. 3. Thus, Fig. 3 shows the membership functions of the fertilizer variable mentioned in Table 2.

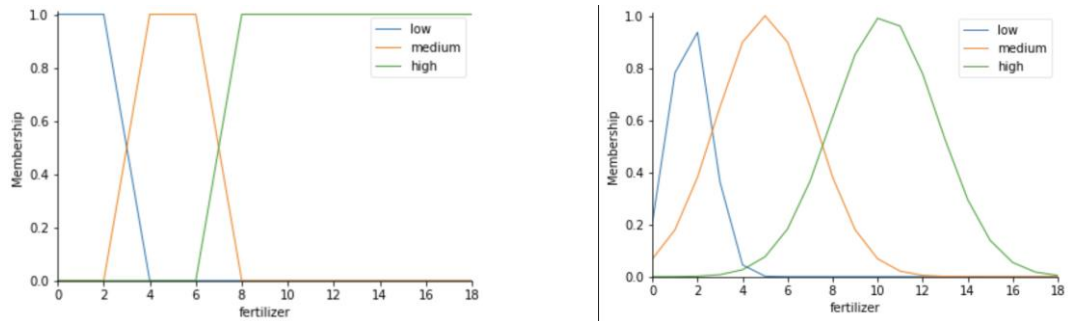


Fig. 3. Example of trapezoidal/Gaussian membership functions for one of the input variables (fertilizer).

4.5 Fuzzy rules

The fuzzy rules were generated using the responses of the experts. The rules are represented as IF-THEN. The input variables are antecedents and the crop yield is the consequent. According to the information in the antecedent, the consequent has a result. Table 3 shows two examples of the rules designed, which we use to show their behavior in the inference process using the different membership functions defined in the previous section. 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. Similarly, rule 2 has a different combination in the antecedent, and as result the crop yield is Medium. Finally, thirty-eight rules were designed for the system.

Table 3

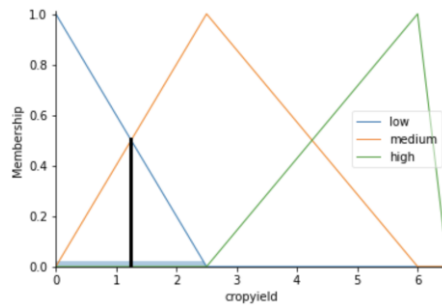
Rule structure (Example of two of them).

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

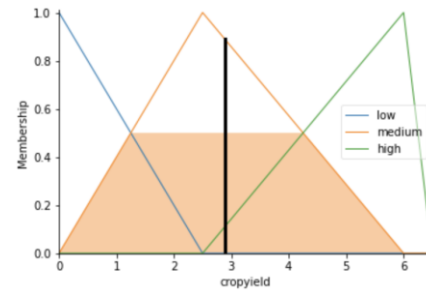
4.6 Defuzzification

The process of obtaining precise outputs (crop yield values) is known as defuzzification. In this subsection, we show two examples of the defuzzification process using the rules shown before, once the yield of the crop with said inputs is inferred. The center of gravity (CoG), also known as the centroid method, was adopted for defuzzification in this study (Cerrada et al., 2005). This approach uses the value of the output variables weighted by their membership functions. For example, Fig. 4.B shows the result of the defuzzification for one of the rules presented in Table 3, for a given set of inputs, with a result of prediction of Medium (2.88 ton/ha).

358



A. Crop yield of 1.23 ton/ha



B. Crop yield of 2.88 ton/ha

359

360

Fig. 4. Examples of defuzzification of the output variable (crop yield).

361

362

363

364

365

A standard fuzzy Mamdani system was developed using the scikit-fuzzy in Python, with a total of 38 if-then rules. Multiple sets of membership functions were examined during the FCS design process. The variables were set with triangular/trapezoidal/Gaussian membership functions according to Table 2.

366

5. Results

367

368

369

370

371

5.1 Determination of the optimal membership functions for each scenario

372

373

374

375

The experts were surveyed about the specific values on the low, medium, and high scales for some variables. The specific value is a number representing the scale. The mean and standard deviation were calculated for each one. Table 4 shows the survey results for each fuzzy variable.

376

377

Another group of experts was used to validate the information, and they were surveyed about the specific values on the low, medium, and high scales of some variables.

The specific value is a number representing the scale. The mean and standard deviation were calculated for each one. Table 4 shows the survey results for each fuzzy variable.

Table 4

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 mean and standard deviation were used to generate the Gaussian shape in the membership function. For each scenario, a triangular/trapezoidal or Gaussian membership function combination was used. These combinations were for five variables: red weevil attack level, black weevil attack level, precipitation, fertilizer, and crop yield. The other three variables (crop stage, pheromone trap, and weevil elimination tube) only have a triangular/trapezoidal membership function. In this sense, 32 possibilities (2^5) were generated for each scenario. In total were generated 288 combinations (9 scenarios x 32 possibilities). Table 5 shows the best performance using the best combination of the membership function for each scenario. In some cases, the trend was triangular/trapezoidal (e.g., scenarios 1 and 6), while in other cases it was Gaussian (e.g., scenario 9). In general, the FCS results are very close to expert opinion (see the last two columns).

Table 5

Evaluation of the best combination of membership functions.

Scenario	Membership Function								Expert System	Mean Expert
	Input							Output		
1	T	T	T	G	T	T	T	T	1,236	1,366
2	T	T	T	G	T	T	T	T	1,236	1,633
3	G	G	T	T	G	T	T	T	2,820	2,666
4	G	G	T	T	G	T	T	G	3,831	4
5	G	G	T	T	G	T	T	G	3,831	4
6	T	T	T	T	T	T	T	T	2,880	2,766
7	G	G	T	T	G	T	T	G	1,655	1,5
8	G	G	T	T	G	T	T	G	3,831	4
9	G	G	T	G	G	T	T	G	1,917	1,833

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

Finally, the results of the validation experts were compared with those of the experts who defined the initial rules for the FCS. The comparison is explained in the following subsection.

5.2 Evaluation of the estimation capabilities of our FCS

A test was carried out with the best models (formats of the membership functions, see Table 5) for each scenario, to check whether the results (crop yield) matched the crop yield established by the experts. For this test, different values of the input variables describing the different scenarios were evaluated (more than 50.000 entries). The results (i.e., FCS outputs) were then compared to the crop yield provided by the experts. To do this, the mean of the responses for each scenario was calculated to arrive at a sole crop yield for each scenario, which was then compared to the crop yield for each scenario established by the experts that

defined the initial rules for the FCS. Thus, the difference between the answers of the experts and the results of the FCS was evaluated.

Three measures were used to evaluate the performances of the estimations of our FCS (see Table 6). We used the *Coefficient of Determination* (R^2), which refers to the percentage of the variability in the dependent variable that can be explained by the independent variables. On the other hand, the *Mean Squared Error* (MSE) was used to evaluate the dissimilarity between the expert and predicted values. In addition, the *Mean Absolute Error* (MAE) was used to calculate the mean absolute difference between the experts and predicted values. R^2 ranges from 0 to 1. R^2 values that approach 1 are desirable, it means that the FCS obtained a very good result. For the MSE and MAE, the values are always positive and a value closer to 0 or a lower value is better. In Table 6, we can see that the error metrics MSE and MAE are very small, which means very good results.

Table 6

Evaluation of estimates

R^2	MSE	MAE
0.9374	0.0661	0.2154

6. Discussion of results

This study defines a FCS with adaptive models (membership functions) for the diagnosis/prediction of cotton yield, validated by experts, with a very low error rate. In this study, we take into account the factors that influence crop yields; for example, the presence of insect pests related to the climate (Toscano-Miranda, Hoyos, et al., 2022; Trebilcok, 2020; Villarreal et al., 2005), the soil conditions related to the fertilizers (Mourhir et al., 2017; Papageorgiou et al., 2013), and other techniques to monitoring and control the insect pests (pheromone traps, boll weevil killing tube). Mendes et al. (2019) and Navinkumar et al. (2021) used AI techniques to predict the level of water needed to improve crop yields. Thus, we include the analysis of rainfall levels (low, medium, and high) according to the needs of

cotton crops at a specific crop stage. As for fertilizers, we include rules that consider their efficient use according to soil analysis and crop stage. Moreover, according to Prabakaran et al. (2018), seasonal factors and pest incidence determine yield and their study can be used to maximize crop production. Thus, we use also these variables. Our results showed that the integration of all these variables is useful for the diagnosis/prediction of cotton crop yield.

Papageorgiou et al. (2013) used FCM and soil variables to predict crop yield. Thus, we found that fuzzy logic can be used to model the experts' knowledge and thus manages the uncertainty of the variables of the context. Uncertainties are unavoidable, such as, for example, climatic data, which we took into account in our proposal.

Unlike Papageorgiou et al. (2013), we use the data generated from expert opinion to fit the model (particularly, the membership function). Also, compared with variables, the approach proposed in the current study uses multi variables that cover different aspects of the crop (pests, stage of the crop, etc.). Other studies used techniques like single-layer neural networks (Maskey et al., 2019), extreme gradient boosting (Obsie et al., 2020), and Random Forest (Ranjan & Parida, 2019), but no one of them use adaptative models. Nor do they consider the data generated from expert opinion to fit the model.

In summary, in Table 7 we present four aspects of comparison. First, whether they proposed adaptive models for diagnosis/prediction. Second, whether they managed the uncertainty from the variables of the model or the context. Third, whether they considered the data that is generated from the opinion of the experts to adapt the model. Fourth, whether they simultaneously used variables of the climate, pests, fertilizers, and crops. As we discussed before, our approach is the first to combine these criteria to propose a diagnostic/predictive approach, which can be extended with the multi-agent systems paradigm (Aguilar et al., 2007; Terán et al., 2017) to give it more adaptability, extendibility and autonomy (Vizcarrondo et al., 2017) to the system.

Table 7
Comparison with other works.

Work	Adaptive model	Uncertainty	Data expert	CLFCT
Papageorgiou et al. (2013)	X	X	X	
Maskey et al. (2019)				
Ali et al. (2018)		X		
Lobell et al. (2013)				
Obsie et al. (2020)		X		
Ranjan & Parida (2019)				
This work	X	X	X	X

Abbreviation: CLFCT= Simultaneous use of Climatic, of pests, of Fertilizer, and of Crop variables.

Lastly, environmental variables such as soil organic matter, weed coverage percentage, and tillage system management, which were not explored in this study, might also explain the diagnosis/prediction of cotton yield observed in this study.

7. Conclusions

An FCS for the diagnosis/prediction of cotton crop yield is proposed in this work. The results showed that the system manages the experts' knowledge very well. This FCS has integrated with the results of previous work and different input variables were considered in the system (climates, pests and crops, among others). Results show that the fuzzy system can generate outputs that align with the assessments made by experts, and the metrics to evaluate the FCS show small errors. Our FCS can be used to help farmers select the best practices in crop management, to obtain the best yield in a specific context.

We used the opinion of experts in the crop, management, and marketing of cotton, which allowed defining good rules that achieve a diagnosis/prediction of the yield of the cotton crop, relating some characteristics of the cotton with the study area. In addition, the model can adapt the fuzzy rules and handle uncertainty. A limitation of the current work refers to the general use of the amount of fertilizer. We did not include each specific fertilizer

class. This could be overcome by specifying the amount of each fertilizer, and adding them to the model.

We have plans to integrate this study and other knowledge models into a metacognitive architecture for smart agriculture that operates as an autonomous cognitive system. Furthermore, we will use the transfer learning paradigm to define models of boll-weevil behavior for various regions through a meta-learning task. In addition, we will extend the FCS with the skill of adaptability using data in real-time. This study did not include other environmental variables such as soil organic matter, weed coverage percentage, and tillage system management, which can be considered in future works. Finally, these models will be integrated into our cognitive architecture using the multi-agent systems approach, leveraging the modeling abilities and implementations within agent theory.

Acknowledgments

We thank the Colombian Agricultural Institute (ICA), Colombian Cotton Confederation (CONALGODON) and Colombian Institute of Hydrology, Meteorology and Environmental Studies (IDEAM).

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

- Aguilar, J., Cerrada, M., & Hidrobo, F. (2007). A Methodology to Specify Multiagent Systems. *Lecture Notes in Computer Science*, 4496, 92–101.
- 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 programming algorithm: A new hybrid copula-driven approach. *Agricultural and Forest Meteorology*, 263, 428–448. <https://doi.org/10.1016/j.agrformet.2018.09.002>
- Boadh, R., Chaudhary, K., Dahiya, M., Dogra, N., Rathee, S., Kumar, A., & Rajoria, Y. K. (2022). Analysis and investigation of fuzzy expert system for predicting the child anaemia. *Materials Today: Proceedings*, 56, 231–236. <https://doi.org/10.1016/j.matpr.2022.01.094>
- Brüngel, R., Rückert, J., Wohlleben, W., Babick, F., Ghanem, A., Gaillard, C., Mech, A., Rauscher, H., Hodoroaba, V. D., Weigel, S., & Friedrich, C. M. (2019). NanoDefiner e-tool: An implemented decision support framework for nanomaterial identification. *Materials*, 12(19). <https://doi.org/10.3390/ma12193247>
- Buchanan, B. G. (1983). Constructing an expert system. *Building Expert Systems*, 127–167.
- 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>
- Ele, S., Umoh, E., & Adesola, W. (2014). An overview of the development principles, stages and building blocks of Expert System. *West African Journal of Industrial & Academic Research*, 11(1), 44.
- Grüger, J., Geyer, T., Bergmann, R., & Braun, S. A. (2022). CGK4PM: Towards a Methodology for the Systematic Generation of Clinical Guideline Process Models and

the Utilization of Conformance Checking. *BioMedInformatics*, 2(3), 359–374.
<https://doi.org/10.3390/biomedinformatics2030023>

Hamedan, F., Orooji, A., Sanadgol, H., & Sheikhtaheri, A. (2020). Clinical decision support system to predict chronic kidney disease: A fuzzy expert system approach. *International Journal of Medical Informatics*, 138. <https://doi.org/10.1016/j.ijmedinf.2020.104134>

Holzman, M. E., & Rivas, R. E. (2016). Early Maize Yield Forecasting from Remotely Sensed Temperature/Vegetation Index Measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(1), 507–519.
<https://doi.org/10.1109/JSTARS.2015.2504262>

Hwang, G. J., Sung, H. Y., Chang, S. C., & Huang, X. C. (2020). A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Computers and Education: Artificial Intelligence*, 1. <https://doi.org/10.1016/j.caeai.2020.100003>

Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. In *Computers and Electronics in Agriculture* (Vol. 177). Elsevier B.V. <https://doi.org/10.1016/j.compag.2020.105709>

Lanzaro, G., & Andrade, M. (2022). A fuzzy expert system for setting Brazilian highway speed limits. *International Journal of Transportation Science and Technology*, xxxx.
<https://doi.org/10.1016/j.ijtst.2022.05.003>

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>

Mandal, M., Mohanty, B. K., & Dash, S. (2021). Understanding consumer preference through fuzzy-based recommendation system. *IIMB Management Review*, 33(4), 287–298. <https://doi.org/10.1016/j.iimb.2021.03.015>

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

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

569 Mendel, J. M. (1995). Fuzzy logic systems for engineering: a tutorial. *Proceedings of the*
570 *IEEE*, 83(3), 345–377. <https://doi.org/10.1109/5.364485>

571 Mendes, W. R., Araújo, F. M. U., Dutta, R., & Heeren, D. M. (2019). Fuzzy control system
572 for variable rate irrigation using remote sensing. *Expert Systems with Applications*, 124,
573 13–24. <https://doi.org/10.1016/j.eswa.2019.01.043>

574 Mourhir, A., Papageorgiou, E. I., Kokkinos, K., & Rachidi, T. (2017). Exploring precision
575 farming scenarios using fuzzy cognitive maps. *Sustainability (Switzerland)*, 9(7).
576 <https://doi.org/10.3390/su9071241>

577 Navinkumar, T. M., Kumar, R. R., & Gokila, P. v. (2021). Application of artificial
578 intelligence techniques in irrigation and crop health management for crop yield
579 enhancement. *Materials Today: Proceedings*, 45, 2248–2253.
580 <https://doi.org/10.1016/j.matpr.2020.10.227>

581 Navinkumar, T. M., Ranjith Kumar, R., & Gokila, P. V. (2020). Application of artificial
582 intelligence techniques in irrigation and crop health management for crop yield
583 enhancement. *Materials Today: Proceedings*, xxxx.
584 <https://doi.org/10.1016/j.matpr.2020.10.227>

585 Obsie, E. Y., Qu, H., & Drummond, F. (2020). Wild blueberry yield prediction using a
586 combination of computer simulation and machine learning algorithms. *Computers and*
587 *Electronics in Agriculture*, 178. <https://doi.org/10.1016/j.compag.2020.105778>

588 Papageorgiou, E. I., Aggelopoulou, K. D., Gemtos, T. A., & Nanos, G. D. (2013). Yield
589 prediction in apples using Fuzzy Cognitive Map learning approach. *Computers and*
590 *Electronics in Agriculture*, 91, 19–29. <https://doi.org/10.1016/j.compag.2012.11.008>

591 Papageorgiou, E. I., Markinos, A. T., & Gemtos, T. A. (2011). Fuzzy cognitive map based
592 approach for predicting yield in cotton crop production as a basis for decision support
593 system in precision agriculture application. *Applied Soft Computing Journal*, 11(4),
594 3643–3657. <https://doi.org/10.1016/j.asoc.2011.01.036>

- Prabakaran, G., Vaithiyanathan, D., & Ganesan, M. (2018). Fuzzy decision support system for improving the crop productivity and efficient use of fertilizers. *Computers and Electronics in Agriculture*, 150, 88–97. <https://doi.org/10.1016/j.compag.2018.03.030>
- Ranjan, A. K., & Parida, B. R. (2019). Paddy acreage mapping and yield prediction using sentinel-based optical and SAR data in Sahibganj district, Jharkhand (India). *Spatial Information Research*, 27(4), 399–410. <https://doi.org/10.1007/s41324-019-00246-4>
- Salman, F. M., & Abu-Naser, S. (2019). Expert System for Castor Diseases and Diagnosis. *International Journal of Engineering and Information Systems (IJEAIS)*, 3(3), 1–10.
- 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. *2022 XLVIII Latin American Computer Conference (CLEI)*, 1–5.
- 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 algodón en Córdoba*.
- Villarreal, N., Lobatón, V., Galindo, J., & Mondragón, V. (2005). *Monitoreo del Picudo del Algodonero Anthonomus grandis Boheman en Colombia*.
- Vizcarrondo, J., Aguilar, J., Exposito, E., & Subia, A. (2017). MAPE-K as a service-oriented Architecture. *IEEE Latin America Transactions*, 15(6), 1163–1175.
- Wang, A. X., Tran, C., Desai, N., Lobell, D., & Ermon, S. (2018, June 20). Deep transfer learning for crop yield prediction with remote sensing data. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018*. <https://doi.org/10.1145/3209811.3212707>

Appendix I. Surveys used for the development of the FCS.

The purpose of this study is to implement an FCS to diagnose/predict the cotton-crop yield in Córdoba, Colombia. This is useful to determine the actions to implement in the future to maximize production. Your viewpoints on various cotton crop scenarios should be expressed in this survey. Nine scenarios will be presented, related to the following variables:

- Attack level of red boll weevil (Low, Medium, and High)
- Attack level of black boll weevil (Low, Medium, and High)
- Crop stage (Vegetative, Flowering, Fruiting, Harvesting, Destruction of soca, and Closing)
- Rainfall (Low, Medium, and High)
- Fertilizer (Low, Medium, and High)
- Pheromone traps (Absent, Adequate)
- Boll-weevil killing tube (Absent, Adequate)

At the conclusion, it's essential to declare whether the suggested output should be classified as low, medium, or high, and designate an integer between 0 and 6 ton/ha as the yield. For example:

Scenario 1

VEGETATIVE STAGE Case 1 - low yield

- Attack level of red boll weevil: High
- Attack level of black boll weevil: High
- Crop stage: Vegetative
- Rainfall: High
- Fertilizer: Low
- Pheromone traps: Absent

652 • Boll-weevil killing tube: Absent

653

654

655 Note 1: The structure is recurrent in all 9 scenarios, but the variable values are
656 modified for each individual case.

657

658 **Ending comments**

659 Are you interested in adding another scenario, not previously mentioned, what is
660 essential for diagnosing the yield?