Computing (2018) 100:689–713
https://doi.org/10.1007/s00607-018-0594-9

Fuzzy based approach for discovering crops plantation knowledge from huge agro-climatic data respecting climate changes

Assem H. Mohammed[1] · Ahmed M. Gadallah[1] · Hesham A. Hefny[1] · M. Hazman[2]

Received: 7 August 2017 / Accepted: 20 February 2018 / Published online: 2 April 2018

Abstract Climate change has noticeable significant impacts on development of most countries because of its direct negative effect on the production and revenue of most crops plantation process. In reality, the ongoing changes in climate variables affect the suitability of planting some crops in their traditional places at their traditional dates. Furthermore, the availability of huge volumes of agro-climatic data that almost incorporates uncertainty increases the complexity of managing and discovering the crops suitable plantation patterns from such data. Accordingly, a need appeared to an efficient approach to handle such uncertainty and to exploit such huge data volume to manage the crops plantation process accurately. This paper presents a fuzzy approach based on Hadoop for discovering crops plantation knowledge from the agro-climatic historical database of the years from 1983 to 2016 of Egypt. Commonly, the proposed approach provides a set of scenarios for plantation dates of each crop with a suitability degree for each scenario. Also, it helps managing crops plantation process from some other aspects such as harvesting dates, candidate diseases and follow up for crops water requirements respecting the data streaming of the prevailing weather data. The proposed approach has been tested on a set of crops with cooperation of researchers

✉ Assem H. Mohammed
eng_asemhm@yahoo.com

✉ Ahmed M. Gadallah
ahmgad10@yahoo.com; a.gadallah@cu.edu.eg

Hesham A. Hefny
hehefny@ieee.org

M. Hazman
maryam.hazman@gmail.com

[1] Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt

[2] Central Laboratory of Agricultural Expert Systems, Agricultural Research Center, Cairo, Egypt

from Cairo University and Agricultural Research Center. The results show the added value of the proposed approach against other works respecting the more suitable crops plantation dates, harvesting dates, expected diseases and follow up for crops water requirements. Furthermore, the proposed approach benefits from Hadoop framework capabilities of handling huge amounts of data streamed from weather stations.

Keywords Fuzzy set theory · Climate change · Crops plantation · Agro-climatic data · Hadoop · Prediction

Mathematics Subject Classification 94D05 (Fuzzy logic; logic of vagueness)

1 Introduction

Generally, climate is considered one of the most important factors in the crops growth process. Actually, each crop needs specific climate conditions for the growth of each stage [1]. So, climate changes are projected to have direct effect on the crops yielding. The ongoing changes in climate variables like temperature, humidity and sun shining make some crops not suitable for planting in its traditional places at its traditional dates. Reasonably, such plantation dates may become suitable by shifting it up or down some days or it may become not suitable at all. Also, according to the climate changes, new dates may become ideal for planting a specific crop although such new dates were not suitable before [2, 3]. Because the imprecision nature of the climate requirements of crops plantation, harvesting and candidate diseases emergency, a need appeared to an efficient approach to cope with such imprecision like fuzzy logic. On the other hand, the huge amount of historical climate data is used in predicting the incoming climate variable values in the underlined incoming period. Also, the ongoing weather data streaming through the crop growth life time are essential for the follow up process of crops in both water requirements and harvesting date. Accordingly, such huge amount of data and the ongoing climate data streaming need a reliable framework to be handled such as Hadoop framework.

Commonly, fuzzy logic is a soft computing approach based on degrees of truth rather than the old traditional Boolean logic that allows only two values: true or false, on which the modern computer is built. Fuzzy sets calibrate vagueness by allowing members to belong partially or totally to them. The main advantage of fuzzy logic when compared with traditional logic is its ability to deal with imprecise data and inexact knowledge [4]. Fuzzy logic had been initiated by Zadeh at 1965 and since then many applications and researches in different fields have been achieved [4–6]. One of such applications is fuzzy query which has been used to cope with the necessity to soften the Boolean logic when querying database [7]. A fuzzy query system is an interface for users to retrieve information from a database using human linguistic words [8].

Hadoop is an open source framework for distributed processing and storage of huge data sets on computer clusters [9]. It provides massive storage for any data type, high performance and the ability to virtually perform unlimited concurrent jobs [10]. As shown in Fig. 1, Hadoop ecosystem consists of a set of tools to deal with different stages of data processing. Hive is part of Hadoop ecosystem for data warehousing. It

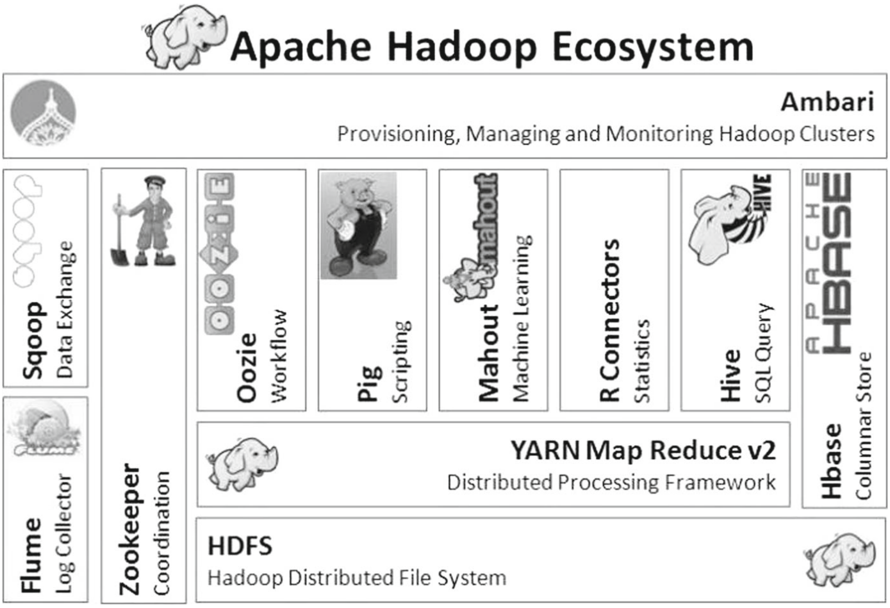

Fig. 1 Hadoop ecosystem [41]

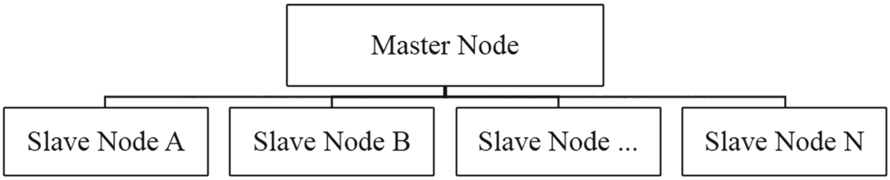

Fig. 2 Hadoop master slave architecture

provides data querying, analysis and summarization using SQL like query language [7, 11]. Consequently, all implemented Hive tasks are converted by the Hive engine into a set of Map Reduce tasks [12].

Commonly, the cluster of Hadoop distribute file system consists of one namenode (the master) and a number of datanodes (slaves). The namenode manages the filesystem namespace as shown in Fig. 2. On the other hand, slaves are used to perform jobs, store and retrieve data blocks in response to a request by the master node. Also, they report back to the master periodically with lists of data blocks that they are storing [13].

This paper proposes a fuzzy-based approach that predicts the more suitable plantation dates, harvesting dates, candidate diseases for specific crop at specific location given the agro-climatic historical data respecting climate changes. On the other hand, it allows a follow up process for water requirements of the crops growth life time.

The rest of this paper is organized as follows: the second section introduces the related works. The proposed approach is presented in the third section. The fourth section shows the illustrative case study. The results and findings are presented in the fifth section. The sixth section shows the conclusion and future works.

2 Related work

Many approaches have been proposed for the problem of the effects of climate changes on agriculture. Some approaches aimed mainly to show the impacts of climate change on the crop production process like [1, 14]. As water resources are one of the most important parameters in most crops plantation process, some approaches have been developed to make adaptation to crop planting dates with climate changes like in [15, 16]. Another approach for crop yield forecasting was presented in [17] to map the relationships between climate data and crop yielding. This approach is based on time series data of both crops requirements and weather data. The works presented in [18, 19] represent fuzzy-based decision support systems for evaluating land suitability for planting a specific crop and selecting the more suitable crops to be planted in a given area. Also, a recommendation approach for wheat plantation based on the average values of monthly climate data was presented in [20]. On the other hand, a set of approaches based on fuzzy set theory were presented in [21–23]. Such approaches predict the more suitable plantation dates of a set of crops respecting climate changes that happened in the last five years in some areas of Egypt.

Commonly, each of these previous works attempts to tackle, at least for some extent, the problem of crops planting and yielding respecting climate change. Unfortunately, some of these works don't provide weights or matching measures for the resulted solutions. Some other approaches depend on the average values of climate data which are erroneous with most of crops that have mini-mum and maximum suitable values of climate variables. Also, most of related works do not take care of expecting the crops harvesting dates that are very crucial to prevent the crop yield from damage. On the other hand, there still a need for searching for the candidate diseases that might emergence during the recommended plantation dates for a given crop in a specific area. Finally, by entering the era of big data all of such works need to be extended to benefit from the available ongoing data streaming to enhance the overall plantation process respecting the changes in climate and the water resources in many areas of the world [24].

Accordingly, this work proposes a fuzzy-based approach for adapting crops plantation dates, predicting the harvesting date and the candidate diseases regarding the ongoing climate changes. Based on the available historical agro-climatic data for Egypt, the proposed approach discovers new plantation dates that are more suitable than old ones for a set of underlined crops. Consequently, it discovers the expected crops diseases for the selected plantation dates. Also, it expects the harvesting dates. On the other hand, the proposed approach performs a follow up for water requirements of crops that have been sowed using the actual data streamed from weather stations.

3 The proposed approach

The proposed approach benefits from the flexibly of fuzzy set theory to deal with uncertainty that exists in the requirements of most crops plantation process. Also, it uses Hadoop Hive as a platform to enhance the performance and processing of agro-climatic and crops databases as shown in in Fig. 3. It aims mainly to discover the

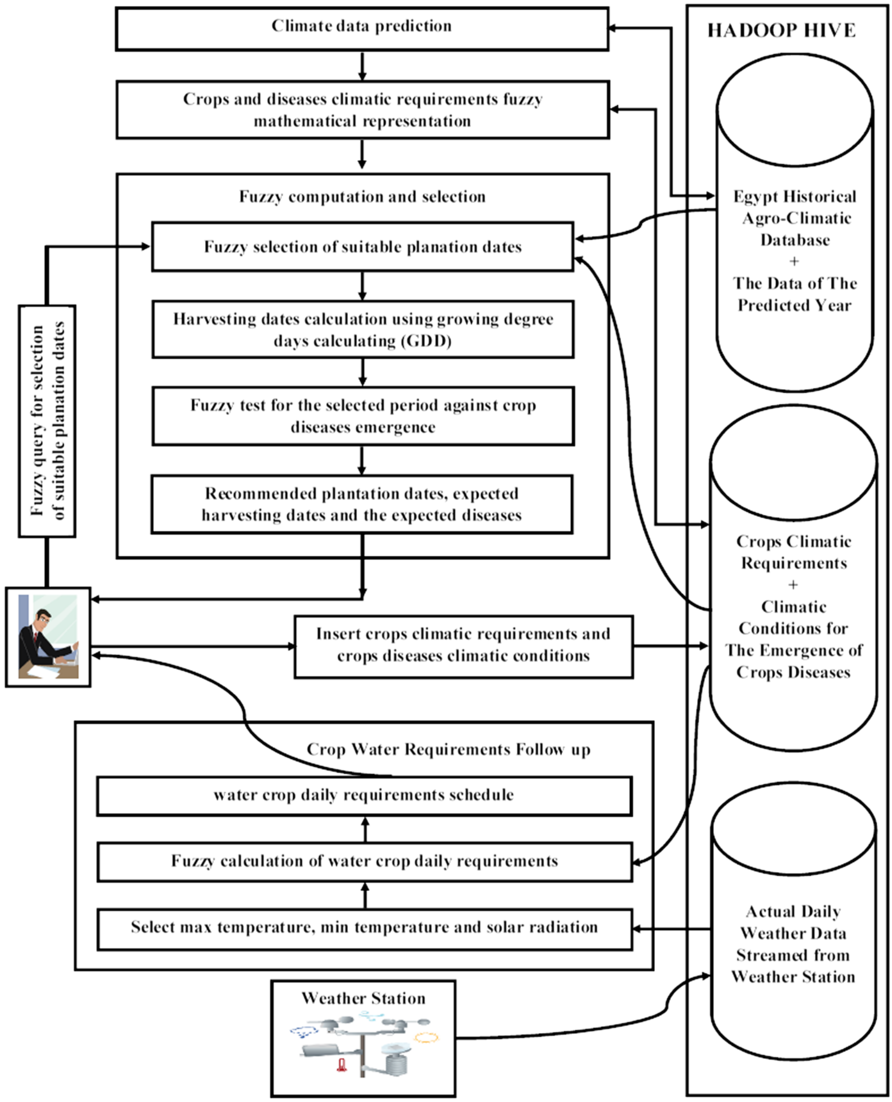

Fig. 3 The proposed approach architecture

more suitable plantation dates, harvesting dates, expected diseases and the crops water requirements. Firstly, the proposed approach uses the historical climate data of Egypt to predict the next year climate data. Secondly, it receives the climatic requirements for crops from experts or from database to form a suitability fuzzy climatic pattern for each crop. Thirdly, it searches the agro-climatic data of the predicted year to find the more suitable plantation dates, harvesting dates and candidate diseases for a given crop. Finally, the proposed approach performs a crops water requirement follow up for crops that have been sowed by using the data streamed from weather stations. Thus,

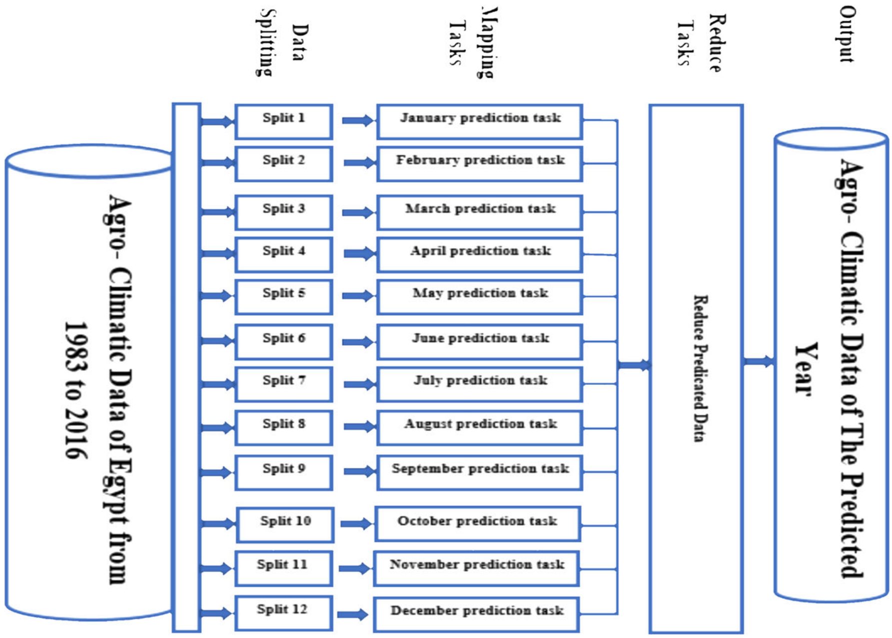

Fig. 4 The data prediction framework

the proposed approach determines, in a dynamic fashion, how much a planted crop needs the water to which extent.

3.1 Prediction for climate data phase

In this phase, the climate data of Egypt for the years from 1983 to 2016 obtained from the Central Laboratory for Agricultural Climate are used to predict the next year using Hadoop MapReduce technique [25]. The climate prediction technique that is used is called climatology [26]. As shown in Fig. 4 the prediction process consists of a set of steps. Firstly, the historical data are divided to twelve classes each class represents only one month of the year. Secondly, the daily data for the corresponding months are passed to a map task to predict climate data of the same month of the future under-prediction year(s) day by day using Eq. (1). Finally, the data of each predicted month is grouped into only one class that presents the predicted climate data of the future predicted year(s).

$$PCVV_d = \frac{\sum_{y=1}^{n} ACVV_{dy}}{n} \tag{1}$$

where $PCVV_d$ represents the predicted value of climate variable of day d, $ACVV_{dy}$ denotes the actual climate variable value of day d at year y in the historical climate data and n is the number of historical years of the climate data.

For example, the predicted value of the temperature at 20 December 2017 calculated as following: PT p $_{20\ \mathrm{December}\ 2017}$ (Temperature) = (value of temperature at 20 December 1983 + value of temperature at 20 December 1984 + value of temperature at 20 December 1985 +……… + value of temperature at 20 December 2016)/33.

3.2 Fuzzy mathematical representation and processing for plantation related data phase

This phase is responsible for gathering the crops climatic requirements and crops diseases climatic emergence conditions from agricultural organizations, resources and experts. After that, such climatic data requirements are mathematically represented depending on the data nature.

3.2.1 Fuzzy mathematical representation for crops climatic requirements

The crops climatic requirements are gathered from the agricultural experts and from a set of agricultural resources including [25, 27–29]. After that such climatic requirements are used in defining the suitability fuzzy membership functions as shown in Fig. 5. Such fuzzy membership functions are used to measure the suitability of each day of the plantation period of each crop.

Generally, the gathered data for crops climatic requirements have optimum values that range between two values, around specific value, less than specific value or more

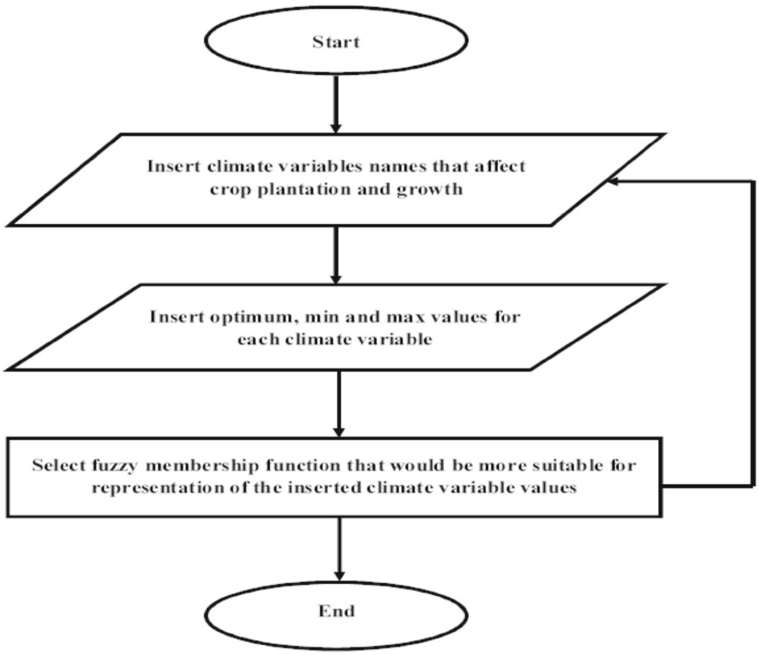

Fig. 5 A flow chart that represents requirements gathering and representation

than specific value. Accordingly, using fuzzy sets to represent such uncertain requirements are rational. Since, most of such climatic requirements have range of values with full matching degrees, so trapezoids and truncated trapezoids are the best fuzzy membership functions to represent them [30].

For each climatic variable, representing its suitability membership functions for any growth stage of any crop is represented using one of the following cases.

Case 1 If the optimum suitable values of a climatic variable for a crop growth are between two values then the trapezoidal fuzzy membership function depicted in Eq. (2) is used to represent such suitability.

Case 2 If the optimum suitable climatic variable values required for crop growth are around specific value then the triangular fuzzy membership function presented in Eq. (3) is used to represent such suitability.

Case 3 If the optimum suitable climatic variable values required for crop growth are more than a specific value then the left shoulder triangular fuzzy membership function depicted in Eq. (4) is used to represent such suitability.

Case 4 If the optimum suitable climatic variable values that are required for crop growth are less than a specific value then the right shoulder triangular fuzzy membership function presented in Eq. (5) is used to represent such suitability.

$$\mu_{CV_i}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & b \le x \le c \\ \frac{d-x}{d-c} & c < x < d \\ 0 & x \ge d \end{cases} \tag{2}$$

where $\mu_{CV_i}(x)$ represents the suitability of climate variable i for planting the specified crop, x is the climate variable value of a specific day, a and d are minimum and maximum threshold values such that any value less than a or greater than d has zero matching degree; while b and c represents the boundaries of the optimum range of climate variable values for crop growth; any value belongs to the range $]a, b[$ has a partial suitability membership value in $]0, 1[$.

$$\mu_{CV_i}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & x = b \\ \frac{c-x}{c-b} & b < x < c \\ 0 & x \ge c \end{cases} \tag{3}$$

where $\mu_{CV_i}(x)$ represents the suitability of climate variable i for planting the specified crop, x is the climate variable value of specific day, any climate variable value less than

a has zero suitability degree for crop growth, c is maximum climate variable value for crop growth, and b is optimum climate variable value for crop growth.

$$\mu_{CV_i}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & x \ge b \end{cases} \tag{4}$$

where $\mu_{CV_i}(x)$ represents the suitability of climate variable i for planting the specified crop, x is the climate variable value of specific day any climate variable value less than a has zero suitability degree for crop growth, and climate variable values greater than or equal b have full suitability degree for crop growth.

$$\mu_{CV_i}(x) = \begin{cases} 1 & x \le a \\ \frac{x-b}{a-b} & b \le x < a \\ 0 & x > b \end{cases} \tag{5}$$

where $\mu_{CV_i}(x)$ represents the suitability of climate variable i for planting the specified crop, x is the climate variable value of specific day, any climate variable value less than or equal to a has full suitability degree for crop growth, and any climate variable value greater than b has zero suitability degree for crop growth.

For example, assuming that the optimum temperature degree for planting a given crop in an area is between 20 and 25° C and the minimum and maximum temperature are 15 and 35° C respectively. Accordingly, the best mathematical representation would be achieved using the Trapezoidal fuzzy membership function presented in Eq. (2) with values 15, 20, 25 and 35 for the function variables a, b, c and d.

3.2.2 Fuzzy mathematical representation for diseases emergence climatic conditions

As presented in Sect. 3.2.1, the emergence climatic conditions of crops diseases are mathematically represented in the same fuzzy-based way as the crops growth climatic requirements are represented. The data of crops diseases is collected from agricultural experts and some related references like [30–33].

3.3 Fuzzy suitability computation and selection phase

Commonly, this phase represents the main phase in the proposed approach and it includes three sub phases. The first sub phase is concerned with selecting the more suitable plantation dates. Consequentially, in the second sub phase, the harvesting dates of the crops that will be planted in such selected plantation dates are calculated respecting the changes in the climate data of the predicted year. Finally, the selected plantation dates are checked in the third sub phase to predict the probable diseases that might harm the specified crop.

3.3.1 Fuzzy suitable plantation dates selection sub phase

This sub phase is concerned with performing fuzzy evaluation for each day of the predicted year to determine its suitability degree of planting a given crop. The processing scenario of this sub phase is achieved through a set of steps as presented in Fig. 6. Firstly, fetch a set of successive continues days (period) from the time series climate data of the year with a length equal to the age of the crop under study. Secondly,

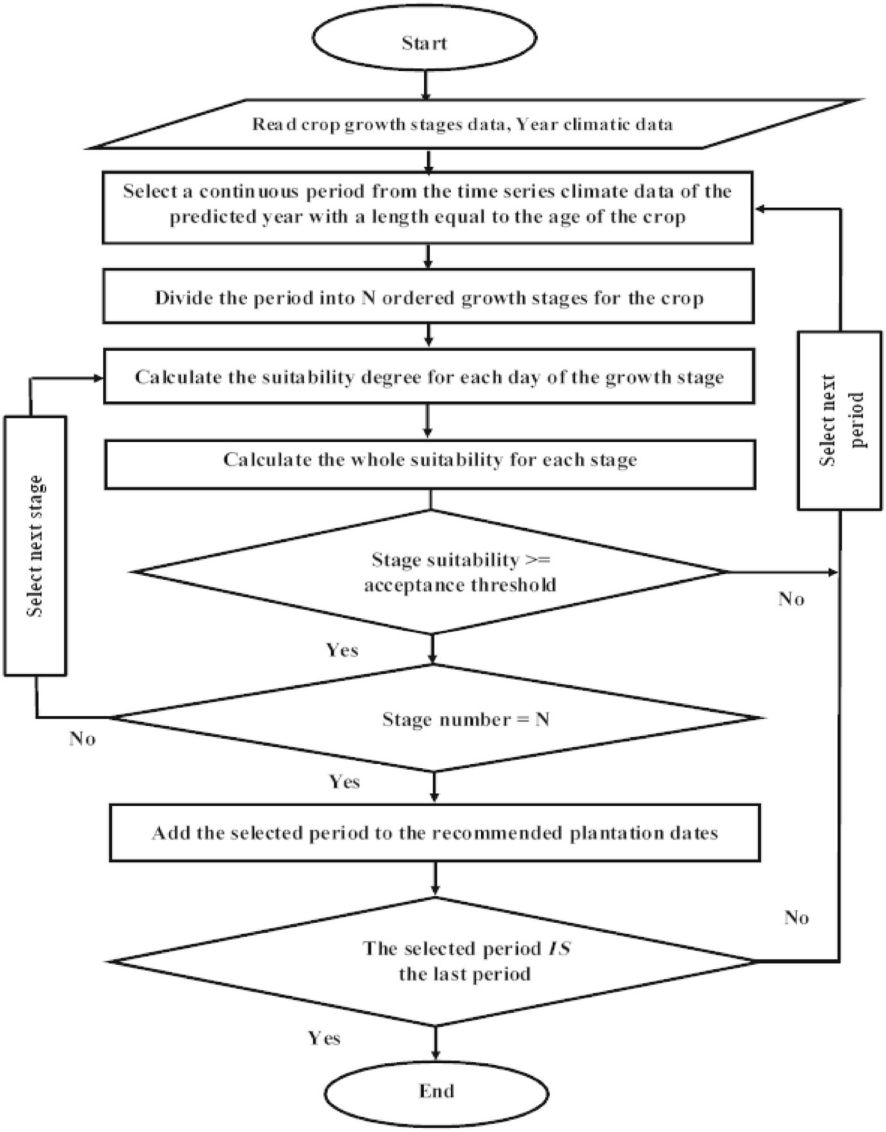

Fig. 6 Fuzzy selection of plantation dates sub phase flow chart

such period is divided into ordered growth stage for such crop. Thirdly, the proposed approach calculates the suitability degree for each period for the growth stage by calling the suitability fuzzy membership functions of climate variables that affect such growth stage. In consequence, the more suitable plantation dates for the considered crop can be determined by executing the following fuzzy query:

SELECT Plantation_Date, Suitability_Degree
FROM Years_Climate_Variables_Values, Crops_Climate_Requirements
WHERE CROP_ID = cropid
AND Temperature_ Fuzzy _Suitability (Plantation_Date) >= acceptance_threshold
AND Humidity_ Fuzzy _Suitability(Plantation_Date) >= acceptance_threshold
AND Sunshine_ Fuzzy _Suitability(Plantation_Date) >= acceptance_threshold
AND PERIOD_LENGTH=crop_age_default_value;

Generally, climate variables have different effects on almost crops. So, each climate variable is given a weight value (affecting factor) reflecting how much it affects the plantation process of the crop. Such affecting factor depends on the crop type. The agricultural experts are responsible for determining the value of the affecting factor (AF) of each climate variable regarding that the sum of all AFs must be Equal 1. Consequently, the total suitability matching value is computed by calculating the summation of the weighted suitability matching values of the affecting climate variables. Accordingly, the suitability membership value of specific day(smvd) for the plantation of a specific crop is computed as depicted in Eq. (6).

$$SMVD = \sum_{j=1}^{n} \mu_{CV_j}(x) * AF_{CV_j} \tag{6}$$

where n is the number of climate variables, CV_j is the climate variable j and x is the climate variable value.

Fourthly, the proposed approach calculates the suitability of each stage using Eq. (7) and the whole period suitability using Eq. (8). Fifthly, it selects the period if each sub period (stage) of such period has a suitability degree equal to or more than such stage acceptance threshold.

$$SatgeSutibility = \frac{\sum_{i=1}^{n} SMVD_i}{n} \tag{7}$$

where $SMVD_i$ is the suitability membership value for day i and n represents the stage length in days.

$$PeriodSuitability = \frac{\sum_{i=1}^{n} SatgeSutibility_i}{n} \tag{8}$$

where n is the number of crop growth stages.

3.3.2 Harvesting date calculation sub phase

Commonly, crops development and growth depend mainly on the temperature climate variable. Each crop needs minimum amount of development temperature which is known as base temperature. In other words, each crop requires a specific amount of heat (thermal units) to develop from one point in their lifecycle to another. The summation of daily mean temperature above the base temperature is known as growing degree days (GDD). If the crop planted at a suitable date obtains the required total amount of GDD for its development, it reaches the maturity stage regardless of the number of days [31].

In this sub phase the expected harvesting dates of the crops are calculated. This is computed based on the plantation date, the base temperature for the crop growth, the max and min temperature during the plantation period. The processing scenario of this phase is presented in the flow chart shown in Fig. 7. Generally, the GDD for one day is calculated using Eqs. (9) and (10). The total GDD of the selected crop is calculated using Eq. (11) by summing GDD for the plantation days. Once the required GDD is reached, the selected crop starts reaching its maturity stage.

$$GDD_{day} = \begin{cases} T_{mean} - T_{base} & T_{mean} > T_{base} \\ 0 & T_{mean} < T_{base} \end{cases} \tag{9}$$

where T_{mean} is the average value for temperature at specific day and T_{base} is the minimum development temperature for the selected crop growth.

$$T_{mean} = \left(\frac{T_{max} + T_{min}}{2}\right) \tag{10}$$

where T_{max} is the maximum temperature at specific day and T_{min} is the minimum temperature at specific day.

$$Total\_GDD = \sum_{i=1}^{n} GDD_i \tag{11}$$

where *Total_GDD* is the required amount of temperature for the crop maturity, GDD_i represents the obtained amount of temperature of the specified crop at day i and n is the number of actual planting days.

3.3.3 Fuzzy prediction of expected diseases

In this sub phase, the stages of each selected plantation date for a specific crop are tested to check if such stages are suitable for emergence of some diseases as shown in the flow chart presented in Fig. 8. It retrieves the expected diseases that might appear at any stage of crop growth regarding the predicted data of the next year and the climatic data that motivate crop diseases. Accordingly, the following fuzzy query is used to retrieve the expected diseases for a given crop.

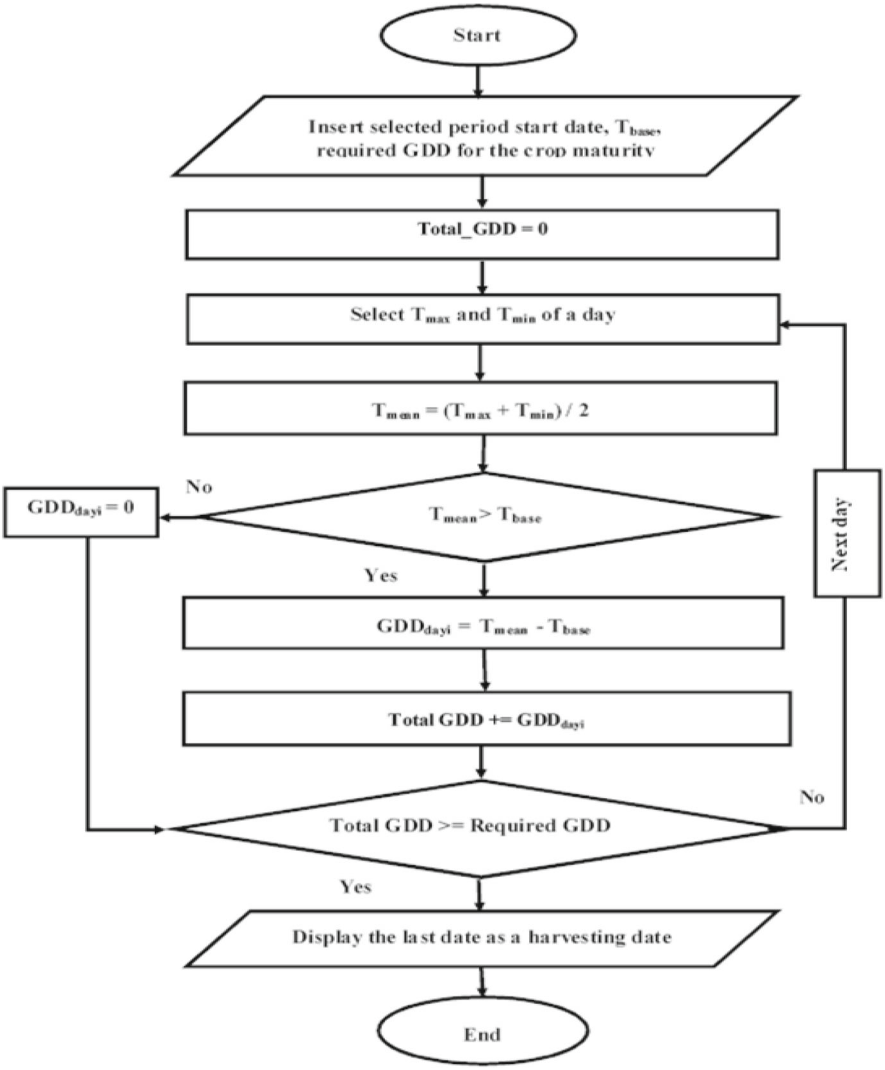

Fig. 7 Harvesting date prediction flow chart

SELECT Disease_Name, Matching_Degree
FROM CROP_DISEASES , YEAR_CLIMATE_DATA
WHERE CROPID = A_Given_CropID
AND Temperature_ Fuzzy _Suitability (Crop_diseases_emergence) >= acceptance_threshold
AND Humidity_ Fuzzy _Suitability (Crop_diseases_emergence) >= acceptance_threshold
AND Sunshine_ Fuzzy _Suitability (Crop_diseases_emergence) >= acceptance_threshold

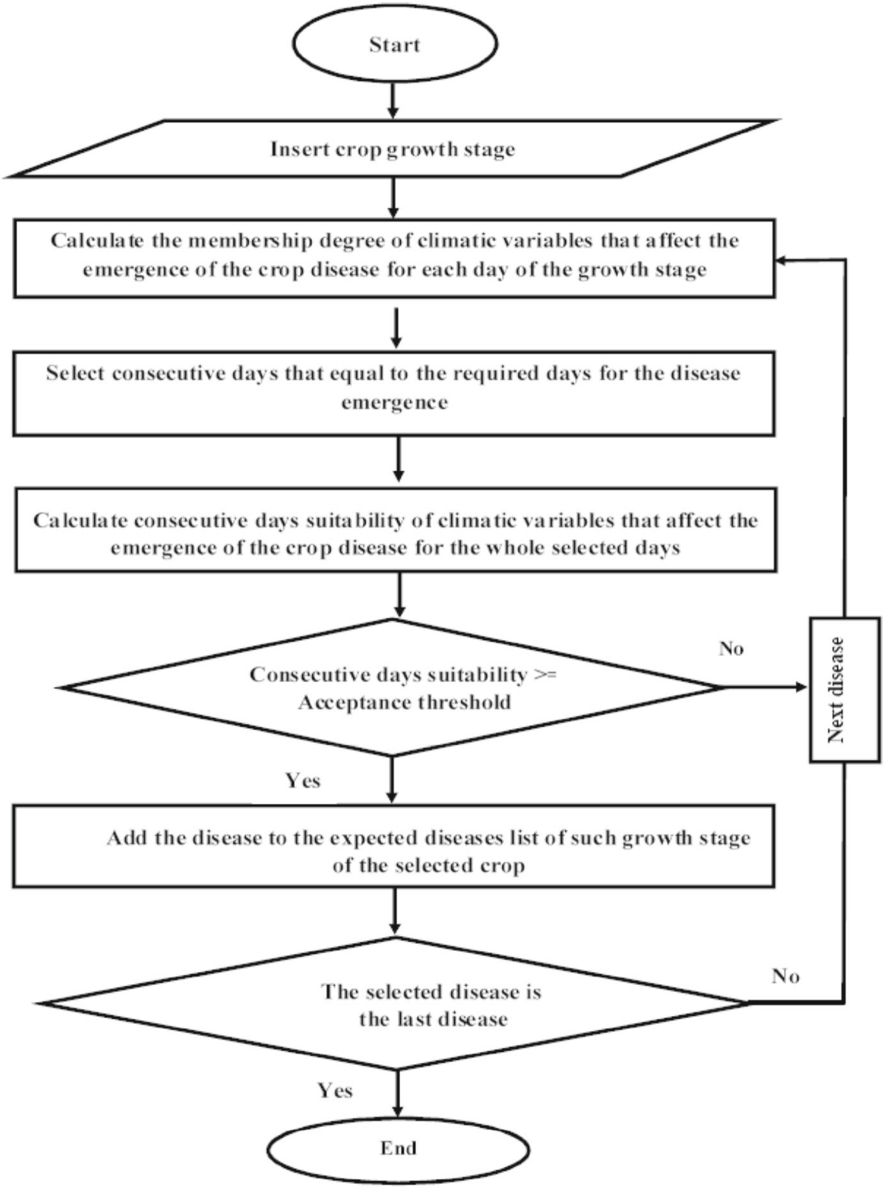

Fig. 8 Fuzzy evaluation of crops plantation periods suitability to the emergence of expected diseases

AND ***STAGE_START*** = *current_stage_start and* ***STAGE_END*** = *current_stage_end;*

Firstly, the disease emergence suitability fuzzy membership value of specific crop growth stage day(dsmvd) is computed day by day using Eq. (12). If the disease emergence requires more than one day then the proposed approach searches for a set of consecutive days that equal to the required days for the emergence of such disease. Then the overall suitability of such consecutive days for disease emergence is computed using Eq. (13). Finally, a disease is added to the crop expected diseases list if its emergence suitability fuzzy membership value of such day(s) is more than or equal to the acceptance threshold. Finally, the proposed approach provides the more suitable plantation dates and the expected diseases that might emergence during such dates.

$$DSV(R) = \frac{\sum_{j=1}^{n} \mu_{RDCV_j}(x)}{n} \tag{12}$$

where $DSV(R)$ is the suitability value of the emergence of disease D to crop R at a specific day, μ_{DCV_j} is the climate variable CV_j suitability membership function for emergence of disease D to, x the value of climate variable CV_j and n is the number of affecting climate variables on the emergence of disease the specified crop.

$$Consecutive\_days\_sutibility = \frac{\sum_{i=1}^{n} DSV(R)}{n} \tag{13}$$

where n is the number of consecutive days.

3.4 Crops water requirements follow up phase

The water is undoubtedly the most important requirement for crop growth. So that, this sub phase is concerned with a daily follow up of crop water requirement (the actual evapotranspiration by the crop) after the crop has been sowed. Such actual evapotranspiration is calculated as in Eqs. (14, 15) using the actual daily data streamed from the weather stations and the sowed crop coefficient [33].

$$ET_C = ET_O \times K_C \tag{14}$$

where, ET_C is the actual evapotranspiration by the crop (mm/day), ET_O is the reference evapotranspiration (mm/day), and K_C is the crop coefficient at a specific growth stage.

$$ET_O = 0.0023(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} R_a \tag{15}$$

where ET_O is Hargreaves ET_O, T_{mean} is the average temperature, T_{max} is the maximum temperature, T_{min} is the minimum temperature and R_a is solar radiation data at specific day

Depending on the data obtained from [32], the crop coefficient (K_C) is changed from crop stage to another and sometimes it ranges inside the development stage. So that to calculate it the proposed approach defines a function to calculate the values of K_C in Eqs. (16) and (17).

$$K_C(x) = \begin{cases} K_{C\_Initial} & x \le LD_{Initial} \\ K_{C\_Initial} + (DD_{num} * I_{co}) & LD_{Initial} < x \le LD_{Development} \\ K_{C\_Mid\text{-}season} & LD_{Development} < x \le LD_{Mid\text{-}season} \\ K_{C\_Late\ season} & LD_{Mid\text{-}season} < x \le LD_{Late\text{-}season} \end{cases} \quad (16)$$

where K_C is the crop growth stage coefficient, $K_{C\_Initial}$ is the crop growth stage coefficient at Initial stage, $K_{C\_Mid\text{-}season}$ is the crop growth stage coefficient at Mid-season stage, $K_{C\_Late\text{-}season}$ is the crop growth stage coefficient at Late-season stage, $LD_{Initial}$ is the last day in the initial stage, $LD_{Development}$ is the last day in the development stage, $LD_{Mid\text{-}season}$ is the last day in the Mid-season stage, $LD_{Late\text{-}season}$ is the last day in Late-season stage, DD_{num} is the day number in the development stage and I_{co} is the daily increase coefficient in K_C during the development stage.

$$I_{co} = \left(\frac{K_{C\_Mid\text{-}season} - K_{C\_Initial}}{Development_{length}} \right) \quad (17)$$

where I_{co} is the daily increase coefficient in K_C during the development stage, $KC_{\_Mid\text{-}season}$ is the crop growth stage coefficient at Mid-season stage, $K_{C\_Initial}$ is the crop growth stage coefficient at Initial stage and $development_{length}$ is the length of the development stage.

3.5 Illustrative case study

The proposed approach has been applied to manage the plantation process of the wheat, soybean and bean crops in Alexandria governorate, Egypt. The climatic requirements of such crops are gathered from the agricultural experts at Agricultural Research Center and from [27–29, 34–39]. On the other hand, the historical agro-climatic data includes temperature, sunshine and humidity as climate variables for the years from 1985 to 2016.

3.6 Crops plantation dates, harvesting dates and expected diseases

Commonly, the proposed approach returns each candidate plantation period with suitability degree more than 75% for any growth stage of the crop and overall suitability more than 80% (for crops plantation or crops diseases emergence). Figures 9, 10 and 11 show the suitability of each day of the year for some crops that are planted in Alexandria. And Table 1 shows the suitability of the traditional plantation dates for such crops in Alexandria. While Tables 2, 3 and 4 present the recommended plantation dates, harvesting dates and the expected diseases that the proposed approach provides.

As shown in tables from 1 to 4, a lot of traditional plantation dates had been affected by the climate changes. At the same time new plantation dates become more suitable although such new dates were not suitable before. For example, for the bean crop the

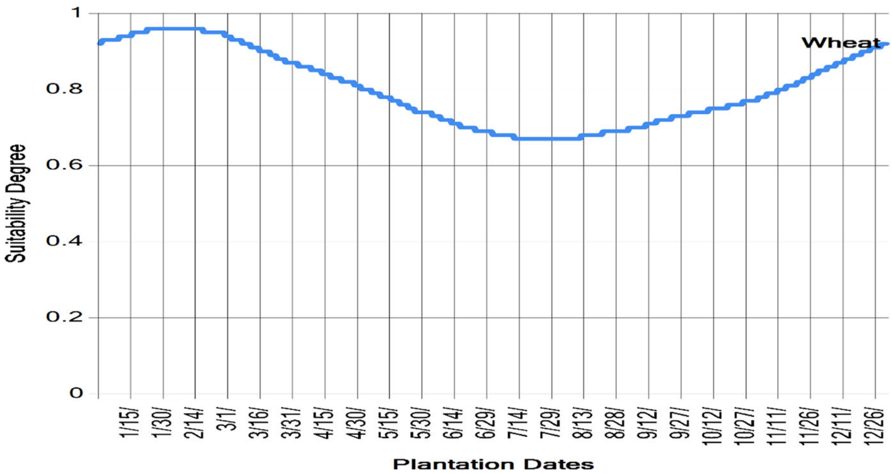

Fig. 9 Year days suitability for Wheat plantation in Alexandria

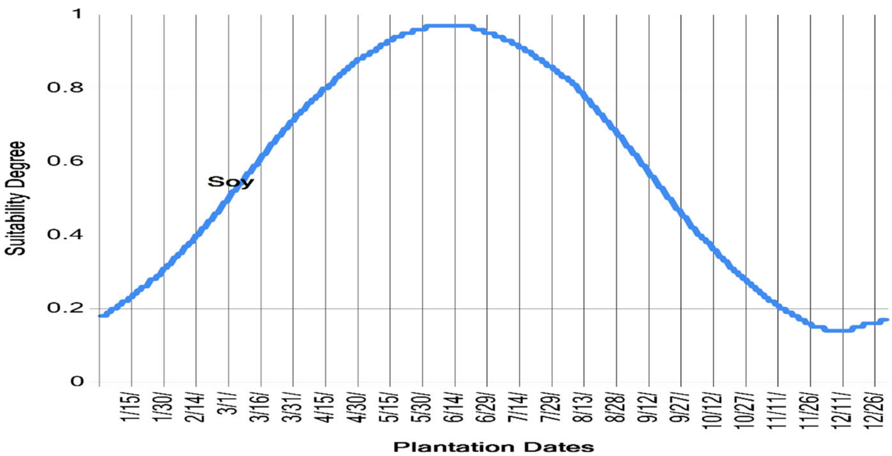

Fig. 10 Year days suitability for Soybean plantation in Alexandria

suitability of Nile buttonhole that start at 15 August has been degraded to became 66%. while other new discovered plantation date has been become more suitable like the plantation date that start at 24 December which has suitability degree 90.56%. Also, the plantation date that start at 1 January which has suitability degree 92.37%.

3.7 Water crops follow up

After the crops have been sowed, the proposed approach follows up the daily water requirements of such crops. For example, it was applied on the wheat in Alexandria governorate in Egypt with using the actual weather data streamed from the weather station, radiation data of Alexandria obtained from [40] and K_C of the wheat gathered

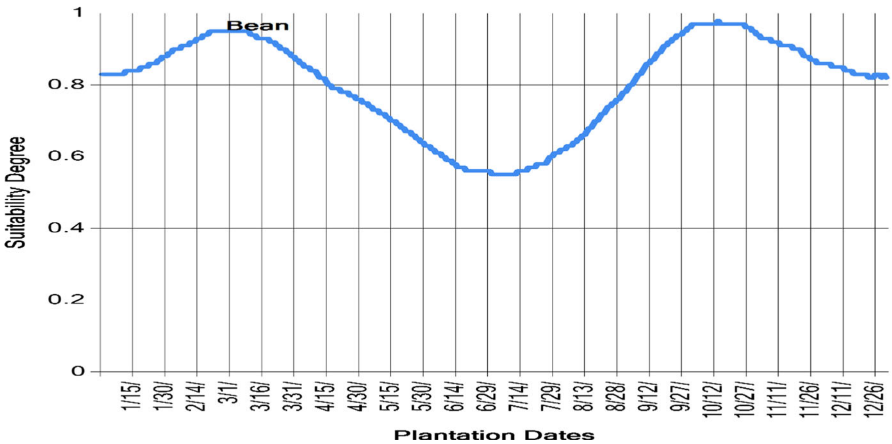

Fig. 11 Year days suitability for Bean plantation in Alexandria

Table 1 Evaluation of the suitabilityof the traditional plantation dates for cropsunder study

Crop name	Buttonholes	Sowing dates	Suitability degree (%)
Wheat	Only one	From 15 November	80
		To 30 November	84
Soybean	Only one	From 1 may	88
		To 30 may	96
Bean	Nile	From 15 August	66
		To 15 October	97
	Summer	From 1 February	90
		To 7 March	95

from [32]. Table 5 and Fig. 12 represent the daily water requirements for the first 50 day of the recommended plantation date of wheat that starts at 1 January Actually, such results confirmed that the proposed approach provides more accurate values for the wheat water requirements. Because the water requirements continuous changes reflect the weather changes and the wheat growth stage change.

4 Results and findings

Considering the results presented in the previous section, it is noted that some of traditional crops plantation dates became not suitable at all. At the same time, the proposed approach discovers new plantation dates that match the crops suitable pattern of climate variables like temperature, humidity and sun shining. Such discovered new plantation dates help Tackling the problem of negative effects of climate change on agriculture. Accordingly, the proposed approach helps decision makers when planning, managing and making a follow up for crops plantation process.

Table 2 Recommended periods for wheat plantation in Alexandria

Scenario number	Scenario sowing dates	Scenario harvesting dates	Germination suitability	Tillering suitability	Stem longation suitability	Flowering suitability	Physiological maturity suitability	Scenario full suitability	Expected diseases
1	24 December	10 October	99.78	93.76	92.26	91.13	75.88	90.562	Stripe rust (yellow rust) Common and dwarf bunt (stinking smut) Septoria diseases
2	5 March	2 June	99.38	76.92	88.93	99.88	100	93.022	Leaf rust (brown rust) Stem rust (black rust) Alternaria leaf blight
3	12 February	19 May	99.6	86.28	98.74	99.45	96.98	96.21	Leaf rust (brown rust) Stem rust (black rust) Alternaria leaf blight
4	22 January	4 May	99.61	90.79	98.08	98.21	89.98	95.334	Leaf rust (brown rust) Stem rust (black rust) Septoria diseases
5	1 January	18 April	99.74	93.34	94.14	95.05	79.57	92.368	Stripe rust (yellow rust) Common and dwarf bunt (stinking smut) Septoria diseases

Table 3 Recommended periods for Soybean plantation in Alexandria

Scenario number	Scenario sowing dates	Scenario harvesting dates	Scenario suitability	Expected diseases
1	30 July	20 November	85.25	Soybean rust
2	9 July	22 October	92.84	Soybean rust
3	18 June	29 September	96.98	Soybean rust
4	28 May	10 September	96.17	Soybean rust
5	7 May	26 August	90.1	Soybean rust
6	16 April	14 August	80.47	Soybean rust Sclerotinia steam rot

Table 4 Recommended periods for Bean plantation in Alexandria

Scenario number	Scenario sowing dates	Scenario harvesting dates	Frist stage suitability	Second stage suitability	Scenario full suitability	Expected diseases
1	24 December	21 May	82.87	81.79	82.33	Angular leaf spot Halo blight Anthracnose Rust
2	3 December	7 May	93.12	79.07	86.095	Angular leaf spot Halo blight Anthracnose Rust
3	12 November	14 April	99.88	82.89	91.385	Angular leaf spot Halo blight Anthracnose Rust
4	22 October	10 March	99.92	93.66	96.79	Angular leaf spot Halo Blight Rust
5	1 October	20 January	97	95.37	96.185	Angular leaf spot Rust
6	10 September	12 January	85.95	83.48	84.715	Angular leaf spot Rust
7	26 March	26 June	99.53	80.67	90.1	Angular leaf spot Rust
8	5 March	20 June	95.07	95.73	95.4	Angular leaf spot Rust
9	12 February	11 June	88.96	95.29	92.125	Angular leaf spot Rust
10	22 January	2 June	84.71	85.88	85.295	Angular leaf spot Halo blight Anthracnose Rust
11	1 January	25 May	81.68	83.32	82.5	Angular leaf spot Halo blight Anthracnose Rust

Table 5 Wheat daily water requirements respecting the weather prevailing conditions and actual evapotranspiration by the crop

Date	Water crop requirements (mm/day)	Date	Water crop requirements (mm/day)
1/1/2017	0.225	1/26/2017	0.355
1/2/2017	0.264	1/27/2017	0.355
1/3/2017	0.264	1/28/2017	0.445
1/4/2017	0.257	1/29/2017	0.417
1/5/2017	0.291	1/30/2017	0.495
1/6/2017	0.314	1/31/2017	0.297
1/7/2017	0.249	2/1/2017	0.305
1/8/2017	0.283	2/2/2017	0.463
1/9/2017	0.249	2/3/2017	0.548
1/10/2017	0.305	2/4/2017	0.621
1/11/2017	0.241	2/5/2017	0.608
1/12/2017	0.249	2/6/2017	0.557
1/13/2017	0.257	2/7/2017	0.535
1/14/2017	0.257	2/8/2017	0.582
1/15/2017	0.323	2/9/2017	0.456
1/16/2017	0.344	2/10/2017	0.452
1/17/2017	0.291	2/11/2017	0.406
1/18/2017	0.283	2/12/2017	0.632
1/19/2017	0.249	2/13/2017	0.7
1/20/2017	0.291	2/14/2017	0.677
1/21/2017	0.376	2/15/2017	0.566
1/22/2017	0.347	2/16/2017	0.508
1/23/2017	0.285	2/17/2017	0.518
1/24/2017	0.394	2/18/2017	0.581
1/25/2017	0.312	2/19/2017	0.657

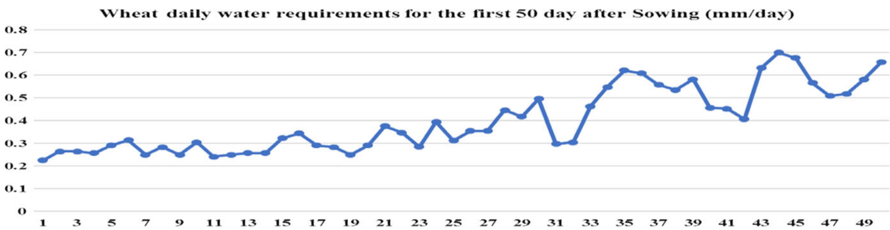

Fig. 12 Wheat daily water requirements assuming 1 January as start plantation date

Such help can be achieved by considering the following aspects of the plantation process: firstly, some crops seeds may be modified for adaptation with climate changes; secondly, some other crops need new convenient plantation dates to be planted in its traditional places or other new places having a suitable climatic pattern for planting

Table 6 Comparative analysis for the proposed approach against other similar approaches

Approach	Recommended plantation dates	Harvesting dates	Expected diseases	Crops water follow up	Scalability for processing huge data sets
Moussa et al. [15]	A single plantation date for the selected crop within two weeks	supported	Not supported	Not supported	Based on traditional processing models so that by the time the processing will be regarded specially with big data
Pankaj [17]	A single plantation date for the selected crop	Not supported	Not supported	Not supported	Based on traditional processing models so that by the time the processing will be regarded specially with big data
Hamid et al. [14]	Suggests a single crop plantations date based on water	Not supported	Not supported	Not supported	Based on traditional processing models so that by the time the processing will be regarded specially with big data
Mohammed et al. [21, 22] and Gadallah et al. [23]	Suggests more than one plantation date for each crop with a suitability degree for each period	Not supported	Not supported	Not supported	Based on traditional processing models so that by the time the processing will be regarded specially with big data
The proposed approach	Suggests more than one plantation date for each crop with a suitability degree for each period	Supported	Supported	Supported	Based on MapReduce model and Hadoop Hive which mean that 1. The approach could handle the huge data sets 2. Better performance

such crop. Thirdly, the problem of water poverty causes the need for a reliable mean to estimate the water consumption and make a follow up of the crops requirements from water to save the water as possible. Finally, it is crucial to predict and explore the ability of some diseases to harm a specific crop under some weather conditions.

Considering Table 6 the proposed approach has many advantages compared with the similar fuzzy based approaches. The first advantage is that the proposed approach base on the Hadoop platform to handle large scale data sets. Secondly, the proposed approach determines the harvesting dates depending on the GDD. Thirdly, the proposed approach performs test to the recommended plantation dates to discover the diseases that may emergence during such plantation dates. Finally, the proposed approach performs follow up to the water requirements for the crops that have been planted based on the actual weather data streamed from weather stations. Accordingly, it has been showed that the proposed approach has more added values compared with the other related works.

5 Conclusion

This paper presented a fuzzy based approach for discovering knowledge that helps managing crops plantation process respecting climate changes. It enables exploring the hidden patterns of suitable requirements for plantation process, predicting the crops diseases emergency and also in achieving a follow up of crops water requirements. Based on the historical agro-climatic database of Egypt, the proposed approach discovers new periods with suitable climatic patterns that are more appropriate than old periods for planting of wheat, soybean and bean. Then, the proposed approach estimates the harvesting dates for the recommended plantation dates depending on the GDD of each crop. Also, the proposed approach provides fuzzy evaluation for recommended planation dates to discover the expected diseases that might harm the crops during such plantation dates. Furthermore, the proposed approach allows a follow up the crops water requirements during crops life using actual data streamed from weather stations. Accordingly, such approach greatly helps in agriculture strategic planning to enhance the plantation plan of any crop in a specific location. Also, it directly increases the profitability of a crop plantation process. Finally, the proposed approach has been built using Hadoop Hive to handle the huge data sets with high performance. As a future work, the proposed approach would be extended to deal with more aspects of crops plantation process

References

1. Defang N, Manu I, Bime M, Tabi O, Defang H (2014) Impact of climate change on crop production and development of Muyuka subdivision—Cameroon. Int J Agric For Fish 2(2):40–45
2. Chen C, Qian C, Deng A, Zhang W (2012) Progressive and active adaptations of cropping system to climate change in Northeast China. Eur J Agron 38:94–103
3. Tao F, Yokozawa M, Zhang Z (2009) Modelling the impacts of weather and climate variability on crop productivity over a large area: a new process-based model development, optimization, and uncertainties analysis. Agric For Meteor 149:831–850

4. Gottwald S (2013) Fuzzy sets and fuzzy logic: the foundations of application—from a mathematical point of view. Springer, Berlin, pp 1–12
5. Zadeh L (1965) Fuzzy sets. Inf Control 8:338–353
6. Zhang J, Goodchild M (2002) Uncertainty in geographical information. Taylor & Francis, London
7. Shvachko K, Kuang H, Radia S, Chansler R (2010). The hadoop distributed file system. In: 2010 IEEE 26th symposium on mass storage systems and technologies (MSST), pp 1–10
8. Branco A, Evsukoff A, Ebecken N (2005) Generating fuzzy queries from weighted fuzzy classifier rules. In: ICDM workshop on computational intelligence in data mining, pp 21–28
9. Fuad, A, Erwin A, Ipung HP (2014) Processing performance on Apache Pig, Apache Hive and MySQL cluster. Information, communication technology and system (ICTS). In: 2014 international conference on. IEEE
10. Bhandarkar M (2010) MapReduce programming with apache Hadoop. Parallel & distributed processing (IPDPS). In: 2010 IEEE international symposium on IEEE
11. Floratou A, Minhas UF, Özcan F (2014) Sql-on-hadoop: full circle back to shared-nothing database architectures. Proc VLDB Endow 7(12):1295–1306
12. Capriolo E, Wampler D, Rutherglen J (2012) Programming hive. O'Reilly Media, Inc
13. Srinivasa KG, Muppalla AK (2016) Guide to high performance distributed computing. Springer, Berlin, pp 33–72
14. Hamid R, Mohammad A, Mohammad H (2013) Consideration of climate conditions in reservoir operation using fuzzy inference system (FIS). Br J Environ Clim Change 3(3):444–463
15. Moussa W, Patrick L, Traoré SB, Moussa S (2014) A crop model and fuzzy rule based approach for optimizing maize planting dates in Burkina Faso, West Africa. J Appl Meteorol Climatol 53:598
16. Mohaddes SA, Mohayidin MG (2008) Application of the fuzzy approach for agricultural production planning in a watershed, a case study of the Atrak Watershed. Iran. Am Eurasian J Agric Environ Sci 3(4):636–648
17. Kumar Pankaj (2011) Crop yield forecasting by adaptive neuro fuzzy inference system. Math Theory Model 1:3
18. Joshi RG, Bhalchandra P, Khmaitkar SD (2013) Predicting suitability of crop by developing fuzzy decision support system. IJETAE 3(2):10
19. Hartati Sri, Sitanggang Imas S (2010) A fuzzy based decision support system for evaluating land suitability & selecting crops. J Comput Sci 6(4):417–424
20. Salam MA, Mahmood MA, Awad YM, Hazman M, El Bendary N, Hassanien AE, Tolba MF, Saleh SM (2014) Climate recommender system for wheat cultivation in North Egyptian Sinai Peninsula. In: Proceedings of the fifth international conference on innovations in bio-inspired computing and applications IBICA 2014, pp. 121–130. Springer International Publishing
21. Mohammed AH, Allah AMG, Hefny HA (2014) Fuzzy time series approach for optimizing crops planting dates with climate changes. In 2014 10th international computer engineering conference (ICENCO), pp 36–41. IEEE
22. Mohammed AH, Gadallah AM, Hefny HA (2014) Fuzzy query approach for crops planting dates adaptation with climate changes. In 2014 9th international conference on informatics and systems (INFOS), pp DEKM-52. IEEE
23. Gadallah, AM, Mohammed AH (2017) Fuzzy-based approach for reducing the impacts of climate changes on agricultural crops. In: Handbook of research on machine learning innovations and trends. IGI Global, pp 272–294
24. Wolfert S, Ge L, Verdouw C, Bogaardt MJ (2017) Big data in smart farming–a review. Agric Syst 153:69–80
25. Central Laboratory for Agricultural Climate. http://www.clac.edu.eg/?lang=en. Accessed 20 Jan 2017
26. Climatology: ww2010.atmos.uiuc.edu/%28Gh%29/guides/mtr/fcst/mth/oth.rxml. Accessed: 20 Jan 2017
27. Yadav S Climatic conditions for growing wheat. https://www.importantindia.com/12612/climatic-conditions-for-growing-wheat. Accessed 20 Nov 2017
28. Kenanaonline: wheat plantation-suitable climate. http://kenanaonline.net/page/1966. Accessed 14 Oct 2016
29. Faculty of agriculture: soybean: www.aun.edu.eg/faculty_agriculture/soy.pdf. Accessed 20 Jan 2017
30. Lee KH (2006) First course on fuzzy theory and applications. Springer, pp 129–150
31. Midwestern Regional Climate Center: Growing Degree Days. http://mrcc.isws.illinois.edu/gismaps/gddinfo.htm. Accessed 20 Oct 2017

32. Chowdhury S, Al-Zahrani Muhammad, Abbas A (2016) Implications of climate change on crop water requirements in arid region: an example of Al-Jouf, Saudi Arabia. J King Saud Univ Eng Sci 28(1):21–31
33. Natural Resources Management and Environment Department: crop evapotranspiration—guidelines for computing crop water requirements. http://www.fao.org/docrep/X0490E/x0490e07.htm. Accessed 20 Nov 2016
34. Prescott JM, Burnett PA, Saari EE, Ranson J, Bowman J, de Milliano W, Singh RP, Bekele G Wheat diseases and pests: a guide for field identification. https://wheat.pw.usda.gov/ggpages/wheatpests.html. Accessed 14 Jan 2017
35. Hagedorn DJ, lnglis DA Handbook of bean diseases. http://learningstore.uwex.edu/assets/pdfs/A3374.PDF. Accessed 14 Jan 2017
36. Pessl Instruments: Soybean Disease. http://info.metos.at/tiki/tiki-index.php?page=Diseases+of+Soybean&structure=Disease+models. Accessed 20 Jan 2017
37. North Dakota Agricultural Weather Network: wheat growth stage prediction using growing degree days (GDD). https://ndawn.ndsu.nodak.edu/help-wheat-growing-degree-days.html. Accessed 20 Sept 2017
38. Western Regional Climate Center: Degree days. https://wrcc.dri.edu/Climate/Education/degdays.php. Accessed 20 Mar 2017
39. Akyuz FA, Kandel H, Morlock D (2017) Developing a growing degree day model for North Dakota and Northern Minnesota soybean. Agric For Meteorol 239:134–140
40. Diab F, Lan H, Zhang L, Ali S (2015) An environmentally-friendly tourist village in egypt based on a hybrid renewable energy system—part two: a net zero energy tourist village. Energies 8(7):6945–6961
41. Bigdata.black: What is Hadoop?. http://bigdata.black/infrastructure/compute/what-is-hadoop/. Accessed 20 Oct 2017