SMART AGRICULTURE: AN IMPROVED FARMER-CENTRIC INFORMATION RETRIEVAL SYSTEM USING MACHINE LEARNING TECHNIQUES AND FUZZY LOGIC FOR FAST CASE RETRIEVAL

Surabhi Solanki[1], Dr. Seema Verma[2] and Kishore Chahar[3]

Article History: Received: 10.05.2022 **Revised:** 15.06.2023 **Accepted:** 20.06.2023

ABSTRACT

To construct an improved farmer-centric information retrieval system smart agriculture using machine learning techniques and fuzzy logic is proposed in this paper. In case-based reasoning systems, the capacity to precisely define cases is crucial. Researchers have extensively analyzed various representations, including linguistic, attribute-value, and ontological models. As a result, retrieving cases from large databases might be time-consuming. In this paper, we propose a strategy for efficient case retrieval based on the concept of related representations. Whether or not two cases share similarities or differences, they are still connected. After a case is reported, it is compared to past data in an effort to find an exact match. Related cases are evaluated for similarities rather than the entire case base. The related case representation and conventional techniques were compared for their respective caseloads and retrieval accuracies. Fuzzy rules are utilized to define the threshold values of the various models, and Independent Recurrent Neural Networks (Ind RNN) are employed to evaluate the effectiveness and similarity of the models. The findings suggest that the idea can be used for very accurate and fast case retrieval. The related case representation strategy outperforms competing approaches in terms of retrieval effectiveness.

Keywords: Information Retrieval, Machine Learning, Ind RNN, Fuzzy.

[1]Deptartment of Computer Science & Engineering, Banasthali Vidyapith, Tonk Rajasthan- 304022
surabhics.solanki9@gmail.com
[2]Deptartment of Electronics, Banasthali Vidyapith, Tonk Rajasthan- 304022
seemaverma3@yahoo.com
[3]Deptartment of Electrical Engineering, Gautam Buddha University, Greater Noida, Uttar Pradesh- 201312
kishore001jan@gmail.com

DOI: 10.31838/ecb/2023.12.si6.598

1. INTRODUCTION

New technologies are causing a revolution in the farming industry, and it looks like it will be a good one because it will help farmers increase their profits and productivity. Timely adjustments to farming inputs are at the heart of precision agriculture. Increased data availability is currently aiding the third phase of the current agricultural revolution. The United States Department of Agriculture (USDA) claims that increased profitability can be attributed to the use of precision agriculture technologies [1]. The environment of modern knowledge is gradually being utilized on farms to guarantee the long-term viability of agricultural produce.

However, energy crops raise issues about how to use best-limited land resources [2][3], especially in light of rising product costs and a growing population [4]. Such worries can push policymakers to promote lignocellulosic biomass which is directed to “surplus” land and can relieve pressure on premium cropland.

Due to the long cropping cycle of crops, they can compete with food production and future food demands [5]. Realizing energy crops can be cultivated and can be grown for a reliable assessment of biomass supply appropriateness and global bioenergy production sustainability [6][7]. Indeed, many policy calculations assume that if sufficient incentives are provided during the start-up period, many farmers can decide to plant energy crops. However, their low adoption rate – about 100,000 ha in Europe – in comparison to tremendous technical potential appears to cast doubt on this notion. When it comes to the adoption of perennial energy crops, it is essential to stay on top of farmers’ attitudes, preferences, and behaviors, especially in the context of agricultural system innovation.

As a result of experience, farmer behavior varies and changes with time. It becomes progressively difficult to foresee when making a decision. Miscanthus has a high yield potential, it requires low input and a superior carbon sequestration capability. Simply expanding the quantity of land dedicated to agriculture cannot be enough to meet the nutritional needs of future populations. The amount of suitable land accessible can be reduced as a result of the loss of suitable agricultural land due to desertification and salinization, The current plant production systems can be made more productive using past agricultural intensification methods. Global food production has increased in the previous 35 years, thanks to the employment of high-yielding plant genetics which paired with extensive usage of synthetic fertilizer, pesticides, and irrigation.

IoT has newly increased traction in the farming business because it can meet the critical need for cross-brand interoperability, traceability, and scalability. As the IoT continues to evolve, it adapts to a wide range of applications. This paper shows the layers in the IoT design to encompass the spectrum of technologies, protocols, standards, and so on. There appears to be a consensus among the authors whose work we've reviewed that there are three distinct layers to the design of the Internet of Things. The ability to measure more than three levels between the network and the device is very relevant in IoT systems that employ fog or edge computing. Despite some variation in terminology, layers are often broken down into device, application, and network categories. Information retrieval has become the primary method for extracting knowledge from vast amounts of data and solving problems [8]. The processes of data in arable agriculture are depicted in Figure 1 [9].

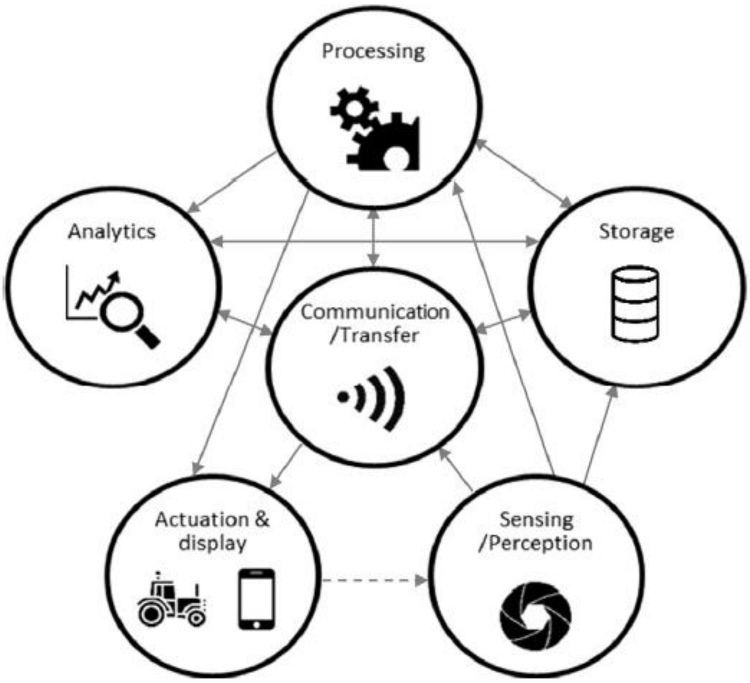


Figure 1: There are many data streams in arable farming [9]

The real meaning of classic keyword matching-based information retrieval is to get material from the Internet using advanced algorithms, construct a mapping link between information and keywords and then sort the retrieval outcomes using algorithms. Users search and extract knowledge from the retrieved results to solve practical problems [10]. Traditional retrieval methods are centered on keyword literal matching rather than the meaning or concept that keywords express. It separates the semantic relationship between words and to a great extent of words. Furthermore, synonym ad polysemy is common in natural language. The web page containing the keywords is returned in the retrieval results. As a result, a substantial number of resources that were missing or misdiagnosed are retrieved. Users have a difficult time finding resources that are closely connected to their demands from the recovered results, which do not match the demands of semantic and precise retrieval in the modern generation of semantic web environment [11][12].

Agricultural modernization refers to the move from traditional labor-based farming to technology-based farming. It's a major difficulty for agricultural policy, particularly in emerging countries where agriculture is less established. For a variety of reasons, agriculture often lags behind other economic sectors when it comes to adopting and implementing current organizational and production technologies, IT solutions, managerial techniques, and institutional frameworks. Competitive advantage is maintained because of its ability to rapidly build farms and undertake modernization. Poor nations are also plagued by primitive technology, which is both a cause and a consequence of their poor incomes. As a consequence, modernization and technological change are considered the key drivers of economic progress. Modernizing a farm in a well-managed way increases the efficiency of working conditions and, as a result, increases employee satisfaction with their jobs. Modernization may have negative consequences if it is done incorrectly [13].

Precision irrigation is the method of applying water in such a way as to meet the specific requirements of particular vegetation or management units with the fewest possible adverse effects on the surrounding environment. Precision irrigation has been used as a solution to the growing problem of water scarcity and the need for more genuine irrigation management because it is a holistic coordinated performance that allows for the emergence of additional elastic and reactive operational systems and the most effective operation and organization of agricultural irrigation systems. Precision irrigation, which includes a wide range of plot as well as scheme-level audience component systems to account for different management and checking realities, has the potential to be a useful allocation of water support tools for agricultural practices in areas that are irrigated and irrigation scheduling. The following are some of the direct benefits of the water allotment system and the integrative management effectiveness of precision irrigation [14]:

- **Collaboration in Irrigation**: To best represent the historical and geographical allocation of water variations throughout the field of agriculture at a resolution appropriate to farming managing activities, modern irrigation equipment design techniques and optimization tools allow water allocation growth with together digital and physical prototyping.
- **Availability of Water Resources**: By deciding to instantly distribute water resources throughout the whole precision irrigation supply chain, farmers are afforded greater leeway and accuracy in their water allocation and implementation schedules, as well as in the mechanisms by which they adapt to unpredictable environmental changes.
- **Real-Time Data on Essential Factors**: In the agricultural production system: It can be acquired by using stage irrigation objectives as water allocation elements. The irrigation system's corroborative behavior can be altered in response to unusual agricultural settings or changes in environmental variables, assisting in the identification of optimal irrigation timing when planning water allocation schemes.

To collect data in smart agriculture, a variety of sensors are used. One of the difficulties in handling these data is converting sensor readings into useful information. Case-Based Reasoning (CBR) has been recognized as a potent approach of developing smart agricultural systems in the context of knowledge management, where metadata descriptions are used for characterizations. [14]. Cases, which consist of problem descriptions and solutions can be used to present these characterizations. As a result, merging case-based reasoning with sensing technology can help intelligent agricultural systems. ML algorithms are progressively more used in the agriculture supply chain's core 4 clusters (production, processing, distribution, and reproduction) is vital. The ML was utilized during the preliminary stages of manufacture. Most commonly, technologies are used for crop forecasting. It's important to think about things like yield, soil quality, and watering needs. Figure 2 demonstrates New Farming Methods.

Figure 2: Improved farmer technique [15]

In the manufacturing phase, ML can be used. Disease identification and prognosis are performed using the third processing stage group. Particularly, ML techniques may be utilized in a distribution cluster to assess production planning for the purpose of guaranteeing the safety and quality of the final product. especially in terms of logistics, shipping, and storage. The agricultural market begins with a customer analysis, which is the initial preproduction cluster. It is mostly concerned with Crop output, soil qualities, and irrigation requirements are all predicted. The significance of this is reported by several studies. production of crop yields in the order generated promotes plant growth management. These precision agricultural solutions use data (equipment) as input and employ machine learning algorithms to foresee efficient models that encourage the best agricultural yield forecasting choice and enhance clever farming techniques, with the goal of better-informing stakeholders and farmers regarding what they require. Clustering, decision tree, Regression, and deep learning are all names that may be utilized to refer to a Bayesian network, which has recently been applied to the task of forecasting crop yields [15-17].

2. RELATED WORK

Here you may see how other writers have used Machine Learning to create superior retrieval systems with a focus on farmers' needs.

ALMarwi et al. (2020) suggested an information retrieval system based on keyword matching that can certainly fail to recover texts with equal meaning but syntactically distinct keywords (form). A well-known approach to bypassing this constraint is query expansion (QE). There are a variety of methods for query expansion including statistical approaches. This technique uses phrase frequency to generate expansion features, but it ignores meaning and term dependency. Other methods include the semantic approach which is based on knowledge with a limited number of ideas and relationships. In the suggested study, a hybrid query expansion strategy that incorporates statistical and semantic approaches is preferred. To choose the finest terms for query expansion, researchers suggested an effective weighting technique based on particle swarm optimization (PSO). A system prototype was created as a proof-of-concept and its accuracy was tested. The experiment was carried out on a real dataset. The suggested strategy improves query expansion accuracy according to the results of the experiments [18].

Saiz-Rubio et al. (2020), help farmers save money, safeguard the environment, and adapt their food production to meet the needs of a growing population in a sustainable way. This data-driven approach to farming is utilized to maximize productivity while minimizing environmental impact. The future of natural farming is bright, thanks to data-driven agriculture coupled with automated methods that employ artificial intelligence techniques. Acquiring field data for different rate applications is a crucial component of modern farm management systems [19].

Chaterji and colleagues (2020) suggested digital agriculture. Internet-of-Things (IoT)-based technology allows a system for the evaluation at various levels and the generation of instruments that permit better decision-making in each sub-process. Recent advancements in IoT hardware, like as networks of diverse embedded gadgets and software, like internet optimization software and lightweight approaches for computer vision, make it possible for a linked (smart) farm to effectively gather and evaluate data from a broad variety of sources. In order to collect data from various farming operations comprising various time frames in near real-time, it is necessary to link such IoT sensors, that are generally distributed across wide swathes of farmland. Insights that can be put into action, such as the best way to apply soil additives to minimize waste and contamination, can be gleaned from such data. The potency of such insights will increase when federated learning methods are utilized to aggregate information from various farms. IoT devices with sensing capabilities of all kinds can talk to one another via a wireless sensor network (WSN). Various field activities in line crop systems generate massive volumes of data. Soil testing Among these tasks is: fertilizer spreading; planting; spraying; scouting; distributing; processing; and harvesting [20].

Precision irrigation's water integrative management and allocation mechanism effectiveness have recently been highlighted by Liang et al. (2020), showcasing the role of irrigation water allocation in enhancing the efficiency of precision irrigation methods; discussing the models, facilities, and management techniques currently in use; and discussing the models, facilities and management techniques currently under development. Some of the most cutting-edge strategies for developing a long-term, integrated, and natural system for irrigation that also provides the superior effectiveness and efficiency required for entirely precise irrigation application are data-driven watering management, cloud-based irrigating management, and performance-proven distribution of water. [21].

Patel et al. (2020) The three facets of Indian conventional agriculture that were investigated were farming, readily available regionally sustainable crop protection measures, and biological pest management. Food security and environmental sustainability need India's most well-known conventional agricultural practices to be revitalized, including rotation of crops, mixed cropping, dual cropping, agroforestry, resource exploitation with host-pathogen interaction, and the use of local varieties. Such methods are crucial to the long-term success of agriculture because they improve crop nutrition [22].

Dubois et al., (2021) centered on growing potatoes, a commodity that requires a lot of water. The author posed the problem of predicting soil water potential as a supervised learning model. Multiple inputs and outputs make this seem like a difficult task. Using data gathered over three years, experiments demonstrate how to automatically build models with relevant traits and excellent performance via feature selection [23].

Wang et al., (2020) looked at detecting estrus in dairy cows utilizing speed, location, and machine learning. Over 12 days, data was gathered from 12 cows. It collected 25,684 measurements of location and velocity. Using a PCA analysis of 12 behavioral characteristics, multiple machine-learning approaches were evaluated to automatically identify estrous cows. The neck tag has a static positioning efficiency of 0.25 m and a dynamic positioning accuracy of 0.45 m [24].

3. BACKGROUND STUDY

Zhai et al. (2019) [12] suggested an associated method of case representation, which allows for quick case recovery in agricultural CBR arrangements. A case usually comprises 2 sections, plus definitions of problems and descriptions of solutions. A further part known as the association part, is expanded on this basis in this typical case structure. As an outcome, a “problem-solution-association” framework accompanies the proposal. Each case is interlinked with many similar cases in the case base, that have good similarities or changes. When a modern agricultural scenario has been identified, its characteristics are compared by similarity measurements with historical data to recognize a comparatively same as past cases. Later evaluations are performed preferentially between the current case and related cases until the most comparable past case is matched.

The most comparable previous case can be recovered by staying a small number of issues because of this association relationship. Thus, beneath the condition that a big amount of data is saved in a case base, the quality of the case retrieval can be greatly improved. CBR system can be nicer support for the product domain of quick agriculture with such enhancements. In the case of retrieval tasks, the corresponding case description approach was contrasted with the standard attribute-value pair representation. There are two parameters considered in the comparison precision and performance. The accuracy of the case retrieval, on the one hand, stated that the previous case retrieved should have significant similarities with the objective one. The success of case retrieval was proven by the amount of time each case was kept. Efficiency increases with reduced repetition.

4. PROBLEM FORMULATION

Case association work is covered in the background research to speed up the data retrieval process over a limited dataset. The system accepts direct user input in the form of cases to analyze and mine for data on the similarity between them. The research shows that the first big hole is that there is no specification for the situations that are generated in response to a user query. Next, machine learning should be used to make similarity evaluation more practical by shifting the emphasis from quick retrieval of sparse data to rapid and thorough assessment of massive datasets. Finally, regardless of the classification and kind of dataset, a consistent set of criteria is used throughout the processing to determine the threshold value.

There are three main areas where this work lacks, and much of the effort here has gone into filling those gaps. Because 70-80% of the platform's intended customers are unfamiliar with the technical terminology associated with agriculture features, query expansion is an essential first step in finding a solution to the problem. Second, since the focus of the previous work is on the efficient retrieval of cases, the algorithm's accuracy is never considered for massive datasets. As the result of each neuron varies in response to the present input and the history of previous concealed state output, IndRNN is recommended for efficient and rapid similarity evaluation. Third, depending on the kind and classification of the dataset, the threshold is defined according to some fuzzy rules that can be added or removed as needed.

5. MATERIAL METHOD

Two classes of information are usually considered in an agricultural case: ecological data and crop/plant-based data. Initially, ecological data such as soil, warmth, moisture, sunlight, wind, pests, illnesses, and many others are typically external variables. It performs an important role in the management of farming operations, as it can affect plant growth and operating efficiency. For instance, if the growing temperature is higher, the crop growth cycle will be reduced. A one-degree Celsius increase in temperature, according to agronomists' projections, would cut the rice growing season by more than a week [25]. Another example can apply to the deployment of farming operations of unmanned aerial vehicles (UAVs). Due to safety concerns, unmanned aerial vehicle (UAV) launches are not possible when wind speeds exceed 10 meters per second. A Few of the main ecological data characteristics in farming instances i.e., summarised in Table 1.

Secondly, plant data includes crop development, crop production, stress, dry weight of plants, flowering period, root biomass index, the thickness of sowing, etc. These factors need to be considered, as agricultural judgment-making differs in various circumstances. The planting density of crops, for example, can affect insecticide dilution concentrations. A lower dilution typically requires extensive planting. In addition, because of their growth stage, crops are susceptible to the insecticides applied. For those crops which are at the seed level, toxic insecticides can be lethal.

Table 1: Critical Pieces of Environment-Related Information for Farming [26]

Features	Specification
Climate	The min and Max climate value
Pests	Location, severity, pest species, population, etc.
Humidity	Humidity Value
Soil	The situation, variety, region, spot, etc. of the soil.
Diseases	Disease name, disease stage, and incidence are the severity
Wind	Wind speed and direction
Sunlight	Quantity of sunlight and duration of radiation.

Table 2: Data Elements Relevant to Crops and Plants in Agricultural Investigations

Features	Specification
Planting	Area and Planting density

Crop/plant-based data in agriculture situations are summarized in Table 2.

Therefore, presents the main characteristics of crop/plant-associated data. The query from the user side is extended at the initial stage and then the expanded query for cases is transformed. The initial user query is extended in request to attain a rich set of related words. This can be achieved by combining candidate words, like WordNet®, word embeddings, and term frequency, from different types of information sources listed above. In three separate steps, the problem is optimized. First, using WordNet®, the same word every time a user query is found. For further expansion in the second level, they are shared with the seed query words.

In the second phase, Word2Vec is used by calculating its cosine gap from the original word in the vector space to derive more semantic-related words from the previous phase each time. For our term embedding process, Word2Vec (skip Gram) is chosen as it has proven helpful in securing intense word representations related to their meaning. The most common terms are determined in the third stage to define more candidate expansion terms. In answer to the initial query, numerous terms are determined utilizing the quick miner instrument on a list of documents that are recovered at the top levels [27]. As far as the relationship between cases is concerned, it not only distinguishes related cases but also dissimilar ones.

5.1 Similar Cases

In environmental and crop/plant-related results, similar instances are considered to have great similarities. Each medical state has three mirror images, all of which have been gathered together according to their respective case IDs and various measures of resemblance. A new example is then linked to an older case, and the connected occurrences of the older case are compared to one another to shed light on why the new case is the most related possible scenario. The association of related cases offered the ability to evaluate the correlation contained by a narrower range as an alternative to looking for the entire case base. Therefore, the number of cases stayed was decreased, prominent to greater efficiency of case recovery.

5.2 Dissimilar Cases

Environmental and crop/plant-related outcomes are expected to differ significantly between cases. Here, we compare three distinct cases by identifying their associated case IDs and degrees of similarity from a database. It was originally planned that the latest case, which may quickly discover a comparably comparable example, would benefit from the organizational relationship between distinct instances at the very outset of case retrieval. If the link between the most recent case and the preceding case exceeded the set threshold, the comparison will be done with those specific rather than exploring succeeding cases.

For the similarity evaluation, IndRNN is applied here for the objective of training and testing the collected data/cases. RNN is a type of deep neural network, unlike a conventional neural network, that has specific characteristics defined like core cell state or memory to survey input features and time dependencies to expect future output from following input data. The output of every neuron, therefore, differs not only on the existing input but also on the background of prior concealed state output.

RNN can efficiently preserve knowledge about the past depending on the number of time levels. By providing characters as input parameters, RNN has remained widely helpful in real language management, speech identification, and system interpretation. Figure 3 shows the IndRNN Model [27].

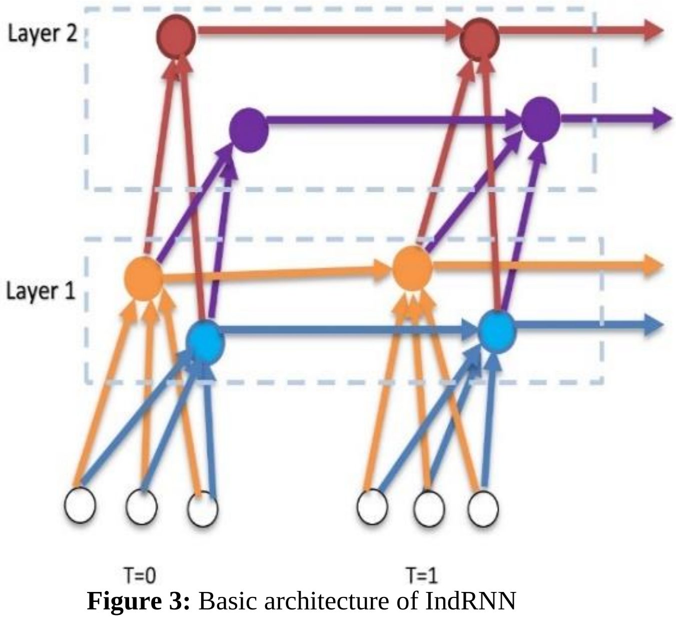


Figure 3: Basic architecture of IndRNN

6. METHODOLOGY

As can be observed in Figure 4, the proposed related case structure enabled rapid case recovery by connecting with the assistance of like and unlike examples, rather than assessing similarities of each case in the data base individually. The attribute-value pair clarifies the technique, whereas the vector formalizes the situation. The vectors' features are compared and the weights are applied if they are similar.

Subsequent instructions for resolved threshold

- **Rule 1**

Identification of cases that are not the same. When there is less than a 50% degree of similarity between the source or target cases, we say that the target event is different from the source case. A special flag indicates the starting point of the chain of events.

- **Rule 2**

Identification of identical situations. When there is greater than a 50% degree of similarity between the two situations, the alternative case is evaluated in the same way as the original case. A similar flag is applied to the source event.

- **Rule 3**

Determination in somewhat close cases. When the target event is very identical to the source case, then the similarity between the two examples is quite high. An extra indicator that the source case is highly similar has been added.

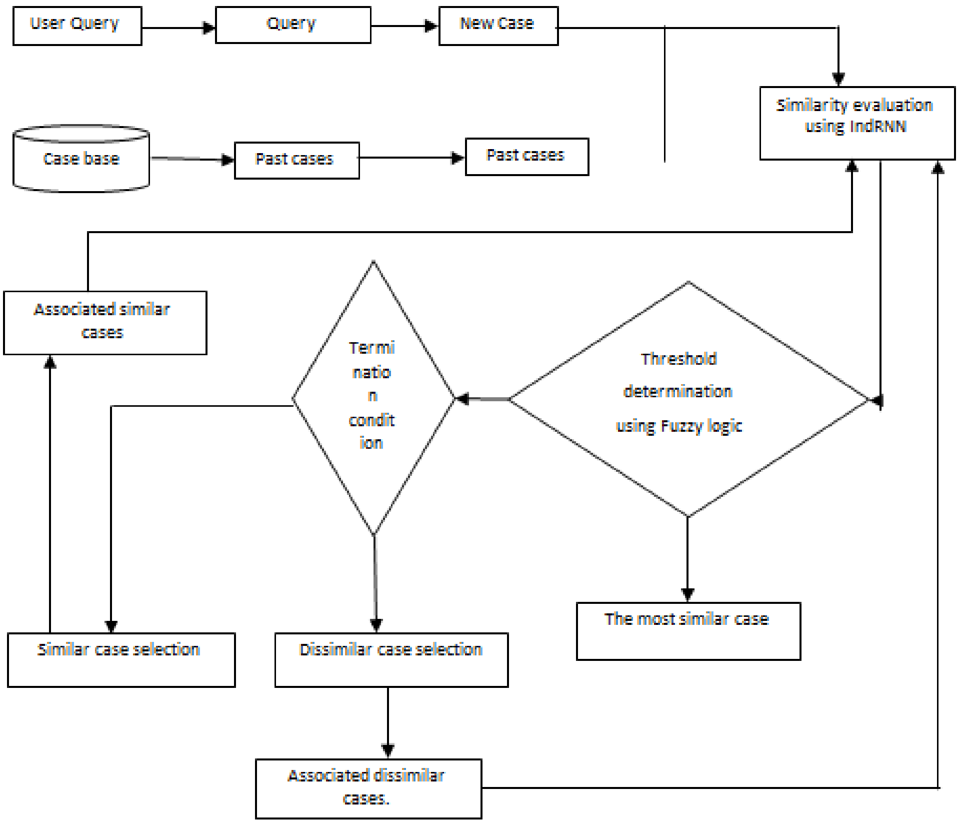


Figure 4: Resultant G

- **Rule 4**

For the next iteration determine whether to prefer identical or dissimilar connections. For the next version, the affiliation of more flags will be selected under the circumstances. If there are more like flags than different ones, the case with the highest comparison value is used as the starting point for the following round of evaluations. If there are more dissimilar flags than similar ones, the case with the strongest connections would be chosen as the source. In the next cycle, we'll compare the similar and different examples.

- **Rule 5**

a series of incidents that are eerily similar. Furthermore, if the level of similarity between two examples exceeds 75%, its analogous connection in the most similar scenario will be selected for comparison in the following instance.

- **Rule 6**

Case history evaluation and selection. All three examples in a given iteration have been compared historically due to the one-to-many nature of the links between cases in the past. Case 547 is connected to the digits 9 and 956. The contrasted scenario for the next iteration will be chosen from the previous phase in this situation. The related cases of the 2 foremost comparable (or dissimilar) cases are selected for comparison centered on Rules 4 and 5 [6].

7. FUZZY RULES

The complete set of fuzzy rules used in fuzzy logic is listed below.

9.1 Fuzzy Control Rules

An expert's knowledge in any topic is best described by the fuzzy control rule. If the closed-loop control method is employed, the fuzzy rule is specified by an IF-THEN statement that specifies the actions to be taken based on the information currently available. The law that controls the design or construction of a set of fuzzy rules depends on the specific circumstances at hand, and as such, it is grounded in the human knowledge and experience with which the designer or constructor is working. Fuzzy sets of data and linguistic parameters are used to define a scenario, which is then linked to a result using a fuzzy IF-THEN rule. In most cases, the THEN section is used to offer an output in the form of a semantic variable, while the IF part is used to gather experience through a range of varying conditions. This IF-THEN rule is commonly used with the fuzzy inference technique to evaluate the extent to which incoming data satisfies the requirement of recalculating the degree of fuzziness between a LOW (temperature) fuzzy condition and an input T (temperature) fuzzy value as depicted in Figure 5 [7].

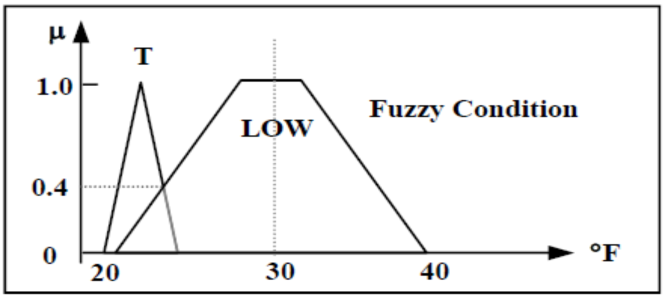


Figure 5: Fuzzy input with the fuzzy condition

This computation might also be expressed using the function.

$$M(T, LOW) = Support\ min\big(\mu_t(\mathrm{X}), \mu_{Low}(X)\big) \qquad (1)$$

Where M is the degree difference and μ_t = is the function that determines the membership of set x over time and μ_{low} is the membership function with the fewest elements in set x.

9.2 Fuzzy Mapping Rules

By making advantage of linguistic variables, fuzzy mapping rules efficiently map input data to output data. A fuzzy mappings rule is based on a fuzzy graph, that is a representation of the connection among a fuzzy input and an output. In the case of real-world products, it may be challenging to isolate a single causal link between output and input, or the nature of that connection may be murky even at the time of creation. Fuzzy mapping rules can be useful in some scenarios. A set of fuzzy mapping rule should be used to assess the entire function, as each fuzzy rule mappings only value a subset of the operation. The following fuzzy rule mappings are built while still using AC as an example. When the temperature is low, the heater's engine should be turned on high speed. There need to be different regulations for various input temperatures [7].

In most practical situations, the input variables will contain multiple levels of detail. For instance, both the current temperature and the rate of change in temperature expressed in AC are included as inputs. It is recommended that the rules for fuzzy control be modified to take more inputs into account when determining the output.

Different rates of change in temperatures are related to IF clauses in IF-THEN rules. The control output is a three-dimensional variable associated with the THEN condition of IF-THEN rules, and it is located at the point of intersection of each row and each column. The speed of the heater engine must be FAST if the current temperature was low and the rate at which the temperature is changing is LOW.

One way to put this is in terms of the IF-THEN rule: if the temperature and the speed of change are both low, then the output must be fast. Several other laws also use a method that is strikingly comparable to human intuition. In this specific instance regarding an AC, nine guidelines are established. More granular separation of input and output and the use of fuzzy rules are required for applications requiring fine-grained control.

9.3 Fuzzy Implication Rules

An input-output correlation based on generalized logic. An implications rule of fuzzy describes it. Fuzzy implication rules are founded on the totality of fuzzy logic's core principles. Fuzzy implication follows some rules that have connections with both multiple-valued reasoning and traditional, two-valued logic. Using the air conditioner as an example further, we can deduce that a fast-heating motor is required if the indoor temperature is low. This leads to the conclusion that, and confirms, the temperature is quite high. The heating engine, it follows, needs to go into coast mode (or SLOW).

Similarity Relation - Fuzzification

When it comes to the immensely important matter of comparison, a straightforward model is often found to be wanting. The gradualness in the fuzzy sets reflects how gradualness is perceived. The fuzzy equivalency relation is applied to show how the members of a fuzzy set are associated with one another. With a fuzzy equivalent relation as our lens. By identifying similarities as the foundation of the fuzzy equivalence relation, fuzzy sets were viewed as a source of inspiration. Two elements that belong to an identical collection or its complement cannot be distinguished from one another in a fuzzy set. Following is an explanation of how the equivalency fuzzy relation is used to generate membership functions for fuzzy sets:

- **Definition 1**

The mappings E: V x V [0,1] satisfying the fuzzy equivalent relations on set V are denoted by

$(E1)\ E(v, v) = 1, v \in V$ (reflexivity)

$(E2)\ E(v_1, v_2) = E(v_2, v_1), v_1 v_2 \in V$ (symmetry)

$(E3)\ E(v_1, v_2) * E(v_2, v_3) \leq E(v_1, v_3), v_1 v_2 v_3 \in V$ (transitivity)

where the usual ordering is the unit interval notation [0, 1]. On rare occasions, E is referred to as a similarity connection. Some theorems and concepts are recalled for this context.

- **Definition 2**

If and only if $\mu_A\ (v_1) * E\ (v_1, v_2) \leq \mu_A(v_2)$ holds for every $v_1, v_2\ \in V$ then the fuzzy set $A \in [0,1]^u$ is extensional via the fuzzy equivalent relation E on V.

- **Definition 3**

Allow 'E' to be a fuzzy equivalence relation on V, and define $A\ \in\ [0,1]^u$.

The extensional hull of A about E is the fuzzy set $\hat{A} = \cap \{B | A\ \subseteq B$ are extensional concerning E.

8. PSEUDOCODE

Following is the Pseudo code, which shows initialization, Similarity evaluation, threshold, and Fuzzification process. Figure 6 shows pseudo-code.

- ➔ Initialization:
  - UQ ← User query
  - CB ← Case base
- ➔ For similarity evaluation:
  - Let the words be UQ and CB
  - Find the similarity for both the words
  - similarity1 = similarity(UQ)
  - similarity2 = similarity(CB)
  - Find the pair (sim1, sim2) where sim1∈similarity1 and sim2∈similsrity2 such that word_similarity(sim1, sim2) is maximized.
  - Find the similarity of both the synsets.
- ➔ Cases to set the threshold
  - Dissimilar ← similarity < 50%
  - Similar ← similarity > 50%
  - Most similar ← similarity > 75%
- ➔ Fuzzification
  - Given n fuzzy sets $A_1, A_2, \ldots, A_n$ defined over some universe of discourse V and $a \in V$.
  - Construct a fuzzy equivalence relation $\mathrm{E}(v_1, v_2)$ from $A_1, A_2, \ldots, A_n$.
  - Set $\delta > 0$. Define a fuzzy set about the point $a\ \in V$ from the fuzzy equivalence relation $\mathrm{E}(v_1, v_2)$ by
    $\mu_a(v_1) = E(v_1, a)$, in the interval $a - \delta\ \leq\ v_1\ \leq a + \delta, \delta > 0$
    $= 0,$ otherwise.

Figure 6: Pseudo code of the methodology

9. RESULTS

This section demonstrates the results of the suggested effort using machine learning and fuzzy logic. The solution was put to the test by extracting agricultural cases from a case base and conducting comprehensive trials to validate its efficiency. According to the suggested "problem–solution–association" representation, all examples were arranged in this way. Currently, simulated data was adopted, which were generated within a given range. The confusion matrix of the proposed methodology is shown below:

$$\begin{bmatrix} 56 & 0 \\ 2 & 45 \end{bmatrix}$$

In this, the absolute mean error (equation 2) is 1.014626457230125739, the root mean squared error (equation 3) is 0.0030972993429502953, and the mean square error (equation 4) is 0.05565338572764729. Figure 7 shows the parameters for the selected confusion matrix. The results obtained are classified into different categories as the most similar model, similar model, and dissimilar model. The most similar model is those that have similarities equal to more than 75 %, whereas the similar model is those that have similarities between 50% to 75% and the dissimilar models are having similarities less than 50%.

$$AME = \frac{\sum_{i=1}^{n}|Y_i - X_i|}{n} \qquad (2)$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x}_i)^2}{N}} \qquad (3)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{Y}_i)^2 \qquad (4)$$

where,

AME = Absolute mean error

RMSD = Root mean square deviation

MSE = Mean squared error

n = Total number of data points

N = Number of non-mixing data points

X_i = Observed value

Y_i = Prediction value

i = variable number

```
Mean Absolute Error: 0.014626457230125739

Root Mean Squared Error: 0.0030972993429502953

Mean Squared Error: 0.05565338572764729
```

Figure 7: Error values of the dataset

Fuzzy Rule: For evaluating the confusion matrix some fuzzy rules are implemented so the actual similarity match or dissimilarity can be measured more precisely. While implementing fuzzy rules some cases have been assumed (like case 0 is considered whenever a confusion matrix does not show any similarity and case 1 is considered whenever a confusion matrix shows similarity). Following such rules, the proposed techniques evaluate the similarity in the model.

rule1=ctrl.Rule(temperature['case 1'] & Pesticide_type['case 1'] & Soil_Type['case 1'],cases['case 1'])
rule2=ctrl.Rule(temperature['case 1'] & Pesticide_type['case 1'] & Soil_Type['case 0'],cases['case 1'])
rule3=ctrl.Rule(temperature['case 1'] & Pesticide_type['case 0'] & Soil_Type['case 1'],cases['case 1'])
rule4=ctrl.Rule(temperature['case 1'] & Pesticide_type['case 0'] & Soil_Type['case 0'],cases['case 0'])
rule5=ctrl.Rule(temperature['case 0'] & Pesticide_type['case 1'] & Soil_Type['case 1'],cases['case 1'])
rule6=ctrl.Rule(temperature['case 0'] & Pesticide_type['case 1'] & Soil_Type['case 0'],cases['case 0'])
rule7=ctrl.Rule(temperature['case 0'] & Pesticide_type['case 0'] & Soil_Type['case 1'],cases['case 0'])
rule8=ctrl.Rule(temperature['case 0'] & Pesticide_type['case 0'] & Soil_Type['case 0'],cases['case 0'])

9.1 **Scenario 1:** Results obtained in the Most Similar model
Input parameters for case 1 are 15 degrees temperature, pesticide type is 3, soil type taken as 2, crop damage is 2, and the wind speed is 12 m/s.

Result 1: In this, the accuracy (Equation 5) is 98.05%, precision (Equation 6) of the model is 100, recall (Equation 7) of the model is 96.55172413793103, F1 score (Equation 8) of the model is 98.24561403508771. Figure 8 shows the accuracy of the model. The results demonstrate the case 0 which means the model is matched.

$$\boldsymbol{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5)$$

$$\boldsymbol{Precision} = \frac{TP}{TP+FP} \qquad (6)$$

$$\boldsymbol{Recall} = \frac{TP}{TP+FN} \qquad (7)$$

$$\boldsymbol{F1\ Score} = 2\frac{Precision \times Recall}{Precision + Recall} \qquad (8)$$

where,

TP = True Positive Value

TN = True Negative Value

FP = False Positive Value

FN = False Negative Value

```
Accuracy of model is 98.05825242718447 Percent
Precision of model is  100.0
Recall of model is  96.55172413793103
F1 Score of model is 98.24561403508771
Your Value Matched  75.0 %
Case is 0
The most similar case
```

Figure 8: Evaluated parameters

9.2 **Scenario 2:** Results obtained in the Most Similar model
Input parameters for case 2 are 20 degrees temperature, pesticide type as 1, soil type taken as 1, crop damage as 1, and the wind speed as 12 m/s.

Result 2: In this, the accuracy is 98.05%, the precision of the model is 100, the recall of the model is 96.55172413793103, F1 score of the model is 98.24561403508771. Figure 9 shows the accuracy of the model.

```
Accuracy of model is 98.05825242718447 Percent
Precision of model is  100.0
Recall of model is  96.55172413793103
F1 Score of model is 98.24561403508771
Your Value Matched  100.0 %
Case is 0
The most similar case
```

Figure 9: Evaluated parameters

9.3 **Scenario 3:** Results obtained in a Similar model
Input parameters for case 3 are 26-degree temperature, pesticide type as 3, soil type taken as 1, crop damage as 1, and wind speed as 10 m/s.

Result 3: In this, the accuracy is 98.05%, the precision of the model is 100, the recall of the model is 96.55172413793103, F1 score of the model is 98.24561403508771. Figure 10(a) shows the accuracy of the model. Here the model does not get matched and thus it is represented by 1 (means not matched). Then the model again recalculates to find any match using the fuzzy rule and gets matched by 86.22% as shown in Figure 10(b).

```
Accuracy of model is 98.05825242718447 Percent
Precision of model is  100.0
Recall of model is  96.55172413793103
F1 Score of model is 98.24561403508771
Your Value Matched  0.0 %
Case is 1
```

```
86.2228670789143
Now The Value Matched
Now The Case is 0
The most similar case
     Case %
0  86.222867
```

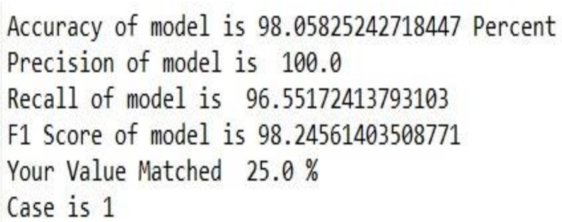
(a) **(b)**

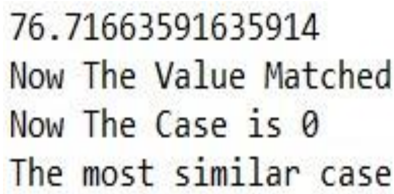
Figure 10: Evaluated parameters

9.4 **Scenario 4:** Results obtained in a Similar model

Input parameters for case 4 are 24-degree temperature, pesticide type as 1, soil type taken as 2, crop damage as 2, and wind speed as 13 m/s.

Result 4: In this, the accuracy is 98.05%, the precision of the model is 100, the recall of the model is 96.55172413793103, F1 score of the model is 98.24561403508771. Figure 11(a) shows the accuracy of the model. Here the model does not get matched and thus it is represented by 1 (means not matched). Then the model again recalculates to find any match using fuzzy logic and gets matched by 76.71% as shown in Figure 11(b).





(a) **(b)**

Figure 11: Evaluated parameters

Figure 12 shows the relationship established between membership and case (either case 1 or case 2) using the fuzzy rule for observing the matched case in the case of nonsimilar models and the system finds the similarity in dissimilar models.

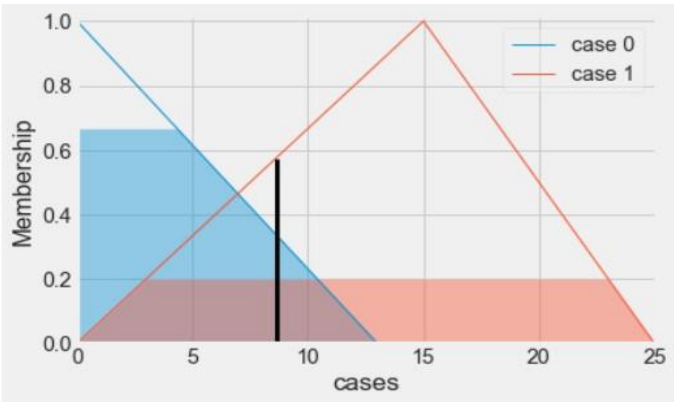


Figure 12: Relationship between membership and cases

9.5 **Scenario 5:** Results obtained in Dissimilar model

Input parameters for case 5 are 25-degree temperature, pesticide type as 1, soil type taken as 2, crop damage as 1, and wind speed as 16 m/s.

Result 5: In this, the accuracy is 98.05%, the precision of the model is 100, the recall of the model is 96.55172413793103, F1 score of the model is 98.24561403508771. Figure 13 shows the result. Here the model does not get matched and thus it is represented by 1 (means not matched). Then the model again recalculates to find any match using fuzzy but was unable to match.

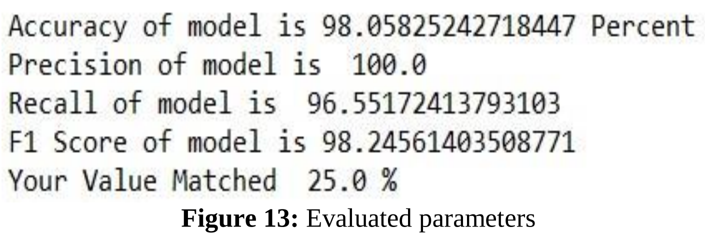


Figure 13: Evaluated parameters

9.6 **Scenario 6:** Results obtained in Dissimilar model

Input parameters for case 6 are 16-degree temperature, pesticide type as 1, soil type taken as 1, crop damage as 1, and wind speed as 10 m/s.

Result 6: In this, the accuracy is 98.05%, the precision of the model is 100, the recall of the model is 96.55172413793103, F1 score of the model is 98.24561403508771. Figure 14 shows the result. Here the model does not get matched and thus it is represented by 1 (means not matched). Then the model again recalculates to find any match but was unable to match.

```
Accuracy of model is 98.05825242718447 Percent
Precision of model is  100.0
Recall of model is  96.55172413793103
F1 Score of model is 98.24561403508771
Your Value Matched  25.0 %
```

Figure 14: Evaluated parameters

Figure 15 shows the relationship established between membership and case (either case 1 or case 2) using the fuzzy rule for observing the matched case in the case of nonsimilar models and the system finds the similarity in dissimilar models.

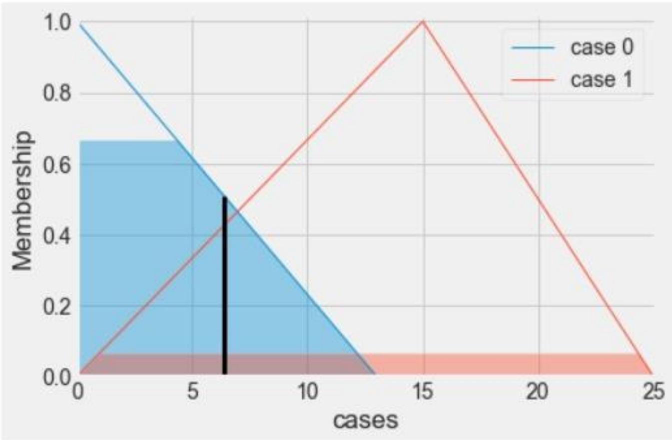


Figure 15: Relationship between membership and cases

Zhai et al., (2020) [28] case base maintenance research and development Work on developing an effective case retrieval algorithm for use in agricultural case-based reasoning systems. The author achieved an MAE equal to 0.02273 and RMSE equal to 0.0252. The results of the research are compared with Zhai et al., 2020 [28] on the bases of MAE and RMSE parameters with the proposed methodology. Table 3 shows the discussed comparison and the comparison clearly shows that the proposed methodology has achieved better results.

Table 3: Comparison of MAE and RMSE values of previous studies

Author	MAE	RMSE
Zhai et al., 2020 [28]	0.02273	0.0252
Proposed Methodology	0.01462	0.0030

10. CONCLUSION

This study proposed a case representation strategy used for a case-based reasoning system, and it focused specifically on agricultural cases. Since conventional case representation methods do not take into account the interconnected nature of cases, a large case base may result in inefficient retrieval. Therefore, a case representation approach was used to implement the methodology outlined in this work. The notion differs from the usual attribute-value pair format in that it investigates the relationships between individual situations. The relationship considers not just examples that are comparable to the target one, but also those that are distinct from it. Because similar cases are selected for comparison, we can quickly get relevant cases. On another hand, making connections that are dissimilar to the case at hand can aid in the early phases of recall. The proposed case representation strategy may be used to expedite case retrieval by visiting fewer instances while maintaining accuracy. Efficient case retrieval is fundamental to case-based reasoning, and the development of the other processes depends on it. A case-based reasoning system might greatly benefit from the approach of related case representation that is being suggested. With the help of case-based reasoning and the suggested case representation method, farmer-centric retrieval systems may be able to better aid farmers in the management of agricultural operations and the making of educated decisions. In addition, the idea offers a system for effectively managing information in the agricultural sector.

In the future, want to get data from actual fields to test the suggested related case representation technique and the quick retrieval methodology. This is further shown by the experiment's results, which show that the proposed quick case retrieval technique does not consider the mining of different associations between instances. As a result, the suggested approach has to be refined even more.

REFERENCES

[1] Schimmelpfennig, D. (2016). Farm profits and adoption of precision agriculture (No. 1477-2016-121190).

[2] Li, R., di Virgilio, N., Guan, Q., Feng, S., & Richter, G. M. (2013). Reviewing models of land availability and dynamics for biofuel crops in the United States and the European Union. Biofuels, Bioproducts and Biorefining, 7(6), 666-684.

[3] Ostwald, M., Jonsson, A., Wibeck, V., & Asplund, T. (2013). Mapping energy crop cultivation and identifying motivational factors among Swedish farmers. Biomass and Bioenergy, 50, 25-34.

[4] Acevedo, M. F. (2011). Interdisciplinary progress in food production, food security and environment research. Environmental conservation, 38(2), 151-171.

[5] Lovett, A. A., Sünnenberg, G. M., Richter, G. M., Dailey, A. G., Riche, A. B., & Karp, A. (2009). Land use implications of increased biomass production identified by GIS-based suitability and yield mapping for Miscanthus in England. Bioenergy Research, 2, 17-28.

[6] Sanscartier, D., Deen, B., Dias, G., MacLean, H. L., Dadfar, H., McDonald, I., & Kludze, H. (2014). Implications of land class and environmental factors on life cycle GHG emissions of Miscanthus as a bioenergy feedstock. Gcb Bioenergy, 6(4), 401-413.

[7] Nassar, A. M., Harfuch, L., Bachion, L. C., & Moreira, M. R. (2011). Biofuels and land-use changes: searching for the top model. Interface Focus, 1(2), 224-232.

[8] Kamilaris, A., Gao, F., Prenafeta-Boldu, F. X., & Ali, M. I. (2016, December). Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications. In 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT) (pp. 442-447). IEEE.

[9] Ahmed, N., De, D., & Hussain, I. (2018). Internet of Things (IoT) for smart precision agriculture and farming in rural areas. IEEE Internet of Things Journal, 5(6), 4890-4899.

[10] Watson, I., & Marir, F. (1994). Case-based reasoning: A review. The knowledge engineering review, 9(4), 327-354.

[11] Marling, C., Montani, S., Bichindaritz, I., & Funk, P. (2014). Synergistic case-based reasoning in medical domains. Expert systems with applications, 41(2), 249-259.

[12] Islam, O., Alfakeeh, A., & Nadeem, F. (2017). A framework for effective big data analytics for decision support systems. Int. J. Comput. Netw. Appl.(IJCNA), 4(5), 129-137.

[13] Kovács, S. (2013, July). Adapting the scale and move FRI for the fuzziness interpolation of the double fuzzy point rule representation. In 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-7). IEEE.

[14] Chakraborty, B., Ghosh, D., Garnaik, S., & Debnath, N. (2010, July). Knowledge management with case-based reasoning applied on fire emergency handling. In 2010 8th IEEE International Conference on Industrial Informatics (pp. 708-713). IEEE.

[15] Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. Computers and electronics in agriculture, 155, 257-282.

[16] Ben Ayed, R., Ennouri, K., Ben Amar, F., Moreau, F., Triki, M. A., & Rebai, A. (2017). Bayesian and phylogenic approaches for studying relationships among table olive cultivars. Biochemical genetics, 55, 300-313.

[17] Goap, A., Sharma, D., Shukla, A. K., & Krishna, C. R. (2018). An IoT based smart irrigation management system using Machine learning and open source technologies. Computers and electronics in agriculture, 155, 41-49.

[18] ALMarwi, H., Ghurab, M., & Al-Baltah, I. (2020). A hybrid semantic query expansion approach for Arabic information retrieval. Journal of Big Data, 7(1), 1-19.

[19] Saiz-Rubio, V., & Rovira-Más, F. (2020). From smart farming towards agriculture 5.0: A review on crop data management. Agronomy, 10(2), 207.

[20] Chaterji, S., Delay, N., Evans, J., Mosier, N., Engel, B., Buckmaster, D., ... & Chandra, R. (2020). LATTICE: Machine Learning, Data Engineering, and Policy Considerations for Digital Agriculture at Scale. IEEE internet of things journal.

[21] Liang, Z., Liu, X., Xiong, J., & Xiao, J. (2020). Water allocation and integrative management of precision irrigation: A systematic review. Water, 12(11), 3135.

[22] Patel, S. K., Sharma, A., & Singh, G. S. (2020). Traditional agricultural practices in India: an approach for environmental sustainability and food security. Energy, Ecology and Environment, 5, 253-271.

[23] Dubois, A., Teytaud, F., & Verel, S. (2021). Short term soil moisture forecasts for potato crop farming: A machine learning approach. Computers and Electronics in Agriculture, 180, 105902.

[24] Wang, J., Bell, M., Liu, X., & Liu, G. (2020). Machine-learning techniques can enhance dairy cow estrus detection using location and acceleration data. Animals, 10(7), 1160.

[25] Wang, G., Wei, Y., Qiao, S., Lin, P., & Chen, Y. (2018). Generalized inverses: theory and computations (Vol. 53). Singapore: Springer.

[26] Leistra, E., Visser, M., van der Hout, A., Langius, J. A., & Kruizenga, H. M. (2014). Systematic screening for undernutrition in hospitals: predictive factors for success. Clinical Nutrition, 33(3), 495-501.

[27] Marling, C., Montani, S., Bichindaritz, I., & Funk, P. (2014). Synergistic case-based reasoning in medical domains. Expert systems with applications, 41(2), 249-259.

[28] Zhai, Z., Ortega, J. F. M., Martínez, N. L., & Xu, H. (2020). An efficient case retrieval algorithm for agricultural case-based reasoning systems, with consideration of case base maintenance. Agriculture, 10(9), 387.