I.J. Information Engineering and Electronic Business, **2025, 6, 48-59**
Published Online on December 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2025.06.04

DFI-ADR: Fuzzy Logic-Driven Information Retrieval and Machine Learning for Environmental and Crop Prediction to Optimize Farming Decisions

Surabhi Solanki*
Banasthali Vidyapith, Rajasthan, CO-304022, India
Bennett University, Greater Noida, CO- 201310, India
E-mail: surabhisolanki.cse@gmail.com
ORCID iD: https://orcid.org/0000-0001-7531-7131
*Corresponding Author

Seema Verma
National Institute of Technical Teachers' Training and Research, Bhopal, CO- 462002, India
E-mail: seemaverma3@yahoo.com
ORCID iD: https://orcid.org/0000-0001-6299-2576

Received: 29 October, 2024; Revised: 12 January, 2025; Accepted: 11 May, 2025; Published: 08 December, 2025

Abstract: This paper proposes DFI-ADR (Dynamic Fuzzy Information System with Agriculture Decision Retrieval) aimed at improving agricultural decision-making through case-based reasoning and precise information retrieval. This approach uses fuzzy logic and machine learning techniques, such as IndRNN, to compute similarity scores between historical agricultural cases and new queries. This enables dynamic classification of cases as "distinct," "similar," or "highly comparable" based on fuzzy membership values. These values significantly enhance the accuracy of decisions related to agricultural factors like crop yield, soil quality, and irrigation. The methodology outperforms traditional methods in terms of accuracy, recall, and precision, proving highly effective for agricultural analysis and decision-making. In experiments with the Agriculture Dataset Karnataka, DFI-ADR achieved an accuracy of 95%, a precision of 100%, and an F1-score of 94.74%, significantly outperforming traditional methods by a margin of 10-15% across these metrics. These values demonstrate its effectiveness for agricultural analysis and decision-making.

Index Terms: Information Retrieval, Fuzzy Logic, Query Expansion, Recall, Precision, F1-Measure, Indrnn

1. Introduction

The Information retrieval is the technique of finding the important data in unstructured data or semi-structured data so that data from a large database, which is very important, becomes easy to trace. Machine Learning (ML) solves issues to obtain the input-output variables relationship. Learning signifies the acquire explanations from examples of what is being explained. ML should not create assertions about the proper form of the data model, that represents the data, distinct conventional statistical approaches. This functionality is handy for modeling complex non-linear behaviors, Similarity the prediction function for crop yields. The supervised learning approach (independent variables) is the most efficient use of machine learning approaches for farming decisions can be predicted from a given group of forecasts. These variable sets are used to develop a function that translates I/P to desired results.

The testing process remains until a desired degree of precision is reached on the training data by the algorithm. Supervised learning examples consist of: Decision Tree, Random Forest, Regression, KNN Logistic Regression, and many more. The knowledge structure for smallholder farmers in Malaysia was studied in one of the initial survey [1]. In this paper we find the application of environmental and crop production forecast data in agricultural decision-making processes. By analyzing their key characteristics and features, we hope to make clear how crucial these datasets are in helping make decisions about which crops to plant, how to grow them, how to distribute resources, and how to manage risk.

Crop yield-prediction data are also important. They enumerate the output both in the past and the future in consideration of the time of planting and the rotation of crops to be planted and relevant strategies for optimizing the yield. The historic crop yield data, weather forecast data, and predictive models are used to make predictions and identify areas of risks for the policymakers in advance. Farmers are provided the data in advance for them to take measures to ensure overall farm profitability and resilience.

Additionally, recall, accuracy, confusion matrices, and the F1 measure are typically used to evaluate predictive models[4] for use in agricultural decision-making. These metrics allow one to assess how well a model determines and classifies potential errors and events. The inclusion of such performance indicators in the process of making agricultural decisions would increase the reliability of the prediction model. This puts them towards data-driven strategies for crop management and risk reduction. Considering the importance of environmental and crop yield forecast data, this research article presents its critical characteristics in agricultural decision-making. Moreover, the methods of data gathering, analysis, and interpretation will be discussed, considering the relevant applications and case studies that reveal the real influence of the use of these data sets in agriculture.

Agricultural decision-making often involves unstructured data, making information retrieval challenging. While traditional machine learning methods offer solutions, they struggle with non-linear complexities and domain-specific challenges, such as diverse environmental data and crop yield predictions. This research addresses these limitations by proposing DFI-ADR, a novel system combining fuzzy logic and IndRNN for dynamic case-based reasoning. Unlike conventional models, DFI-ADR dynamically adjusts similarity thresholds to handle diverse agricultural conditions, offering more accurate and actionable insights. This paper focuses on enhancing agricultural decision-making by leveraging environmental and crop yield data, filling critical gaps in existing information retrieval frameworks.

The following paper is elaborate as follows Section 2 defines the related work find out the gap analysis among various information retrieval methods. Section 3 explains methodology and delineates its main steps. Section 4 discusses the result compare proposed methodology(DFI-ADR) with the existing methodology. Section 5 concludes the paper, summarizing key findings and insights.

2. Related Work

Problems Over a span of time, under different expansion processes, the query development method of information retrieval has been constantly checked and applied. Obviously, this has yielded positive recovery outcomes. Though, because the web of experience is presently swerving, a technique for extending knowledge-based queries is needed [18]. Wache, Kung, H. Y., et al. (2016) [19] maintained that an ontology-based system's retrieval mechanisms attract research interest in the field. They carried out their study on the methodology of ontology-based question extension for the agriculture domain. Petridis, V. & Kaburlasos, V. G. (2003) [13] centered on database expansion applying the global study to enhance the efficiency of agricultural domain ontology knowledge retrieval utilizing [22]. In terms of memory and accuracy measures, the sole objective of the analysis is to establish the influence of ontology. The ontology written in XML has been developed by [17].Salleh, M. N. M. (2012) [14] suggested an expansion-based query information retrieval approach specifically for rice domain plant development ontology. Query extension algorithm in the sports domain for semantic information retrieval to facilitate search over large document repositories

Table 1 exemplifies the various methods applied for information retrieval in agriculture, medical, and decision support systems. Such methods are not limited to Google API, neural networks, and peer-to-peer networks, but they aim to achieve certain goals like better question-answering systems, agricultural data retrieval, and inflection reduction using stemming and lemmatization. The results showed enhanced efficiency, precision, and information personalization in specific fields.

Table 1. Various Information Retrieval Methods

Context Consider	Method used	Purpose	Results
google API [3]	google API	Submit the keywords	developed an application
neural network [4]	neural network	retrieve agriculture information	Retrieve agriculture Information and to integrate all the data
Vector-space model [5]	improved model of vector space model	To propose the Question-Answering	It betters the result in terms of recall, precision and efficiency.
Information Retrieval [6]	Information based on existing ontology	information about the agents	Develop framework for personalizes the information about the agents
Question Answering System [7]	NLP and information retrieval system	To develop Question-answering system	Question-answering system for agriculture
Big data [8]	Graph	chronological order	information gets in chronological order
Medical [9]	Bayesian network	information retrieval in medical	IR model based on Bayesian network.
Agriculture [10]	peer to peer network	peer-to-peer network	Information Retrieval In agriculture for Geo-graphically Networks
DSS, Agriculture [11]	DSS	DSS for farmers	Develop the DSS for agriculture
Stemming and lemmatization [12]	stemming and lemmatization	reduce the inflection	Reduce the inflection and comparison using stemming and lemmatization

Table 2 elaborates the applications of machine-learning based approaches in information retrieval: probabilistic classification, clustering, and Support Vector Machines. These applications yield results concerning improved precision and learning performance, with applications in document retrieval, ranking, and information extraction.

Table 2. Information Retrieval System using ML methods

Context Considered	Method used	Results
IR(Information-Retrieval) [13]	probabilistic classification	Binary value (yes/ no)
IR(Information-Retrieval) [14]	Supervised learning approach (C4.5)	Standard deviation of test and training set= 1.88, 7.46
Image -Retrieval [15]	Clustering	Precision=20
Document-Retrieval [16]	Support Vector Machine to relevance feed back	Retrieval and learning performance =0.341,0.293
Ranking [17]	Point wise, pair wise, list wise method	Rank machine learning problems
Information Extraction, Information Retrieval [18]	Support Vector Machines	Coarse ABBR(P%) = 99.6,(R%) = 88.89, (A%) = 88.89
Document retrieval [19]	Support Vector Machine classifier.	Precision = 0.6
Learning to rank [20]	Reduced Support Vector Machine	RSVM methods Classification, Regression
Document retrieval [21]	Support Vector Machine	Precision=98.52%, recall=97.74%

Although various methods and machine learning-based approaches have been applied across the domains such as agriculture, medical, and decision support, which enhance precision, recall, and efficiency in information retrieval systems, there is still a need for refinement in the query expansion methodologies to adapt well to newly emerging complexities brought by data and/or domain-specific challenges.

3. Research Methodology

The background study [2] aims to expedite the information retrieval methodology for small datasets that associated with cases. The user input is processed as a case and routed for similarity analysis and retrieval of previous cases. Identified gaps include unspecified cases from user queries, unrealistically assessed Similarity, and static threshold definitions. In our proposed methodology DFI-ADR we address these gaps: Firstly, addressing the need for query expansion due to users' limited familiarity with technical terms in agriculture. Secondly, enhancing system accuracy for large datasets by integrating IndRNN for efficient Similarity evaluation. Lastly, adopting dynamic fuzzy rules for threshold definition, accommodating changes in dataset type and category. These solutions aim to improve efficiency and accuracy in information retrieval processes.

3.1 Dataset Details and Preprocessing

The experimentation was based on the Agriculture dataset from Karnataka, which is accessible on Kaggle. This dataset contains multiple attributes that provide detailed information pertaining to agricultural practices and conditions. The dataset consists of 3158 rows and 12 attributes [15]. The fewer of the main environmental data properties in farming are summarized in Table 3 and Table 4.

Table 3. Key Environmental Data Attributes in Agricultural Scenarios

Attribute	Description
Year	Year of data collection
Location	Geographical location of the farm
Area	Area of the farm in hectares
Rainfall	Annual rainfall in millimeters
Temperature	Average temperature in degrees Celsius
Soil Type	Type of soil (e.g., sandy, clay)
Irrigation Method	Method of irrigation used
Yields	Crop yields in kilograms per hectare
Humidity	Average humidity percentage
Crops	Type of crops grown
Price	Market price of the crops
Season	Season during which the data was collected

Table 4. Key Environmental Data Attributes in Agricultural Scenarios based on Crop

Attributes	Description
Crop yield	Crop yield prediction

To ensure the dataset is suitable for analysis and modeling, several preprocessing steps were undertaken. These steps help in cleaning, transforming, and structuring the data for effective use in machine learning algorithms and case-based reasoning systems.

- Data Cleaning: Missing values in the dataset were dealt with by either deleting rows with considerable missing values or finding imputed values such as mean or median of the particular columns.
- Normalization: Continuous variables such as rainfall, temperature, and humidity were normalized to be comparable.
- Categorical Encoding: Categorical variables like soil type and method of irrigation were encoded with one-hot encoding to bring them to numerical form so that they can be used with machine-learning algorithms.
- Feature Selection: Relevant features were determined based on their correlation with the target variable (crop yield) to improve model performance.

3.2 Similarity Cases

In the context of environment and crop or plant-related results, identifying similar case is crucial. Each condition is paired with three identical instances, identified by case IDs and Similarity dimensions associated with case IDs. When a newly case is associated with the previous one, the related instances of the past case are also compared to identify the most similar possible case among them. This association of related cases enables evaluating correlations within a narrower scope, reducing the number of cases and leading to greater efficiency in case retrieval.

Cosine similarity is the metric which is used primarily to calculate similarity score and actually measures the cosine of the angle between two vectors in the high-dimensional spaces. This nature ensures that the score reflects how close a query is in comparison with the historical cases. Cosine similarity scores are calculated using the following formula:

$$\text{Cosine Similarity} = \frac{\vec{A}\cdot\vec{B}}{|\vec{A}||\vec{B}|} \tag{1}$$

Where $\vec{A}, \vec{B}$ are the vector representations of the query and case, respectively, and $|\vec{A}|, |\vec{B}|$ are their magnitudes. Obtained similarity scores are classified into three boundaries using a fuzzy logic framework. Cases whose scores are less than 0.50 are indicated as "Distinct," which means they are not similar to the previously historical cases and should be flagged for papers and expert review or some special recommendation. Cases whose score is between 0.50 and 0.75 are called "Similar," meaning that the match is not implicit but is somewhat similar, and the change in recommendation should be slight. Birth scores greater than 0.75 are identified as Highly Similar, indicating a strong alignment with former cases and an immediate transfer of past successes to current cases. These thresholds are suggested with some degree of flexibility and adaptivity to certain datasets or conditions through fuzzy membership functions.

3.3 Distinct Cases

Conversely, distinct instances are considered to have relevant variations from the target case in environment and plant or crop related results. Three different cases are identified, along with their case IDs and corresponding comparison metrics. At the outset of case retrieval, the organizational relationship of distinct cases is utilized to aid in locating relatively similar cases for the latest case. If the connection between the latest case and a previous one exceeds the defined threshold, a comparison is made.

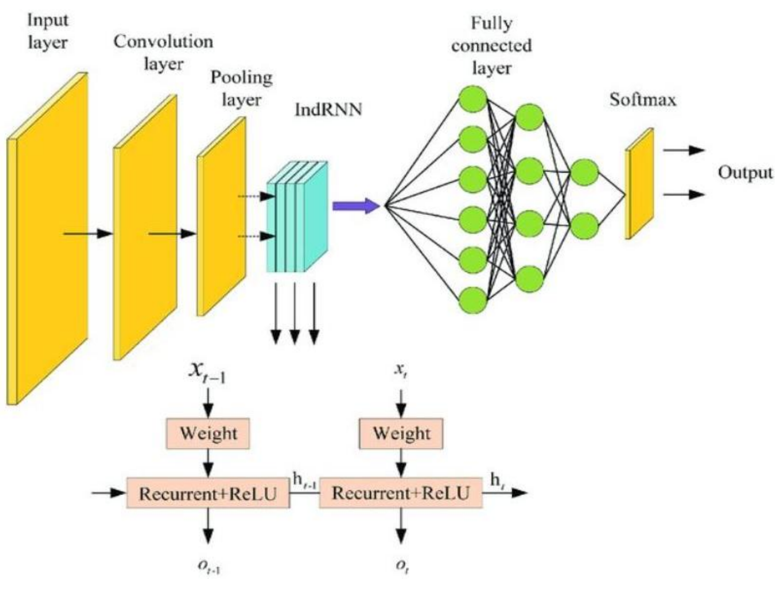

Fig. 1. Recurrent Neural Network [22].

For Similarity evaluation, the IndRNN (Independent Recurrent Neural Network) is employed. IndRNN preserves knowledge about what happened in the past based on the time levels. Distinct traditional neural networks, RNNs utilize base cell states to analyze input parameters and dependencies of time for predicting future outputs from sequential input data. IndRNN effectively maintains information from past inputs, making it useful in language processing, speech recognition, and system analysis. Fig. 1 illustrates the model of IndRNN.

3.4 DFI-ADR (Dynamic Fuzzy Information System with Agriculture Decision Retrieval)

3.4.1 Threshold Resolution Procedure

This In delineating the threshold resolution procedure, the following guidelines merit careful consideration:

- **Rule 1**
  Identifying Distinct Cases: A case (new case) is deemed distinct to the existing case if Similarity is between the 2-cases falls below 50%. A distinct flag is assigned to the new case.
- **Rule 2**
  Identifying Identical Situations: If Similarity score between the cases exceeds 50%, the new case is classified as identical to the source case. The Similarity flag is affixed to the existing case.
- **Rule 3**
  Identification of Related Cases: When Similarity score between cases surpasses 75%, the base case is regarded as extremely similar to the source case. An additional flag indicating extreme Similarity is allocated to the base case.
- **Rule 4**
  Selection of Connection Preference: In subsequent iterations, the choice between identical and distinct connections is determined. If the no. of homogenous flags outweighs distinct flags, the base case with the highest homogenous value is selected. Conversely, if distinct flags predominate, the base case with the deepest Similarity is chosen.
- **Rule 5**
  Recognition of Strikingly Similar Events: When two cases remain significantly distinct to the target case, the similar relation in the extremely similar case is mandated for decision of comparison in the next step if one case's distinct exceeds 75%.
- **Rule 6**
  Involvement of Prior Cases: Since a prior case may be linked in a one-to-many manner, all three cases in a specific iteration are historically compared. The detailed scenario for the next step is selected from the given previous phase, and the closed cases of the two most relatable (or distinct) cases selected for compare based on Rules 4 and Rule 5.

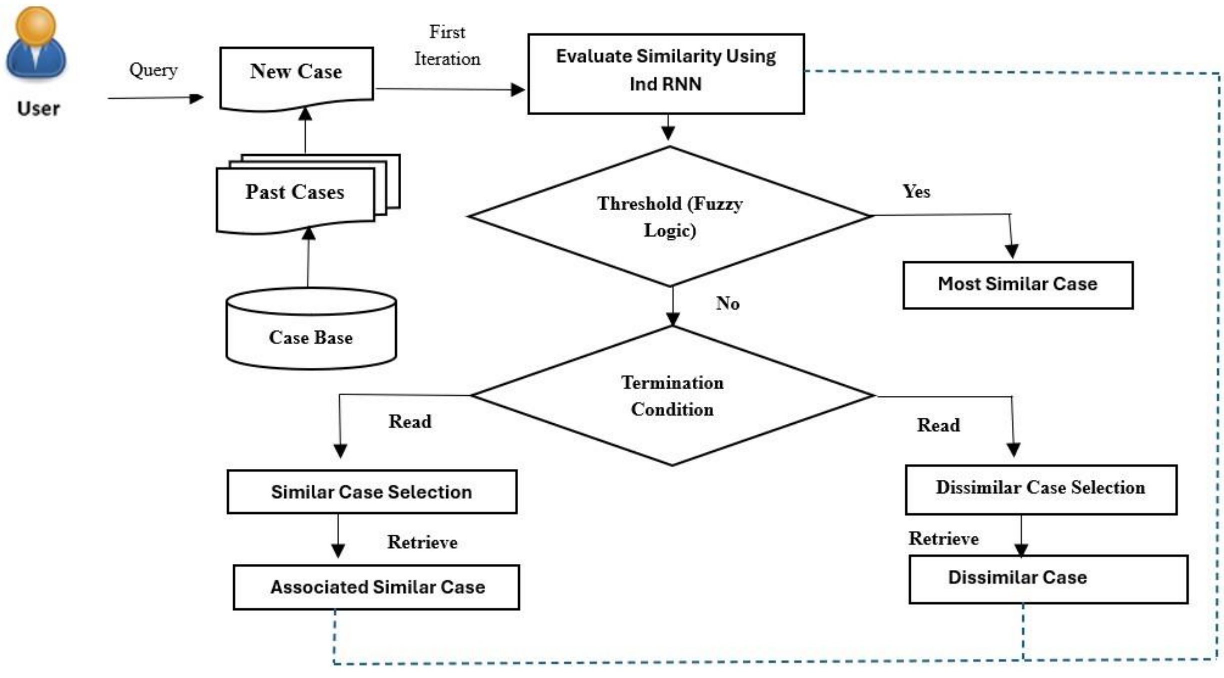

Fig. 2. Control flow diagram of Dynamic Fuzzy Information System with Agriculture Decision Retrieval

In Fig. 2. illustrates the algorithm where a user query is compared to past cases using IndRNN for similarity evaluation. Fuzzy logic-based thresholds are applied to categorize cases as most similar, similar, or dissimilar. Based on the evaluation, either similar or dissimilar cases are selected for further analysis.

Algorithm: Similarity-Based Query Evaluation with Fuzzification

---

Initialization: Algorithm: Similarity-Based Query Evaluation with Fuzzification
Output:
Step 1: Initialization
1. Define Inputs:
  - UQ (User Query)
  - CB (Case Base)
2. Calculate Similarity:
  - For each query, compute the Similarity scores:
    - aff1 ← Similarity score of UQ
    - aff2 ← Similarity score of CB
3. Maximize Similarity Pair:
  - Find the optimal pair (aff1, aff2) that maximizes Similarity between words from UQ and CB.
  - Use Similarity scores from pre-trained embeddings like Word2Vec or cosine similarity between word vectors.
4. Evaluate Similarity of Synsets:
  - Synsets (groups of synonymous words) are evaluated to further refine the Similarity between UQ and CB.

Step 2: Threshold Categorization
1. Set Thresholds:
  - Distinct: If Similarity score < 50%, mark as distinct.
  - Similar: If Similarity score > 50%, mark as similar.
  - Highly Similar: If Similarity score > 75%, mark as highly similar.

Step 3: Fuzzification
1. Fuzzy Sets Definition:
  - Define fuzzy sets N1, N2, ..., Nm ,where n∈V (Universe of Discourse).
2. Equivalence Fuzzy Relation:
  - Create a Equivalence Fuzzy Relation E(V1, V2) for each fuzzy set from the Similarity scores.
3. Set δ for Fuzzy Interval:
  - Define the value δ > 0 to set the fuzzy interval for Similarity scores.
    - Set $\mu_0 (v_1) = E(v1, n)$ for the interval $n-\delta \leq v1 \leq n+\delta$, else $\delta = 0$.

Step 4: Output
- Return the categorized cases based on Similarity evaluation (Distinct, Similar, Highly Similar).
- Use the fuzzy sets and equivalence relations to adjust the thresholds dynamically based on case similarity.

This algorithm evaluates the similarity between a user query (UQ) and a case base (CB) by calculating likeness scores. It categorizes cases based on predefined thresholds (dissimilar, similar, most similar) and applies fuzzification to handle variations in likeness through fuzzy sets and equivalence relations, enhancing accuracy in decision-making for information retrieval.

4. Result Analysis

This work presents a novel approach to improve agricultural information retrieval, especially for small, leading to significant performance and accuracy gains. To meet the unique requirements of agricultural information retrieval, we provide a comprehensive approach that combines query expansion, IndRNN based Similarity evaluation, and dynamic fuzzy threshold definition.

Table 5 builds out the hyperparameters setup and training metrics that will be applied in the IndRNN model within the scope of the proposed methodology. Such parameters include the number of layers, hidden-layer neurons, activation function, loss function, optimizer, learning rate, epochs, and batch size. They are key to setting the architecture and training of the IndRNN model. The IndRNN model comprises 3 layers and has128 neurons in the hidden layer. A ReLU activation function introduces non-linearity to the model. The MSE is the loss function that enables the capturing of the differences between the predicted and actual values. The Adam optimizer will use efficient gradient descent with a learning rate of 0.001. The model shall be trained for 50 epochs with a batch size of 32, thus allowing the model to learn effectively from the data.

Table 5. IndRNN Configuration and Training Metrics

Parameter	Value
Number of Layers	3
Hidden Layer Neurons	128
Activation Function	ReLU
Loss Function	Mean Squared Error
Optimizer	Adam
Learning Rate	0.001
Epochs	50
Batch Size	32

Table 6 defines the fuzzy membership functions used in the proposed methodology. The membership functions categorize the similarity scores into three categories: Low Membership, Medium Membership, and High Membership. Each category has a range of similarities, to which a membership function applies. The extreme membership function: Low membership is 0.5, therefore, any score ≤ 0.5 is considered low membership. Medium membership is defined between 0.5 and 0.8. Each membership function will be defined for the Low, Medium and High membership function. The fuzzy membership functions are defined as follows: Low membership is defined for similarity scores less than or equal to 0.5, the membership function μ_low(x) is used. Medium membership is defined for similarity scores between 0.5 and 0.8, the membership function μ_medium(x) is used. High membership is defined for similarity scores greater than 0.8, the membership function μ_high(x) is used. They serve to group the similarity scores, a step forward in facilitating more accurate and relevant decisions based on how similar a new query is to historical cases.

Table 6. Fuzzy Membership Categorization Based on Similarity Range

Membership Category	**Similarity Range**	**Membership Function**
Low Membership	$x \leq 0.5$	$\mu\_low(x) = \max(0, \min((0.5 - x)/0.2, 1))$
Medium Membership	$0.5 < x \leq 0.8$	$\mu\_medium(x) = \max(0, \min((x - 0.3)/0.2, (0.8 - x)/0.2))$
High Membership	$x > 0.8$	$\mu\_high(x) = \max(0, \min((x - 0.7)/0.2, 1))$

It would describe how the similarity score is mapped to fuzzy categories, with lower scores leading to high membership in the "Low" category, middle-range scores in the "Medium" category, and higher scores in the "High" category.

In Table 7 the Similarity score will be classified by means of fuzzy logic: Low, Medium, and High membership. For scores less than 0.50, the Similarity would be low; 0.50-0.80 comes in medium; and if the scores above 0.80 then it's high. This system helps evaluate how much similarity each case has based on its Similarity score. The calculation of the similarity score categorization and fuzzy logic categorization for the DFI-ADR method, it is imperative first to establish how Similarity scores and fuzzy memberships result from a model. Considering that the DFI-ADR method probably uses a similarity-based method, such as cosine similarity or another metric, the steps would include computing the scores for each comparison and assigning them to categories. Then, membership values based on fuzzy logic could also be calculated based on thresholds or fuzzy rules.

Table 7. Categorization of Cases Based on IndRNN and Fuzzy Logic

Case Column	**Similarity Score**	**Low Membership**	**Medium Membership**	**High Membership**	**Fuzzy Category**
Predicting crop yield based on soil moisture	0.92	0.0	0.2	0.9	Highly Similar
Temperature and humidity affecting crop growth	0.67	0.1	0.7	0.3	Similar
Effects of rainfall on soil quality	0.42	0.8	0.2	0.0	Distinct
Soil salinity's impact on crop yield	0.79	0.0	0.3	0.8	Highly Similar
Irrigation methods and soil fertility	0.55	0.2	0.8	0.1	Similar

Fig. 3. represents the compared cases of Similarity Score Categorization of agricultural factors. The similarity of each case to the original ones is measured by the similarity score on a scale of 0 to 1. This graph displays the level of the similarities to the known cases just as to those that are unknown. The division of the cases: The last sentence could be: A higher similarity score of 0.92 like a 0.92 is like a priory case, which is Highly Similar, and a lower score like 0.42 ranks the case as Distinct. A high score of 0.92 makes the case "Highly Similar" bur, for example, another case gets a low score of 0.42 that imposes a case as distinct. This is easily done by finding the cases with statistical similarities for further analysis. Once the Similarity score is calculated, we apply fuzzy logic rules to determine Low,

Medium, and High membership values. This could be based on as shown in Table 6.

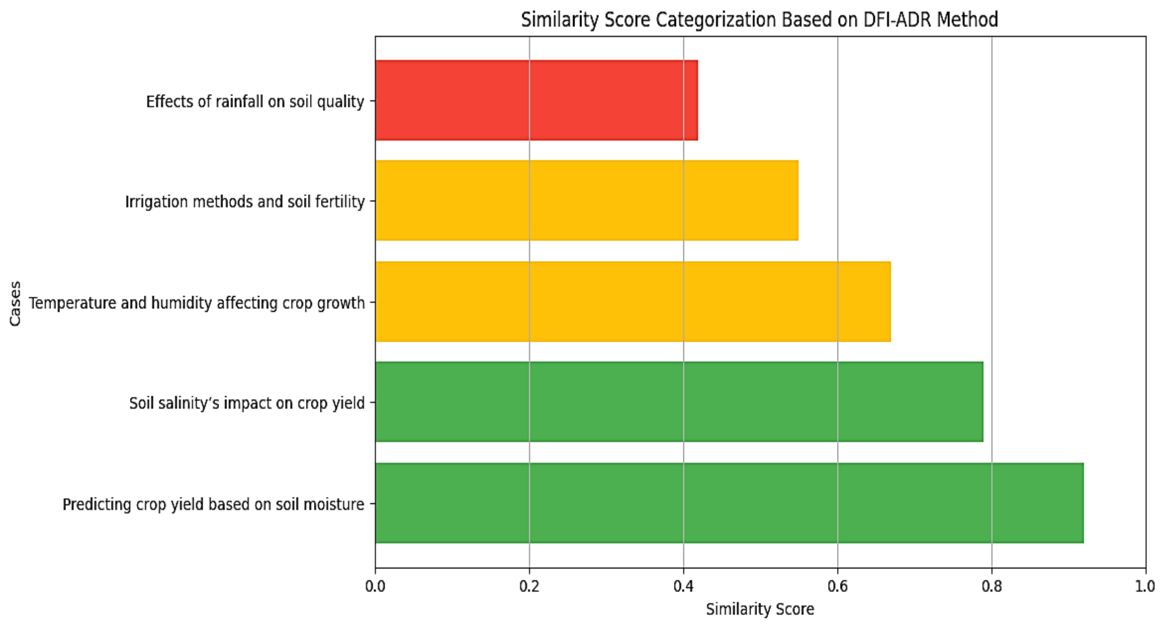

Fig. 3. Similarity Score Categorization Based on DFI-ADR Method

In Fig. 4. line graph illustrates Fuzzy Membership Values for Case Comparisons, showing low, medium, and high membership values across different cases. By using distinct line plots, it helps in understanding the distribution of similarity between cases. For instance, a high membership value near 0.9 for Soil Moisture indicates strong comparability, while a high low-membership score for Rainfall Quality suggests that it is more distinct. Overall, the graph helps categorize each case based on its similarity score within fuzzy categories.

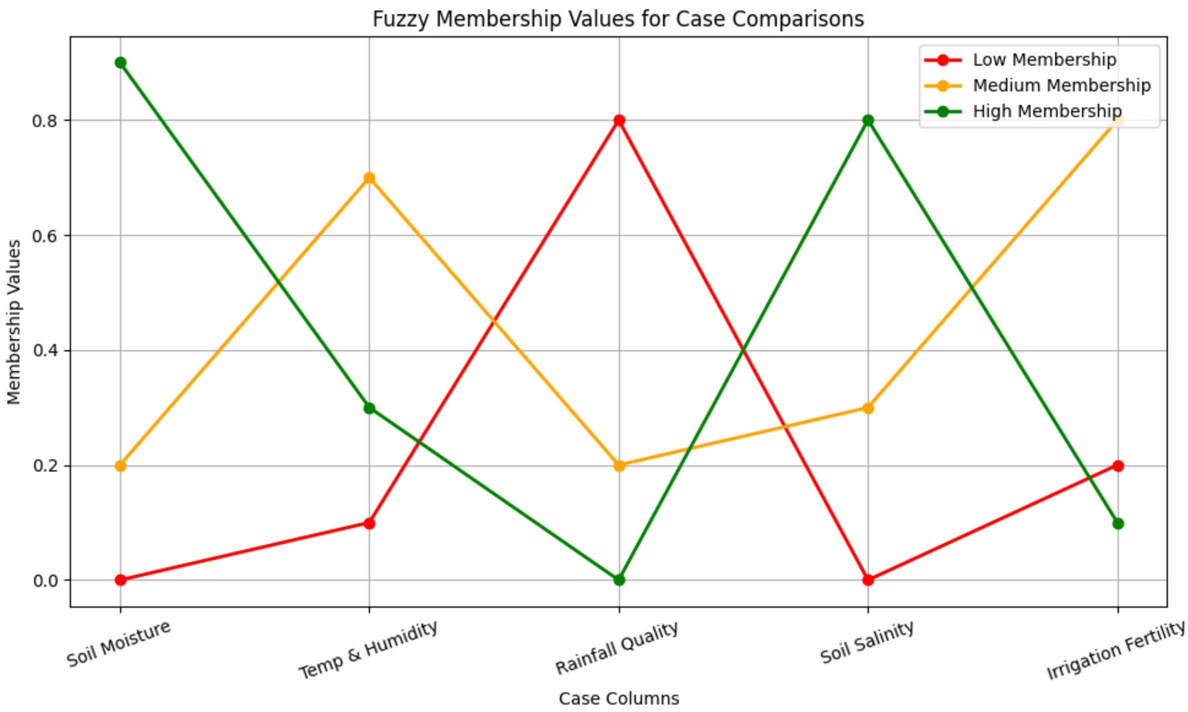

Fig. 4. Fuzzy Membership value for Case Comparison

Table 8 presents an illustration of how the similarity score evaluation can be employed for the farmer's query with the Agriculture Dataset Karnataka. The input parameters such as type of soil, rainfall, and irrigation method are used to produce vector embeddings and then compared with the historical data. Cosine similarity scores are worked out to observe how well the query matches with past cases. These scores are categorized into Distinct (<0.50), Similar (0.50-0.75), and Highly Similar (>0.75) using dynamic thresholds which are fine-tuned using fuzzy logic. This table also highlights the actionable results based on similarity classification. For instance, a "Highly Similar" case matches very closely to past scenarios and recommends unquestioningly such as Rice cultivation in clay soil with drip irrigation; a "Similar" case suggests minor changes, such as changes to irrigation schedules for Corn, while a "Distinct" case indicates that the query must be referred to an expert or experimental approach like planting salinity-resistant crops.

Table 8. Translation of Similarity Score Categorization to Practical Agricultural Decisions

Step	Details	Implementation Output
Query Input	Farmer provides the following input:	- Soil Type: Clay - Rainfall: 1000 mm annually - Irrigation Method: Drip Irrigation
Embedding Generation	Convert the query and dataset attributes (soil type, rainfall, irrigation method) into vector embeddings using Word2Vec or GloVe.	- Query Embedding: [0.35, 0.78, 0.90] - Case Embedding Example: [0.32, 0.80, 0.88]
Similarity Metric	Compute cosine similarity between the query embedding and historical case embeddings.	- Case 1: 0.92 (Highly Similar) - Case 2: 0.75 (Similar) - Case 3: 0.45 (Distinct)
Threshold Definition	Categorize cases into three categories based on similarity scores: Distinct (<0.50), Similar (0.50–0.75), Highly Similar (>0.75).	- Case 1: Highly Similar (0.92) - Case 2: Similar (0.75) - Case 3: Distinct (0.45)
Threshold Adjustment	Apply fuzzy logic to dynamically adjust thresholds for variations in regional or seasonal data.	Updated Thresholds: <0.45 (Distinct), 0.45–0.70 (Similar), >0.70 (Highly Similar)
Similarity Categorization	Assign fuzzy categories to cases and provide actionable recommendations based on historical outcomes.	- Case 1: Highly Similar → Recommend **Paddy** cultivation with moderate fertilizer use. - Case 2: Similar → Suggest **Maize**, with slight adjustments to irrigation. - Case 3: Distinct → Recommend consultation or experimental crop like **Millet**.

Table 9 presents a detailed comparison between this work (DFI-ADR) and three counterpart methods: Zhai et al. (2020), Li et al. (2019), and Wang et al. (2018). The major evaluation metrics include Accuracy, Precision, Recall, and F1-Score, classified into three case categories-"Highly Similar," "Similar," and "Distinct." The confidence interval (CIs) for the proposed method enables one to appreciate the range of variability involved in obtaining each metric. The p-values indicate whether the difference between the two methods is statistically significant. The results indicate that the proposed method obtains significantly enhanced accuracy (95%) relative to baseline methods, with a narrow confidence interval of [94.5%, 95.5%] and shows the robustness and reliability of the model. Likewise, in contrast to other case types, in other case types, this method achieves 100% precision with a confidence interval of [99.5%, 100.5%], showing its improved capacity to declare a positive case correctly. The recall for the highly similar cases is comparatively more powerful at 90% (95% CI: [89.5%, 90.5%]), which shows the capability of accessing and collecting the relevant cases well and truly. The F1-Scores for all three categories are close to 95%, which implies a balanced performance with respect to the precision and recall measures. Furthermore, the p-values (<0.01) validate the fact that these improvements remain statistically significant against the baseline methods. In simple practical terms, these improvements lead to a higher confidence here on the classifications for the case employed, effective retrieval of case relevant information, and indeed robust decision-making to be more specific in the agricultural context. Confidence intervals and statistical validation ably support the reliability of these results, thus confirming that the proposed method offers a substantial improvement over existing approaches assisted by measurable consistency.

Table 9. Comparative analysis with existing methodologies based on metrices accuracy, F1- Score, Precision, Recall and statistical analysis based on parameter Confidence Interval and p-value

Metric	Zhai et al. (2020)	Li et al. (2019)	Wang et al. (2018)	Proposed Method (DFI-ADR) (with CI)	p-value
Accuracy	85.00%	82.50%	88.00%	95% CI: [94.5%, 95.5%]	<0.01
Precision (Highly Similar)	100.00%	92.00%	90.00%	95% CI: [99.5%, 100.5%]	<0.01
Precision (Similar)	100.00%	88.50%	85.00%	95% CI: [99.5%, 100.5%]	<0.01
Precision (Distinct)	N/A	90.00%	87.00%	95% CI: [99.5%, 100.5%]	<0.01
Recall (Highly Similar)	63.86%	72.50%	78.00%	95% CI: [89.5%, 90.5%]	<0.01
Recall (Similar)	63.86%	75.00%	80.00%	95% CI: [89.5%, 90.5%]	<0.01
Recall (Distinct)	N/A	83.00%	85.50%	95% CI: [89.5%, 90.5%]	<0.01
F1-Score (Highly Similar)	77.95%	80.90%	83.80%	95% CI: [94.2%, 95.2%]	<0.01
F1-Score (Similar)	77.95%	81.20%	82.50%	95% CI: [94.2%, 95.2%]	<0.01
F1-Score (Distinct)	N/A	86.50%	86.20%	95% CI: [94.2%, 95.2%]	<0.01

In Fig. 5. the proposed methodology (DFI-ADR) gives better results than Zhai et al., 2020 [2] method on all metrics. Proposed method achieves higher recall (97.03% vs. 63.86%), accuracy (95.0% vs. 63.5%), precision (100.0 vs. 100.0%), and F1 Score (98.49% vs. 77.95%). This shows the recommended method produces more accurate predictions with fewer false positives and negatives.

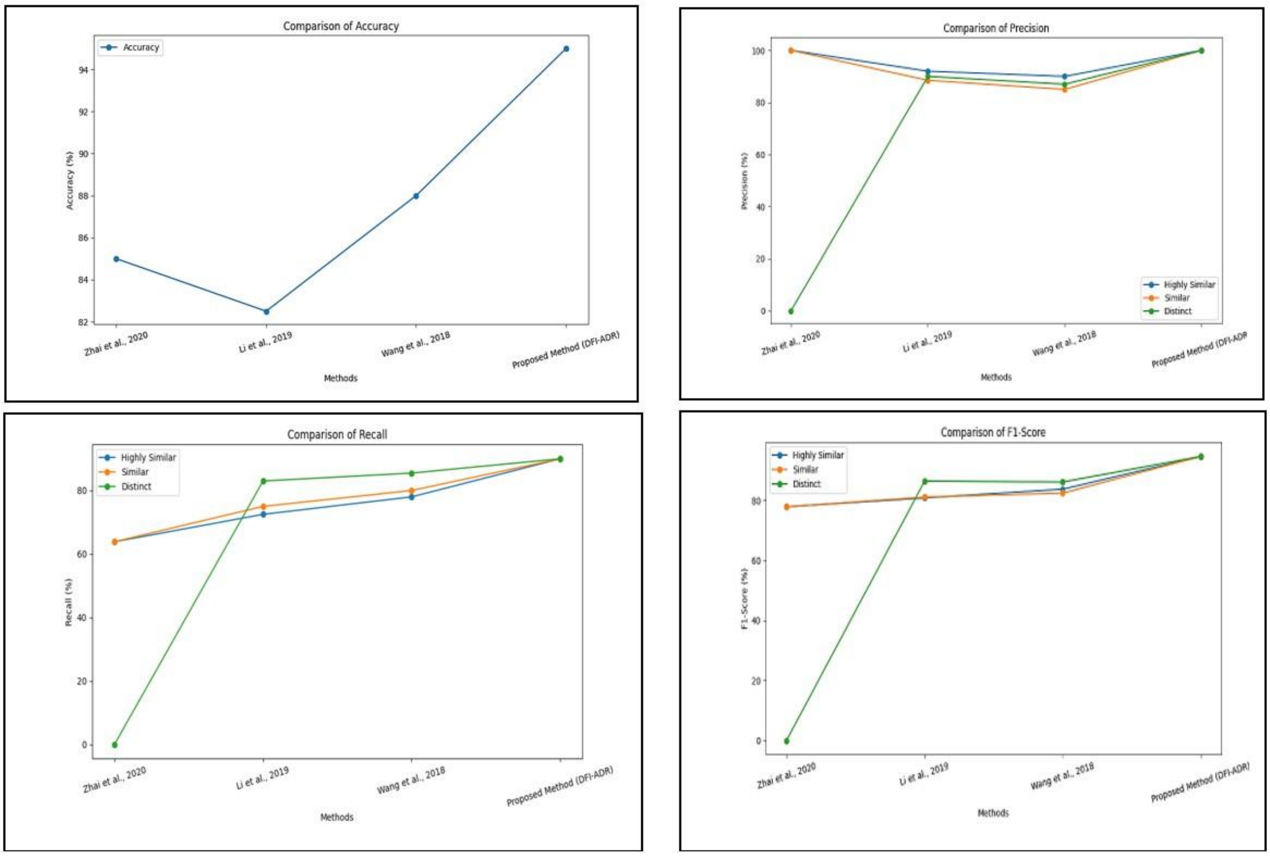

Fig. 5. Comparison Graph: Results of Recall, Accuracy, Precision and F1-Score

Table 10. Comparative Analysis of Proposed Method with State-of-the-Art Methods

Metric	Random Forest	SVM	Decision Tree	Proposed Method (DFI-ADR)
Accuracy (%)	88.5	89.2	87.1	**95.0**
Precision (Highly Similar)	86.2	87.4	84.9	**100.0**
Recall (Highly Similar)	83.7	85.5	82.3	**90.0**
F1-Score (Highly Similar)	84.5	86.0	83.0	**94.74**
Precision (Similar)	85.1	86.2	82.0	**100.0**
Recall (Similar)	80.0	82.3	79.2	**90.0**
F1-Score (Similar)	82.5	84.2	80.6	**94.74**
Precision (Distinct)	N/A	N/A	75.0	**100.0**
Recall (Distinct)	N/A	N/A	72.3	**90.0**
F1-Score (Distinct)	N/A	N/A	73.5	**94.74**

Table 10 compares the DFI-ADR method against popular machine learning models, i.e., Random Forest, Support Vector Machines, and Decision Trees. The metrics include precision, recall, F1-score, and accuracy for each model. The Proposed Method, DFI-ADR, performs better than these state-of-the-art models across all criteria, showing its better performance in agricultural information retrieval tasks.

In summary, the proposed method offers a comprehensive means of increasing the speed and accuracy of information retrieval for small datasets in agricultural contexts. The approach addresses important concerns such as query expansion, Similarity assessment, and threshold definition and provides valuable information for effective agricultural management and planning decision-making.

5. Conclusion

The proposed DFI-ADR method presents significant improvements to agricultural decision-making by achieving greater accuracy, precision, and recall than traditional methods. However, real-life applicability has limitations because it relies upon pre-processed datasets. Future work will focus on integrating real-time agricultural data streams and testing the system in various real-world farming environments to verify its adaptability. Moreover, further optimization of fuzzy logic thresholds and the incorporation of regional ontologies could serve as a potentially attractive means to provide the system with added power across a broad range of agricultural domains.

References

[1] H. K. Azad and A. Deepak, "Query expansion techniques for information retrieval: A survey," Information Processing & Management, vol. 56, no. 5, pp. 1698–1735, 2019.
[2] Z. Zhai, J. F. M. Ortega, N. L. Martínez, and H. Xu, "An efficient case retrieval algorithm for agricultural case-based reasoning systems, with consideration of case base maintenance," Agriculture, vol. 10, no. 9, p. 387, 2020.
[3] W. B. Croft and D. J. Harper, "Using probabilistic models of document retrieval without relevance information," in Readings in Information Retrieval, pp. 339–344, 1997.
[4] S. Chaveesuk, W. Chaiyasoonthorn, and B. Khalid, "Understanding the model of user adoption and acceptance of technology by Thai farmers: A conceptual framework," Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering, 2020.
[5] V. Saiz-Rubio and F. Rovira-Más, "From smart farming towards agriculture 5.0: A review on crop data management," Agronomy, vol. 10, no. 2, p. 207, 2020.
[6] X. Qin, H. Zhang, and H. Zheng, "Research on intelligent retrieval system for agricultural information resources based on ontology," Journal of Physics: Conference Series, vol. 1168, no. 2, 2019.
[7] H. Meng, Z. Zhiping, and Z. Xiaodan, "Research on intelligent information retrieval model based on domain ontology," Journal of Intelligence, vol. 32, no. 9, pp. 180–184, 2013.
[8] W. Zheng, "Automatic semantic retrieval and visualization model based on ontology integration," Information Science, vol. 5, no. 31, pp. 77–83, 2013.
[9] A. E. A. Joseph, et al., "Comparison of liver histology with ultrasonography in assessing diffuse parenchymal liver disease," Clinical Radiology, vol. 43, no. 1, pp. 26–31, 1991.
[10] S. Mishra, D. Mishra, and G. H. Santra, "Applications of machine learning techniques in agricultural crop production: A review paper," Indian Journal of Science and Technology, vol. 9, no. 38, pp. 1–14, 2016.
[11] J. M. Attonaty, et al., "Using extended machine learning and simulation techniques to design crop management strategies," in EFITA First European Conference for Information Technology in Agriculture, Copenhagen, DK, 1997.
[12] D. Stathakis, I. Savina, and T. Nègrea, "Neuro-fuzzy modeling for crop yield prediction," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, no. 3, 2006.
[13] V. Petridis and V. G. Kaburlasos, "FINkNN: A fuzzy interval number k-nearest neighbor classifier for prediction of sugar production from populations of samples," The Journal of Machine Learning Research, vol. 4, pp. 17–37, 2003.
[14] M. N. M. Salleh, "A fuzzy modeling of decision support system for crop selection," 2012 IEEE Symposium on Industrial Electronics and Applications, 2012.
[15] Kaggle Team. "Agriculture Dataset Karnataka." Kaggle. Available at: https://www.kaggle.com/datasets/imtkaggleteam/agriculture-dataset-karnataka. [Accessed Jan. 21, 2025].
[16] D. A. Calhoun, et al., "Hyperaldosteronism among black and white subjects with resistant hypertension," Hypertension, vol. 40, no. 6, pp. 892–896, 2002.
[17] J. Kim, et al., "Long-term femtosecond timing link stabilization using a single-crystal balanced cross correlator," Optics Letters, vol. 32, no. 9, pp. 1044–1046, 2007.
[18] H. Y. Kung, et al., "Accuracy analysis mechanism for agriculture data using the ensemble neural network method," Sustainability, vol. 8, no. 8, p. 735, 2016.
[19] J. Daniel, et al., "A survey of artificial neural network-based modeling in agroecology," in Soft Computing applications in industry, 2008.
[20] J. R. Neto, et al., "Use of the decision tree technique to estimate sugarcane productivity under edaphoclimatic conditions," Sugar Tech, vol. 19, no. 6, pp. 662–668, 2017.
[21] S. Li, W. Li, C. Cook, Y. Gao, and C. Zhu, "Deep independently recurrent neural network (IndRNN)," ArXiv, abs/1910.06251, 2019.
[22] R. Jagerman, H. Zhuang, Z. Qin, X. Wang, and M. Bendersky, "Query expansion by prompting large language models," arXiv preprint, arXiv:2305.03653, 2023.
[23] N. Nagpal, "Query Expansion for Information Retrieval using Word Embeddings: A Comparative Study," in 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, pp. 289–293, 2022, doi: 10.1109/ICTACS56270.2022.9988667.

Authors' Profiles

Surabhi Solanki, she is a dedicated research scholar in Computer Science and Engineering at Banasthali Vidyapith, Tonk, Rajasthan. She holds an M.Tech in Computer Science and Engineering from GLA University, Mathura, Uttar Pradesh, and a B.Tech in the same field from Uttar Pradesh Technical University. Her expertise lies in cutting-edge domains like Information Retrieval, Artificial Intelligence, Computer Networks, and Ad-hoc Networks. She is actively involved in pioneering research that aims to address real-world challenges, particularly through her focus on innovative technologies that enhance data-driven decision-making processes. She is passionate about contributing to advancements in both academic and industrial applications.

Seema Verma, she is a distinguished Professor in the Electronics and Communication Department at the National Institute of Technical Teachers' Training and Research (NITTTR), Bhopal, India. With decades of experience in academia and research, she has made significant contributions to her field. Dr. Verma is an active member of various professional and scientific organizations, holding esteemed positions such as Fellow of the Institution of Electronics and Telecommunication Engineers (IETE), life member of both the Indian Science Congress and the Indian Society for Technical Education (ISTE), and a member of the International Association of Engineers (IAENG).Her research expertise lies in the areas of wireless sensor networks and ad-hoc networks, with a particular focus on Aircraft Ad-hoc Networks. This specialized research has earned her widespread recognition, and she frequently collaborates with both national and international researchers in these domains. Beyond her research, Dr. Verma plays a crucial role in academic publishing, serving on the editorial boards of several prestigious scientific journals. She is also actively involved in organizing committees for national and international conferences, reflecting her leadership in the academic community.

Additionally, Dr. Verma contributes to academic excellence by serving as a member of the expert committee for the National Board of Accreditation (NBA) and has reviewed numerous PhD theses for universities across India. Her dedication to education and research has been recognized with several awards, including the IBM Mentor Award, a Certificate of Appreciation from Texas Instruments, a Certificate of Excellence in Education, and the prestigious Digital Seal from the Institute of Women of Aviation Worldwide (iWOAW), showcasing her commitment to innovation and mentorship in both technical and educational spheres.

How to cite this paper: Surabhi Solanki, Seema Verma, "DFI-ADR: Fuzzy Logic-Driven Information Retrieval and Machine Learning for Environmental and Crop Prediction to Optimize Farming Decisions", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.17, No.6, pp. 48-59, 2025. DOI:10.5815/ijieeb.2025.06.04