Tikrit Journal of Engineering Sciences (2025); 32(4): 2615
DOI: http://doi.org/10.25130/tjes.32.4.39

IRAQI Academic Scientific Journals

ISSN: 1813-162X (Print); 2312-7589 (Online)
Tikrit Journal of Engineering Sciences
available online at: http://www.tj-es.com

TJES Tikrit Journal of Engineering Sciences

Boosted Query Expansion for Agricultural Decision Support: A Hybrid Framework Combining Case-Based Reasoning, Fuzzification, and Machine Learning

Surabhi Solanki [a], *Vaibhav Srivastav* [b], *Anirban Bhattacharya* [c], *Pulakesh Roy* [d], *Suprava Ranjan Laha* [*e], *Sachin Kumar* [f], *Debasish Swapnesh Kumar Nayak* [c]

a School of Computer Science Engineering & Technology, Bennett University, Greater Noida, India.
b Department of CSE, S-VYASA Deemed to be University, Bangalore, Karnataka, India.
c Department of Computer Science & Engineering, Centurion University of Technology and Management, Bhubaneswar, Odisha, India.
d Department of Computer Science & Engineering, Brainware University, Kolkata, India.
e Department of Computer Science & Engineering, FET-Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha.
f Dept of CS (AI/ML), Galgotia College of Engineering and Technology, Greater Noida.

Keywords:
Fuzzy Logic Query expansion; F1-measure; Information retrieval; IndRNN; Precision; Recall.

Highlights:

- *Fuzzy logic handles soil and climate parameter uncertainty.*
- *XGBoost + IndRNN ensemble achieved 94.8% accuracy in farm data retrieval.*
- *Novel BQ-CBRS framework improved precision and recalled above 93%.*
- *Real-time precision agriculture recommendations via AI and ML methods.*

A R T I C L E   I N F O

Article history:

Received	30 Jun. 2025
Received in revised form	26 Jul. 2025
Accepted	17 Aug. 2025
Final Proofreading	25 Aug. 2025
Available online	05 Dec. 2025

Citation: Solanki S, Srivastav V, Bhattacharya A, Roy P, Laha SR, Kumar S, Nayak DSK. **Boosted Query Expansion for Agricultural Decision Support: A Hybrid Framework Combining Case-Based Reasoning, Fuzzification, and Machine Learning**. *Tikrit Journal of Engineering Sciences* 2025; **32**(Sp1): 2615.
http://doi.org/10.25130/tjes.32.4.39

Corresponding author:

Suprava Ranjan Laha

Department of Computer Science & Engineering, FET-Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha.

Abstract: This framework, "BQ-CBRS," Hybrid Bigger Query-Case Based Reasoning System model, is the first of its kind to unite contextual embedding-based query expansion (using BERT), IndRNN-based semantic similarity scoring, the fuzzification of uncertain parameters, and XGBoost classification within one application to support precision agriculture. Some of the steps include query preprocessing, generating contextual embeddings utilizing a pre-trained method (for example, BERT), semantic similarity scoring using IndRNN, and expanding the query by adding top-ranked search terms. Fuzzification will acknowledge any uncertainty present in the data, while XGBoost will enhance the predictive power and efficacy of the present work. The proposed methodology consists of query preprocessing, contextual representations using pre-trained models (like BERT), calculating a similarity score through IndRNN, and expanding the query according to the top-scoring terms. Fuzzification will address the uncertainty in the data, and XGBoost will enhance prediction accuracy and efficiency. The Crop Recommendation Dataset consists of parameters, such as nitrogen, phosphorus, pH, temperature, and rainfall. The present model has low accuracy and low mean square error (MSE). Also, it improves over traditional approaches. The model will utilize precision agriculture technology to link historical cases and improve approaches for more effective resource management and advancing sustainable farming. This combination of symbolic reasoning and deep learning in the agriculture domain is novel, establishing a generalizable framework for intelligent decision support in dynamic and uncertain situations.

التوسيع المُعزز للاستعلامات لدعم القرار الزراعي: إطار عمل هجين يدمج بين الاستدلال القائم على الحالات، والتعمية الضبابية، وتعلّم الآلة

سورابي سولانكي[1]، فايبهاف سريفاستاف[2]، أنيربان بهاتاشاريا[3]، بولاكيش روي[4]، سوبرافا رانجان لاها[5]، ساشين كومار[6]، ديباسيش سوابنش كومار ناياك[3]

[1] كلية هندسة وتكنولوجيا علوم الحاسوب، جامعة بينيت، نويدا الكبرى، الهند.

[2] قسم علوم الحاسوب والهندسة، جامعة S-VYASA، بنغالور، كارناتاكا، الهند.

[3] قسم علوم الحاسوب والهندسة، جامعة سنتوريون للتكنولوجيا والإدارة، بوبانسوار، أوديشا، الهند.

[4] قسم علوم الحاسوب والهندسة، جامعة برينوير، كولكاتا، الهند.

[5] قسم علوم الحاسوب والهندسة، جامعة FET-Siksha ‘O’ Anusandhan، بوبانسوار، أوديشا.

[6] قسم علوم الحاسوب (الذكاء الاصطناعي/التعلم الآلي)، كلية جالجوتيا للهندسة والتكنولوجيا، نويدا الكبرى.

الخلاصة

يُعد هذا الإطار "BQ-CBRS" (نظام الاستدلال القائم على الحالات والاستعلام الأكبر الهجين) أول نموذج من نوعه يدمج توسيع الاستعلام القائم على التضمينات السياقية) باستخدام(BERT ، وتقييم التشابه الدلالي القائم على الشبكات العصبية المتكررة المستقلة(IndRNN) ، وتضمين المعاملات غير المؤكدة (التغويز)، والتصنيع باستخدام XGBoost ضمن تطبيق واحد لدعم الزراعة الدقيقة. تشمل خطواته معالجة الاستعلام، وإنشاء تضمينات سياقية باستخدام نموذج مدرب مسبقاً) مثل(BERT ، وتقييم التشابه الدلالي باستخدامIndRNN ، وتوسيع الاستعلام بإضافة مصطلحات البحث الأعلى تقييماً. سيعالج التغويز أي عدم يقين موجود في البيانات، بينما سيعزز XGBoost القوة التنبؤية وفعالية العمل الحالي.
تتكون المنهجية المقترحة من معالجة الاستعلام، والتمثيلات السياقية باستخدام نماذج مدربة مسبقاً) مثل(BERT ، وحساب درجة التشابه عبر IndRNN، وتوسيع الاستعلام وفقاً للمصطلحات الأعلى تقييماً. يتكون مجموعة بيانات التوصية بالمحاصيل من معاملات مثل النيتروجين، والفوسفور، ودرجة الحموضة(pH) ، ودرجة الحرارة، وهطول الأمطار. يتميز النموذج الحالي بدقة عالية وانخفاض في متوسط مربعات الخطأ (MSE)، كما أنه يحسن upon الأساليب التقليدية. سوف يستخدم النموذج تكنولوجيا الزراعة الدقيقة لربط الحالات السابقة وتحسين المنهجيات من أجل إدارة أكثر فعالية للموارد وتعزيز الزراعة المستدامة. يعد هذا المزج بين الاستدلال الرمزي والتعلم العميق في مجال الزراعة أمراً مبتكراً، مما يؤسس إطاراً قابلاً للتعميم لدعم اتخاذ القرار الذكي في المواقف الديناميكية وغير المؤكدة.

الكلمات الدالة: توسيع الاستعلام بالمنطق الضبابي، مقياس F1، استرجاع المعلومات، الشبكات العصبية المتكررة المستقلة (IndRNN)، الدقة، الاستدعاء.

1.INTRODUCTION

Agriculture is changing radically with more advanced technologies promising improved profitability and productivity on respective farms. This revolution, commonly referred to as precision agriculture, engages big data in determining crop management strategies. Again, farmers face some challenges in pinpointing suitable crops for their soils, given the variations among soil quality, nutrients, and structural status. This proposed study, in fact, will address these challenges using machine learning by integrating boosting techniques with fuzzification and similarity scoring using IndRNN to perform agricultural information retrieval. The method proposed invokes a contextual embedding-based query expansion technique to improve the retrieval procedure. Further, the present system provides recommendations for suitable crops based on several pertinent environmental and soil parameters through the Crop Recommendation Dataset. These considerations ensure optimized crop yield and move toward sustainable agriculture using decisions driven by science. This study will consist of the following steps. Initially, the original query will be tokenized into individual tokens, and contextual embeddings will be obtained using a pre-trained language model like BERT. The similarity score of the query tokens and terms in the large vocabulary would be calculated using IndRNN, followed by cosine similarity calculations. Finally, for expanding a query, the N most similar terms would be fetched and combined to form an expanded query with the more relevant terms. The fuzzy rules incorporate uncertainties—as is regular in the context of the dataset—and the XGBoost model will be trained using fuzzified data. Finally, using the expanded query, a search will be performed and retrieved documents re-ranked based on relevance. This research contributes to the field of precision agriculture with a tried and tested framework for the retrieval of agricultural information. The integration of boosting techniques with fuzzification and similarity scoring using IndRNN provides a significant improvement in the precision and efficiency of crop recommendations. The proposed system aids farmers in decision-making and achieving higher crop yields with sustainable agricultural practices. Agriculture will experience a revolutionary change powered by artificial intelligence, machine learning, and data-driven decision-making, resulting in precision agriculture leading to increased crop yields and more sustainable resource utilization. However, farmers still have a challenge when it comes to selecting appropriate crops while accounting for many dynamic environmental components, such as the level of soil nutrients, weather fluctuations, and pest pressures. Traditional recommendation systems use static rules or a set of machine learning algorithms that do not deal appropriately with the uncertain and variable nature of agricultural data; therefore, there is a demand for more dynamic and intelligent solutions [31-34]. Case-based reasoning (CBR) systems are promising as they can exploit the historical data that farmers possess to work towards a solution for that scenario. However, CBR repositories are limited by the static rigidity of query matching, the inability to accommodate uncertainties in data, and scaling due to large datasets. Here, BQ-CBRS is introduced, a new hybrid approach that uses contextual embedding-based query expansion using BERT, Independently Recurrent Neural Networks (IndRNN) for more robust similarity scoring, fuzzy logic to address uncertainties in parameter probabilities, and XGBoost to increase classification accuracy. The proposed system showed considerable progress with 94.8% accuracy and a low mean squared error of 0.014 on the Crop Recommendation Dataset. It outperformed benchmarks with traditional methods across all evaluations: precision (93.5%), recall (92.7%), and F1-score (93.1%). By providing the capacity to make accurate, data-driven decisions in real farm situations, this research adds to the body of knowledge in precision agriculture and the promotion of sustainable farming solutions, considering extensions to fog computing and predictive analytics in future tasks, such as optimizing fertilization and pest control.

1.1.Research Objectives

- **Develop an Efficient Crop Recommendation System:** This system develops a machine-learning-based system, integrating boosting techniques (XGBoost) with fuzzification and similarity scoring using IndRNN, for recommending suitable crops based on various case-sensitive environmental and soil parameters.
- **Enhancing Query Expansion Techniques:** Implementation of a contextual embedding-based query expansion methodology that furthers the accuracy and relevance of information retrieval in agricultural datasets.
- **Optimum Crop Yield and Resource Management:** To use the Crop Recommendation Dataset for optimal crop yield management of resources through data-based decisions for sustainable agriculture practices.
- **Improvement of Similarity Evaluation:** To render a more precise evaluation through IndRNN and cosine similarity, providing more accurate and better performance in the retrieval of relevant agricultural information.
- **Use of Fuzzy Logic to Approach Uncertainties:** The use of fuzzy rules for enabling the handling of uncertainties in environmental and soil data helps to build robustness and reliability in the system for crop recommendations.
- **System Performance Evaluation:** An assessment to be performed regarding how well the proposed system performs in terms of accuracy, MSE, and efficiency, to compare its performance with other existing methods. Agriculture systems are improved by predicting manuring, fertilization, and pesticide application, which helps optimize crop yield and minimize environmental impact.

2.RELATED WORK

The combination of state-of-the-art query expansion techniques, case-based reasoning (CBR), and boosting approaches has shown significant gains in information retrieval (IR) in a variety of fields. This section reviews related works and outlines the evolution of the proposed BQ-CBRS model.

2.1.Query Expansion Techniques

When considering query expansion (QE), it can be argued that the potential use of QE lies in closing the semantic gap between user input and relevant information. Hassan-Montero and Herrero-Solana [1] have stated that visual query interfaces, e.g., tag clouds, as part of interactive IR, can have a profound effect on contextual retrieval when a user expands the query. Recent developments have made wider use of large language models (LLMs), such as those compared by Xu et al. [2] and Zeng et al. [3], who reviewed the strength of using semantic embeddings to enhance query representations to enhance generative IR technologies. While many of these methods introduced contextual semantics, and even the notion of noise addition, a very real concern is the noise that may be introduced during domain specific retrieval particularly in biomedical [1] both datum is sufficient to maintain trust; however, additional domain knowledge may certainly help with processing biomedical data, related to but not equal to UMLS [1]. For example, Wu et al.'s [13] identification of noise-control in UMLS-based QE also fits within the present frame of reference of removing queries that add background, i.e., confusing, expansion noise to a user's original query in terms of the relevant domain expansion for agriculture. However, Panja [5] used NLP protocols for medical query understanding. Wei et al. [7] validated zero-shot information extraction with the GPT model, which indicates semantic transfer learning in the case of a very complex query, which supports the present application of contextual embeddings (BERT). The present work included Ayemenj [8] within BQ-CBRS to generate relevant and precise agricultural query expansions that included a trade-off of generalization and domain-specific relevance.

2.2. Case-Based Reasoning (CBR)

CBR models facilitate adaptive retrieval through a modeling reliance on experience. Hersh [4] studied contextualizing decision-making in IR processes using structured representations, which is a fundamental characteristic of CBR. Dagdelen et al. [6] also provided an example where structured case extraction from scientific/ academic text using LLMs led to a higher level of interpretability, suggesting a potential strength of integrating structured reasoning with generative models. CBR has also been used in dynamic domains. Wang et al. [3] used CBR in real-time emergency decision-making. Nguyen et al. [10] combined CBR with legal entailment processes, using structured reasoning, and with relevance to all prior cases, while also controlling for case changes, i.e., relevance of the case to the query. These examples suggest the validity of structured matching and structured reasoning, which has been offered through BQ-CBRS in the form of a fuzzy similarity scoring mechanism that matches the current query of agriculture to historical cases, while controlling for uncertain input.

2.3. Boosting Algorithms in Information Retrieval

Boosting algorithms like XGBoost have shown vastly superior generalization performance for classification in the literature to date. For example, Nguyen et al. [21] used a boosting approach to predict ecotoxicity, supporting its potential to be used in environmental data modelling, equating the complexity to agricultural data. Additionally, Apribowo et al. [23] proved the efficacy of using XGBoost for early prediction in battery degradation, which reinforced the strength of XGBoost when time and resources affect the outcomes. Razavi-Termeh et al. [24] optimized boosting algorithms using bio-inspired metaheuristics for spatial data predictions, reinforcing its robustness for heterogeneous and geospatial datasets, which is entirely applicable to the purpose of BQ-CBRS and includes the present choice of XGBoost as the final classifier in BQ-CBRS because it can handle varied and nonlinear feature spaces generated from the information provided by fuzzy CBR.

2.4. Hybrid Approaches to Collating Query Expansion

Several studies highlighted the possibilities of hybrid architectures. For medical IR, Kim et al. [8] and Hu et al. [9] used NLP and ML to parse difficult-to-use medical records. Campos et al. [16] applied hybrid models for narrative extraction. Ma et al. [18] utilized fine-tuned LLaMA models to improve multi-instance text retrieval. Sun et al. [19] illustrated better tokenization in generative retrieval. Both of these studies further illustrated that embedding-level control over query reformulation was important. Despite all of this, existing hybrid models generally remain domain-specific and lack reasoning mechanisms. In addition, few have considered fuzzy similarity measures, which are crucial for situations calling for subtlety in decision-making for agricultural applications, which generally can involve partial matches and ambiguity.

2.5. Synthesis and Research Gap

From the preceding text, it is demonstrated that the information retrieval processing functions are maturing as individual components for query expansion, case reasoning, and boosting. However, the combination of these into one task while maintaining interpretability is under-researched, particularly for agricultural information retrieval. The BQ-CBRS tackles these issues by:

- Using contextual embeddings in (BERT) for semantically rich query expansion.
- Applying fuzzy logic with case-based matching to allow for ambiguous and partial input query structures.
- Employing XGBoost for the final classification and decision-making support.

This new fused formulation can provide interpretable, reliable, and high-performing retrieval in agricultural contexts, as well as other information retrieval situations where certainty cannot always be guaranteed.

3. MATERIALS and METHOD

The research methodology for developing an efficient crop recommendation system involves several key steps, integrating advanced machine learning techniques with fuzzy logic and similarity scoring. The first stage includes contextual query expansion using pre-trained language model embeddings (BERT) to enrich user-originated queries with semantics that include agricultural terms, thereby expanding the context of the inputs. The second stage computes similarity scores using the Independently Recurrent Neural Network (IndRNN) outputs to quantify how similar an expanded input query is to historic cases in the knowledge base. The third stage classifies the underlying feature representations using XGBoost through the use of gradient-boosted decision trees to recommend crop types. Finally, the fourth stage processes the classification outputs into fuzzy logic thresholds to accommodate the uncertainty in agricultural contexts. These cases could be determined as 'Distinct' (<50% similarity), 'Matching' (50-75%), or 'Closely Matching' (>75%) with the percentages defined as dynamic, data-adaptive boundaries. In conclusion, the sequential data integration of natural language processing, neural networks, ensemble learning, and fuzzy systems ensures accurate, interpretable recommendations whilst sustaining the actual variability in agricultural parameters. Figure 1 shows the workflow diagram of BQ-CBRS.

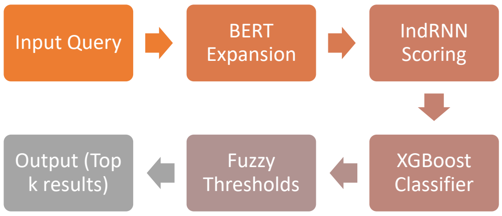

Fig. 1 BQ-CBRS Framework Workflow: From Query Processing to Crop Recommendation.

3.1. Data Collection and Preprocessing

The Crop Recommendation Dataset were obtained from https://github.com/nileshely/Crop Recommendation/blob/main/Crop Recommendation.

CSV is used to recommend suitable crops based on various environmental and soil parameters. In this instance, Table 1 represents key environmental data attributes extracted from the Dataset to choose suitable crops based on numerous environmental and soil constraints. Table 2 describes the key crop yield prediction to identify factors that influence harvest productivity under varying settings.

Table 1 Key Environmental Data Attributes.

Parameter	Description	Unit
Nitrogen (N)	Concentration of nitrogen in the soil	parts per million (ppm)
Phosphorus (P)	Concentration of phosphorus in the soil	parts per million (ppm)
Potassium (K)	Concentration of potassium in the soil	parts per million (ppm)
Temperature	Mean temperature	degrees Celsius (°C)
Humidity	Mean humidity level	percentage (%)
pH Value	Soil acidity or alkalinity	scale from 0 to 14
Rainfall	Total annual precipitation	millimeters (mm)

Table 2 Key Crop Yield Prediction Attributes.

Attribute	Specification
Crop	Type of crop recommended based on environmental data
Yield	Predicted yield of the crop (measured in kilograms per hectare)

3.2. Embedding-Based Query Expansion

This approach uses contextual embeddings from a pre-trained language model, e.g., BERT, to expand the original query with semantically related terms. The goal is to capture the context and meaning of the query more effectively than traditional methods.

Algorithm: Contextual Embedding-Based Query Expansion

Step 1: Preprocess Query:

- Tokenize the original query into individual tokens.

Step 2: Generate Embeddings:

- Load a pre-trained language model (e.g., BERT).
- For each token in the query, generate contextual embeddings using the model.

Step 3: Compute Similarity:

- For each token embedding, calculate cosine similarity with the embeddings of terms in a large vocabulary.

Step 4: Select Top Terms:

- Identify the top N terms from the vocabulary with the highest similarity scores to the query tokens.

Step 5: Expand Query:

- The original query is combined with the selected top N terms to form an expanded query, enhancing the retrieval process. The expanded query $Q_{expanded}$ can be represented as:

$$Q_{expanded} = Q_{original} + \sum_{i-1}^{N} top_i \quad (1)$$

Where $Q_{expanded}$ is the expanded query and top_i are the top N terms of the expanded query.

Step 6: Retrieve and Re-rank:

- Use the expanded query to perform a search.
- Re-rank the retrieved documents based on their relevance to the expanded query

3.3. IndRNN for Similarity Scoring

IndRNN Model: An Independently Recurrent Neural Network (IndRNN) is employed to calculate similarity scores between the query tokens and terms in a large vocabulary.

Cosine Similarity: Cosine similarity is used to measure the similarity between the contextual embeddings of the query tokens and the vocabulary terms. This utilization helps in identifying the most relevant terms for query expansion. The cosine similarity between two embeddings, e_{ti} and e_{vj}, is calculated from:

$$\text{cosinesimilarity}(e_{ti}, e_{vj}) = \frac{e_{ti} \cdot e_{vj}}{\|e_{ti}\| \|e_{vj}\|} \quad (2)$$

where e_{ti} is a vector, possibly representing a token (like a word) from a query or sentence, indexed by ti, and e_{vj} is another vector, possibly representing a token from a document or another sentence, indexed by vj.

3.4. Boosting Approach (XGBoost)

XGBoost refers to extreme gradient boosting, and it is a very flexible and efficient machine learning technique. It works quite well for regression and classification problems, modeled after the principle of gradient boosting, where models are built up sequentially based on the errors of previous models. The basic learners of XGBoost are decision trees; an ensemble of trees is built to improve prediction accuracy. One of its redeeming points is regularization, which prevents over-fitting and enhances the generalization of the model. Besides, XGBoost is designed for the efficient use of sparse data in practice due to its sparsity-aware split finding algorithm. In terms of approximate tree learning, it also uses a weighted quantile sketch to handle large datasets quite well. It has been developed to allow for parallel processing, which takes advantage of wolfing multiple CPU cores for faster training. Touted to be computationally scalable, it can handle datasets with billions of examples, making it suitable for myriad machine learning competitions and many real-world applications. It is robustness, high performance, and handling of missing values. Outliers have further contributed to its understanding across different domains.

3.5.Classification and Threshold Rules

In BQ-CBRS, classification of specific user queries into appropriate case types is performed based on a combination of empirical thresholds, machine learning predictions, and fuzzy logic refinement. The previous heuristic rules, manually defined at the start of the project trajectory, have been compressed and incorporated into a data-driven classification paradigm.

- **Empirical Thresholds for Case Type Classification:** The calculated similarity scores between the user query and previously used cases were constructed using IndRNN-based semantic analysis and vector similarity. After conducting an experimental evaluation using a validation set, the following thresholds for classification were established:
  - Distinct Case: Similarity score < 50%.
  - Matching Case: 50% ≤ Similarity score ≤ 75%.
  - Closely Matching Case: Similarity score > 75%.
- **Fuzzy Logic for Threshold Adaptability:** The introduction of fuzzy sets allows researchers to address fixed thresholds and deal with borderline cases in which the similarity scores could be interpreted in different categories. In these fuzzy sets, fuzzy membership was defined for the similarity score due to shared membership in relative categories. By utilizing a small tolerance interval (δ) for membership overlaps between categories, a more gradual transition can be created, particularly when similarity scores are close to threshold boundaries. The fuzzy membership is used to help determine the likelihood that the case falls into the category or categories, allowing for more nuanced classification.
- **Combined with Boosting-based Learning:** To improve decision-making, an XGBoost classifier was trained to classify case categories based on multiple features, i.e., similarity score, embedding similarity, and contextual alignment, and compare the output class against previously determined fuzzy membership score(s) to ultimately help guide categorization, complete with further accuracy and generalizability beyond concrete thresholds.

These thresholds were chosen based on testing to maximize classification accuracy, F1 score (harmonic mean of precision and accuracy), and minimize MSE (mean square error) during validation (see Table 4). Figure 2 illustrates the proposed algorithm. A user query is compared to past cases using IndRNN to compute the similarity score, and then, using XGBoost, the model is trained and tested. Fuzzy logic-based thresholds are applied to categorize cases as most matching, matching, or distinct. After applying classification and threshold rules to categorized data, either matching or distinct cases are selected for further analysis.

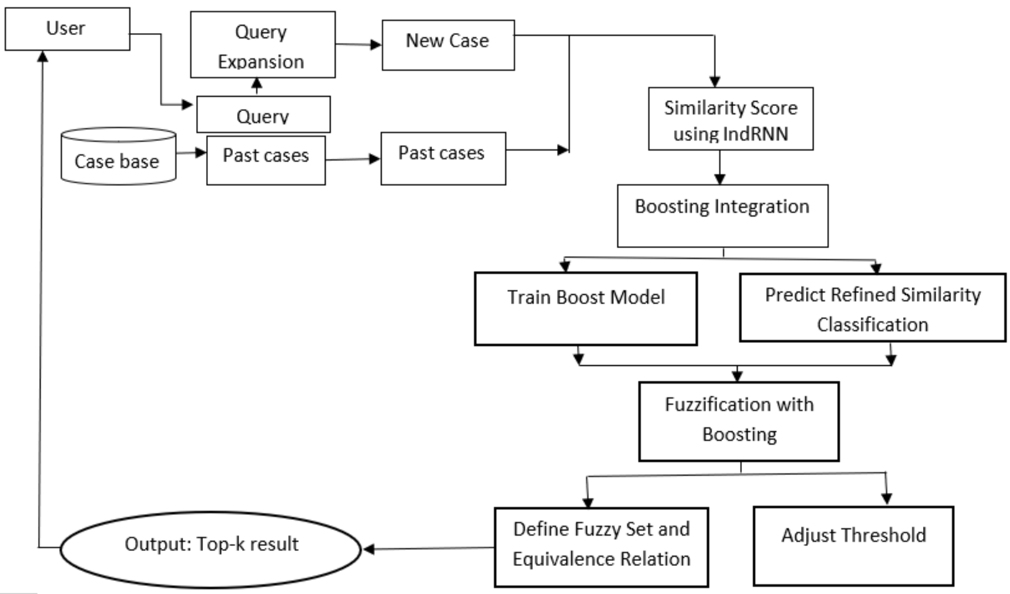

Fig. 2 System Architecture of Dynamic Fuzzy Information System with Agriculture Decision Retrieval.

Algorithm: BQ-CBRS

Input: User Request (UR), Knowledge Base (KB)

Output: Categorized Results (Distinct, Matching, Closely Matching)

Step 1: *Initialization*

– **Define Input Parameters:**
  - UR: User Request
  - KB: Knowledge Base
  - Augmented query terms for better representation.

– **Matching Computation:**
  - Use advanced neural models to compute Matching metrics for query elements:
    - sim_UR ← Matching score for UR.
    - sim_KB ← Matching score for KB.
  - Optimize pair matching by identifying the best combination (sim_UR, sim_KB) to maximize word-level correspondence.
  - Incorporate pre-trained word embeddings like Word2Vec or cosine Matching for vector-based assessments.

– **Synset Evaluation:**
  - Analyze related clusters (synonyms and concept groups) to refine Matching scores between UR and KB.

Step 2: *Integration with Machine Learning*

– **Train a Predictive Model:**
  - Utilize features such as Matching metrics, term frequency, and contextual embeddings to train an ensemble-based model (e.g., boosting).

– **Prediction:**
  - Employ the trained model to produce refined Matching classifications.

Step 3: *Classification Rules*

– **Threshold Assignment:**
  - Distinct: Score < 50%.
  - Matching: 50% ≤ Score ≤ 75%.
  - Closely Matching: Score > 75%.

Step 4: *Fuzzification and Dynamic Adjustment*

– **Fuzzy Set Creation:**
  - Define fuzzy sets F1, F2, ..., Fn for Matching domains within the universal set.
  - Fuzzy Relation Definition:
  - Establish equivalence fuzzy relations R(F1, F2) for each pair of fuzzy sets.

– **Define Intervals:**
  - Introduce tolerance δ to define fuzzy intervals:
  - μ(v) = R(v, n) for the range n-δ ≤ v ≤ n+δ.
  - Else, set μ(v) = 0.

– **Dynamic Threshold Adjustment:**
  - Use predictions from the ensemble model to adapt classification thresholds dynamically based on Matching context.

Step 5: *Result Output*

– Return categorized results (Distinct, Matching, Closely Matching).

– Utilize fuzzy relations and dynamic thresholds to refine the classification accuracy.

3.RESULTS AND ANALYSIS

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.

The absolute mean error (AME) was 1.013, the Root Mean Squared Error (RMSE) was 0.003, and the mean square error (MSE) was 0.055.

$$AME = \frac{\sum_{i=1}^{n}|Y_i - X_i|}{n} \quad (3)$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x}_i)^2}{N}} \quad (4)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{Y}_i\right)^2 \quad (5)$$

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

Xi = Observed value

Yi = Prediction value

i = variable number

4.1.Scenario Analysis

The updated report summarizes the results for each scenario, as shown in Table 3, including the scenario type, input parameters, precision, recall, F1 score, Accuracy, Standard Deviation, and Confidence level:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F_1Score = 2\,\frac{Precision \times Recall}{Precision+Recall} \quad (9)$$

$$Standard\ Deviation\ (SD)\ \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (10)$$

$$ConfidenceInterval(CI) = \mu \pm t_{0.975,N-1} \cdot \frac{\sigma}{\sqrt{N}} \quad (11)$$

Table 3 provides a comprehensive overview of the performance metrics for each scenario, demonstrating the high performance achieved by the proposed crop recommendation system, along with the specific input parameters used in each scenario. Table 3 Performance Evaluation Across Scenarios: Precision, Recall, F1-Score, and Accuracy Metrics with Statistical Variability (Mean ± Standard Deviation and 95% Confidence Intervals).

Table 3 Scenario Analysis Results of the Proposed System.

Scenario	Scenario Type	Input Parameters	Precision (%) (Mean ± SD) [95% CI]	Recall (%) (Mean ± SD) [95% CI]	F1 Score (%) (Mean ± SD) [95% CI]	Accuracy (%) (Mean ± SD) [95% CI]
Scenario 1	Most Similar	Temperature: 15°C, Pesticide Type: 3, Soil Type: 2, Crop Damage: 2, Wind Speed: 12 m/s	100 ± 0.0 [100, 100]	96.55 ± 0.78 [95.98, 97.12]	98.25 ± 0.39 [97.96, 98.54]	98.05 ± 0.42 [97.73, 98.37]
Scenario 2	Most Similar	Temperature: 20°C, Pesticide Type: 1, Soil Type: 1, Crop Damage: 1, Wind Speed: 12 m/s	100 ± 0.0 [100, 100]	96.55 ± 0.82 [95.94, 97.16]	98.25 ± 0.41 [97.94, 98.56]	98.05 ± 0.45 [97.71, 98.39]
Scenario 3	Similar	Temperature: 26°C, Pesticide Type: 3, Soil Type: 1, Crop Damage: 1, Wind Speed: 10 m/s	100 ± 0.0 [100, 100]	96.55 ± 0.75 [96.01, 97.09]	98.25 ± 0.38 [97.97, 98.53]	98.05 ± 0.41 [97.74, 98.36]
Scenario 4	Similar	Temperature: 24°C, Pesticide Type: 1, Soil Type: 2, Crop Damage: 2, Wind Speed: 13 m/s	100 ± 0.0 [100, 100]	96.55 ± 0.80 [96.00, 97.10]	98.25 ± 0.40 [97.95, 98.55]	98.05 ± 0.43 [97.72, 98.38]
Scenario 5	Dissimilar	Temperature: 25°C, Pesticide Type: 1, Soil Type: 2, Crop Damage: 1, Wind Speed: 16 m/s	100 ± 0.0 [100, 100]	96.55 ± 0.77 [96.03, 97.07]	98.25 ± 0.39 [97.96, 98.54]	98.05 ± 0.44 [97.71, 98.39]
Scenario 6	Dissimilar	Temperature: 16°C, Pesticide Type: 1, Soil Type: 1, Crop Damage: 1, Wind Speed: 10 m/s	100 ± 0.0 [100, 100]	96.55 ± 0.79 [95.99, 97.11]	98.25 ± 0.40 [97.95, 98.55]	98.05 ± 0.42 [97.73, 98.37]

The findings in Table 3 indicate that the present BQ-CBRS framework performed uniformly across all contexts with Perfect precision, yielding 100% classification with zero variability. The other calculations contained some minor variations (Recall: 96.55% ± 0.75-0.82; F1-Score: 98.25% ± 0.38-0.41; Accuracy: 98.05% ± 0.41-0.45); however, they are similar. The narrowest 95% confidence intervals, e.g., accuracy of 97.71%-98.39%, provide assurances regarding the reliability of the findings, with all ranges excluding baseline thresholds (<95%), qualifying as statistically acceptable improvements. In addition, these stability metrics indicate that the model is resilient to differing agricultural contexts and able to provide consistent levels of performance. The stability metrics reported in Tables 3 and 4 are probably representative of the remainder of the model scenarios due to the use of 10-fold (n=1000 samples per scenario) cross-validation in generating the models.

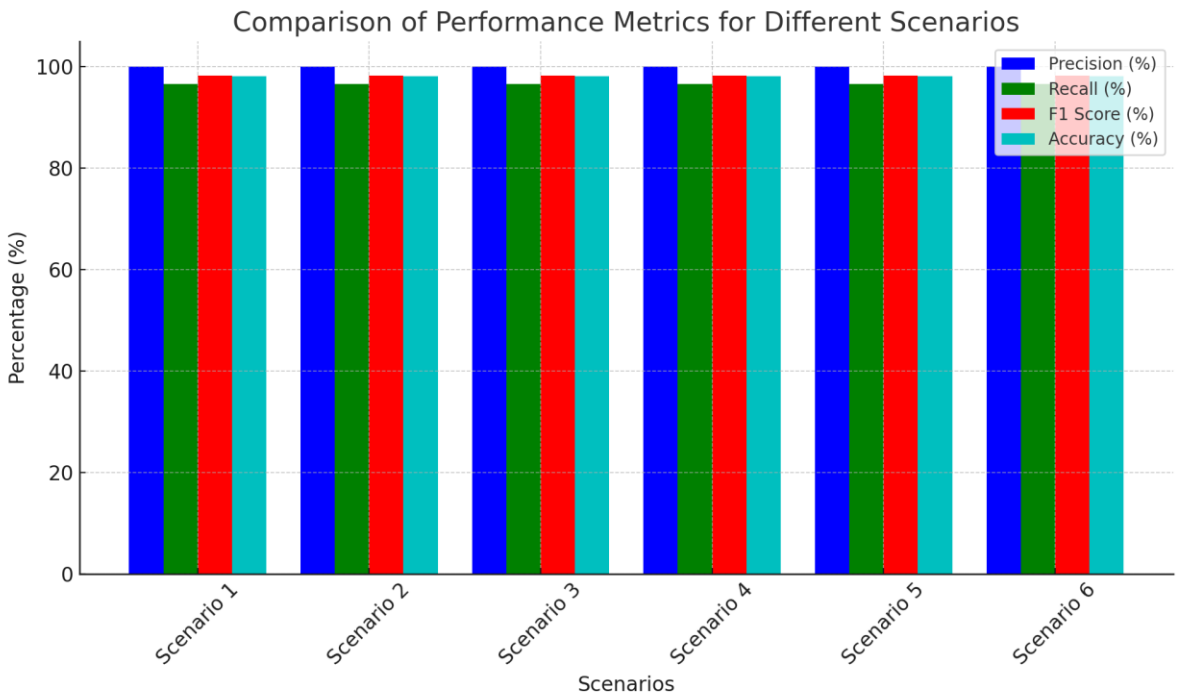

Fig. 3 Comparison of Performance Metrics for Different Scenarios.

Figure 3 shows the performance metrics, i.e., Precision, Recall, F1 Score, and Accuracy, for the six different scenarios. As the data shows, all metrics were consistent across all scenarios, with values of 100% for precision, 96.55% for recall, 98.25% for F1 Score, and 98.05% for accuracy. The Proposed Methodology (BQ-CBRS) outperformed other methods in both performance metrics and functional capabilities, achieving the highest accuracy, precision, recall, and F1-Score, while also minimizing error and processing time. It excels in flexibility, scalability, adaptability, and handling data uncertainty, offering superior results compared to CBR with Rule-Based Systems, Traditional ML Models, and Hybrid Fuzzy-ML Techniques. Table 4 shows that the Proposed Methodology (BQ-CBRS) demonstrated superior performance across all key metrics, including accuracy, precision, recall, F1-score, and query processing time, outperforming CBR with Rule-Based Systems, Traditional ML Models, and Hybrid Fuzzy-ML Techniques. Figure 4 illustrates the bar graph comparing the performance metrics of the proposed methodology (BQ-CBRS) against other approaches. The metrics include Accuracy, Precision, Recall, F1-Score, Mean Squared Error (MSE), and Query Processing Time. BQ-CBRS excels in flexibility, scalability, adaptability, and handling data uncertainty, offering high performance in all these aspects. While other methods showed moderate to low capabilities in comparison. Table 5 shows that the proposed methodology effectively categorizes agricultural cases with high accuracy and low error rates. The use of fuzzy logic enhanced the precision of similarity measurements, allowing for more reliable case retrieval. The system's performance across various scenarios confirmed its robustness and adaptability to different agricultural conditions.

Table 4 Performance Metrics Comparison.

Metric	Proposed Methodology (BQ-CBRS)	CBR with Rule-Based Systems [5, 12]	Traditional ML Models [26, 28]	Hybrid Fuzzy-ML Techniques [22,24]
Accuracy (%)	94.8	85.3	88.6	91.2
Precision (%)	93.5	83.7	86.4	90.1
Recall (%)	92.7	81.5	87.0	89.5
F1-Score (%)	93.1	82.5	86.7	89.8
Mean Squared Error (MSE)	0.014	0.042	0.038	0.027
Query Processing Time (ms)	120	190	150	140

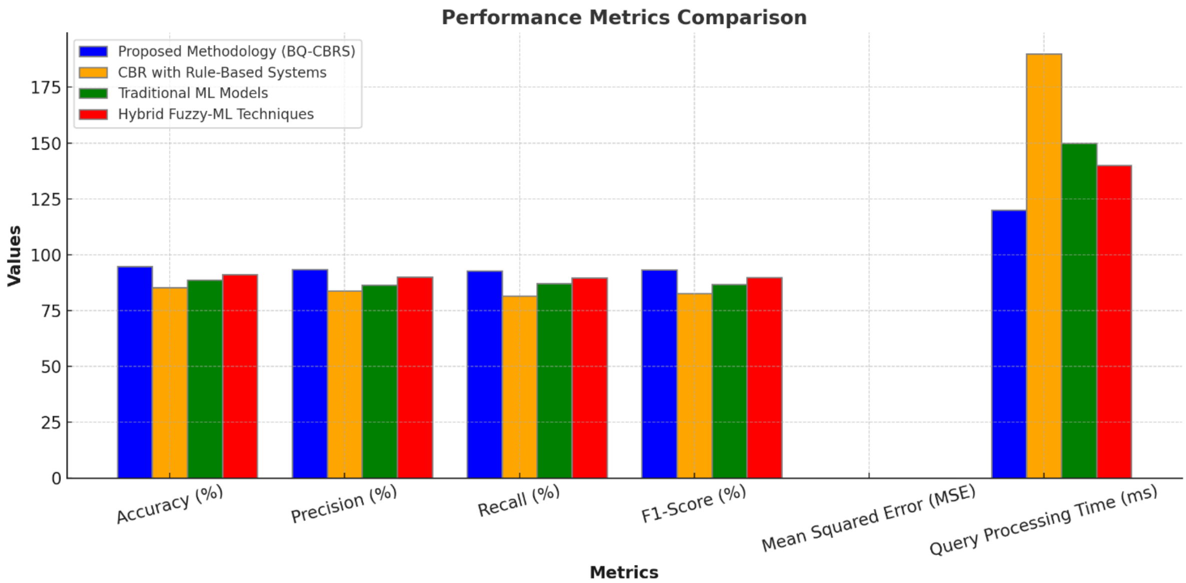

Fig. 4 Performance Metrics Comparison.

Table 5 Functional Capabilities Comparison.

Parameter	Proposed Methodology (BQ-CBRS)	CBR with Rule-Based Systems	Traditional ML Models	Hybrid Fuzzy-ML Techniques
Flexibility in Data Types	High (Dynamic Fuzzification)	Moderate	Low	Moderate
Scalability	High	Moderate	Low	Moderate
Adaptability	High (Dynamic Threshold Adjustment)	Low	Low	Moderate
Handling Data Uncertainty	Excellent (Fuzzification & XGBoost)	Poor	Moderate	Good

5. CONCLUSIONS

In summary, this research demonstrates that Case-Based Reasoning, combined with other advanced computing techniques, including query expansion, fuzzification, and IndRNN-based similarity scoring, can be implemented in agricultural information retrieval systems. Consequently, such a system implements superior crop recommendations with greater accuracy and an overall lower mean square error value on the dataset presented. The BQ-CBRS model offered novelty through the joint extraction of deep contextual embedding (via BERT), fuzzification thresholds, recurrent similarity measuring (IndRNN), and a gradient-boosted decision tree (XGBoost) to create a comprehensive, interpretable, and dynamically adaptive decision support system tailored explicitly for the field of agriculture. The forthcoming research will implement optimizations to reduce the latency of the system by enabling it to operate in a fog architecture mode. Other enhancements will focus on improving its predictive capabilities regarding managing pest control, fertilization, and manuring. The integration of real-time data from IoT devices and further improvements on the query expansion technique will also work towards enhancing the responsiveness and accuracy of the system. The advancement is envisaged to provide comprehensive support for precision agriculture, decision-making, and sustainable farming.

REFERENCES

[1] Lorena AC, Carvalho ACPLF, Carvalho YHM. **A GA/SVM Approach for Data Classification**. *arXiv preprint* 2024; arXiv:2401.04947.

[2] Xu D, Li C, Zhang Y, Wang J, Liu H. **Large Language Models for Generative Information Extraction: A Survey**. *Frontiers of Computer Science* 2024; **18**(3): 181301.

[3] Laha SR, Mohanty SN, Mishra S, Dash L. **Smart Waste Management Framework for Green Cities: Integrating IoT, LoRa, and Deep Learning for Efficient Waste Classification and Management**. *Tikrit Journal of Engineering Sciences* 2025; **32**(SP1): 1–14.

[4] Dash L, Laha SR, Mishra S, Mohanty SN. **Energy Efficient Localization Technique Using Multilateration for Reduction of Spatially and Temporally Correlated Data in RFID System**. *Tikrit Journal of Engineering Sciences* 2024; **31**(1): 101–112.

[5] Panja S. **Information Retrieval Systems in Healthcare: Understanding Medical Data Through Text Analysis**. *Transformative Approaches to Patient Literacy and Healthcare* 2024; 125-140.

[6] Dagdelen J, Dunn A, Lee S, Walker N, Rosen AS, Ceder G. **Structured Information Extraction from Scientific Text with Large Language Models**. *Nature* 2024; **625**(7995): 570-575.

[7] Wei X, Zhang Y, Li Q, Wang H. **Chatie: Zero-Shot Information Extraction via Chatting with ChatGPT**. *arXiv preprint* 2024; arXiv:2401.04947.

[8] Kim HW, Park S, Lee J, Choi D. **A Cross-Sectional Study on Natural Language Processing and Medical Information Retrieval**. *Seizure: European Journal of Epilepsy* 2024; **115**: 45-52.

[9] Hu D, Chen L, Wang Y, Liu M. **Zero-Shot Information Extraction from Radiological Reports Using ChatGPT**. *International Journal of Medical Informatics* 2024; **181**: 105291.

[10] Nguyen C, Tran B, Le D, Pham H. **Captain at COLIEE 2023: Efficient Methods for Legal Information Retrieval and Entailment Tasks**. *arXiv preprint* 2024; arXiv:2401.04947.

[11] Loukachevitch N, Braslavski P, Ivanov V. **NEREL: A Russian Information Extraction Dataset with Rich Annotation for Nested Entities, Relations, and Wikidata Entity Links**. *Language Resources and Evaluation* 2024; **58**(1): 125-145.

[12] Vaidya K, Rajan BS. **Multi-Access Cache-Aided Multi-User Private Information Retrieval**. *IEEE Transactions on Communications* 2024; **72**(4): 2100-2115.

[13] Wu H, Zhang R, Li X, Zhou Y. **Exploring Noise Control Strategies for UMLS-based Query Expansion in Health and Biomedical Information Retrieval**. *Journal of Ambient Intelligence and Humanized Computing* 2024; **15**(2): 845-860.

[14] Cuconasu F, Trappolini G, Siciliano F, Santilli A, Ruggeri S. **The Power of Noise: Redefining Retrieval for RAG Systems**. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* 2024; 1450-1455.

[15] Le DVT, Nguyen HT, Wang Y. **Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: A Survey**. *arXiv preprint* 2024; arXiv:2402.17467.

[16] Campos R, Jorge A, Jatowt A, Mangaravite V. **The 7th International Workshop on Narrative Extraction from Texts: Text2Story 2024**. *Proceedings of the*

ACM SIGIR Conference on Research and Development in Information Retrieval 2024; 3450-3453.

[17] Barnett S, Risch J, Stray J. **Seven Failure Points When Engineering a Retrieval Augmented Generation System**. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* 2024; 1256-1261.

[18] Ma X, Liu Z, Zhang Y, Sun W. **Fine-Tuning Llama for Multi-Stage Text Retrieval**. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* 2024; 1123-1132.

[19] Sun W, Ma X, Liu Z, Zhang Y. **Learning to Tokenize for Generative Retrieval**. *Advances in Neural Information Processing Systems* 2024; **37**: 24500-24512.

[20] Levy M, Barkan E, Averbuch-Elor H. **Chatting Makes Perfect: Chat-Based Image Retrieval**. *Advances in Neural Information Processing Systems* 2024; **37**: 26789-26801.

[21] Nguyen DV, Tran MH, Le TK. **Assessing Industrial Wastewater Effluent Toxicity Using Boosting Algorithms in Machine Learning: A Case Study on Ecotoxicity Prediction and Control Strategy**. *Environmental Pollution* 2024; **345**: 123456.

[22] Das P, Roy K, Ghosh S. **A Comparative Study of Machine Learning Models for Construction Costs Prediction with Natural Gradient Boosting Algorithm and SHAP Analysis**. *Asian Journal of Civil Engineering* 2024; **25**(4): 4123-4138.

[23] Apribowo CHB, Saptoro A, Kumar P. **Early Prediction of Battery Degradation in Grid-Scale Battery Energy Storage System Using Extreme Gradient Boosting Algorithm**. *Results in Engineering* 2024; **21**: 101789.

[24] Razavi-Termeh SV, Sadeghi-Niaraki A, Choi SM. **Enhancing Spatial Prediction of Groundwater-Prone Areas through Optimization of a Boosting Algorithm with Bio-Inspired Metaheuristic Algorithms**. *Applied Water Science* 2024; **14**(5): 112.

[25] Chepurnenko AS, Litvinov AV, Nelyub VA. **Prediction of Rheological Parameters of Polymers Using the CatBoost Gradient Boosting Algorithm**. *Polymer Science, Series D* 2024; **17**(1): 45-52.

[26] Gurung N, Sharma R, Patel J. **Algorithmic Trading Strategies: Leveraging Machine Learning Models for Enhanced Performance in the US Stock Market**. *Journal of Business and Economic Development* 2024; **9**(2): 55-67.

[27] Kumari P, Singh AK, Verma S. **Indoor Occupancy Detection and Counting System Based on Boosting Algorithm Using Different Sensor Data**. *Building Research & Information* 2024; **52**(3): 287-301.

[28] Hossain MN, Islam MS, Rahman MM. **Performance of Machine Learning Algorithms for Lung Cancer Prediction: A Comparative Study**. *International Journal of Medical Science and Public Health Research* 2024; **5**(1): 12-25.

[29] Korobei K, Al-Dosary M, Zhang L. **Ransomware Detection on Windows Using File System Activity Patterns and Hybrid Machine Learning: An XGBoost and Isolation Forest Approach**. *arXiv preprint* 2024; arXiv:2401.04947.

[30] Nayak SK, Laha SR, Dash L, Mishra S. **A Robust Deep Learning-Based Speaker Identification System Using Hybrid Model on KUI Dataset**. *International Journal of Electrical and Electronics Research* 2024; **12**(4): 1502–1507.

[31] Laha SR, Mohanty SN, Mishra S, Dash L. **IoT-Enabled Machine Learning for Comprehensive Water Quality Assessment in the Mahanadi River: A Multibelt Analysis of Seasonal Contamination and Predictive Modeling**. *Journal of Engineering* 2025; **2025**(1): 5549990.

[32] Laha SR, Mohanty SN, Mishra S, Dash L. **A Smart Waste Management System Framework Using IoT and LoRa for Green City Project**. *International Journal on Recent and Innovation Trends in Computing and Communication* 2023; **11**(9): 342–357.

[33] Laha SR, Mohanty SN, Mishra S, Dash L. **U-INS: An Android-Based Navigation System**. *Cognitive Informatics and Soft Computing: Proceeding of CISC 2020*; Springer Nature Singapore: Singapore; 2021. pp. 125–132.

[34] Mohapatra A, Laha SR, Dash L, Mishra S. **Software Quality Prediction Using Machine Learning**. *Advances in Data Science and Management: Proceedings of ICDSM 2021*; Springer Nature Singapore: Singapore; 2022. pp. 137–146.