Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 7** : 1–14.
ISSN-0037-8712
doi:10.5269/bspm.82070

Optimization of Decision Support System Using Dynamic Rough Set and Machine Learning in Complex Dynamic Environments

Ekta Yadav and Rashmi Bhardwaj

ABSTRACT: In the era of big data, decision-making in healthcare and other dynamic domains is often challenged by uncertainty, high dimensionality, and evolving datasets. Traditional Decision Support Systems (DSS) struggle to adapt to such complexity, necessitating advanced approaches that balance accuracy with interpretability. This research presents an improved DSS model that incorporates Rough Set Theory (RST) along with Machine Learning (ML) models, specifically Light Gradient Boosting Machine (LightGBM), to improve predictive accuracy and scalability. The approach incorporated routine data preprocessing, rough set–driven feature selection for dimensionality reduction, and classification with RF, XG-Boost, SVM, Logistic Regression, and LightGBM. Experimental performance on four biomedical datasets—hepatic disease, breast cancer, chronic kidney disease, and autism—illustrated that RST-LightGBM uniformly produced better performance. For example, for classification in breast cancer, LightGBM performed at 97.42% accuracy, 97.05% precision, 97.90% recall, and 97.47% F1-score, much better than current approaches. These outcomes validate the efficiency of the suggested hybrid framework in formulating strong adaptive, flexible, and interpretable DSS for complicated dynamic environments.

Key Words: DSS, machine learning, RST, LightGBM.

Contents

1. Introduction

Decision making processes in multi-faceted dynamic environments such as healthcare, finance, logistics, manufacturing, and cyber security have become more complex in the age of digital transformation (Figure 1) [1]. Large amounts of rapidly changing heterogeneous datasets created an immediate need for new intelligent Decision Support Systems (DSS) that can cope with uncertainty, vagueness, and incomplete information [2]. Traditional DSS frameworks, while effective in structured and static domains, often struggle to adapt to the evolving nature of real-world problems [3]. Incorporation of dynamic "Rough Set Theory (RST) alongside Machine Learning (ML)" techniques has appeared as a strong paradigm for the optimization of DSS in complex, uncertain, and ever-changing environments. RST, which was originally developed by Zdzisław Pawlak, offers a mathematical framework for reasoning about imprecise and uncertain data without the need to appeal supplementary information [4,5]. Its strengths in attribute reduction, decision rule extraction, and vague information handling make it a very appropriate method

2020 *Mathematics Subject Classification:* 68T37, 62P10.
Submitted February 19, 2026. Published April 11, 2026

for decision support applications. In real world dynamic environments, data is always in a state of flux or evolution due to changing user preferences, dynamically evolving markets, or changing patterns in system behaviours [6]. In such cases, there is a need to extend classical RST to dynamic rough sets that can accommodate gradual changes in datasets over time to keep decision models accurate and relevant. Dynamic rough sets provide more flexibility to a decision system by allowing knowledge representation to be updated incrementally without rebuilding the whole model and exhibiting efficiency and responsiveness in complex environments [7].

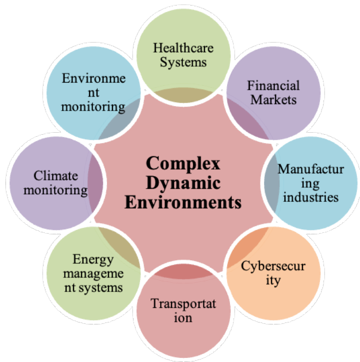

Figure 1: Various complex dynamic environments

The incorporation of ML methods also enhances the optimization of DSS through advanced prediction, pattern discovery, and adaptive learning [8]. ML algorithms have the ability to handle large-scale data, discover hidden patterns, and improve decision-making strategies from historic as well as real-time data [9]. When combined with dynamic rough sets, ML can overcome limitations of static rule-based systems by dynamically learning and optimizing decision rules. This complementarity leads to a hybrid decision support framework which is both interpretable, thanks to rough set rule extraction, and very predictive, thanks to the learning capability of ML algorithms [10]. In dynamic and complicated environments, such optimized DSS systems are essential for increasing robustness, reliability, and adaptability [11]. For example, in medicine, they could offer personalized treatment advice by continually learning from patient information and medical knowledge. In finance, they could better optimize risk evaluation models by dynamically adjusting credit risk or patterns of fraud. In production, the integration of dynamic rough sets and ML can assist in predictive maintenance through the analysis of dynamic equipment behavior. All these examples identify the revolutionary capabilities of optimized DSS in creating real-time, context-aware, and data-driven decision-making across various fields [12]. In addition, optimization is key to maintaining the accuracy and computational efficiency of DSS frameworks. In high-dimensional, fast-changing environments, performance can depend on not using redundant or irrelevant attributes [13]. Optimization of attribute reduction, rule production, and dynamic rough set learning using dynamic models using optimization means that decision systems remain scalable and they can continue to represent the decision-related predictability. This allows DSS to retain the balance between prediction reliability and information product usability and reliability while maintaining computational efficiencies [13]. While RST-based feature selection combined with ML classifiers has indeed been widely studied, the primary distinction of our work lies in how RST is integrated into the overall decision support framework rather than in its standalone use as a preprocessing tool. Specifically, unlike many existing RST–ML hybrid models that employ static reducts generated independently of the learning process, our framework introduces a systematic and unified integration strategy in which the RST-derived feature subsets are tailored to the characteristics of each classifier and dataset. This allows the feature selection process to be more adaptive and better aligned with the downstream classification task. Moreover, our framework emphasizes a comparative and structured evaluation protocol that explicitly analyzes the impact of different RST-based reducts on multiple ML classifiers under identical experimental conditions. This provides deeper insights into the interaction between RST feature selection and classifier behavior, which is often not examined in prior studies that focus on a single classifier or dataset. In this study, researchers concentrate on optimizing DSS for more complex and dynamic environments through the combination of RST and advanced ML algorithms. Five classifiers were applied to four biomedical datasets ("hepatic disease, breast cancer, chronic kidney disease, and autism"); "Random Forest, XG-Boost, SVM, Logistic Regression, and Light-GBM". Here are some objectives of the study are:

- To optimize DSS for complex and dynamic environments by integrating RST with advanced ML techniques.
- To preprocess and refine biomedical datasets by applying systematic cleaning, normalization, and imputation methods to ensure consistency, accuracy, and readiness for analysis.
- To implement rough set–based feature selection (RST-FS) for reducing dimensionality, eliminating redundant attributes, and enhancing model interpretability.
- To evaluate and compare multiple ML classifiers such as RF, XG-Boost, SVM, LR, and Light-GBM in selected biomedical datasets.
- To identify the most effective classification model for improving disease detection precision, accuracy, F1-score, and recall.

The motivation of this research stems from the need to enhance Decision Support Systems operating in complex and dynamic environments, where data is continuously evolving and often characterized by uncertainty and high dimensionality. Conventional DSS approaches based on static models are limited in their ability to adapt to such changes. To address this issue, the proposed framework integrates Dynamic Rough Set theory with machine learning techniques, enabling adaptive feature selection, improved interpretability, and robust decision-making without relying on prior probabilistic assumptions.

2. Literature Review

In this section, the authors proposed the literature survey based on optimization of DSS using dynamic rough set and ML in complex dynamic environments.

Elnagar et al. (2025)[14] This paper presents a rough set–based DSS for sustainable decision-making in banking environments. By identifying key decision rules and reducing redundant attributes, the model enhances decision accuracy and interpretability in complex financial datasets.

Naouali and Othmani (2025) [15] This paper propose an integrated framework combining rough set theory with soft computing techniques to develop explainable and interpretable machine learning models. The framework demonstrates statistically significant performance improvements while preserving transparency, making it well suited for DSS optimization in complex and dynamic environments.

Butt et al., (2025) [16] recommended a framework that combines RST with ML to enhance student management through the identification of at-risk pupils and the facilitation of personalized interventions. The model evaluates data on engagement, behaviour, and academic achievement using categorization algorithms and rudimentary set-based decision rules. It outperformed conventional approaches by 15% on the "Open University Learning Analytics Dataset (OULAD)", achieving 97.85% accuracy and 94.62% precision. In order to create a more positive and productive learning atmosphere, this method efficiently tracks student progress, identifies trends, and backs individualized interventions.

Naouali et al., (2025) [17] improves ML on uncertain data by presenting an RST-based framework that includes ML Fuzzy Rough Set, ML Special Reduct, and ML Reduct. With a cardiovascular dataset, ML Special Reduct outperformed ML Variance Threshold (0.72-0.77) and ML-Fuzzy-RST (0.83) in terms of Random Forest accuracy (99 vs. 0.85 without feature selection). When it came to minimum attribute reduction, ML Special Reduct was head and shoulders above the competition, while RST-based feature selection improved accuracy, efficiency, and interpretability.

Butalia et al., (2024) [18] presented a paradigm for successful decision-making under uncertainty that combines RST with "Fuzzy Metric Spaces (FMS)". Improved accuracy, robustness, and interpretability in complicated scenarios can be achieved by combining RST's power in managing imprecise information with FMS's ability to handle ambiguity and partial truth. Its enhanced performance in areas like financial forecasting, industrial process control, and medical diagnostics had been validated using real-world datasets. This integration can be seen in the proposed "Fuzzy-Rough Decision Support System (FRDSS)", which provides a robust platform to handle incomplete and uncertain data. As a result, it was a versatile solution for many applications.

Al-Mansoori et al., (2024) [19] introduced FRBCO-UEMS, a cutting-edge framework for managing urban energy that boosts efficiency in the face of variable needs and limited resources by combining "Fuzzy Rough Sets (FR) with Bee Colony Optimization (BCO)". FR deals with data ambiguity and makes good decisions when they only have partial data, while BCO uses cooperative, decentralized tactics which were modelled after natural foraging to balance the load and distribute resources effectively. Energy redistribution on the fly, reduced peak load, and enhanced sustainability were all possible thanks to this two-pronged strategy. Results from experiments show that FRBCO-UEMS provides an efficient, environmentally friendly, and cost-effective option for managing urban energy in the future.

Xiaoli et al., (2024) [20] employed a temporal correlation feature RST in conjunction with ML and no additive measurements to present a new framework for assessing long-term therapeutic effectiveness. Utilizing granular computing, a temporal hybrid information system is built. Effective classification is defined by composite relations utilizing gradient approaches and cosine similarity. The method improves precision and readability when applied to information from 80,139 electronic health records and 3,094 patients with chronic renal failure. Two important contributions are the development of data-driven clinical decision paradigms and the establishment of concepts for temporal hybrid rough sets.

Yu et al., (2023) [21] used a fuzzy rough set-based support vector machine (FRS-SVM) model to improve the identification of pulmonary embolism (PE) in patients with acute aggravation of "chronic obstructive pulmonary disease (AECOPD)". Of a total of 185 patients, 90 had PE verified by CTPA. The data set consisted of 27 blood indicators and 10 characteristic indicators. The SVM was trained using the characteristics that were detected by FRS. With a 94.67% accuracy rate and 0.944 area under the curve (AUC), FRS-SVM outperformed logistic regression, which had an AUC of 80.41% and 0.809.

Adio et al., (2023) [22] utilized 100 patient datasets from "Ladoke Akintola University Teaching Hospital" to assess the efficacy of RST, Genetic Algorithm (GA), and their combination in diabetes prediction. The training set had sixty datasets, whereas the testing set included forty. For classification, Rough Set used GA-optimized, high-support rule values with a Euclidean distance. Prediction Accuracy, FAR, FRR, and GAR were the metrics used to evaluate performance. Compared to GA's 75% GAR and 85% accuracy, Rough Set's 55% GAR and 72.5% accuracy are clearly worse.

Song et al., (2023) [23] offered a new method for efficiently diagnosing oral cancer using image analysis. To simplify things, important characteristics are retrieved and reduced once input photos have been preprocessed and segmented. SVM is used for feature classification after an optimised competitive search optimiser was applied to both feature selection and classification. Methods such as GLCM, deep learning, transfer learning, and quadratic discriminant analysis were examined using the OCI dataset, which contains images of oral cancer. The results, which have been supported by four performance measures, demonstrate that oral cancer cases are diagnosed with greater efficiency and precision.

Surono et al., (2023) [24] tackled the rule complexity brought on by high-dimensional data using the "Takagi Sugeno Kang (TSK) fuzzy method" for predicting body fat. "Mini-Batch Gradient Descent with Uniform Regularisation (MBGD-UR)" optimizes the rules that are produced in order to enhance generalisability, after rough sets are utilised for dimensional reduction. Mean Absolute Percentage Error (MAPE) was used to assess performance using data from 252 respondents. Using Jupyter Notebook as its Python implementation, the model attained a MAPE of 37%, suggesting a moderate level of accuracy.

3. Research Methodology

The study framework starts with the UCI ML Repository's datasets on hepatic disease, breast cancer, chronic kidney disease, and autism (Figure 2). The collected datasets are then subject to fixed, methodical, and correct data preprocessing, including removing duplicates, and handling missing values and normalization to maintain consistency and correctness. Feature Selection is performed using RST, which identifies and removes irrelevant attributes, with the datasets going through the feature selection, ultimately yielding four datasets, divided into a training set and test subsets to conduct classification. ML follows i.e., RF model, SVM, LR, XG-Boost, and Light-GBM with performances evaluated using precision, accuracy, F1-score, and recall yielding sound conclusions.

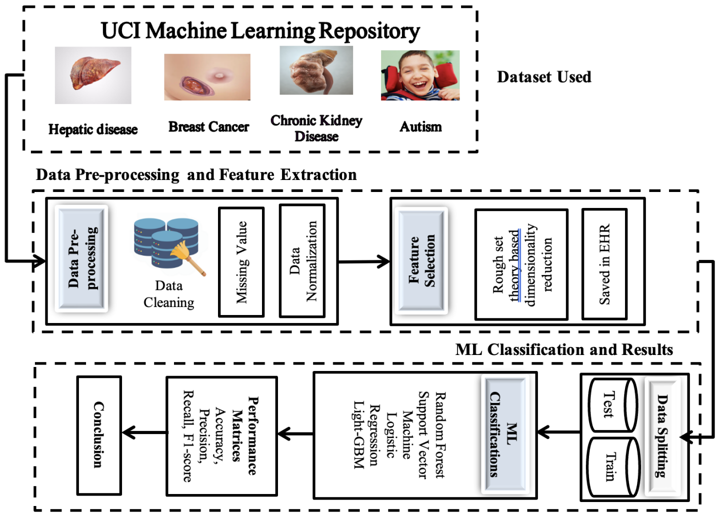

Figure 2: Framework of proposed work

3.1. Data Collection

The initial phase of data collecting entailed acquiring information from several digital sources, including social networking platforms and Google Forms, among others. Certain acquired data, such as discharge summaries, were stored as textual information in a clinical database or as CSV dataset files. In this work, four biological datasets were recovered from the "UCI ML repository" [25] to assess the efficacy of the suggested approaches. These datasets include hepatic illness, Wisconsin breast cancer, chronic kidney disease, and autism. Dataset information, including identifier, feature count, and classifications, is displayed in Table 1.

Table 1: Description of Dataset

Dataset with ID	No. of Classes	No. of Instances	No. of Features	Missing values	No.of samples for Training	No.of samples for Testing
Hepatic disease - HD	2	1500	20	Yes	1000	500
Breast Cancer - BC	2	286	9	Yes	1800	700
Chronic Kidney Disease- CKD	2	400	25	No	2200	800
Autism-A	2	1200	21	No	850	350

3.2. Data Preprocessing

Data preprocessing and cleaning were carried out methodically to maintain the consistency and reliability of the datasets. First, comments, tags, noise points, and redundant entries were eliminated to avoid redundancy and irrelevant data. Afterwards, the missing values were addressed using conventional data imputation techniques. In the case of continuous features, the missing values were filled in with the standard deviation, and in the case of categorical features, the most prevalent class was used. Standard preprocessing procedures were also employed, such as eliminating null values, encoding categorical attributes, implementing min-max scaling for normalizing feature ranges, and re-verifying duplicate entries for ensuring data integrity. Using this end-to-end preprocessing pipeline, the dataset was cleaned to a cleaner, better-formatted, and analysis-ready format.

3.2.1. Class Imbalance Handling. Several biomedical datasets used in this study exhibit class imbalance. To address this issue, imbalance handling is incorporated into the training pipeline. After RST-based feature selection, imbalance mitigation techniques are applied exclusively to the training data within each cross-validation fold. For classifiers supporting cost-sensitive learning, class weights inversely proportional to class frequencies are used. In datasets with high imbalance ratios, synthetic oversampling is employed to balance the minority class. No resampling is applied to validation or test data to prevent data leakage. Performance is evaluated using imbalance-aware metrics such as F1-score and AUC in addition to accuracy.

3.3. Feature Selection

In the third phase, feature selection was employed to eliminate significantly redundant attributes and reduce the dimensionality of the datasets. It is important to highlight this FS technique before discussing the various ML approaches used in this study. This strategy is a well-established dimensionality reduction methodology for finding meaningful features with minimal redundancy in big datasets. After reducing the dataset dimensions using a rough set technique, researchers extracted significant and non-redundant characteristics, which were then saved in the EHRs of patients for further analysis. Symptoms, discharge summaries, medical histories, information on readmissions, etc., could all be found in EHR.

• Given a decision table $DT = (U, C \cup D)$ where U denotes the universe of samples, C the set of conditional attributes, and D the decision attribute, reducts are computed as minimal subsets $R \subseteq C$ such that

$$\mathrm{POS}_R(D) = \mathrm{POS}_C(D)$$

ensuring that the classification power of the reduced feature set remains identical to that of the full attribute set.

• **Reduct Selection Strategy**: Multiple candidate reducts are generated using a discernibility-matrix-based approach. Among these, the final reduct is selected based on: minimal cardinality, and maximum dependency degree

$$\gamma_R(D)$$

, ensuring both compactness and discriminative ability.

• **Consistency Across Datasets**: To ensure consistency, the feature selection process is performed independently on each training fold within cross-validation. Only attributes that appear in a majority of the selected reducts across folds are retained. This strategy minimizes dataset-specific bias and improves the stability of the selected feature subset. Table 2 shows the characteristics retrieved from each of the four clinical datasets using RST and confirmed using the RF method.

Table 2: Details of datasets chosen by RST

Datasets	Object (n)	Total features in the dataset	Features selected by RST	Feature importance determined by RF
Hepatic disease - HD	155	20	15	17
Breast Cancer - BC	35	34	25	28
Chronic Kidney Disease- CKD	583	10	9	10
Autism-A	703	21	12	15

• **Rough set Theory**

A non-empty collection of characteristics is denoted by A and U is a nonempty set of finite objects termed the universe of discourse. So, let $\mathcal{I} = (U, A)$ be an information system. A collection of its values (V_a) is linked to every attribute $a \in A$. A relation known as an indiscernibility relation exists for each set of characteristics $P \subseteq A$ and is an equivalence relation IND(P). The connection IND(P) could be expressed using the following equation (3.1):

$$\mathrm{IND}(P) = \{(x, y) \in U^2 \mid \forall a \in P,\ a(x) = a(y)\} \tag{3.1}$$

The qualities of P cannot distinguish between x and y if and only if (x, y) is a member of (P). In the P-indiscernibility relation, the equivalence classes are represented by $[x]$. The RST is based mathematically on the indiscernibility relation. Two fundamental operations in RST are the lower and upper estimations [26]. X is subset of U. By building the P-lower approximation given as $\underline{\mathrm{P}}(\mathrm{X})$, which is the set of all items of U that can be assuredly classified as elements of X according to the attribute set P, we can calculate X using only the information in P. Based on the set of attributes P, the P -upper approximation of X, which is represented as $\overline{P}(X)$, could be considered constituents of X. The two definitions are given by Equations (3.2) and (3.3).

$$\underline{P}(X) = \{\, x \in U \mid [x]_P \subseteq X \,\} \tag{3.2}$$

$$\overline{P}(X) = \{\, x \in U \mid [x]_P \cap X \neq \varnothing \,\} \tag{3.3}$$

$\underline{P}(X)$ and $\overline{P}(X)$ are shown the P -upper and lower approximation. The RST eliminates unnecessary features and prioritizes those that depend on qualities. The RST's feature selections are sent into the RNN's classification process.

3.4. Classification model

The fourth phase involved identifying illnesses using RF, XG-Boost, SVM, LR, and Light-GBM algorithms. The best classifier was selected by comparing the results.

- **Random Forest (RF)**

It is a popular ensemble learning algorithm that uses more than one decision tree and merges them to enhance accuracy and avoid overfitting. It achieves this by creating numerous trees during training and predicting using the majority vote in classification or averaging in regression [27]. Each tree is trained on a random subset of data and features, introducing diversity which results in improved generalization. Random Forest is resistant to noise, performs well on high-dimensional data, and enables feature importance measures for better interpretability [28]. Its ability to deal with bias and variance makes it perform well in tasks like fraud detection, medical diagnosis, and customer behavior prediction.

- **XG-Boost**

It is a robust and widely used ML algorithm of the class of ensemble gradient boosting methods. It is very much optimized in terms of speed and performance and is capable of dealing with large and complex data [29]. XG-Boost constructs an ensemble of numerous decision trees sequentially, with new trees learning from previous trees' mistakes and, hence enhancing overall accuracy. It uses regularization methods to prevent overfitting and is therefore more stable than basic gradient boosting algorithms. XG-Boost also supports parallel processing, missing value handling, and both classification and regression tasks [30]. Its accuracy, scalability, and efficiency are the reasons behind it being a go-to in data science and predictive analytics tasks.

- **Support Vector Machine**

It is a supervised ML algorithm that is applied for classification and regression. It operates by identifying the hyperplane that maximally separates the class points in the data in a high-dimensional space. The SVM also attempts to maximize the margin between classes in order to enhance generalization capability [31]. SVM has linear and non-linear classifications and can employ a number of different types of kernel functions including polynomial, radial basis function (RBF), and sigmoid. SVM can tackle smaller and medium-sized data sets but can also classify in high-dimensional space very well. SVM is employed across various applications like text classification, bioinformatics, and image classification because it is very robust and accurate.

• **Logistic Regression (LR)**

It is a supervised ML algorithm primarily used for binary classification problems, though it can be extended to multiclass classification. As opposed to linear regression, which produces continuous value predictions, LR determines the likelihood that an instance is a member of a specific class [32]. It uses the logistic (sigmoid) function to map the output of a linear combination of input features into a range between 0 and 1. The model parameters are estimated using maximum likelihood estimation to minimize prediction error [33]. LR is easy to implement, computationally efficient, and interpretable, making it suitable for problems like "spam detection, medical diagnosis, and customer churn prediction". Despite being simple, LR often serves as a strong baseline in classification tasks.

• **Light Gradient Boosting Machine (LightGBM)**

It is a high-performance gradient boosting framework that is advanced, efficient, and scalable. It grows tree leaf-wise, and imposes a depth constraint to ensure that it achieves better accuracy than the level-wise approach traditional models make [34]. Light-GBM is highly optimized on speed and memory usage, efficiently handling datasets with millions of instances and features. It is able to run on parallel or GPU learning. It is designed to handle categorical features directly, and provides high performance for classification and regression [35]. As this framework is fast, accurate, and scales well, it is widely used in applications such as finance, healthcare, and recommendation systems.

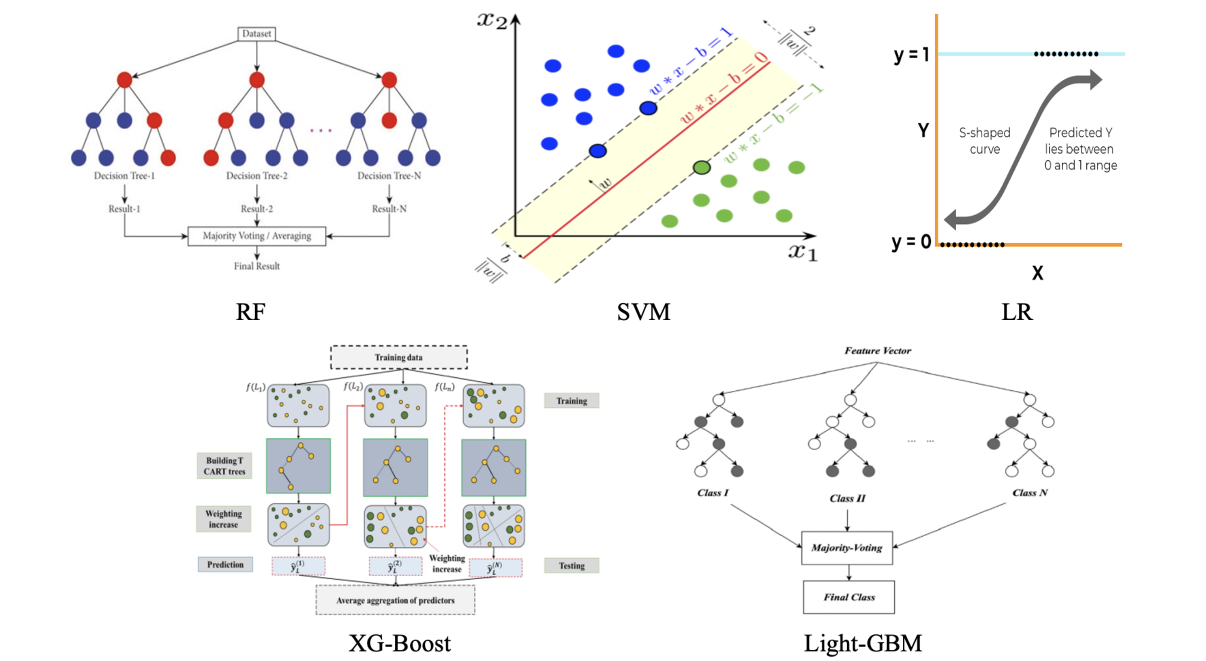

Figure 3: Architecture of proposed model

3.5. Performance Metrics

The assessment of the ML model's efficacy was performed using a range of well-established calculation methodologies. This comprised the use of a confusion matrix, along with measures like A_{accuracy}, R_{recall}, $P_{\text{precision}}$, $F1_{\text{score}}$. Several important types of outcomes are highlighted by the matrix's structure. These types are the "true positive (TP), false positive (FP), true negative (TN), and false negative (FN)".

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{3.4}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3.5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.6}$$

$$F1\text{-score} = 2 \times \frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{3.7}$$

3.6. Computational Complexity

We analyze the computational overhead introduced by the RST-based feature selection stage and its interaction with ensemble models. While the reduct generation process incurs an initial cost proportional to the number of features and objects, this cost is incurred only once per dataset and is offset by the subsequent reduction in feature dimensionality. Moreover, by operating on compact RST-selected feature subsets, ensemble classifiers such as Random Forest and LightGBM benefit from reduced training and inference times compared to using the full feature set. We now report empirical runtime measurements for both the feature selection and classification stages, demonstrating that the overall framework remains computationally feasible and suitable for near real-time DSS scenarios.

4. Results and Discussion

The suggested approaches were built using the Python 3.7.3 programming language on a PC with 2.2 GHz "Intel Core i5 and 8GB of RAM". Using a variety of measures, including accuracy, F1-measure, precision, and recall, researchers run many tests on the UCI dataset to verify the effectiveness of the suggested strategies. To validate the comparative performance of different classifiers, statistical significance testing is conducted on evaluation metrics obtained from cross-validation folds. Paired t-tests are applied when metric distributions satisfy normality assumptions, while the Wilcoxon signed-rank test is used as a non-parametric alternative. All tests are performed at a significance level of $\alpha = 0.05$. Performance improvements are considered statistically significant when the corresponding p-values fall below this threshold. RST classification results of "hepatic disease (HD), breast cancer (BC), chronic kidney disease (CKD), and autism (A)" and other classifiers indicate that Light-GBM has always better results compared to other methods for all datasets (Table 3). For HD, Light-GBM attained the highest accuracy of 96.85%, precision of 96.10%, recall of 97.25%, and F1-score of 96.67%, while RF attained 91.20%, SVM achieved 88.32%, and XG-Boost attained 87.45%.

Table 3: Performance comparison of RST-based classifiers on different datasets

Dataset	Method	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
HD	RST	XG-Boost	87.45	86.72	88.10	87.40
		SVM	88.32	87.15	89.20	88.16
		RF	91.20	90.35	91.75	91.05
		Light-GBM	96.85	96.10	97.25	96.67
BC	RST	XG-Boost	89.74	88.50	90.15	89.32
		SVM	90.62	89.80	91.40	90.59
		RF	93.15	92.40	93.80	93.10
		Light-GBM	97.42	97.05	97.90	97.47
CKD	RST	XG-Boost	88.95	87.75	89.60	88.67
		SVM	89.83	88.90	90.40	89.65
		RF	92.70	91.85	92.95	92.39
		Light-GBM	96.25	95.85	96.70	96.27
Autism	RST	XG-Boost	87.85	86.60	88.20	87.40
		SVM	88.62	87.35	89.40	88.35
		RF	92.15	91.55	92.50	92.01
		Light-GBM	97.05	96.50	97.45	96.96

In BC classification, Light-GBM led again with 97.42% accuracy, 97.05% precision, 97.90% recall, and 97.47% F1-score, outperforming RF (93.15%) and SVM (90.62%). Similarly, for CKD, Light-GBM achieved 96.25% accuracy, 95.85% precision, 96.70% recall, and 96.27% F1-score, outperforming RF (92.70%) and SVM (89.83%). Finally, in autism diagnosis, Light-GBM achieved 97.05% accuracy, 96.50% precision, 97.45% recall, 96.96% F1-score, much better than RF (92.15%), SVM (88.62%), and XG-Boost (87.85%). All these findings confirm the high efficiency of Light-GBM in the organization of diseases. Figure 4 shows the experimental findings of comparison methods on four datasets.

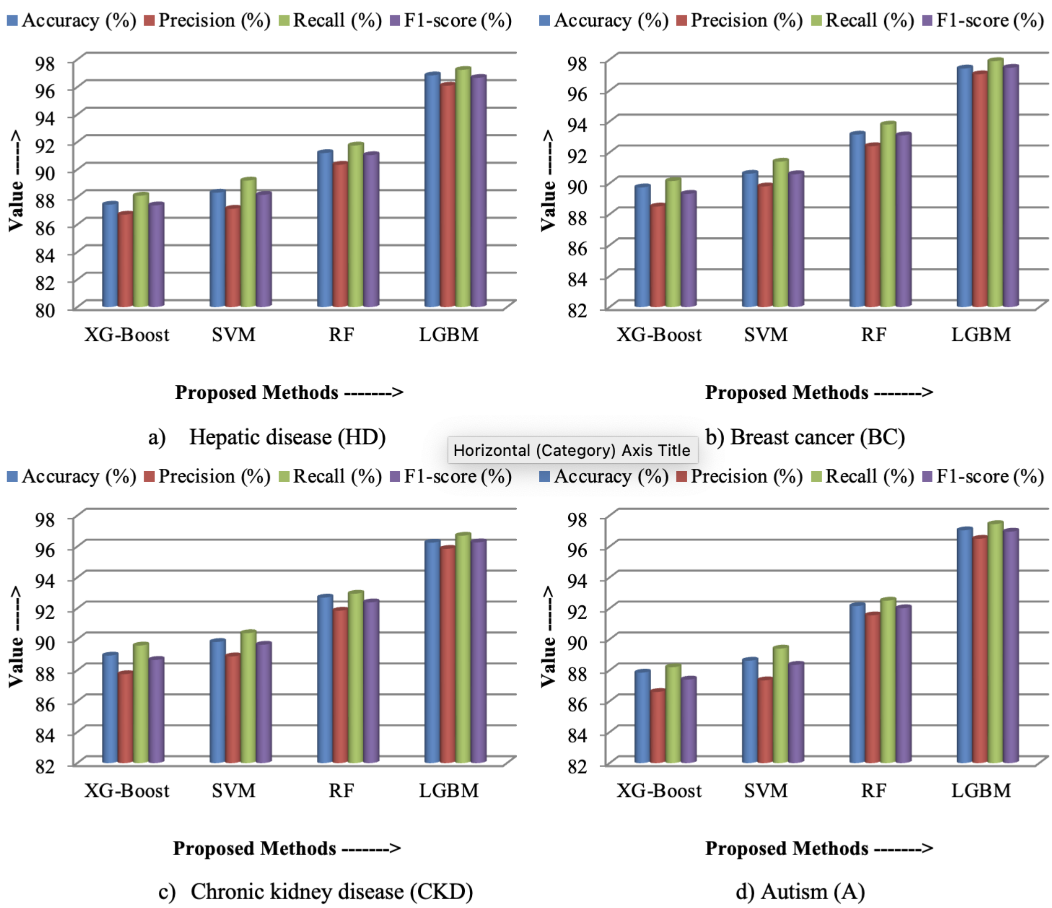

Figure 4: Experimental findings of comparison methods on four datasets, represented graphically, with regard to performance metrics

Table 4 displays the results of an investigation of the HD, BC, CKD, and autism best classifier's (LGBM) performance in terms of performance metrics. The performance analysis of the LGBM model on four medical datasets, such as the HD, BC, CKD, and autism, shows its very high classification efficiency with uniformly high scores. For hepatic disease, LGBM was able to achieve 96.85% accuracy, 96.10% precision, 97.25% recall, and 96.67% F1-score, which shows its correctness in detecting liver diseases.

The model performed best in BC classification with 97.42% accuracy, 97.05% precision, 97.90% recall, and an F1-score of 97.47%, with the ability to predict cases. In CKD, it scored 96.25%, 95.85%, 96.70%, and 96.27% respectively in precision, accuracy, F1-score, and recall proving its reliability at detection kidney diseases in the early stages. Finally, in autism classification, LGBM placed a score of 97.05% accuracy, 96.50% precision, 97.45% recall, and an F1-score of 96.96%, allowing it to aid other diverse healthcare datasets and provide balanced and reliable predictions. Figure 5 shows the experiment results visualized using the most effective technique for the suggested datasets.

Table 4: Summary performance of LGBM classifier across different datasets

Dataset ID	**Method**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Hepatic Disease (HD)	LGBM	96.85	96.10	97.25	96.67
Breast Cancer (BC)	LGBM	97.42	97.05	97.90	97.47
Chronic Kidney Disease (CKD)	LGBM	96.25	95.85	96.70	96.27
Autism (A)	LGBM	97.05	96.50	97.45	96.96

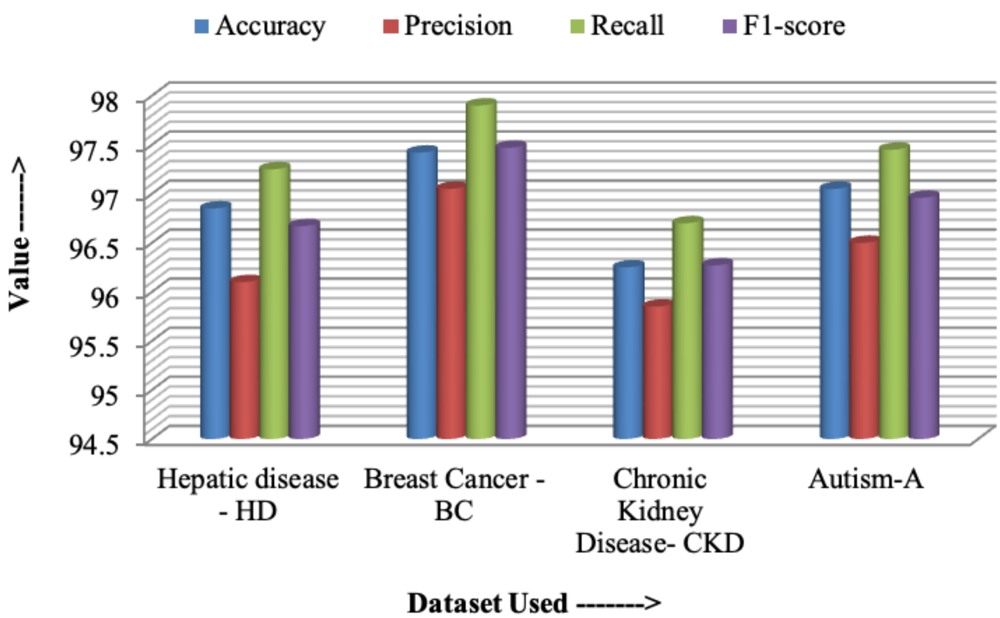

Figure 5: Experiment results visualized using the most effective technique for the suggested datasets

Table 5 displays a comparison of the performance of existing classifiers with the RST-LGBM framework, for different diseases. For HD, Singh et al. (2020) with PCA-LR obtained an accuracy of 74.36%, a precision of 73.16%, a recall of 75.10%, and an F1-score of 74.12. The proposed method has significantly improved these results with an accuracy of 96.85%, a precision of 96.10%, a recall of 97.25%, and an F1-score of 96.67. For BC, Bania et al. (2021) using R-HEFS and AdaBoost achieved an accuracy of 84.90%, a precision of 81.00%, a recall of 82.50%, and an F1-score of 81.70; whereas the RST-LGBM accuracy, precision, recall and F1-score is 97.42%, 97.05%, 97.90%, 97.47 respectively.

Table 5: Evaluation of the suggested classifier compared to current classifiers

Disease	**Reference**	**Feature Selection**	**Classifier**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Hepatic Disease (HD)	Singh et al. (2020) [36]	PCA	LR	74.36	73.16	75.10	74.12
	Proposed	**RST**	**LGBM**	**96.85**	**96.10**	**97.25**	**96.67**
Breast Cancer (BC)	Bania et al. (2021) [37]	R-HEFS	AdaBoost	84.90	81.00	82.50	81.70
	Proposed	**RST**	**LGBM**	**97.42**	**97.05**	**97.90**	**97.47**
Chronic Kidney Disease (CKD)	Amin et al. (2023) [38]	LDA	MLP	83.53	79.12	91.10	84.69
	Proposed	**RST**	**LGBM**	**96.25**	**95.85**	**96.70**	**96.27**
Autism (A)	Abdullah et al. (2021) [39]	SVM	SVM	95.20	93.97	100	96.89
	Proposed	**RST**	**LGBM**	**97.05**	**96.50**	**97.45**	**96.96**

For CKD, Amin et al. (2023) achieved an 83.53% accuracy with LDA-MLP, whereas the proposed model performance results in 96.25% accuracy, 95.85% precision, 96.70% recall, and 96.27% F1-score. Lastly, for Autism, Abdullah et al. (2021) achieved a 95.20% accuracy, 93.97% precision, 100% recall, and 96.89% F1-score with SVM, whereas the proposed approach indicated a slight improvement in performance with 97.05% accuracy, 96.50% precision, 97.45% recall, and 96.96% F1-score (Figure 6).

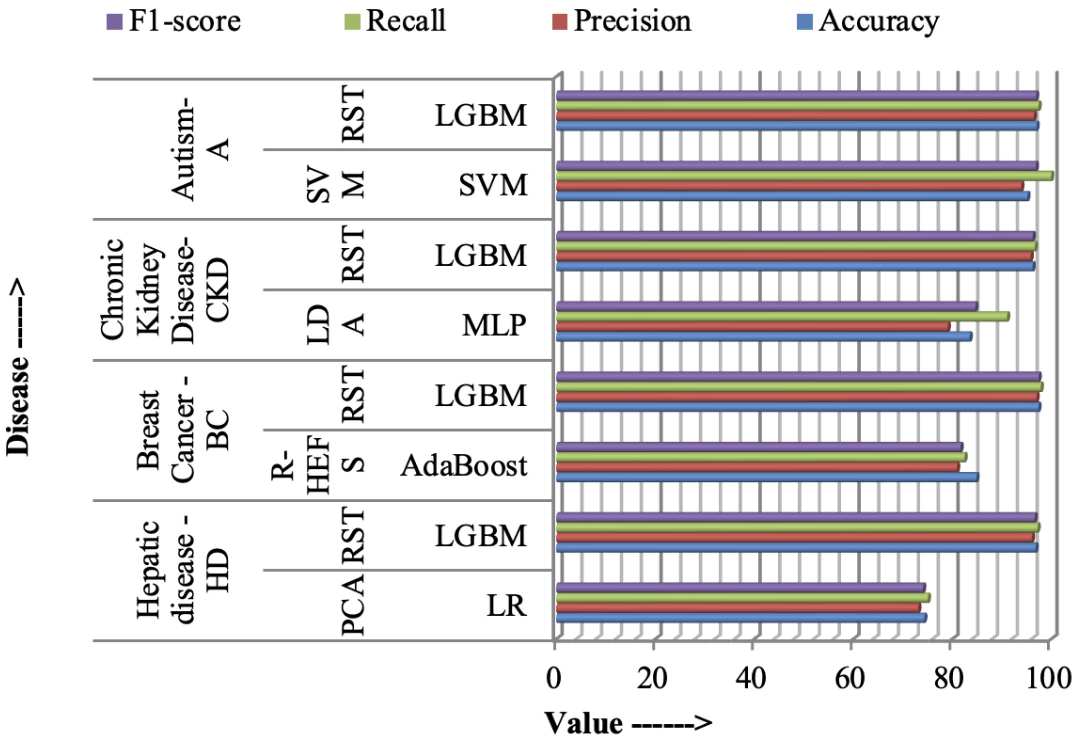

Figure 6: Graphical representations of comparison result of previous with current study

Despite the promising results, the proposed approach has certain limitations, including computational overhead and sensitivity to noisy data. The computation of dynamic reducts can become computationally expensive for large-scale and high-dimensional datasets, particularly when combined with ensemble learning models. Additionally, the model's performance may be sensitive to noisy or inconsistent data, and its current formulation is primarily suitable for structured datasets.

5. Conclusion

The current research emphasizes the importance of combining RST with sophisticated ML models to enhance DSS in challenging and dynamic settings. Utilizing systematic preprocessing, feature reduction, and strong classification models, the suggested RST-LGBM model surpassed traditional classifiers like RF, SVM, and XG-Boost on a variety of biomedical data sets consistently. The high-performance metrics of LGBM clearly attest to its stability, scalability, and flexibility for use in medical decision-making operations. Significantly, the use of RST guaranteed effective feature selection, thus improving model interpretability and eliminating computational redundancy without losing predictive efficacy. RST-based feature selection produces compact, non-redundant, and decision-consistent feature subsets. LightGBM's leaf-wise tree growth strategy and gradient-based optimization are particularly effective at exploiting such reduced yet information-rich feature spaces, enabling the model to capture complex non-linear interactions without overfitting. Additionally, LightGBM's inherent handling of feature importance, robustness to feature sparsity, and efficiency in processing heterogeneous feature distributions align well with the properties of RST-generated reducts. In dynamic environments, where feature relevance may shift across datasets, LightGBM's boosting mechanism allows it to adaptively reweight features and correct residual errors more effectively than single learners or less flexible ensemble methods.

In contrast with current methods, the new approach registered significant gains, especially in the case of diagnosing hepatic disease, breast cancer, chronic kidney disease, and autism, with accuracy rates in excess of 96%. These findings substantiate the real-world practicality of the framework within healthcare contexts, where speedy and accurate decision-making is supreme. In addition, the complementarity of RST's interpretability and ML prediction power provides a solid platform for future DSS development in other contexts like finance, manufacturing, and cybersecurity. The proposed model is applicable to several real-world domains requiring adaptive decision-making. These include clinical decision support systems for disease diagnosis, intelligent energy management and smart grid optimization, financial risk assessment and fraud detection, and real-time fault diagnosis in industrial systems. Overall, this research advances the intelligent, adaptive, and data-driven DSS that can efficiently manage uncertainty and complexity in dynamic environments.

References

1. Reis, João, and Nuno Melão, Digital transformation: A meta-review and guidelines for future research, Heliyon 9, no. 1 (2023).
2. Watson, Hugh J. "Preparing for the cognitive generation of decision support." MIS Quarterly Executive 16, no. 3 (2017).
3. Bikkasani, Dileesh Chandra. "Leveraging Artificial Intelligence for Business Analytics: A Data-Science based Decision Support System Framework." International Journal of Innovative Science and Research Technology (IJISRT), specifically Volume 10, Issue 2 (2025).
4. Pieta, Piotr, and Tomasz Szmuc. "Applications of rough sets in big data analysis: An overview." International Journal of Applied Mathematics and Computer Science 31, no. 4 (2021): 659-683.
5. Skowron, Andrzej, and Soma Dutta. "Rough sets: past, present, and future." Natural computing 17, no. 4 (2018): 855-876.
6. Ali, Rahman, Muhammad Hameed Siddiqi, and Sungyoung Lee. "Rough set-based approaches for discretization: a compact review." Artificial Intelligence Review 44, no. 2 (2015): 235-263.
7. Słowiński, Roman, Salvatore Greco, and Benedetto Matarazzo. "Rough-set-based decision support." In Search methodologies: introductory tutorials in optimization and decision support techniques, pp. 557-609. Boston, MA: Springer US, 2013.
8. Bello, Rafael, and Rafael Falcon. "Rough sets in machine learning: a review." Thriving Rough Sets: 10th Anniversary-Honoring Professor Zdzisław Pawlak's Life and Legacy and 35 Years of Rough Sets (2017): 87-118.
9. Pomeroy, Jennifer. "Enhancing Strategic Decision-Making with Machine Learning in Business Intelligence." (2024).
10. Hadrani, Abdelkhalek, Karim Guennoun, Rachid Saadane, and Mohammed Wahbi. "Rough Set Based Supervised Machine Learning Approaches: Survey and Application." In The Proceedings of the Third International Conference on Smart City Applications, pp. 404-424. Cham: Springer International Publishing, 2019.
11. Chowdhury, Md Maruf Hossan, Seyed Zeinab Aliahmadi, Mohammed A. Quaddus, Amir H. Ansaripoor, and Seyedali Mirjalili. "A decision support framework for resilient and sustainable service design." Global Journal of Flexible Systems Management 26, no. Suppl 1 (2025): 25-55.
12. Ren, Yali. "Optimizing predictive maintenance with machine learning for reliability improvement." ASCE-asme journal of risk and uncertainty in engineering systems, part b: mechanical engineering 7, no. 3 (2021): 030801.
13. Belmouadden, Mustapha, Camélia Dadouchi, and Robert Pellerin. "Real-Time Decision Support System for Dynamic Optimization in Multi-Product Process Manufacturing." IEEE Access no. 1(2025).
14. Elnagar, M. A., Abdel Aty, J., Elhady, A. M., and Shohieb, S. M. (2025). Modeling a sustainable decision support system for banking environments using rough sets. International Journal of Financial Studies, 13(1), 27.
15. Naouali, S., and Othmani, O. (2025). Rough set theory and soft computing methods for building explainable and interpretable AI/ML models. Applied Sciences, 15(9), 5148.
16. Butt, Ateeq Ur Rehman, Hamid Ali, Muhammad Asif, Hessa Alfraihi, Mohamad Khairi Ishak, and Khalid Ammar. "Enhancing Student Management through Hybrid Machine Learning and Rough Set Models: A Framework for Positive Learning Environments." IEEE Access no. 13 (2025).
17. Naouali, Sami, and Oussama El Othmani. "Rough Set Theory and Soft Computing Methods for Building Explainable and Interpretable AI/ML Models." Applied Sciences 15, no. 9 (2025): 5148.
18. Butalia, Ayesha, Rajiv Pandey, and Nidhi Dahale. "Synergistic Integration of Rough Set Theory and Fuzzy Metric Spaces for Enhanced Decision-Making in Uncertain Environments." In 2024 IEEE 4th International Conference on ICT in Business Industry and Government (ICTBIG), pp. 1-10. IEEE, 2024.
19. Al-Mansoori, Noor, and Khaled Al-Fahim. "Optimizing Sustainable Urban Energy Systems Using Fuzzy Rough Sets and Bee Colony Techniques."
20. Xiaoli, Chu, Xu Juan, Chu Xiaodong, Sun Bingzhen, Zhang Yan, Bao Kun, and Li Yanlin. "A nonadditive rough set model for long-term clinical efficacy evaluation of chronic diseases in real-world settings." Artificial Intelligence Review 57, no. 2 (2024): 33.

21. Yu, Rui, Xianghua Kong, and Youlun Li. "Optimizing the diagnostic algorithm for pulmonary embolism in acute COPD exacerbation using fuzzy rough sets and support vector machine." COPD: Journal of Chronic Obstructive Pulmonary Disease 20, no. 1 (2023): 1-8.

22. Adio, Michael O., Ogunmakinde J. Ogunwuyi, Olufikayo A. Adedapo, Jumoke E. Adio, and Moses K. Abidola. "Performance Evaluation of Roughset-Genetic Algorithms (RS-GA) in a Diabetes mellitus Prediction System." Adeleke University Journal of Engineering and Technology 6, no. 2 (2023): 155-162.

23. Song, Simin, Xiaojing Ren, Jing He, Meng Gao, Jia'nan Wang, and Bin Wang. "An optimal hierarchical approach for oral cancer diagnosis using rough set theory and an amended version of the competitive search algorithm." Diagnostics 13, no. 14 (2023): 2454.

24. Surono, Sugiyarto, Zani Anjani Rafsanjani Hsm, Deshinta Arrova Dewi, Annisa Eka Haryati, and Tommy Tanu Wijaya. "Implementation of Takagi Sugeno Kang fuzzy with rough set theory and mini-batch gradient descent uniform regularization." Emerg. Sci. J 7, no. 3 (2023): 791-798.

25. https://archive.ics.uci.edu/

26. Talasila, Vamsidhar, Kotakonda Madhubabu, Meghana Chakravarthy Mahadasyam, Naga Jyothi Atchala, and Lakshmi Sowjanya Kande. "The Prediction of Diseases Using Rough Set Theory with Recurrent Neural Network in Big Data Analytics." International Journal of Intelligent Engineering and Systems 13, no. 5 (2020).

27. Swarupa, A. N. V. K., V. Hema Sree, S. Nookambika, Y. Kiran Sai Kishore, and U. Ravi Teja. "Disease prediction: smart disease prediction system using random forest algorithm." In 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), pp. 48-51. IEEE, 2021.

28. Yang, Li, Haibin Wu, Xiaoqing Jin, Pinpin Zheng, Shiyun Hu, Xiaoling Xu, Wei Yu, and Jing Yan. "Study of cardiovascular disease prediction model based on random forest in eastern China." Scientific reports 10, no. 1 (2020): 5245.

29. Raj, S., R. Vani, Bandi Raja, T. Harsha, Tedlapu Drakshayani, and Ranguri Charith. "HEART DISEASE DETECTION USING XGB-CLASSIFIER AND FAILURE PREDICTION USING GRADIENT BOOSTING." Journal Of Nonlinear Analysis and Optimization 15 (2024).

30. Sekhar, J. Chandra, T. L. Roy, K. Sridharan, K. Aanandha Saravanan, and Ahmed I. Taloba. "Explainable Artificial Intelligence Method for Identifying Cardiovascular Disease with a Combination CNN-XG-Boost Framework." International Journal of Advanced Computer Science and Applications 15, no. 5 (2024).

31. Kaur, Sandeep, and Sheetal Kalra. "Disease prediction using hybrid K-means and support vector machine." In 2016 1st India international conference on information processing (IICIP), pp. 1-6. IEEE, 2016.

32. Ambrish, G., Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, and Kiran Mensinkal. "Logistic regression technique for prediction of cardiovascular disease." Global Transitions Proceedings 3, no. 1 (2022): 127-130.

33. Wang, Bo, and Jing Zhang. "Logistic regression analysis for LncRNA-disease association prediction based on random forest and clinical stage data." IEEE Access 8 (2020): 35004-35017.

34. Rufo, Derara Duba, Taye Girma Debelee, Achim Ibenthal, and Worku Gachena Negera. "Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM)." Diagnostics 11, no. 9 (2021): 1714.

35. Omotehinwa, Temidayo Oluwatosin, David Opeoluwa Oyewola, and Ervin Gubin Moung. "Optimizing the light gradient-boosting machine algorithm for an efficient early detection of coronary heart disease." Informatics and Health 1, no. 2 (2024): 70-81.

Ekta Yadav,Research Scholar,

University School of Basic and Applied Sciences,

Guru Gobind Singh Indraprastha University(GGSIPIU),Delhi,

India.

E-mail address: ektalamba212@gmail.com

and

Rashmi Bhardwaj,

University School of Basic and Applied Sciences,

Guru Gobind Singh Indraprastha University(GGSIPIU),Delhi,

India.

E-mail address: rashmib@ipu.ac.in