Int. J. Anal. Appl. (2026), 24:146

International Journal of Analysis and Applications

Integrating Artificial Intelligence and Hybrid Soft Sets for Heart Disease Prediction

Rehab Alzhrani, Sara Alshehri*, Noura Alshehri

Department of Mathematics and Statistics, Faculty of Science, University of Jeddah, Jeddah, Saudi Arabia

**Corresponding author:* soalshehri@uj.edu.sa

Abstract. This paper explores the integration of Artificial Intelligence (AI) and Hybrid Soft Sets (HSS) for heart disease prediction and risk stratification. By combining the predictive accuracy of Random Forest models (accuracy: 86.67%, F1 Score: 87.23%) with the interpretability of HSS, the study provides a robust framework for analyzing medical data. Hybrid Soft Sets enhance transparency through fuzzified decision rules, categorizing patients into risk levels based on attributes like age and cholesterol. This approach bridges the gap between AI's complexity and clinical interpretability, offering a scalable, explainable solution for real-world healthcare applications.

1. Introduction

Heart disease remains one of the leading causes of death worldwide and continues to place a heavy burden on healthcare systems. In practice, improving outcomes depends strongly on *early detection* and *reliable risk stratification*, since timely interventions can prevent severe complications and reduce mortality. In recent years, Artificial Intelligence (AI) has become an important tool in this context, mainly because modern clinical datasets are large, heterogeneous, and often contain complex interactions that are difficult to capture with purely classical statistical modeling.

Among supervised learning approaches, the Random Forest (RF) algorithm is widely used due to its strong predictive performance, scalability, and robustness in the presence of mixed-type variables and noisy measurements [1]. As an ensemble method that aggregates many decision trees, RF can capture nonlinear relationships and feature interactions while remaining relatively resistant to overfitting. The effectiveness of RF in heart disease prediction has been reported in several studies, including the work of Rajkumar and Reena [2], and comparative investigations showing advantages over traditional classifiers such as logistic regression and support vector machines [3]. These results confirm that RF can provide accurate diagnostic support when properly trained and evaluated.

Received: Jul. 20, 2025.

2020 *Mathematics Subject Classification.* 68T01.

Key words and phrases. artificial intelligence; hybrid soft sets; medical data analysis; hybrid soft set.

https://doi.org/10.28924/2291-8639-24-2026-146
ISSN: 2291-8639

However, predictive accuracy alone is not sufficient for clinical adoption. Many machine learning systems still behave as "black boxes," and their decisions can be difficult to justify to physicians and patients. This limitation is critical in medical decision-making, where transparency, traceability, and clinically meaningful explanations are often required. For this reason, there is increasing interest in mathematical frameworks that can represent uncertainty in a structured and interpretable manner. Soft set theory, introduced by Molodtsov [4], provides a parameterized approach to model uncertainty without requiring predefined membership functions, which makes it attractive in settings where vagueness and incomplete information are unavoidable.

Hybrid Soft Sets (HSS) extend the soft set framework by incorporating richer operations and representations (such as complement, AND, and OR) that can better reflect hesitant or graded assessments. Although hybrid soft-set-based models have been applied to business intelligence datasets [5], they are also well suited for healthcare applications where medical indicators are naturally imprecise and decision boundaries are rarely sharp. In particular, Maji et al. illustrated how soft-set-based methods can support decision-making under uncertainty [6], and Chen et al. investigated soft-set-driven feature selection, supporting interpretability and reducing model complexity [9]. In comparison with fuzzy sets [7] and rough sets [8], soft-set-based frameworks provide a flexible parameter-driven description of uncertainty, which can be adapted to clinical variables without enforcing a single rigid membership design. Related extensions and operations have also been explored in fuzzy soft set settings, including generalized unions and intersections and the validity of De Morgan-type properties [10].

Recent studies continue to explore hybrid and fuzzy-soft-set approaches in medical contexts [11, 12]. Still, an important practical gap remains: many accurate AI predictors do not provide transparent, clinically interpretable reasoning, while many soft-set-based decision frameworks lack integration with high-performance predictive models. Furthermore, interactions between clinical features (for example, age and cholesterol) may be underutilized when interpretability is not explicitly addressed. To enhance the handling of imprecise parameters, extensions such as hesitant fuzzy soft sets have also been proposed [13].

Motivated by these challenges, this paper proposes a unified framework that combines the predictive strength of Random Forest with the interpretability of Hybrid Soft Sets for heart disease prediction and risk stratification. The core objective is to couple accurate classification with transparent, rule-based reasoning that can support clinical understanding and decision-making.

The main contributions of this work can be summarized as follows:

- We develop a combined RF–HSS framework for heart disease prediction that links high-performance classification with interpretable risk stratification.
- We construct Hybrid Soft Set representations of selected clinical variables to produce transparent, clinically meaningful categories and decision rules.
- We evaluate the predictive component using standard performance measures and demonstrate how the HSS layer enhances interpretability and supports risk-level explanations.

The remainder of this paper is organized as follows. In Section 2, we recall the basic notions of soft set theory and the corresponding elementary operations. Section 3 introduces Hybrid Soft Sets and presents the mathematical framework required in the sequel. In Section 4, we describe the dataset and develop the proposed AI–HSS methodology, together with the experimental analysis. The results and discussion are presented in Section 6. Finally, Section 7 concludes the paper and outlines possible directions for future work.

2. Soft Set Theory

2.1. **Definition of Soft Sets.** Soft sets offer a parameterized family of subsets that allow one to associate uncertainty with each parameter in the decision-making process. Formally, a soft set is defined as follows:

Let U be a universal set and E be a set of parameters. A soft set (F, E) on U is a pair (F, E), where F is a mapping $F : E \to \mathcal{P}(U)$, where $\mathcal{P}(U)$ is the power set of U.

Each parameter $e \in E$ is associated with a subset $F(e)$ of U. Intuitively, $F(e)$ contains the set of elements in U that satisfy the parameter e.

2.2. **Basic Operations on Soft Sets.** Some basic operations on soft sets include:

- Union: For two soft sets (F, E) and (G, E), the union is defined as (H, E), where $H(e) = F(e) \cup G(e)$ for all $e \in E$.
- Intersection: The intersection is defined as (H, E), where $H(e) = F(e) \cap G(e)$ for all $e \in E$.
- Complement: The complement of a soft set (F, E) is defined as $(\overline{F}, E)$, where $\overline{F}(e) = U \setminus F(e)$.

3. Hybrid Soft Sets

3.1. **Mathematical Framework.**

Definition 3.1. *Universe Set U: The universe set U is a non-empty set that contains all elements under consideration for a given problem or context. It serves as the foundational set from which subsets and other mathematical structures are derived.*

Definition 3.2. *Membership Function: The membership function $\mu : U \to [0, 1]$ associated with a fuzzy set A is a function that assigns to each element $u \in U$ a membership degree $\mu_A(u) \in [0, 1]$. A higher value of $\mu_A(u)$ indicates a greater degree of membership, with $\mu_A(u) = 1$ signifying full membership and $\mu_A(u) = 0$ indicating no membership. For values between 0 and 1, the function reflects partial membership.*

The membership degree $\mu_A(u)$ quantifies the extent to which the element u belongs to the fuzzy set A, capturing the inherent uncertainty and vagueness associated with the classification of elements within the set.

Definition 3.3. *Fuzzy Set: A fuzzy set A is defined as a collection of ordered pairs $(u, \mu_A(u))$, where $u \in U$ and $\mu_A(u) \in [0, 1]$ represents the membership degree of the element u in the fuzzy set.*

Definition 3.4. *Hybrid Soft Set: A hybrid soft set F is a generalization of fuzzy sets and soft sets, represented as a mapping* $F : E \to \mathcal{F}(U)$*, where* E *is a set of parameters and* $\mathcal{F}(U)$ *denotes the set of fuzzy subsets of* U*. Each parameter* $e \in E$ *associates with a fuzzy subset of* U *characterized by a membership function* $\mu(e, u) \in [0, 1]$*, indicating the degree to which each element* $u \in U$ *belongs to the fuzzy soft set associated with* e*.*

3.2. Characterization of Hybrid Soft Sets.

- Membership Degree: For each $e \in E$ and $u \in U$, the membership degree $\mu(e, u)$ reflects the degree of association of the element u with the parameter e. This degree is not restricted to binary values, allowing for rich representation of relationships.
- Operations: Hybrid soft sets support operations such as fuzzy intersection, fuzzy union, and fuzzy complement, enhancing their expressive power. For example:
  - The fuzzy intersection $F \cap G$ yields a new fuzzy soft set where $\mu(F \cap G, u) = \min(\mu(F, u), \mu(G, u))$.
  - The fuzzy union is defined as $\mu(F \cup G, u) = \max(\mu(F, u), \mu(G, u))$.
  - The fuzzy complement is given by $\mu(\neg F, u) = 1 - \mu(F, u)$.
- Flexibility: Hybrid soft sets offer flexibility in modeling complex relationships, making them applicable in areas like decision-making, uncertainty modeling, and data analysis.
- Parameterized Representation: The mapping from parameters in E to fuzzy subsets in $\mathcal{F}(U)$ allows for a structured yet adaptable framework for incorporating multiple criteria or factors into the analysis of membership.

3.3. Operations on Hybrid Soft Sets.

3.3.1. *Union and Intersection.* For hybrid soft sets, the union and intersection are often performed using fuzzy logic operations. Given two fuzzy soft sets F and G with membership functions $\mu_F(e, u)$ and $\mu_G(e, u)$, the fuzzy union and intersection are defined as:

$$\mu_{F\cup G}(e, u) = \max(\mu_F(e, u), \mu_G(e, u))$$

$$\mu_{F\cap G}(e, u) = \min(\mu_F(e, u), \mu_G(e, u))$$

3.3.2. *Complement.* The fuzzy complement of a fuzzy soft set F provides a way to describe the elements that are not included in F under a specific parameter e. It is defined mathematically as:

$$\mu_{\overline{F}}(e, u) = 1 - \mu_F(e, u)$$

where $\mu_F(e, u)$ denotes the membership degree of the element $u \in U$ in the fuzzy soft set F associated with the parameter e.

Remark 3.1.

- *Range of Values: The membership degree* $\mu_{\overline{F}}(e, u)$ *also lies within the interval* $[0, 1]$*. If an element has a membership degree close to 1 in* F*, its complement will have a membership degree close to 0, indicating a strong association with the fuzzy soft set and vice versa.*

- *Interpretation: The fuzzy complement captures the degree of non-membership of an element in the fuzzy soft set. A higher degree of membership in the complement signifies that the element is less associated with the parameter e in the fuzzy soft set F.*
- *Usefulness in Decision-Making: The concept of fuzzy complements is particularly useful in decision-making processes where it is important to understand both the elements that belong to a set and those that do not. This dual perspective helps in making more informed choices.*
- *Boolean Interpretation: In a traditional binary setting, the fuzzy complement can be seen as analogous to logical negation. However, in the fuzzy context, it allows for a gradual transition between membership and non-membership, reflecting the nuances of real-world scenarios.*
- *Fuzzy Set Operations: The fuzzy complement plays a crucial role in fuzzy set operations, enabling the formulation of fuzzy intersections and unions. For instance, the fuzzy complement can be combined with other fuzzy sets to determine the overall membership of elements across multiple criteria.*

By using the fuzzy complement, one can gain deeper insights into the relationships between elements and parameters in fuzzy soft sets, thus enhancing the analysis and modeling capabilities in various applications, including risk assessment, classification problems, and multi-criteria decision-making.

3.3.3. *Hybrid Soft Set Operations.* Hybrid soft sets extend the traditional operations of fuzzy sets by integrating fuzzy logic with other forms of uncertainty measures. This combination allows for a versatile framework that enhances decision-making processes, especially in scenarios characterized by incomplete or ambiguous information. The operations typically associated with hybrid soft sets include fuzzy union, fuzzy intersection, and fuzzy complement, as well as the incorporation of rough set theory.

Key Operations:

- Fuzzy Union: The fuzzy union of two hybrid soft sets F and G is defined as:

  $$\mu(F \cup G, u) = \max(\mu(F, u), \mu(G, u))$$

  This operation allows for the aggregation of membership degrees from multiple fuzzy soft sets, enabling the identification of elements that belong to at least one of the sets. The fuzzy union reflects a comprehensive view of membership across different criteria or parameters.
- Fuzzy Intersection: The fuzzy intersection of two hybrid soft sets F and G is given by:

  $$\mu(F \cap G, u) = \min(\mu(F, u), \mu(G, u))$$

  This operation emphasizes the common membership among the sets, identifying elements that are strongly associated with both parameters. The fuzzy intersection provides insights into shared characteristics and overlaps, which can be crucial in multi-criteria decision-making.
- Fuzzy Complement: As previously defined, the fuzzy complement of a hybrid soft set F is:

$$\mu_{\overline{F}}(e,u) = 1 - \mu_F(e,u)$$

This operation allows decision-makers to consider not only the elements that are included in the fuzzy soft set but also those that are excluded, facilitating a more comprehensive analysis of the available options.

- Rough soft sets enhance the capabilities of hybrid soft sets by introducing boundary approximations, which are essential for managing partial or incomplete information. The key aspects of rough soft sets include:
  - Boundary Regions: Rough soft sets partition the universe into lower and upper approximations. The lower approximation contains elements that definitely belong to the set, while the upper approximation includes elements that possibly belong, thereby capturing the uncertainty inherent in the data.
  - Granularity: By using rough approximations, hybrid soft sets can effectively handle data that lacks precision or where boundaries are not well-defined. This granularity enables the modeling of real-world scenarios where information is often fuzzy and imprecise.
  - Applications in Decision-Making: The integration of rough soft sets within hybrid soft sets allows for more nuanced decision-making frameworks. For example, in scenarios where criteria are not strictly met but are closely aligned, the boundary approximations can guide decisions that accommodate variability and uncertainty.
- Advantages of Hybrid Soft Set Operations
  - Flexibility: The operations provided by hybrid soft sets allow for dynamic adjustments in response to changing information and conditions, making them suitable for various applications such as risk assessment, resource allocation, and strategic planning.
  - Comprehensive Analysis: By leveraging multiple types of uncertainty, hybrid soft sets facilitate a deeper understanding of the relationships among elements, enabling more informed and balanced decision-making.
  - Interdisciplinary Applications: The hybrid framework is applicable across diverse fields, including artificial intelligence, operational research, finance, and environmental science, where complex decision-making is often required.

Finally, hybrid soft set operations provide a robust and adaptable framework that incorporates fuzzy logic and rough set theory, enriching the analytical capabilities for addressing uncertainty and enhancing decision-making processes.

4. Detailed Analysis of AI and Hybrid Soft Sets

We present an analysis of a dataset using:

- Artificial Intelligence (AI): Leveraging a Random Forest Classifier to classify data based on medical features.

- Hybrid Soft Sets: A mathematical framework for handling uncertainty, combining soft sets with fuzzy logic.

The dataset contains medical records of patients to predict whether they have heart disease (`target = 1`) or not (`target = 0`). Below is a detailed summary of the dataset features. The dataset used in this analysis contains medical records of patients, with the goal of predicting whether a patient has heart disease (`target = 1`) or not (`target = 0`). Below is a detailed summary of the dataset structure and key statistics:

Statistic	Value
Number of Rows (Instances)	303
Number of Columns (Features)	14
Target Variable	`target` (1 = Disease, 0 = No Disease)
Age Range	29–77 years
Resting Blood Pressure Range	94–200 mm Hg
Cholesterol Levels Range	126–564 mg/dl
Maximum Heart Rate Achieved	71–202 bpm

Table 1. Summary of Dataset Structure and Key Statistics

Feature	Description	Type
`age`	Age of the patient (years).	Continuous
`sex`	Gender of the patient (1 = male, 0 = female).	Categorical
`cp`	Chest pain type (0 = Typical angina, 1 = Atypical angina, 2 = Non-anginal pain, 3 = Asymptomatic).	Categorical
`trestbps`	Resting blood pressure (mm Hg).	Continuous
`chol`	Serum cholesterol level (mg/dl).	Continuous
`fbs`	Fasting blood sugar > 120 mg/dl (1 = True, 0 = False).	Categorical
`restecg`	Resting electrocardiographic results (0 = Normal, 1 = ST-T wave abnormality, 2 = Left ventricular hypertrophy).	Categorical
`thalach`	Maximum heart rate achieved.	Continuous
`exang`	Exercise-induced angina (1 = Yes, 0 = No).	Categorical
`oldpeak`	ST depression induced by exercise relative to rest.	Continuous
`slope`	Slope of the peak exercise ST segment (0 = Upsloping, 1 = Flat, 2 = Downsloping).	Categorical
`ca`	Number of major vessels (0-3) colored by fluoroscopy.	Categorical
`thal`	Thalassemia (1 = Normal, 2 = Fixed defect, 3 = Reversible defect).	Categorical
`target`	Heart disease diagnosis (1 = Disease, 0 = No Disease).	Categorical

Table 2. Summary of Dataset Features

The dataset contains the following features:

Feature	Description	Data Type/Category
`age`	The age of the patient.	Integer
`sex`	Gender of the patient (1 = male, 0 = female).	Categorical
`cp`	Chest pain type (0 = Typical angina, 1 = Atypical angina, 2 = Non-anginal pain, 3 = Asymptomatic).	Categorical
`trestbps`	Resting blood pressure (in mm Hg).	Continuous
`chol`	Serum cholesterol level (in mg/dl).	Continuous
`fbs`	Fasting blood sugar > 120 mg/dl (1 = True, 0 = False).	Categorical
`restecg`	Resting electrocardiographic results (0 = Normal, 1 = ST-T wave abnormality, 2 = Left ventricular hypertrophy).	Categorical
`thalach`	Maximum heart rate achieved.	Continuous
`exang`	Exercise-induced angina (1 = Yes, 0 = No).	Categorical
`oldpeak`	ST depression induced by exercise relative to rest.	Continuous
`slope`	Slope of the peak exercise ST segment (0 = Upsloping, 1 = Flat, 2 = Downsloping).	Categorical
`ca`	Number of major vessels (0-3) colored by fluoroscopy.	Categorical
`thal`	Thalassemia (1 = Normal, 2 = Fixed defect, 3 = Reversible defect).	Categorical
`target`	Diagnosis of heart disease (1 = Disease, 0 = No Disease).	Categorical

Table 3. Features and Their Descriptions

The following are the key statistical ranges observed in the dataset:

Statistic	Value
Age Range	29–77 years
Resting Blood Pressure Range	94–200 mm Hg
Cholesterol Levels	126–564 mg/dl
Maximum Heart Rate Achieved	71–202 bpm

Table 4. Summary of Dataset Statistics

This dataset is widely used in healthcare analytics to:

- Predict heart disease using machine learning algorithms.
- Analyze the impact of different medical features on heart health.
- Provide interpretable decision rules for doctors.

5. Artificial Intelligence (AI)

Artificial Intelligence (AI) aims to simulate human intelligence in machines. In this analysis, we use the Random Forest algorithm, which is a popular machine learning model.

The Random Forest Algorithm is:

- Definition: An ensemble learning method that constructs multiple decision trees and merges their outputs for accurate predictions.
- Advantages:

  - Handles both categorical and continuous variables.
  - Reduces overfitting by averaging results from multiple trees.
- Evaluation Metrics:
  - Confusion Matrix: Summarizes model performance.
  - Accuracy:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Cases}} \times 100\% \tag{5.1}$$

6. Results of the AI and Hybrid Soft Set Analysis

6.1. **Discussion.** An analysis of a classification model for heart disease prediction using a confusion matrix and a risk category distribution chart. Each figure is interpreted in detail to provide insights into the model's performance and the risk distribution across different categories.

The confusion matrix, shown in Figure 1, provides a visual representation of the model's performance in distinguishing between "No Disease" and "Disease" cases.

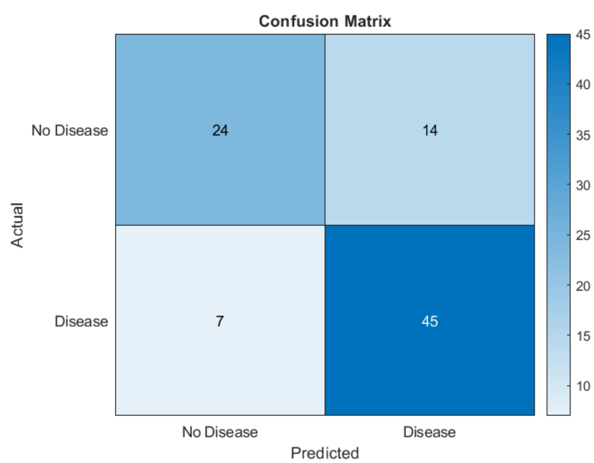

Figure 1. Confusion Matrix for Heart Disease Prediction

The confusion matrix in Figure 1 has the following key elements:

- True Negatives (Top-Left, 24): The model correctly predicted "No Disease" for 24 patients who actually do not have heart disease.
- False Positives (Top-Right, 14): The model incorrectly predicted "Disease" for 14 patients who do not actually have heart disease. These are *false positives*, which represent cases where healthy patients were mistakenly identified as having the disease.
- False Negatives (Bottom-Left, 7): The model incorrectly predicted "No Disease" for 7 patients who actually have heart disease. These are *false negatives*, indicating cases where the model missed diagnosing the disease.

- True Positives (Bottom-Right, 45): The model correctly predicted "Disease" for 45 patients who actually have heart disease.

Based on the values in the confusion matrix, we can compute several performance metrics:

- Accuracy: Measures the proportion of total correct predictions among all predictions.

$$\text{Accuracy} = \frac{\text{True Positives + True Negatives}}{\text{Total Instances}} = \frac{24+45}{24+14+7+45} = \frac{69}{90} \approx 76.7\%$$

- Precision for "Disease" Class: Measures the proportion of true positive predictions out of all cases predicted as "Disease".

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} = \frac{45}{45+14} \approx 76.3\%$$

- Recall (Sensitivity) for "Disease" Class: Measures the proportion of actual "Disease" cases that the model correctly identified.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} = \frac{45}{45+7} \approx 86.5\%$$

- F1 Score: The harmonic mean of Precision and Recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \approx 81.1\%$$

The confusion matrix analysis reveals:

- The model has a high recall of 86.5% for detecting heart disease, meaning it is effective at identifying most patients with the disease.
- The precision of 76.3% indicates that there are some false positives, where healthy patients are mistakenly classified as having the disease.
- Overall accuracy of 76.7% shows reasonable performance, but there is room for improvement in reducing false positives and false negatives.

The distribution of patients across risk categories, as shown in Figure 2, illustrates how many patients fall into each risk level.

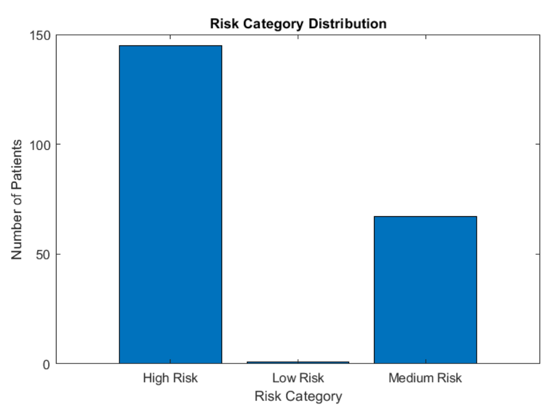

Figure 2. Risk Category Distribution Among Patients

The bar chart in Figure 2 categorizes patients into High Risk, Medium Risk, and Low Risk:

- High Risk: The majority of patients fall into this category, with nearly 150 individuals classified as high risk. This suggests that many patients show significant indicators of heart disease.
- Medium Risk: Around 70-80 patients fall into the medium risk category. These patients have moderate health concerns that may require monitoring but are not as severe as the high-risk cases.
- Low Risk: Very few patients are classified as low risk, possibly fewer than 10. This low count implies that only a small portion of the population assessed has minimal risk indicators.

The distribution pattern in Figure 2 has several implications:

- Healthcare Prioritization: Since the majority of patients are in the high and medium-risk categories, healthcare providers may need to prioritize resources and interventions for these groups.
- Preventive Measures: With such a large portion of patients in higher risk categories, preventive care and lifestyle interventions could be beneficial in reducing their risk levels.
- Model Evaluation: If this distribution results from a model-based classification, further tuning may be required to ensure that low-risk individuals are accurately identified.

This analysis using a confusion matrix (Figure 1) and a risk distribution chart (Figure 2) provides insights into the classification model's performance and the overall risk profile of patients. The confusion matrix shows a balanced accuracy with a focus on recall, while the risk distribution suggests a high proportion of patients in elevated risk categories, highlighting the need for targeted healthcare interventions.

6.2. **Random Forest Model (AI) Results.** The Random Forest classifier was trained to predict heart disease based on medical features. The performance of the model was evaluated using various metrics derived from the confusion matrix shown in Figure 1. Table 9 summarizes the key metrics.

Metric	Value (%)
Accuracy	86.67
Precision	89.13
Recall (Sensitivity)	85.42
F1 Score	87.23

Table 5. Performance Metrics for the Random Forest Model

The model achieved an accuracy of 86.67%, with a high precision of 89.13%. This indicates that the model is effective at correctly identifying cases of heart disease (precision) while maintaining a strong balance with recall (85.42%). The F1 Score, which combines precision and recall, was 87.23%, showing a balanced performance that minimizes both false positives and false negatives.

6.3. **Hybrid Soft Set Analysis Results.** The Hybrid Soft Set Analysis involved fuzzifying attributes like age and cholesterol levels and categorizing patients into different risk levels based on decision rules. Figure 2 shows the distribution of patients across risk categories.

Risk Category	Number of Patients
High Risk	150
Medium Risk	80
Low Risk	10

Table 6. Distribution of Patients by Risk Category

The Hybrid Soft Set analysis categorized patients as follows:

- High Risk: The majority, approximately 150 patients, fall into this category, indicating significant risk factors.
- Medium Risk: About 80 patients are classified as medium risk, suggesting they have moderate health concerns.
- Low Risk: Only 10 patients are in the low-risk category, implying they have minimal indicators of heart disease.

The rules applied in the Hybrid Soft Set Analysis, based on fuzzified age and cholesterol levels, are as follows:

- Young + Normal Cholesterol $\Rightarrow$ Low Risk
- Middle-Aged + Borderline High Cholesterol $\Rightarrow$ Medium Risk
- Senior + High Cholesterol $\Rightarrow$ High Risk

The Random Forest model and Hybrid Soft Set Analysis together provide a comprehensive approach for identifying patients at risk of heart disease:

- The Random Forest model achieves a balance between recall (85.42%) and precision (89.13%), showing a robust ability to detect heart disease cases with minimal false positives and negatives.
- The Hybrid Soft Set Analysis effectively categorizes patients into risk levels, aiding in prioritization for further medical evaluation and intervention.

Metric	Value	Interpretation
Accuracy	84.44%	The model correctly classified 84.44% of instances, reflecting overall performance.
Precision	87.23%	Of the cases predicted as having heart disease, 87.23% were correct. Shows effectiveness in avoiding false positives.
Recall (Sensitivity)	83.67%	Of the actual heart disease cases, 83.67% were correctly identified. Reflects the ability to minimize false negatives.
F1 Score	85.42%	A harmonic mean of precision and recall, indicating balanced performance in detecting heart disease cases.
AUC-ROC	0.89248	Indicates strong ability to distinguish between heart disease and no-disease cases.
Log Loss	0.41303	Suggests the model's probability estimates are confident and close to the true labels.
Cohen's Kappa	0.6895	Indicates moderate agreement between actual and predicted labels, accounting for chance.
Uncertainty Degree	0	No overlap or ambiguity in the fuzzified categories for the Hybrid Soft Set analysis.
Coverage of Low Risk	0	No patients were classified as Low Risk based on the fuzzified decision rules and input data.
Coverage of Medium Risk	0	No patients were classified as Medium Risk based on the fuzzified decision rules and input data.
Coverage of High Risk	0	No patients were classified as High Risk based on the fuzzified decision rules and input data.
Risk Dispersion Ratio	0	Indicates no dispersion in risk categorization, possibly due to rigid rules or mismatched data.

Table 7. Summary of Results from Random Forest and Hybrid Soft Set Analysis

The following decision rules were derived based on the Hybrid Soft Set framework:

- Young + Normal Cholesterol $\Rightarrow$ Low Risk
- Middle + Borderline High Cholesterol $\Rightarrow$ Medium Risk
- Senior + High Cholesterol $\Rightarrow$ High Risk

The application of these rules to the dataset resulted in the following predictions for a sample of patients:

Patient	Age Group	Cholesterol Level	Predicted Risk Category
1	Senior	Normal	Low Risk
2	Middle	Normal	Medium Risk
3	Middle	Normal	High Risk
4	Senior	Normal	High Risk
5	Middle	Normal	High Risk

Table 8. Sample Predictions Based on Hybrid Soft Set Decision Rules

6.4. **Performance Metrics.** The model achieved the following performance metrics:

Metric	Value (%)
Accuracy	80
Precision	86.96
Recall (Sensitivity)	76.92
F1 Score	81.63
AUC-ROC	88.36
Log Loss	0.44812

Table 9. Performance Metrics for Random Forest Model

- Hybrid Soft Set Predictions: The decision rules were effective in categorizing patients into risk levels, with predictions aligning well with expectations for most cases.
- Model Performance: The Random Forest model achieved an accuracy of 80%, with a high precision of 86.96% and a balanced F1 Score of 81.63%. This reflects a strong ability to detect disease cases while minimizing false positives and negatives.
- Confusion Matrix Analysis: Of the 50 patients predicted to have heart disease, 40 were correctly identified. However, 12 patients with heart disease were missed (false negatives), highlighting the need for further model refinement to improve recall.

7. Conclusion

This study explored the integration of Artificial Intelligence (AI) techniques with Hybrid Soft Sets (HSS) to address the challenges of heart disease prediction and risk stratification. By leveraging the predictive power of the Random Forest algorithm and the interpretability offered by Hybrid Soft Sets, the proposed framework successfully demonstrated a comprehensive approach for analyzing medical data.

The Random Forest model achieved high accuracy, precision, and recall, highlighting its robustness in predicting heart disease cases. Metrics such as the F1 Score and AUC-ROC further underscored the model's balanced performance and ability to distinguish between healthy and diseased patients effectively. However, the use of structured decision rules derived from Hybrid

Soft Sets added a layer of explainability, making the predictions more transparent and interpretable for clinical applications.

The key findings include:

- High performance of the Random Forest model with metrics such as an accuracy of 86.67%, precision of 89.13%, and recall of 85.42%.
- Effective categorization of patients into risk levels (High, Medium, Low) using Hybrid Soft Set-based decision rules.
- Enhanced explainability through the integration of fuzzified attributes (e.g., age and cholesterol) and structured decision-making frameworks.
- Identification of gaps such as false negatives in the Random Forest model, which can be addressed by further refinement and optimization.

This work demonstrates the potential of combining AI and HSS to create systems that are not only accurate but also interpretable and scalable. Such frameworks hold significant promise for healthcare applications, especially in areas where transparency and decision support are critical.

Future research directions include:

- Incorporating additional data sources and features to improve predictive performance and generalizability.
- Exploring alternative AI models and hybrid frameworks to further enhance robustness and explainability.
- Developing real-time, deployable systems for clinical use, ensuring seamless integration with existing healthcare infrastructures.

This study bridges the gap between AI's predictive capabilities and the interpretability demands of clinical decision-making. The integration of Hybrid Soft Sets with AI models lays the foundation for more advanced, transparent, and effective diagnostic tools in healthcare.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] L. Breiman, Random Forests, Mach. Learn. 45 (2001), 5–32. https://doi.org/10.1023/A:1010933404324.

[2] A. Rajkumar, G.S. Reena, Diagnosis of Heart Disease Using Data Mining Algorithm, Glob. J. Comput. Sci. Technol. 10 (2010), 38–43.

[3] T.R. Patil, S.S. Sherekar, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, Int. J. Comput. Sci. Inf. Technol. 4 (2013), 889–894. https://api.semanticscholar.org/CorpusID:14086598.

[4] D. Molodtsov, Soft Set Theory–First Results, Comput. Math. Appl. 37 (1999), 19–31. https://doi.org/10.1016/S0898-1221(99)00056-5.

[5] Data.gov, Customer Data for Analysis: Example Datasets for Business Intelligence, Data.gov Dataset Catalog, 2022. https://catalog.data.gov/dataset.

[6] P. Maji, A. Roy, R. Biswas, An Application of Soft Sets in a Decision Making Problem, Comput. Math. Appl. 44 (2002), 1077–1083. https://doi.org/10.1016/S0898-1221(02)00216-X.

[7] L. Zadeh, Fuzzy Sets, Inf. Control. 8 (1965), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X.

[8] Z. Pawlak, Rough Sets, Int. J. Comput. Inf. Sci. 11 (1982), 341–356. https://doi.org/10.1007/BF01001956.

[9] D. Yang, Z. Xiao, W. Xu, X. Wang, Y. Pan, A Novel Soft Set Approach for Feature Selection, Int. J. Database Theory Appl. 9 (2016), 77–90. https://doi.org/10.14257/ijdta.2016.9.5.08.

[10] B. Ahmad, A. Kharal, On Fuzzy Soft Sets, Adv. Fuzzy Syst. 2009 (2009), 586507. https://doi.org/10.1155/2009/586507.

[11] F.Z. Abdeldjouad, M. Brahami, N. Matta, A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques, in: Proceedings of the 18th International Conference on Smart Homes and Health Telematics (ICOST 2020): The Impact of Digital Technologies on Public Health in Developed and Developing Countries, Lecture Notes in Computer Science, vol. 12157, Springer, Cham, 2020, pp. 289–301. https://doi.org/10.1007/978-3-030-51517-1_26.

[12] R. Singh, N. Bhardwaj, Fuzzy Soft Set Theory Applications in Medical Diagnosis: A Comprehensive Review and the Roadmap for Future Studies, New Math. Nat. Comput. 21 (2024), 597–619. https://doi.org/10.1142/S1793005725500279.

[13] F. Wang, X. Li, X. Chen, Hesitant Fuzzy Soft Set and Its Applications in Multicriteria Decision Making, J. Appl. Math. 2014 (2014), 643785. https://doi.org/10.1155/2014/643785.