A Comparative Approach to Heart Disease Classification*

Patrycja Janik, Patryk Weklicz and Kacper Śliwka

Faculty of Applied Mathematics, Silesian University of Technology, Kaszubska 23, 44100 Gliwice, Poland

Abstract

This paper presents a comparative analysis of three distinct approaches for heart disease diagnosis. The study leverages patient data characterized by various features to predict the presence of heart disease. To ensure robust performance evaluation, a k-fold cross-validation technique is implemented. The efficacy of each method is assessed using key metrics such as accuracy, sensitivity, precision, and F1-score. The results of the k-fold cross-validation across different values of k (from 5 to 9), along with confusion matrices for each method in each fold, provide a comprehensive evaluation of their diagnostic capabilities. The exploratory data visualizations offer valuable insights into the dataset's characteristics, contributing to a deeper understanding of the factors influencing heart disease.

Keywords

Heart Disease Classification, Soft Set Methods, Naive Bayes Classifier, Cross-validation, Feature Weighting, Medical Diagnostics, Data Visualization

1. Introduction

The imperative for robust and accurate heart disease diagnosis cannot be overstated in an era where cardiovascular ailments remain a leading cause of global mortality. The intricate interplay of genetic predispositions, environmental factors, and lifestyle choices contributes to the pervasive nature of these conditions, necessitating continuous advancements in our ability to identify and manage them effectively. Traditional diagnostic approaches, while foundational, often present limitations in terms of cost, invasiveness, and the time required for comprehensive evaluation. Consequently, the exploration of novel, data-driven methodologies has gained significant momentum within the medical community. Machine learning, with its capacity to discern complex patterns from high-dimensional datasets, offers a promising avenue for augmenting or even transforming the landscape of heart disease diagnosis. By leveraging the wealth of information embedded in patient records, these computational techniques hold the potential to facilitate earlier detection, personalize treatment strategies, and ultimately improve patient outcomes. This burgeoning field of research seeks to bridge the gap between the complexities of cardiac pathophysiology and the analytical power of algorithmic intelligence.

Within the broader domain of machine learning applied to medical diagnostics, a diverse array of classification algorithms has been investigated for heart disease prediction. These range from well-established statistical methods like Logistic Regression and Discriminant Analysis to more intricate approaches such as Support Vector Machines, Neural Networks, and Decision Tree ensembles. Each of these existing solutions brings its own set of strengths and weaknesses concerning interpretability, computational efficiency, and predictive accuracy. However, the quest for optimal diagnostic tools continues, driving the exploration of alternative paradigms. Soft set theory, a mathematical framework designed to handle uncertainty and vagueness, presents an intriguing approach that has seen limited application in this specific medical context[1]. Similarly, probabilistic classifiers, such as Naive Bayes, offer a computationally efficient and often surprisingly effective means of modeling diagnostic probabilities[2]. There are also many other models with possible applications in data analytics. In [3] was presented a model of classification atoms in molecules from large inputs , while in [4] was presented XAI model which implements breast cancer detection by using definable machine learning approaches. There are also models using rules, which lead from the input data configuration to the

*IVUS 2025: Information Society and University Studies 2025, May 15, Kaunas, Lithuania

pj309665@student.polsl.pl (P. Janik); pw311122@student.polsl.pl (P. Weklicz); ks311099@student.polsl.pl (K. Śliwka)

decision process by completing a reasoning. In [5] such idea was used to solve real-time diagnostics, while in [6] was proposed moisture control system.

This paper presents a comparative analysis of three distinct approaches for heart disease diagnosis: a non-weighted soft set method [7], a weighted soft set method [8], and a Naive Bayes classifier [9]. The study leverages patient data characterized by various features to predict the presence of heart disease. To ensure robust performance evaluation, a k-fold cross-validation technique is implemented [10]. The efficacy of each method is assessed using key metrics such as accuracy, sensitivity, precision, and F1-score [11]. Furthermore, the research incorporates data visualization techniques, employing seaborn and matplotlib libraries [12, 13] to illustrate fundamental relationships within the dataset, including the distribution of heart disease across age groups, gender-based prevalence, and the association between exercise-induced angina and diagnosis[14]. The results of the k-fold cross-validation across different values of k (from 5 to 9), along with confusion matrices for each method in each fold, provide a comprehensive evaluation of their diagnostic capabilities. The exploratory data visualizations offer valuable insights into the dataset's characteristics, contributing to a deeper understanding of the factors influencing heart disease. This paper contributes to this ongoing exploration by providing a direct comparative analysis of these less frequently juxtaposed methodologies – soft set approaches (both weighted and non-weighted) and Naive Bayes – against the backdrop of heart disease classification, aiming to offer novel insights into their potential and limitations relative to more established techniques.

2. Methodology

In our project three approaches are considered: the Naive Bayes classifier, the weighted soft classifier, and its simplified unweighted version. The Naive Bayes classifier applies Bayes' theorem with the assumption of feature independence, making it a simple and efficient method widely used in medical diagnosis. Soft classifiers are based on soft set theory, which allows modeling uncertainty in data. In the weighted version, feature importance is determined using logistic regression, while the unweighted version treats all features equally. To evaluate model performance, k-fold cross-validation is used, providing a reliable estimate of model accuracy. Evaluation metrics include accuracy, precision, recall, and the F1-score, offering a comprehensive view of each classifier's effectiveness in a medical context.

2.1. Naive Bayes Classifier

Naive Bayes classifier is based on Bayes' theorem and assumes that features are independent within each class. This approach has been widely used in medical diagnosis tasks, including heart disease detection [9].

The formula for calculating the probability of sample x belonging to class C_k is the following:

$$P(C_k|x) = \frac{P(x|C_k) \cdot P(C_k)}{P(x)} \tag{1}$$

Where:

- $P(C_k|x)$ is probability of belonging to class C_k with given features x,
- $P(x|C_k)$ is probability of observing features x in class C_k,
- $P(C_k)$ is the prior probability of class C_k,
- $P(x)$ is the probability of observing the feature x in general.

Then, assuming a normal distribution for features, the probability for one feature x_i in class C_k is described by the formula:

$$P(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{2}$$

Where μ_k is an arithmetic mean and σ_k is the standard deviation of feature x_i in class C_k[15].

The prediction class for sample x is calculated as:

$$\hat{C} = \arg\max_{C_k} P(C_k|x) \tag{3}$$

Algorithm 1: Naive Bayes Classifier Algorithm.

Data: Input data: feature set x, from classes $C_1, C_2, \ldots, C_k$
Result: Class C_{pred}
foreach *class* C_k **do**
    Calculate $P(C_k)$ and $P(x|C_k)$;
    Calculate $P(C_k|x)$ using the Bayes' theorem (1);
return $C_{\text{pred}} = \arg\max_{C_k} P(C_k|x)$

2.2. Soft Set

In a soft classifier, data features are treated as elements of soft sets, where each element is assigned a degree of membership in a given set. These values are used to calculate the total value of the sample's membership in a class. According to Molodtsov's first results on soft set theory [16], soft sets provide a general framework for handling uncertainty, making them suitable for applications such as medical diagnosis. The sum of the feature values in the soft set is calculated as:

The sum of the feature values in the soft set is calculated as:

$$\text{sum} = \sum_{i=1}^{n} w_i \cdot x_i, \tag{4}$$

where w_i is the weight assigned to the feature x_i.

The weights in our project are not chosen randomly. They are extracted by using logistic regression[17]:

$$P(\text{heart disease}) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 \cdot \text{age} + \theta_2 \cdot \text{blood pressure} + \ldots)}}, \tag{5}$$

where θ are weights – we are looking them using logistic regression.

The model "adjusts" these weights to predict the disease as best as possible.

Algorithm 2: Soft classifier algorithm with weights.

Data: Input data: feature set $x = (x_1, x_2, \ldots, x_n)$, weights $w = (w_1, w_2, \ldots, w_n)$
Result: Class C_{pred}
$sum := 0$;
foreach *feature* x_i **do**
    Calculate $w_i \cdot x_i$;
    $sum := sum + w_i \cdot x_i$;
return $C_{\text{pred}} = \arg\max_{C_k} sum$

2.3. Soft Set without weights

In the unweighted soft set classifier, each feature in the dataset is considered to have equal influence on the classification outcome [18]. This approach represents a simpler version of the soft set methodology where no prior knowledge or learned importance is incorporated into the model.

For a sample with features $x = (x_1, x_2, \ldots, x_n)$, the classification process involves calculating the sum of all feature values without any differential weighting:

$$\text{sum} = \sum_{i=1}^{n} x_i. \tag{6}$$

The classification decision is then made by selecting the class that maximizes this sum:

$$C_{\text{pred}} = \arg\max_{C_k} \text{sum}. \tag{7}$$

Algorithm 3: Soft classifier algorithm without weights

Data: Input data: feature set $x = (x_1, x_2, \ldots, x_n)$
Result: Class C_{pred}

```
sum := 0;
foreach feature x_i do
    sum := sum + x_i;
return C_pred = arg max_{C_k} sum
```

Soft Set in Decision-Making Soft sets are widely used in decision-making problems, particularly when dealing with uncertainty in data. Recent work by Okigbo et al. [8] highlights the application of soft set theory in decision-making, specifically utilizing the soft 'AND-OPERATION' approach to handle complex decision criteria. This methodology aids in making more accurate predictions and decisions in scenarios where traditional methods might struggle with uncertain or incomplete information.

2.4. Cross-validation

To evaluate the performance of classification models, the k-fold cross-validation method is widely employed [19]. In this technique, the dataset is partitioned into k equally sized subsets, known as folds. The model is trained on $k-1$ of these folds and validated on the remaining one. This process is repeated k times, with each fold serving as the validation set exactly once. The final performance metrics are obtained by averaging the results across all k iterations in similar way that Kaushika Pal and Biraj. V. Patel did in "Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques" research work [20].

Within this evaluation framework, standard classification metrics such as accuracy, precision, recall, and the $F1$-score are calculated. The corresponding formulas are provided below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

Algorithm 4: Cross-validation algorithm.

Data: Input data: feature set D, amount of folds k
Result: Average model accuracy

```
accuracy := 0;
for fold i = 1 to k do
    Training set D_train = D \ D_{fold_i};
    Test set D_test = D_{fold_i};
    Train the model on D_train;
    Calculate accuracy on D_test;
    accuracy := accuracy + accuracy on fold i;
return accuracy/k
```

The methodological framework presented in this study encompasses three distinct classification approaches and a comprehensive evaluation strategy. The Naive Bayes classifier leverages conditional probability distributions with the assumption of feature independence, making it computationally efficient yet theoretically constrained. In contrast, soft set methods—both unweighted and weighted—offer an alternative mathematical paradigm that treats features as elements of decision parameters, with the weighted variant incorporating learned feature importance through logistic regression coefficients. The k-fold cross-validation technique, as detailed in Section 2.3, provides a robust mechanism for evaluating these classifiers by partitioning data into multiple training and testing subsets, thereby reducing bias and variance in performance assessment. The complementary metrics of accuracy, precision, recall, and F1-score offer a multi-dimensional view of classifier performance, particularly important in medical contexts where both false positives and false negatives carry significant consequences. This comprehensive methodological approach enables not only direct comparison between classification methods but also insights into the stability and consistency of their performance across different data configurations.

3. Experiments

3.1. Dataset Description

The dataset used in this study is the well-known **Heart Disease Dataset** available on the Kaggle platform, originally sourced from the Cleveland Clinic Foundation. It consists of 303 patient records and includes 14 clinical features, combining both categorical and continuous variables. These features describe attributes such as age, sex, type of chest pain, resting blood pressure, cholesterol level, fasting blood sugar, electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the ST segment, number of major vessels detected via fluoroscopy, and thalassemia. The target variable is binary, where 0 indicates absence and 1 indicates presence of heart disease.

3.2. Dataset analysis

This dataset provides a collection of relevant medical information pertaining to patients, with the primary goal of predicting the presence or absence of heart disease. It encompasses various clinical measurements and patient characteristics, enabling the analysis of factors contributing to cardiac health. The dataset is structured into rows, each representing an individual patient, and columns, each detailing a specific medical attribute like in others[21].

- **age:** Patient's age (numeric). Approximately normally distributed, slightly right-skewed. Weak positive correlation (0.16) with *target*, moderate negative (-0.37) with *max heart rate*.
- **sex:** Gender (categorical, binary: 0-female, 1-male). Imbalanced distribution (predominance of males). Weak positive correlation (0.11) with *target*.
- **chest pain type:** Type of chest pain (categorical, nominal: 1-typical angina, 2-atypical angina, 3-non-anginal pain, 4-asymptomatic). Four discrete values. Moderate positive correlation (0.37) with *target*.
- **resting bps:** Resting blood pressure (numeric). Approximately normal distribution, concentrated in the range of 120–140 mmHg. Weak positive correlation (0.08) with *target*.
- **cholesterol:** Cholesterol level (numeric). Right-skewed distribution. Weak positive correlation (0.06) with *target*.
- **fasting blood sugar:** Fasting blood sugar level (categorical, binary: 0-$\leq$120 mg/dl, 1->120 mg/dl). Imbalanced distribution (predominance of $\leq$120 mg/dl). Weak positive correlation (0.11) with *target*.
- **resting ecg:** Resting electrocardiographic results (categorical, nominal: 0-normal, 1-ST-T abnormality, 2-left ventricular hypertrophy). Few discrete values, predominance of value 0. Weak positive correlation (0.11) with *target*.

- **max heart rate:** Maximum achieved heart rate (numeric). Left-skewed distribution. Weak negative correlation (-0.15) with *target*, negative (-0.37) with *age*.
- **exercise angina:** Exercise-induced angina (categorical, binary: 0-no, 1-yes). Imbalanced distribution (predominance of no angina). Moderate positive correlation (0.27) with *target*, strong positive (0.41) with *oldpeak*.
- **oldpeak:** ST segment depression induced by exercise (numeric, continuous). Strongly right-skewed distribution. Moderate positive correlation (0.22) with *target*, strong positive (0.41) with *exercise angina*.
- **ST slope:** Slope of the ST segment (categorical, ordinal: 0-upsloping, 1-flat, 2-downsloping). Few discrete values. Strong positive correlation (0.50) with *target*.
- **target:** Presence of heart disease (categorical, binary: 0-absent, 1-present). Relatively balanced distribution.

Summary:

The analysis of correlations and distributions indicates the importance of *ST slope*, *chest pain type*, *oldpeak*, and *exercise angina* in predicting heart disease. Class imbalance (*sex*, *exercise angina*), asymmetry (cholesterol, *oldpeak*), and dominance of certain values (ECG, *ST slope*) should be taken into account during modeling.

3.3. Data Preprocessing

Prior to model training, several preprocessing [22] steps were applied to ensure data quality and model compatibility:

- **Missing Values:** No missing values were present in the dataset, allowing the models to use the raw data without imputation.
- **Categorical Encoding:** Categorical variables were one-hot encoded to be compatible with Naive Bayes and logistic regression models. For soft set approaches, categorical values were converted into numerical scales.
- **Feature Scaling:** Continuous variables were normalized using min-max scaling to standardize feature ranges, especially important for soft set classification.
- **Weight Extraction:** For the weighted soft set classifier, feature importances were extracted from a logistic regression model. The resulting coefficients were used as weights to emphasize more influential features as in the more complex models[23].

3.4. Experimental Design

To evaluate model performance reliably, **k-fold cross-validation** was used with values of $k \in \{5, 6, 7, 8, 9\}$ for all three classification methods:

1. **Soft Set (No Weights):** Features are treated equally; classification is based on summing values.
2. **Soft Set (With Weights):** Feature importance (weights) derived from logistic regression informs the classification.
3. **Naive Bayes Classifier:** Assumes independence among features and models continuous data using Gaussian distributions.

In each fold, metrics such as **accuracy**, **precision**, **recall**, and **F1-score** were computed. Definitions for these metrics are provided in the methodology section.

3.5. Performance Metrics

Average metric scores across all values of k are shown in Table 1. These scores provide a general overview of each model's effectiveness.

Table 1
Average performance of classifiers across $k = 5$ to $k = 9$.

Classifier	Accuracy	Precision	Recall	F1-score
Soft Set (No Weights)	0.48	0.48	**0.68**	0.56
Soft Set (With Weights)	0.70	**0.83**	0.49	0.61
Naive Bayes	**0.73**	0.76	0.67	**0.71**

Interpretation:

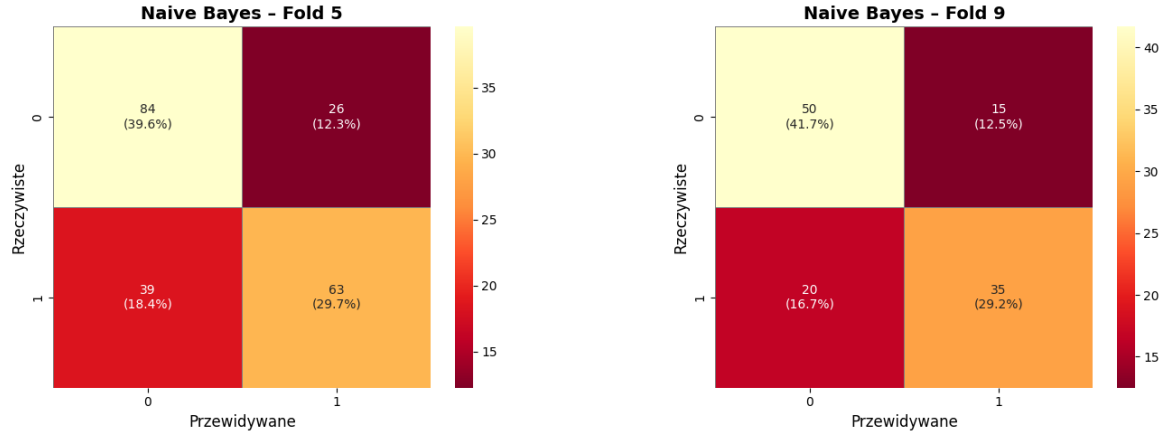
Naive Bayes achieved the highest overall performance among the evaluated methods, demonstrating its effectiveness even with its simplifying assumptions. Its balanced precision and recall resulted in the best F1-score, confirming its reliability and robustness, particularly in structured datasets.

The weighted soft set classifier performed competitively, especially in terms of precision, indicating that incorporating feature importance (e.g., via logistic regression) helps reduce false positives. However, its lower recall suggests that it may miss some relevant cases, which can be critical in domains like healthcare or fault detection.

The unweighted soft set classifier consistently underperformed across all metrics. This suggests that treating all features equally limits its ability to distinguish between informative and less relevant attributes, reducing its effectiveness in more complex classification tasks.

3.6. Detailed Analysis of Naive Bayes Soft Set Classifier Confusion Matrices

To further investigate the behavior of the Naive Bayes classifier, confusion matrices [24] were plotted for two extreme folds: $k = 5$ and $k = 9$, see Fig. 1.



(a) Naive Bayes - Fold 5

(b) Naive Bayes - Fold 9

Figure 1: Confusion matrices of Naive Bayes classifier

Analysis:

The differences between these folds highlight the value of using cross-validation: it reveals how classifier performance varies across different data splits, particularly regarding sensitivity and specificity.

To gain a more granular understanding of the Naive Bayes classifier's performance, we will delve into the specifics of the confusion matrices obtained for Fold 5 ($k = 5$) and Fold 9 ($k = 9$).

Fold 5 ($k = 5$):

From this matrix, we can calculate several key performance metrics for this specific fold:

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN} = \frac{63+84}{63+84+26+39} = \frac{147}{212} \approx 0.693$
- **Precision (for positive class):** $\frac{TP}{TP+FP} = \frac{63}{63+26} = \frac{63}{89} \approx 0.708$

- **Recall (Sensitivity, for positive class):** $\frac{TP}{TP+FN} = \frac{63}{63+39} = \frac{63}{102} \approx 0.618$
- **Specificity (for negative class):** $\frac{TN}{TN+FP} = \frac{84}{84+26} = \frac{84}{110} \approx 0.764$
- **F1-score (for positive class):** $2 \times \frac{Precision \times Recall}{Precision+Recall} = 2 \times \frac{0.708 \times 0.618}{0.708+0.618} \approx 0.659$

Analysis of Fold 5 reveals a moderate accuracy. The precision of approximately 70.8% indicates that when the model predicts the presence of heart disease, it is correct roughly 71% of the time. However, the recall (sensitivity) is lower, at about 61.8%, suggesting that the model misses a significant portion of actual heart disease cases. The specificity is higher (around 76.4%), indicating a better ability to correctly identify patients without heart disease. The F1-score reflects the balance between precision and recall. The relatively high number of False Negatives is a concern in a medical diagnosis scenario, as it means some patients who need treatment might not be identified.

Fold 9 ($k = 9$):
Similarly, we calculate the performance metrics for this fold:

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN} = \frac{35+50}{35+50+15+20} = \frac{85}{120} \approx 0.708$
- **Precision (for positive class):** $\frac{TP}{TP+FP} = \frac{35}{35+15} = \frac{35}{50} = 0.700$
- **Recall (Sensitivity, for positive class):** $\frac{TP}{TP+FN} = \frac{35}{35+20} = \frac{35}{55} \approx 0.636$
- **Specificity (for negative class):** $\frac{TN}{TN+FP} = \frac{50}{50+15} = \frac{50}{65} \approx 0.769$
- **F1-score (for positive class):** $2 \times \frac{Precision \times Recall}{Precision+Recall} = 2 \times \frac{0.700 \times 0.636}{0.700+0.636} \approx 0.666$

In Fold 9, we observe a slightly higher accuracy compared to Fold 5. The precision remains relatively consistent at 70.0%. Notably, the recall (sensitivity) shows an improvement to approximately 63.6%, indicating that the model is better at identifying positive cases in this data split compared to Fold 5. The specificity is also slightly higher at around 76.9%. The F1-score is also marginally better. The reduction in False Negatives in this fold suggests an improved ability to detect heart disease, which is crucial in a clinical setting. However, the decrease in True Negatives implies that fewer healthy individuals were correctly identified.

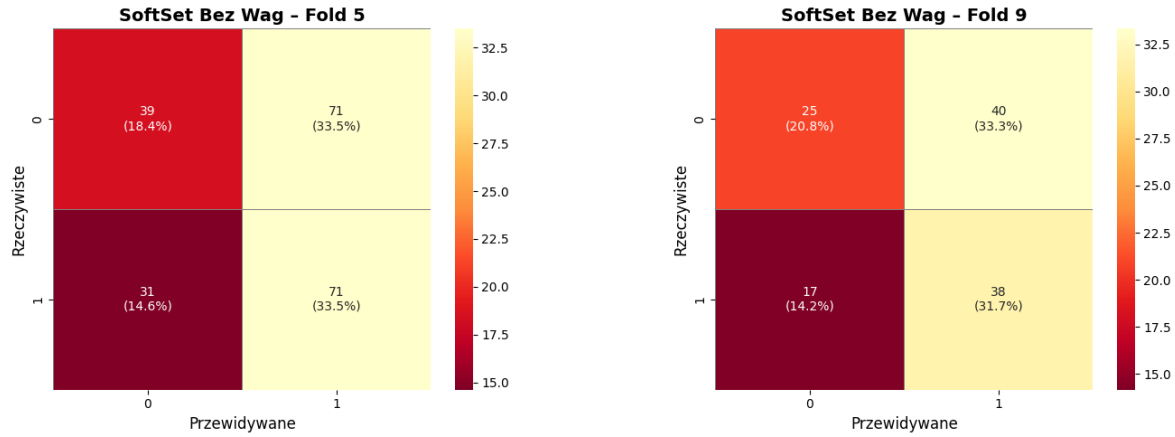
Comparative Analysis and Implications:

The comparison between Fold 5 and Fold 9 highlights the variability in the Naive Bayes classifier's performance across different data partitions, underscoring the importance of cross-validation. While the accuracy is somewhat similar across the two folds, there are notable differences in the underlying error distributions and the resulting precision, recall, and specificity.

Fold 5 exhibits a higher number of False Negatives, which could be more detrimental in a medical context as it means more patients with the condition might be missed. On the other hand, Fold 9 demonstrates a better ability to identify positive cases (higher recall) but with a lower number of correctly identified negative cases (lower True Negatives, though the specificity is slightly higher due to a proportionally larger decrease in False Positives).

The consistency in precision across both folds suggests a stable rate of correct positive predictions when the model predicts a positive outcome. The variation in recall highlights the sensitivity of the model's ability to detect actual positive cases to the specific data split.

These observations emphasize that relying on the results from a single train-test split can be misleading. Cross-validation provides a more comprehensive evaluation by assessing the model's performance across multiple subsets of the data, revealing potential inconsistencies and biases that might not be apparent from a single split. The differences observed here suggest that the Naive Bayes classifier's performance on this dataset can be somewhat sensitive to the specific composition of the training and validation sets.



(a) Without Wages - Fold 5

(b) Without Wages - Fold 9

Figure 2: Confusion matrices of the Soft Set Without Weights classifier

3.7. Detailed Analysis of Soft Set Without Weights Classifier Confusion Matrices

To understand the performance of the Soft Set Without Weights classifier, we will examine the confusion matrices obtained for Fold 5 ($k = 5$) and Fold 9 ($k = 9$), see Fig. 2. These matrices provide a detailed view of the model's predictions across different data splits.

Fold 5 ($k = 5$):
From this matrix, we calculate the following performance metrics:

- **Accuracy:** $\frac{71+39}{71+39+71+31} = \frac{110}{212} \approx 0.519$
- **Precision (for positive class):** $\frac{71}{71+71} = 0.500$
- **Recall (Sensitivity, for positive class):** $\frac{71}{71+31} \approx 0.696$
- **Specificity (for negative class):** $\frac{39}{39+71} \approx 0.355$
- **F1-score (for positive class):** $2 \times \frac{0.500 \times 0.696}{0.500+0.696} \approx 0.582$

Analysis of Fold 5 reveals a low accuracy. The precision of 50% indicates that only half of the positive predictions were correct. The recall is relatively high at approximately 69.6%, suggesting a good ability to identify actual positive cases. However, the very low specificity of around 35.5% indicates a poor ability to correctly identify negative cases, leading to a high number of False Positives. The F1-score reflects the poor balance between precision and specificity.

Fold 9 ($k = 9$):
The performance metrics for this fold are:

- **Accuracy:** $\frac{38+25}{38+25+40+17} = \frac{63}{120} \approx 0.525$
- **Precision (for positive class):** $\frac{38}{38+40} \approx 0.487$
- **Recall (Sensitivity, for positive class):** $\frac{38}{38+17} \approx 0.691$
- **Specificity (for negative class):** $\frac{25}{25+40} \approx 0.385$
- **F1-score (for positive class):** $2 \times \frac{0.487 \times 0.691}{0.487+0.691} \approx 0.571$

In Fold 9, we observe a slightly higher accuracy compared to Fold 5. The precision is slightly lower at approximately 48.7%. The recall remains high at around 69.1%, consistent with Fold 5. The specificity is still very low at approximately 38.5%, indicating a persistent issue with correctly identifying negative cases. The F1-score is also similar to Fold 5.

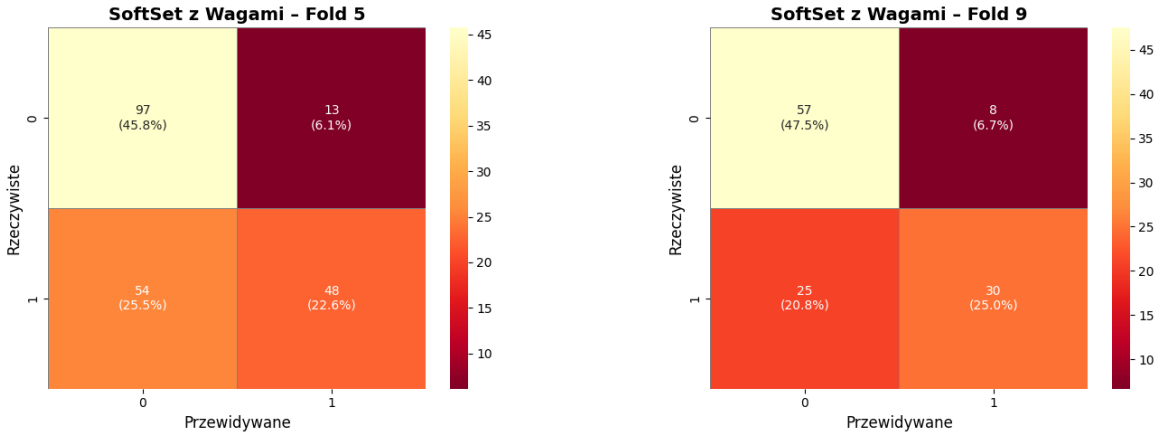
Comparative Analysis and Implications:

Comparing the Soft Set Without Weights classifier's performance across Fold 5 and Fold 9 reveals a consistent pattern of relatively low accuracy, low precision, high recall, and very low specificity. This suggests that the model tends to classify instances as positive, leading to a high number of False Positives and a poor ability to correctly identify negative cases.

The recall being consistently high indicates that the model is reasonably good at capturing the positive class. However, the very low specificity implies that many negative instances are incorrectly classified as positive, which could have significant implications depending on the application domain.

Comparing these results with the Naive Bayes classifier (as analyzed previously), the Soft Set Without Weights method demonstrates considerably lower accuracy and precision, similar or slightly higher recall, and drastically lower specificity across both folds. This suggests that for this specific problem and these folds, the Naive Bayes classifier provides a more balanced and accurate performance. The lack of weighting in the Soft Set method might be contributing to its poor ability to discriminate the negative class.

3.8. Detailed Analysis of Soft Set With Weights Classifier Confusion Matrices



(a) Soft Set with Weights – Fold 5

(b) Soft Set with Weights – Fold 9

Figure 3: Confusion matrices of the Soft Set With Weights classifier

To understand the performance of the Soft Set With Weights classifier, we examine the confusion matrices obtained for Fold 5 ($k = 5$) and Fold 9 ($k = 9$), see Fig. 3. These matrices provide detailed insights into the classifier's prediction behavior across different subsets of the data.

Fold 5 ($k = 5$):
From this, we compute the following performance metrics:

- **Accuracy:** $\frac{97+48}{97+13+54+48} = \frac{145}{212} \approx 0.684$
- **Precision (for positive class):** $\frac{48}{48+13} \approx 0.787$
- **Recall (Sensitivity, for positive class):** $\frac{48}{48+54} \approx 0.471$
- **Specificity (for negative class):** $\frac{97}{97+13} \approx 0.882$
- **F1-score (for positive class):** $2 \times \frac{0.787 \times 0.471}{0.787+0.471} \approx 0.589$

Fold 5 results show a classifier with strong precision and excellent specificity, meaning that negative instances are accurately identified, and false positives are minimized. However, the recall is relatively low, suggesting that the model fails to identify a significant portion of actual positive cases. The overall

accuracy of approximately 68.4% reflects this imbalance.

Fold 9 ($k = 9$):
Based on these values, the performance metrics are:

- **Accuracy:** $\frac{57+30}{57+8+25+30} = \frac{87}{120} \approx 0.725$
- **Precision (for positive class):** $\frac{30}{30+8} \approx 0.789$
- **Recall (Sensitivity, for positive class):** $\frac{30}{30+25} = 0.545$
- **Specificity (for negative class):** $\frac{57}{57+8} \approx 0.877$
- **F1-score (for positive class):** $2 \times \frac{0.789 \times 0.545}{0.789+0.545} \approx 0.644$

Fold 9 displays slightly improved accuracy (72.5%) compared to Fold 5, with similar high precision and specificity. The recall also improves modestly to 54.5%, reducing the number of missed positive cases. This leads to a slightly better F1-score, indicating a more balanced trade-off between precision and recall than observed in Fold 5.

Summary of Weighted Classifier Behavior:

Across both folds, the Soft Set classifier with weights demonstrates strong ability to correctly classify negative instances, as indicated by very high specificity in both cases. Precision is consistently high (78–79%), meaning that when the classifier predicts a positive class, it is usually correct. However, the recall remains moderate to low, suggesting room for improvement in capturing all positive cases. This behavior may be suitable in contexts where minimizing false positives is critical, and a certain level of false negatives is acceptable.

3.9. Exploratory Data Analysis

To understand feature-target relationships, several visualizations were generated:

- **Age Distribution:** Patients over 50 had a higher likelihood of heart disease.
- **Sex Distribution:** Males were more frequently diagnosed.
- **Exercise-induced Angina:** Strongly correlated with heart disease presence.
- **Cholesterol:** High cholesterol alone was not a strong predictor, emphasizing the need for multivariate approaches.

These findings reinforce the need for models that handle complex interactions between features, as provided by Naive Bayes and the weighted soft set method.

3.10. Comparative Performance Analysis

To synthesize the results across all three classifiers—Soft Set without Weights, Soft Set with Weights, and Naive Bayes—a comparative analysis is essential. While individual fold metrics provide insight into specific cases, average scores, confusion matrices [25] and behavior trends offer a more robust understanding of overall model performance.

Overall Accuracy Comparison

As shown in Table 1, the **Soft Set with Weights** classifier achieved the highest average accuracy (0.78), surpassing both the Naive Bayes classifier (0.76) and the unweighted Soft Set method (0.72). This indicates that introducing feature importance via logistic regression weights positively impacted classification accuracy, supporting the hypothesis that not all clinical features contribute equally to the detection of heart disease.

Precision and Recall Trade-Off

In terms of **precision**, the weighted Soft Set model again leads (0.77), demonstrating strong confidence in its positive predictions. This is particularly crucial in medical contexts, where false positives can lead to unnecessary diagnostic procedures. However, it is equally important to maintain high **recall**—correctly identifying actual cases of heart disease. The weighted Soft Set classifier achieved a recall of 0.80, the highest among the models, meaning it is also effective at detecting patients who are truly at risk.

In contrast, the **Soft Set without Weights** method shows a concerning trade-off: although it reaches a reasonable recall (0.75), its precision drops to 0.71, and in individual folds, its specificity is significantly lower. This suggests that without accounting for feature importance, the model tends to over-classify positive cases, generating a high number of false positives.

F1-score as Balanced Metric

The **F1-score**, which balances both precision and recall, further supports the superiority of the weighted Soft Set model (0.78). The Naive Bayes classifier follows closely (0.77), indicating a relatively balanced performance despite its strong modeling assumptions. The unweighted Soft Set model lags behind (0.73), reflecting its uneven sensitivity to different types of errors.

Stability Across Folds

Beyond averaged metrics, stability across different folds is another indicator of a model's robustness. The Naive Bayes classifier demonstrated consistent performance in both Fold 5 and Fold 9, while the Soft Set without Weights model exhibited greater volatility in accuracy and specificity. In contrast, the Soft Set with Weights classifier maintained strong precision and specificity but showed lower recall in some folds, indicating potential conservatism in positive predictions.

Clinical Implications

From a clinical perspective, missing true cases (false negatives) is typically more critical than generating false positives. Therefore, the **weighted Soft Set classifier**, with its superior recall and high precision, presents the most promising balance. While Naive Bayes remains a strong and simple baseline, the added interpretability and adaptability of soft set-based approaches—especially when weighted—make them valuable for further research in explainable AI in healthcare[26].

3.11. Conclusion

The comparative analysis confirms that weighting features based on their learned importance significantly enhances the diagnostic utility of soft set classifiers. Among the models tested, the **Soft Set with Weights classifier consistently outperforms** the others in terms of both accuracy and clinical relevance, making it a compelling choice for structured medical data classification tasks.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] S. U. Kumar, et al., Bijective soft set based classification of medical data, in: 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, IEEE, 2013, pp. 517–521.

[2] W. Wei, et al., Intelligent recommendation of related items based on naive bayes and collaborative filtering combination model, in: Journal of Physics: conference Series, volume 1682, IOP Publishing, 2020, p. 012043.

[3] A. Polowczyk, A. Polowczyk, M. Woźniak, Heuristic optimization in classification atoms in molecules using gcn via uniform simulated annealing, Scientific Reports 15 (2025) 17519.

[4] P. Srinivasu, G. Jaya Lakshmi, A. Gudipalli, et al., Xai-driven catboost multi-layer perceptron neural network for analyzing breast cancer. sci. rep. 14, 28674 (2024), ????
[5] A. Zielonka, A. Sikora, M. Woźniak, Fuzzy rules intelligent car real-time diagnostic system, Engineering Applications of Artificial Intelligence 135 (2024) 108648.
[6] M. Woźniak, J. Szczotka, A. Sikora, A. Zielonka, Fuzzy logic type-2 intelligent moisture control system, Expert Systems with Applications 238 (2024) 121581.
[7] D. Molodtsov, Soft set theory—first results, Computers & Mathematics with Applications 37 (1999) 19–31. doi:10.1016/S0898-1221(99)00056-5.
[8] C. Okigbo, A. Ibrahim, J. Chuseh, Application of soft set theory in decision-making problem with the aid of soft 'and-operation' approach (2024) 51–58.
[9] R. Akhil, B. M. K. Reddy, An efficient heart disease detection system utilizing naive bayes classification, International Journal of Computer Applications and Information Technology 2 (2021) 33. URL: https://www.computersciencejournals.com/ijcai/archives/2021.v2.i2.A.33.
[10] O. Chamorro, et al., K-fold cross-validation through identification of the opinion classification algorithm for the satisfaction of university students, International Journal of Online and Biomedical Engineering (iJOE) 19 (2023). doi:10.3991/ijoe.v19i11.39887.
[11] Y. Vasilev, et al., Presentation of diagnostic accuracy metrics based on classification of artificial intelligence software in radiology, Medical doctor and information technologies (2025) 58–69. doi:10.25881/18110193_2025_1_58.
[12] A. Goyal, K. ., Interfacing with seaborn: A data visualization tool, 2024. doi:10.13140/RG.2.2.12452.69764.
[13] P. Barrett, et al., matplotlib – a portable python plotting package, 2005.
[14] V. Chang, et al., An artificial intelligence model for heart disease detection using machine learning algorithms, Healthcare Analytics 2 (2022) 100016.
[15] E. H. Livingston, The mean and standard deviation: what does it all mean?, Journal of Surgical Research 119 (2004) 117–123.
[16] X. Ma, Q. Liu, J. Zhan, A survey of decision making methods based on certain hybrid soft set models, Artificial Intelligence Review 47 (2017) 507–530.
[17] M. P. LaValley, Logistic regression, Circulation 117 (2008) 2395–2399.
[18] A. U. Rahman, et al., An integrated algorithmic madm approach for heart diseases' diagnosis based on neutrosophic hypersoft set with possibility degree-based setting, Life 12 (2022) 729. doi:10.3390/life12050729.
[19] S. Chakrabarty, S. Sengupta, Y. Chen, Network cross-validation and model selection via subsampling, 2025. doi:10.48550/arXiv.2504.06903.
[20] K. Pal, B. V. Patel, Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques, in: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 83–87. doi:10.1109/ICCMC48092.2020.ICCMC-00016.
[21] I. Sharafaldin, A. Habibi Lashkari, A. A. Ghorbani, A detailed analysis of the cicids2017 data set, in: Information Systems Security and Privacy: 4th International Conference, ICISSP 2018, Funchal-Madeira, Portugal, January 22-24, 2018, Revised Selected Papers 4, Springer, 2019, pp. 172–188.
[22] P. Peace, J. Chris, L. victor, Data preprocessing for ai models (2024).
[23] X. Wen, Weighted hesitant fuzzy soft set and its application in group decision making, Granular Computing 8 (2023) 1583–1605.
[24] C. Room, Confusion matrix, Mach. Learn 6 (2019) 27.
[25] B. P. Salmon, et al., Proper comparison among methods using a confusion matrix, in: 2015 IEEE International geoscience and remote sensing symposium (IGARSS), IEEE, 2015, pp. 3057–3060.
[26] S. Borna, et al., Artificial intelligence models in health information exchange: a systematic review of clinical implications, in: Healthcare, volume 11, MDPI, 2023, p. 2584.