Increasing Accuracy of Heart Disease Classification on C4.5 Algorithm Based on Information Gain Ratio and Particle Swarm Optimization Using Adaboost Ensemble

Mohammad Qois Syafi [1,*], Alamsyah [1]

[a] Departement of Computer Science, Faculty of Mathematics and natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: goissyafieee123@students.unnes.ac.id

ARTICLE INFO

Article history
Received 13 Maret 2022
Revised 4 April 2022
Accepted 25 April 2022

Keywords
Data Mining
Heart Disease
Classification
C4.5 Algorithm
Information Gain Ratio
Particle Swarm Optimization
Adaboost

ABSTRACT

The heart is a vital organ of the body that has an important role in the process of blood flow(nambahin tujuan). Data mining is a process to get very useful information from a very large data warehouse to facilitate the decision-making process. In the data mining process, the first stage performs data processing called preprocessing by handling data formatting. Then, the feature selection stage is carried out using the Information Gain Ratio and Particle Swarm Optimization algorithms to find the best attributes. Then the Adaboost ensemble was applied to optimize the accuracy results. Next, it is done by classifying the dataset. The algorithm used for classification is the C4.5 algorithm. Based on the research that has been done, using the k-fold = 5 model test with three trials, the best accuracy results are obtained for the C4.5 algorithm without feature selection and the Adaboost ensemble produces an accuracy rate of 95.87%, while the C4.5 algorithm with Information Gain Ratio and Particle Swarm Optimization then applying the Adaboost ensemble produces an accuracy rate of 96.68%. This shows that the feature selection algorithm, namely, Information Gain Ratio and Particle Swarm Optimization by applying the Adaboost ensemble is considered to be able to improve the performance of the C4.5 classification algorithm.

1 Introduction

The heart is a vital organ of the body that has an important role in the process of blood flow. The importance of maintaining and caring for heart health has a major influence on human survival. This is because the heart is the main controller in the blood circulation system, where this organ works in pumping blood. The relationship between the work of the heart and other organs of the body makes the quality of the work of the heart can affect the work of other organs, so that if a problem occurs it can interfere with the work of other organs (Rusdiana et al., 2019). The perspective of medical science and data mining is used to find various kinds of metabolic syndrome. Data mining with classification plays an important role in heart disease prediction and data investigation (Prasetio & Susanti, 2019).

Classification is the process of obtaining a model or function that will describe and distinguish data classes or concepts, where the model is derived from a combined analysis of the training data. The model is used to be able to predict object classes where the class label is unknown (Han et al., 2012). Classification is a technique can be used to predict data or describe data classes (Alamsyah & Fadila, 2021). Decision Tree is the most powerful approach in science discovery and data mining, as well as a very effective tool in areas such as data and text mining, information extraction, machine learning, and pattern recognition (Mathuria et al., 2013). The C4.5 Algorithm is one of the algorithms

developed by J. Ross Quinlan which is the development of the Iterative Dichotomiser 3 algorithm (Muzakir & Wulandari., 2016). The C4.5 Algorithm has been used to predict a data (Prihanditya & Alamsyah, 2020).

Problems in classification often occur when finding a large number of features in the dataset, but not all of them are used. Things that can reduce classification performance are irrelevant and redundant features (Xue et al., 2012). To maximize the accuracy of the classification, you can use feature selection to select the features to be used. After the preprocessing process, a feature selection was applied to reduce attributes that did not increase the results of classification accuracy (Lestari & Alamsyah, 2020). The previous research conducted by Sari, this research is entitled "Increasing Accuracy of SVM Algorithm Using PSO and Information Gain for Heart Disease Diagnosis." and discusses the prediction of heart disease and the method used is Support Vector Machine (SVM) with feature selection methods are Information Gain Ratio (IGR) and Particle Swarm Optimization (PSO). The resulting accuracy of the SVM method with IGR and PSO is 89.66% (Sari, 2019).

Accuracy in the classification algorithm is important because it is used to measure how accurately the classification algorithm is successful in classifying data based on learning from training data. The importance of accuracy in the classification algorithm is very influential in determining a decision. Based on the description of the problem above, the research focuses on the application of feature selection, it is IGR, PSO and applying the Adaboost ensemble to improve diagnostic accuracy in the C4.5 Algorithm.

2 The Proposed Algorithm

2.1 C4.5 Algorithm

The C4.5 Algorithm was introduced by Quinlan as an improved version of ID3. In ID3, Decision Tree induction can only be performed on features of categorical type (nominal/ordinal), while numerical types (internal/ratio) cannot be used. Improvements that distinguish the C4.5 Algorithm from ID3 is that can handle features with numeric types, pruning Decision Tree, and deriving rule sets. The C4.5 algorithm also uses gain criteria in determining the features that break the nodes in the induced tree (Prasetyo, et al., 2014).

The C4.5 Algorithm includes a continuation of the ID3 algorithm with several developments such as processing continuous data, being able to overcome missing values (Irena & Erwin Budi Setiawan, 2020). This algorithm is a very powerful and well-known method in classification and prediction techniques in data mining. The C4.5 algorithm can handle discrete data (Irawan, 2021).

2.2 Information Gain Ratio (IGR)

Attributes with multivariables are enough to speed up predictions in the calculation process (Mahalingam & Vivek, 2016). However, it will be a loss if all calculation processes are assessed in the same way, because some transactions must be given additional weight over others. Information Gain can be thought of as a parameter to measure the impact of a particular item in a set. For this reason, the ratio parameter taken from the ID3 Decision Tree is added to the Information Gain. IGR is the ratio of the acquisition of Information Gain with intrinsic information. Quoted from Quinlan, to reduce bias towards multi value attributes by taking the number and size of branches in a calculation when selecting attributes.

2.3 Particle Swarm Optimization (PSO)

Algorithm PSO is an optimization or search technique based on random (stochastic) variables developed by (Kennedy and Eberhart, 1995), which was inspired by a group of birds or fish (Saputra, 2020). PSO is an optimization method that can be used to determine process parameters that produce optimal response values (Sateria A et.al., 2018). PSO is used because the simplicity, easy to implement, and relatively fast convergence. In the PSO technique there are several ways to optimize, including increasing the attribute weight of all attributes or variables used, selecting attributes (attribute selection) and feature selection (Giovani et al., 2020).

The difference between PSO and other optimization techniques is that each PSO particle is associated with a velocity. These particles tend to move to a better search area with each passing search process. PSO is almost the same as Genetic Algorithm, but PSO does not use evolution operators such as mutation and recombination (crossover) (Muslim M. A et al., 2019).

2.4 Adaptive Boosting (Adaboost)

In improving classification performance, it is necessary to have Adaboost which is a boosting algorithm from ensemble learning (Listiana, 2017). Boosting is an approach to machine learning to improve accurate predictive rules by combining many relatively weak and inaccurate rules and Adaboost is a popular learning ensemble and is often used in boosting algorithms (Nurzahputra & Muslim M. A, 2017).

3 Methods

In this study, the classification of heart disease is carried out by applying the C4.5 Algorithm as a classification method for data mining, IGR and PSO is used in the feature selection process to obtain the selected attribute, then Adaboost is used to improve the results of classification performance in diagnosing heart disease. The result of this research is the accuracy resulting from the proposed method. The flow chart of the method proposed in this study can be seen in Figure 1.

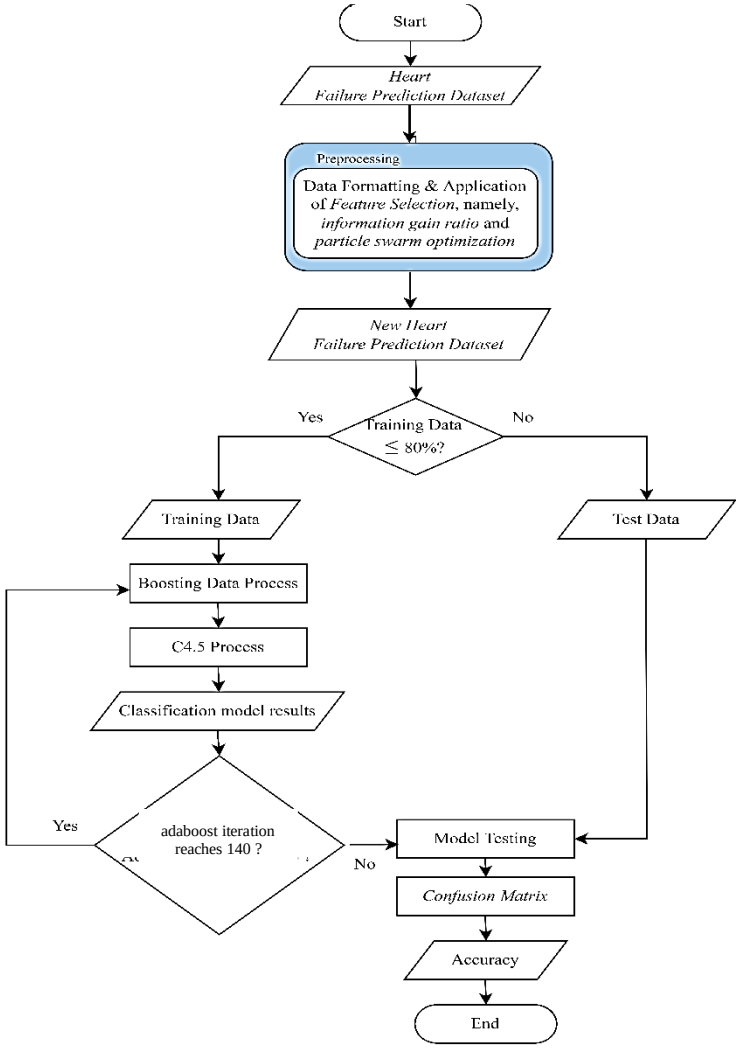

Figure 1. Flowchart Of Proposed Method

Based on Figure 1 before the classification process is carried out, the Heart Failure Prediction Dataset used is processed first at the preprocessing stage with data formatting. The results of the formatting of the data are processed by feature selection with IGR to select the attributes that influence the classification. After that, the next feature selection is continued by using PSO to maximize the feature selection process. The Particle Swarm Optimization flowchart can be seen in Figure 2. The selected attributes obtained from the feature selection process are used in the classification process by dividing them into training data and test data. The next stage is applying Adaboost to improve classification performance and get maximum accuracy. The classification process is applied using the C4.5 algorithm. The flowchart of the C4.5 algorithm can be seen in Figure 3.

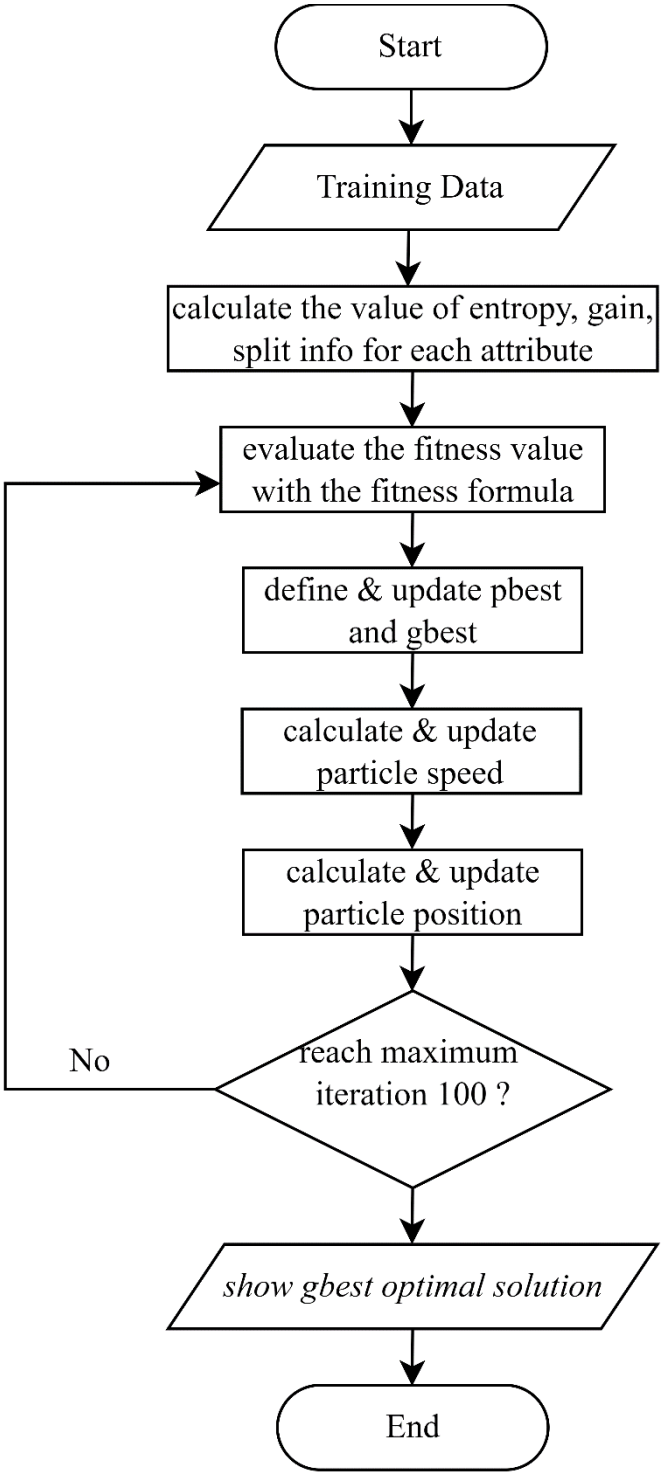

Figure 2. Flowchart Of Particle Swarm Optimization

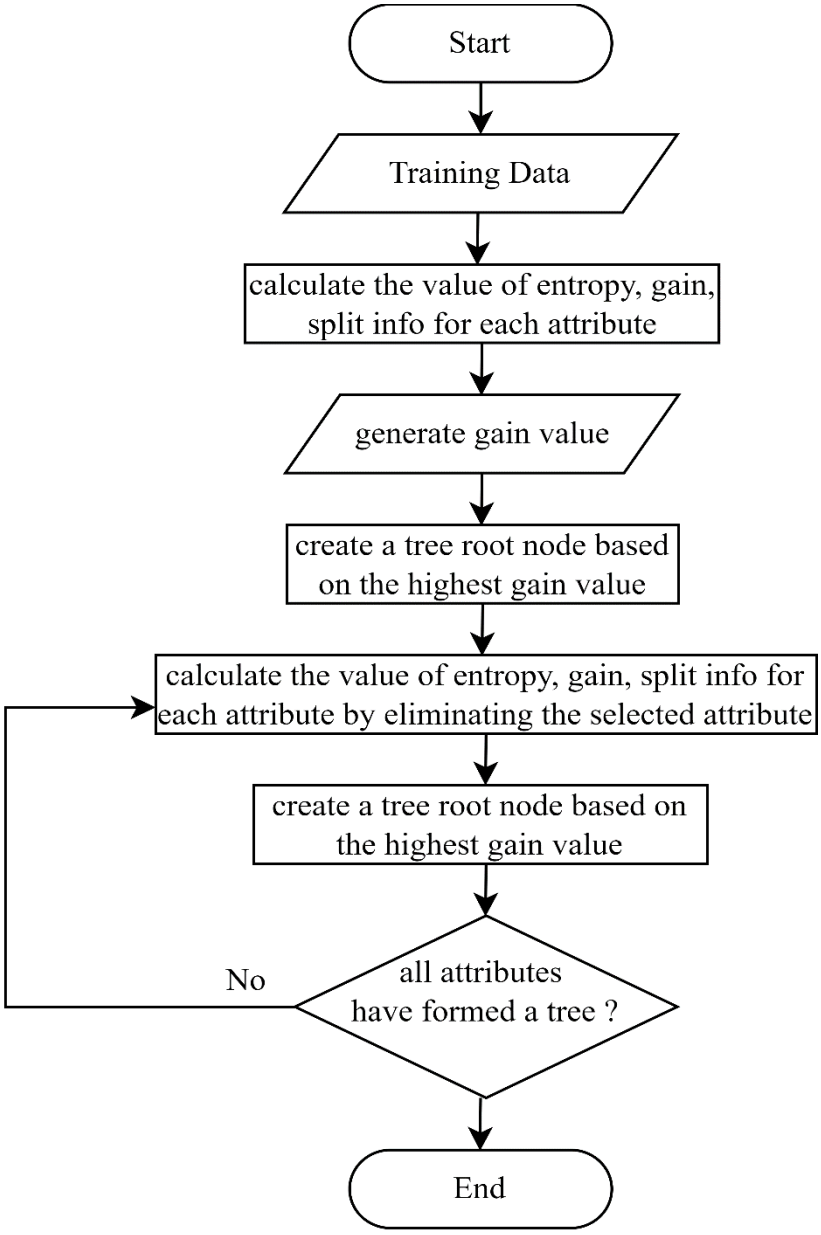

Figure 3. Flowchart Of C4.5 Algorrithm

4 Results and Discussion

This section is divided into two parts, results and discussion. The results are a description of the data and findings obtained using the methods and procedures described in the data collection method. The discussion is an explanation of the results that answer research questions more comprehensively.

4.1 Results

From this research, the method used is IGR, PSO, Algorithm C4.5 and Adaboost ensemble to diagnose heart disease. In this study, the application of feature selection in the Heart Failure Prediction Dataset aims to select attributes that match certain criteria in order to improve quality so that optimal results are obtained. Feature selection selected in this study are IGR and PSO. The implementation of the Adaboost ensemble to optimize the accuracy of the C4.5 Algorithm as a learning ensemble on the Heart Failure Prediction Dataset dataset. The research was carried out in several stages, namely the data preprocessing stage, the feature selection stage, and the data mining stage. The following is a more complete explanation of the research results.

4.1.1 *Data Collection*

The results of data collection in this study were taken from www.kaggle.com. The Heart Failure Prediction Dataset consists of 918 rows with 12 attributes and 1 target used in classifying people with heart disease and not for someone. The following describes the attributes and data types used in the Heart Failure Prediction Dataset in the Table 1.

Table 1. Heart Failure Prediction Dataset

No.	Attribute	Attribute Type	Description
1.	Age	Numeric	Age in year
2.	Sex	Nominal	Male, Female
3.	Chest pain type	Nominal	Typical angina, Atypical angina, Non-anginal pa, Asymptomatic
4.	Resting blood pressure	Numeric	In/mm/Hg
5.	Serum Cholestoral	Numeric	Mg/dl
6.	Fasting blood sugar	Nominal	True, False
7.	Resting electrocardiographic results	Nominal	Normal, Having ST-T wave abnormality, Showing probable or definite left ventricular hypertrophy by Estes' criteria
8.	Maximum heart rate achieved	Numeric	Maximum heart rate achieved
9.	Exercise angina	Nominal	Yes,No
10.	Oldpeak	Numeric	ST depression induced by exercise relative to rest
11.	The slope of the peak exercise ST segment	Nominal	Up sloping, Flat, Down sloping
12.	Heart Desease	Nominal	diagnosis of heart disease : Healthy, possible heart disease

4.1.2 ***Data Preprocessing***

At this preprocessing stage, it can be used to identify and improve the data to be studied. In improving the heart disease data that will be examined comes from the fact that raw data is often not suitable for mining. At this stage the data is converted and combined into the required format or the selected data is transformed into a form suitable for the data mining process by initializing each data that has a letter type so that it can be transformed into numbers. Because the classification process can only accept input data in the form of numeric numbers. The data formatting stage aims to create a standard format for the datasets used in the research. Formatting performed on the Heart Failure Prediction Dataset is shown in Table 2.

Tabel 2. Formatting Heart Failure Prediction Dataset

No.	Attribute	Formatting
1	Sex	Change the Sex attribute to 0 for female and 1 for male
2	Chestpaintype (Chest pain type)	Change the Chestpaintype attribute to: 0 for ASY: Asymptomatic, 1 for ATA: Atypical Angina, 2 for NAP: Non-Anginal Pain, 3 for TA: Typical Angina.
3	RestingECG (Resting electrocardiographic results)	Change the Sex attribute to 1 for Normal and 2 for ST: having ST-T wave abnormality
4	ExerciseAngina (Exercise angina)	Change the ExerciseAngina attribute to 0 for No and 1 for yes
5	ST_Slope (The slope of the peak exercise ST segment)	Change the ST_Slope attribute to 0 for Down: downsloping, 1 for Flat: flat and 2 for Up: upsloping

4.1.3 ***Feature Selection***

At this stage, the IGR is used for the feature/attribute selection process which is then continued with classification using the C4.5 Algorithm. IGR will find the best number of features that will be used in the classification process. IGR will weight each attribute. The attribute selected for the classification process is the attribute with the largest gain value. Table 3 shows the results of the weighting of each attribute using the information gain ratio.

Tabel 3. Weighting Results with Information Gain Ratio

No.	Attribute	Ratio
1.	Age	0.06332025
2.	Sex	0.04622387
3.	ChestPainType	0.14333598
4.	RestingBP	0
5.	Cholesterol	0.07398648
6.	FastingBS	0.02755282
7.	RestingECG	0
8.	MaxHR	0.11061584

9.	ExerciseAngina	0.14975398
10.	Oldpeak	0.11821292
11.	ST_Slope	0.2039712

Determination of the number of features selected for the classification process using C4.5 Algoritm, determined by the parameter k. Determining the number of attributes used or the value of k, the researcher tested each number of k with a value of k = 10, 9, 8, 7, 6, 5. In each trial, the researcher assessed the quality of the number of k, using the classification C4.5 Algorithm and selected the number of k that produces the highest accuracy. From the results of experiments that have been carried out, the number of k = 7 has the highest average accuracy of 77.22%. Accuracy results for each k can be seen in Table 4.

Tabel 4. Accuracy Results for Every k Process Information Gain Ratio

Experiment	Amount k					
	K5	K6	K7	K8	K9	K10
1	80.43	80.23	82.60	81.52	81.52	80.60
2	77.17	79.89	77.17	79.34	78.88	77.34
3	81.52	75.23	76.63	79.34	78.80	78.06
4	73.77	73.22	75.40	73.22	75.40	74.86
5	67.75	72.67	74.31	71.03	71.03	73.77
Average	76.12	76.24	77.22	76.89	77.12	76.92

Then the selection of the next feature is the PSO which is used to optimize the C4.5 classification algorithm. PSO aims to find the best features to be used in the classification process. This is done by specifying several parameters. Determination of this parameter is based on research has been done by (Chiu et al., 2009) and (Kennedy & Eberhart, 1995). These parameters include cognitive learning factors (*C1*) and social learning factors (*C2*), inertia weight (*w*), and number of iterations (*N*). This parameter value was chosen because it is considered to give convergence results.The following parameter values were selected among them $C1 = C2 = 1.49$, $w = 0.72$, $N = 100$.

The value of *w* is used as an inertia weight parameter whose function is to limit, if the optimization process has exceeded the inertia weight value, the process will automatically stop. In this feature selection process, the number of parameters for the particles used following the research of (Sabrila et al., 2022) is 30 particles. It mean, there will be 30 times the optimization process in every one iteration. Then for the number of iterations used in this study is 100 iterations. After 100 iterations, the final cost will be obtained by taking the lowest cost value or called the best cost. The cost value of the C4.5 algorithm with PSO for the three experiments can be seen in Table 5.

Tabel 5. Algorithm best cost value C4.5

Best Cost Value		
1st experiment	2nd experiment	3rd experiment
0.1701863	0.1721739	0.1654037

4.1.4 ***Data Mining***

At the data mining stage, there are two mining processes. First, the classification process using the C4.5 Algorithm on the Heart Failure Prediction Dataset. Second, the classification process using the C4.5 Algorithm with IGR and PSO, then applying the Adaboost ensemble. The results were evaluated using a confusion matrix. From the experiments that have been carried out, the accuracy results for the C4.5 Algorithm are shown in Table 6. And Table 7 shows the highest results from the classification calculated using the confusion matrix.

Table 6. Final Accuracy Results on the C4.5 Algorithm

Experiment	Accuracy C4.5
1	80.43%
2	81.52%
3	80.98%
Best Accuracy	81.52%

Table 7. Confusion Matrix using C4.5 Algorithm

Actual	Predicted		
	True Positive	True Negative	Amount
Pred Positive	99	16	115
Pred Negative	18	51	69
Amount	115	69	184

Accuracy = $\frac{TP+TN}{P+N} \times 100\%$

Accuracy = $\frac{99+51}{115+69} \times 100\% = 81.52\%$

The results of the highest accuracy from the application of the C4.5 Algorithm without using feature selection and Adabost ensemble in the classification of heart disease obtained the best accuracy of 81.52%. The accuracy obtained was considered low, so the next testing process was carried out.

The next classification result is the application of the C4.5 Algorithm classification with IGR, PSO and Adaboost ensemble using the Heart Failure Prediction Dataset. From the experiments have been carried out, the results obtained for the accuracy of the C4.5 Algorithm with feature selection and

applying Adaboost as shown in Table 8, also Table 9 shows the highest results from the classification calculated using the confusion matrix.

Table 8. Final Accuracy Results on the C4.5 Algorithm with IGR, PSO and applying Adaboost

Experiment	Accuracy C4.5 + IGR + PSO + ADABOOST
1	88.04%
2	90.76%
3	91.85%
Best Accuracy	91.85%

Table 9. Confusion Matrix using C4.5 Algorithm

Actual	Predicted		
	True Positive	True Negative	Amount
Pred Positive	100	15	115
Pred Negative	0	69	69
Amount	100	84	184

Accuracy = $\frac{TP+TN}{P+N} \times 100\%$

Accuracy = $\frac{100+69}{115+69} \times 100\% = 91.84\%$

The results of the highest accuracy from the application of the C4.5 Algorithm without using feature selection and Adabost ensemble in the classification of heart disease obtained the best accuracy of 81.52%. Thus, it can be concluded that the application of the IGR and PSO feature selection algorithm, then applying Adaboost can increase the accuracy of the C4.5 algorithm by 10.33%. in the diagnosis of heart disease.

4.2 Discussion

In this study, the calculation of the accuracy of the C4.5 algorithm without feature selection and Adaboost ensemble compared with the calculation of the accuracy of the C4.5 algorithm using IGR, PSO, then applying the Adaboost ensemble to improve accuracy in the Heart Failure Prediction Dataset. taken from Kaggle. Then, the data is carried out in a preprocessing stage, namely handling data formatting before the process is carried out using an algorithm. The accuracy results obtained from the use of the proposed method can be seen in Table 10.

Table 10. Accuracy Comparison

Method	Accuracy
C4.5	81.52%
C4.5 + IGR + PSO + ADABOOST	91.85%

From the results of the research on the classification of heart disease diagnosis using the C4.5 algorithm without a feature selection process, the best accuracy is 81.52%. While the results of the C4.5 algorithm classification by applying the IGR feature selection, followed by PSO and then applying the Adabost Ensemble gain the best accuracy is 91.85%. Thus, it can be concluded that the application of the IGR and PSO feature selection algorithm can increase the accuracy of the C4.5 algorithm by 10.33%. in the diagnosis of heart disease.

Based on the research that has been done, the IGR and PSO can be applied to the C4.5 algorithm to produce the best attributes and then apply the Adaboost ensemble to get the best accuracy. This is evidenced by the accuracy results obtained. The application of C4.5 + IGR + PSO + Adaboost resulted in a 2.19% higher accuracy rate compared to SVM + IG + PSO which was previously carried out by (Sari, 2019). By applying the method proposed in this study, we can maximize classification performance in order to obtain maximum and effective attribute selection accuracy. From the results obtained, it shows that the accuracy of the C4.5 algorithm without feature selection and Adaboost ensemble can be increased after the application of feature selection, namely, IGR and PSO, also applying the Adaboost ensemble. Based on the final results of this study, which were taken from the highest accuracy in each experiment that had been carried out, there was an increase in accuracy of 10.33%. Comparison of the accuracy used by previous studies can be seen in Table 11.

Table 11. Comparison of the Accuracy by Previous Studies

Research	Methods	Accuracy
(Sari, 2019)	SVM + IGR + PSO	89.66%
(Arul Jothi et al., 2021)	KNN	67%
(Annisa, 2019)	RF	80.38%
	DS	80.38%
(Dharmawan, 2021)	SVM + PSO	84.81%
	C4.5 + PSO	80%
(Nurmasani & Pristyanto, 2021)	Stacking	81%
	KNN	62%
	C4.5	78%
	SVM	78%

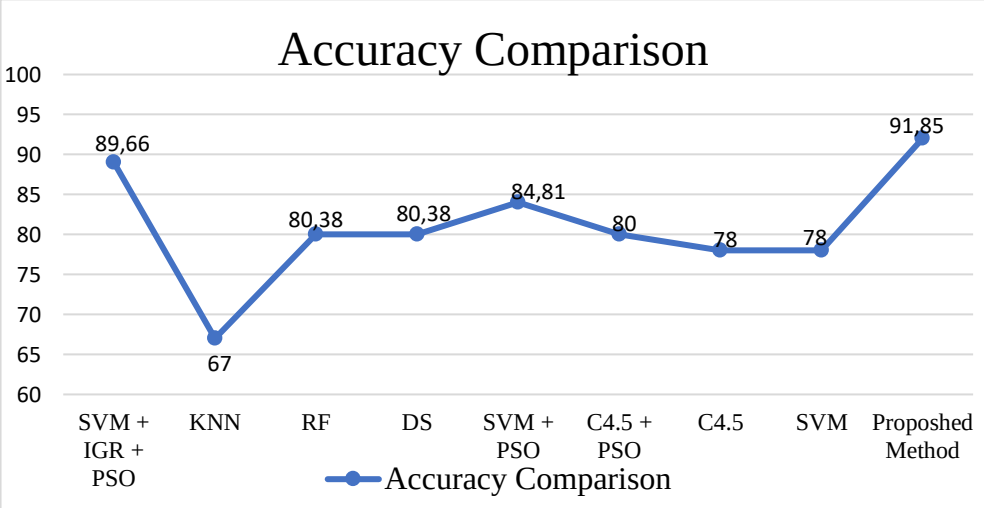

Figure 4. Chart Comparison of the Accuracy by Previous Studies

Based on Table 11 the proposed method to get better accuracy results than previous studies. The proposed method gets better accuracy because the proposed method uses the C4.5 Algorithm with

feature selection namely IGR, PSO, then applies Adaboost. The research method conducted by previous researchers did not use the C4.5 algorithm and Adaboost.

5 Conclusion

Based on the results of research that the accuracy of the C4.5 algorithm in the diagnosis of heart disease can be applied and improved using the IGR and PSO feature selection algorithm, then applying the Adaboost ensemble. The application of the PSO algorithm as feature selection uses parameters $C1 = C2 = 1.49$,$w = 0.72$; number of particles = 50; and the $N = 100$. The selected feature set has the highest fitness value which is represented by number 1. The application of the IGR calculates the gain value for each feature, the selection of features uses 7 attributes that have the highest gain value. The application of Adaboost to this classification aims to give more weight to inappropriate observations. The default iteration in Adaboost is 140 iterations. After the feature selection, the training data and testing data will be divided using a data splitting ratio of 80: 20 to obtain a classification model. Then apply Adaboost to optimize results accuracy. After that, the results obtained will be tested using test data and evaluated using a confusion matrix. Then from the results of the research on the classification of heart disease diagnosis using the C4.5 algorithm without the feature selection process, the best accuracy is 81.52%. While the results of the C4.5 algorithm classification by applying the IGR selection, followed by PSO and then applying the Adabost ensemble the best accuracy is 91.85%. Thus, it can be concluded that it can increase the accuracy of the C4.5 algorithm by 10.33% in the diagnosis of heart disease.

1. References

Alamsyah, A., & Fadila, T. (2021). Increased accuracy of prediction hepatitis disease using the application of principal component analysis on a support vector machine. Journal of Physics: Conference Series, 1968(1). https://doi.org/10.1088/1742-6596/1968/1/012016

Giovani, A. P., Haryanti, T., & Kurniawati, L. (2020). *Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi*. *14*(2), 116–124.

Han, J., Kamber, M., & Pei, J. (2012). Data mining: concepts and techniques, Waltham, MA. *Morgan Kaufman Publishers*, *10*, 978-1.

Irawan, Y. (2021). Penerapan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelayakan Calon Pendonor Darah Dengan Klasifikasi Data Mining (Application of the C4 . 5 Decision Tree Algorithm to Predict the Eligibility of Prospective Blood Donors by Classification Data Mining. *Jurnal Teknologi Informasi Dan Multimedia*, *2*(4), 181–189.

Irena, B., & Erwin Budi Setiawan. (2020). Fake News (Hoax) Identification on Social Media Twitter using Decision Tree C4.5 Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(4), 711–716.

Lestari, A., & Alamsyah. (2020). Increasing Accuracy of C4 . 5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease. Journal of Soft Computing Exploration, 1(1), 32–38.

Listiana, E. M. (2017). Penerapan Adaboost Untuk Klasifikasi Support Vector Machine Guna Meningkatkan Akurasi Pada Diagnosa Chronic Kidney Disease. *Prosiding SNATIF Ke-4 Tahun 2017*, *2015*, 153–160.

Mahalingam, P. R., & Vivek, S. (2016). Predicting Financial Savings Decisions Using Sigmoid Function and Information Gain Ratio. *Procedia Computer Science*,*93*(September), 19–25. https://doi.org/10.1016/j.procs.2016.07.176

Mathuria, M., Alam, L., an Haq, N. F., Mamun, T., Bhargava, N., Sharma, G., & Bhargava, R. (2013). Decision Tree Analysis on J48 Algorithm for Data Mining Cite this paper Related papers An Analyt ical Comparison on Filt er Feat ure Ext ract ion Met hod in Dat a Mining using J48 Classi… Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(6), 2277.

Muslim, Much Aziz, Prasetiyo, B., Harum, E. L., Juli, A., Mirqotussa'adah, Hardiyanti, S., & Nurzahputra, A. (2019). *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer*.

Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, *3*(1), 19–26.

Nurzahputra, A & Muslim, M. (2017). *Peningkatan Akurasi Pada Algoritma C4. 5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit*. *1*, 243–247.

Prasetio, R. T., & Susanti, S. (2019). Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor. *Jurnal Responsif*, *1*(1), 64–69.

Prasetyo,E., 2014. Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab,

Prihanditya, H. A., & Alamsyah, A. (2020). The Implementation of Z-Score Normalization and Boosting Techniques to Increase Accuracy of C4.5 Algorithm in Diagnosing Chronic Kidney Disease. Journal of Soft Computing Exploration, 1(1), 63–69.

Rusdiana, T., Putriana, N. A., Sopyan, I., Gozali, D., & Husni, P. (2019). Pemberian Pemahaman Mengenai Sediaan herbal yang Berfungsi untuk Pemeliharaan Kesehatan jantung dan Ginjal di Desa Cibeusi, Sumedang, Jawa Barat. *Jurnal Pengabdian Kepada Masyarakat*, *4*(6), 139–141.

Saputra, R. H. (2020). *Optimasi Algoritma C4.5 Menggunakan Seleksi Fitur Particle Swarm Optimization (Pso) Dan Teknik Bagging Pada Diagnosis Penyakit Kanker Payudara*.

Sari, S. P. (2019). *Increasing Accuracy of SVM Algorithm Using PSO and Information Gain for Heart Disease Diagnosis*. *6*(1), 1–10.

Xue, B., Zhang, M., & Browne, W. N. (2014). Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied soft computing*, *18*, 261-276.