❏   1128

Classification of cardiac disorders based on electrocardiogram data using a decision tree classification approach with the C45 algorithm

Sumiati[1], Viktor Vekky Ronald Repi[2], Penny Hendriyati[3], Anharudin[4], Afrasim Yusta[5], Agung Triayudi[6]

[1]Department of Informatic, Faculty of Information Technology, Serang Raya University, Serang, Indonesia
[2]Department of Engineering Physics, Faculty of Engineering, National University, Jakarta, Indonesia
[3]Information System, Faculty of Engineering, STTIKOM Excellent Humans, Cilegon, Indonesia
[4]Computer System Department, Faculty of Information Technology, Serang Raya University, Serang, Indonesia
[5]Information Management, Faculty of Engineering, STTIKOM Excellent Humans, Cilegon, Indonesia
[6]Informatic Department, Faculty of Communication and Informatics Technology, National University, Jakarta, Indonesia

Article Info

ABSTRACT

The limitations of medical personnel, especially heart disease, cause difficulties in diagnosing heart disorders, so diagnosing heart disorders is not easy, it takes the ability and experience of a cardiologist who has the expertise and experience to be able to accurately diagnose heart disorders. Several studies in the field of computing have been carried out in diagnosing cardiac abnormalities in patients. This study was conducted to accurately test the results of the classification of heart disorders using electrocardiogram medical record data with a C.45 decision tree approach. The results showed that the classification of heart defects obtained a mean squared error (MSE) value of 0.24, a root mean squared error (RMSE) value of 0.49, and an accuracy value of 75.33% with the C4.5 algorithm.

Corresponding Author:

Sumiati
Department of Informatic, Faculty of Information Technology, Serang Raya University
Jl. Raya Cilegon No.Km. 5, Taman, Drangong, Kec. Taktakan, Kota Serang, Banten 42162, Indonesia
Email: sumiatiunsera82@gmail.com

1. INTRODUCTION

The information technology industrial revolution 4.0 is growing very rapidly, which can be seen from various sciences such as expert systems, machine learning, fuzzy logic. Artificial intelligence applications are widely used in the world of health, where the field of artificial intelligence, especially expert systems can adopt the expertise of an expert in an application. The shortage of medical personnel, especially heart disease specialists, is the highest cause of death due to heart disease. This study aims to provide an alternative solution to solve the limitations of heart disease specialist medical personnel by using a classification model of heart disorders using computational methods. The computational method with the C4.5 Algorithm approach to diagnose heart disorders has advantages, namely the accuracy of prediction of the model's ability to predict class labels for new data, besides that it has advantages in terms of speed and efficiency of computation time.

In addition to using a computational method with a C4.5 decision tree approach to find cardiac disorders, many approaches are used to find cardiac disorders, such as the use of an expert system with a certainty factor approach that is built based on a person's expertise. which was adopted into an application

[1]–[3] diagnosis of heart disease with decision tree algorithm for prediction and classification [4]–[16], development of an android-based diagnosis system for heart disorders [17], Fuzzy Expert System for Diagnosing Heart Disease [18], authentication method in utilizing electrocardiogram (ECG) wave features [19], application of IoT for disease diagnosis [20]–[23], single lead ECG classification with approach deep learning [24]–[26], signal analysis system for ECG authentication system [27], improve the diagnosis of heart disease with particle swarm optimization (PSO) Algorithm evolutionary approach and neural network [28], improving the heart disease diagnosis by evolutionary algorithm of PSO and feed forward neural network [29]. ECG classification using the k-nearest neigbor (KNN) approach [30]. An expert system for automated identification of obstructive sleep apneafrom single-lead ECG using random under sampling boosting [31]. Cardiology expert system based on electrocardiogram data using factors with multiple rules [32], cognitive map of certainty (CCM) to assess the causality of cognitive maps using the factor certainty for heart failure [33]. The use of support vector machine (SVM) method and adaboost such as detection of ventricular fibrillation rhythm by using a supported vector machine amplified with optimal combination of variables [34]. Two novel methods for multiclass ECG arrhythmias classification based on principal component analysis (PCA), fuzzy support vector machine and unbalanced clustering [35], A multi-class ECG beat classifier based on the truncated a truncated karhunen-loeve transform (KLT) Representation [36], multi-class ECG beat classification based on a gaussian mixture model of karhunen-loève transform [37], Research on rhythmic ventricular fibrillation (VF) detected using a new approach involving support vector machine algorithm SVM, adaptive enhancement (Boost) and evolutionary differential algorithm (DE) with the help of optimal combination of variables. The end of our method is that it takes up less memory and can be implemented in real-time. In the health sector, especially diseases that are used to diagnose someone with an indication of heart disease, one of them is by looking at the patient's medical record in the form of electrocardiogram (ECG) data.

Research with the classification method using the decision tree has been done by many previous researchers. Comparative studies have been carried out to determine the decision tree model as done [38]. The decision tree approach is to determine the results of patients diagnosed with heart disease based on electrocardiogram medical record data. This research is expected to be able to make a positive contribution to strengthening doctor's confidence when diagnosing the type of heart defect correctly from the results of the electrocardiogram medical record.

This study aims to provide alternative solutions to overcome the limitations of medical personnel in the field of heart disease specialists, the data used is electrocardiogram medical record data, so this study uses a classification model of heart disorders using computational methods. Computing method with decision tree approach with C45 Algorithm. System testing 250 data as training data and 50 data used as test data. From the results of calculations and testing, the mean squared error (MSE) value is 0.24, the root mean squared error (RMSE) value is 0.49, and the accuracy value with the C4.5 algorithm is 75.33%. Medical electrocardiogram, so we need the right method to perform the data classification process accurately and ensure data validity.

2. METHOD
2.1. Decision tree C.45
Decision tree is a classification technique that is widely used in the concept of data mining [39]. A decision tree is a flow chart graphic that represents the decision-making process, where the flow chart graph resembles the shape of a tree. A decision tree can be used by a person to make difficult decisions by simplifying them into easier choices. Each decision tree has a node and a branch connecting each node (nodes).

3. ECG COMPONENTS
An electrocardiogram (EKG) is a graph of the results of the ECG medical record that is generated from the patient's heart rate. The ECG signal is a signal (time and voltage) that describes the activity of the heart during sequential contact. Patients who experience symptoms of heart disease such as chest pain, fatigue, weakness, palpitations will undergo an EKG test. This test aims to detect heart abnormalities, such as heart attacks, coronary heart disease, and evaluate the effectiveness of the pacemaker used. ECG examination has an important role in helping to diagnose heart disease, but other tests are still needed to be more certain.

4. RESEARCH METHOD
The initial stage of research is to analyze the problem. The next step is data collection, the research data is in the form of electrocardiogram medical record data, the data used is 300 patient data that has been

processed for data preparation, so the data is ready for analysis. Sixteen parameters used in this study were based on electrocardiogram medical record data. The next process is 300 electrocardiogram medical record data which is divided into two parts, 250 training data and 50 test data. The next step is to classify the cardiac abnormality model using the C.45 decision tree approach. The final stage is interpreting the results of the classification.

5.   RESULTS AND DISCUSSION

Determination of dominant (significant) attributes can be done by calculating the value of information Gain (Information gain),

$$Information\ Gain\ (S, F_j) =\ Entropy\ (S) - \sum_{V_i \in V_{F_j}} \frac{|S_{V_i}|}{|S|} . Entropy\ (S_{V_i})$$

where Vfj is the set of all possible values of an attribute Fj and S_{vi} is a subset of S, where F_j has the value v_i. Calculation of the value of Entropy can be done using the Shannon Entropy equation,

$$Entropy\ (S) =\ \sum_{i=1}^{c} -p_i . log_2(p_i)$$

where pi is the proportion of the training sample for class i.

The stages of making a decision tree are,

The first step, determine the number of positive (Normal) and negative (Abnormal) sets. In this case, the number of positive sets (Normal) is 139 while the negative set (Abnormal) is 111.

The second step, calculate the entropy value of the training sample S based on its positive and negative decisions.

$$Entropy\ (S) =\ -\frac{139}{250}log_2\left(\frac{139}{250}\right) - \frac{111}{250}log_2\left(\frac{111}{250}\right) = 0.99$$

The third step, calculate the entropy value of each attribute against entropy S. The entropy calculation is carried out in the decision attribute class. Entropy calculation for each attribute value for each calculation result is shown in Table 1.

Table 1. Entropy calculation for each attribute value

Attribute		Entropy value	Attribute		Entropy Value	Attribute		Entropy value
HR attribute	HR < 120	0.88	AXIS attribute:	AXIS < 147	0.96	II-III AVF Wave T:	Normal =169	0.96
	HR >= 120	0.2		AXIS >= 147	0.91		Abnormal = 81	0.95
P-R attribute	P-R < 268.5	0.96	RV6 attribute	RV6 < 2.05	0.95	T Wave I-AVL Lead	Normal =194	0.96
	P-R >= 268.5	1		RV6 >= 2.05	0.95		Abnormal = 56	0.95
QRS attribute	QRS < 233.5	0.96	SV1 attribute	SV1 < 0.5	0.97	T Wave V2-V6 Lead	Normal =198	0.95
	QRS >= 233.5	0.91		SV1>= 0.5	0.96		Abnormal = 52	0.96
QT attribute	QT >= 266	0.50	R+S attribute	R+S < 2.56	0.93	P Wave II_VI Lead	Normal =164	0.97
				R+S >= 2.56	0.95		Abnormal = 85	0.95

a)   Calculating the value of entropy (HR): The HR attribute has two types of value <120 as many as 198 pieces, > = 120 as many as 52 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, <120 = 198 pieces consisting of: normal = 137 Abnormal = 61

$$Entropy\ (HR < 120) =\ -\frac{137}{198}log_2\left(\frac{137}{198}\right) - \frac{61}{198}log_2\left(\frac{61}{198}\right) = 0.88$$

> = 120 = 52 pieces consisting of: normal = 2 abnormal = 50

$$Entropy\ (HR \geq 120) =\ -\frac{2}{52}log_2\left(\frac{2}{52}\right) - \frac{50}{52}log_2\left(\frac{50}{52}\right) = 0.2$$

b) Calculating the value of entropy (P-R): The P-R attribute has two types of value <268.5, as many as 247 units, > = 268.5 as many as 4 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, <268.5 = 247 pieces consisting of: normal = 137 abnormal = 109

$$\text{Entropy (PR} < 191.5) = -\frac{137}{247}\log_2\left(\frac{137}{247}\right) - \frac{109}{247}\log_2\left(\frac{109}{247}\right) = 0.96$$

> = 268.5 = 4 pieces consisting of: normal = 2 Abnormal = 2

$$\text{Entropy (PR} < 191,5) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$

c) Calculating the entropy value (QRS): The QRS attribute has two types of value <233.5 as many as 247 pieces, > = 233.5 as many as 3 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, < 233.5 = 247 pieces consisting of: normal = 137 Abnormal = 110

$$\text{Entropy (QRS} < 233.5) = -\frac{137}{247}\log_2\left(\frac{137}{247}\right) - \frac{110}{247}\log_2\left(\frac{110}{247}\right) = 0.96$$

QRS > = 233.5 = 3 pieces consisting of: normal = 2 abnormal = 1

$$\text{Entropy (QRS} >= 233.5) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.91$$

d) Calculating the value of entropy (QT): The QT attribute has two types of Value <266, 9 pieces, > = 266 as many as 3 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, > = 266 = 247 pieces consisting of: normal = 137 abnormal = 104

$$\text{Entropy (QT,} >= 266) = -\frac{137}{247}\log_2\left(\frac{137}{247}\right) - \frac{104}{247}\log_2\left(\frac{104}{247}\right) = 0.95$$

<266 = 9 pieces consisting of: normal = 2 abnormal = 7

$$\text{Entropy (QT} < 266) = -\frac{2}{7}\log_2\left(\frac{2}{7}\right) - \frac{7}{7}\log_2\left(\frac{7}{7}\right) = 0.50$$

e) Calculating the entropy value (QTC): The QTC attribute has two types of value <425.5 as many as 106 pieces, > = 425.5 as many as 3 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, <425.5 = 106 pieces consisting of: normal = 62 abnormal = 43

$$\text{Entropy (QTC} < 425.5) = -\frac{62}{106}\log_2\left(\frac{62}{106}\right) - \frac{43}{106}\log_2\left(\frac{43}{106}\right) = 0.96$$

> = 425.5 = 146 pieces consisting of: normal = 78 abnormal = 68

$$\text{Entropy (QTC} >= 425.5) = -\frac{78}{146}\log_2\left(\frac{78}{146}\right) - \frac{68}{146}\log_2\left(\frac{68}{146}\right) = 0.97$$

f) Calculating the value of entropy (AXIS): The AXIS attribute has two types of value <147, 247, > = 147 as many as 3 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, < 147 = 247 pieces consisting of: normal = 138 abnormal = 109

$$\text{Entropy (AXIS} < 147) = -\frac{138}{247}\log_2\left(\frac{138}{247}\right) - \frac{109}{247}\log_2\left(\frac{109}{247}\right) = 0.96$$

> = 147 = 3 pieces consisting of: normal = 1 abnormal = 2

$$\text{Entropy (AXIS} >= 147) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.91$$

g) Calculating the entropy RV6 value: The attribute RV6 has two types of value <2.05, as many as 213 pieces, > = 2.05 as many as 40 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, <2.05 = 213 pieces consisting of: normal = 124 abnormal = 89

$$\text{Entropy (RV6} < 2.05) = -\frac{124}{213}\log_2\left(\frac{124}{213}\right) - \frac{89}{213}\log_2\left(\frac{89}{213}\right) = 0.95$$

$> = 2.05 = 37$ pieces consisting of: normal = 15 abnormal = 22

$$\text{Entropy (RV6} >= 2.05) = -\frac{15}{37}\log_2\left(\frac{15}{37}\right) - \frac{22}{37}\log_2\left(\frac{22}{37}\right) = 0.95$$

h) Calculating the value of entropy SV1: The SV1 attribute has two types of value <0.5, as many as 92 pieces, > = 0.5 as many as 158 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, <0.5 = 92 pieces consisting of: normal = 48 abnormal = 44

$$\text{Entropy (SV1} < 0.5) = -\frac{48}{92}\log_2\left(\frac{48}{92}\right) - \frac{44}{92}\log_2\left(\frac{44}{92}\right) = 0.97$$

$> = 0.5 = 158$ pieces consisting of: normal = 92 abnormal = 66

$$\text{Entropy (SV1} >= 0.55) = -\frac{92}{158}\log_2\left(\frac{92}{158}\right) - \frac{66}{158}\log_2\left(\frac{66}{158}\right) = 0.96$$

i) Calculating the entropy R + S value: The attribute R + S has two types of value <2.56 totaling 188 pieces, > = 2.56 totaling 62 pieces. Thus, the calculation of the entropy value for each attribute value is obtained, <2.56 = 188 pieces consisting of: normal = 113 abnormal = 75

$$\text{Entropy (R + S} < 2.56) = -\frac{113}{188}\log_2\left(\frac{113}{188}\right) - \frac{75}{188}\log_2\left(\frac{75}{188}\right) = 0.93$$

$> = 2.56 = 62$ pieces consisting of: normal = 36 abnormal = 26

$$\text{Entropy (R + S} >= 2.56) = -\frac{36}{62}\log_2\left(\frac{36}{62}\right) - \frac{26}{62}\log_2\left(\frac{26}{62}\right) = 0.95$$

j) Calculating the value of Entropy Lead II VI P Wave: Attribute Lead II VI P wave has two types of value, namely normal as many as 165 pieces, abnormal 85 as many as fruit. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 164 pieces consisting of: normal to normal = 90 normal to abnormal = 74

$$\text{Entropy (Normal)} = -\frac{90}{164}\log_2\left(\frac{90}{164}\right) - \frac{74}{164}\log_2\left(\frac{74}{164}\right) = 0.97$$

Abnormal = 85 fruit consisting of: normal to abnormal = 50 abnormal to abnormal = 35

$$\text{Entropy (Abnormal)} = -\frac{50}{85}\log_2\left(\frac{50}{85}\right) - \frac{35}{85}\log_2\left(\frac{35}{85}\right) = 0.95$$

k) Calculating the value of entropy II_III_AVF wave T: Attribute II_III_AVF T wave has two types of value, namely normal as many as 169 fruits, abnormal 81 as many as fruit. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 169 pieces consisting of: normal to normal = 89 normal to abnormal = 80

$$\text{Entropy (Normal)} = -\frac{89}{169}\log_2\left(\frac{89}{169}\right) - \frac{80}{169}\log_2\left(\frac{80}{169}\right) = 0.96$$

Abnormal = 81 pieces consisting of abnormal to normal = 50 abnormal to abnormal = 31

$$\text{Entropy (Abnormal)} = -\frac{50}{81}\log_2\left(\frac{50}{81}\right) - \frac{31}{81}\log_2\left(\frac{31}{81}\right) = 0.95$$

l) Calculating the T wave, I_AVL entropy lead value: Attribute lead I_AVL T wave has two types of value, namely normal as many as 194 pieces, abnormal 56 as many as fruit. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 194 pieces consisting of: normal to normal = 116 normal to abnormal = 78

$$\text{Entropy (Normal)} = -\frac{116}{194}\log_2\left(\frac{116}{194}\right) - \frac{78}{194}\log_2\left(\frac{78}{194}\right) = 0.96$$

Abnormal = 56 pieces consisting of abnormal to normal = 23 abnormal to abnormal = 33

$$\text{Entropy (Abnormal)} = -\frac{23}{56}\log_2\left(\frac{23}{56}\right) - \frac{33}{56}\log_2\left(\frac{33}{56}\right) = 0.95$$

m) Calculating the value of entropy Lead V2-V6 Wave T: Lead V2-V6 T waves have two types of value, namely 198 normal and 52 abnormal. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 198 pieces consisting of: normal to normal = 118 normal to abnormal = 80

$$\text{Entropy (Normal)} = -\frac{118}{198}\log_2\left(\frac{118}{198}\right) - \frac{80}{198}\log_2\left(\frac{80}{198}\right) = 0.95$$

Abnormal = 52 pieces consisting of abnormal to normal = 20 abnormal to abnormal = 32

$$\text{Entropy (Abnormal)} = -\frac{20}{52}\log_2\left(\frac{20}{52}\right) - \frac{32}{52}\log_2\left(\frac{32}{52}\right) = 0.96$$

n) Calculating the value of entropy Lead II_III_AVF segment ST: Lead II_III_AVF segment ST has two types of value, namely 248 normal and 2 abnormal. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 248 pieces consisting of: normal to normal = 138 normal to abnormal = 110

$$\text{Entropy (Normal)} = -\frac{138}{248}\log_2\left(\frac{138}{248}\right) - \frac{110}{248}\log_2\left(\frac{110}{248}\right) = 0.96$$

Abnormal = 2 pieces consisting of abnormal to normal = 1 abnormal to abnormal = 1

$$\text{Entropy (Abnormal)} = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$$

o) Calculating the value of entropy Lead I_AVL Segment ST: The I_AVL lead segment ST has two types of value, namely 248 normal and 2 abnormal. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 248 pieces consisting of: normal to normal = 138 normal to abnormal = 110

$$\text{Entropy (Normal)} = -\frac{138}{248}\log_2\left(\frac{138}{248}\right) - \frac{110}{248}\log_2\left(\frac{110}{248}\right) = 0.96$$

Abnormal = 2 pieces consisting of abnormal to normal = 1 abnormal to abnormal = 1

$$\text{Entropy (Abnormal)} = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$$

p) Calculating the value of entropy V2-V6 segment ST: V2-V6 segment ST has two types of value, namely 241 normal and 9 abnormal. Thus, the calculation of the entropy value for each attribute value is obtained, Normal = 241 pieces consisting of: normal to normal = 134 normal to abnormal = 107

$$\text{Entropy (Normal)} = -\frac{134}{241}\log_2\left(\frac{134}{241}\right) - \frac{107}{241}\log_2\left(\frac{107}{241}\right) = 0.96$$

abnormal = 9 pieces consisting of abnormal to normal = 5 abnormal to abnormal = 4

$$\text{Entropy (Abnormal)} = -\frac{5}{9}\log_2\left(\frac{5}{9}\right) - \frac{4}{9}\log_2\left(\frac{4}{9}\right) = 0.96$$

The fourth step, calculate the information gain value for each attribute to determine which attribute is more dominant and will serve as the root tree. The results of the calculation of gain information for each attribute are obtained,

$$\text{Information Gain}\,(S, F_j) = \text{Entropy}\,(S) - \sum_{V_i \in V_{F_j}} \frac{|S_{V_i}|}{|S|} \times \text{Entropy}\,(S_{V_i})$$

HR attribute:Information Gain $(S, \text{HR}) = 0.97 - \left\{\frac{198}{250} \times 0.88 + \frac{52}{250} \times 0.2\right\} = 0.24$

P-R attribute:Information Gain $(S, \text{P} - \text{R}) = 0.97 - \left\{\frac{247}{250} \times 0.96 + \frac{4}{250} \times 1\right\} = 0.01$

QRS attribute:Information Gain $(S, \text{QRS}) = 0.97 - \left\{\frac{247}{250} \times 0.96 + \frac{3}{250} \times 0.91\right\} = 0.02$

QT attribute:Information Gain $(S, \text{QT}) = 0.97 - \left\{\frac{247}{250} \times 0.95 + \frac{9}{250} \times 0.50\right\} = 0.01$

QTC attribute:Information Gain $(S, \text{QTC}) = 0.97 - \left\{\frac{106}{250} \times 0.96 + \frac{146}{250} \times 0.97\right\} = 0.01$

AXIS attribute:Information Gain $(S, \text{AXIS}) = 0.97 - \left\{\frac{247}{250} \times 0.96 + \frac{3}{250} \times 0.91\right\} = 0.01$

RV6 attribute:Information Gain $(S, \text{RV6}) = 0.97 - \left\{\frac{213}{250} \times 0.95 + \frac{37}{250} \times 0.95\right\} = 0.02$

SV1 attribute:Information Gain $(S, \text{SV1}) = 0.97 - \left\{\frac{92}{250} \times 0.97 + \frac{158}{250} \times 0.96\right\} = 0.01$

R+S attribute:Information Gain $(S, \text{R} + \text{S}) = 0.97 - \left\{\frac{188}{250} \times 0.93 + \frac{62}{250} \times 0.95\right\} = 0.04$

P Wave II_VI Lead Attribute: $(S, \text{P Wave II\_VI Lead}) = 0.97 - \left\{\frac{164}{250} \times 0.97 + \frac{85}{250} \times 0.95\right\} = 0.01$

T Wave AVF Lead II_IIIAttribute: $(S, \text{T Wave AVF Lead II\_III:}) = 0.97 - \left\{\frac{169}{250} \times 0.96 + \frac{81}{250} \times 0.95\right\} = 0.01$

T Wave Lead I-AVL Attribute: $(S, \text{T Wave I} - \text{AVL Lead}) = 0.97 - \left\{\frac{194}{250} \times 0.96 + \frac{56}{250} \times 0.95\right\} = 0.02$

T Wave V2-V6 Lead Attribute: $(S, \text{T Wave V2} - \text{V6 Leads}) = 0.97 - \left\{\frac{198}{250} \times 0.95 + \frac{52}{250} \times 0.96\right\} = 0.03$

Attribute Lead II-III AVF Segment ST: $(S, \text{Lead II} - \text{III AVF Segment ST}) = 0.97 - \left\{\frac{248}{250} \times 0.96 + \frac{2}{250} \times 1\right\} = 0.01$

Lead I-AVL Segment ST Attributes:$(S, \text{Lead I} - \text{AVL Segment ST}) = 0.97 - \left\{\frac{248}{250} \times 0.96 + \frac{2}{250} \times 1\right\} = 0.012$

Attribute V2-V6 Segment ST: $(S, \text{V2} - \text{V6 Segment ST}) = 0.97 - \left\{\frac{241}{250} \times 0.96 + \frac{9}{250} \times 0.96\right\} = 0.02$

The fifth step, derive a subset for the root tree based on the attribute values of the attributes selected as the root tree. In this step, the attributes that will be used as root are selected based on the highest gain information value. The gain information for each attribute from the first calculation is shown in Table 2.

Table 2. Information on the gain of each yield attribute

Attribute	Gain information	Attribute	Gain information	Attribute	Gain information
HR attribute	0.24	AXIS attribute:	0.01	II_III AVF Wave T:	0.01
P-R attribute	0.01	RV6 attribute	0.02	T Wave I-AVL Lead	0.04
QRS attribute	0.02	SV1 attribute	0.01	T Wave V2-V6 Leads	0.01
QT attribute	0.01	R+S attribute	0.01	Lead II-III AVF Segment ST	0.01
QTC attribute	0.01	P Wave II_VI Lead	0.02	Attribute V2-V6 Segment ST	0.02

Based on the calculation of gain information, it can be seen that the highest gain value is owned by the HR attribute, so that the HR attribute is used as the root. Meanwhile, the decision node is seen based on the attribute value of HR. The initial shape of the decision tree is based on the information gain value as shown in Figure 1.

The calculation is continued until a complete decision tree is formed so that it represents the pattern of classification of heart disease data according to available ECG data. Figure 2 overall decision tree. The calculation for the formation of a decision tree in more detail is shown in Figure 2.
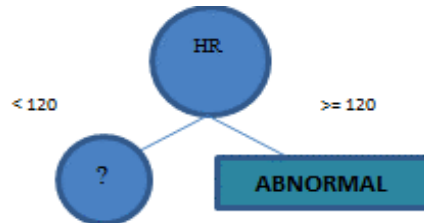
Figure 1. Initial decision tree
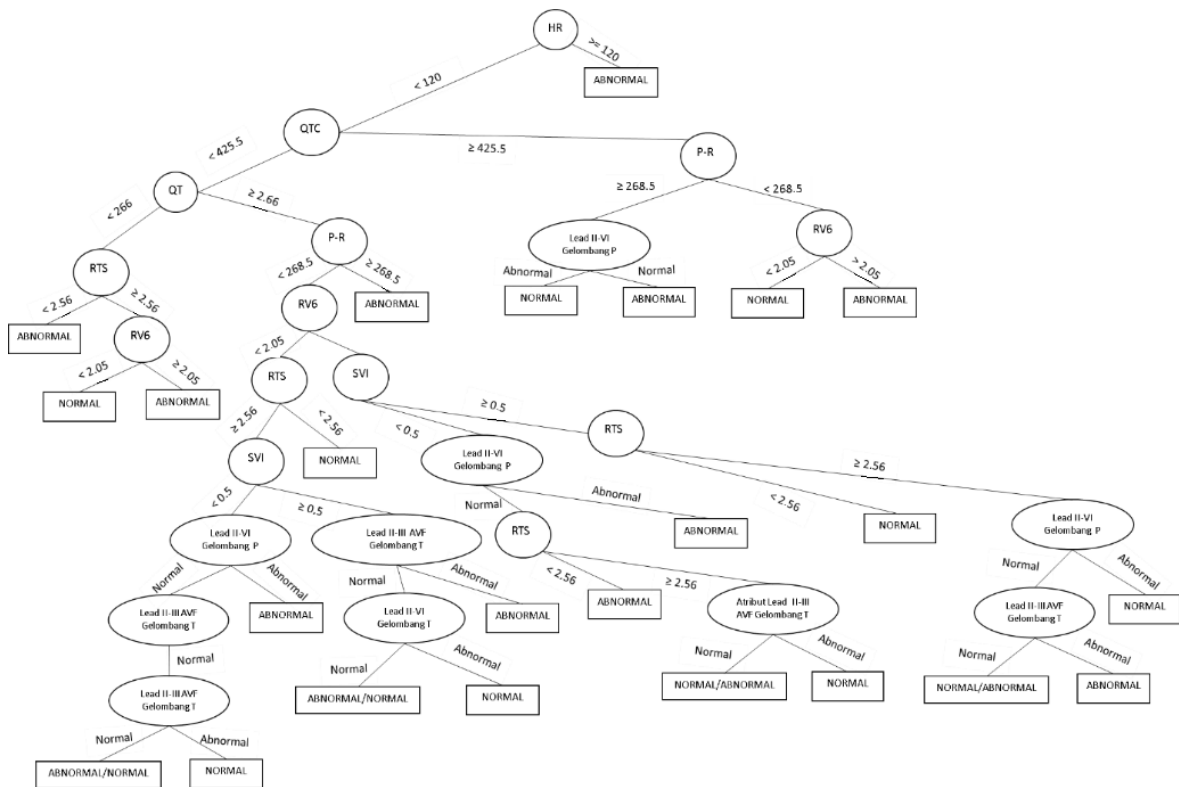
Figure 2. Overall decision tree

5.1. Results evaluation

C4.5 data mining algorithm is used to classify and predictive. The C4.5 decision tree is easy to interpret and the fastest compared to other algorithms. Evaluation of results is assisted by Python language. The following results from the python language are shown in Table 3.

Table 3. Result evaluation

Mean square error (MSE)	Root means square error (RMSE)	Success rate
0.24	0.49	75.33 %

6.   CONCLUSION

The pattern of classification rules formed from training data is influenced by the distribution of data referring to cardiac abnormalities. The more data distribution, the better the classification rule pattern extraction will be, conversely, if the less data distribution, then the classification rule pattern extraction is not formed well. In this study, the data were tested using computational methods. The method uses the C4.5 decision tree approach. Tests were carried out from the results of system validity, system testing 250 data as training data and 50 data used as test data. From the results of calculations and testing, the MSE value is 0.24, the RMSE value is 0.49, and the accuracy value with the C4.5 algorithm is 75.33%.

REFERENCES

[1]   T. Lindow, O. Pahlm, and K. Nikus, "A patient with non-ST-segment elevation acute coronary syndrome: Is it possible to predict the culprit coronary artery?," *Journal of Electrocardiology*, vol. 49, no. 4, pp. 614–619, Jul. 2016, doi: 10.1016/j.jelectrocard.2016.05.001.

[2]   I. Fernández-Varela, E. Hernández-Pereira, D. Álvarez-Estévez, and V. Moret-Bonillo, "Combining machine learning models for the automatic detection of EEG arousals," *Neurocomputing*, vol. 268, pp. 100–108, Dec. 2017, doi: 10.1016/j.neucom.2016.11.086.

[3]   M. Al-Ani and A. A. Ayal, "A rule-based system for automated ECG diagnosis," *International Journal of Advances in Engineering & Technology, IJAET*, 2013.

[4]   M. Abdar, "Using decision trees in data mining for predicting factors influencing of heart disease," *Journal of Electronic and Computer Engineering*, vol. 8, no. 2, pp. 31–36, 2015.

[5]   A. Hamoud, "Selection of best decision tree algorithm for prediction and classification of student," *Action American International Journal of Research in Science, Technology, Engineering & Mathematics*, vol. 16, no. 1, pp. 26–32, 2016.

[6]   N. A. Setiawan *et al.*, "Classification of arrhythmia's ECG signal using cascade transparent classifier," *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 2, pp. 1015–1025, Jan. 2022, doi: 10.3233/JIFS-189768.

[7]   M. Mohanty, A. Subudhi, and M. N. Mohanty, "Detection of supraventricular tachycardia using decision tree model," *International Journal of Computer Applications in Technology*, vol. 65, no. 4, p. 378, 2021, doi: 10.1504/IJCAT.2021.117285.

[8]   S. Ketu and P. K. Mishra, "Empirical analysis of machine learning algorithms on imbalance electrocardiogram based arrhythmia dataset for heart disease detection," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 1447–1469, Feb. 2022, doi: 10.1007/s13369-021-05972-2.

[9]   T. A. Munandar, V. Rosalina, and Sumiati, "Cardiovascular disease classification based on echocardiography and electrocardiogram data using the decision tree classification approach," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 1, pp. 50–59, 2020.

[10] A. Joshi, E. J. Dangra, and M. K. Rawa, "A decision tree based classification technique for accurate heart disease classification & prediction," *International Journal of Technology Research and Management*, vol. 3, no. 11, pp. 1–4, 2016.

[11] A. S. Karthiga, S. Mary, and M. Yogasini, "Early prediction of heart disease using decision tree algorithm," *International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)*, vol. 3, no. 3, pp. 1–17, 2017.

[12] S. Vachiravel, "Diagnosis of heart disease using decision tree," *International Journal of Research in Computer Applications & Information Technology*, vol. 2, no. 6, pp. 74–79, 2014.

[13] Z. Mašetic and A. Subasi, "Detection of congestive heart failures using C4.5 Decision Tree," *Southeast Europe Journal of Soft Computing*, vol. 2, no. 2, 2013, doi: 10.21533/scjournal.v2i2.32.

[14] S. Sahoo, A. Subudhi, M. Dash, and S. Sabut, "Automatic classification of cardiac arrhythmias based on hybrid features and decision tree algorithm," *International Journal of Automation and Computing*, vol. 17, no. 4, pp. 551–561, Aug. 2020, doi: 10.1007/s11633-019-1219-2.

[15] M. MOHANTY, P. BISWAL, and S. SABUT, "Ventricular tachycardia and fibrillation detection using DWT and decision tree classifier," *Journal of Mechanics in Medicine and Biology*, vol. 19, no. 03, p. 1950008, May 2019, doi: 10.1142/S0219519419500088.

[16] M. Mohanty, S. Sahoo, P. Biswal, and S. Sabut, "Efficient classification of ventricular arrhythmias using feature selection and C4.5 classifier," *Biomedical Signal Processing and Control*, vol. 44, pp. 200–208, Jul. 2018, doi: 10.1016/j.bspc.2018.04.005.

[17] Sumiati and H. Triono Sigit, "Design of android application for telemedicine system to improve public health," *MATEC Web of Conferences*, vol. 218, p. 3005, Oct. 2018, doi: 10.1051/matecconf/201821803005.

[18] A. Adeli and M. Neshat, "A fuzzy expert system for heart disease diagnosis," 2010.

[19] A. Y. Shdefat, M.-I. Joo, S.-H. Choi, and H.-C. Kim, "Utilizing ECG waveform features as new biometric authentication method," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 2, p. 658, Apr. 2018, doi: 10.11591/ijece.v8i2.pp658-665.

[20] S. Abbas, "An innovative IoT service for medical diagnosis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, p. 4918, Oct. 2020, doi: 10.11591/ijece.v10i5.pp4918-4927.

[21] O. Vermesan and P. Friess, *Digitizing the industry internet of things connecting the physical, digital and virtual worlds*. River Publishers, 2016.

[22] P. Chatterjee and R. L. Armentano, "Internet of things for a smart and ubiquitous ehealth system," in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Dec. 2015, pp. 903–907, doi: 10.1109/CICN.2015.178.

[23] P. Verma and S. K. Sood, "Cloud-centric IoT based disease diagnosis healthcare framework," *Journal of Parallel and Distributed Computing*, vol. 116, pp. 27–38, Jun. 2018, doi: 10.1016/j.jpdc.2017.11.018.

[24] D. Magrì *et al.*, "QT spatial dispersion and sudden cardiac death in hypertrophic cardiomyopathy: Time for reappraisal," *Journal of Cardiology*, vol. 70, no. 4, pp. 310–315, Oct. 2017, doi: 10.1016/j.jjcc.2017.01.006.

[25] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Computers in Biology and Medicine*, vol. 99, pp. 53–62, Aug. 2018, doi: 10.1016/j.compbiomed.2018.05.013.

[26] A. M. Abdu, M. M. M. Mokji, and U. U. U. Sheikh, "Machine learning for plant disease detection: an investigative comparison between support vector machine and deep learning," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, p. 670, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp670-683.

[27] S. J. Kang, S. Y. Lee, H. Il Cho, and H. Park, "ECG authentication system design based on signal analysis in mobile and wearable devices," *IEEE Signal Processing Letters*, vol. 23, no. 6, pp. 805–808, Jun. 2016, doi: 10.1109/LSP.2016.2531996.

[28] K. Singh, A. Singhvi, and V. Pathangay, "Dry contact fingertip ECG-based authentication system using time, frequency domain features and support vector machine," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 526–529, doi: 10.1109/EMBC.2015.7318415.

[29] M. G. Feshki and O. S. Shijani, "Improving the heart disease diagnosis by evolutionary algorithm of PSO and feed forward neural network," in *2016 Artificial Intelligence and Robotics (IRANOPEN)*, Apr. 2016, pp. 48–53, doi: 10.1109/RIOS.2016.7529489.

[30] S. Faziludeen and P. Sankaran, "ECG beat classification using evidential k-nearest neighbour," *Procedia Computer Science*, vol. 89, pp. 499–505, 2016, doi: 10.1016/j.procs.2016.06.106.

[31] A. R. Hassan and M. A. Haque, "An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting," *Neurocomputing*, vol. 235, pp. 122–130, Apr. 2017, doi: 10.1016/j.neucom.2016.12.062.

[32] S. Sumiati, H. Saragih, T. Abdul Rahman, and A. Triayudi, "Expert system for heart disease based on electrocardiogram data using certainty factor with multiple rule," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, p. 43, Mar. 2021, doi: 10.11591/ijai.v10i1.pp43-50.

[33] Sumiati *et al.*, "Certainty cognitive map (CCM) for assessing cognitive map causality using cartainty factors for cardiac failure," *ICIC Express Letters*, vol. 15, no. 1, pp. 27–36, 2021.

[34] D. Panigrahy, P. K. Sahu, and F. Albu, "Detection of ventricular fibrillation rhythm by using boosted support vector machine with an optimal variable combination," *Computers & Electrical Engineering*, vol. 91, p. 107035, May 2021, doi: 10.1016/j.compeleceng.2021.107035.

[35] M. C. Nait-Hamoud and A. Moussaoui, "Two novel methods for multiclass ECG arrhythmias classification based on PCA, fuzzy support vector machine and unbalanced clustering," in *2010 International Conference on Machine and Web Intelligence*, Oct. 2010, pp. 140–145, doi: 10.1109/ICMWI.2010.5647931.

[36] G. Biagetti, P. Crippa, A. Curzi, S. Orcioni, and C. Turchetti, "A multi-class ECG beat classifier based on the truncated KLT representation," in *2014 European Modelling Symposium*, Oct. 2014, pp. 93–98, doi: 10.1109/EMS.2014.31.

[37] P. Crippa, A. Curzi, L. Falaschetti, and C. Turchetti, "Multi-class ECG beat classification based on a gaussian mixture model of karhunen-loève transform," *International Journal of Simulation Systems Science & Technology*, Feb. 2020, doi: 10.5013/IJSSST.a.16.01.02.

[38] J. S. Lee and E. S. Lee, "Exploring the usefulness of a decision tree in predicting people's locations," *Procedia - Social and Behavioral Sciences*, vol. 140, pp. 447–451, Aug. 2014, doi: 10.1016/j.sbspro.2014.04.451.

[39] E. Paryad, L. R. Balasi, E. Kazemnejad, and S. Booraki, "Predictors of illness perception in patients undergoing coronary artery bypass surgery," *Journal of Cardiovascular Disease Research*, vol. 8, no. 1, pp. 16–18, Mar. 2017, doi: 10.5530/jcdr.2017.1.3.

BIOGRAPHIES OF AUTHORS

Sumiati, ST., MM., Ph. D 🆔 8️⃣ SC ◐ is a lecturer and researcher at Informatic Department, Universitas Serang Raya, Indonesia. Strong educational background in Artificial Intelligence, Data Mining. Currently developing research in the field of medicine combined with Artificial Intelligence. She can be contected at email: sumiatiunsera82@gmail.com.

Dr. Viktor Vekky Ronald Repi 🆔 8️⃣ SC ◐ is a lecturer and researcher at the Department of Engineering Physics at the Universitas Nasional Jakarta. Strong educational background in Engineering Physics. The field of research that has been occupied is materials for energy, synthesis, characterization of materials using artificial intelligence. Currently developing research in the field of medicine combine with Artificial Intelligence. He can be contected at email: vekky_repi@civitas.unas.ac.id.

Penny Hendriyati, M. Kom. 🆔 8️⃣ SC ◐ She holds a master's degree in information technology at STMIK Eresha Jakarta in 2014. Currently dedicating himself as a lecturer at College of Computer schience Technology Insan Unggul (STTIKOM Insan Unggul) in the information system study program. The Subject taught are Management Information System, Human, And Computer Interaction, Computers and Society. He is also interested in Research in the field of decision support system, Data Science, and Artificial Intelligence. She can be contected at email: pennyhendriyati@gmail.com.

Anharudin, M. Kom. ⓘ 🅖 ꜱᴄ ↻ He holds a master's degree in information systems at STMIK LIKMI Bandung in 2015. Currently dedicating himself as a lecturer at Serang Raya University in the computer system study program. The Subject taught are Algorithm and Programming, Web Programming, and Android Programming. He is also interested in Research in the field of Data Science, web programming, and mobile programming. He can be contacted at email: anhar.dean@gmail.com.

Afrasim Yusta, M. Kom. ⓘ 🅖 ꜱᴄ ↻ He holds a master's degree in information technology at STMIK Eresha Jakarta in 2014. Currently dedicating himself as a lecturer at STTIKOM Insan Unggul in the information management. The Subject taught are analysis of system design, computer application, accounting information system, management information system and computer architecture. He Also interested in Research in the field od Data Science, computer application and artificial intelligence. He can be contected at email: yafras777@gmail.com.

Assoc. Prof. Agung Triayudi, S. Kom., M. Kom., Ph. D ⓘ 🅖 ꜱᴄ ↻ is a lecturer and researcher at the Information Systems Department, Faculty of Communication and Information Technology, Indonesia. Strong educational background in Educational Data Mining. He can be contacted at email: agungtriayudi@civitas.unas.ac.id.