# CLASSIFICATION OF ELECTROCARDIOGRAM (ECG) WAVES OF HEART DISEASE USING THE XGBOOST METODE METHOD

**Serly Yunarti Butarbutar, Christian Deniro Napitupulu, Nessa Sanjaya Ginting, Evta Indra, Delima Sitanggang***

Program Studi Sistem Informasi Falkultas Teknologi Dan Ilmu Komputer Unieversitas Prima Indonesia

delimasitanggang@unprimdn.ac.id

**Abstract**

| Article Info | In life, the heart is always in good condition because it works to cope with the blood it carries throughout the body. Impaired heart function can be fatal to human health, even heart problems can cause death. In previous studies, to detect abnormalities or changes in the work of the heart, it is necessary to first know the pattern of the same heart rate. It is important to detect cardiovascular disease as early as possible for counseling and drug management. One of the preventions is to perform early detection of Electro Cardiography (EKG) wave analysis. This research applies machine learning can help make this diagnostic process more efficient. In this study, machine learning can accurately classify abnormal ECG up to a maximum of 97, |
|---|---|
| Received, 01 Juni 2022 Revised 20 Juni 2022 Accepted 22 Juni 2022 | |

Keywords: Machine learning, Xgboost, Electrocardiogram, Data Mining

## 1.    INTRODUCTION

The World Health Organization (WHO) 1948 defined health as "a state of physical, mental and social well-being and not merely disease or infirmity. The heart is the main organ of the cardiovascular system (heart rhythm disorders) and one of the vital organs of the human body. In life, the heart is always in good condition because it works to cope with the blood it carries throughout the body. Impaired heart function can be fatal to human health, even heart problems can cause death. In previous studies, to detect abnormalities or changes in the work of the heart, it is necessary to first know the pattern of the same heart rate. [1].

Cardiovascular disease remains a major threat worldwide. Data from the Global Burden of Cardiovascular Disease show that there were 271 million cases of cardiovascular disease in 1990 and almost doubled to 523 million in 2019 [2]. In the last 70 years, cases of diseases with cardiovascular disorders continue to occupy the highest rank. According to data from the World Health Organization, about 33 percent of the causes of death worldwide are cardiovascular disease. An estimated 18.6 million people die from cardiovascular disease, 75 percent of which are heart attacks and strokes. More than three-quarters of deaths from cardiovascular disease occur in developing and low- and middle-income countries.Cardiovascular disease can strike anyone, especially in the current situation. Increased air pollution and unhealthy lifestyle trends increase the likelihood of developing cardiovascular disease [4]. In this case, early detection is the right step to reduce risk.        An    electrocardiogram (ECG) is    a biological signal that results from the electrical activity of the heart. This signal appears to trigger the contraction of the heart muscle which will pump blood throughout the body. The ECG signal is a representation of the level of a person's heart health which is described through the rhythm, shape, and orientation of the ECG signal. [5].

It is important to detect cardiovascular disease as early as possible for counseling and drug management. One of the preventions is to perform early detection of Electro Cardiography (EKG) wave analysis. Electrocardiograph (ECG) is a device used to capture and record changes in heart potential with the help of cables attached to the patient's body at certain locations. The results of the ECG examination are in the form of an image called an electrocardiogram [6].

In accordance with a previous study entitled The Use of Artificial Neural Networks on Electrocardiogram Signals to Detect Supraventricular Arrhythmic Heart Disease. The ANN structure used for the classification of the ECG signal is a Feed Forward Network (FFN), specifically the Multi-Layer Perceptron (MLP) (Marius-Constantin et al. using 2 hidden layers used with the Tansig activation function with Levenberg-Marquardt training function which produces a precision of 84.3% [7] And research on ECG Signal Recognition Using Wavelet Packet Decomposition and K-Means-Clustering. The ECG signal in this study is decomposed to level 5 using Daubechies 2 wavelet, then the energy of each subband is calculated for the periodogram. As a classifier, the K-Means group is used which uses the cityblock method to calculate the distance between data. From the 3 data classes tested, normal sinus rate (NSR), congestive heart failure (CHF) and atrial fibrillation (FA) obtained a precision of 94.4%. [8]. So that more accurate accuracy is needed in detecting ECG waves in classifying cardiovascular disease with other methods. The author is interested in conducting a cardiovascular classification study with the title "Classification of Electrocardiogram (ECG) Waves of Heart Disease Using the XGBoost Method". So that more accurate accuracy is needed in detecting ECG waves in classifying cardiovascular disease with other methods. The author is interested in conducting a cardiovascular classification study with the title "Classification of Electrocardiogram (ECG) Waves of Heart Disease Using the XGBoost Method". So that more accurate accuracy is needed in detecting ECG waves in classifying cardiovascular disease with other methods. The author is interested in conducting a cardiovascular classification study with the title "Classification of Electrocardiogram (ECG) Waves of Heart Disease Using the XGBoost Method".

## 2.METHOD
### 2.1 Types of Research

It is important to detect it as early as possible so that treatment advice can be offered immediately. One of the prevention is through early detection of electrocardiographic wave (ECG). A diagnostic tool in the form of an electrocardiogram (ECG) can record the human heart rate. Then, the doctor analyzes the results of the heart rate to determine the patient's heart condition and problems. Heart rate is shown in the form of a sinusoidal curve. The ECG curve consists of P, Q, R, S, and T curves. The P curve represents atrial depolarization, the QRS complex curve represents ventricular depolarization, and the T curve represents ventricular repolarization. A normal sinusoidal rhythm has the following characteristics: [9]:

1. Regular rhythm
2. Heart rate ranges from 60-100 x/min
3. The P curve is normal (width < 0.12 s, height < 0.3 mv, positive in lead II and negative in lead aVR) and is followed by a QRS and T complex curve
4. Normal QRS curve (between 0.06 to less than 0.12 seconds)
5. Normal PR curve interval (0.12 - 0.20 sec)

One solution to overcome this is to use machine learning technology so that the initial diagnostic process can be carried out more quickly and easily. One method that can be done is to apply a machine learning algorithm to identify abnormal ECG waves. In this study, using Xgboost, the best results of this method will be sought to minimize early diagnostic errors.

XGBoost is a gradient-enhanced algorithm that is enhanced in line with decision trees and can build driven trees efficiently and run in parallel [10]. XGBoost is a machine learning technique for

solving regression and classification problems based on a gradient augmenter decision tree (GBDT). In the regression tree, the inner node represents the attribute test value and the leaf node with the score represents the decision.[11].

## 2.2 Steps of Research Work

It takes a flow or flowchart so that the research runs well, is completed on time, and does not go out of bounds. The research flow flow chart can be seen in figure 1:
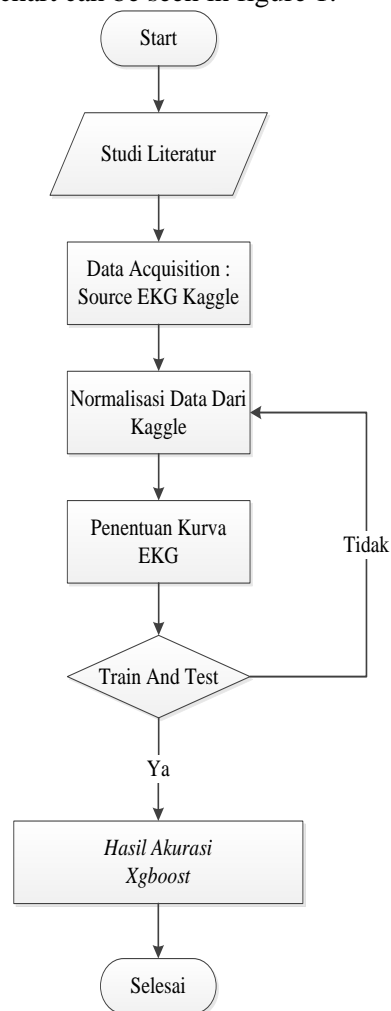


Figure 1. Flowchart

1. Study of literature

The library study method is an activity related to library data collection, reading and taking notes, and material management. This technique is carried out to reveal various theories based on the problems studied as reference material in the discussion of research results [12].

2. Data Acquisition

Data acquisition, abbreviated as DAS or DAQ, is a tool or process used to gather information to analyze physical or electrical phenomena, such as voltage, current, temperature, pressure, or sound. [13]. The process of retrieving datasets from Kaggle. This dataset is used in exploring heart rate classification. The signal corresponds to an electrocardiogram (ECG) heartbeat to detect abnormalities in the heart.

3.  Kaggle Data Normalization

Data normalization is the process of making variables that have the same range of values, nothing is too large or too small to facilitate analysis. At this stage, data that is not needed in the study is removed.

4.  ECG Curve Determination

The ECG curve consists of P, Q, R, S and T curves. The P curve represents atrial depolarization, the QRS complex curve represents ventricular depolarization, and the T curve represents ventricular repolarization.

5.  Train and Test

The process of training and testing on the Extreme Gradient Boosting (XGBoost) method to get accurate results in detecting heart abnormalities. In this study, to conduct training and testing using the Python Google Colab programming.

6.  Result Accuracy

The final result is machine learning data processing accuracy using the Xgboost method to classify heart rate electrocardiogram waves.

## 3.RESULTS AND DISCUSSION
### 3.1 Problem Analysis

Identification of ECG signals is carried out by experts to avoid errors in diagnosing heart disease. Because of this, the diagnostic process is quite time consuming. Not to mention the long queues at the hospital because many people panic so that if they experience a little complaint they immediately flock to the hospital.

### 3.2 Data Analysis

Early detection of cardiovascular disease can be done by analyzing ECG waves, in this case Machine learning can help make early detection easier and faster. To do this Machine learning requires a dataset to train the network. The dataset used in this study was sourced from [14].

The dataset is in the form of numerical ECG signal data taken from 87554 subjects consisting of 72471 normal, 2223 supraventricular ectopic disorders, 5788 ventricular ectopic disorders, 641 fusion disorders and 6431 unknown disorders.

| | 9.78E-01 | 9.26E-01 | 6.81E-01 | 2.45E-01 | 1.54E-01 | 1.91E-01 | 1.52E-01 | 8.58E-02 | 5.88E-02 | 4.90E-02 | 4.41E-02 | 6.13E-02 | 6.62E-02 | 6.13E-02.1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.960 | 0.863 | 0.462 | 0.1970 | 0.0940 | 0.1250 | 0.0997 | 0.08830 | 0.0741 | 0.0826 | 0.0741 | 0.0627 | 0.0655 | 0.0655 |
| 1 | 1.000 | 0.659 | 0.186 | 0.0703 | 0.0703 | 0.0595 | 0.0568 | 0.04320 | 0.0541 | 0.0459 | 0.0568 | 0.0432 | 0.0676 | 0.0622 |
| 2 | 0.925 | 0.666 | 0.541 | 0.2760 | 0.1960 | 0.0773 | 0.0718 | 0.06080 | 0.0663 | 0.0580 | 0.0635 | 0.0829 | 0.0856 | 0.0939 |
| 3 | 0.967 | 1.000 | 0.831 | 0.5870 | 0.3570 | 0.2490 | 0.1460 | 0.08920 | 0.1170 | 0.1500 | 0.1310 | 0.1220 | 0.1410 | 0.1600 |
| 4 | 0.927 | 1.000 | 0.627 | 0.1930 | 0.0950 | 0.0725 | 0.0432 | 0.05350 | 0.0933 | 0.1900 | 0.3280 | 0.4180 | 0.4270 | 0.4420 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 87548 | 0.807 | 0.495 | 0.537 | 0.5300 | 0.4910 | 0.4840 | 0.4560 | 0.39600 | 0.2840 | 0.1370 | 0.0526 | 0.1400 | 0.1190 | 0.0912 |
| 87549 | 0.718 | 0.605 | 0.487 | 0.3620 | 0.2320 | 0.1200 | 0.0517 | 0.00167 | 0.0000 | 0.0133 | 0.0783 | 0.1830 | 0.2750 | 0.3150 |
| 87550 | 0.906 | 0.624 | 0.596 | 0.5760 | 0.5310 | 0.4820 | 0.4450 | 0.38800 | 0.3220 | 0.1920 | 0.0980 | 0.0490 | 0.0367 | 0.0531 |
| 87551 | 0.858 | 0.646 | 0.846 | 0.2480 | 0.1670 | 0.1320 | 0.1220 | 0.12200 | 0.1190 | 0.1040 | 0.1010 | 0.1010 | 0.0886 | 0.0810 |
| 87552 | 0.902 | 0.846 | 0.801 | 0.7490 | 0.6870 | 0.5990 | 0.5120 | 0.42800 | 0.3950 | 0.4020 | 0.3710 | 0.2870 | 0.2360 | 0.3240 |

87553 rows × 188 columns

Figure 2. ECG Wave Dataset
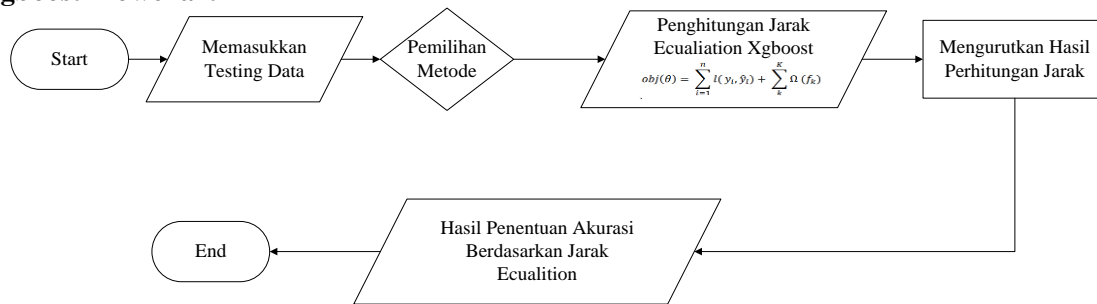
### 3.3 Xgboost Flowchart



Figure 3. Flowchart Xgboost

Xgboost formula:

$$obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_l) + \sum_{k}^{K} \Omega(f_k)$$

[11]

Where $\sum_{i=1}^{n} l(y_i, \widehat{y})$ is the differentiable loss function to measure whether the model is suitable for the training data set and is the item that determines the complexity of the model. As the complexity of the model increases the corresponding score is reduced $\sum_{k}^{k} \Omega(f_k)$ [11-16].

### 3.4 Data Processing

A system that can classify heart rate conditions based on the results of an electrocardiography (EKG) examination using the Xgboost method to assist heart examinations is expected to be one way to overcome the problem of detecting heart disease in Indonesia, and can help medical personnel ranging from nurses to doctors to classify results. ECG examination. The steps followed in data processing:

1.    **Import Data**

The first step is to import the required libraries, initialize the headers:

Source Code Import Library:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import numpy as np
import matplotlib.pyplot as plt
from xgboost.sklearn import XGBClassifier
df = pd.read_csv("drive/MyDrive/Dataset/ECG/EKG.csv", header=None)names =
```

```
['t1','t2','t3','t4','t5','t6','t7','t8','t9','t10','t11','t12','t13
','t14','t15','t16','t17','t18','t19','t20','t21','t22','t23','t24',
't25','t26','t27','t28','t29','t30','t31','t32','t33','t34','t35','t
36','t37','t38','t39','t40','t41','t42','t43','t44','t45','t46','t47
','t48','t49','t50','t51','t52','t53','t54','t55','t56','t57','t58',
't59','t60','t61','t62','t63','t64','t65','t66','t67','t68','t69','t
70','t71','t72','t73','t74','t75','t76','t77','t78','t79','t80','t81
','t82','t83','t84','t85','t86','t87','t88','t89','t90','t91','t92',
't93','t94','t95','t96','t97','t98','t99','t100','t101','t102','t103
','t104','t105','t106','t107','t108','t109','t110','t111','t112','t1
13','t114','t115','t116','t117','t118','t119','t120','t121','t122','
t123','t124','t125','t126','t127','t128','t129','t130','t131','t132'
,'t133','t134','t135','t136','t137','t138','t139','t140','t141','t14
2','t143','t144','t145','t146','t147','t148','t149','t150','t151','t
152','t153','t154','t155','t156','t157','t158','t159','t160','t161',
't162','t163','t164','t165','t166','t167','t168','t169','t170','t171
','t172','t173','t174','t175','t176','t177','t178','t179','t180','t1
81','t182','t183','t184','t185','t186','t187','class']
```

Pemberian Label Header Pada Dataset

```
df.columns = names
print(df.head(2))
print(df.info())
print(df.describe())
```

```
           t1      t2      t3      t4      t5      t6       t7      t8      t9     t10  \
0       0.978   0.926   0.681   0.245   0.154   0.191   0.1520  0.0858  0.0588  0.0490
1       0.960   0.863   0.462   0.197   0.094   0.125   0.0997  0.0883  0.0741  0.0826

        ...  t179  t180  t181  t182  t183  t184  t185  t186  t187   class
0       ...   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0  normal
1       ...   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0  normal

[2 rows x 188 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 87554 entries, 0 to 87553
Columns: 188 entries, t1 to class
dtypes: float64(187), object(1)
memory usage: 125.6+ MB
None
                  t1            t2            t3            t4            t5  \
count   87554.000000  87554.000000  87554.000000  87554.000000  87554.000000
mean        0.890360      0.758161      0.423971      0.219103      0.201127
std         0.240908      0.221812      0.227306      0.206879      0.177057
min         0.000000      0.000000      0.000000      0.000000      0.000000
25%         0.922000      0.682250      0.251000      0.048500      0.082300
50%         0.991000      0.826000      0.429000      0.166000      0.148000
75%         1.000000      0.911000      0.579000      0.342000      0.259000
max         1.000000      1.000000      1.000000      1.000000      1.000000

                  t6            t7            t8            t9           t10  \
count   87554.000000  87554.000000  87554.000000  87554.000000  87554.000000
mean        0.210399      0.205807      0.201771      0.198690      0.196756
std         0.171909      0.178481      0.177241      0.171777      0.168357
min         0.000000      0.000000      0.000000      0.000000      0.000000
25%         0.088400      0.073300      0.066100      0.065000      0.068600
50%         0.159000      0.145000      0.144000      0.150000      0.149000
75%         0.288000      0.298000      0.295000      0.291000      0.284000
max         1.000000      1.000000      1.000000      1.000000      1.000000
```

Figure 4. Header Initialization Dataset

## 2.    AssessmentData

The second step is data assessment. At this stage, we check whether the dataset that we use has empty values, filled with 0 or with the mean value of the data. As well as checking whether it contains outlier data, if any, it can be removed. Because the dataset that we use is numerical data of the ECG signal, which is numeric, null data and outliers are not found.

Source Code for Assessing Data:
```
df.isnull().sum()
```

```
 1  df.isnull().sum()

t1        0
t2        0
t3        0
t4        0
t5        0
          ..
t184      0
t185      0
t186      0
t187      0
class     0
Length: 188, dtype: int64
```

Figure 5. Empty Dataset

**3.    Convert Class**

The third step, convert the class to numeric. Normal = 0, supraventricular = 1, ventricular = 2, fusion = 3 and unknown = 4. Next, create x_values as input data and y_values as target data.

Source Code Convert Class

```
def label_encode(df):
df_labelled = df.copy()
# Dictionary to input the different numbers for different classes
label_encode = {"class": {"normal":0, "supraventricular ectopic":1, "ventricular ectopic":2, "fusion":3, "unknown":4}}

# Use .replace to change the different classes into numbers
df_labelled.replace(label_encode,inplace=True)
return df_labelled
```

**4.    Machine Learning Input**

The fourth step, the process of entering data into machine learning to get the results of abnormalities in the heart through the classification of the electrocardiogram wave dataset using the XGBoost method.

Source Code Machine Learning

a.   Normal Predict

Source Code Normal Predict:

```
normalpredict = pd.read_csv("drive/MyDrive/Dataset/ECG/predict normal2.csv")
Xgboost = XGBClassifier()
x = df
```

```
[['t1','t2','t3','t4','t5','t6','t7','t8','t9','t10','t11','t12','t1
3',
't14','t15','t16','t17','t18','t19','t20','t21','t22','t23','t24','t
25',
't26','t27','t28','t29','t30','t31','t32','t33','t34','t35','t36','t
37',
't38','t39','t40','t41','t42','t43','t44','t45','t46','t47','t48','t
49',
't50','t51','t52','t53','t54','t55','t56','t57','t58','t59','t60','t
61',
't62','t63','t64','t65','t66','t67','t68','t69','t70','t71','t72','t
73',
't74','t75','t76','t77','t78','t79','t80','t81','t82','t83','t84','t
85',
't86','t87','t88','t89','t90','t91','t92','t93','t94','t95','t96','t
97',
't98','t99','t100','t101','t102','t103','t104','t105','t106','t107',
't108',
't109','t110','t111','t112','t113','t114','t115','t116','t117','t118
','t119',
't120','t121','t122','t123','t124','t125','t126','t127','t128','t129
','t130',
't131','t132','t133','t134','t135','t136','t137','t138','t139','t140
','t141',
't142','t143','t144','t145','t146','t147','t148','t149','t150','t151
','t152',
't153','t154','t155','t156','t157','t158','t159','t160','t161','t162
','t163',
't164','t165','t166','t167','t168','t169','t170','t171','t172','t173
','t174',
't175','t176','t177','t178','t179','t180','t181','t182','t183','t184
','t185','t186','t187'
]]
```

Pemanggilan Inisiasi Header

```
y = df['class']
Xgboost = Xgboost.fit(x,y)
test = normalpredict
predictXGB = Xgboost.predict(test)
print(predictXGB)
```

```
['normal']
```

b.  Supraventricular EctopicPredict
Source CodeSupraventricularEctopicPredictions:

```
SEpredict = pd.read_csv("drive/MyDrive/Dataset/ECG/predict supraventricular ectopic2.csv")
Xgboost = XGBClassifier()
x2 = df
```

```
[['t1','t2','t3','t4','t5','t6','t7','t8','t9','t10','t11','t12','t13'
,'t14',
't15','t16','t17','t18','t19','t20','t21','t22','t23','t24','t25','t26
','t27',
't28','t29','t30','t31','t32','t33','t34','t35','t36','t37','t38','t39
','t40',
't41','t42','t43','t44','t45','t46','t47','t48','t49','t50','t51','t52
','t53',
't54','t55','t56','t57','t58','t59','t60','t61','t62','t63','t64','t65
','t66',
't67','t68','t69','t70','t71','t72','t73','t74','t75','t76','t77','t78
','t79',
't80','t81','t82','t83','t84','t85','t86','t87','t88','t89','t90','t91
','t92',
't93','t94','t95','t96','t97','t98','t99','t100','t101','t102','t103',
't104',
't105','t106','t107','t108','t109','t110','t111','t112','t113','t114',
't115',
't116','t117','t118','t119','t120','t121','t122','t123','t124','t125',
't126',
't127','t128','t129','t130','t131','t132','t133','t134','t135','t136',
't137',
't138','t139','t140','t141','t142','t143','t144','t145','t146','t147',
't148',
't149','t150','t151','t152','t153','t154','t155','t156','t157','t158',
't159',
't160','t161','t162','t163','t164','t165','t166','t167','t168','t169',
't170',
```

Pemanggilan Inisiasi Header

```
y2 = df['class']
Xgboost = Xgboost.fit(x2,y2)
test = SEpredict
predictXGBSE = Xgboost.predict(test)
print(predictXGBSE)
```

```
['normal']
```

c. Ventricular EctopicPredict

Source Code Ventricular Ectopic Predict :

```
VEpredict = pd.read_csv("drive/MyDrive/Dataset/ECG/predict ventricular ectopic2.csv")
Xgboost = XGBClassifier()
x3 = df
```

```
[['t1','t2','t3','t4','t5','t6','t7','t8','t9','t10','t11','t12','t
13','t14',
't15','t16','t17','t18','t19','t20','t21','t22','t23','t24','t25','
t26','t27'
't28','t29','t30','t31','t32','t33','t34','t35','t36','t37','t38','
t39','t40'
't41','t42','t43','t44','t45','t46','t47','t48','t49','t50','t51','
t52','t53'
't54','t55','t56','t57','t58','t59','t60','t61','t62','t63','t64','
t65','t66'
't67','t68','t69','t70','t71','t72','t73','t74','t75','t76','t77','
t78','t79'
't80','t81','t82','t83','t84','t85','t86','t87','t88','t89','t90','
t91','t92'
't93','t94','t95','t96','t97','t98','t99','t100','t101','t102','t10
3','t104'
't105','t106','t107','t108','t109','t110','t111','t112','t113','t11
4','t115'
't116','t117','t118','t119','t120','t121','t122','t123','t124','t12
5','t126'
't127','t128','t129','t130','t131','t132','t133','t134','t135','t13
6','t137'
't138','t139','t140','t141','t142','t143','t144','t145','t146','t14
7','t148'
't149','t150','t151','t152','t153','t154','t155','t156','t157','t15
8','t159'
't160','t161','t162','t163','t164','t165','t166','t167','t168','t16
9','t170'
't171','t172','t173','t174','t175','t176','t177','t178','t179','t18
0','t181'
't182','t183','t184','t185','t186','t187']]
```
Pemanggilan
Inisiasi
Header

```
y3 = df['class']
Xgboost = Xgboost.fit(x3,y3)
test = VEpredict
predictXGBVE = Xgboost.predict(test)
print(predictXGBVE)
```

```
['fusion']
```

d. FusionPredict

Source Code Ventricular Ectopic Predict:

```
fusionpredict = pd.read_csv("drive/MyDrive/Dataset/ECG/predict fusion2.csv")
Xgboost = XGBClassifier()
# Extract out the x values and y values. x will be sepal_length and y will be classes
x4 = df
```

```
[['t1','t2','t3','t4','t5','t6','t7','t8','t9','t10','t11','t12
','t13','t14',
't15','t16','t17','t18','t19','t20','t21','t22','t23','t24','t2
5','t26','t27'
't28','t29','t30','t31','t32','t33','t34','t35','t36','t37','t3
8','t39','t40'
't41','t42','t43','t44','t45','t46','t47','t48','t49','t50','t5
1','t52','t53'
't54','t55','t56','t57','t58','t59','t60','t61','t62','t63','t6
4','t65','t66'
't67','t68','t69','t70','t71','t72','t73','t74','t75','t76','t7
7','t78','t79'
't80','t81','t82','t83','t84','t85','t86','t87','t88','t89','t9
0','t91','t92'
't93','t94','t95','t96','t97','t98','t99','t100','t101','t102',
't103','t104'
't105','t106','t107','t108','t109','t110','t111','t112','t113',
't114','t115'
't116','t117','t118','t119','t120','t121','t122','t123','t124',
't125','t126'
't127','t128','t129','t130','t131','t132','t133','t134','t135',
't136','t137'
't138','t139','t140','t141','t142','t143','t144','t145','t146',
't147','t148'
't149','t150','t151','t152','t153','t154','t155','t156','t157',
't158','t159'
't160','t161','t162','t163','t164','t165','t166','t167','t168',
't169','t170'
't171','t172','t173','t174','t175','t176','t177','t178','t179',
't180','t181'
't182','t183','t184','t185','t186','t187']]
```
Pemanggilan
Inisiasi
Header

```
y4 = df['class']
```

```
Xgboost = Xgboost.fit(x4,y4)
test = fusionpredict
predictXGBF = Xgboost.predict(test)
print(predictXGBF)
```

```
['fusion']
```

## 5.Training and Test

The fifth step is to build a network and conduct training and testing. In this experiment, we use the Xgbost algorithm. We each apply a split data train and test with a ratio of 70% and 30%. We do this until we get the best accuracy from the algorithm.

### 5.1 Data Split

The function of training is to train the model to recognize patterns in the data, while testing is to ensure that the trained model is able to predict tags from new observations that have not been studied by previous models.[15].

Split Data source code:

```
x_train, x_test, y_train,
y_testXgBoost = train_test_split(x_values,y_values,test_size=0.3,random_state=10)
scaler = StandardScaler()
x_train_scale = scaler.fit_transform(x_train)
x_test_scale = scaler.transform(x_test)
```

### 5.2. Conversion Matrix Xgboost

Split Data source code:

```
from sklearn.metrics import confusion_matrix
matrix = (confusion_matrix(y_testXgBoost, y_pred))
print(matrix)
benar =(matrix[0,0]+matrix[1,1]+matrix[2,2]+matrix[3,3]+matrix[4,4])
Total=(matrix[0,0]+matrix[0,1]+matrix[0,2]+matrix[0,3]+matrix[0,4]+matrix[1,0]+
matrix[1,1]+matrix[1,2]+matrix[1,3]+matrix[1,4]+matrix[2,0]+matrix[2,1]+
matrix[2,2]+matrix[2,3]+matrix[2,4]+matrix[3,0]+matrix[3,1]+matrix[3,2]+
matrix[3,3]+matrix[3,4]+matrix[4,0]+matrix[4,1]+matrix[4,2]+matrix[4,3]+
matrix[4,4])
print(benar)
print(Total)

akurasi = ((matrix[0,0]+matrix[1,1]+matrix[2,2]+matrix[3,3]+matrix[4,4])/
(matrix[0,0]+matrix[0,1]+matrix[0,2]+matrix[0,3]+matrix[0,4]+matrix[1,0]
+matrix[1,1]+matrix[1,2]+matrix[1,3]+matrix[1,4]+matrix[2,0]+matrix[2,1]
+matrix[2,2]+matrix[2,3]+matrix[2,4]+matrix[3,0]+matrix[3,1]+matrix[3,2]
+matrix[3,3]+matrix[3,4]+matrix[4,0]+matrix[4,1]+matrix[4,2]+matrix[4,3]
+matrix[4,4])*100)

print(accuracy)
```

```
[[  128     47     0     0    12]
 [    1  21691    31    12    53]
 [    0    225   420     3     8]
 [    0     93     1  1786    12]
 [   10    211     3    13  1507]]
25532
26267
Akurasi =  97.20181215974416 %
```

Figure 6. Predict Accuracy

In the matrix conversion with the Xgboost algorithm, 97.20% of the data accuracy in detecting heart disease can be obtained.

## 3.5     Results

To record the activity of the heartbeat, a diagnostic tool in the form of an Electrocardiogram (ECG) can be used, the results of the heart record are then analyzed by the doctor to determine the conditions and problems with the patient's heart. The rhythm of the heart rate is shown in the form of a sine curve.

### 3.5.1 Normal

In general, the ECG curve consists of P, Q, R, S and T curves. The P curve represents atrial depolarization, the QRS complex curve describes ventricular depolarization, and the T curve describes ventricular repolarization. Seen in figure 7.

The normal sine wave rhythm has the following characteristics:
1.  Regular rhythm
2.  Heart rate ranges from 60-100 beats/minute
3.  The P curve was normal (width < 0.12 sec, height < 0.3 mv, positive in lead II and negative in lead aVR) and the QRS and T complex curves followed.
4.  Normal QRS curve (between 0.06 to less than 0.12 seconds)
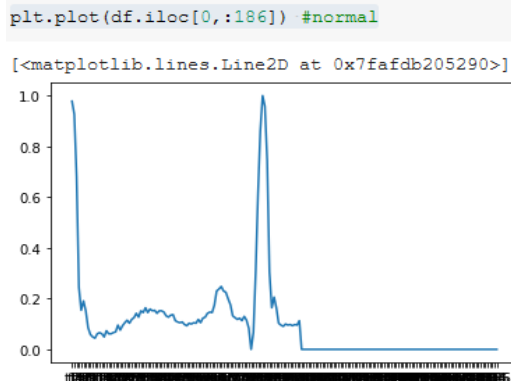5.  The PR curve interval is normal (0.12 - 0.20 sec).



Figure 7.Normal

## 3.5.2     Supraventricular ectopic

Supraventricular ectopicis a condition in which the heart beats very fast, much more than the normal heart rate. If the normal heart beats 60-100 times per minute, the heart of a patient with supraventricular ectopic beats 140-250 times per minute. Seen in figure 8.
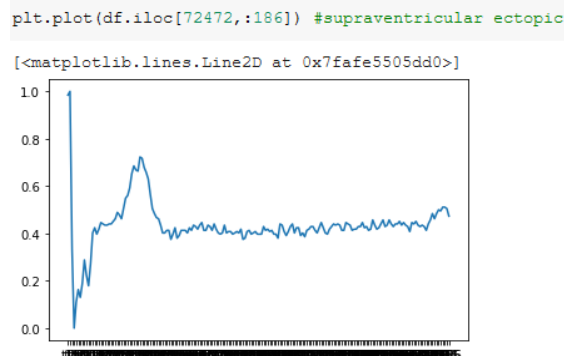


Figure 8.Supraventricular

### 3.5.3 Ventricular ectopic

Ventricular ectopicis a disorder of the heart's pumping work in the ventricles or ventricles, in which the chambers beat too fast and cause less blood to be pumped by the heart. Seen in figure 9.
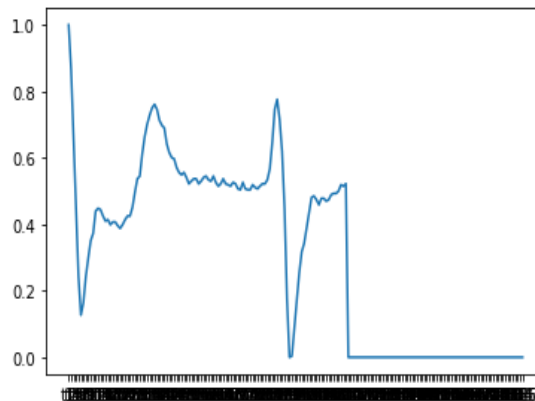


Figure 9.Ectopic Ventricular ECG Waveform

### 3.5.4 Fusion beat

Fusion beatis a condition in which impulses from the sinus node are delivered to the ventricles through the atrioventricular node and are combined with impulses from the ventricles. This is dangerous in acute ischemic situations. Seen in figure 10.
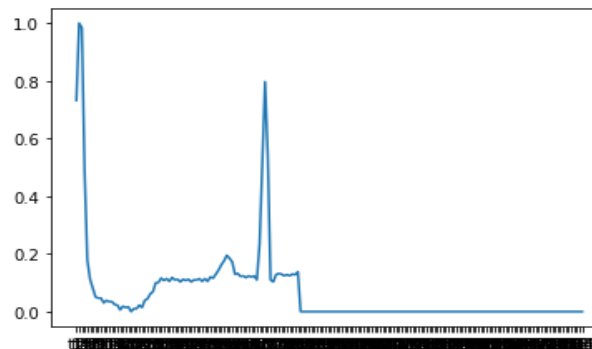


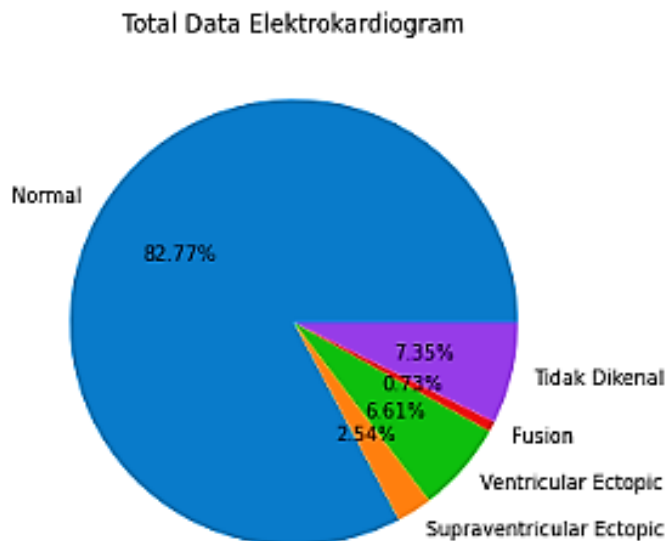Figure 10.Fusion Beat EKG wave

### 3.6     Accuracy Results



Figure 10.Processing Results

From the data processing carried out in classifying electrocardiogram waves with normal data sharing 72471 or 82.77%, 2223 supraventricular ectopic abnormalities or 2.54%, 5788 ventricular ectopic abnormalities or 6.61%, 641 fusion disorders or 0.73%, and 6431 data or 7.35% unknown abnormality. As many as these results are known, Xgboost gives the best results with an accuracy of 97.20% or is able to correctly classify 25577 data from 26267 data.

## 4.     CONCLUSIONS

Currently, cardiovascular disease can affect anyone and can also be caused by: an unhealthy lifestyle. So cardiovascular disease really needs to be done early prevention is important Early treatment can be done to prevent the worst effects of this disease. This research applies machine learning can help make this diagnostic process more efficient. In this study, machine learning was able to accurately classify abnormal ECGs up to a maximum of 97.20 percent using the Xgboost algorithm.

## reference

[1]    V. Octaviani, "Klasifikasi Detak Jantung Normal Dan Abnormal Berbasis Sinyal Ekg Menggunakan Metode Backpropagation," p. 143, 2017.
[2]    G. A. Roth et al., "Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study," J. Am. Coll. Cardiol., vol. 76, no. 25, pp. 2982–3021, Dec. 2020, doi: 10.1016/J.JACC.2020.11.010.
[3]    "Cardiovascular Disease Infographic - World Heart Federation," 2019. https://world-heart-federation.org/resource/cardiovascular-disease-infographic/ (accessed Feb. 18, 2022).
[4]    N. Estananto, "Klasifikasi Sinyal Elektrokardiogram Menggunakan Renyi Entropy," J. Elektro dan Mesin Terap., vol. 4, no. 2, pp. 11–18, 2018, doi: 10.35143/elementer.v4i2.2139.
[5]    M. Sofie and A. Rizal, "Klasifikasi Citra Rekaman Sinyal Elektrokardiogram Menggunakan Metode Analisis Tekstur, K-Nn Dan Multilayer Perceptron," Simetris J. Tek. Mesin, Elektro dan Ilmu Komput., vol. 7, no. 1, p. 233, 2016, doi: 10.24176/simet.v7i1.509.
[6]    A. K. Nastiti, E. Purwanti, and A. Supardi, "Klasifikasi Kelainan Jantung Dengan Metode

Transformasi Fourier Dan Jaringan Saraf Tiruan," Klasifikasi Kelainan Jantung Dengan Metod. Transform. Fourier Dan Jar. Saraf Tiruan, 2013.

[7]    Niendy Alexandra Yosephine and Ratnadewi, "Penggunaan Artificial Neural Network pada Sinyal Elektrokardiogram untuk Mendeteksi Penyakit Jantung Aritmia Supraventrikular," Inf. (Jurnal Inform. dan Sist. Informasi), vol. 13, no. 1, pp. 14–23, 2021, doi: 10.37424/informasi.v13i1.69.

[8]    A. Rizal and V. Suryani, "Pengenalan Signal EKG Menggunakan Dekomposisi Paket Wavelet dan K-Means Clustering," Proceeding Semin. Nas. Apl. Teknol. Inofrmasi 2008(SNATI 2008), vol. 2008, no. Snati, pp. 5–8, 2008.

[9]    D. Evanko, "ALGORITMA KLASIFIKASI DATA MINING NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK DETEKSI PENYAKIT JANTUNG," Nat. Methods, vol. 7, no. 1, p. 34, 2010, doi: 10.1038/nmeth.f.284.

[10]   T. Chen and C. Guestrin, "XGBoost : A Scalable Tree Boosting System," pp. 785–794, 2016.

[11]   I. Muslim and K. Karo, "Implementasi Metode XGBoost dan Feature Important untuk Klasifikasi pada Kebakaran Hutan dan Lahan," J. Softw. Eng. Inf. Commun. Technol., vol. 1, no. 1, pp. 10–16, 2020, [Online]. Available: https://ejournal.upi.edu/index.php/SEICT/article/view/29347.

[12]   K. Kelvin, J. Banjarnahor, E. Indra, and M. N. Nababan, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA), vol. 5,       no.       2,       pp.       47–52,       Feb.       2022,       doi: 10.34012/JURNALSISTEMINFORMASIDANILMUKOMPUTER.V5I2.2365.

[13]   "Apa itu Data Acquisition?" https://www.dataloggerindonesia.com/apa-itu-data-acquisition-152 (accessed Feb. 19, 2022).

[14]   "Heart       Attack       Analysis       &       Prediction       Dataset       |       Kaggle." https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset (accessed May 13, 2022).

[15]   "Machine Learning dengan Python. Untuk membuat model, tidak bisa… | by Muchamad Nur Kholis | Program Studi Statistika UII | Medium." https://medium.com/statistics-uii/machine-learning-dengan-python-47e004eaa727 (accessed Apr. 13, 2022).

[16]   Tamba, S.P., 2022. PREDIKSI PENYAKIT GAGAL JANTUNG DENGAN MENGGUNAKAN RANDOM FOREST. Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA), 5(2), pp.176-181.