Electrodiogram Signal Classification by Using XGBoost in Different Discrete Wavelet Transform

Bibit Waluyo Aji, Sri Nur Chasanah, Fihris Aulia Sa'adah, Bambang Irawanto

***Abstract*—Electrocardiogram (ECG) is electrical signal from heart. ECG can use for Detection or tracking the hearth health. The one of method can use is machine learning. Machine learning is Algorithm which can learning from data and is used for classifying and predicting. Machine Learning can use for signal classification, in this case is for ECG classification. In signal processing, wavelet transform is common method for analyzing signal. Wavelet transform has many familly. The aim from this research is to find the best wavelet transform in the classification of Electrocardiogram (ECG) signals on XGBoost. The Discrete Wavelet Transform which is used for the research is daubechies, coiflets, symlets, biorthogonal, reverse biorthogonal, haar. Finally, the best wavelet transform in the classification is biorthogonal (3.1) with F1 score 1.0.**

Keywords: Signal Classification, Wavelet, Electrocardiogram, XGBoost.

I. INTRODUCTION

ACCORDING to the World Health Organization (WHO) in 1948, health is defined as "a state of complete physical, mental, and social well-being and not merely disease or infirmity. The heart is the main organ of the cardiovascular system (cardiac arrhythmia) and one of the vital organs of the human body. In life, the heart is always in good condition because it works to survive together with the blood it carries throughout the body. Heart failure can threaten human health, even heart problems can cause death. In previous studies, to detect abnormalities or changes in heart function, it was necessary to know the same pattern of heartbeats first. Cardiovascular disease remains a major threat worldwide. Data on the global burden of cardiovascular disease shows that there were 271 million cases of cardiovascular disease in 1990 which nearly doubled to 523 million in 2019 [1].

Over the past 70 years, the incidence of cardiovascular disease has remained at a high level. According to data from the World Health Organization, around 33 percent of deaths worldwide are caused by cardiovascular disease. As a result of cardiovascular disease, an estimated 18.6 million people died, of whom 75 percent died of heart attacks and strokes. More than three-quarters of deaths from cardiovascular disease occur in developing and low- and middle-income countries. In situations like this, cardiovascular disease can strike anyone. Increasing air pollution and lifestyle may be the causes of cardiovascular disease [2].

B. W. Aji, S. N. Chasanah, B. Irawanto are with Department of Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: bibitwaji@gmail.com, asasrinurwsb@gmail.com

F. A. Sa'adah are with Department of Nursing, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: fihrisaulia@gmail.com

Manuscript received July 18, 2023; accepted September 14, 2023.

It is important to detect cardiovascular disease as early as possible so that counseling and drug management can be given promptly. This ECG diagnostic tool can record the human heart rate. The electrocardiogram (ECG) also reflects various aspects of the heart's electrical activity, and is an important bio-electrical signal. Thus, this ECG is an effective noninvasive and inexpensive clinical tool for the treatment of cardiovascular disease. However, during the ECG signal acquisition process it is often damaged by various noise and artifacts, consequently affecting the subsequent processing and analysis of ECG signals, and greatly reducing the diagnostic accuracy [3].

To avoid mistakes in diagnosing heart disease, experts identify ECG signals. In order to identify ECG signals, a machine learning data set is needed to train the network. Machine learning technology is used to find out the initial diagnosis to make it faster and easier. One method that can be used is to apply a machine learning algorithm to identify abnormal ECG waves. A system capable of classifying pulse conditions based on electrocardiographic (ECG) results can be performed using the Xgboost method. The XGboost method helps examine the heart and helps medical personnel classify results. In a previous study entitled "Classification of Electrocardiogram (ECG) Waves of Heart Disease Using The XGBoost Method", it was found that machine learning using the XGBoost method can accurately classify abnormal ECGs up to a maximum of 97.20 percent.

Machine learning for monitoring electrocardiograms can be designed with various systems. One of them can be designed based on the wavelet transform. Wavelet analysis can be used to describe low frequencies and high frequencies to represent signal characteristics. In addition, the wavelet transformation has an effect on improving the quality of the ECG signal. Wavelet-based procedures are important in genetic algorithms, neural networks, digital and adaptive screening techniques, and detection accuracy. Signal analysis is greatly benefited by this wavelet transform, which was developed in recent years. This wavelet transform has resulted in thirteen well-documented wavelet families [4]. This becomes the basis of the argument in establishing a reasonable and well-founded basis regarding the choice of the best and most suitable wavelet for machine learning with the XGBoost Technique to obtain the most accurate diagnosis possible maximum. Therefore, the author is interested in conducting a classification study entitled "Electrocardiogram (ECG) Signal Classification By Using XGBoost In Different Discrete Wavelet Transform".

II. Electrocardiogram

Electrocardiogram or commonly abbreviated as ECG / EKG is a simple test to check or measure the electrical activity of a person's heart using a tool called an electrocardiograph. The tool translates electrical impulses into a graph that is displayed on the monitor screen. This procedure is classified as safe, fast and painless because it is done without electricity and without incisions (non-invasive). The ECG is recognized as a powerful clinical screening and diagnostic tool and is used globally in almost every line of health [5]. An electrocardiogram is recorded by measuring the potential difference between two electrodes placed on the patient's skin. The ECG detects the length of time of electrical waves by measuring the interval and amount of electrical activity in the heart muscle [6]. The results of measuring the length of time of electrical waves in the heart are used to detect the normality of electrical activity in the heart organs. Meanwhile, the amount of electrical activity in the heart muscle can indicate abnormalities in this vital organ.

III. XGBoost

In 2016 Tianqi Chen developed XGBoost which is a development of GBDT (Gradient Boosting Decision Tree) [7]. This algorithm is an extension of the 3 classic Gradient Boosting Machine (GBM) algorithms and is only used for labeled data in the training process. XGBoost (Extreme Gradient Boosting) is a regression and classification algorithm with the ensemble method. XGBoost is also a variant of the Tree Gradient Boosting algorithm which was developed with an optimization that is 10 times faster than Gradient Boosting [8]. XGBoost function is [9]:

$$obj(\theta) = L(\theta) + \Omega(\theta) \tag{1}$$

With $L(\theta)$ is loss function, and $\Omega(\theta)$ is a regularization function. Which regularization function is formulated by

$$\Omega(\theta) = \gamma N + \frac{1}{2}\lambda \, ||w||^2 \tag{2}$$

Where γ and λ is regularization function parameter, N represents the number of leaf nodes in the decision tree and w represents the weights of the node.

IV. Wavelet Transform

Wavelet is a mathematical function that divides data into several different frequency components, then an analysis is carried out for each component using a resolution that is appropriate to its scale. Wavelet transform is a multi-resolution decomposition technique to solve modeling problems that produce good local representation of signals in the time domain and frequency domain. The wavelet transform uses a flexible modulation window, which means it has adjustments for frequency and time, this is the most different from the short-time Fourier transform (STFT), which is a development of the Fourier transform, which means it cannot analyze frequency and time simultaneously [10]. DWT decomposes the EMG signal into multi-resolution wavelet coefficients. the signal is displayed in time and frequency representation so that the extracted features contain time information on different frequency sub-bands [11]. Wavelet decomposition involves digital filters, low-pass and high-pass filters. Any finite energy analog signal can be decomposed into small waveforms and through scaling functions:

$$x(t) = \sum_{n=-\infty}^{infty} c(n)\phi(t-n) + \sum_{j=0}^{\infty}\sum_{n=-\infty}^{\infty} d(j,n)2^{\frac{j}{2}}\psi(2^j - n) \tag{3}$$

where $\psi(t)$ is a real-valued bandpass wavelet and $\phi(t)$ is a shifts of a real-valued low-pass scalling function [12].

The scalling coefficient $c(n)$ and wavelet coefficient $d(j,n)$ are computed via the inner product

$$c(n) = \int_{-\infty}^{\infty} x(t)\phi(t-n)dt \tag{4}$$

$$d(j,n) = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} x(t)\psi(2^j n)dt \tag{5}$$

Wavelets have many families including haar, Daubechies (dbN), symlet (SymN), Coiflet (CoifN), etc. The core of the wavelet is haar, which was first introduced by Alfred Haar in 1909. The **haar is the peak and mother wavelet**, then derifiated on to other families.

Wavelet Daubechies is a development of the Haar wavelets. Daubechies 1 (db1) with filter length 2 is a Haar wavelet. Daubechies 2 abbreviated (db2) is a wavelet Daubechies with many filters 4 which is db3 is a Daubechies wavelet with many filters 6 etc. Modification of Daubechies wavelets are Symlets wavelets, that increase symmetry.

Coiflet wavelets are another wavelet family commonly abbreviated coif, where the scaling function has (N/3-1) lost moments, and the wavelet function has N/3 lost moments. The biorthogonal wavelet (abbreviated bior) has two scaling functions and two wavelet functions, which are used for decomposition and reconstruction. reverse biorthogonal wavelet is not orthogonal (or from the same side/area), so the reverse biorthogonal wavelet provides inner freedom design on any system. moreover, the design more flexible and the threshold value can be measured from different level [13].

A. Evaluation

In this article, to assess how well the Wavelet Transform is used in classification using XGBoost, several evaluation methods are used. Evaluation method for this research is using 4 methods, precision, recall, F1 score, accuracy. Where the formulas for the four methods are as follows.

a. Precision

$$precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

b. Recall Score

$$recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

c. F1 Score

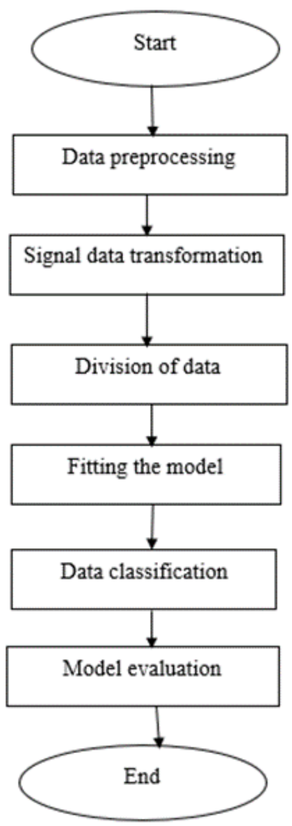
$$F1\ Score = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

d. Accuracy

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

V. RESEARCH METHODS

The aim of the proposed research is to find the best wavelet transform in the classification of Electrocardiogram (ECG) signals on XGBoost. Initially, the ECG data was taken from the secondary data from the physionet, with three data classes, ARR : Abnormal Arrhythmia, NSR : normal sinus rhythm, CHF : congestive heart failure . The ECG signal is affected by various types of noise, namely electrosurgical noise, instrumentation noise, baseline drift, etc. Therefore, the next step is to denoise the ECG signal. The six types of wavelet families that will be tested to denoise ECG signals are daubechies, coiflets, symlets, biorthogonal, reverse biorthogonal, haar, and mayer. The limitation of each Wavelet family is that it uses the first 5 orde. Normal, Arrhythmia, and data are separated into 50 training data and 50 test data. The data is trained and then tested. After that, the data is evaluated and then classified using XGBoost.



VI. RESULTS AND DISCUSSION

Below is result for Classification performance by using XGboost with different Discrete Wavelet Transform.

The use of the Wavelet Daubechies family in signal processing to be classified using the XGBoost method has increased performance from order 1 to order 4 then its performance drops to order 5. The result in two different outputs, namely good and bad daubechies signal processing performance. This can be proven from the results of the accuracy score and f1 score. The best score was obtained using 4th order Daubechies, with its accuracy and f1 score being 0.976.

TABLE I: Performance Daubechies Wavelet transform on XGBoost

Daubechies family	Accuracy	Precision	Recall	F1 Score
db1	0.901	0.901	0.901	0.901
db2	0.918	0.918	0.918	0.918
db3	0.948	0.948	0.948	0.948
db4	0.976	0.976	0.976	0.976
db5	0.920	0.920	0.920	0.920

TABLE II: Performance Coiflet Wavelet transform on XGBoost

Coiflet family	Accuracy	Precision	Recall	F1 Score
coif1	0.952	0.952	0.952	0.952
coif2	0.973	0.973	0.973	0.973
coif3	0.988	0.988	0.988	0.988
coif4	0.962	0.962	0.962	0.962
coif5	0.950	0.950	0.950	0.950

The use of the Coiflet Wavelet family in signal processing to be classified using the XGBoost method has increased performance from order 1 to order 3 then its signal performance has decreased to order 5. This can be seen from its accuracy score and f1 score. That a good coiflet processing signal was obtained using a 3rd order coiflet with accuracy and an f1 score of 0.988.

TABLE III: Performance Symplet Wavelet transform on XGBoost

Symlet family	Accuracy	Precision	Recall	F1 Score
sym2	0.925	0.925	0.925	0.925
sym3	0.975	0.975	0.975	0.975
sym4	0.945	0.945	0.945	0.945
sym5	0.947	0.947	0.947	0.947
sym6	0.960	0.960	0.960	0.960

The use of the Symplet Wavelet family in signal processing to be classified using the XGBoost method has increased performance from order 1 to order 2 then its performance drops to order 3 and rises again 4 and 5, this can be seen from its accuracy score and f1 score. The best score was obtained using Symplet 3, with its accuracy and f1 score being 0.975. The use of the Biorthogonal Wavelet family in signal processing to be classified using the XGBoost method has a decreased performance from order 1 to order 3 then its performance increases in order 4, this can be seen from its accuracy score and f1 score. The best score is obtained using

TABLE IV: Performance Biorthogonal Wavelet transform on XGBoost

Biorthogonal family	Accuracy	Precision	Recall	F1 Score
bior1.1	0.913	0.913	0.913	0.913
bior1.3	0.912	0.912	0.912	0.912
bior1.5	0.869	0.869	0.869	0.869
bior2.2	0.951	0.951	0.951	0.951
bior3.1	1.00	1.00	1.00	1.00

Biorthogonal 3.1, with its accuracy and f1 score is 1.00.

TABLE V: Performance Reverse Biorthogonal Wavelet transform on XGBoost

Reverse Biorthogonal family	Accuracy	Precision	Recall	F1 Score
rbio1.1	0.896	0.896	0.896	0.896
bior1.3	0.988	0.988	0.988	0.988
bior1.5	0.962	0.962	0.962	0.962
bior2.2	0.954	0.954	0.954	0.954
bior3.1	0.960	0.960	0.960	0.960

The use of the Reverse Biorthogonal Wavelet family in signal processing to be classified using the XGBoost method has an increased performance from order 1 to order 4 then its performance drops to order 5, this can be seen from the accuracy score and f1 score. The best score was obtained using Reverse Biorthogonal 1.3, with its accuracy and f1 score being 0.976.

TABLE VI: Performance Other Wavelet transform on XGBoost

Wavelet family	Accuracy	Precision	Recall	F1 Score
Dmey	0.988	0.988	0.988	0.988
Haar	0.880	0.880	0.880	0.880

The use of mayer and haar wavelets in signal processing to be classified using the XGBoost method has an F1 Score of 0.980 and 0.880 respectively.

TABLE VII: The best wavelet transform each family in Classification with XGBoost

Wavelet family	Accuracy	Precision	Recall	F1 Score
Daubechies (db4)	0.976	0.976	0.976	0.976
Coiflet (coif3)	0.988	0.988	0.988	0.988
Symplet (sym3)	0.975	0.975	0.975	0.975
Biorthogonal (bior3.1)	1.00	1.00	1.00	1.00
Reverse Biorthogonal	0.988	0.988	0.988	0.988
Mayer(dmey)	0.988	0.988	0.988	0.988
Haar(haar)	0.880	0.880	0.880	0.880

The use of various types of wavelets in signal processing to be classified using the XGBoost method obtained the best score for Biorthogonal Wavelet 3.1 with an Accuracy score and f1 score is 1.00 and the worst is Haar with an accuracy score and f1 score is 0.880.

VII. CONCLUSIONS

The best discrete wavelet transform from this study is the biorthogonal wavelet type with the biorthogonal 3.1 type. XGBoost performance with Biorthogonal 3.1 for accuracy and F1 Score of 1.00 The results in this study are still limited by the use of only three data classes. The wavelet transform used focuses on the first 5 orders of discrete wavelet transforms. So, to find out which discrete wavelet transform is the best in signal classification, especially EGG it is necessary to further develop more data classes, different machine learning methods, and more wavelet orders.

REFERENCES

[1] Roth, G.A., et.al Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study, (2020). https://doi.org/10.1016/j.jacc.2020.11.010.

[2] Yunarti Butarbutar, S., Deniro Napitupulu, C., Sanjaya Ginting, N., Indra, E., Sitanggang, D.: CLASSIFICATION OF ELECTROCARDIOGRAM (ECG) WAVES OF HEART DISEASE USING THE XGBOOST METODE METHOD. JURNAL INFOKUM. 10, (2022).

[3] Yang, S., Chen, H.-C., Chen, W.-C., Yang, C.-H.: Student Enrollment and Teacher Statistics Forecasting Based on Time-Series Analysis. Comput Intell Neurosci. 2020, 1–15 (2020). https://doi.org/10.1155/2020/1246920.

[4] Kumar, A., Komaragiri, R., Kumar, M.: Design of wavelet transform based electrocardiogram monitoring system. ISA Trans. 80, 381–398 (2018). https://doi.org/10.1016/j.isatra.2018.08.003.

[5] Too, J., Abdullah, A.R., Saad, N.M., Kejuruteraan, F., Teknikal, U., Melaka, M., Elektronik, F.K., Komputer, K.: Classification of Hand Movements based on Discrete Wavelet Transform and Enhanced Feature Extraction. (2019).

[6] Chatterjee, S., Thakur, R.S., Yadav, R.N., Gupta, L., Raghuvanshi, D.K.: Review of noise removal techniques in ECG signals, (2020). https://doi.org/10.1049/iet-spr.2020.0104.

[7] Ogunleye, A., Wang, Q.G.: XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Trans Comput Biol Bioinform. 17, 2131–2140 (2020). https://doi.org/10.1109/TCBB.2019.2911071.

[8] Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. Association for Computing Machinery (2016). https://doi.org/10.1145/2939672.2939785.

[9] Maleki, A., Raahemi, M., Nasiri, H.: Breast cancer diagnosis from histopathology images using deep neural network and XGBoost. Biomed Signal Process Control. 86, (2023). https://doi.org/10.1016/j.bspc.2023.105152.

[10] Nobre, J., Neves, R.F.: Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets. Expert Syst Appl. 125, 181–194 (2019). https://doi.org/10.1016/j.eswa.2019.01.083.

[11] Manhas, P., Thakral, S.: Image Processing by Using Different Types of Discrete Wavelet Transform. (2018).

[12] Rhif, M., Abbes, A. Ben, Farah, I.R., Martínez, B., Sang, Y.: Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review, (2019). https://doi.org/10.3390/app9071345.

[13] Zhang, D.: Wavelet Transform. Presented at the (2019). https://doi.org/10.1007/978-3-030-17989-2_3.