A Comparison of Heart Abnormalities Detection on ECG using KNN and Decision Tree

Evta Indra, Stiven H. Sinurat, Maggie A. Fong, Calvin Tiopan, Saut P. Tamba, Oloan Sihombing, Ivan K. Laksono, Delima Sitanggang, and Andrian.

***Abstract*— In life, the heart is always required to always be in good condition because the heart serves to pump blood that carries nutrients throughout the body. Impaired heart function can be fatal to human health, even some heart disorders can lead to death. To be able to detect the presence of abnormalities or disorders of the heart, it must be known in advance the working rhythm or signal pattern of the heart itself. In this study the algorithm KNN and the decision tree of the three algorithms was used to search for the best results so that the initial diagnosis error can be minimized. K-NN and the decision tree give the best results with an accuracy of 97.373% and 95.87%, respectively. Early detection of cardiovascular disease can be done with ECG wave analysis based on Artificial Intelligence to make the process more efficient. In this research, machine learning can classify abnormal ECGs with maximum accuracy of 97.373%.**

***Index Terms*— KNN, Decision Tree, ECG, Wave Identification**

I. INTRODUCTION

BASED on data from the World Health Organization, Cardiovascular disease is the leading cause of death globally. In 2019, an estimated 18.6 million people died from the disease, representing 33% of all global deaths. Of these deaths, 75% were caused by heart attacks and strokes [1]. More than three-quarters of deaths from cardiovascular disease occur in low and middle-income countries. Cardiovascular disease is a disease that occurs due to impaired heart and blood vessel function [2].

A previous study entitled "Predictions Based on Support Vector Machine for High-Speed Rail Passenger Travel Options in China" discusses the prediction of trains selected by ferry. This can improve competitiveness in the transportation market and lead to better high-speed rail service. Then another paper ``Detect Pedestrian'' Orientation by Integrating Use of Multiclass SVM Binary Decision Tree" This research is to predict the possibility of a collision by detecting pedestrians directions using the same method namely binary decision tree-SVM using INRIA dataset consisting of 2000 images. Results accuracy for Non-Ped (Non-pedestrian) is 100%, front and left by 99%, right by 95%, and behind about 90%. Apart from research above, also conducts a study of research with the title The second research on the application of Fuzzy K-NN method to determine quality yield of sugarcane by Rahmi, et al, shows that FK-NN is a method good for solving classification problems, where in this study the accuracy of finding 98% using 175 training data and 80 data tests.

Heart is the main organ in the cardiovascular system as well as one of the vital organs in the human body [3]. In life, the heart is always required to always be in good condition because the heart serves to pump blood that carries nutrients throughout the body [4]. Impaired heart function can be fatal to human health, even some heart disorders can lead to death. To be able to detect the presence of abnormalities or disorders of the heart, it must be known in advance the working rhythm or signal pattern of the heart itself.

Because the trigger for cardiovascular disease is closely related to human lifestyle, implementing a healthy lifestyle is the best way to prevent it. Such as not smoking and not consuming alcohol, keeping the ideal weight, eating healthy foods, exercising regularly, and detecting cardiovascular disease as early as possible [5].

According to previous research mentioned earlier, to reduce farmer mobility when detecting leaf disease, the writer proposes using a method that starts with collecting a dataset from the Plant Village repository in Kaggle [4], then doing a test & train split on the dataset. Pre-processing will be done after that. Pre-processing covers resizing & rescaling the picture, rotating the picture by 35°, and adjusting the picture saturation to get more features from a single picture (Data Augmentation). After that, a Convolutional Neural Network model will be built using ResNet-50 architecture, which is expected to have 85% of minimum accuracy. The difference of our method with the previous research is that we let the machine learn itself to predict the disease, not using an expert system.

II. METHOD

The main reason for ECG analysis and monitoring is to obtain information about the heart condition [6]. The attack of heart disease is sudden and even the symptoms are rarely shown. Long term monitoring is usually suggested, due to its simplicity and safety reasons. It is important to detect cardiovascular disease as early as possible so that counseling and medications can be managed. One of the preventions is to

Manuscript received October 9, 2020.

E. Indra, SH. Sinurat, M. A. Fong, C. Tiopan, S. Tamba, O. Sihombing, D. Sitanggang, Andrian. Information System and Technology Universitas Prima Indonesia Medan, Indonesia. (e-mail: maggieamelia23@gmail.com)

I. K. Laksono. Mathematics Department, Universitas Negeri Surabaya. (e-mail: ivanlaksono@mhs.unesa.ac.id)

conduct early detection of Electrocardiographic (ECG) Wave Analysis [7]. To record heart rate activity can be used diagnostic instruments in the form of electrocardiograms (ECG), the doctor then analyzes the results of the heart record to find out the condition and problems in the patient's heart [8].

Heart rate rhythm is displayed in the form of sinus curves. Generally, ECG curves consist of curves P, Q, R, S and T [9]. The P curve describes the depolarization of the atrium, the QRS complex curve describes the depolarization of the ventricle, the T curve describes the repolarization of the ventricle. The research work procedures in this study are given in Fig. 1.

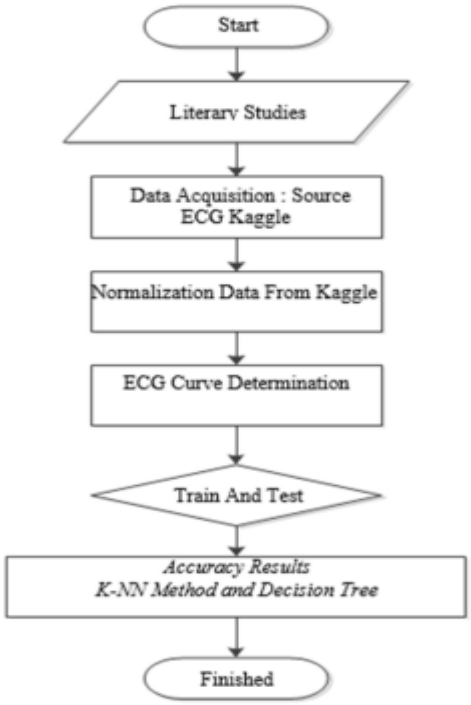

Fig. 1. System Diagram of Proses Application

Fig. 2A indicates a normal sinus wave rhythm with the characteristics of [10]:

- Regular rhythm
- Heart frequency ranges from 60-100 x/min
- Normal P curve (width < 0.12 seconds, height <0.3 mv, positive in lead II and negative in leadaVR) and followed by complex curves QRS and T
- Normal QRS curve (between 0.06 to less than 0.12seconds)
- Normal PR curve interval (0.12 - 0.20 seconds)

If the rhythm of the sinus waves is not the same as these criteria, it indicates abnormalities in the heart performance. Abnormalities that can occur include supraventricular ectopic, ventricular ectopic, and fusion beat.

Fig. 2. B shows the supraventricular ectopic with a condition where the heart beats very fast, far exceeding the normal heartbeat [11]. If the normal heart beats 60-100 times per minute, then the supraventricular ectopic heart beats 140-250 times per minute [12].

Fig. 2.C Ectopic ventricular is a performance disorder of the heart pump in the chamber or ventricle part, which pulses too fast and results in less blood pumped by the heart [13].

Fig. 2.D Fusion beats an impulse condition of the sinus nodus delivered to the ventricle through the atrioventricular nodes and joins the impulse of the ventricle. It is dangerous in acute ischemic situations [14].

Identification of ECG signals must be made by an expert so that there is no error in the diagnosis. Because of this, the diagnosis process is quite time-consuming. Moreover, the number of experts available is limited. Under certain conditions, it is necessary to diagnose quickly, such as the current pandemic period. Where many medical experts focus on handling covid 19, whereas cardiovascular disease is no less deadly. Not to mention the length of the queue at the hospital because many people are panicking, so if they experience a little complaint, they immediately flock to the hospital. If the handling process does not run quickly, then it can be fatal.

Literature study is research conducted by researchers by collecting several books and magazines related to the problem and research objectives. This technique is carried out with the aim of revealing various theories that are in accordance with the problem being studied as reference material in the discussion of research results.

The retrieval of the dataset were processed from kaggle. This dataset was used to explore heart rate classification using deep neural network architecture and observe some of its transfer learning capabilities. The signal corresponds to an electrocardiogram (ECG) heartbeat to detect abnormalities in the heart. Normalization of data is the process of making variables that have the same range of values, nothing is too big or too small so that it can make statistical analysis easier.

The process of making 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. The process of training and testing on the K-NN and Decision Tree methods to get the accuracy of each method's detecting abnormalities in the heart. The final result of machine learning data processing using the K-NN method and Decision Tree.

One solution to overcome this is to make use of Artificial Intelligence (AI) technology so that the initial diagnostic process can be done faster and easier [15]. One method that can be done is to apply machine learning algorithms to identify the normality of ECG waves. In this study the algorithm used is KNN, decision tree, of the three algorithms will be searched for the best results so that the initial diagnosis error can be minimized.

Actually, the heat from the liquid chiller is absorbed by the refrigerant through another heat exchanger. That is why liquid chiller is used in air conditioners in large buildings because the air circulation in the building has a large volume, it will be more efficient if you use a liquid chiller. For, the humidity is the concentration of water vapor in the air, which its number can be expressed in terms of absolute humidity, specific humidity, or relative humidity. The relative humidity is a term

used to describe water vapor contained in the water-air mixture in the gas phase, while the temperature is a measure of the state of the object, which determines the speed at which the object receives or releases heat to its surroundings which are different from that of the object.

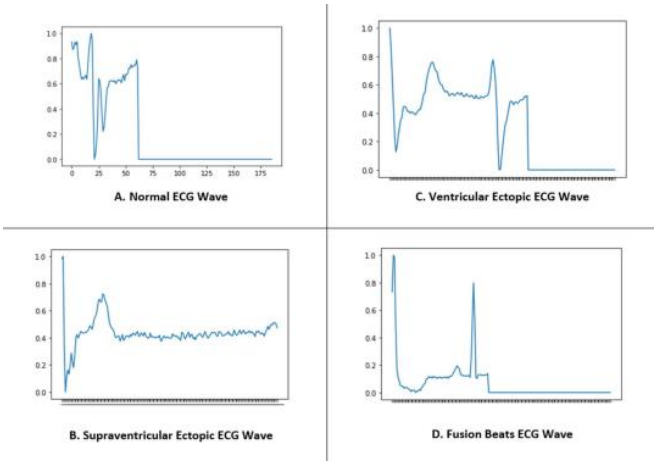

Fig. 2. Characteristics ECG Wave

III. Results and Discussions

Cases of cardiovascular disease from year to year show an increasing number, everyone can be infected with this disease. Early treatment or detection is important in order to reduce the risk of death that can be caused. Early detection of cardiovascular disease can be done with ECG wave analysis, in which case AI can help early detection more easily and quickly. To do so AI requires a dataset used to train the network. The datasets used in this study are sourced: https://www.kaggle.com/shayanfazeli/heartbeat.

The dataset is ECG numeric signal data taken from87554 subjects consisting of 72471 normal, 2223 ectopic supraventricular abnormalities, 5788 ventricular ectopic abnormalities, 641 fusion abnormalities and 6431 unknown abnormalities.

The first step is to import the required libraries and initialize the headers.

```
[ ] import pandas as pd
    from sklearn.neighbors import KNeighborsClassifier
    from sklearn.model_selection import train_test_split
    from sklearn.preprocessing import StandardScaler
    import numpy as np
```

```
df = pd.read_csv("./EKG.csv",header=None)
names = ['t1','t2','t3','t4','t5','t6','t7','t8','t9'
, 'class']
df.columns = names
print(df.head(2))
print(df.info())
print(df.describe())
```

Fig. 3. Import the required library

The second step is data assessment. At this stage, check whether the dataset we use is empty, if it can be filled with the number 0 or the mean value of the data. As well as checking whether it contains outliers data; if it does, it can be eliminated. Because the dataset we use numerical data of ECG signal that is numerical, null data and pencilan data are not found

```
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
```

Fig. 4. The second step data assessment

Step three, convert the class to numeric. Normal = 0, supraventricular = 1, ventricular = 2, fusion = 3 and unknown = 4. Next, create a x_values variable as input data and y_values as target data.

```
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
```

```
x_values = df[['t1','t2','t3','t4','t5','t6','t7','t8','t9','t10','t11','t12',
]]
y_values = df['class']
```

Fig. 5. Step three, convert the class to numeric.

Step four, build the network and train and test. In this experiment, we used two algorithms: KNN and decision tree. Each of us applies a split train and test data with a ratio of 7:3. For KNN, we did some experiments by changing the value of K while for the decision tree, experiments are done by changing the depth. We do this to get the best accuracy of each algorithm.

Train and test (KNN)

```
[8] x_train, x_test, y_train, y_test = 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)
```

```
[ ] #Train a K-Nearest Neighbours algorithm with the dataset.

    KNN = KNeighborsClassifier(n_neighbors=3)
    KNN = KNN.fit(x_train_scale,y_train)
```

```
[ ] y_predict = KNN.predict(x_test_scale)
```

Fig. 6. Train and test (KNN)

CONV MATRIX(K-NN)

```
from sklearn.metrics import confusion_matrix
matrix = (confusion_matrix(y_test, y_predict))
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(akurasi)
```

```
[[  129    38     0     0    20]
 [   15 21617    57    14    85]
 [    1   190   456     0     9]
 [    0    55     0  1828     9]
 [   16   162    13     6  1547]]
25577
26267
97.37312978261697
```

Fig. 7. Conv matrix (K-NN)

Train and test (Decision tree)

```
from sklearn import tree

dt = tree.DecisionTreeClassifier(max_depth = 16, min_samples_split = 6)
dt.fit(x_train_scale,y_train)
```

```
DecisionTreeClassifier(max_depth=16, min_samples_split=6)
```

```
y_predict = dt.predict(x_test_scale)
```

Fig. 8. Train and test (Decision Tree)

The results of ECG signal classification using K-NN and Decision Tree algorithms are given in Fig. 9 and Table 1.

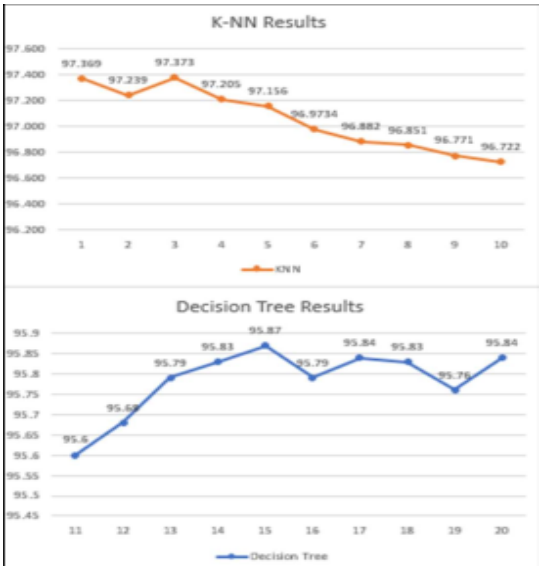

Fig. 9. K-NN and Decision Tree Results

TABLE I
ACCURACY RESULTS OF K-NN AND DECISION TREE

K-NN		Decision Tree	
K	Acc	depth	Acc
1	97.369	11	95.6
2	97.239	12	95.68
3	97.373	13	95.79
4	97.205	14	95.83
5	97.156	15	95.87
6	96.9734	16	95.79
7	96.882	17	95.84
8	96.851	18	95.83
9	96.771	19	95.76
10	96.722	20	95.84

IV. CONCLUSION

From these results, it is known that K-NN gives the best results when the value k=3 with an accuracy of 97.373% oracle to correctly classify 25577 data from 26267 data. The Decision tree gives the best results when depth = 15 with an accuracy of 95.87 or able to correctly classify 25178 data from 26267 data. So in this study, the K-NN algorithm performs better than the decision tree. At the time, cardiovascular disease can attack anyone due to an unhealthy lifestyle. Early detection of cardiovascular disease is important so that early treatment can be done to prevent the worst effects of this disease. Implementing Machine learning can help this process to be more efficient.

REFERENCES

[1] A. Turnip, Andrian, M. Turnip, A. Dharma, D. Paninsari, T. Nababan and C.N. Ginting, "An Application of Modified Filter Algorithm Fetal Electrocardiogram Signals with Various Subjects," *International Journal of Artificial Intelligence*, vol. 18, no. 1, pp. 207-217, 2020.

[2] Martiningsih and A. Haris, "Cardiovascular Disease Risk Program for Participants Management of Chronic Disease (Prolanis) in Bima City Puskesmas: Correlation with AnkleBrachial Index and Obesity," *Indonesian Journal of Nursing*, vol. 22, no. 3, pp. 200–208, 2019.

[3] A. Turnip, M. Ilham Rizqywan, D. E. Kusumandari, M. Turnip, and P. Sihombing, "Classification of ECG signal with Support Vector Machine Method for Arrhythmia Detection," *IOP Conf. Series: Journal of Physics: Conf.* Series 970 (2018) 012012.

[4] B. Mallick and A. K. Patro, "Heart rate Monitoring System Using Finger Tip Through Arduino and ProcessingSoftware," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 5, Issue 1, January, 2016.

[5] D. Armalina, B. Witjahjo, N. Susilanigsih, R. D. Purnawati, A. Ismail, and F. Saktini, "Screening for Hypertension and Electrocardiography (ECG) for Cardiovascular Diseases Risk." *Jurnal Kesehatan Masyarakat, KEMAS* vol. 15 no. 3, pp. 303-308, 2020.

[6] C. Wijaya, Andrian, M. Harahap, Christnatalis, M. Turnip, and A. Turnip, "Abnormalities State Detectionfrom P-Wave, QRS Complex, and T-Wave in Noisy ECG," vol. 1230, no. 01, 2015.

[7] M. Risnasari, "Noise Empowerment on ECG Signal using Transformation Wave," *Edutic Scientific Journal*, vol. 1, no. 1, November, 2014.

[8] Sugiyanto, T. Indriyani, and M. heru Firmansyah, "Electrocardiogram Arrhythmia Classification System usingSupport Vector Machine Based Fuzzy Logic," *Journal of Computer Science and Information*, vol. 9, no. 1, pp. 35-42, 2016.

[9] E. B. Setiawan and M. Y. Tahir, "Electrocardiogram (ECG) Application for Checking HeartConditions Based on Android," *IJNMT*, vol. VI, no. 1, June, 2019.
[10] S. Setiawidayat, A.Y. Rahman, and R. Hidayat, "Detection of Peak Amplitude and Duration of Electrocardiogram QRS Waves Using Discrete Data," *Journal Resti*, vol. 4, no. 3, pp. 438–446, 2020.
[11] N.A. Yosephine and Ratnadewi, "Using Artificial Neural Neworks on Electrocardiogram Signals to DetectSupraventicular Arrhythmic Heart Disease," *INFORMATION (Journal of Informatics and Information Systems)*, vol. 13, no.1, May, 2021.
[12] P.J. Michalek, "An Authentic Egg Simulator," Electric Thesesand Dissertations, Electronic Theses and Dissertations, 2004-2019, 966.
[13] P.S.C. Mei and A. Ahmad, "A Review on Trial Fibrillation and its Related Analysis," *Journal of Technology (Sciences & Engineering)*, vol. 78, pp. 67–73, 2016.
[14] J. Pardede, M. Turnip, D.R. Manalu and A. Turnip, "Adaptive Recurrent Neural Network forReduction of Noise and Estimation of Source from Recorded EEG Signals," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 3, February, 2015.
[15] A. Gultom, E. Girsang, and S.L.R. Nasution, "Design a Predictive Analytics Model of Hospital Accreditation Continuty from Employee Readiness based on Artificial Intelligence," *In Proceedings of the International Conference on Health Informatics, Medical, Biological Engineering, and Pharmaceutical (HIMBEP 2020)*, pp. 96-103, 2020.
[16] D. S. Wibowo, R. Maulana, and F. Utaminingrum, "Implementation of The K-Nearest Neighbors Method in The Sleep Apnea Detection System," *Journal of Information Technology Development and Computer Science*, vol. 3, no. 11, pp. 10894-10899, 2019.
[17] A. Bataineh, W. Batayneh, T. Harahseh, K. Hijazi, A. Alrayes, M. Olimat, and A. Btaineh, "Early Detection of Cardiac Diseases fromElectrocardiogram Using Artificial Intelligence Techniques," *International Review on Modelling and Simulations*, vol. 14, no. 2, 2021.
[18] S. Zokaee and K. Faez, "Human Identification Based onElectrocardiogram and Palmprint," *International Journal of Electrical and Computer Engineering*, vol. 2, no. 2, 2012.
[19] L.A. Abdullah and M. Shaban Al – Ani, "A Review Study for Electrocardiogram Signal Classification," *UHD Journal of Science and Technology*, vol. 4, 2020.
[20] R. Rani, V. S. Chouhan and H. P. Sinha, "AutomatedDetection og QRS Complex in ECG Signal Using Wavelet Transform", *IJCSNS International Journal of Computer Scienceand Network Security*, VOL.15 No.1, January 2015.

Evta Indra is currently the Chair of the Information Systems Study Program, Faculty of Technology and Computer Science, Universitas Prima Indonesia. and also currently the Editor in Chief of Jurnal Sistem Informasi dan Ilmu Komputer PRIMA (JUSIKOM PRIMA). Received a Bachelor's Degree in Computers in 2012 at the Potential Utama University of Medan, and a Master's Degree in Computers in 2015 at UPI YPTK Padang. His research interests include decision support systems, expert systems, geographic information systems, artificial intelligence, data mining.