Decision Support System for Heart Disease Diagnosing Using K-NN Algorithm

Tito Yuwono
Department of Electrical Engineering
Islamic University of Indonesia
Yogyakarta
Address: Kaliurang Street KM 14
Yogyakarta, Indonesia
Email:tito@uii.ac.id

Noor Akhmad Setiawan, Hanung Adi Nugroho, Anugrah Galang Persada
Department of Electrical and Information Technology
Gadjah Mada University
Address: Jalaan Grafika No 2 Yogyakarta
Email:noorwewe@ugm.ac.id, adinugroho@ugm.ac.id

Ipin Prasojo, Sri Kusuma Dewi, Ridho Rahmadi
Department of Informatic
Islamic University of Indonesia
Yogyakarta
Address: Kaliurang Street KM 14
Yogyakarta, Indonesia
Email:rm_prasojo@yahoo.com, cicie, ridhorahmadi@uii.ac.id

***Abstract*— Heart disease is a notoriously dangerous disease which possibly causing the death. An electrocardiogram (ECG) is used for a diagnosis of the disease. It is often, however, a fault diagnosis by a doctor misleads to inappropriate treatment, which increases a risk of death. This present work implements *k*-nearest neighbor (K-NN) on ECG data to get a better interpretation which expected to help a decision making in the diagnosis. For experiment, we use an ECG data from MIT BIH and zoom in on classification of three classes; normal, myocardial infarction and others. We use a single decision threshold to evaluate the validity of the experiment. The result shows an accuracy up to 87% with a value of K = 4.**

Keywords—K-NN; ECG; Diagnosis; Heart Disease

I. Introduction

Heart disease (cardiovascular) is a notoriously dangerous disease. According to World Health Organization (WHO), the disease is a major cause of death [1]. Moreover, The Department of Health of the Republic of Indonesia also states that the disease is one of the main causes of the death in Indonesia [2].

Some alternatives have been suggested to decrease the risk of death caused by this disease; an early screening of heart activity, and an accurate diagnosis. An electrocardiogram (ECG) is used to check the disease. It is often, however, a diagnosis that is conducted by a doctor based on the ECG data misleading to an inappropriate decision. A computer-aided diagnosis is an alternative to solve this problem.

Several previous studies have been conducted. For example, to use the Wavelet and template matching method to measure the QT interval of the ECG signal [3], or to use pattern recognition [4][5][6], or to use EMD (Empirical Mode Decomposition) combined with R peak detection

method and CWT (Continuous Wavelet Transform)[7], or to use generalized tensor rank one discriminant analysis on feature extraction [8], and a software development for automatic detection [9].

This present work provides an implementation of computer assistance to help the diagnosis. We implement a method called *k*-nearest neighbor (K-NN) on ECG data which intended to give a better and accurate interpretation.

Furthermore, some practical purposes that motivate our study are: (1) to help Physicians in interpreting the ECG signal, (2) To transform the ECG signal into a form that is ready for a further process of computation.

The remainder of the paper is organized as follows: the methodology is described in Section II, the experimental results is discussed in Section III, and the conclusion of this study is summarized in Section IV.

II. methodology

A. *k-Nearest Neighbor*

K-nearest neighbor (k-NN) is a non-parametric technique for classification and regression. It is an instance-based learning method to group data points based on the majority class in a neighborhood with size *k*. The basic idea is, to find the *k* nearest neighbors for each data point, and to classify it based on the majority. In order to do that, the distances between this particular data point to the other *k* data points are computed.

Technically *k*-NN works as the following steps. Suppose we are given a training data set (X_i,Y_i),....,(X_n,Y_n) (with m-dimensional attribute vector X and its corresponding class Y. Each X is classifiable into P classes, namely C_1,....,C_p Then we are given a new instance X_{n+1} to classify.

The *k*-NN first computes the distances between X_{n+1} and X_1,...,X_n by means of Euclidean distance measurement.

$$d_i = \sqrt{\sum_{j=1}^{m} (X_{n+1} - X_{ij})^2} \quad (1)$$

A. Block Diagram

In this section we provide a block diagram as a visualization of the steps of our implementation. Figures 1 and 2 give sketches of this block diagram. First we collect the ECG signal and ECG files (containing prior knowledge base) as the input. The distance metric is then computed between the new ECG signals and the prior knowledge base, based on an attribute vector, namely PR interval, PR segment, QRS interval, ST segment and QT interval. This step assigns a new instance of ECG signal into a class.

The prior knowledge base that we plug in contains three classes, representing three different types of patient states: disease Healthy Controls (normal), Myocardial Infarction, and the other. The distribution of the disease as shown in Figure 3 and Appendix A.

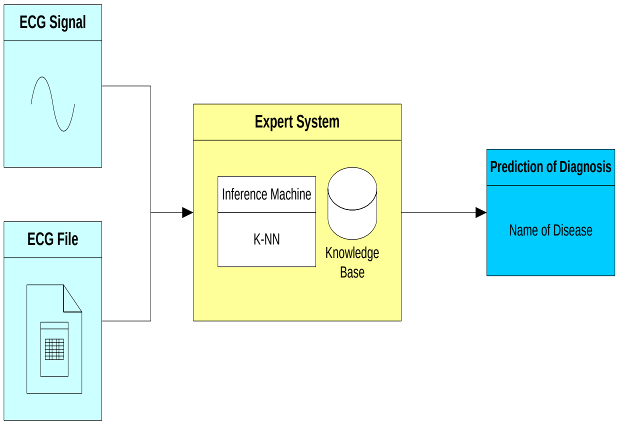


Fig. 1. Block Diagram of the intelligence system for ECG interpretation.

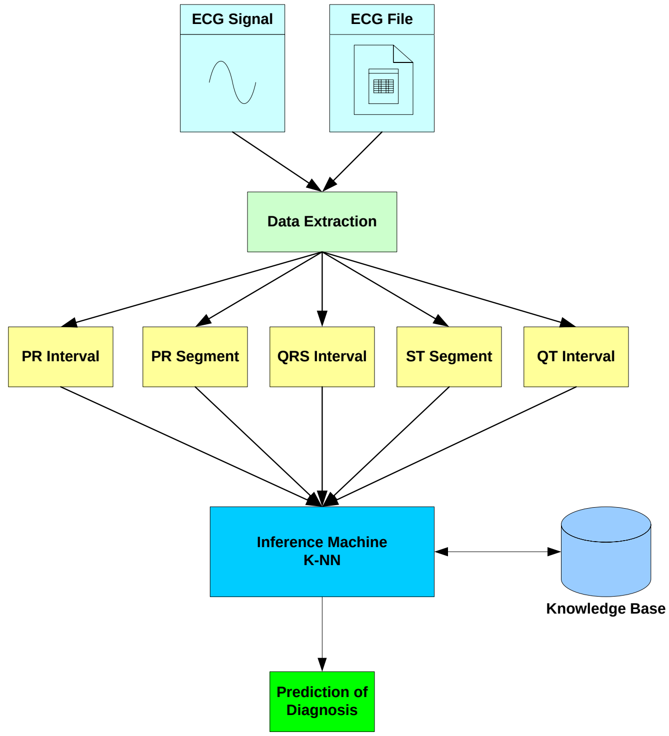


Fig. 2. The model of the decision support system

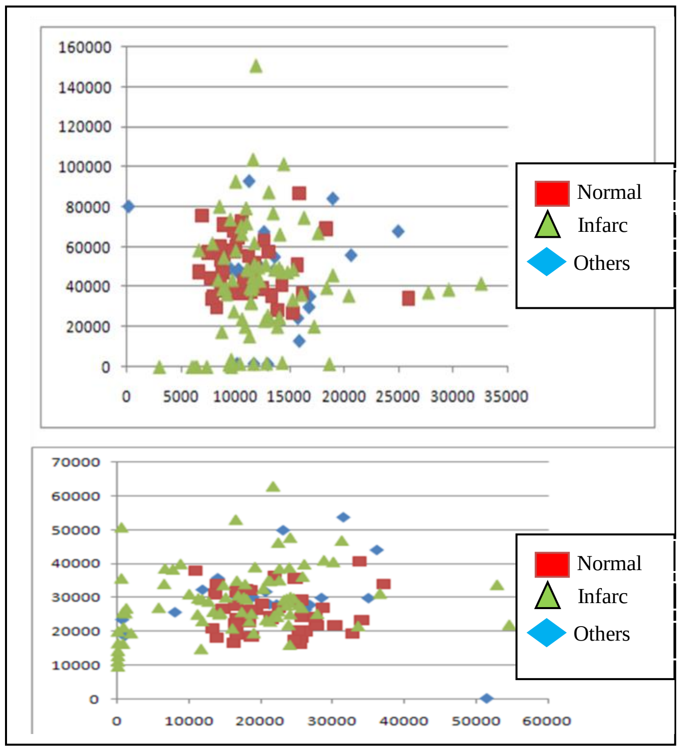


Fig. 3. Heart disease distribution ((PRinterval x QRSinterval), (STsegment x QTinterval)) and ((PRinterval x STsegment), (QRSinterval x QTinterval)).

III. Experimental result

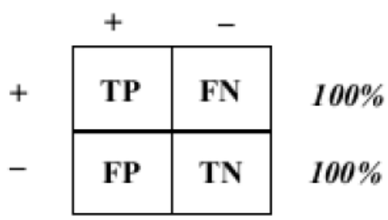
Table 1 lists the classes of diseases in the prior knowledge base.

TABLE 1. THE DISTRIBUTION OF DISEASE CLASSES

No	Disease class	Number of instancess
1	Health Control (Normal)	41
2	Myocardial infarction	79
3	Others	23

In order to test the performance of our implementation, we conduct a validity test. This test is to compare the result of prediction to the real outcome. The result of the validity test is described in Table 4 in Appendix A.

Furthermore we use single decision threshold to determine the accuracy. The computation is based on the confusion matrix as follows.



1. TP (*True Positive*) is for correct prediction of a patient's disease.
2. TN (*True Negative*) is for a correct prediction that a patient does not suffer from the disease.
3. FP (*False Positive*) is for a prediction that states a patient suffers from a disease, which actually not.
4. FN (*False Negative*) is for a prediction that states a patient does not suffers from a disease, which actually yes.

Table 2 shows comparison between reality with the diagnosis system

TABLE 2. THE COMPARISON BEETWEN THE REAL DIASEASE WITH THE DIAGNOSIS

	Diagnosis using K-NN			
Real Disease	***Diagnosis***	***Normal***	***Myocardial Infarction***	***Others***
	Normal	32	8	1
	Myocardial Infarction	8	70	1
	Others	2	6	15

The accuracy of the system can be determined by calculating the value of TP, TN, FP and FN of Table 4.

TP= 32 + 70 + 15 = 117

TN= (70 + 15) + (32 + 15) + (32 + 70) = 234

FP= (8 + 2) + (8 + 6) + (1 + 1)= 26

FN= (8 + 1) + (8 + 1) + (2 + 6) = 26

The accuracy $T = \frac{TP + TN}{TP + TN + FP + FN} x100\%$

$$T = \frac{117 + 234}{117 + 234 + 26 + 26} x100\% = 87\%$$

Based on the above calculation can be seen that the level of accuracy of the expert system that has been created is 87%. Several algorithms have been used with each its accurattion are presented in Table 3.

TABLE 3. THE COMPARATION OF ALGORITHM [10]

Algorithm	Control Patiens N=382	Ventricular Hypertrophy N=291	Myocardial Infarction N=547	Total of Accuration (%) N=1220
Padova	89.8	61.3	47.1	62.0
Nagoya-Fukuda	89.3	42.6	63.7	65.6
IBM Medis	91.3	49.4	62.5	67.6
HP (Agilent)	93.5	51.0	64.5	69.3
Glasgow	94.0	51.0	67.7	69.7
GE (Marquette)	86.3	61.1	69.7	69.8
Means	97.1	42.5	67.2	69.8
Hannover	86.6	72.1	79.0	75.8
Louvaine (Louven)	91.5	67.0	82.1	77.3
8 Cardiologists Combined Scores	97.1	60.4	80.3	79.2

Table 3 is a publication of QRS diagnostic using variation of algorithms. This table shows Louvaine has the best total accuracy with 77.3%. The accuracy for diagnosing Control Patiens (normal) is 91.5%, Ventricular Hypertrophy is 67.0% and Myocardial Infarction is 82.1%.

IV. Conclusion

The accuracy achieved in this study is 87% with *k*=4. Although this percentage is incomparable with the results provided in Table 5 (as the sample size is different), but our result shows that *k*-NN is able to return a reasonably good result. As the future work, we suggest to increase both number of diseases and members of each class.

Acknowledgment

Thanks to the minister of research-technology and higher education that has funded this research.

Appendix A

TABLE 5. THE DIAGNOSIS VALIDATION FOR K=4

No	Real Deasease	Diagnosis
1	Myocardial Infarction	Myocardial Infarction
2	Myocardial Infarction	Myocardial Infarction
3	Myocardial Infarction	Myocardial Infarction
4	Myocardial Infarction	Myocardial Infarction
5	Myocardial Infarction	Myocardial Infarction
6	Myocardial Infarction	Myocardial Infarction
7	Myocardial Infarction	Myocardial Infarction
8	Myocardial Infarction	Myocardial Infarction
9	Myocardial Infarction	Myocardial Infarction
10	Myocardial Infarction	Myocardial Infarction
11	Myocardial Infarction	Myocardial Infarction
12	Myocardial Infarction	Myocardial Infarction
13	Myocardial Infarction	Myocardial Infarction
14	Myocardial Infarction	***Normal***
15	Myocardial Infarction	Myocardial Infarction
16	Myocardial Infarction	***Normal***
17	Myocardial Infarction	Myocardial Infarction
18	Myocardial Infarction	***Normal***
19	Myocardial Infarction	***Normal***
20	Myocardial Infarction	***Others***
21	Myocardial Infarction	Myocardial Infarction
22	Myocardial Infarction	Myocardial Infarction
23	Myocardial Infarction	Myocardial Infarction
24	Myocardial Infarction	Myocardial Infarction
25	Myocardial Infarction	Myocardial Infarction
26	Myocardial Infarction	Myocardial Infarction
27	Myocardial Infarction	Myocardial Infarction
28	Myocardial Infarction	Myocardial Infarction
29	Myocardial Infarction	Myocardial Infarction
30	Myocardial Infarction	Myocardial Infarction
31	Myocardial Infarction	Myocardial Infarction
32	Myocardial Infarction	Myocardial Infarction
33	Myocardial Infarction	Myocardial Infarction
34	Myocardial Infarction	Myocardial Infarction
35	Myocardial Infarction	Myocardial Infarction
36	Myocardial Infarction	Myocardial Infarction
37	Myocardial Infarction	Myocardial Infarction
38	Myocardial Infarction	***Normal***
39	Myocardial Infarction	Myocardial Infarction
40	Myocardial Infarction	Myocardial Infarction
41	Myocardial Infarction	Myocardial Infarction
42	Myocardial Infarction	Myocardial Infarction
43	Myocardial Infarction	Myocardial Infarction
44	Myocardial Infarction	Myocardial Infarction
45	Myocardial Infarction	Myocardial Infarction
46	Myocardial Infarction	Myocardial Infarction
47	Myocardial Infarction	Myocardial Infarction
48	Myocardial Infarction	Myocardial Infarction
49	Myocardial Infarction	Myocardial Infarction
50	Myocardial Infarction	Myocardial Infarction
51	Normal	Normal
52	Normal	Normal
53	Others	***Myocardial Infarction***
54	Normal	Normal
55	Normal	Normal
56	Others	***Myocardial Infarction***
57	Normal	Normal
58	Normal	***Myocardial Infarction***
59	Others	Others
60	Others	Others
61	Myocardial Infarction	***Normal***
62	Others	***Myocardial Infarction***
63	Normal	Normal
64	Myocardial Infarction	Myocardial Infarction
65	Myocardial Infarction	Myocardial Infarction
66	Myocardial Infarction	Myocardial Infarction
67	Myocardial Infarction	***Normal***
68	Myocardial Infarction	Myocardial Infarction
69	Others	Others
70	Myocardial Infarction	Myocardial Infarction
71	Normal	***Myocardial Infarction***
72	Normal	***Myocardial Infarction***
73	Myocardial Infarction	Myocardial Infarction
74	Others	***Normal***
75	Myocardial Infarction	Myocardial Infarction
76	Others	***Myocardial Infarction***
77	Normal	Normal
78	Normal	***Myocardial Infarction***
79	Others	Others
80	Normal	Normal
81	Normal	Normal
82	Normal	Normal
83	Others	Others
84	Normal	Normal
85	Normal	Normal
86	Normal	Normal
87	Others	Others
88	Myocardial Infarction	Myocardial Infarction
89	Others	Others
90	Others	Others
91	Others	Others
92	Myocardial Infarction	Myocardial Infarction
93	Others	Others
94	Myocardial Infarction	***Normal***
95	Others	***Myocardial Infarction***
96	Myocardial Infarction	Myocardial Infarction
97	Others	Others
98	Others	Others
99	Normal	***Myocardial Infarction***
100	Others	Others
101	Others	Others
102	Myocardial Infarction	Myocardial Infarction
103	Others	Others
104	Normal	Normal
105	Normal	Normal
106	Normal	Normal
107	Normal	Normal
108	Normal	Normal
109	Normal	Normal
110	Normal	Normal
111	Normal	Normal
112	Normal	***Myocardial Infarction***

113	Normal	Normal
114	Normal	Normal
115	Normal	***Myocardial Infarction***
116	Normal	Normal
117	Normal	***Myocardial Infarction***
118	Normal	Normal
119	Myocardial Infarction	Myocardial Infarction
120	Normal	Normal
121	Myocardial Infarction	Myocardial Infarction
122	Normal	Normal
123	Normal	Normal
124	Myocardial Infarction	Myocardial Infarction
125	Normal	***Others***
126	Normal	Normal
127	Myocardial Infarction	Myocardial Infarction
128	Myocardial Infarction	Myocardial Infarction
129	Myocardial Infarction	Myocardial Infarction
130	Myocardial Infarction	Myocardial Infarction
131	Myocardial Infarction	Myocardial Infarction
132	Others	***Normal***
133	Normal	Normal
134	Normal	Normal
135	Myocardial Infarction	Myocardial Infarction
136	Myocardial Infarction	Myocardial Infarction
137	Myocardial Infarction	Myocardial Infarction
138	Normal	Normal
139	Myocardial Infarction	Myocardial Infarction
140	Others	***Myocardial Infarction***
141	Myocardial Infarction	Myocardial Infarction
142	Myocardial Infarction	Myocardial Infarction
143	Myocardial Infarction	Myocardial Infarction

Appendix B

TABLE 6. THE VARIATION OF K VALUES

No	Real diasease	K Variation										
		K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
1	M	M	M	M	M	M	M	M	M	M	M	M
2	M	M	M	M	***N***	***N***	***N***	M	***N***	***N***	***N***	***N***
3	M	M	M	M	***N***	M	***N***	***N***	***N***	***N***	***N***	***N***
4	M	M	M	M	M	M	M	M	M	M	M	M
5	M	M	***N***	M	M	M	M	M	M	M	M	M
6	M	M	***N***	M	M	M	M	M	M	M	M	M
7	M	M	M	M	M	M	M	M	M	M	M	M
8	M	M	M	M	M	M	M	M	M	M	M	M
9	M	M	M	M	M	M	M	M	M	M	M	M
10	M	M	M	M	M	M	M	M	M	M	M	M
.												
.												
.												
135	M	M	M	M	M	M	***N***	***N***	M	M	M	M
136	M	M	M	M	M	M	M	M	M	M	M	M
137	M	M	M	M	***N***	***N***	***N***	***N***	***N***	***N***	***N***	***N***
138	N	N	***M***	N	N	N	N	N	N	N	N	N
139	M	M	M	M	M	M	M	M	M	M	M	M
140	L	L	***M***	***M***	L	L	***N***	L	***N***	***N***	L	L
141	M	M	M	M	M	M	M	M	M	M	M	M
142	M	M	M	M	M	M	M	M	M	M	M	M
143	M	M	M	M	M	M	***L***	***L***	***N***	***L***	***L***	***N***
Perbedaan Diagnosis		0	37	26	40	43	55	46	53	50	48	51

References

[1] WHO, Cardiovascular Disease,2012.

[2] Menkes RI,"Situasi Kesehatan Jantung," Pusat Data dan Informasi Kemenkes RI, 2014.

[3] Kinckmerova, Metodhes for Detection and Classification in ECG analysis. Doctoral Dissertation, Brno University of Technology, Brno, 2009.

[4] K.O. Gupta & P.N. Chatur, "ECG Signal Analysis and Classification using Data Mining and Artificial Neural Networks, "International Journal of Emerging Technology and Advanced Engineering, Vol 2, Issue 1, 2012.

[5] J.Kaur & J.P.S Raina, "An Intelligent Diagnosis System for Electrocardiogram (ECG) Images Using Artificial Neural Network (ANN)," International Journal of Electrical, Electronics and Computer Engineering 1(1):47-51(2012)

[6] J.Kaur & J.P.S Raina, "An Intelligent Diagnosis System for Electrocardiogram (ECG) Images Using Artificial Neural Network (ANN)," International Journal of Electrical, Electronics and Computer Engineering 1(1):47-51(2012)

[7] Priyadarshini, ECG Signal Analysis: Enhancment and R-Peak detection. Natioal Institut of Rourkela, India, 2010.

[8] K.Huang & L. Zhang, "Cardiology knowledge free ECG feature extraction using generalized tensor rank one discriminant analysis,"EURASIP Journal and Advances Signal Processing, 2014.

[9] M.H.Imtiaz & M.A. Kiber, "Design and Implementation of a Real-Time Automated ECG Diagnosis (AED) System," Gobal Journal of Researches in Engineering and Electronics Engineering, Vol 13 Issue 11,2013.

[10] J.L. Willems, et al., "The Diagnostic Performance of Computer Programs for the Interpretation of Electrocardiograms", New England Journal of Medicine, 325:1767-1773,1991.