JSiI | Jurnal Sistem Informasi
Vol. 11 No. 2 September 2024, Hal.139-142

DOI 10.30656/jsii.v11i2.9164
p-ISSN: 2406-7768
e-ISSN: 2581-2181

UTILIZATION OF K-NEAREST NEIGHBOR ALGORITHM TO ANALYZE AND CLASSIFY HEART DISORDERS BASED ON ELECTROCARDIOGRAM RECORDING DATA

Sumiati [1*], Hanif Nurmajid [2], Muhammad Ibrohim[3] Hendry Gunawan[4]

[1, 2, 3, 4] Teknik Informatika Fakultas Teknologi Informasi Universitas Serang Raya
Jln. Raya Cilegon Serang – Drangong Kota Serang

[1] sumiatiunsera82@gmail.com
[2] hanifnurmajid@gmail.com
[3] *Muhamad.ibrohim01@gmail.com*
[4] hendrygunawan@unsera.ac.id.

Abstract

This study develops a system to classify heart conditions based on electrocardiogram (ECG) medical records using the K-Nearest Neighbor (KNN) method. This system aims to assist medical personnel, especially doctors, in analyzing ECG results more efficiently, considering the limited number of doctors and practice schedules, with the KNN method, the system can classify heart conditions based on the proximity of the patient's ECG data to other ECG data whose conditions are already known. The results of this study have an accuracy of 80%, a value of 0.88 on the Success Rate and 0.54 on Kappa. This study provides a significant contribution in the use of technology to improve the efficiency of heart examinations. This KNN-based system can be used as a tool in the diagnostic process, considering the limited medical resources. In the future, the development of this system can be done by increasing the amount of data, more complete features, or trying other more complex classification methods to improve accuracy and Kappa.

Keyword: Heart Disorders, Classification, K-Nearest Neighbor, Success Rate and Kappa Statistic

I.INTRODUCTION

Health is a very valuable asset for every individual, because everyone is at risk of experiencing health problems. Therefore, it is important for us to maintain our health to avoid various diseases, including heart disease [1]. The heart is a vital organ in the human body that plays a major role in pumping blood throughout the body. The blood pumped by the heart functions as a transporter of nutrients and oxygen needed by other organs of the body [2][3]. Heart disorders refer to conditions that affect the blood vessels of the heart, heart rhythm, heart valves, or disorders caused by congenital factors. Some types of heart disorders include coronary heart disease, heart failure, arrhythmia, and congenital heart disease. According to the World Health Organization (WHO) report in 2005, of the total 58 million deaths worldwide, around 17.5 million (30%) were caused by heart and blood vessel disease, with heart attacks (7.5 million) and strokes (5.7 million) being the main causes. It is estimated that in 2015, deaths from heart and blood vessel disease will increase to 20 million[5][6].

An electrocardiogram (ECG) is a tool used to check the health of the heart. This tool is usually only available in certain hospitals or clinics, and only a few nurses or specialist doctors have the ability to operate and diagnose the results of an ECG examination. At Sari Asih Hospital, Serang, Banten, a heart examination using an electrocardiogram takes about 15 to 20 minutes. After the ECG results are available, the doctor will analyze the medical records to determine the diagnosis of the patient's heart condition. Patients who are diagnosed with heart abnormalities will be scheduled for further examination[7][8].

Various studies in the field of computing have been carried out to contribute to the world of health, especially in analyzing heart disorders. A heart disease prediction system using data mining has a number of benefits, such as prevention, diagnosis, treatment planning, and reducing patient management costs[9][10]. The research conducted by Gunsai Dineshgar and Lolita Singh succeeded in developing a heart disease prediction system modeling, which can help health practitioners in making more appropriate clinical decisions [12]. the K-Nearest Neighbors (K-NN) algorithm is a method for classifying objects based on training data that has the closest distance to the object being analyzed. K-Nearest Neighbor is based on the concept of "learning by analogy". Some of the advantages of the K-NN algorithm include: its resistance to training data containing a lot of noise and its effectiveness when used on large amounts of training data [13][14].

This study develops a system to classify heart conditions based on electrocardiogram (ECG) medical records using the K-Nearest Neighbor (KNN) method. This system aims to help medical personnel, especially doctors, in analyzing ECG results more efficiently, considering the limited number of doctors and

Received: July 15th, 2024| Published: September 15th, 2024

DOI 10.30656/jsii.v11i2.9164
p-ISSN: 2406-7768
e-ISSN: 2581-2181

practice schedules, with the KNN method, the system can classify heart conditions based on the proximity of the patient's ECG data to other ECG data whose conditions are already known.

II.RESEARCH METHODOLOGY

The research stages consist of Literature Study, Data Collection, Initial Classification of Electrocardiogram Data, System Analysis, System Design and Creation, System Testing, and Accuracy Level Testing.

III. RESULTS AND DISCUSSION

This study examines how to diagnose heart disorders based on Electrocardiogram (ECG) recording results by applying the K-Nearest Neighbor method.In this study, heart disorders can be predicted using Electrocardiogram (ECG) data through the K-Nearest Neighbor method. A total of 100 data were used, with 80 data for training and 20 data for testing, the results ofwhich were calculated manually using the K-Nearest Neighbor method.

a. Determining the K Value

In this study, the K value used is 5.

b. Calculate the Distance Between New Data and Each Labeled Data or Training Data.

The distance calculation uses the Euclidean Distance formula.

$$d_i = \sqrt{\sum_{i=1}^{P}(x_{2i} - x_{1i})}^{\ 2}$$

Table 1. Testing Data Distance Calculation 16

No	HR	P-R	QRS	QT	QTC	AXIS	RV6	SV1	R+S	Jarak	Diagnosa Pakar
1	134	157	95	318	433	153	0.58	0.32	0.9	209.075	AbNormal
2	106	169	99	329	439	41	2.36	1.59	3.95	96.1609	AbNormal
3	137	271	113	385	420	66	0.89	0.96	1.85	178.394	AbNormal
4	140	212	99	336	412	57	1.08	0.12	1.2	139.378	AbNormal
5	136	158	98	390	433	17	1.1	0.5	3.32	103.3914	AbNormal
6	134	136	110	272	406	40	1.19	2.13	3.32	142.886	AbNormal
7	118	169	100	314	441	61	2.36	1.59	3.95	120.5691	AbNormal
8	124	151	99	378	458	67	2.38	1.28	3.66	127.9336	AbNormal
9	140	266	80	394	420	51	0.91	0.22	1.13	166.0758	AbNormal
10	125	152	108	373	417	63	2.26	0.94	3.2	132.6627	AbNormal
11	100	150	86	358	462	38	0.25	-1.12	1.66	87.66497	AbNormal
12	84	92	398	470	470	26	-0.74	1	3	360.1269	AbNormal
13	114	104	96	282	388	-11	2.86	-0.424	3.77	124.834	AbNormal
14	102	140	110	340	442	43	1.633	-0.551	0.57	100.9921	AbNormal
15	115	135	84	340	471	54	1.43	0.97	2.4	109.784	AbNormal
16	66	130	98	380	398	26	2.341	-1.904	1.65	107.843	AbNormal
17	85	136	102	365	435	-4	2.6	-1.096	1.97	64.44331	AbNormal
18	133	112	92	290	430	42	1.9	-1.096	86	156.6127	AbNormal
19	100	174	119	349	450	69	1.24	0.42	1.66	123.4634	AbNormal
20	75	126	135	398	447	99	2.26	0.11	2.37	167.8602	AbNormal
21	122	116	87	291	415	44	0.93	0.28	1.21	130.5262	AbNormal
22	88	146	91	343	417	31	2.05	1.9	3.95	89.74325	AbNormal
23	163	112	93	267	441	122	0.72	1.04	1.76	210.3626	AbNormal
24	150	131	98	365	391	65	1.29	0.34	1.63	155.1266	AbNormal
25	132	61	73	301	128	131	0.4	0.4	0.8	395.2424	AbNormal
26	133	112	82	290	430	43	1.9	1.98	1.36	131.6024	AbNormal
27	64	258	96	442	456	73	1.19	-2.16	3.36	180.1226	AbNormal
28	141	82	100	278	426	20	1.32	-1.95	3.27	145.4195	AbNormal
29	114	104	96	282	388	-11	2.86	-0.42	3.77	124.8339	AbNormal
30	128	82	100	360	524	43	0.35	-0.32	0.96	147.8451	AbNormal
31	127	80	110	372	380	53	1.49	-0.75	1.66	165.2818	AbNormal
32	100	82	100	366	366	51	2.8	-2.36	3.77	161.5509	AbNormal
33	160	81	100	349	570	69	3.94	-0.86	1.66	198.4198	AbNormal
34	132	91	73	320	475	13	1.04	0.96	3.36	111.929	AbNormal
35	126	148	98	285	418	40	2.25	1.45	3.7	124.4311	AbNormal
36	102	207	100	353	420	38	1.26	1.29	1.55	106.7557	Normal
37	102	142	87	316	413	57	0.61	1.08	1.63	119.6973	Normal
38	77	180	110	376	428	39	1.14	1.54	2.68	100.9494	Normal
39	85	170	91	357	425	6	0.99	0.41	1.4	64.46725	Normal
40	87	153	90	346	412	51	0.95	0.41	1.36	107.7266	Normal
41	121	61	73	301	128	131	0.4	0.4	0.8	393.8065	Normal
42	110	117	81	294	398	24	1.2	0.61	1.81	119.5504	Normal
43	102	156	97	355	464	-36	1.02		1.52	37.68343	Normal
44	99	136	98	337	433	37	1.01	0.73	1.74	94.98519	Normal
45	103	155	81	349	458	0	0.98	0.88	1.86	52.61523	Normal
46	91	146	98	343	422	-36	1.02	0.56	1.52	50.57708	Normal
47	76	153	102	362	403	20	0.64	0.26	0.9	91.32557	Normal
48	69	157	83	369	397	80	0.76	0.28	1.04	140.9155	Normal
49	76	150	96	446	338	422	-0.74	1.2	1.8	492.05	Normal
50	95	144	100	252	316	65	1.38	-0.42	3.15	205.534	Normal
51	98	178	98	338	430	32	1.88	-0.88	2.37	88.16995	Normal
52	101	98	98	420	420	97	1.72	2.14	1.82	180.2652	Normal
53	85	146	94	344	408	26	1.5	2.29	1.98	90.647	Normal
54	64	82	432	446	446	115	-0.73	1.88	1.18	411.2778	Normal
55	93	146	90	338	420	126	0.45	-0.5	1.66	176.1882	Normal
56	71	142	88	374	406	13	2.53	-0.42	2.95	87.18987	Normal
57	60	146	98	408	408	43	1.58	-0.42	3.4	123.4278	Normal
58	73	166	92	362	414	39	1.09	0.47	1.56	97.0983	Normal
59	74	137	104	365	407	40	1.43	0.59	2.02	107.9524	Normal

Received: July 15th, 2024| Published: September 15th, 2024

DOI 10.30656/jsii.v11i2.9164
p-ISSN: 2406-7768
e-ISSN: 2581-2181

60	100	124	124	266	342	115	1.14	-1.71	1.82	222.5	Normal
61	70	130	88	371	411	42	-1.31	0.62	1.98	108.0977	Normal
62	87	150	99	371	448	87	1.07	1.42	1.18	136.539	Normal
63	63	129	102	375	403	48	2.04	1.07	1.66	120.5726	Normal
64	103	152	99	374	492	31	1.5	0.47	1.97	94.11554	Normal
65	94	150	96	338	422	43	1.2	-0.39	1.97	99.16791	Normal
66	93	121	97	357	445	44	1.82	0.75	1.56	102.8493	Normal
67	91	149	114	297	297	-8	1	0.92	1.92	178.6854	Normal
68	87	119	98	325	325	54	1.21	0.47	1.97	175.067	Normal
69	77	146	104	360	360	43	1.18	-0.88	1.66	136.9895	Normal
70	98	175	135	380	380	111	2.82	0.25	3.02	188.5458	Normal
71	93	146	90	338	420	126	0.456	-0.5	86	195.4509	Normal
72	71	142	88	374	406	13	2.53	-0.424	2.954	87.18998	Normal
73	60	146	98	408	408	43	1.58	-0.424	3.4	123.4279	Normal
74	82	143	97	367	429	-3	1.57	0.5	2.07	63.41887	Normal
75	79	158	93	360	414	14	0.62	0.94	1.56	77.21572	Normal
76	73	114	86	416	403	64	2.23	2.72	0.72	149.9771	Normal
77	76	146	110	418	472	-4	2.04	1.34	3.38	92.88704	Normal
78	94	184	86	372	465	78	0.48	0.08	0.56	127.7186	Normal
79	73	159	105	387	428	14	1.02	0.73	1.75	83.61935	Normal
80	105	130	100	334	442	54	1.23	2.42	3.65	111.6864	AbNormal

c. Determine K Labeled Data That Has the Minimum Distance

Table2.Classification of K Values of Testing Data 16

No	HR	P-R	QRS	QT	QTC	AXIS	RV6	SV1	R+S	Jarak	Diagnosa Pakar
43	102	156	97	355		-36	1.02	0.56	1.52	37.68343	Normal
46	91	146	98	343	422	-36	1.02	0.56	1.52	50.57708	Normal
45	103	155	81	349	458	0	0.98	0.88	1.86	52.61523	Normal
74	82	143	97	367	429	-3	1.57	0.5	2.07	63.41887	Normal
39	85	170	91	357	425	6	0.99	0.41	1.4	64.46725	Normal

d. Classification of New Data into Majority Labeled Data.

From the calculation results above, it is known that the tested data falls into the "Normal" class, because the closest distance value from the five neighbors (K) shows the "Normal" class.

Calculating the Level of Accuracy

After analyzing the data, it is necessary to find the level of accuracy to find out whether the results of the analysis that have been done have a high or low level of accuracy. The method used to calculate the accuracy value in this study is the success rate method and kappa statistic.

Success Rate

Success rate (Sr) is a measure of performance measurement that is usually used in classification techniques. How successful the model formed is in classifying a number of data based on the rules generated from the model formed. The measure of success is seen if the class of a data can be predicted correctly, conversely if the class of a data cannot be predicted correctly, then it is said to be a failure/Error.

To measure the SR value, the confusion matrix approach is used as follows:

		Predicated Class	
		Normal	Abnormal
Actual Class	Normal	12(TP)	0 (FN)
	Abnormal	4 (FP)	4 (TN)

True Positive rate =TP/(TP+FN)
= 12 / (12+0)= 1

False Positive Rate =FP/ (FP+TN)
= 4 / (4+4)= 0.5

Success rate = (TP+TN)/(TP+TN+FP+FN)
= (12+4)/ (12+4+4+0)= 0.8

Error rate =1 – Success Rate.
= 1- 0.8= 0.2

Kappa Statistics

Kappa statistics is a method used to compare the level of accuracy of observed data with predicted data. Kappa is usually used to evaluate the classification results formed. The Kappa value interval ranges from -1 to +1. If the Kappa value = 1 then it is certain that the classification results are very accurate, the further away from the number 1 it can be said that the accuracy is decreasing. The following is the equation for calculating the Kappa value.

$$Kappa = \frac{observed\ accuracy - expected\ accuracy}{(1 - expected\ accuracy)}$$

$$Kappa = \frac{(1 - 0,5)}{(1 - 0,5)}$$

$$Kappa = 1$$

IV.CONCLUSIONS

Based on research findings and the results of testing the classification of heart disorders using the K-Nearest Neighbor method, it can be concluded that:

1. The classification system developed using the K-Nearest Neighbor method was successfully built using the PHP programming language. This system is able to perform the classification process well. After testing using the black box testing method, it can be concluded that the system built functions well and is in accordance with expectations.

Received: July 15th, 2024| Published: September 15th, 2024

DOI 10.30656/jsii.v11i2.9164
p-ISSN: 2406-7768
e-ISSN: 2581-2181

2. 2. The components of the Electrocardiogram (ECG) used as variables in this study include HR value, P-R value, QRS value, QT value, QTC value, AXIS value, RV6 value, SV1 value, and R+S value. The data used in this study amounted to 100 samples, with 80 data for training and 20 data for testing. The results of this study showed an accuracy of 80%, with a Success Rate value of 0.88 and a Kappa value of 0.54.

V. SUGGESTION

1. Increase the amount of training data, because in the application of the K-Nearest Neighbor method, the more training data used, the higher the level of accuracy of the results. In this study, 80 data were used for training and 20 data for testing, where from 20 testing data, there were 4 data whose classification results did not match the predictions of the experts.
2. The web-based classification system that was built has several shortcomings and still requires development, both in terms of features and security aspects. Therefore, further research can lead to the development of this system in the form of a desktop or Android application.

REFERENCE

[1] Anggara, Mario. (2016). Pemilihan Distance Measure Pada *K-Means Clustering* Untuk Pengelompokkan Member Di Alvaro Fitness. Jurnal Sistem dan Teknologi Informasi (JUSTIN), Vol.1 (No.1), 1-6.

[2] Anthony, J.V dan Joannem, M.G. (2005). *Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine*, Vol.37 (No.5),360-363.

[3] Dineshgar, Gunsai Pooja dan Singh, Lolita. (2016). *A Reveiw On Data Mining For Heart Disease Prediction*. *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, Vol.5 (No.2) 462-466.

[4] Dr. Eng, R. H., dan Sianipar, ST, M.Eng.(2018). Pemograman Database Menggunakan MySql. Yogyakarta : ANDI OFFSET.

[5] Iswanto, ST,.(2007). Membangun Aplikasi Berbasis PHP 5 Dan Firebird 1.5. Yogyakarta : ANDI OFFSET

[6] Kementrian Kesehatan RI. (2011), Pedoman Pengendalian Faktor Resiko Peyakit Jantung Dan Pembuluh Darah, Surat Keputusan Menteri Kesehatan Republik Indonesia Nomor 854/Menkes/SK/IX/2009. Diakses pada 4 April 2020 pukul 20:00 WIB. www.kemenkes.go.id

[7] Mariana, N., dkk.(2015). Penerapan Algoritma K-NN (*Nearest Neighbor*) Untuk Deteksi Penyakit (Kanker Serviks). Dinamika Informatika. Vol.7 (No 1), 26-34.

[8] Maulana, dkk. (2018). Rancang Bangun Instrumentasi Elektrokardiograf (EKG) dan Klasifikasi Kenormalan Jantung Pada Pola Sinyal EKG Menggunakan *Learning Vector Quantization* (LVQ). Publiser : *JITCE (Journal of Information Technology and Computer Engineering)*, Vol.2 (No.1) , 20-26.

[9] Musa, O., dan Alang. (2017). Analisis Penyakit Paru-Paru Menggunakan Algoritma K-Nearest Neighbors Pada Rumah Sakit Aloei Saboe Kota Gorontalo. ILKOM Jurnal Ilmiah, Vol.9 (No 3), 348-352.

[10] Napitupulu, Darmawan (2014), Studi Validitas Dan Realibilitas Faktor Sukses Implementasi EGoverentment Berdasarkan Pendekatan Kappa. Jurnal Sistem Informasi *(Journal of Information Systems)*, Vol.2 (No.10), 70—74.

[11] Prasetyo, Eko. (2012). DATA MINING Konsep Dan Aplikasi Menggunakan MATLAB. Yogyakarta: ANDI.

[12] Rohman, A.,(2016). Komporasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit

[13] Safriandono, A. N., (2017). Algoritma *K-Nearest Neighbors* Berbasis *Forward Selection* Untuk Mendiagnosis Penyakit Jantung Koroner. Komputaki. Vol.3 (No 1). 1-16.

[14] Sinaga, I. R. P. dkk. (2017). Analisa Kelainan Jantung Menggunakan Sinyal Elektrokardiogram Metode Transformasi Wavelet Dan *K-Nearest Neighbors*. *e-Proceeding of Engineering*. Vol 4 (No 3). 3518-3523.

Received: July 15th, 2024| Published: September 15th, 2024