K-Nearest Neighbor (K-NN) Method for Optimizing Data Training on Diabetes Diagnosis and Chronic Complications

Risky Aswi Ramadhani[1], Ratih Kumalasari Niswatin[2]

[1,2] *Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri*

[1]risky_aswi@unpkediri.ac.id , [2] Ratih.workmail@gmail.com

Abstract— Information technology has entered various fields, one of which is the health sector. Many researchers develop expert systems, medical records, and hospital registration systems. The diagnostic system is a concern for researchers because, with a diagnosis system, patients can consult through the system without visiting a doctor (expert). To make the diagnosis system, need medical record data from patients who have had previous treatment. The data will be used as a source of diagnostic of system knowledge as data training. Chronic diabetes patient data to be tested are called training data. The K-Nearest Neighbor (K-NN) method is used to detect diabetes and chronic complications. By developing this system, it is expected that the number of patients who have chronic complications and diseases that accompany Diabetes can be suppressed. The results of the K-NN method are very optimal if tested inpatient data training with optimal values of y = 0.73 and x = 0.62.

Keywords— *expert system; data training; K-Nearest Neighbor; diabetes diseases; chronic complications*

Teknologi informasi telah masuk ke berbagai bidang, salah satunya adalah bidang kesehatan. Banyak peneliti mengembangkan sistem pakar, catatan medis, dan sistem pendaftaran rumah sakit. Sistem diagnostik menjadi perhatian bagi para peneliti karena dengan sistem diagnosis, pasien dapat berkonsultasi melalui sistem tanpa mendatangi dokter (ahli). Untuk membuat sistem diagnosis dibutuhkan data rekam medis dari pasien yang pernah menjalani pengobatan sebelumnya. Data ini akan digunakan sebagai sumber pengetahuan sistem diagnose sebagai data training. Data pasien diabetes kronis yang akan diuji disebut sebagai data training. Metode K-Nearest Neighbor (K-NN) digunakan untuk mendeteksi diabetes dan komplikasi kronis. Dengan mengembangkan sistem ini diharapkan jumlah pasien yang mengalami komplikasi kronis dan penyakit yang menyertai Diabetes dapat ditekan. Hasil metode K-NN sangat optimal jika diujikan pada data training pasien dengan nilai optimal y=0.73 dan x=0,62.

Keywords— *expert system; data training; K-Nearest Neighbor; diabetes diseases**;** chronic complications*

I. INTRODUCTION

Today information technology has entered into various fields, the number of smartphone users in Indonesia in 2018 reached 100 million people, this number is a very large number [1]. Smartphone has its own appeal because with smartphones Indonesian people can do many things such as listening to music, looking for an address, online transactions, and perform a simple health check. Many applications have been developed by android developers specifically for health such as ideal body weight checks, fertile period checks (for women), and simple diagnoses for some diseases.

Nowadays, researchers are competing to develop technology especially to promote the world of health. Here is an application developed by researchers in the field of IT to advance the world of health systems such as experts, medical records, and hospital registration system. These applications have their own function. One of them is spatial data modeling through geographic information system technology used for mapping tropical diseases in East Java, Indonesia [2] and Decision Tree Classifier early detection of heart disease [3]. This research will focus on expert systems.

The diagnostic system gained special attention among ITU researchers because by developing a system of diagnosing the number of medical personnel who were still inadvertently [4]. Given the very small number of medical personnel in Indonesia and the archipelagic geography of Indonesia, the population spread and the number of medical personnel is uneven. While the number of diabetics in Indonesia reached 10 million. 1.67 million are aged under 40, 4.65 are 40-59 years old, and the rest are over 60 years [5]. The number of medical personnel, when compared to diabetics, is very less. By making a diagnosis system of chronic complications and comorbidities in diabetes

DOI: 10.25139/inform.v3i2.1042

the number of diabetics can be suppressed. With a large number of diabetic patients, surely medical personnel often make mistakes when diagnosing, the purpose-built this application is to accompany medical personnel so it can minimize mistakes. This system can also be used to make an early diagnosis without being accompanied by medical personnel, this is done if the patient is in the hinterland. If the patient is declared diabetic or with complications then the patient is advised to seek a doctor.

The diagnostic system utilizes previously treated patient data stored on the hospital medic record as a source of knowledge. Medic record data is referred to as training data. Diagnosis of disease will be more accurate if the number of sets of training data used more and more [6]. The data of chronic diabetic patients to be tested is referred to as data testing.

The method used to develop a system of chronic complications and diseases in diabetes is K-NN. The K-NN method has the advantage of separating training data and searching for nearest neighbors (data that are similar to data testing) [7].

The diagnostic system to be developed has more value that is the system is able to diagnose diabetes and chronic disease. Because diabetes is a parent disease that can cause disorders of some vital organs such as heart, kidney, eye damage, and nerves [8]. Therefore it is very necessary applications capable of diagnosing diabetes and its complications.

With developed a system capable of diagnosing chronic complications and diseases accompanying Diabetes. It is expected that the number of diabetics can be suppressed. This system requires surveillance due to the development of a diabetic diagnosis system and comorbidities, this system is the first time.

II. RESEARCH METHODOLOGY

The research methodology is the steps that will be used to conduct research on a system capable of diagnosing chronic complications and diseases accompanying Diabetes. Here is a Scheme of research methodology to be used.

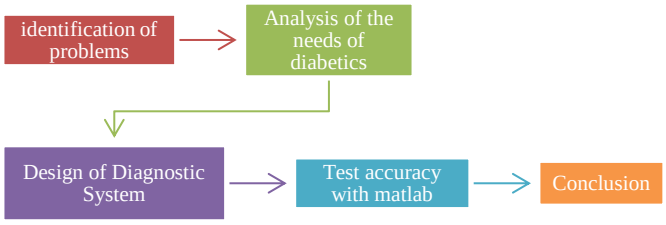

Figure 1. Scheme of Research Methodology

The scheme in Figure 1 describes the diagnostic system for diagnosing chronic complications and accompanying Diabetes.

a. Identification of problems.

From the background written, the identification of problems that will be obtained as research material as follows:

- Lack of medical personnel in Indonesia.
- The development of diabetes diagnostic system is only limited to diagnose diabetes does not speak chronic complications and comorbidities.

b. Analysis of the needs of diabetics

This research activity is used to identify the requirement of the diagnosis system of chronic complication and diabetes accompanying disease, the following is the result of required identification.

- Functional needs analysis, functional requirements analysis covers matters relating to the development of chronic complication diagnosis system and diabetic disease and processes
- Analysis of the need for non-functional, discussing things that are not related to technology such as user behaviour, user environment, and the behaviours that are owned by the diagnosis system of chronic complications and diseases accompanying diabetes.

c. Design of Diagnostic System

The design of this system is divided into two processes namely the flowchart of a system that will be designed with the application of Power Designer and System Interface design.

d. Test the accuracy

Benefits of accuracy tests with MATLAB software to test whether the K-NN method is truly capable of confronting the diagonal system of chronic complications and diabetic comorbidities

e. Conclusion

This section is used to deduce whether using the K-NN method has high accuracy when used to diagnose chronic copies and diabetes disease.

III. RESULTS AND DISCUSSION

Here is a discussion of diagnosing chronic complications and diseases accompanying Diabetes, in this section, will be discussed about the processes undertaken at the time of research.

A. System planning

a.1 Network Diagram

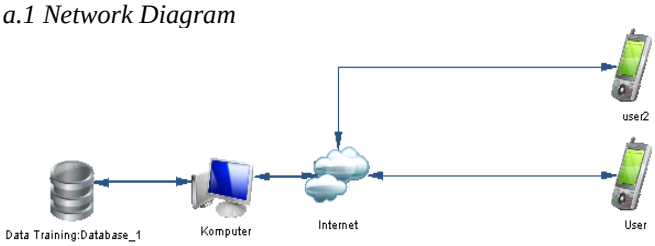

Figure 2. Network Diagram

In Figure 2 it is explained that the developed diagnostic system can be accessed via handphone because the diagnostic system used is web-based. This diagnostic system has training data stored on the server, this training data comes from the medic record of patients with diabetes before.

a.2 Flowchart

DOI: 10.25139/inform.v3i2.1042

The flowchart is a chart that is used to describe certain procedures, especially those related to chronic complication diagnosis and comorbidities in diabetes. The following is the system flowchart used. Figure 3 describes the flowchart of the diagnosis system of chronic complications and comorbidities of Diabetes, the system has a data store used to store training data, all the diagnostic processes undertaken by the system must take the training data from the data store. This system has several processes namely the input of diabetes symptoms, diabetes diagnoses, making decisions about diabetes, symptomatic inputs related to diabetic complications, diabetic process diagnosis, and results of diagnosing complications.

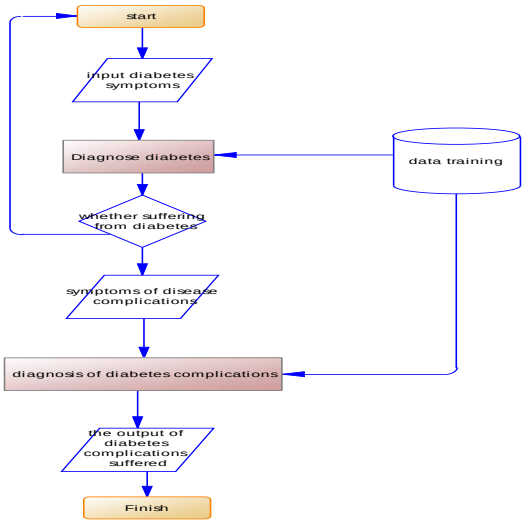

Figure 3. Flowchart of diagnosis system

a.3 Data Flow Diagram

Data Flow diagrams (DFDs) are models used to describe functional processes that are interconnected with each other. Each entity in the DFD has different tasks and tasks. For more dampers can be seen as in Figure 4.

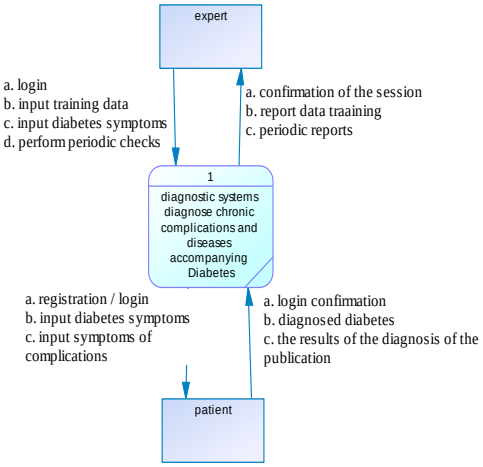

Figure 4. Data flow diagram

Data Flow Diagram (DFD) in Figure 4 has 2 entities ie patient and doctor. patients want to diagnose whether suffering from diabetes, patients are required to register first if not yet have an account, after which the patient inputs the symptoms that are felt and the system gives the diagnosis result. While experts are on duty to monitor the results of diagnosis, the diagnosis of this system is still dangerous if released alone because it is directly related to humans. This system is a first-of-its-kind diabetic diagnostic system that features chronic complications and comorbidities. So when this system operates the need for supervision.

B. Diagnosa Diabetes

To diagnose diabetes are used some of the symptoms among others

- Often BAK at night although not drinking much water
- Weight loss drastically, because the fat in the body lost because converted into sugar in the blood
- In case of injury or irritation to the skin for healing takes a long time.
- The body often experience fatigue
- Gum Often Bloody
- Often Haus and easily hungry.
- A numb leg like a needle stabbed

C. K-NN

K-NN method is a method that finds the closest distance between data training (sample) with data testing. The training data that has the closest distance is considered to have the same high properties so that the K-NN Method can decide the diagnosis.

$$d = \sum_{i=1}^{p} (x_i - x_{|i})^2 \quad \text{.....(1)}$$

$x_i = Sample\ data$
$x_2 = Testing$
$i = Variable$
$d = Distance$

D. Diabetes Diseases

d.1 Data Training diabetes

Data training is data derived from, data diagnosis results of previous diabetes patients. Data training serve as a source of knowledge for the system, the more training data used the diagnosis of the system will be better. The following is the training data used.

Table 1. Diabetes Training Data

id	Symptoms suffered							results
	a	b	c	d	e	f	g	
12	y	n	y	n	n	y	y	+
13	ya	y	t	y	y	y	n	+
14	n	n	n	n	n	y	n	-

where,

a= Often BAK
b=Weight loss
c=The wound is hard to heal
d=Tired easily

DOI: 10.25139/inform.v3i2.1042

e=Bleeding gums
f=Often thirsty
g=Numb legs

d.2Diabetes Testing Data

Data testing on chronic complication diagnoses and comorbidities in diabetes is patient data to be diagnosed, the diagnosis of this patient is unknown.

Table 2. Data testing table

id	Symptoms suffered							results
	a	b	c	d	e	f	g	
40	y	y	y	n	n	y	y	?

where,
a= Often BAK
b=Weight loss
c=The wound is hard to heal
d=Tired easily
e=Bleeding gums
f=Often thirsty
g=Numb legs

d.3 Results of treatment with K-NN for diabetes

Training data with Id = 40 will be diagnosed, patients with id = 40 positive with diabetes or not. The following is the result of processing K-NN method done on matlab tools

- Source code of diabetes diagnosis

  To prevent symptoms of diabetes, so a diabetic diagnosis system used the following sourcode.

```
%% load data from file "ha.txt"
clear; clc;load('ha.txt')
N = ha(:,1:2);
%% plot point data N(x,y)
gscatter(N(:,1),N(:,2));  xlabel('x');ylabel('y');
%% Mengambil data
label=cellstr(num2str([1:length(N)]'));
text(N(:,1),N(:,2),label,'VerticalAlignment','bottom','HorizontalAlignment','right');grid on; %show label for all pts;
%% menapilkan 5 nilai yang paling dekat
[n,d] = knnsearch(N,newpoint,'k',5);
```

- The plot of diabetes diagnosis

  The following is an image of training data and data testing, the x mark is data testing and the red colored point is training data, the closest training data is given a red mark and a circle is given. For more details can be seen in the picture below.

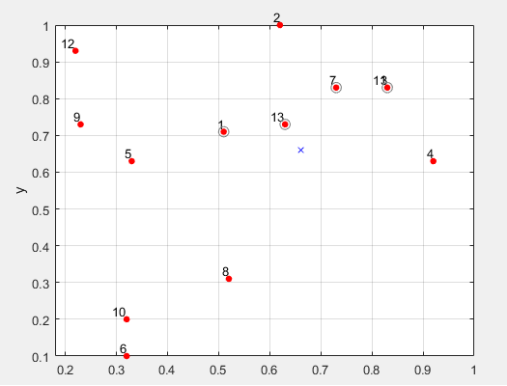

Figure 5. Diabetes Plots

In Figure 5 it is explained that patients with those used as data testing are said to have diabetes, data testing is marked with a blue cross. The patient has a very close proximity to a patient marked 13 having a coordinate value of y = 0.73 and x = 0.62. The patient was tested positive for diabetes. From the results of this diagnosis patient who are not as data testing stated suffering from diabetes, but for the sake of security needs to be consulted to the doctor first.

E. Complications of diabetes

e.1 Data Testing of diabetes complications

Diabetes is the mother of all diseases, there are some diseases that arise due to diabetes such as heart, neuropathy, and kidney. The following are examples of symptoms and training data.

Table 3 data training diabetic complications

no	a	b	c	d	e	f	g	h	i	j	k	l	kom
1	1	1	1	1	0	0	0	0	0	0	0	1	ne
2	1	0	1	1	1	1	1	1	0	0	0	0	gj
3	1	0	0	0	0	0	0	0	1	1	1	1	jt

where,
A = no appetite
B = weight down
C = difficulty sleeping
D = muscle cramps
E = dizziness nausea
F = foot clearing
G = morning sickness
H = itch throughout the body
I = sick head
J = chest pain
K = difficulty breathing
L = easy lethargy
ne = neurofati
gj = kidney
jt = heart

e.2 Data Testing of diabetes complications

Data testing on chronic complications and comorbidities in diabetes is patient data to be diagnosed, the patient's diagnosis is unknown

Table 4 testing diabetic complications

no	a	b	c	d	e	f	g	h	i	j	k	l	kom
1	1	1	1	1	0	0		0	0	0	0	0	?

where,
A = no appetite
B = weight down
C = difficulty sleeping
D = muscle cramps
E = dizziness nausea
F = foot clearing
G = morning sickness
H = itch throughout the body

DOI: 10.25139/inform.v3i2.1042

I = sick head
J = chest pain
K = difficulty breathing
L = easy lethargy

e.3 Results of treatment with K-NN for diabetic complications

The following is an image of training data and data testing, the x mark is data testing and the red colored point is training data, the closest training data is given a red mark and a circle is given. For more details can be seen in the figure 6. In Figure 6 it is explained that patients with those used as data testing are said to suffer from compilation Nephropathy and heart, the patient's value is y = 0.4. And x = 0.42. This diagnosis is not appropriate because the points obtained between nefopam and heart are the same. Expected in the study, can know one of the complications. The K-NN method is less effective when used to select one of the complications.

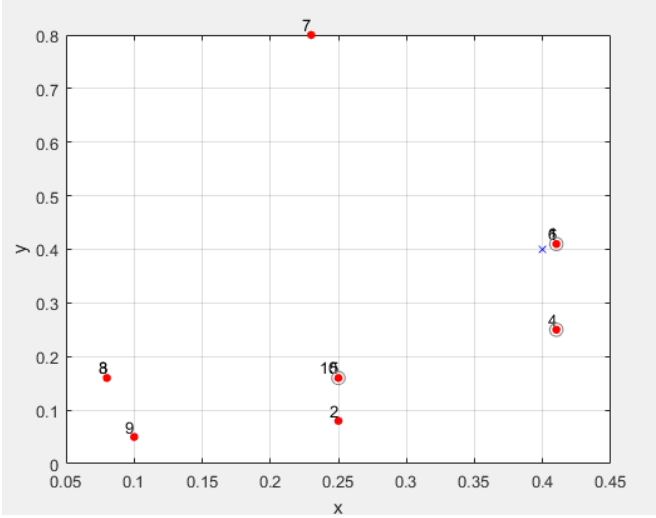

Figure 6. The plot of diabetes complications

IV. CONCLUSION

K-NN method is very optimal when used to diagnose diabetics. Patients who act as data testing are very close to the patient marked 13 with y = 0.73 and x = 0.62. from the diagnosis, the patient is stated suffering from diabetes. for the latter, the pessimist is diagnosed whether to suffer from complications. KNN method is less than maximal if used for determine diabetics also suffer complication because when deciding result there are two decision of disease that is kidney and nephropathy this can happen because both diseases have the same coordinate that is y = 0.4. And x = 0.42. To improve this research, there needs to be a second stage improvement that is the process to diagnose chronic complications of diabetics.

THANK-YOU NOTE

Alhamdulillah this research can run smoothly, although there are some findings that are less appropriate and need to be improved. I thank the Kemenristek-DIKTI who helped funded this research through Beginner Lecturer Research program and I thank the parties who helped in this study which cannot be mentioned one by one.

REFERENCE

[1] Novalius, Feby, "Indonesia Pengguna Smartphone Ke-4 Dunia, Begini Tekad Menperin Dongkrak Industri Telematika," Okezone. Sabtu 17 Februari 2018 11:10 WIB

[2] V. Vitianingsih, D. Cahyono, and A. Choiron, "Analysis and design of web-geographic information system for tropical diseases-prone areas: A case study of East Java Province, Indonesia," in *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2017, pp. 255–260.

[3] Kacung, S., & Prihartono, E. (2016). Sistem Cerdas untuk Mendeteksi Dini Penyakit Jantung Dengan Decision Tree. *Jurnal INFORM*, *1*(2), 108–111. ISSN: 2502-3470. E-ISSN: 2581-0367.

[4] S. Supartini and Hindarto, "Sistem Pakar Berbasis Web Dengan Metode Forward Chaining Dalam Mendiagnosis Dini Penyakit Tuberkulosis di JawaTimur," Universitas Muhamadiah Sidoarjo , vol. I no.3 , ISSN : 2503-2259, E-ISSN : 2503-2267, 2016

[5] W.W, Sherine , "Miris, Indonesia Peringkat 7 Pasien Diabetes Terbanyak di Dunia," Kompas. 11/11/2017, 17:04 WIB

[6] N.A, Aprelia, Identifikasi Kualitas Beras dengan Citra Digital, vol. 2. No.1 2015

[7] H. Risa, R. Siti, and R.A. Ramadhani," Perancangan Sistem Diagonosa Penyakit Hepatitis Menggunakan Metode K-NN,"Ilkom. Volume 9 Nomor 2 Agustus 2017

[8] E. A.Dwi , A Maswan,E Efrida ," Pola Komplikasi Kronis Penderita Diabetes Melitus Tipe 2 Rawat Inap di Bagian Penyakit Dalam RS. Dr. M. Djamil Padang Januari 2011 - Desember 2012", Jurnal Kesehatan Andalas, 2015

DOI: 10.25139/inform.v3i2.1042