The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00001 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Application of the *K-Nearest Neighbor* Method for Classification of Leiomyoma (Myoma)

Selly Alfika[1],Mukti Qamal[2],Zahratul Fitri [3]
[1,2,3]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia,
[1]selly.200170075@mhs.unimal.ac.id , [2]mukti.qamal@unimal.ac.id , [3]zahratulfitri@unimal.ac.id

Abstract
Information technology is very important in the process of human life. Along with the growth opens up opportunities for relatively large data growth, one of which is hospitals. Mioma is a disease that continues to increase and has a major impact on health female reproduction. Myoma is a benign tumor that grows in or around the uterus. Mioma is a medical condition experienced by women of all ages, but is often experienced by women who have entered pre-menopause, myoma is also the second benign tumor in Indonesia by age range sufferers 20-50 years old. Sufferers rarely cause specific symptoms so women are rarely aware of them the presence of myoma growth in their uterus. This research classifies patient data with a purpose to classify types of myoma disease using the K-Nearest Neighbor method. There are several The attributes used in this research are diastolic blood pressure, systolic blood pressure, hemoglobin, ever been pregnant, symptoms 1 and symptom 2. The data used for this research amounted to 288 myoma patient data which will be divided into 2, namely 70% training data and 30% testing data. Then it is divided into 3 classes, namely intramural myoma, submucosal myoma, and subserosal myoma. Results of myoma classification using the K-Nearest algorithm Neighbor at Aceh Tamiang Regional Hospital used 87 test data or patient data, indicating people with the disease Intramural myoma are more numerous with 48 data, subserosal myomas 15 data and for subserosal myomas there are 25 data with a high accuracy rate of 93%.
Keywords: Classification,*K-Nearest Neighbor,Myoma.*

Introduction
Data mining is a data processing process using statistical techniques, mathematics, artificial intelligence and machine learning, to extract and identify useful information and relevant knowledge from large database[1]. There are several techniques that can be used in data mining to find information new information, one of the data mining techniques is classification[2].

Classification is a method of systematically grouping data according to rules and regulations that has been determined, Classification is also defined as grouping new data or objects that have not yet been determined class or category is known[3]. Classification is a statistical method for grouping or classify data that is arranged systematically. Grouping rules are used for classify new data into existing groups[4]. In this research the classification system is used is the K-Nearest Neighbor algorithm.

Myoma, leiomyoma, or uterine myoma is a medical condition in the form of abnormal tissue growth or tumors inside or around the uterus. Uterine myoma consists of muscle and fibrous tissue, Unlike malignant cancer, uterine myoma is a benign tumor. The size of a benign tumor that grows on an organ These uteruses generally vary, ranging from the size of a marble to a tennis ball[5]. Women's reproductive health has a huge impact and plays an important role in the continuity of the next generation, uterine myoma/mioma is one of the reproductive health problems that occurs in women continues to increase and is the second most common benign tumor in Indonesia, wrong the only one is in the Aceh Tamiang area. Myoma patients at the Aceh Tamiang Regional Hospital continue increases every year with an age range of around 20-50 years, the sufferers Myoma disease rarely causes specific symptoms so it is rare in women aware of the presence of myoma growth in their uterus.

For this reason, it is necessary to classify uterine myoma using the KNN method and can be done applied at the Aceh Tamiang Regional Hospital so that it can facilitate the resolution of existing problems and provide more accurate and effective information. Based on the background that has been described, then The proposed title is "Application of the K-Nearest Neighbor Method for Classification of Leiomyoma Disease(Myoma)".

Literature Review
K-Nearest Neighbor Algorithm

The K-Nearest Neighbor algorithm is one of the methods used for classification within an object. K-Nearest Neighbor works with looking for the shortest distance between data and other data, K-Nearest Neighbor too including a supervised learning algorithm, where the results of a new query instance are classified based on the majority of categories in K-Nearest Neighbor[6].

Which performs classification based on the proximity of the location of one data to other data is the K-Nearest Neighbor method. Purpose of the K-Nearest algorithm Neighbor is to classify new objects based on features and training samples[7]. There are several steps to calculate the K Nearest Neighbor algorithm as follows:

1. Determine the K parameter (number of closest neighbors)
2. Calculate the Eulidean Distance of each object to the data training provided.
3. Then sort the objects into groups has the smallest distance.
4. Collecting category Y (K-Nearest Neighbor classification based on value k).

The advantages of the K-Nearest Neighbor algorithm for classification are:

1. Have strong consistency. When the amount of data approaches infinity The K-Nearest Neighbor algorithm guarantees a better error rate than Bayes error rate.
2. Resilient to noisy test data.
3. Highly nonlinear

The disadvantages of the K-Nearest Neighbor algorithm are:

1. Need to determine the parameter K (number of nearest neighbors).
2. Quite high computational load due to the distance between test data and all data training must be counted

Euclidean Distance

Euclidean distance has a function to test possible measurements used as an interpretation of the closeness of the distance between two objects. Euclidean distance is also often used to calculate distance. Method used in this research is the Euclidean Distance approach method. Following
the formula:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (1)$$

Information:

d(x,y) = Distance
xi = Test Data
yi = Training Data
N = Data Dimension
I = Data Variable

Confusion Matrix

Confusion matrix is a testing stage to estimate objects true or false. The confusion matrix testing sequence is the predicted class will be displayed at the top of the matrix and the observed class on the left. Each cell has a number indicating how many cases there are the actual value of the observed class to predict.

Table 1. *Confusion Matrix*

Mark	Actual Value	
Prediksi	TP	FP
	FN	TN

Materials & Methods

The research steps carried out by the author were to develop This final project was made using the waterfall method. First step starting with a literature study to system evaluation. The waterfall method consists of from several stages including: Literature study, needs analysis, design system, system implementation, system testing, and system evaluation. can be seen as follows:

1. Literacy Studies
   Before conducting any research, it is necessary to conduct a literature study means to find the underlying theory. The theoretical basis can be found Obtain from various sources such as journals, articles, the internet, and various books.
2. Needs Analysis
   Requirements analysis functions to obtain requirements specifications generality of the system being built. In this research, a needs analysis was carried out through data collection. After getting general needs data, the data will be used at the next stage of system design.
3. System Design
   System design is carried out in several stages, namely design architecture, database design, and interface design.

Planning architecture involves creating context diagrams for database design, and interface design creating wireframes.

4. System Implementation
   At the implementation stage of this system, a design is made at stage previously implemented in the programming language that will be used.
5. System Testing
   This stage carries out system testing that has been carried out at stage system implementation. This stage is useful to help determine whether system implementation that meets the requirements.
6. System Evaluation
   After the system software is complete, system testing will be carried out that has been made. The purpose of testing is to find errors in system and then improve it in the hope that it is in accordance with the analysis done..

System Schema

The classification system scheme in the research entitled "Application of the K-Nearest Neighbor Method for Classification of Leiomyoma Disease (Myoma)" as follows:

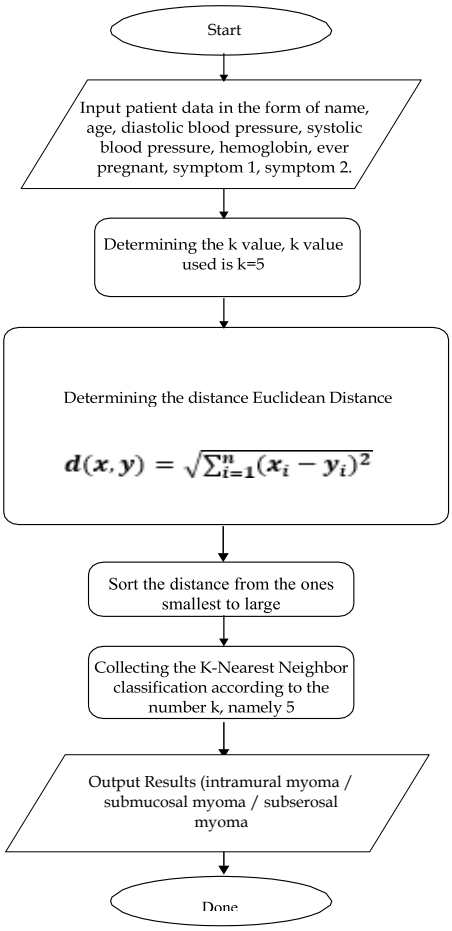

Figure 1. System Schema

Information:

1. The system process starts.
2. User inputs attributes, in the form of name, age, diastolic blood pressure, systolic blood pressure,hemoglobin, ever pregnant, symptom 1, symptom 2.
3. Then the user determines the k value, the k value used is k=5.
4. Then the user calculates the Euclidean distance.
5. Then the user sorts the Euclidean distance from the smallest to the smallest biggest
6. Next, the user collects the K-Nearest Neighbor classification according to the number k is 5.
7. After that, the results appear in the form of a diagnosis in the form of Intramural Mioma, Mioma Submucosal, Subserosal Mioma according to the symptoms suffered.
8. The system work process is complete.

Result and Discussion

The data used in this study is data on patients with myoma disease obtained from the Aceh Tamiang District Hospital for 5 years, from 2019-2023 with the classification of intramural myoma, submucosal myoma, and subserosal myoma. The following is the normalization table:

Table 2. Normalization Table

Normalization Table	
Classification	Weight
Intramural Myoma	1
Submucosal Myoma	2
Subserosal Myoma	3

The following is a table of criteria codes:

Table 3. Criteria Table

Criteria Code	Criteria
X1	Diastolic Pressure
X2	Systolic Pressure
X3	Hemoglobin
X4	Ever Pregnant
X5	Symptom 1
X6	Symptom 2

The following is a table of diastolic blood pressure criteria:

Table 4. Diastolic Blood Pressure Criteria

Diastolic Blood Pressure		
Attribute	Value	Weight
Diastolic Blood Pressure	100	1
	110	2
	120	3
	130	4
	140	5
	150	6

The following is a table of systolic blood pressure criteria:

Table 5. Systolic Blood Pressure Criteria

Systolic Blood Pressure		
Attribute	Value	Weight
Systolic Blood Pressure	70	1
	80	2
	90	3

The following is a table of hemoglobin criteria:

Table 6. Hemoglobin Criteria

Hemoglobin		
Attribute	Value	Weight
Hemoglobin	6	1
	7	2
	8	3
	9	4
	10	5
	11	6
	12	7

The following is a table of ever pregnant criteria:

Table 7. Ever Pregnant Criteria

Ever Pregnant		
Attribute	Value	Weight
Ever Pregnant	Yes	1
	No	2

The following is a table of symptom 1 criteria:

Table 8. Symptom 1 Criteria

Symptom 1		
Attribute	Value	Weight
Symptom 1	Bleeding	1
	Stomach Enlargement	2

The following is a table of symptom 2 criteria:

Table 9. Symptom 2 Criteria

Symptom 2		
Attribute	Value	Weight
Symptom 2	Pelvic Pain	1
	Stomach Pain	2
	Changes in menstrual patterns	3

From all the tables above, training data and testing data are obtained which have been weighted as follows:

Table 10. Training Data

No	Patient Name	X1	X2	X3	X4	X5	X6	Classification
1	Nurjanah	3	3	6	1	1	1	Intramural Myoma
2	Fatimah Hasan	4	2	7	1	1	1	Intramural Myoma
3	Rasiyah	1	2	3	1	1	1	Intramural Myoma
….	….	…	…	…	…	…	…	……..
200	Jubaedah	4	3	2	1	1	1	Intramural Myoma

Table 11. Data Testing

No	Patient Name	X1	X2	X3	X4	X5	X6
1	Imrana	1	2	7	1	2	2
2	Sari Ayu	3	3	6	1	1	3
3	Lili Yuliani	5	3	3	1	1	1
….	….	…	…	…	…	…	…
88	Iin Supreh	4	3	7	1	2	2

From the training data and testing data above, distance testing was carried out from the 1st testing data to the 88th testing data with all training data using the Euclidean distance formula. After getting the distance value, the data will be grouped based on the predetermined k, namely k=5 and the data will be sorted based on the smallest distance value for each distance according to the number of k. The following is a table of test results for the value k=5:

Table 12. Table of Test Results K=5

No	Name	Actual	Predictions
1	Imrana	Subserosal Myoma	Subserosal Myoma
2	Sari Ayu	Submucosal Myoma	Submucosal Myoma
3	Lili Yuliani	Intramural Myoma	Intramural Myoma
4	Siti Ruqoyah	Intramural Myoma	Intramural Myoma
5	Nur Dewi	Intramural Myoma	Intramural Myoma
….	……..	……..	……..
88	Iin Supreh	Subserosal Myoma	Subserosal Myoma

From the distance calculation results above, 5 data were obtained with different results. This indicates that calculating classification using the k-nearest neighbor method and using the Euclidean distance formula has a very low error value. The following confusion matrix performs tests to photograph the correct and incorrect objects:

Table 13. Confusion Matrix Testing

Actual	Prediction		
	Intramural Myoma	Submucosal Myoma	Subserosal Myoma
Intramural Myoma	45	0	0
Submucosal Myoma	0	15	3
Subserosal Myoma	3	0	22

Based on the confusion matrix above, the levels of precision, recall and accuracy are as follows:

Table 14. Table of Recall, Precision and Accuracy Values

Class	Recall	Precision	Accuracy

1	100%	94%	
2	83%	100%	93%
3	88%	88%	

The results are visualized with a bar graph in the following image:

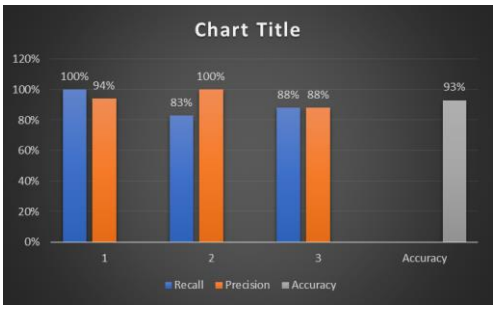

Figure 2. Bar Graph Result

System Implementasion

The following are the results of system implementation using the PHP programming language according to the system scheme that was previously designed:

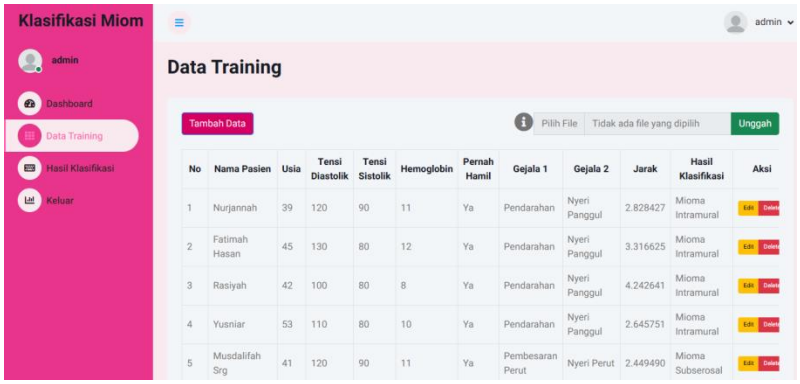

Figure 3. Training Data Menu

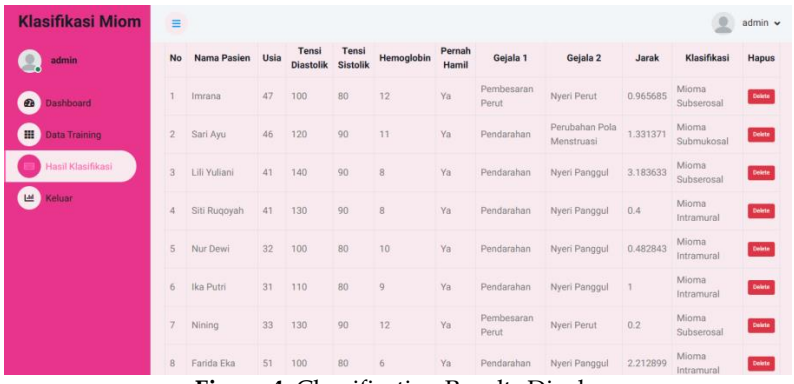

Figure 4. Classification Results Display

Figure 3 shows the training data menu which contains patient data that can be added, edited and deleted. Then in Figure 4 shows the classification results menu which can be deleted and managed by the admin.

Conclusions

From the results of the tests that have been carried out, it can be concluded that the data used was 288 patient data from the Aceh Tamiang District Hospital. There are 200 data used as training data and 88 data used for testing data and using 6 (six) criteria, namely diastolic blood pressure, systolic blood pressure, hemoglobin, ever pregnant, symptom 1, and symptom 2. Using the K-Nearest Neighbor algorithm , using the Euclidean Distance formula and Confusion Matrix calculations with a value of K=5, 88 data were tested to obtain classification output results in the form of 48 intramural myoma data, 15 submucosal myoma data and 25 subserosal myoma data in patient data with myoma at the Aceh Tamiang District Hospital. In its application, the myoma disease classification system which uses the K-Nearest

Neighbor algorithm and uses the Euclidean Distance formula produces a high accuracy value of 93%.

Acknowledgments

I would like to express my deepest gratitude to all parties for their support during the research and until now. I would like to express special thanks to my supervisors, Mr. Mukti Qamal and Mrs. Zahratul Fitri, for their invaluable guidance and input, as well as their continuous encouragement. Their dedication was critical in shaping this all. I am also very grateful to the collaborators for their technical assistance, which contributed to the success of this project. Finally, I would like to thank myself, family and friends for their support and encouragement all this time.

References

[1] A. Ariani, K. Kunci-Penyakit, and G. Kronis, "Klasifikasi Penyakit Ginjal Kronis menggunakan K-Nearest Neighbor," *Pros. Annu. Res. Semin.*, vol. 5, no. 1, pp. 148–151, 2019, [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Dise

[2] I. N. Atthalla, A. Jovandy, and H. Habibie, "Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 148–151, 2018.

[3] H. A. Dwi Fasnuari, H. Yuana, and M. T. Chulkamdi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus," *Antivirus J. Ilm. Tek. Inform.*, vol. 16, no. 2, pp. 133–142, 2022, doi: 10.35457/antivirus.v16i2.2445.

[4] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 104–111, 2020, doi: 10.33096/ilkom.v12i2.539.104-111.

[5] S. S. Yosi Apriyani, "ANALISA FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN MIOMA UTERI DI RSUD dr. ADHYATMA SEMARANG," *J. Kebidanan*, vol. 2, no. 5, pp. 36–46, 2019, [Online]. Available: http://ejournal.poltekkes-smg.ac.id/ojs/index.php/jurkeb/article/view/105

[6] R. Dwi Yulian Prakoso, B. Soejono Wiriaatmadja, and F. Wahyu Wibowo, "Sistem Klasifikasi Pada Penyakit Parkinson Dengan Menggunakan Metode K-Nearest Neighbor," *Semin. Nas. Teknol. Komput. Sains* , no. 2016, pp. 63–68, 2020, [Online]. Available: https://prosiding.seminar-id.com/index.php/sainteks

[7] M. Sadli, F. Fajriana, W. Fuadi, E. Ermatita, and I. Pahendra, "Penerapan Model K-Nearest Neighbors Dalam Klasifikasi Kebutuhan Daya Listrik Untuk Masing-Masing Daerah Di Kota Lhokseumawe," *J. ECOTIPE*, vol. 5, no. 2, pp. 11–18, 2018, doi: 10.33019/ecotipe.v5i2.646.