

Diagnosis of Preeclampsia in Pregnant Women Based on K-Nearest Neighbor Algorithm

Rifki Hidayat ^{1,*}, Tri Astuti ²

^{1,2} Program Studi Sistem Informasi, Universitas Amikom Purwokerto, Indonesia

¹ ripki09.id@gmail.com*; ² tri_astuti@amikompurwokerto.ac.id

* corresponding author

(Received: May 12, 2020; Revised: June 26, 2020; Accepted: August 3, 2020; Available online: December 1, 2020)

Abstract

Maternal deaths are divided into two namely direct and indirect deaths. Globally 80% of direct maternal deaths, preeclampsia are included in direct maternal deaths. Preeclampsia conditions of pregnancy with hypertension occur after the 20th week in women who previously had normal blood pressure. Preeclampsia can also be characterized by hypertension (systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg) accompanied by proteinuria (≥ 300 mg / dl in tamping urine 24 hours). In this study, an analysis of medical records in the Purbalingga and Banyumas areas using 8 attributes, namely age, body weight, blood pressure, edema, multiple pregnancy, history of hypertension, how many children, urine protein, and preeclampsia class. From calculations using the K-NN (K-Nearest Neighbor) algorithm, the Sensitivity performance value of 98.19%, Specificity 100%, and Accuracy 98.33%.

Keywords: Confusion Matrix; Data Mining; KNN Algorithm; Preeclampsia

1. Introduction

Health problems about pregnant women are things that must be considered. Because pregnant women take many lives. Maternal deaths are divided into direct deaths, namely due to complications of pregnancy, childbirth or the postpartum period and any intervention or inappropriate handling of these complications, and indirect deaths as a result of pre-existing diseases and or diseases that arise during pregnancy that affect pregnancy. Globally 80% of maternal deaths are classified as direct maternal deaths [6]. Which is where preeclampsia goes into indirect maternal death.

Preeclampsia is a pregnancy-specific condition with hypertension occurring after the 20th week in women who previously had normal blood pressure. Preeclampsia is a vasospastic disease, which involves many systems and is characterized by hemoconcentration, hypertension, and proteinuria. The diagnosis of preeclampsia has traditionally been based on the presence of hypertension with proteinuria or edema [3]. Preeclampsia is characterized by hypertension (systolic blood pressure 140 mmHg or diastolic blood pressure 90 mmHg) accompanied by proteinuria (≥ 300 mg/dl in 24-hour urine collection) at gestational age of more than 20 weeks or immediately after delivery. While preeclampsia is characterized by hypertension, proteinuria and seizures [8].

Data on preeclampsia in Banyumas Regency in 2011 were 551 people (32.1%) of all 1714 pregnant women. While the incidence of preeclampsia in 2012 was 930 people (50.9%) of all 1826 pregnant women. It was reported that the number of maternal deaths in 2013 was 33 cases, with 9 cases due to eclampsia according to [12].

Data mining is a series of processes to obtain useful information from large database warehouses [Han]. Data mining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses mathematical

statistical techniques, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases [13].

Various commercial and non-commercial tools are circulating and are used for various purposes related to data mining. One of them is WEKA, which is a tool in data mining, especially in the application to solve classification problems [2].

The research conducted is data mining analysis on medical record data in Banyumas and Purbalingga areas using the KNN (K-Nearest Neighbor) algorithm to diagnose preeclampsia. Why use the algorithm because KNN (K-Nearest Neighbor) is one of many popular text classification methods that provides accurate and easy-to-understand results [1]. In addition, the KNN (K-Nearest Neighbor) technique is a classification model that has several advantages, its application is simple but effective in many cases [9].

The purpose of this study was to find out the performance of medical record data in the Banyumas and Purbalingga regions as much as 120 data consisting of 11 data with preeclampsia and 109 data without preeclampsia. The data performance consists of three, namely sensitivity, specificity, and accuracy from the calculation of the confusion matrix. To get the value of the confusion matrix, the researchers used the WEKA software.

2. Literature Review

K-Nearest Neighbor is a algorithm for the construction of systems that rely on probabilistic knowledge. Teles et al. [7] propose the use of a context-aware platform based on K-Nearest Neighbor algorithm to support the experts' decision-making in public health systems. This study is focused on scenarios of dengue. Results show that the use of ontologies together with a K-Nearest Neighbor approach makes the prediction more refined. Bobba et al. [8] present a data based DSS that uses a K-Nearest Neighbor to merge gene expression data into prognostic models. This system integrates information from earlier experiments to predict the disease state. Future works proposes the implementation of this research on other pathologies for prevention and treatment. Huang et al. [9] use data mining techniques to extract rules and relationships between diseases using patient medical records. This study used an ensemble of classifiers like K-Nearest Neighbors and J-48 to try to improve the prediction performance of several diseases. Results showed a small improvement of the accuracy, sensitivity, and F-measure. Hannan et al. [10] describe an intelligent DSS built with artificial intelligence mechanisms. This system uses a K-Nearest Neighbor to identify the confidence level for each possible cause. The preliminary results show that this system is able to assist people in decision-making.

Kachroo et al. [11] compared three classifiers using machine-learning techniques to project cancer incidence and mortality. This research evaluates the performance of these classifiers, examining the accuracy of each method. The results show that K-Nearest Neighbor provides the best results. Researchers plan advances in developing nonlinear models for projection of future cancer occurrences. Qian et al. [12] developed a framework that uses the naive K-Nearest Neighbor classifier to measure the risk probability caused by lesions in the coronary artery. Results show a potential to reduce medical expenses avoiding unnecessary tests and treatment. In [13], the authors propose a computational model to classify stages of heart failure. They evaluate the best classifier like K-Nearest Neighbor, Support Vector Machine, and Radial Basis Function Network to offer a quantitative tool to facilitate the early diagnosis. Authors proposed that further works should analyze other intelligent algorithms to keep improving the best found classifier. In [14], the authors use K-Nearest Neighbor to assist the diagnosis of social anxiety disorder. This approach is modeled using conditional probability tables. Results show the model can be efficient for diagnosis of anxiety disorder.

Based on the related literature analysis, the next section will describe a mathematical model based on K-Nearest Neighbor that can assist decision makers in uncertain times to diagnose and evaluate the gravity of hypertension in pregnant women.

3. Research Model

The first step research method that must be carried out is to identify the problem and then the data collection method used is as follows:

- 1) Literature review

Literature studies are related to theoretical studies and other references related to values, culture and norms that develop in the social situation under study, besides that library research is very important in conducting research, this is because research cannot be separated from scientific literature according to [11].

- 2) Primary data

In this study, primary data is used which is the source of data obtained by researchers directly, namely through a literature study of the data that will be needed in the Purbalingga and Banyumas areas. The data is in the form of medical records of pregnant women which are then copied into an excel file. From the data collection process, 120 data were obtained. The data obtained consisted of age, weight <45kg, blood pressure $\geq 160/90$ mmHg, edema, urine protein, how many pregnancies, how many births, history of hypertension, preeclampsia.

In this study, the researcher will conduct a research concept in the form of a flow chart in this study. In the research, the research material used medical record data for pregnant women in the Purbalingga and Banyumas areas. This stage is carried out after the problem identification process and data collection, the next are the stages to obtain the output of the KNN (K-Nearest Neighbor) Algorithm method in diagnosing preeclampsia. The research concept will be explained in Figure 1 as follows:

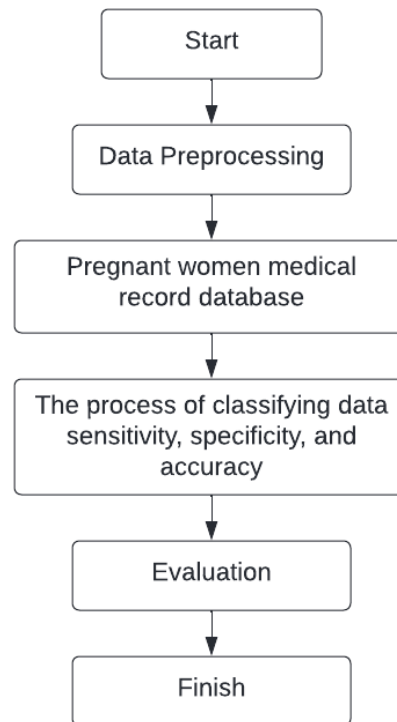


Figure. 1. Research Concept

3.1. Data Preprocessing

The initial research concept that will be carried out in this study is the preprocessing stage of data on the medical record dataset of pregnant women. This preprocessing stage will be used to generate rules based on the KNN (K-Nearest Neighbor) Algorithm, and then carry out data classification and evaluation of the results obtained from data testing.

Data preprocessing is a very important stage because at this stage the data selection process is carried out so that the data can still be controlled and not out of line. Thus the data will be ready to be used as research material [9]. In the preprocessing stage, the first thing is to prepare medical record data for pregnant women in the Purbalingga and Banyumas areas which consists of 120 data. The first step is the selection of the attributes used for calculations and changing the data format to make it easier to calculate the Weka software. After passing through the following stages, the medical record data for the Purbalingga and Banyumas areas has been processed and can enter the next stage.

3.2. Classification

Classification algorithms can determine predictions on the potential of the data [10]. The classification results will be compared with the actual dataset using the confusion matrix value between the results of the KNN (K-Nearest Neighbor) method using the Weka software. The confusion matrix values that will be used in this study are accuracy, specificity, and sensitivity.

Confusion Matrix is an evaluation of a data mining classification which is represented in a table [4]. Confusion matrix contains information on the comparison of the classified labels with the actual labels. Table 2 describes a confusion matrix with two labels, namely yes and no.

Table. 1. Confusion Matrix

Classification		Predicted class	
		Class : YES	Class : NO
Observed class	Class YES	True Positive (TP)	False Negative (FN)
	Class NO	False Positive (FP)	True Negative (TN)

The grouping of data in this confusion matrix is used to calculate the value of sensitivity, specificity, recall, precision, f-measure and accuracy can be calculated using the following equation [15]:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{Sensitivity} = \frac{TN}{TN+FP} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Description :

TP : Number of positive cases classified as positive

FP : Number of negative cases classified as positive

TN : Number of negative cases classified as negative

FN : Number of positive cases classified as negative

3.3. Evaluation

At this stage, a classification evaluation will be carried out based on testing on yes and no objects [4] as well as measuring the accuracy of the results obtained using the KNN (K-Nearest Neighbor) method. In this study, it will be assessed from the value obtained from the confusion matrix which will be generated as a parameter. At this stage, the accuracy, specificity, and sensitivity values will be evaluated which are used to determine the performance value of the test using the KNN (K-Nearest Neighbor) method.

4. Result and Discussion

4.1. Problem Identification

In this stage the authors identify the disease preeclampsia which is a pregnancy syndrome suffered by pregnant women. To find out the cause of preeclampsia, the author studied the literature related to the author's research, namely to diagnose preeclampsia. To carry out the diagnosis, the writer chose the K-NN (K-Nearest Neighbor) Algorithm that was suitable for the author's research, where the K-NN (K-Nearest Neighbor) Algorithm could be used to analyze the dataset.

4.2. Data Collection

In this research, the data that will be used is by taking data on pregnant women in the Purbalingga and Banyumas areas, with a total of 120 data and using 9 attributes as supporters. Where the attributes and their explanations are as follows:

- 1) Age: this input variable shows the age/age of the patient.
- 2) Mother's weight <45 (Maternal weight <45kg): this input variable shows the weight of pregnant women who are at risk of being affected is in the range of less than 45 kg.
- 3) BP 160/90 mmHg (Blood Pressure 160/90 mmHg): This input variable shows the patient's blood pressure that exceeds 160/90mmHg.
- 4) Significant edema: this input variable represents the term used to refer to swelling of soft tissues such as skin in pregnant women.
- 5) Twin pregnancy: this input variable shows the condition of pregnant women having more than one fetus.
- 6) History of hypertension: this input variable indicates whether the patient has a history of hypertension or high blood pressure
- 7) Number of children: this input variable shows how many children the patient is pregnant with.
- 8) PU (Urine Protein): This input variable shows that pregnant women who experience abnormal urine protein will have a positive value (T) and pregnant women who do not experience abnormal urine protein will have a negative value (F).
- 9) Class Variable Preeclampsia: this variable is the output consisting of two possibilities, namely yes and no. T means that one record variable from 8 input variables will result in an instance of the 9th variable (T) that the patient suffers from preeclampsia. F means that in one record of the 8 input variables will produce an instance of the 9th variable (F) that the patient does not suffer from preeclampsia. Class This variable is the purpose of the class to state the diagnosis of preeclampsia.

Table. 2. Medical Record Data for Pregnant Women in Purbalingga and Banyumas Regions.

No	Age	Weight <45 Kg	TD>=160/90 mmHg	Odema	Twin pregnancy	History of hypertension	Child order	Urine Protein	Preeclampsia
1	24	F	F	F	?	?	0	F	F
2	36	F	F	F	?	?	3	F	F
3	20	F	F	F	?	?	0	F	F
4	30	T	F	F	?	?	1	F	F
5	34	F	F	F	?	?	1	F	F
6	35	F	F	F	?	?	0	F	F
7	28	F	F	F	?	?	1	F	F
8	29	F	F	F	?	?	1	F	F
9	26	T	F	F	?	?	1	F	F
10	35	F	F	F	?	?	1	F	F
...
118	38	F	F	F	?	?	?	F	F

119	31	F	F	F	?	?	?	F	F
120	33	F	F	F	?	?	?	F	F

In table 2. is a table of medical record data for pregnant women with a total of 120 data and using 9 attributes before processing, the explanation for each attribute is the age attribute has the lowest value of 17 and the highest is 42, the attribute BB (weight) <45kg has category T and F, the attribute TD (Blood Pressure) 160/90 mmHg has categories T and F, the real Edema attribute has categories T and F, the attribute of twin pregnancy does not have a category, the attribute history of hypertension does not have a category, the attribute of which child has the lowest value 0 and the highest value is 4, the Urine Protein attribute has T and F, the preeclampsia attribute has the T and F instance categories. The preeclampsia attribute is the classification reference in this study.

Table. 3. Explanation of Pregnant Women's Medical Record Data

1st data	Information	120 th data	Information
24	Age	33	Age
F	Weight <45 Kg	F	Weight <45 Kg
F	TD \geq 160/90mmHg	F	TD \geq 160/90mmHg
F	Odema	F	Odema
?	Twin pregnancy	?	Twin pregnancy
?	History of hypertension	?	History of hypertension
0	Child order	?	Child order
F	Urine Protein	T	Urine Protein
F	Preeclampsia	T	Preeclampsia

Table 3 is an explanation of two records from the Maternity Medical Record dataset, namely data 1 and 120 along with a description of each value in the attribute and record based on the dataset. The 1st data contains age : 24, weight <45kg : F, BP 160/90mmHg : F, Edema : F, Twin Pregnancy : ?, History of Hypertension : ?, How many children : 0, Urine protein : F, Preeclampsia : F and the 120th data contains age : 33, weight <45kg : F, BP 160/90mmHg : F, Edema : F, Twin Pregnancy : ?, History of Hypertension : ?, Multiple children : 0, Urine protein : T, Preeclampsia : T. For the use of T is yes, F is no, and ? is empty data.

4.3. Preprocessing Result

At the preprocessing stage the data must prepare the data, namely by converting the medical record data of pregnant women from Microsoft Excel (XLSX) to CSV (Comma-Separated Values) by saving as from the data set which is still in xlsx format and converts it to csv. Samples of pregnant women's medical record data in CSV form can be seen in Figure 2 below.

Age, Weight <45 Kg, TD>=160/90 mmHg, Odema, Twin pregnancy, History of hypertension, Child order, Urine Protein, Preeclampsia

25, F, F, F, ?, ?, 2, F, F
23, F, F, F, ?, ?, 1, F, F
28, F, F, F, ?, ?, 2, F, F
30, T, F, F, ?, ?, 3, F, F,
22, T, F, F, ?, ?, 2, F, T

Figure. 2. Example of Medical Record Data in the form of CSV.

In Figure 2 above which displays medical record data in the form of CSV (Comma Separated Values) where this file is only in the form of attribute names and data from the table is not in tabular form but only in the form of writing attribute names and data separated by commas (,) and without space.

The explanation for the top part of Figure 2 which reads: AGE, BB, TD, edema, Twin Pregnancy, Hypertension, Child_Ke, Protein_Urine, Preeclampsia is the name of the table attribute. While at the bottom which reads: 24, F, F, F, ?, ?, 0, F, F and up to data 120 is the data content of each attribute. How to read using the first data 24, F, F, F, ?, ?, 0, F, F it reads age (24), weight (no), blood pressure (no), edema (no), twin pregnancy (?) , hypertension (?), child_to (0), protein_urine (no), preeclampsia (no).

After changing the format to csv, there needs to be more data formatting so that it can be processed in the WEKA software. Changing the data format can be done using Notepad++ software from a csv file to an arff file by saving as. Before changing the data format, there needs to be an additional script in the data file, firstly, @relation functions for file naming, secondly @attribute functions to explain the attributes used and also the declaration of the attribute data type used must match the data, for example numbers, the data type is numeric, third @data serves to declare the data as a whole according to the attributes used. In Figure 3 is a script that is in arff format in the Notepad++ software editor.

```
@Attribute Age
@Attribute Weight <45 Kg
@Attribute TD>=160/90 mmHg
@Attribute Odema
@Attribute Twin pregnancy
@Attribute History of hypertension
@Attribute Child order
@Attribute Urine Protein
@Attribute Preeclampsia

@Data
25, F, F, F, ?, ?, 2, F, F
23, F, F, F, ?, ?, 1, F, F
28, F, F, F, ?, ?, 2, F, F
.....
30, T, F, F, ?, ?, 3, F, F,
22, T, F, F, ?, ?, 2, F, T
```

Figure. 3. Sample of Medical Record Data in the form of ARFF.

In Figure 3, the medical record data is in the form of ARFF (Attribute-Relation File Format). In this arff form, there are 3 parts, namely @relation, @attribute and @data which have been described above. In the first part in Figure 3, namely @relation data, this is a file called data. The second part of @attribute numeric age to @attribute preeclampsia {T, F} is the name of each attribute and the data type used for each attribute is described. Figure 3 uses 2 types of data types, numeric data types and nominal specifications data types. Numeric data type is data in the form of numbers such as @attribute AGE numeric, and @attribute child_to numeric. The nominal specification data type is a data type whose data is based on specific data such as {T, F} which must be enclosed in curly braces and commas as data-specific separators without spaces. The implementation is like @attribute BB {T, F}, @attribute TD {T, F}, @attribute oedema {T, F}, @attribute Pregnancy_twins {T, F}, @attribute hypertension {T, F}, @attribute Protein_Urine {T, F}, and @attribute Preeclampsia {T, F} from this application the attribute has data specifications of T and F only. The third part of @data is the data for each attribute that has been described in the @attribute section. In writing, the separator uses a comma and no spaces. Applications such as: @data 24, F, F, F, ?, ?, 0, F, F to 33, F, F, F, ?, ?, ?, T, T the way to read the data is still the same as in the csv form which The difference is that there is a file name, namely @relation and the use of data types for each attribute is described.

4.4. Data Classification

The researcher tested the data on the WEKA software in accordance with the research using the K-NN (K-Nearest Neighbor) Algorithm. In this test, the TP (true positive), TN (true negative), FP (false positive), and FN (false negative) values will be obtained. From this test, confusion matrix values will be sought including the values of accuracy, sensitivity, and specificity resulting from testing using the K-NN (K-Nearest Neighbor) method.

How to test data on the WEKA Algorithm software, first enter the Classify tab because the K-NN (K-Nearest Neighbor) Algorithm is a classification group. The next step is to enter the classifier, click the choose button, select the lazy folder, enter in the folder there are 3 files, select IBk. Continue to test options, select cross-validation with folds 10. Next, make sure the classification attribute is a preeclampsia attribute because this attribute is a classification of this data, then start and the results will appear. So the test got the result, namely the value of the confusion matrix based on the K-NN (K-Nearest Neighbor) Algorithm. The results of the test are in Figure 4 below.

```
=== Confusion Matrix ===
      a    b   <-- classified as
      9    2 |   a = Y
      0 109 |   b = T
```

Figure. 4. Confusion Matrix results from WEKA software.

Figure 4.4 is the result of calculations from the Weka software using the K-NN (K-Nearest Neighbor) algorithm method. These results produce a confusion matrix on the TP (true positive) value of 109 based on column b in the second row, TN (true negative) of 9 based on column a in the first row, FP (false positive) of 0 based on column a in the second row, and FN (false negative) by 2 based on column b first row. To be clearer read in table 4.

Table. 4. The results of the confusion matrix K-NN method.

109 true positive	2 false negative
0 false positive	9 true negative

Table 4 above is the result of the confusion matrix from the Weka software using the K-NN (K-Nearest Neighbor) method. From table 4.3, the results of the confusion matrix include: TP (true positive) value of 109, TN (true

negative) of 9, FP (false positive) of 0, and FN (false negative) of 2. And the next step is to perform performance calculations such as looking for sensitivity, specificity, and accuracy values. Here is how it is calculated:

$$\begin{aligned}
 \text{a. Sensitivity} &= \frac{TP}{P} \times 100\% \\
 &= \frac{109}{111} \times 100\% \\
 &= 98,19 \% \\
 \text{b. Specificity} &= \frac{TN}{N} \times 100\% \\
 &= \frac{9}{9} \times 100\% \\
 &= 100\% \\
 \text{c. Accuracy} &= \frac{TP+TN}{P+N} \times 100\% \\
 &= \frac{109+9}{111+9} \times 100\% \\
 &= \frac{118}{120} \times 100\% = 98,33 \%
 \end{aligned}$$

From the results of the confusion matrix calculation above, it can be concluded that the K-NN (K-Nearest Neighbor) method has the following performance:

Table. 5. Performance results of the K-NN method.

Performance	Value
<i>Sensitivity</i>	98,19 %
<i>Specificity</i>	100 %
<i>Accuracy</i>	98,33%

In table 5 regarding the performance results obtained using the K-NN (K-Nearest Neighbor) method, it can be concluded as follows for the sensitivity value of 98.19%, the specificity value of 100%, and the accuracy value of 98.33%. These results were tested based on the medical records of pregnant women in the Purbalingga and Banyumas areas.

4.5. Evaluation

The results of the study entitled Diagnosis of Preeclampsia in Pregnant Women Based on the K-Nearest Neighbor Algorithm. In the research, what was done was processing the medical record data of pregnant women in the Purbalingga and Banyumas areas with a lot of 120 data. For processing the medical record data of pregnant women using 9 attributes, namely age, bb (weight), td (blood pressure), edema, twin pregnancy, hypertension, child_to, protein_urine, and preeclampsia. The attribute as a classification class is the last attribute, namely the preeclampsia attribute because the attribute already has results related to preeclampsia based on the existing attributes in the medical record data of pregnant women. Data processing uses data mining methods based on the KNN (K-Nearest Neighbor) algorithm to find performance values from medical record data.

In searching for the performance value of medical record data, you must look for the values of accuracy, specificity, and sensitivity based on the value of the confusion matrix whose calculations are carried out on the Weka software with the KNN (K-Nearest Neighbor) algorithm. The confusion matrix value from the calculation of the Weka software is the TP (true positive) value of 109, TN (true negative) of 9, FP (false positive) of 0, and FN (false negative) of 2. Seeing from the evaluation stage that the medical record data Pregnant women in the Purbalingga and Banyumas areas have data performance values, namely the accuracy value of 98.33%, specificity of 100%, sensitivity of 98.19%. Based on the performance value and referring to the accuracy value in the medical record data, pregnant

women in the Purbalingga and Banyumas areas can diagnose preeclampsia with the KNN (K-Nearest Neighbor) algorithm.

5. Conclusion

It can be concluded that the performance value of the accuracy, specificity, and sensitivity values of 120 data and 9 attributes. The performance results from the calculation of medical record data that have been processed are sensitivity of 98.19%, specificity of 100%, and accuracy of 98.33%. From these results it can be concluded that the KNN (K-Nearest Neighbor) algorithm can be used to diagnose preeclampsia with an accuracy value of 98.33%. Suggestions for further research are to add medical record data in order to get more accurate accuracy, and use another algorithm as an accuracy comparison material.

References

- [1] Y. Gotoh and C. Okubo, "A proposition of querying scheme with network Voronoi diagram in bichromatic reverse k-nearest neighbor," *Int. J. Pervasive Comput. Commun.*, vol. 13, no. 1, pp. 62–75, Jan. 2017, doi: 10.1108/IJPCC-01-2017-0009.
- [2] S. Yang and J. K. Kim, "Nearest Neighbor Imputation for General Parameter Estimation in Survey Sampling," in *The Econometrics of Complex Survey Data*, vol. 39, Emerald Publishing Limited, 2019, pp. 209–234.
- [3] A. Maity, P. Prakasam, and S. Bhargava, "Robust dual-tone multi-frequency tone detection using k-nearest neighbor classifier for a noisy environment," *Appl. Comput. Informatics*, vol. ahead-of-p, no. ahead-of-print, Jan. 2021, doi: 10.1108/ACI-10-2020-0105.
- [4] J. Ahamed, R. N. Mir, and M. A. Chishti, "Industry 4.0 oriented predictive analytics of cardiovascular diseases using machine learning, hyperparameter tuning and ensemble techniques," *Ind. Robot Int. J. Robot. Res. Appl.*, vol. ahead-of-print, no. ahead-of-print, Jan. 2022, doi: 10.1108/IR-10-2021-0240.
- [5] H. Khan, A. Srivastav, and A. K. Mishra, "Use of Classification Algorithms in Health Care," in *Big Data Analytics and Intelligence: A Perspective for Health Care*, P. Tanwar, V. Jain, C.-M. Liu, and V. Goyal, Eds. Emerald Publishing Limited, 2020, pp. 31–54.
- [6] J. Godara, R. Aron, and M. Shabaz, "Sentiment analysis and sarcasm detection from social networks to train health-care professionals," *World J. Eng.*, vol. 19, no. 1, pp. 124–133, Jan. 2022, doi: 10.1108/WJE-02-2021-0108.
- [7] S. Khodabandehlou, S. A. Hashemi Golpayegani, and M. Zivari Rahman, "An effective recommender system based on personality traits, demographics and behavior of customers in time context," *Data Technol. Appl.*, vol. 55, no. 1, pp. 149–174, Jan. 2021, doi: 10.1108/DTA-04-2020-0094.
- [8] Y. Gotoh and C. Okubo, "A proposition of querying scheme with network Voronoi diagram in bichromatic reverse k-nearest neighbor," *Int. J. Pervasive Comput. Commun.*, vol. 13, no. 1, pp. 62–75, Jan. 2017, doi: 10.1108/IJPCC-01-2017-0009.
- [9] Y. Deng, "Construction of higher education knowledge map in university libraries based on MOOC," *Electron. Libr.*, vol. 37, no. 5, pp. 811–829, Jan. 2019, doi: 10.1108/EL-01-2019-0003.
- [10] H. Kaur and V. Kumari, "Predictive modeling and analytics for diabetes using a machine learning approach," *Appl. Comput. Informatics*, vol. 18, no. 1/2, pp. 90–100, Jan. 2022, doi: 10.1016/j.aci.2018.12.004.
- [11] F. Abdi, K. Khalili-Damghani, and S. Abolmakarem, "Solving customer insurance coverage sales plan problem using a multi-stage data mining approach," *Kybernetes*, vol. 47, no. 1, pp. 2–19, Jan. 2018, doi: 10.1108/K-07-2017-0244.
- [12] E. Mohammed Abdelkader, "On the hybridization of pre-trained deep learning and differential evolution algorithms for semantic crack detection and recognition in ensemble of infrastructures," *Smart Sustain. Built Environ.*, vol. ahead-of-print, no. ahead-of-print, Jan. 2021, doi: 10.1108/SASBE-01-2021-0010.