khazanah informatika
Jurnal Ilmu Komputer dan Informatika

Implementing Bayes' Theorem Method in Expert System to Determine Infant Disease

Herman[1], Sunardi[2], Virasanty Muslimah[1*]
[1]Master Program of Informatics
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
[2]Department of Electrical Engineering
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
*virasanty2107048003@webmail.uad.ac.id

Abstract-The infant's immune system requires parental attention and is recommended to be checked regularly by health professionals. These diseases suffered by infants are Acute Respiratory Infections (ARI), Diarrhea, Acute Pharyngitis, Scabies, and Allergic Contact Dermatitis (ACD). Treatment can be provided at the public health center (Puskesmas), although there is still a general shortage of specialist doctors and no system to help diagnose diseases suffered by infants. Bayes' Theorem is a rule that uses probability to make the best decision based on available information. This study makes a diagnosis of the disease suffered by the baby with the aim that the disease can be treated early by using the Bayes' Theorem method. Based on the scenario that babies who experience symptoms of cold cough, itchy, and runny nose are then calculated using the Bayes' Theorem method, it is concluded that the baby is suffering from Scabies. Bayes' Theorem method, which was tested on 30 data, was found to have an accuracy value of 0.89 or 89%. The infant disease expert system using the Bayes' Theorem method makes it easier for parents to find out the disease in their baby so that they can take action on the symptoms that appear.

Keywords: disease, baby, expert system, bayes' theorem

Article info: *submitted January 13, 2023, revised September 15, 2023, accepted October 10, 2023*

1. Introduction

Children are very susceptible to diseases, which causes anxiety for parents. Infants aged 0-12 months are very susceptible to diseases that can affect the process of growth and development [1]. According to a survey by the Indonesian government, the diseases and disorders that affect infants still revolve around perinatal diseases, infectious diseases, and malnutrition [2].

The infant's immune system requires parental attention and it is recommended that the infant be checked regularly by the health care provider [3]. The most common diseases in infants are Acute Respiratory Infection (ARI), Diarrhea, Acute Pharyngitis, Scabies, and Allergic Contact Dermatitis (ACD) [2], [4]. Disease treatment can be done at the public health center (Puskesmas). Regulation of the Minister of Health of the Republic of Indonesia No. 43 of 2019 on Community Health Centers defines a Puskesmas as a health service facility that organizes public health efforts and first-level individual health efforts, with priority to promotive and preventive efforts in its working area [5]. The obstacle that is often encountered is that the Puskesmas still lacks specialist doctors and does not have a system to help diagnose diseases suffered by babies. An expert system is a system that combines human knowledge into a computer so that it can solve problems like experts do [6], [7]. A good expert system is designed to solve some problems by simulating someone's work. Expert systems allow the general public to solve complex problems that can only be solved with the help of an expert [8], [9]. This system is used to investigate medical diagnoses or as an educational tool [10].A comparative analysis of the methods used, namely Certainty Factor, Dempster Shafer, and Bayes' Theorem in the diagnosis of inflammatory disease of immune dermatitis in children by calculating the method. The results of the calculation are the certainty factor has the highest probability values of the other two methods [11]. The method used in designing an expert system for diagnosing acute respiratory tract infection in children is Bayes' Theorem which is implemented based on 30 data that have been tested with validation results obtained a percentage of 83,33% of appropriate case data and 16,67% of inappropriate case data [12].

The design of an expert system for diagnosing malnutrition in infants using the Bayes' Theorem method and implemented with the accuracy value of manual calculations and the expert system of this study is 99%, so it is declared successful [13]. The Certainty Factor method can predict more diseases with 80% accuracy compared to Bayes' Theorem with 60% accuracy in the analysis of expert systems for diagnosing diseases [14].

Expert system analysis for diagnosis of viral diseases uses the Bayes' Theorem method, which can be implemented to help diagnose skin diseases due to viral infection. The results of expert and system validation obtained a success rate of 85% of the 20 data that had been tested [15]because through the skin we can feel the feel and the surface of an object. If the skin is attacked, the disease may cause many consequences, such as from mild itching or to severe, the skin can not feel anything. Based on existing problems in the community created an application that can help people to recognize skin diseases suffered, so the level of awareness will keep the cleanliness higher, namely the application of expert system of viral skin disease using bayes theorem. In this research will be designed a system using method bayes theorem this expert system will be able to produce a diagnosis of skin diseases and treatment that needs to be done. With this application of course can help people to recognize the illnesses that have been allegedly allegedly to refer to injured. Based on 20 data that have been tested against experts and systems, for patients affected by viral skin diseases were 17 patients and those not infected were 3 patients. As for the suitability of testing between the validation results of experts (doctors.

Researchers [11]–[15] have in common with this research the use of Bayes' Theorem method through the diagnosis of various diseases. Bayes' Theorem has been used to diagnose a variety of medical conditions, including stomach disorders [16] and skin disorders [17].

The main purpose of this research is to provide diagnostic results by the calculation of the Bayes' Theorem method. The results of calculations using Bayes' Theorem method can be used as a service for diagnosing diseases in infants based on their symptoms

2. Methods

This section discusses the method used, the research framework which contains steps in conducting research, and a flowchart of the Bayes' Theorem method in diagnosing infant diseases.

a. Bayes' Theorem Method

Bayes' Theorem is a method based on data training, based on conditional probability [18], [19]. Bayes' Theorem is a rule that uses probabilities to make the best decisions based on the available information [20], [21]so children have a good nutrition status. Nutrition status monitoring for toddlers could be done with Anthropometry calculations, based on 3 index, weight for age (WFA. Using this method has the main advantage of simplifying classical methods that can be difficult to understand [22].

The name Bayes' Theorem comes from its discoverer, the Reverend Thomas Bayes from 1701 to 1761, an English statistician, philosopher, and presbyterian clergyman. Bayes' Theorem published in 1763 completes the conditional probability theorem [23]–[25]. Bayes' Theorem is a mathematical formula that helps calculate the probability value of the evidence given to generate a hypothesis value [26]–[28]. The Bayes is shown in Equation 1 [16], [29].

$$P(H_k|E) = \frac{P(E|H_k)P(H_k)}{\sum_{k=1}^{n} P(E|H_k)P(H_k)} \quad (1)$$

where:
$P(H_k|E)$ is the probability of the hypothesis H_k if given evidence E.
$P(E|H_k)$ is the probability of the emergence of evidence E if it is known that the H_k hypothesis is true.
$P(H_k)$ is the probability of the H_k hypothesis, regardless of any evidence.
n is number of possible hypotheses.
from the Bayes' Theorem, it can be developed if testing the hypothesis appears more than one evidence, then the equation becomes Equation 2:

$$P(H|E,e) = \frac{P(H|E)P(e|E,H)}{P(e|E)} \quad (2)$$

where:
e is old evidence.
E is new evidence.
P(H|E,e) is the probability that hypothesis H is true if given new evidence E from old evidence e.
P(e|E,H) is probability of the relationship between e and E if hypothesis H is true
P(e|E) is the probability of the relationship between e and E regardless of any hypothesis
P(E|H) is the probability that the evidence E appears if the hypothesis is known

b. Research Framework

To achieve research objective, activities of this research are organized as shown in figure 1:

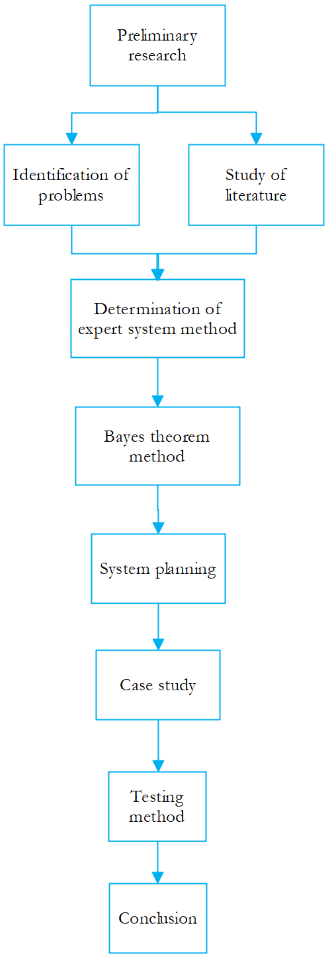

Figure 1. Research framework

This research preceded a preliminary study to match the research problem with the proposed solutions. Identification of problems was conducted in the real world regarding the limitation of existing health infrastructure in underdeveloped regions. An expert system that can assist general health practitioners or doctors in-house in diagnosing infant diseases is highly required. The study of literature involves the investigation of similar works in previous research and the study of the latest technology and techniques in the expert system and artificial intelligence that potential to be implemented to solve the problems.

Determination of the method to be used in determining the infant disease is a crucial step in this research. A review of the literature found there are several methods have been implemented by previous research from the era of expert system to sophisticated machine learning system. This research adopts Bayes' Theorem as an alternative method. This research designs the system to assist health practitioners to diagnose and determine infant disease using a case study that involves 120 data. The data are real data that were collected from Puskesmas Sorong between September 2021 to December 2022. A model of the system was built using Bayes' Theorem by training 90 data out of the 120 data. The rest of the 30 data is used for data testing. Results from the testing were analyzed and concluded at the end of the research framework.

c. Algorithm For Determining Infant Disease Using Bayes' Theorem Method

A model of the system to assist health practitioner to diagnose and determine infant disease is designed as shown in Figure 2.

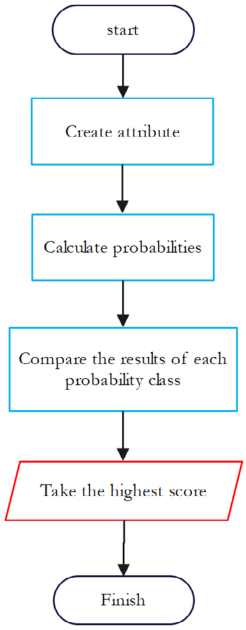

Figure 2. Flowchart

Figure 2 shows that the model was built using Bayes' Theorem. It starts by determining attributes of the data that will be trained. Attributes represent symptoms of the disease. Each data can comprise several attributes of symptoms and one label disease. The label of disease is given by an expert – a pediatrician. All attributes and labels of the data are transformed into binary. These data are fed to the calculations of probabilities in Bayes' Theorem. Firstly, it is the calculation of prior and likelihood of disease.

The prior is determined before new data is included. The likelihood is determined to show the extent to which the data support the given hypothesis. The calculations to determine the prior value as shown in Equation 3.

$$P(Hk) = \frac{number\ of\ hypotheses}{sum\ of\ training\ sets} \quad (3)$$

The calculation to get the likelihood is as shown in Equation 4.

$$P(E|Hk) = \frac{the\ amount\ of\ data}{sum\ of\ hypoth} \quad (4)$$

Calculations to obtain the normalization factor values are performed after obtaining the prior and likelihood values. The calculation to obtain the normalizing factor values is shown in Equation 5.

$$Normalization\ Factor = \sum[P(E|Hk)P(Hk)] \quad (5)$$

Once the value of the normalization factor has been obtained, the calculation is performed to obtain the posterior value. Mathematically, the posterior can be calculated using Bayes' Theorem with equation 1 explained earlier. Thirdly, the results obtained from the calculation of equation 1 are tested. Finally, the highest test result is concluded as the result of the diagnosis of infant disease.

3. Results and Discussion

a. Data Acquisition

The data was obtained from interviews with pediatricians at Puskesmas Sorong. Data acquisition refers to the process of collecting the data necessary to perform Bayes' Theorem modeling. The acquisition data includes information on five diseases that can affect infants. It contains the patient's ID, the age of the baby when suffering from the disease, the name of the disease, and the symptoms that appear. Training data is shown in Table 1.

Table 1. Acquired Data

Patient id	Age (Month)	Attribute							Disease code
		SOD01	SOD02	SOD03	...	SOD28	SOD29	SOD30	
8	1	yes	no	no	...	yes	no	no	BS01
67	1	no	no	no	...	yes	no	no	BS05
115	1	no	no	no	...	no	yes	no	BS04
120	1	no	no	no	...	no	no	yes	BS04
83	1	no	no	no	...	yes	no	no	BS02
...	...	...	...	...	...	...	...	...	...
36	53	no	no	no	...	yes	no	no	BS03
55	54	no	no	no	...	no	yes	no	BS04
60	56	no	no	no	...	yes	no	no	BS04

Table 1 displays information about several patients, their ages in months, unique attributes (symptoms that the patient is experiencing), and the disease codes associated with them. The patient id distinguishes each patient from one another. Age (month) represents the patient's age in months. Attributes (SOD01-SOD30) are characteristics related to the patient, such as the patient's experienced symptoms. The disease code corresponds to the disease associated with the patient.

b. Knowledge Base

This knowledge base was obtained from 120 infant disease data, concluding that there were five disease types and also found 30 disease symptoms. Disease types are shown in Table 2.

Table 2. Types of diseases

Code	Types of diseases	Description
BS01	Acute Respiratory Infection (ARI)	ARI is caused by a viral or bacterial infection in the respiratory tract. ARI is most often caused by a viral infection and most often occurs in the upper respiratory
BS02	Diarrhea	Diarrhea is a condition when the sufferer defecates more often than usual and can be said to be diarrhea if he defecates more than three or more times in one day. In addition, the feces that are excreted are also more watery
BS03	Acute Pharyngitis	Pharyngitis is inflammation of the mucous membrane that lines the back of the throat or pharynx. This inflammation can caouse discomfort, dryness, and difficulty swallowing. Pharyngitis can be caused by infection or no-infection

Code	Types of diseases	Description
BS04	Scabies	Scabies or commonly called sccabies is a contagious skin disease caused by the entry of tiny mites into the outer layer of the skin
BS05	Allergic Contact Dermatitis (ACD)	ACD is an inflammatory skin disease caused by allergic factors

Table 3. Disease symptoms

Code	Symptom name
SOD01	Cold cough
SOD02	Dehydration
SOD03	Nasal congestion
SOD04	Cough
SOD05	Sore throat
SOD06	Phlegm cough
SOD07	Dry cough
SOD08	Stridor sound
SOD09	Sneezing
SOD10	Flu
SOD11	Watery stools that are more than four times in one day
SOD12	Thick snot
SOD13	Bitter mouth
SOD14	Shiver
SOD15	Nauseous vomit
SOD16	There are bumps or rashes in the area between the fingers, around the navel, anus, stomach, and folds
SOD17	Blast around the folds
SOD18	Boil
SOD19	Purulent wound
SOD20	Sore neck
SOD21	Loose feces
SOD22	Watery stools
SOD23	Shortness of breath
SOD24	Itchy
SOD25	Bloated
SOD26	Water lute
SOD27	Itching to sores
SOD28	Fever
SOD29	Rash
SOD30	Runny nose

Table 3 describes the symptoms obtained from experts on Acute Respiratory Infection (ARI), Diarrhea, Acute Pharyngitis, Scabies, and Allergic Contact Dermatitis (ACD).

From the data in tables 2 and 3, it can be concluded that are be mentioned in Rule 1 - 5 [30].

Rule 1: IF cold cough (SOD01) is true
AND fever (S0D28) is true
AND watery stools (SOD22) are true
AND nauseous vomit (SOD15) is true
AND nasal congestion (SOD03) is true
AND cough (SOD04) is true
AND phlegm cough (SOD06) is true
AND sneezing (SOD09) is true

THEN Acute Respiratory Infection (BS01)

Rule 2: IF watery stools more than four times a day (SOD11) is true
AND loose feces (SOD21) are true
AND watery stools (SOD22) are true
AND dehydration (SOD02) is true
AND runny nose (SOD30) is true
AND nauseous vomit (SOD15) is true
AND fever (SOD28) is true
AND bloated (SOD25) is true
THEN Diarrhea (BS02)
Rule 3: IF cold cough (SOD01) is true
AND fever (SOD28) is true
AND nauseous vomit (SOD15) is true
AND dry cough (SOD07) is true
AND stridor sound (SOD08) is true
AND sore throat (SOD05) is true
AND flu (SOD10) is true
AND cough (SOD04) is true
AND sore neck (SOD20) is true
AND mucus cough (SOD06) is true
AND thick snot (SOD12) is true
AND bitter mouth (SOD13) is true
AND shiver (SOD14) is true
THEN Acute Pharyngitis (BS03)
Rule 4: IF itchy (SOD24) is true
AND purulent wound (SOD19) is true
AND water lute (SOD26) is true
AND fever (SOD28) is true
AND There are bumps or rashes in the area between the fingers, around the navel, anus, stomach, and folds (SOD16) is true
AND flu (SOD10) is true
THEN Scabies (BS04)
Rule 5: IF itchy (SOD24) is true
AND rash (SOD29) is true
AND itching to sores (SOD27) is true
AND fever (SOD28) is true
AND blast around to folds (SOD17) is true
AND boil (SOD18) is true
AND purulent wound (SOD19) is true
THEN Allergic Contact Dermatitis (BS05)

c. Training

The training dataset of 120 cases of infant diseases was successfully obtained from the Puskesmas Sorong between September 2021 to December 2022. The dataset is computed using Bayes' Theorem by first determining prior and likelihood. Prior refers to the initial knowledge or belief about a hypothesis before new data is incorporated. A prior can be based on knowledge, experience, or assumptions. The prior is very important in Bayes' Theorem. It allows relevant prior information to be considered when updating beliefs about hypotheses based on new data.

For example, to find the prior value for disease Scabies (BS04), the prior calculation uses Equation 3 based on the information in Table 1 is.

$$P(Hk) = \frac{23}{120} = 0.19$$

The prior value for the five diseases discussed in this study are as in Table 4.

Table 4. Prior

Disease code	Prior P(H)
BS01	0.26
BS02	0.18
BS03	0.17
BS04	0.19
BS05	0.21

The process continues to obtain the likelihood value using Equation 4. Likelihood refers to measuring how well the observed data fits a given hypothesis. Likelihood is a probability distribution describing the data's likelihood of being true.

For example, to find the likelihood values for cold cough symptoms due to scabies, use Equation 4 by referring to the data in Table 1.

Table 5 shows the probability of disease, which is called likelihood in Bayes' Theorem, for five diseases out of 30 symptoms.

Table 5. Disease probability (Likelihood)

No	Symptom code	Diseases code				
		BS01	BS02	BS03	BS04	BS05
1	SOD01	0.77	0.19	0.65	0.22	0.04
2	SOD02	0	0.24	0	0	0
3	SOD03	0.10	0	0.05	0	0
4	SOD04	0.03	0.10	0.15	0	0.04
5	SOD05	0.03	0	0.25	0.04	0
6	SOD06	0.19	0.05	0.05	0	0
7	SOD07	0.03	0	0.10	0	0
8	SOD08	0.03	0	0.15	0	0
9	SOD09	0.06	0	0.05	0	0

No	Symptom code	Diseases code				
		BS01	BS02	BS03	BS04	BS05
10	SOD10	0.10	0.05	0.05	0.13	0
11	SOD11	0.03	1.00	0	0	0
12	SOD12	0.03	0	0.05	0	0
13	SOD13	0	0	0.05	0	0
14	SOD14	0	0	0.10	0	0
15	SOD15	0.16	0.43	0.20	0.04	0
16	SOD16	0	0	0	0.35	0
17	SOD17	0	0	0	0	0.12
18	SOD18	0	0.05	0.10	0.04	0.12
19	SOD19	0	0	0	0.13	0.32
20	SOD20	0	0	0.30	0	0
21	SOD21	0	0.43	0	0	0
22	SOD22	0	0.48	0	0	0
23	SOD23	0.29	0	0	0	0
24	SOD24	0	0	0	0.74	0.68
25	SOD25	0	0.05	0	0	0
26	SOD26	0	0	0	0.35	0
27	SOD27	0	0	0.05	0.09	0.40
28	SOD28	0.77	0.62	0.80	0.48	0.64
29	SOD29	0	0	0	0.04	0.24
30	SOD30	0.03	0.14	0	0.17	0.08

The evidence or probability that the observed data is independent of the hypothesis is also known as the normalization factor. The process to obtain the normalization factor using Equation 5 is performed after the process to obtain the prior and the likelihood.

For example, to find the value of the normalization factor for cold cough symptoms is.

$$\sum P(E|H_k)P(H_k) = (0.77x0.26) + (0.19x0.18) + (0.65x0.17) + (0.22x0.19) + (0.04x0.21)$$

$$\sum P(E|H_k)P(H_k) = 0.20 + 0.03 + 0.11 + 0.04 + 0.01 = 0.39$$

The table displaying the normalization factor can be found in Table 6.

Table 6. Normalization factor

No	Symptom code	P(E|H)P(H)					Sum
		BS01	BS02	BS03	BS04	BS05	
1	SOD01	0.20	0.03	0.11	0.04	0.01	0.39
2	SOD02	0.00	0.04	0.00	0.00	0.00	0.04
3	SOD03	0.03	0.00	0.01	0.00	0.00	0.03
4	SOD04	0.01	0.02	0.03	0.00	0.01	0.06
5	SOD05	0.01	0.00	0.04	0.01	0.00	0.06
6	SOD06	0.05	0.01	0.01	0.00	0.00	0.07
7	SOD07	0.01	0.00	0.02	0.00	0.00	0.03
8	SOD08	0.01	0.00	0.03	0.00	0.00	0.03
9	SOD09	0.02	0.00	0.01	0.00	0.00	0.03
10	SOD10	0.03	0.01	0.01	0.03	0.00	0.07
11	SOD11	0.01	0.18	0.00	0.00	0.00	0.18
12	SOD12	0.01	0.00	0.01	0.00	0.00	0.02
13	SOD13	0.00	0.00	0.01	0.00	0.00	0.01
14	SOD14	0.00	0.00	0.02	0.00	0.00	0.02
15	SOD15	0.04	0.08	0.03	0.01	0.00	0.16

No	Symptom code	P(E|H)P(H)					Sum
		BS01	BS02	BS03	BS04	BS05	
16	SOD16	0.00	0.00	0.00	0.07	0.00	0.07
17	SOD17	0.00	0.00	0.00	0.00	0.03	0.03
18	SOD18	0.00	0.01	0.02	0.01	0.03	0.06
19	SOD19	0.00	0.00	0.00	0.03	0.07	0.09
20	SOD20	0.00	0.00	0.05	0.00	0.00	0.05
21	SOD21	0.00	0.08	0.00	0.00	0.00	0.08
22	SOD22	0.00	0.08	0.00	0.00	0.00	0.08
23	SOD23	0.08	0.00	0.00	0.00	0.00	0.08
24	SOD24	0.00	0.00	0.00	0.14	0.14	0.28
25	SOD25	0.00	0.01	0.00	0.00	0.00	0.01
26	SOD26	0.00	0.00	0.00	0.07	0.00	0.07
27	SOD27	0.00	0.00	0.01	0.02	0.08	0.11
28	SOD28	0.20	0.11	0.13	0.09	0.13	0.67
29	SOD29	0.00	0.00	0.00	0.01	0.05	0.06
30	SOD30	0.01	0.03	0.00	0.03	0.02	0.08

The posterior represents the confidence in a hypothesis following its updating with new data. It arises from the combination of the prior and likelihood. Mathematically, the posterior can be calculated using Bayes' Theorem with Equation 1.

For example, to find the posterior value of cold cough symptoms of scabies is.

$$P(H_k|E) = \frac{0.04}{0.39} = 0.11$$

The table displaying the posterior can be found in Table 7.

Table 7. Posterior

No	Symptom code	P(H|E)				
		BS01	BS02	BS03	BS04	BS05
1	SOD01	0.51	0.09	0.28	0.11	0.02
2	SOD02	0.00	1.00	0.00	0.00	0.00
3	SOD03	0.75	0.00	0.25	0.00	0.00
4	SOD04	0.14	0.29	0.43	0.00	0.14
5	SOD05	0.14	0.00	0.71	0.14	0.00
6	SOD06	0.75	0.13	0.13	0.00	0.00
7	SOD07	0.33	0.00	0.67	0.00	0.00
8	SOD08	0.25	0.00	0.75	0.00	0.00
9	SOD09	0.67	0.00	0.33	0.00	0.00
10	SOD10	0.38	0.13	0.13	0.38	0.00
11	SOD11	0.05	0.95	0.00	0.00	0.00
12	SOD12	0.50	0.00	0.50	0.00	0.00
13	SOD13	0.00	0.00	1.00	0.00	0.00
14	SOD14	0.00	0.00	1.00	0.00	0.00
15	SOD15	0.26	0.47	0.21	0.05	0.00
16	SOD16	0.00	0.00	0.00	1.00	0.00
17	SOD17	0.00	0.00	0.00	0.00	1.00
18	SOD18	0.00	0.14	0.29	0.14	0.43
19	SOD19	0.00	0.00	0.00	0.27	0.73
20	SOD20	0.00	0.00	1.00	0.00	0.00
21	SOD21	0.00	1.00	0.00	0.00	0.00
22	SOD22	0.00	1.00	0.00	0.00	0.00
23	SOD23	1.00	0.00	0.00	0.00	0.00
24	SOD24	0.00	0.00	0.00	0.50	0.50
25	SOD25	0.00	1.00	0.00	0.00	0.00
26	SOD26	0.00	0.00	0.00	1.00	0.00
27	SOD27	0.00	0.00	0.08	0.15	0.77
28	SOD28	0.30	0.16	0.20	0.14	0.20
29	SOD29	0.00	0.00	0.00	0.14	0.86
30	SOD30	0.10	0.30	0.00	0.40	0.20

d. Testing

The testing process in this study is performed by determining the patient's disease, which becomes the test data, using the Bayes' Theorem model. This model was previously built using training data. The type of disease determined by the calculation of Bayes' Theorem model is then compared with the type of disease determined by the physician, which is referred to as the label for the test data.

For example, for the test data of patient ID 119, who has a cold cough (SOD01), itchy (SOD24), and runny nose (SOD30), the doctor has labeled this patient with scabies. Implementing the Bayes' Theorem model and referring to the posterior value in Table 7, the results can be outlined as shown in Table 8.

Table 8. Probability on patient ID 119

No	Symptom code	P(H|E)				
		BS01	BS02	BS03	BS04	BS05
1	SOD01	0.51	0.09	0.28	0.11	0.02
2	SOD24	0.00	0.00	0.00	0.50	0.50
3	SOD30	0.10	0.30	0.00	0.40	0.20
	Sum	0.61	0.39	0.28	1.01	0.71

For symptom cold cough (SOD01), the patient's probability of Acute Respiratory Infection (BS01) is 0.51, while the likelihood of suffering from Diarrhea (BS02) is 0.09. The chances of Acute Pharyngitis (BS03) is 0.28, and Scabies (BS04) is 0.11. The probability of Allergic Contact Dermatitis (BS05) is also 0.02.

As for symptom itchy (SOD24), the patient's probability of having Acute Respiratory Infection, Diarrhea, Acute Pharyngitis, Scabies, and Allergic Contact Dermatitis is 0.00, 0.00, 0.00, 0.50, and 0.50 respectively. Similarly, for symptom runny nose (SOD30), the probability is 0.10, 0.30, 0.00, 0.40, and 0.20.

If we add up the probability values for the three symptoms, the probability of the patient having Acute Respiratory Infection is 0.61. The probability of the patient having Diarrhea is 0.39, while the probability of having Acute Pharyngitis is 0.28. The probability of the patient having Scabies is 1.01, and the probability of the patient having Allergic Contact Dermatitis is 0.71. Thus, according to the Bayes' Theorem model calculation, patient ID 119 is diagnosed with Scabies due to its highest probability value, which is 1.01.

Therefore, the model is considered accurate in predicting infant diseases. To evaluate the accuracy of the Bayes' Theorem method in diagnosing infant diseases, tests were performed on 30 patients. The test is shown in Table 9.

Table 9. Testing

Patient id	Symptom	Expert	Bayes' Theorem	Accuracy
119	SOD01, SOD24, SOD30	BS04	BS04	Accurate
93	SOD10, SOD11, SOD15	BS02	BS02	Accurate
90	SOD01, SOD11, SOD21, SOD22, SOD28	BS02	BS02	Accurate
100	SOD26, SOD28	BS04	BS04	Accurate
31	SOD01, SOD08, SOD23	BS01	BS01	Accurate
9	SOD01, SOD11, SOD15, SOD28	BS01	BS02	Inaccurate
73	SOD01, SOD24, SOD28	BS05	BS01	Inaccurate
74	SOD01, SOD29	BS05	BS05	Accurate
78	SOD11, SOD18, SOD21, SOD22, SOD30	BS02	BS02	Accurate
58	SOD28, SOD29	BS05	BS05	Accurate
40	SOD01, SOD06, SOD20	BS03	BS03	Accurate

Patient id	Symptom	Expert	Bayes' Theorem	Accuracy
96	SOD01, SOD11, SOD15, SOD28	BS02	BS02	Accurate
79	SOD02, SOD11, SOD15, SOD28	BS02	BS02	Accurate
51	SOD04, SOD05, SOD15	BS03	BS03	Accurate
99	SOD19, SOD24, SOD28	BS04	BS05	Inaccurate
32	SOD01, SOD15, SOD28	BS03	BS01	Inaccurate
2	SOD01, SOD28	BS01	BS01	Accurate
117	SOD16, SOD26	BS04	BS04	Accurate
44	SOD01, SOD12, SOD28	BS03	BS01	Inaccurate
39	SOD01, SOD18, SOD28	BS03	BS01	Inaccurate
111	SOD01, SOD16, SOD18, SOD24, SOD28	BS04	BS05	Inaccurate
95	SOD06, SOD11, SOD21, SOD22	BS02	BS02	Accurate
102	SOD10, SOD24, SOD26	BS04	BS04	Accurate
67	SOD19, SOD24, SOD26	BS05	BS05	Accurate
37	SOD01, SOD18, SOD28	BS03	BS01	Inaccurate
94	SOD04, SOD11, SOD30	BS02	BS02	Accurate
21	SOD01, SOD28	BS01	BS01	Accurate
86	SOD11, SOD21, SOD22, SOD28	BS02	BS02	Accurate
101	SOD26, SOD28	BS04	BS04	Accurate

Patient id	Symptom	Expert	Bayes' Theorem	Accuracy
28	SOD07, SOD23	BS01	BS01	Accurate

e. Confusion Matrix

The test results in Table 9 were tested using a confusion matrix. The confusion matrix is a performance evaluation tool commonly used in machine learning to describe the performance of classification models [31]. Confusion matrix testing is performed with a 5x5 matrix because there are five diseases in this study. The 5x5 confusion matrix is shown in Table 10.

Table 10. Confusion matrix

		True Label				
		BS01	**BS02**	**BS03**	**BS04**	**BS05**
Predicted Label	**BS01**	4	0	4	0	1
	BS02	0	0	2	0	0
	BS03	0	0	0	2	3
	BS04	1	8	0	0	0
	BS05	0	0	0	5	0

Confusion Matrix consists of four main elements, namely True Positive (TP) is the amount of positive data correctly predicted by the model; True Negative (TN) is the amount of negative data correctly predicted by the model; False Positive (FP) is the amount of negative data incorrectly predicted as positive by the model; False Negative (FN) is the amount of positive data incorrectly predicted as negative by the model [32].

Evaluation matrix calculations with a 5x5 confusion matrix such as accuracy, precision, recall, and f1-score are performed separately for each label. Accuracy is the proportion of total correct predictions from all data as shown in Equation 6 - 9.

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

Precision is a measure of the accuracy of model's results.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Recall is a measure of the completeness of model.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score is combined measure of precision and recall, describing their balance.

$$F1-Score = 2 * \frac{precision*recall}{precision+recall} \quad (9)$$

The scoring matrix for each label is calculated as follows.

1) Acute Respiratory Infection (BS01)

$$Accuracy = \frac{4+20}{4+20+5+1} = 0.80$$

$$Precision = \frac{4}{4+4} = 0.44$$

$$Recall = \frac{4}{4+1} = 0.80$$

$$F1-Score = 2 * \frac{0.44 * 0.80}{0.44 + 0.80} = 0.57$$

2) Diarrhea (BS02)

$$Accuracy = \frac{8+21}{8+21+1+0} = 0.97$$

$$Precision = \frac{8}{8+1} = 0.89$$

$$Recall = \frac{8}{8+0} = 1.00$$

$$F1-Score = 2 * \frac{0.89 * 1.00}{0.89 + 1.00} = 0.94$$

3) Acute Pharyngitis (BS03)

$$Accuracy = \frac{2+24}{2+24+0+4} = 0.87$$

$$Precision = \frac{2}{2+0} = 1.00$$

$$Recall = \frac{2}{2+4} = 0.33$$

$$F1-Score = 2 * \frac{1.00 * 0.33}{1.00 + 0.33} = 0.50$$

4) Scabies (BS04)

$$Accuracy = \frac{5+23}{5+23+0+2} = 0.93$$

$$Precision = \frac{5}{5+0} = 1.00$$

$$Recall = \frac{5}{5+2} = 0.71$$

$$F1-Score = 2 * \frac{1.00 * 0.71}{1.00 + 0.71} = 0.83$$

5) Allergic Contact Dermatitis (BS04)

$$Accuracy = \frac{3 + 24}{3 + 24 + 2 + 1} = 0.90$$

$$Precision = \frac{3}{3 + 2} = 0.60$$

$$Recall = \frac{3}{3 + 1} = 0.75$$

$$F1 - Score = 2 * \frac{0.60 * 0.71}{0.60 + 0.71} = 0.67$$

The average rating matrix can be calculated after calculating the rating matrix for each label. The average value is as follows.

$$Avg\ accuracy = \frac{0.80 + 0.97 + 0.87 + 0.93 + 0.90}{5} = 0.89$$

$$Avg\ precision = \frac{0.44 + 0.89 + 1.00 + 1.00 + 0.60}{5} = 0.79$$

$$Avg\ recall = \frac{0.80 + 1.00 + 0.33 + 0.71 + 0.75}{5} = 0.72$$

$$Avg\ f1 - score = \frac{0.57 + 0.94 + 0.50 + 0.83 + 0.67}{5} = 0.70$$

The average accuracy value is 0.89, denoting the percentage of correctly predicted infant diseases. Precision has a mean value of 0.79. This indicates that it's the percentage of correct out of all positive predictions made by the model (TP+FP). Recall has a mean value of 0.72. This indicates that it is the percentage of correctly predicted positive cases out of all available positive cases (TP+FN). The average f1-score is 0.70, representing the balanced percentage between precision and recall values.

4. Conclusion

Newman and Khon have investigated the evidence-based diagnosis thoroughly [33]particularly those who are academically minded, but it should be helpful and accessible to anyone involved with selection, development, or marketing of diagnostic, screening, or prognostic tests. Although we admit to a love of mathematics, we have restrained ourselves and kept the math to a minimum – a little simple algebra and only three Greek letters, ϰ (kappa. The diagnosis is an integration of information from medical history, observation, and examination by clinicians, testing using a combination of knowledge, experience, pattern recognition, and doctor intuition to refine the possibilities. Research by Bours [34] simply explains how Bayes' Theorem has been used for diagnostic processes based on evidence. It combines information from the prior to determine the posterior. The method that implements Bayes' rules to diagnose disease was also elaborated in detail by several authors [35]. Knottnerus et. al evaluated diagnostics procedures and concluded that diagnostic dealing with complex relations of parameters and appropriate implementation of Bayes depend on the research questions or types of diseases [36]. So this paper proposes an implementation of Bayes' Theorem to diagnose diseases that are usually suffered by infants, similar to some research conducted previously [37]–[39]. However, this research attempts to increase the effectiveness and efficacy of the model by using a dataset collected from real cases in Puskesmas Sorong. According to Knottnerus et. al, it is aimed to make the model generated by this study more suitable and applicable to be implemented in the real target, that in this case is in healthcare facilities in Sorong. A region in Indonesia that has limited expert clinicians, especially pediatricians.

A confusion matrix with a 5x5 matrix is shown in the test results of the Bayes' Theorem method. The test results were performed on 30 data. 22 data were successfully predicted correctly and 8 data were not successfully predicted correctly or were predicted incorrectly. This test succeeded in obtaining an accuracy value of 0.89 or 89%.

Bayes' Theorem method in diagnosing diseases suffered by infants has the advantage that the diagnosis does not require a weight value because the method performs calculations based on the probability of a symptom, while the drawback is that the method does not always produce a diagnosis according to an expert diagnosis, but can be considered if the symptoms experienced are based on available knowledge and rules.

An expert system for infant disease using Bayes' Theorem makes it easier for parents to know about their baby's disease so that they can take action in response to symptoms. Further research is expected to use different methods or a combination of methods, and data can be replicated for comparison or evaluation.

References

[1] V. Viviliani dan R. Tanone, "Perancangan Sistem Pakar Diagnosis Penyakit pada Bayi dengan Metode Forward Chaining Berbasis Android," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 1, hal. 1–13, Mei 2019, doi: 10.28932/jutisi.v5i1.1577.

[2] M. A. Arif dan A. R. Nugraha, "Sistem Pakar Penyakit Bayi Di Bawah Umur Lima Tahun," *Jurnal Manajemen dan Teknik Informatika*, vol. 02, no. 01, hal. 71–80, 2018.

[3] H. Herman, S. Sunardi, dan V. Muslimah, "Metode Dempster Shafer pada Sistem Pakar Penentuan Penyakit Bayi," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, hal. 1643, Okt 2022, doi: 10.30865/jurikom.v9i5.4908.

[4] D. P. Sari, "Sistem Pakar untuk Diagnosa Penyakit Anak yang Didukung oleh Bahasa Pemograman Visual," *Jurnal Sains dan Informatika*, vol. 3, no. 2, hal. 108–117, Nov 2017, doi: 10.22216/jsi.v3i2.2791.

[5] Kemenkes, "Pusat Kesehatan Masyarakat," *Badan Pemeriksa Keuangan Republik Indonesia*. https://peraturan.bpk.go.id/Home/Details/138635/permenkes-no-43-tahun-2019

[6] D. Susanto, A. Fadlil, dan A. Yudhana, "Application of the Certainty Factor and Forward Chaining Methods to a Goat Disease Expert System," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 6, no. 2, hal. 158–164, 2020.

[7] D. T. Yuwono, A. Fadlil, dan S. Sunardi, "Comparative Analysis of Dempster-Shafer Method and Certainty Factor Method on Personality Disorders Expert Systems," *Scientific Journal of Informatics*, vol. 6, no. 1, hal. 12–22, Mei 2019, doi: 10.15294/sji.v6i1.17216.

[8] A. I. Friska, T. Rismawan, dan S. Bahri, "Aplikasi Sistem Pakar Diagnosa Penyakit pada Anak dengan Inference Forward Menerapkan Metode Dempster Shafer Berbasis Web," *Coding: Jurnal Komputer dan Aplikasi*, vol. 06, no. 02, hal. 25–35, 2018.

[9] M. M. Al-Masawabe dan S. S. Abu-Naser, "Expert System for Short-term Abdominal Pain (Stomach Pain) Diagnosis and Treatment," *International Journal of Academic Information Systems Research (IJAISR)*, vol. 5, no. 5, hal. 37–56, 2021.

[10] A. S. A. Mettleq, I. M. Dheir, A. A. Elsharif, dan S. S. Abu-Naser, "Expert System for the Diagnosis of Seventh Nerve Inflammation (Bell's palsy) Disease," *International Journal of Academic Information Systems Research (IJAISR)*, vol. 3, no. 4, hal. 27–35, 2019.

[11] P. S. Ramadhan dan U. F. S. Sitorus Pane, "Analisis Perbandingan Metode (Certainty Factor, Dempster Shafer dan Teorema Bayes) untuk Mendiagnosa Penyakit Inflamasi Dermatitis Imun pada Anak," *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 17, no. 2, hal. 151, Agu 2018, doi: 10.53513/jis.v17i2.38.

[12] B. Sasangka dan A. Witanti, "Sistem Pakar Diagnosa Penyakit Infeksi Saluran Pernafasan Akut Pada Anak Menggunakan Teorema Bayes," *JMAI (Jurnal Multimedia & Artificial Intelligence)*, vol. 3, no. 2, hal. 45–51, Sep 2019, doi: 10.26486/jmai.v3i2.83.

[13] R. Noviani dan S. Sulindawaty, "Sistem Pakar Mendiagnosa Gizi Buruk Pada Balita Menggunakan Teorema Bayes," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 3, no. 2, hal. 163–169, Okt 2020, doi: 10.32672/jnkti.v3i2.2383.

[14] A. A. Maulia, K. Kusrini, dan A. M. Wijaya, "Analisa Metode Certainty Factor dan Teorema Bayes Dalam Mendeteksi Suatu Penyakit," *Zeta - Math Journal*, vol. 6, no. 1, hal. 6–10, Des 2020, doi: 10.31102/zeta.2021.6.1.6-10.

[15] P. T. Prasetyaningrum dan N. B. Hangesti, "Sistem Pakar Diagnosa Penyakit Kulit Akibat Virus Menggunakan Teorema Bayes," *Telematika*, vol. 15, no. 2, hal. 117–125, Okt 2018, doi: 10.31315/telematika.v15i2.3128.

[16] S. Murni dan F. Riandari, "Penerapan Metode Teorema Bayes pada Sistem Pakar untuk Mendiagnosa Penyakit Lambung," *Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)*, vol. 1, no. 2, hal. 19–25, Okt 2018, doi: 10.34012/jutikomp.v1i2.226.

[17] S. Nainggolan dan F. A. Sianturi, "Sistem Pakar Mendiagnosa Penyakit Kulit Menggunakan Metode Metode Teorema Bayes," *JIKOMSI [Jurnal Ilmu Komputer dan Sistem Informasi]*, vol. 3, no. 3, hal. 192–196, 2021.

[18] H. T. Sihotang, E. Panggabean, dan H. Zebua, "Sistem Pakar Mendiagnosa Penyakit Herpes Zoster dengan Menggunakan Metode Teorema Bayes," *Journal Of Informatic Pelita Nusantara*, vol. 3, no. 1, hal. 33–40, 2019, doi: 10.31227/osf.io/rjqgz.

[19] R. Rachman dan S. Moritami, "Sistem Pakar Deteksi Penyakit Refraksi Mata dengan Metode Teorema Bayes Berbasis Web," *Jurnal Informatika*, vol. 7, no. 1, hal. 68–76, 2020, doi: 10.31294/ji.v7i1.7267.

[20] R. A. Putri, S. Sendari, dan T. Widiyaningtyas, "Classification of Toddler Nutrition Status with Anthropometry Calculation using Naïve Bayes Algorithm," in *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, Nov 2018, hal. 66–70. doi: 10.1109/SIET.2018.8693169.

[21] C.-S. Chang, Y.-T. Yeh, T.-W. Chien, J.-C. J. Lin, B.-W. Cheng, dan S.-C. Kuo, "The computation of case fatality rate for novel coronavirus (COVID-19) based on Bayes theorem," *Medicine*, vol. 99, no. 21, hal. 1–9, Mei 2020, doi: 10.1097/MD.0000000000019925.

[22] N. Rosmawanti dan G. P. Kusumawardhani, "Model Sistem Pakar Diagnosa Penyakit Gagal Ginjal Menggunakan Metode Teorema Bayes," *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, hal. 205, Jan 2021, doi: 10.35889/jutisi.v9i3.560.

[23] F. J. Nababan, "Sistem Pakar Mendiagnosa Kelahiran Bayi Prematur Dengan Menggunakan Metode Teorema Bayes," *Jurnal Pelita Informatika*, vol. 6, no. 2, hal. 233–237, 2017.

[24] A. Trihartati S. dan C. K. Adi, "An Identification

of Tuberculosis (Tb) Disease in Humans using Naïve Bayesian Method," *Scientific Journal of Informatics*, vol. 3, no. 2, hal. 99–108, Nov 2016, doi: 10.15294/sji.v3i2.7918.

[25] A. A. Perbawawati, E. Sugiharti, dan M. A. Muslim, "Bayes Theorem and Forward Chaining Method On Expert System for Determine Hypercholesterolemia Drugs," *Scientific Journal of Informatics*, vol. 6, no. 1, hal. 116–124, Mei 2019, doi: 10.15294/sji.v6i1.14149.

[26] P. S. Ramadhan, "E-Pediatric dalam Mendiagnosis Leprosy pada Anak Menggunakan Analisis Kombinasi K-Nearest Neighbor dan Teorema Bayes," in *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2019, vol. Juli, hal. 540–544.

[27] B. Mondal, "Artificial Intelligence: State of the Art," in *Intelligent Systems Reference Library*, vol. 172, 2020, hal. 389–425. doi: 10.1007/978-3-030-32644-9_32.

[28] F. Ramadhana, F. Fauziah, dan W. Winarsih, "Aplikasi Sistem Pakar untuk Mendiagnosa Penyakit ISPA menggunakan Metode Naive Bayes Berbasis Website," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 4, no. 3, hal. 320, 2020, doi: 10.30998/string.v4i3.5441.

[29] K. E. Setyaputri, A. Fadlil, dan S. Sunardi, "Comparative Analysis of Certainty Factor Method and Bayes Probability Method on ENT Disease Expert System," *Scientific Journal of Informatics*, vol. 5, no. 2, hal. 205–212, Nov 2018, doi: 10.15294/sji.v5i2.16151.

[30] P. S. Ramadhan, "Sistem Pakar Pendiagnosaan Dermatitis Imun Menggunakan Teorema Bayes," *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, vol. 3, no. 1, hal. 43–48, Sep 2018, doi: 10.30743/infotekjar.v3i1.643.

[31] M. R. A. Nasution dan M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *Jurnal Informatika*, vol. 6, no. 2, hal. 226–235, Sep 2019, doi: 10.31311/ji.v6i2.5129.

[32] A. Desiani, R. Primartha, M. Arhami, dan O. Orsalan, "Naive Bayes classifier for infant weight prediction of hypertension mother," *Journal of Physics: Conference Series*, vol. 1282, no. 1, hal. 1–9, Jul 2019, doi: 10.1088/1742-6596/1282/1/012005.

[33] T. B. Newman dan M. A. Kohn, *Evidence-Based Diagnosis*. 2020. doi: 10.1017/9781108500111.

[34] M. J. Bours, "Bayes' rule in diagnosis," *Journal of Clinical Epidemiology*, vol. 131, hal. 158–160, 2021, doi: 10.1016/j.jclinepi.2020.12.021.

[35] J. A. Knottnerus dan F. Buntinx, *The Evidence Base of Clinical Diagnosis: Theory and methods of diagnostic research*, Second Edi. 2009. doi: 10.1002/9781444300574.

[36] J. A. Knottnerus, F. Buntinx, dan C. van Weel, "General Introduction: Evaluation of Diagnostic Procedures," in *The Evidence Base of Clinical Diagnosis: Theory and methods of diagnostic research*, 2009, hal. 1–19.

[37] S. Manalu, "Expert System for Identification of Diseases of Toddlers using Bayes' Theorem Method," *International Journal of Basic and Applied Science*, vol. 10, no. 1, hal. 9–16, 2021.

[38] R. G. Tsiko, "Bayesian spatial analysis of childhood diseases in Zimbabwe," *BMC Public Health*, vol. 15, no. 1, hal. 1–24, Des 2015, doi: 10.1186/s12889-015-2182-7.

[39] S. Wang, M. Li, dan S. B. Ng, "Research on Infant Health Diagnosis and Intelligence Development Based on Machine Learning and Health Information Statistics," *Frontiers in Public Health*, vol. 10, no. June, hal. 1–11, 2022, doi: 10.3389/fpubh.2022.846598.