Comparison of Dempster Shafer AND Certainty Factor Methods in Expert System for Early Diagnosis of Stroke Disease

Pretty Fujianti Febrivia a,1, Alvi Syahrini Utami b,2,* , Desty Rodiah b,3

aInformatics, Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia
1prettyff11@gmail.com; 2 alvi.syahrini@ilkom.unsri*; 3 desty.rodiah@unsri.ac.id
* corresponding author

ABSTRACT

Stroke is one of endangering disease if not treated properly and could lean to death. Most people unwilling to check their health because of high cost, lack of medical service, medical staff of neurologist and their limited working time. Therefore, we need an expert system that can help in early diagnosis of stroke. The Dempster Shafer and Certainty Factor methods are expert systems methods used in many cases to support uncertainty from the expert. The aim of this study is to compare two methods to determine the best method in the expert system for diagnosing stroke, by calculating symptoms so as to produce CF values in the Certainty Factor method and density values in the Dempster Shafer method. The data used in the study to diagnose stroke consisted of data on eighteen disease symptoms and two types of stroke identified. Based on the results of testing on 105 test data, the accuracy value of the expert system for diagnosing stroke using the Dempster Shafer method is 95.2% and the accuracy value of the expert system for diagnosing stroke with the Certainty factor method is 98.1%.

Keywords
Dempster Shafer
Certainty Factor
Expert System
Comparison
Stroke Disease

1. Introduction

Stroke is an emergency condition that needs to be treated as soon as possible, because brain cells can die in just a matter of minutes. Stroke is the main cause of death in almost all hospitals in Indonesia and occupies the third position after heart disease and cancer [1]. In definitively diagnosing stroke, one usually has to use supporting tools in the form of a CT scan [2] and examination with magnetic resonance imaging (MRI), a general physical examination, and a neurological examination [3]. This obviously takes a long time and is also expensive. In addition, not all hospitals in Indonesia have CT Scan equipment so that for further examination it is necessary to refer to another hospital that has this tool [4].

The high cost, the lack of service to patients, the limitations of medical personnel, especially doctors and the limited working hours of doctors have resulted in the general public being reluctant to carry out health checks so they do not pay attention to health [5]. Meanwhile, stroke if not detected early or treated quickly and appropriately can cause other complications that can even lead to death [6]. Therefore, it is very necessary to detect stroke early. To overcome this, we need an expert system to help diagnose stroke as an alternative information and more practical communication media, in which there is information about diagnosing stroke in order to make it easier for doctors, medical personnel, and the general public in carrying out a temporary diagnosis. so that it can also help the doctor in making a decision. The sooner a stroke is diagnosed, the better the stroke management will be so that the costs incurred are not too much and the death rate due to stroke can also be reduced.

Many methods have been applied in building an expert system, including the certainty factor method and the dempster shaper method. The certainty factor method and the dempster shaper method are methods that are both used to deal with the problem of uncertainty and ambiguity in expert systems [7]. According to (N. A. Sari, 2013) and [8], the certainty factor method and the Dempster Shafer method have similarities in the information to be calculated, that is, each piece of information or data from the two methods has an assessment taken from a person's belief. experts, but the concept and
process of completion of the two methods are different. Therefore, the dempster shafer and certainty factor methods need to be compared to be able to find out the difference between the two methods.

Previous expert system research has also been conducted to diagnose stroke. Research conducted by [9] obtained an accuracy of 80% using the fuzzy logic method. Furthermore, research conducted by [10] using the certainty factor method combined with the naive Bayes method to diagnose stroke can only produce an accuracy of 84%.

Another study conducted by [11] managed to achieve an accuracy of 95% by using the certainty factor method to temporarily diagnose diseases of children under five years (toddlers) in the Coastal Area of Bengkulu City with a total of 20 test data. In addition, in 2019 an expert system research was also conducted by [12] using the Dempster Shafer method for early diagnosis of gastric disease. This study provides an accuracy value of 95% on 20 test data and 94% on 100 test data.

Expert system research to compare the dempster shafer method and certainty factor has also been carried out. In a study conducted by [13] to diagnose ENT (Ear Nose Throat) disease concluded that the Dempster Shafer method is better and more accurate than the certainty factor method to deal with this problem with an accuracy value of 99.2% compared to with the certainty factor method which produces an accuracy value of 98.9%. However, in another study conducted by [14] in early diagnosis of postnatal depression, it concluded that the certainty factor method produces a better and more accurate accuracy value, which is 90% compared to the Dempster Shafer method which produces an accuracy value by 70%.

Based on the description above, in this final project the author wants to compare the expert system methods, namely the Dempster Shafer method and the certainty factor method in diagnosing stroke. By comparing the two expert system methods, it is hoped that a better and more accurate method of diagnosing stroke will be obtained with the same number of inputs through the symptoms of stroke.

2. Literature Study / Hypotheses Development

2.1 Artificial intelligence

One part of computer science that can make machines (computers) perform an activity like and as well as humans do, namely artificial intelligence or known as artificial intelligence (AI). At first the computer was used as a calculating tool, but along with the development of technology, computers are needed and used to carry out an activity that can also be done by humans [15].

2.2 Expert system

An expert system or other terms expert system is one of the fields of science from artificial intelligence (artificial intelligence) which is related to the scientific method of making machines that are useful for obtaining knowledge from an expert in solving a problem [16].

2.3 Dempster Shafer Method

The Dempster Shafer method is a mathematical theory that is used to prove probability based on belief functions and plausible reasoning, and is used to combine separate pieces of information so that the probability of an event will be calculated [17]. Dempster Shafer theory is written in an interval as follows:

\[
[\text{Belief, Plausibility}] \quad (1)
\]

Where :

1. Belief (Bel) is a measure of the strength of evidence in supporting a set of propositions. If it is 0 it means that there is no strength of evidence, while if it is 1 it means that there is certainty
2. Plausibility (Pls) is a measure of distrust of evidence. Plausibility (Pls) will reduce the level of confidence of the evidence where plausibility (Pls) is denoted as follows:

\[
\text{Pls}(X) = 1 - \text{Bel}(X)
\]  

(2)

In the Dempster Shafer theory, it is known that there is a probability density function denoted by \( m \) because not all evidence directly supports each item. If it is known that \( X \) is a subset of with \( m_1 \) as a density function and \( Y \) is also a subset of with \( m_2 \) as a density function, then the combination function of \( m_1 \) and \( m_2 \) as \( m_3 \) can be formed with the following formula:

\[
m_3(Z) = \frac{\sum_{xy=x\cap y} m_1(x)\cdot m_2(y)}{1 - \sum_{x=y=\emptyset} m_1(x)\cdot m_2(y)}
\]  

(3)

Where:

- \( m_3(Z) \) = mass function of evidence (Z), where Z is the new density value of the slices of \( m_1(X) \) and \( m_2(Y) \) divided by 1 minus the empty slice of \( m_1(X) \) and \( m_2(Y) \).
- \( m_1(X) \) = mass function of evidence (X)
- \( m_2(Y) \) = mass function of evidence (Y)

2.4 Certainty Factor Method

The Certainty Factor (CF) method or known as the certainty factor is a method used when facing a problem whose answer is uncertain and the uncertainty can be a probability [18].

There are several formulas for the combination of certainty factor rules used in diagnosing disease:

1. Certainty factor for rules with a single premise or symptom (single premise rules):

\[
\text{CF}_{\text{symptom}} = \text{CF}_{\text{user}} \times \text{CF}_{\text{pakar}}
\]  

(4)

2. If there are rules with similar conclusions or more than one symptom (similarly concluded rules):

\[
\text{CF}_{\text{combine}} = \text{CF}_{\text{old}} + \text{CF}_{\text{symptom}} \times (1 - \text{CF}_{\text{old}})
\]  

(5)

3. To calculate the percentage against disease:

\[
\text{CF}_{\text{percentage}} = \text{CF}_{\text{combine}} \times 100\%
\]  

(6)

2.5 Stroke Disease

Stroke is a condition in which brain cells are damaged due to lack of oxygen caused by impaired blood flow to the brain [19].

a. Types of Stroke and Symptoms and Treatment/Solutions

1. Ischemic Stroke

Ischemic stroke with other terms, namely stroke blockage is a stroke caused by a blockage in the blood vessels causing brain tissue to lack oxygen [19]. Symptoms or signs experienced by patients with ischemic stroke, namely:

- Loss of consciousness
- Weakness and or tingling on one side of the body either the left or the right
- Speak quietly
- Goofy face
- Difficulty swallowing
- Suddenly can't see

Therapy or solutions that can be given to patients with ischemic stroke are hospitalization by a neurologist with the aim of saving lives, reperfusion of disturbed brain areas, controlling risk factors and complications, rehabilitation, and preventing recurrent strokes. In addition, patients with ischemic stroke are usually given rtPA, antiplatelet, anticoagulant, brain vitamins and other drugs according to their condition.

2. Hemorrhagic stroke

Hemorrhagic stroke or known as hemorrhagic stroke is a stroke caused by bleeding associated with the rupture of blood vessels in the brain [19]. [20] Symptoms or signs of hemorrhagic stroke sufferers, namely:
- Loss of consciousness
- Disruption of neurological signs such as a sneezing mouth
- Paralysis of one side of the body with or without tingling on one side of the body
- Swallowing disorders
- Impaired vision

Therapy or solutions that can be given to patients with hemorrhagic stroke are treated for reperfusion to the brain area, controlling risk factors, rehabilitation and preventing recurrent strokes. In addition, if there are indications, surgery / surgery can be carried out to remove the blood that is in the head and reduce the pressure inside the head.

b. Stroke Risk Factors

There are various things or conditions that cause or exacerbate stroke which are called stroke risk factors. There are two risk factors for stroke, both hemorrhagic stroke and ischemic stroke [21]:

1. Non-modifiable risk factors:
- Age
- Gender
- Race
- Family History
- History of previous stroke

2. Modifiable risk factors
- Hypertension
- Diabetes
- Smoking
- Dyslipidemia

c. Stroke Prevention

1. Primary prevention

Primary prevention can be done in people who have never had atherosclerosis, including:
- Implement a healthy lifestyle
- Exercising
- Reduce stress
- Regulate cholesterol levels so that they are not high
- Quit smoking

2. Secondary prevention

Secondary prevention can be done when clinical symptoms of atherosclerosis have occurred, which is called the abbreviation ABCDEFG, namely:
- A: Acetosal, ace-inhibitors, anticoagulants: take drugs to control disease risk factors
- B: Beta blocker, body weight reduction: take medication and lose weight body
- C: Cholesterol control and cigarette smoking cessation: control cholesterol and quitsmoking
- D: Diabetes control and diet: diabetes control and sports food and increase knowledge
- E: Exercise and education: exercise and increase knowledge
- F: Family support: family support
- G: Glucose oxidation preservation: maintain the body's glucose oxidation

3. Methodology
3.1 Data Collection

The method of data collection in this study was obtained from data collection of medical recordsof stroke patients at the hospital. Bhayangkara Mayang Mangurai Jambi City as the test data in the study which amounted to 105 data. Then conduct interviews with experts, the purpose of interviewswith experts who aim to get accurate data about the type of stroke and its symptoms. The following are the types of stroke data, symptoms, the value of trust / belief in symptoms, stroke rules, and userconfidence values:

<table>
<thead>
<tr>
<th>Symptom Code</th>
<th>Disease Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>Decreased consciousness/tend to drowsy</td>
</tr>
<tr>
<td>G2</td>
<td>Difficulty speaking/understanding speech (speech slurred)</td>
</tr>
<tr>
<td>G3</td>
<td>Kelemahan di lengan atau di tungkai/kaki secara tiba-tiba</td>
</tr>
<tr>
<td>G4</td>
<td>Sudden weakness in the arms or legs/</td>
</tr>
<tr>
<td>G5</td>
<td>Loss of balance</td>
</tr>
<tr>
<td>G6</td>
<td>Loss of vision, blurred vision, or visual field disturbances</td>
</tr>
<tr>
<td>G7</td>
<td>Sudden severe headache</td>
</tr>
<tr>
<td>G8</td>
<td>Nausea or vomiting</td>
</tr>
<tr>
<td>G9</td>
<td>Difficulty swallowing</td>
</tr>
<tr>
<td>G10</td>
<td>Seizures without a history of previous seizures</td>
</tr>
<tr>
<td>G11</td>
<td>Loss of consciousness</td>
</tr>
<tr>
<td>G12</td>
<td>Loss of fine motor skills (movement) such as being unable to walk, unable to write</td>
</tr>
<tr>
<td>G13</td>
<td>Difficulty writing or reading</td>
</tr>
<tr>
<td>G14</td>
<td>Abnormalities in taste</td>
</tr>
<tr>
<td>G15</td>
<td>Weakness in the face suddenly (mouth drooping, mouth drooping, numbness in the cheeks)</td>
</tr>
<tr>
<td>G16</td>
<td>Have a history of diabetes mellitus (DM)</td>
</tr>
<tr>
<td>G17</td>
<td>Have a history of hypertension</td>
</tr>
<tr>
<td>G18</td>
<td>Have a history of previous stroke</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Type of disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Hemorrhagic Stroke</td>
</tr>
<tr>
<td>P2</td>
<td>Ischemic Stroke</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symptom Code</th>
<th>Value of Trust/Belief</th>
<th>Hemorrhagic Stroke Disease</th>
<th>Ischemic Stroke Penyakit</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>0.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>0.8</td>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td>G3</td>
<td>0.7</td>
<td></td>
<td>0.7</td>
</tr>
</tbody>
</table>
Febrivia, Pretty Fujianti, et.al (Comparison of Dempster Shafer AND Certainty Factor Methods in Expert System)

### Table 4. Stroke Disease Rule

<table>
<thead>
<tr>
<th>No</th>
<th>Rule</th>
<th>Rule</th>
</tr>
</thead>
</table>

### Table 5. User Certainty Value (CF)

<table>
<thead>
<tr>
<th>Expert Statement</th>
<th>Scale or User Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don't know</td>
<td>0.2</td>
</tr>
<tr>
<td>A little sure</td>
<td>0.2</td>
</tr>
<tr>
<td>Pretty Sure</td>
<td>0.6</td>
</tr>
<tr>
<td>Certain</td>
<td>0.8</td>
</tr>
<tr>
<td>Very certain</td>
<td>1.0</td>
</tr>
</tbody>
</table>

#### 3.2 Research Testing

Testing begins by entering test data into the system, then the system will calculate the probability value of each case using the dempster shader and certainty factor methods. The output of the system is the type of stroke, the percentage value of belief/belief in the natural disease, and the solution to the disease. The following is a comparative research framework for the Dempster Shafer method and certainty factor in an expert system for diagnosis of stroke.
4. Result and Discussion

Software testing is done by checking whether the software output is the same as the output that has been determined by the expert. If the output produced by the software is the same as the results determined by the expert, then the test result is "Appropriate" and if the output produced by the software is different from the results determined by the expert, then the test result is "Not Appropriate". The software testing was carried out using 105 data that the author had obtained from the data in 2021. Table 6 shows the results of the tests that have been carried out in the study as many as 3 samples of random test data from 105 data.

Table 6. Sample Research Test Results

<table>
<thead>
<tr>
<th>Trial to-</th>
<th>Test Data</th>
<th>Expert Diagnosis</th>
<th>System Diagnosis by Method</th>
<th>Dempster Shafer</th>
<th>Certainty Factor</th>
<th>Dempster Shafer Conclusion</th>
<th>Certainty Factor Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Difficulty speaking / understanding speech (speech slurred) / Sudden weakness in the arms or legs / Tingling/numbness in the face, arms or legs / Have a history</td>
<td>Ischemic Stroke</td>
<td>Hemorrhagic Stroke, Ischemic Stroke</td>
<td>99.565%</td>
<td>93.8597</td>
<td>Not appropriate</td>
<td>Appropriate</td>
</tr>
</tbody>
</table>

Fig 1. Comparative Research Framework of Dempster Shafer Method and Certainty Factor in Stroke Disease Diagnosis Expert System

Febrivia, Pretty Fujianti, et.al (Comparison of Dempster Shafer AND Certainty Factor Methods in Expert System)
From the results obtained, it is possible to calculate the accuracy of the Dempster Shafer method and the certainty factor using the accuracy level testing formula where the accuracy value is equal to the amount of appropriate test data divided by the total number of test data multiplied by 100 percent. So from the results of the research that has been done, it can be obtained the accuracy values of the two methods which are detailed in table 7.

### Table 7. Analysis of Test Results

<table>
<thead>
<tr>
<th>Method Used</th>
<th>Amount of “Appropriate” Data</th>
<th>Percentage Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dempster Shafer</td>
<td>100</td>
<td>95.2%</td>
</tr>
<tr>
<td>Certainty Factor</td>
<td>103</td>
<td>98.1%</td>
</tr>
</tbody>
</table>

Based on the results of research testing on 105 test data obtained, it shows that if using the Dempster Shafer method there are 100 data that are in accordance with the expert's diagnosis and 5 data that are not appropriate. Meanwhile, if using the certainty factor method there are 103 data that are in accordance with the expert's diagnosis and 2 data that are not in accordance with the expert's diagnosis. This may be due to other considerations used by experts that cannot be determined by the expert system software in diagnosing a disease.

Then, the results of the diagnosis using the Dempster Shafer method are still not able to
classify the disease properly where the diagnosed stroke consists of two diagnosed strokes, namely hemorrhagic and ischemic strokes which are not in accordance with the results of the expert's diagnosis. One example is shown in the first experiment in table 6 where the output or diagnosis using the Dempster Shafer method is hemorrhagic and ischemic stroke, while the output or diagnosis using the certainty factor method is ischemic stroke. This may be due to other considerations used by experts that cannot be determined by expert system software in diagnosing a disease and can also be caused by fundamental differences in the calculation process of the two methods, where the calculation of each possible disease in the certainty factor method is not mutually exclusive, are related and calculated separately, while the calculation of every possible disease in the Dempster-Shafer method is always interrelated which will form a new rule base from the random selection of symptoms.

In addition, judging from the value of confidence / trust in the results of disease diagnosis produced by the Dempster Shafer method and the certainty factor method, it shows that the certainty factor method can provide the highest belief / confidence value of 99.0387553042432%, while the highest belief / confidence value is produced by the Dempster method of 98.4%.

5. Conclusion

Based on the results of a comparative study of the Dempster Shafer method and certainty factor in an expert system for early diagnosis of stroke, it can be concluded that:

1. An expert system for early diagnosis of stroke using the Dempster Shafer method and the certainty factor was successfully applied.
2. The results of trials on 105 data carried out by comparing the accuracy value of the diagnosis results from the expert system software that has been developed with the diagnosis results from the experts give the result that the expert system accuracy value using the Dempster Shafer method is 95.2%, while the accuracy value is 95.2%. Expert system using the certainty factor method is 98.1%.
3. The certainty factor method is better and more accurate in diagnosing stroke because the results of the diagnosis of the disease given can be classified better than the Dempster Shafer method, the value of confidence / trust in the diagnosis results with the certainty factor method can reach confidence of 99.0387553042432%, then the resulting accuracy value expert system with certainty factor method is greater than the Dempster Shafer method.

References


Febrivia, Pretty Fujianti, et.al (Comparison of Dempster Shafer AND Certainty Factor Methods in Expert System)