ABSTRACT

Objective: Diagnosis process of Thalassemia requires several types of medical test, and results of this test together identify the stage of Thalassemia. The objective of this study is to design a Fuzzy Inference System to diagnose the severity of the Thalassemia disease of a patient by using Fuzzy Logic.

Methods: In this paper, a new approach based on fuzzy inference system was presented for prediction of Thalassemia disease in patients. The proposed Fuzzy model combined the expert's knowledge and the fuzzy logic approach which is then combined in fuzzy rule base to diagnose the presence of the disease. The performances of the system graphically represented by fuzzy inference system tools in MATLAB.

Results: It was found that our program matched the doctor's diagnosis in 12 cases perfectly. The other 3 were marginally off. This results with an accuracy of about 80%.

Conclusion: The result suggests that the model provides the most effective way to identify Thalassemia type in patients. The results in this work can be obtained by a simple and inexpensive method. This would generate, in economic terms, significant savings.

Keywords: CBC Test, Fuzzy Logic, Mamdani Fuzzy Inference System, Thalassemia Disease

INTRODUCTION

Anemia is a condition where the number of healthy RBC in the blood is lower than normal. It is due to low RBC’s, destruction of RBC’s or loss of too many RBC’s. If your blood does not have enough RBC’s, your body doesn’t get enough oxygen it needs. As a result, you may feel tired and other symptoms. But sometimes it is very difficult to detect Thalassemia on the basis of symptoms only. In the domain of Thalassemia, there is no such boundary between what is healthy and what is diseased. Having so many factors to detect Thalassemia makes doctor’s work difficult. So, experts require an accurate tool that considering these risk factors and give some certain result in uncertain terms. Some biochemical tests (HGB, HCT, HbA2, RBC, MCV, and MCH) are useful for identifying carriers of the Thalassemia trait [1-3]. In the presence of Thalassemia parameters in the CBC, an accurate and precise quantification of hemoglobin HbA2 is essential for the diagnosis of the Thalassemia trait. When biochemical tests are not exhaustive, it is necessary to study the molecular globin genes [4]. HPLC and electrophoresis are a gold standard for the diagnosis of ß-Thalassemia trait, but it is not available at all places.

Thus, several attempts have been made to diagnose the condition by using red cell indices. In the present study Hemoglobin (HGB), Mean Corpuscular Volume (MCV) and Mean corpuscular hemoglobin (MCH) values were able to detect cases of type of Thalassemia. In this study, we focus on a development of the first knowledge representation corresponding to the CBC test results to create a mathematical model for Thalassemia disease diagnosis using FIS. The developed FIS model for Thalassemia disease will be useful and beneficial for many informatics related Thalassemia tasks in the future. The objective of this study is to create a fuzzy inference system to predict the severity involve in Thalassemia disease. Three steps are used to monitor general health and Thalassemia. But we are focusing only on the Tests and Procedures. Three steps are as follows:

- Medical and Family Histories
- Physical Exam
- Tests and Procedures.
Ranges for Input/output fields of the system

Linguistic Variables:

- For Input Variables

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Input variables</th>
<th>Linguistic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Hemoglobin (g/dl)</td>
<td>HGB</td>
</tr>
<tr>
<td>2.</td>
<td>Mean Corpuscular Volume (fl)</td>
<td>MCV</td>
</tr>
<tr>
<td>3.</td>
<td>Mean corpuscular hemoglobin (pg)</td>
<td>MCH</td>
</tr>
</tbody>
</table>

- For Output Variable

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Output variable</th>
<th>Linguistic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Thalassemia</td>
<td>Type_of_Thalassemia</td>
</tr>
</tbody>
</table>

Define Fuzzy Sets:

- Input Variables and Value Ranges

**Membership function for all input variables**

For calculating the membership function (MF) we scale the range of 0-100 and based on the severity we calculate the MF and the range are given as follows:

**Output variables and value ranges**

The output will be a value within the range [0, 10]. The value, 0, means that no Thalassemia problems exist as of yet. We have divided this range into smaller fuzzy sets to make a cluster of the type of Thalassemia disease. ‘Thalassemia_Minor’ is given to those patients whose output value is in between 0 and 3.4 ‘Thalassemia_Intermedia’ is given to the patients who gets a value between 3.5 and 7 also ‘Thalassemia_Major’ is given to the patients who gets a value between 7.1 and 9.9 and so on as shown in the table 5. The basic relationship is that the higher the severity of Thalassemia disease, the lower the output value.
Table 4: Membership function for all input variables

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Linguistic variables</th>
<th>Membership function</th>
<th>Values</th>
</tr>
</thead>
</table>
| 1.     | HGB                  | $\mu_{Very\_Low}(x) = \begin{cases} 
\frac{1}{x - 6} & x < 6 \\
\frac{6 - x}{0.9} & 6 \leq x \leq 6.9 \\
\frac{x - 7}{0.9} & 7 \leq x \leq 8.4 \\
\frac{x - 0.5}{8.5} & x = 8.5 \\
\frac{x - 10}{1} & 8.5 \leq x \leq 10 \\
\frac{x - 10}{1.5} & 10 \leq x \leq 12 \\
\frac{x - 15}{5} & 12 \leq x \leq 15 \\
\end{cases}$ | Very_Low |
| 2.     | MCV                  | $\mu_{Low}(x) = \begin{cases} 
\frac{10}{x - 60} & x \leq 30 \\
\frac{60 - x}{10} & 30 \leq x \leq 40 \\
\frac{x - 70}{1} & x = 70 \\
\frac{x - 70}{10} & 70 \leq x \leq 80 \\
\frac{x - 100}{10} & 100 \leq x \leq 100 \\
\end{cases}$ | Low |
| 3.     | MCH                  | $\mu_{Medium}(x) = \begin{cases} 
\frac{5}{x - 10} & x \leq 5 \\
\frac{10 - x}{5} & 5 \leq x \leq 10 \\
\frac{x - 15}{5} & 15 \leq x \leq 20 \\
\frac{x - 25}{5} & 25 \leq x \leq 30 \\
\end{cases}$ | Medium |

Table 5: Values for all output linguistic variables

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Linguistic variables</th>
<th>Ranges</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Type_of_Thalasemia</td>
<td>HGB is 9–12 g/dl</td>
<td>Thalassmia_Minor [12, 13]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCV is &lt;80 fl</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCH is &lt;27 pg</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td>HGB is 7–10 g/dl</td>
<td>Thalassemia_Intermedia [14]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCV is 50–80 fl</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCH is 16–24 pg</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td>HGB is &lt;7 g/dl</td>
<td>Thalassemia_Major [14]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCV is 50–&lt;70 fl</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCH is 12–&lt;20 pg</td>
<td></td>
</tr>
</tbody>
</table>

[HGB = Hemoglobin, MCV = Mean corpuscular volume, MCH = Mean corpuscular hemoglobin]

Table 6: Classification of output

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Linguistic variable</th>
<th>Ranges</th>
<th>Fuzzy set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Type_of_Thalasemia</td>
<td>&lt;3.5</td>
<td>Thalassmia_Minor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.5–7</td>
<td>Thalassemia_Intermedia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10&gt;</td>
<td>Thalassemia_Major</td>
</tr>
</tbody>
</table>
Define fuzzy rules

In this section fuzzy inference rules generated; relevant inference rules can be determined by experience human operators well; we use IF-ELSE conditions for fuzzy inference rules, as we have three input variables. Also, there are 15 rules conveniently are represented in IF-ELSE Form.

First 3 rules are for Symptoms based testing:
1. If (Symptoms is Absent) and (Family_History is Present) and (Disorder is Thalassemia_Minor) then HGB is Medium.
2. If (Symptoms is Present) and (Family_History is Present) and (Disorder is Thalassemia_Intermedia) then HGB is Low.
3. If (Symptoms is Present) and (Family_History is Present) and (Disorder is Thalassemia_Major) then HGB is Very_Low.

Further, 3 rules are in the classification of Thalassemia on the basis of MCV only:
1. If (HGB is Medium) and (MCV is Very_Low) then Thalassemia_Minor.
2. If (HGB is Low) and (MCV is Low) then Thalassemia_Intermedia.
3. If (HGB is Very_Low) and (MCV is Medium) then Thalassemia_Major.

At last 15 rules are for the further classification of Thalassemia on the basis of all three parameters such as HGB, MCV, and MCH:
1. If (HGB is Medium) and (MCV is Very_Low) and (MCH is Very_Low) then Type_of_Thalassemia is Thalassemia_Minor.
2. If (HGB is Medium) and (MCV is Very_Low) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Minor.
3. If (HGB is Medium) and (MCV is Very_Low) and (MCH is Medium) then Type_of_Thalassemia is Thalassemia_Minor.
4. If (HGB is Medium) and (MCV is Low) and (MCH is Very_Low) then Type_of_Thalassemia is Thalassemia_Minor.
5. If (HGB is Medium) and (MCV is Low) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Minor.
6. If (HGB is Medium) and (MCV is Low) and (MCH is Medium) then Type_of_Thalassemia is Thalassemia_Minor.
7. If (HGB is Medium) and (MCV is Medium) and (MCH is Very_Low) then Type_of_Thalassemia is Thalassemia_Minor.
8. If (HGB is Medium) and (MCV is Medium) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Minor.
9. If (HGB is Medium) and (MCV is Medium) and (MCH is Medium) then Type_of_Thalassemia is Thalassemia_Minor.
10. If (HGB is Low) and (MCV is Low) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Intermedia.
11. If (HGB is Low) and (MCV is Low) and (MCH is Medium) then Type_of_Thalassemia is Thalassemia_Intermedia.
12. If (HGB is Low) and (MCV is Medium) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Intermedia.
13. If (HGB is Low) and (MCV is Medium) and (MCH is Medium) then Type_of_Thalassemia is Thalassemia_Intermedia.
14. If (HGB is Very_Low) and (MCV is Low) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Major.
15. If (HGB is Very_Low) and (MCV is Medium) and (MCH is Low) then Type_of_Thalassemia is Thalassemia_Major.

The AND operator used in the rule infers that the minimum criterion is used in the resultant and aggregated to indicate the condition of Thalassemia: Thalassemia_Minor, Thalassemia_Intermedia, and Thalassemia_Major. Also, the combination of each rule can be written in the following (table 7) form:

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Antecedent</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Medium (HGB) and (MCV is Very_Low) and (MCH is Very_Low)</td>
<td>Type_of_Thalassemia is Thalassemia_Minor</td>
</tr>
<tr>
<td>2</td>
<td>Medium (HGB) and (MCV is Very_Low) and (MCH is Low)</td>
<td>Type_of_Thalassemia is Thalassemia_Minor</td>
</tr>
<tr>
<td>3</td>
<td>Medium (HGB) and (MCV is Very_Low) and (MCH is Medium)</td>
<td>Type_of_Thalassemia is Thalassemia_Minor</td>
</tr>
<tr>
<td>4</td>
<td>Medium (HGB) and (MCV is Low) and (MCH is Very_Low)</td>
<td>Type_of_Thalassemia is Thalassemia_Minor</td>
</tr>
<tr>
<td>5</td>
<td>Medium (HGB) and (MCV is Low) and (MCH is Low)</td>
<td>Type_of_Thalassemia is Thalassemia_Minor</td>
</tr>
<tr>
<td>6</td>
<td>Medium (HGB) and (MCV is Low) and (MCH is Medium)</td>
<td>Type_of_Thalassemia is Thalassemia_Minor</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>15</td>
<td>Very_Low (HGB) and (MCV is Medium) and (MCH is Low)</td>
<td>Type_of_Thalassemia is Thalassemia_Major</td>
</tr>
</tbody>
</table>

Fig. 3: Rule viewer for generated rules
Experimental results

For our system, we use the mamdani type FIS. For each rule, a degree is calculated. For aggregation, the maximum is taken from all the degrees. In the defuzzification process, the centroid method is used.

Now from Matlab2014b using this input values we derive the rule viewer for giving input as demonstrated in fig. 3 and surface plots using the surface viewer as demonstrated in fig. 4.

Fuzzy system is used to obtain the severity level which is the only output variable of the system.

The risk determines the level of severity of risk given the inputs.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Input variable</th>
<th>Value ranges</th>
<th>Ranges selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>HGB</td>
<td>7.9 g/dl</td>
<td>7&lt;7.9&lt;10</td>
</tr>
<tr>
<td>2.</td>
<td>MCV</td>
<td>54.1 fl</td>
<td>40&lt;54.1&lt;60</td>
</tr>
<tr>
<td>3.</td>
<td>MCH</td>
<td>21.9 pg</td>
<td>20&lt;21.9&lt;30</td>
</tr>
</tbody>
</table>

[HGB = Hemoglobin, MCV = Mean corpuscular volume, MCH = Mean corpuscular hemoglobin].

RESULTS AND DISCUSSION

The patient's health risk was found from the given input of a linguistic variable of HGB, MCV, and MCH. Using Mamdani FIS method to construct the membership function for assigned the linguistic variation in the gasification process. Using If... then rule and inference strategies are chosen for processing the rule base to determine the type of Thalassemia in a patient by logical decision-making analysis. Through the defuzzification, fuzzy system provides an objective process of the risk factor, also to view the surface view of the risk determination using simulation. We tested our fuzzy expert system against the following input variable values.

To test the accuracy of our system we have taken 15 random patients' data from our dataset of 40 patients and used it to give their diagnoses. The randomness of the patients was important as that would account for a proper sample to test with, statistically. The idea is to send 15 patients' Data to the doctor who will give us his diagnosis. We implemented a system which ranges from Thalassemia, Minor (1) to Thalassemia, Major (3) and compared our program results with that of a doctor from registered patients from the Thalassemia Welfare Society, Bhilai, Chhattisgarh, India.

It was found that our program matched the doctor’s diagnosis in 12 cases perfectly. The other 3 were marginally off. This results with an accuracy of about 80%.

CONCLUSION

Because of uncertainty involve in the diagnosis of Thalassemia disease a new method for Thalassemia disease diagnostic problem solving based on fuzzy inference system is constructed in this paper. According to the fuzzy inference system step by step actions is performed. The constructed system in this paper is an efficient attempt to solve the Thalassemia disease problem. The proposed model detects Thalassemia on the basis of Thalassemia both Symptoms and CBC Test. The results of this work can facilitate laboratory work by reducing the time and cost.

ACKNOWLEDGMENT

I wish to express my deep sense of gratitude to Mr. Pramod Puri, General Secretary of Thalassemia Welfare Society, Bhilai (Chhattisgarh, India), for his excellent guidance, the valuable suggestion that greatly helped me to complete the work successfully.

ABBREVIATION

CBC = Complete blood count, FIS = Fuzzy inference system, HGB = Hemoglobin, HbF = Fetal hemoglobin, HbA2 = Hemoglobin A2, HPLC = High-performance liquid chromatography, MCV = Mean corpuscular volume, MCH = Mean corpuscular hemoglobin, RBC = Red blood cell.

CONFLICT OF INTERESTS

Declared none
REFERENCES