International Journal of Engineering & Technology, 7 (3.29) (2018) 674-676

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

AI based Differentiation of Beta-Thalassemia and Iron Deficiency Anemia.

Aravinda Kandadai [1*], Prof. Ashok Shigli [2], Dr. Shaik Ayeesha Begam [3]

[1,2] B.V.R.I.T (N), [3] AmPath
**Corresponding Author E-mail : Aravinda.kandadai@gmail.com*

Abstract

Thalassemia is a genetic disorder that leads to variation in the protein structure of the hemoglobin molecule, which causes profound anemia that kills untreated affected children before the age of 3 years. Unfortunately, the traditional RBC indices based methods do not distinguish between β-thalassemia and a normal iron deficiency anemia, causing wrong treatment to be administered. Techniques such as HPLC, serum ferritin analysis and molecular studies are needed to help differentiate them. The main drawbacks of these techniques are that, it takes a long time to issue the results, it is expensive and is not available everywhere. Here, we attempt to use a heuristic algorithm as well as AI based techniques to accurately predict the occurrence of thalassemia from the easily available RBC data itself. The AI technique is based on support vector machine. An accuracy of 60 % and 96% are obtained for the heuristic algorithm and the AI technique respectively. Details of this work are presented in this paper.

Keywords: *Artificial intelligence; Beta-thalassemia; Iron deficiency anemia; Support Vector machine.*

1. Introduction

Thalassemia refers to a spectrum of diseases characterized by the reduction or absence in the synthesis of the globular chains of hemoglobin. The disease was first noted in the Mediterranean population [1]. Hemoglobin is made up of 4 protein chains (i.e., two alpha and two beta chains). Based on the chain that is not synthesized, thalassemia is of two types. They are:
1.Alpha-Thalassemia, and
2.Beta-Thalassemia.
As their names state if the alpha chains are not synthesized, then its alpha-thalassemia and if the beta chains are not synthesized, then its beta-thalassemia. Thalassemia is also classified based on the number of chains synthesized and the way it is mutated. If only one of the two beta chains are synthesized, then its minor thalassemia. But if both the beta chains are not synthesized then it is known as a major thalassemia. Alpha thalassemia is also classified in a similar way. The real danger of non-diagnosis or misdiagnosis of carriers of the thalassemia trait (thalassemia minor) is the potential homozygous offspring (a major thalassemia) [2].
In order to differentiate a Beta-thalassemia minor (BT) from Iron deficiency anemia (IDA), apart from a normal CBC (Complete Blood Count), a test to identify the different hemoglobin variants must be conducted along with molecular studies. This process consumes a lot of time to get the results and is also expensive.
In this work, support vector machine is used to come up with a classification in order to differentiate between Beta thalassemia and IDA, using only the Red cell indices (as shown in Table 1). This method will help in reducing the time consumed by performing HPLC (High Performance Liquid Chromatography) and molecular studies.

Support vector machine (SVM) was initially designed for separating the binary classes (k=2), with a maximum margin criterion. Margin separation means, in a binary class, a line separating two binary states such that, most of the state one are on one particular side and most of the state two are on the other side of the margin line [3]. SVM, in a multiclass system, works based on maximizing margin criterion where a hyper plane separates the different states.

Table 1: RBC indices that are considered

Indices	Normal Range
RBC	4.20-5.70
Hb	13.2-16.9
Hct	38.5-49.0
MCV	80-100 Fl
MCH	27-31 pg/cell
MCHC	32-36 g/Dl
RDW	11-15

2. Methodology

2.1. Developing a Deterministic Algorithm

The HPLC data for a given set of samples are taken along with their RBC indices, to determine whether the patient is suffering from BT or IDA. Reports and chromatogram were interpreted by observing HbA2 and HbF concentration for the identification of BT. These samples are used as control data.
The 12 secondary indices, mentioned in Table 2, which are a combination of the RBC indices, are computed for each individual sample. The correlation between the results of these secondary indices and the occurrence of BT is analyzed for each sample to check the accuracy of prediction of each index. An heuristic

algorithm, which is a combination of the best correlating indices, is developed. The developed algorithm is used to increase the percentage of accuracy to identify and differentiate between BT and IDA.

Table 2: List of secondary indices used to differentiate between BT and IDA and their relation to the primary RBC indices.

SECONDARY INDEX	FORMULA
Mentzer's index	MCV/RBC
RDWI	MCV X RDW/RBC
Shine and Lal (S&L)	MCV X MCV X MCH/100
Srivastava	MCH/RBC
Green and King (G&K)	MCV X MCV X RDW/HB X 100
Sirdah	MCV – RBC – (3 X HB)
Ehsani	MCV – (10 X RBC)
England and Fraser (E&F)	MCV – (5 X HB) – RBC – 3.4
Ricerca	RDW/RBC
MDHL	(MCH/MCV) X RBC
MCHD	MCH/MCV
RBC index	RBC count

The developed Algorithm is given in Figure 1.

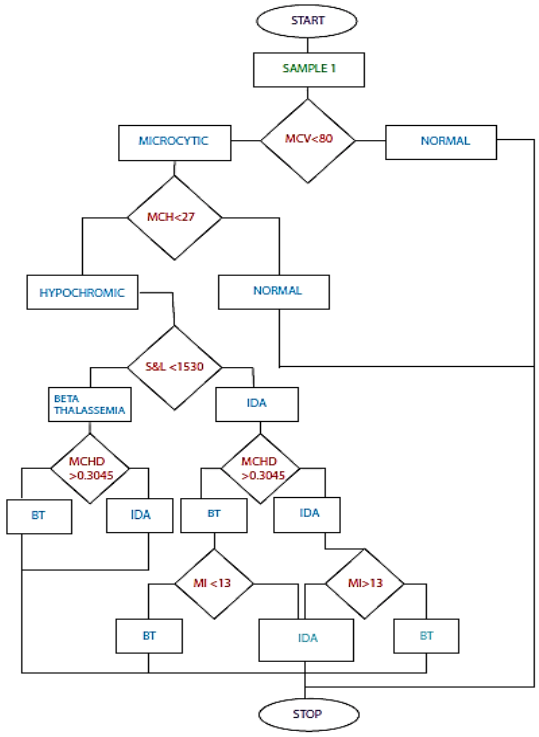

Figure 1: Algorithm to identify and differentiate between BT and IDA.

2.2. Developing an Artificial Intelligence based Algorithm.

A set of data, containing the RBC indices and HPLC results for each patient, were collected in order to determine the characteristics of BT and IDA for 1,500 samples. This data, using machine learning, is classified into three classes (i.e., BT, IDA and Normal). In order to classify the data, first, a classifier must be developed. To do this the machine must be trained.

1. Training the Machine: To train the machine, the data must be analyzed by the system. To do this, a set of training data, with the RBC indices for each sample along with the corresponding output, is taken and given to the machine. The machine uses this data to compute a 'Classifier' that can be used to analyze new data
This scheme is shown in Figure 2.

Figure 2: Training Model of SVM.

2. Using the Classifier for Identification: To check the predictability of the classifier a new set of sample data is given to the classifier as shown in Figure 3.

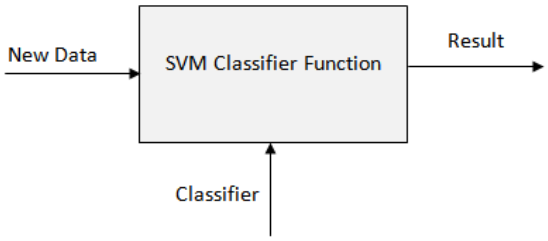

Figure 3: Testing Model of SVM to test the developed classifier.

2.3. Validation

1. Validation of the Algorithm: To validate the developed algorithm, a fresh set of sample data is collected. The data was run through the algorithm to check the accuracy of prediction of the algorithm. The areas in which this algorithm is lacking are noted and to compensate for these, variations in range limits were made. To do this, first the HPLC and RBC indices were taken into account to determine whether the patient is suffering from BT or IDA. This will help to validate the results by using the algorithm with the actual results to find the accuracy of prediction.

2. Validation of the AI Technique: To check the accuracy of prediction of the SVM based classifier, a fresh set of sample data that is used is validated separately at AmPath using their respective HPLC data.
The scheme for validation of the AI model is shown in Figure 4.

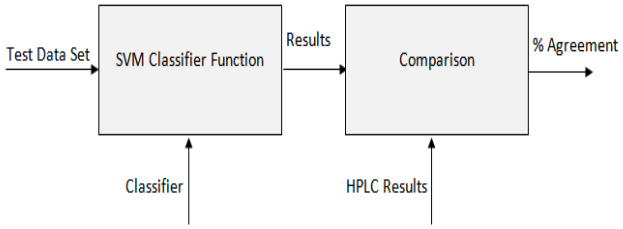

Figure 4: Validation of SVM based classifier

3. Results and Observation

3.1. For The Algorithm

The data pertaining to the RBC indices of the sample set are run through the developed heuristic algorithm. This algorithm has a success rate of 80%.
A fresh set of samples were collected and analyzed. These samples were given to the algorithm and the following was found.
• There was a variation in the probability of prediction.
• The predictability of the algorithm dropped to 60%.
• Variations were found with each index, and the hemoglobin percentages.

3.2. For the Ai based Technique

The AI based technique to differentiate between BT and IDA developed by using the support vector machine has an accuracy of prediction of 95.98%.
The representative output of AI inference engine, along with HPLC results is given in Table 3.

Table 3: Output of AI inference engine, along with HPLC results

S.NO	PREDICTED OUTPUT	HPLC RESULTS
1	'BT'	'BT'
2	'BT'	'BT'
3	'IDA'	'IDA'
4	'IDA'	'IDA'
5	'IDA'	'normal'
6	'normal'	'normal'
7	'normal'	'normal'
8	'IDA'	'IDA'
9	'IDA'	'IDA'
10	'IDA'	'IDA'

Various other experiments were conducted to check the efficiency of the machine by changing the training data set size. The size of the data set was taken as 250, 500, 1000, 1500 and 1667 respectively and a sample set of 600 is used for testing the algorithm developed by the machine for each of the 5 training sets. The resulted accuracy of detection for each of the given 5 training sizes is given in Table 4.

Table 4: The accuracy of detection for varying training sizes.

S.NO	TRAINING SIZE	ACCURACY
1	250	91.5
2	500	93.3
3	1000	94.7
4	1500	95.5
5	1667	95.6

The results shown in Table 4 are represented graphically in Figure 5.

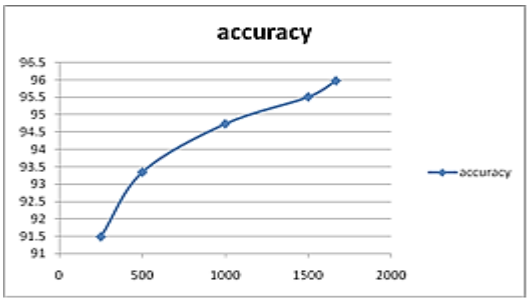

Figure 5: Graphical representation of Table 4.

A two dimensional classification using four different RBC indices are shown in Figure 6 and Figure 7.

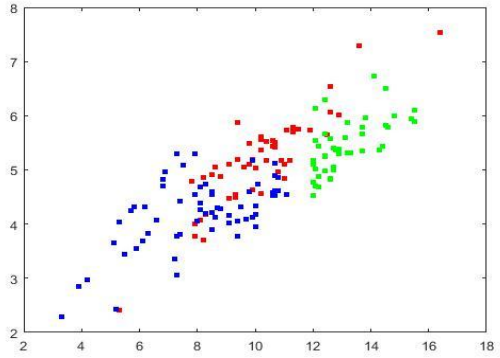

Figure 6: Two dimensional classification1

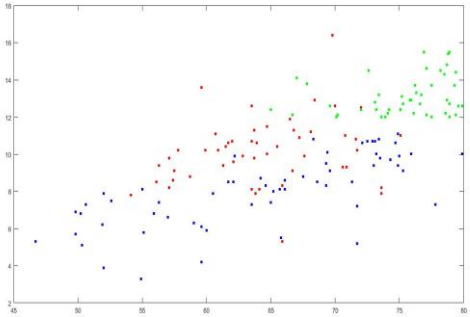

Figure 7: Two dimensional classification 2

4. Conclusion

The developed heuristic algorithm has an accuracy of 60%, and the SVM based classifier is used to get an accuracy of 96% in identifying and differentiating between BT and IDA.
Thus, the AI based classification is useful and efficient in identifying BT and IDA and is more accurate. Using this method will help reduce the time delay in producing the results and also deduct the cost of performing HPLC.

References

[1] Epidemiology of β-thalassemia in Western India: mapping the frequencies and mutations in sub-regions of Maharashtra and Gujarat Roshan Colah, Ajit Gorakshakar, Supriya Phanasgaonkar, Edna D'Souza, Anita Nadkarni, Reema Surve, Pratibha Sawant, Dilip Master, Ramesh Patel, Kanjaksha Ghosh, Dipika Mohanty, British Journal of Haematology; Volume 149, Issue 5, pages 739–747, June 2010

[2] Ehsani MA, Shahghol E, Rahiminejad MS, et al. A new index for discrimination between iron deficiency anemia and beta-thalassemia minor: results in 284 patients. Pak J Biol Sci.2009;12:473-475

[3] Wang Z., Xue X. (2014) Multi-Class Support Vector Machine. In: Ma Y., Guo G. (eds) Support Vector Machines Applications. Springer, Cham