Beta Thalassemia Major and Minor Classification using Artificial Neural Network

Anand Upadhyay
Department of Information Technology
Thakur College of Science & Commerce
Thakur village, Kandivali (E), Mumbai-101

ABSTRACT
The thalassemias are now the very dangerous group of anemia's caused by mutation affecting the synthesis of hemoglobin. The thalassemias are a bag of heterogeneous group of inherited anemias. The thalassemias are very common in the region of the persons of Mediterranean, African, and Southeast Asian descent. Thalassemia trait affects 5 to 30 percent of persons in these ethnic groups Scientists and public health officials predict that thalassemia will become a worldwide issue in the next century. Thalassemia consists of a number of different forms of anemia. The two main types are called Alpha & Beta Thalassemias. The Alpha and Beta thalassemia depends on which part of hemoglobin is lacking in the red blood cells. The impact of alpha thalassemias is concentrated in South East Asia, Malaysia and Southern China. The problem of beta thalassemia is seen primarily in the areas surrounding Mediterranean Sea, Africa and South East Asia. The main objective of this ANN model is to classify the Beta Thalassemias person based on there quantitative blood test. This model helps those people who are organizing the camp for Thalassemia person detection to spread awareness among people and prevent the birth of beta thalassemia major child by pre-marriage counseling.

General Terms
Data classification, Artificial Neural Network, Supervised Learning

Keywords
Thalassemias, Mutation, Alpha Thalassemia, Beta Thalassemia, Anemia

1. INTRODUCTION
The problem of Thalassemia is caused due to the group of inherited autosomal recessive blood disorders that originated in the Mediterranean region. Thalassemia is the genetic defect which is transmitted from one generation to the other generation. The genetic defect of thalassemia could cause either mutation or deletion, results in reduced rate of synthesis or no synthesis of one of the globin chains that make up hemoglobin's. This can cause the formation of abnormal hemoglobin molecules which causing anemia. The abnormal hemoglobin molecules present the symptom of the thalassemias. Thalassemia is a quantitative problem of too few globins synthesized, whereas sickle-cell disease (a hemoglobinopathy) is a qualitative problem of synthesis of an incorrectly functioning globin. Thalassemia is usually resulting in underproduction of normal globin proteins, often through mutations in regulatory genes. Hemoglobinopathy means structural abnormalities in the globin proteins themselves. Here, the two conditions may overlap since some conditions that cause abnormalities in globins' proteins (hemoglobinopathy) also affect their thalassemia. Thus, some thalassemias are hemoglobinopathy, but most are not. Either or both of these conditions may cause anemia [5].

1.2 Alpha Thalassemia
The Alpha thalassemia is one of the types of the thalassemia which is caused by mutations in the alpha chain of the hemoglobin molecule. Normally, there are two alpha chain genes located on each #16 chromosome, for a total of 4. The alpha chain is an important component of fetal hemoglobin. The fetal hemoglobin is usually made before birth the birth of the child where as hemoglobin A and hemoglobin A2 are produces after the birth of the child. How these genes are altered determines the specific type of alpha thalassemia in a child. In Alpha thalassemia major affected person all four alpha chain genes are deleted or mutated. This is so severe that death of child can occur in prior to birth of the child. The Alpha thalassemia patients require the intrauterine transfusions are often necessary to carry a baby to term and after birth. All these babies are dependent on red blood cell transfusions. The mother is carrying such kind of baby also under the health risk. Alpha thalassemia carriers are the person whose two alpha chain genes are deleted either both from the same #16 chromosome called a "cis deletion" or one from both #16 chromosomes, called a "trans deletion" .When the parents are carriers of the cis deletion #16 chromosome, then they are one out of the four person .the percentage of the cis deletion alpha thalassemia patients are 25% chance with each pregnancy, to have a baby with alpha thalassemia major. Such types of carriers of the cis deletion versus the Trans deletion can be distinguished by DNA analysis only. DNA testing is usually done from a blood sample to look at the alpha chain genes on each #16 chromosome, to determine which are deleted.

1.3 Beta Thalassemia
The Beta thalassemia is another type of the thalassemia which caused due to defect in the production of beta globin protein from the beta genes. It is the most common cause of beta thalassemia. In the human cell both types of goblin genes are present but fail to produce hemoglobin adequately. If one of the beta globin gene fails, the amount of beta globin the cells is reduced by half. There are hundreds of mutations within the beta globin gene, but approximately 20 different alleles comprise 80% of the mutations found worldwide. If we find out within each geographic population there are unique mutations among the each. Individuals who have beta thalassemia major are usually homozygous for one of the common mutations or heterozygous for one of the common mutations and one of the geographically unique mutations.

But the common things between this both lead to the absence of beta globins chain production. The beta thalassemia syndromes are much more diverse than the alpha thalassemia syndromes due to the diversity of the mutations that produce the defects in the beta globin gene. Unlike the deletions that constitute most of the alpha thalassemia syndromes, beta thalassemia are caused by mutations on chromosome 11 that affect all the aspects of beta globin production, transcription, translation and the stability of the beta globin product. Most hematologists feel there are three general categories of beta thalassemia which are β Thalassemia Trait, β Thalassemia Intermedia and β thalassemia Major. A β Thalassemia Trait person with this condition has one normal gene and one with a mutation. They will usually experience no health problems other than microcytosis and a possible mild anemia that will not respond to iron supplements. This gene mutation can be passed on to an individual's children. Thalassemia Intermedia, in this condition an affected person has two abnormal genes but is still producing some beta globin. The severity of the anemia and health problems experienced depends on the mutations present in that affected person. The dividing line between thalassemia Intermedia and thalassemia major is the degree of anemia and the number and frequency of blood transfusions required to each people. Those with thalassemia Intermedia may need occasional transfusions but do not require them on a regular basis where as the person with Thalassemia major require frequent blood transfusion. Thalassemia Major is also called Cooley's anemia. This is the most severe form of beta thalassemia which is found in the person. The person has two abnormal genes that cause either a severe decrease or complete lack of beta globin production or preventing the production of significant amounts of Hb A. This condition usually appears in an infant after three months of age and causes life-threatening anemia. This particular anemia requires lifelong regular blood transfusions and considerable ongoing medical care. Over time these frequent transfusions of the blood in the patients lead to excessive amounts of iron in the body. This excess amount of iron deposit in the liver, heart and other organs of the patients body and can lead to a premature death from organ failure. Other forms of thalassemia occur when a gene for beta thalassemia is inherited in combination with a gene for a hemoglobin variant [4].

1.4 How is thalassemia inherited?

The two main types of thalassemia are alpha and beta. Both types are inherited in the same manner. Parents who carry the mutated thalassemia gene can pass it on to their child. A child who inherits one mutated gene is considered to be a carrier, which is sometimes called thalassemia trait. Most carriers lead completely normal, healthy lives[4].

2. ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase [3].

2.1 Feed forward neural network with back propogation of error.

The feed forward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers[3].

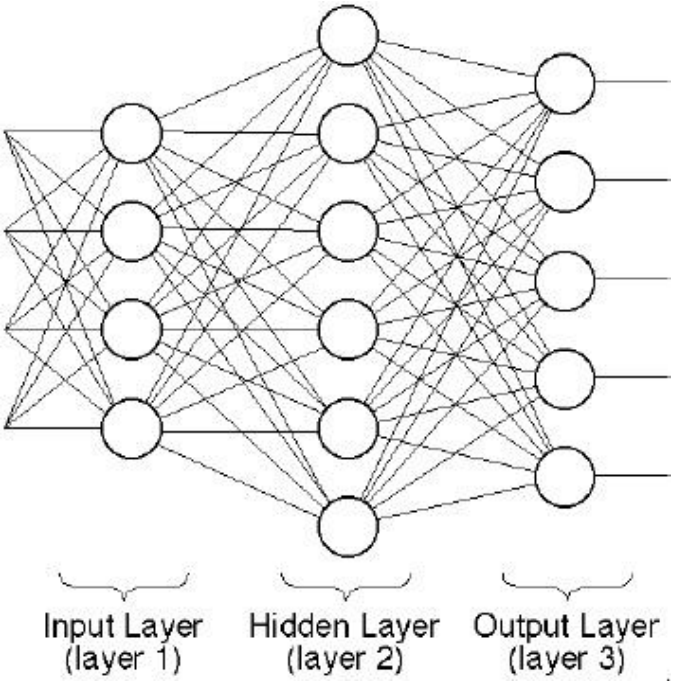

Fig 1: Artificial Neural Network Model

2.2 Feed forward neural network using back propogation of error algorithm.

2.2.1 Decide input, target and testing data.

2.2.2 Initialize the weight and bias.

2.2.3 Calculate the feed forward Neural Network output.

2.2.4 Match the output with target.

2.2.5 Calculate the error= difference between actual & desired output.

2.2.6 Update all the weight and bias of the Neural Network.

2.2.7 Repeat the steps until the error will not reduced.

2.3 Training of artificial neural networks

A neural network has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to 'train' the neural network by feeding it teaching patterns and letting it change its weights according to some learning

rule. Supervised learning or Associative learning in which the network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised) [3].

3. PROPOSED METHOD

The algorithm is developed in MATLAB R2009. It uses a Feed forward algorithm. Algorithm for classification using ANN:

The code builds a classifier that can identify that the particular people are Thalassemia major, minor or normal person. To classify the people having Thalassemia for that we have taken all the test sample in the people and accordingly we have classified them in the class of the Thalassemia major, minor and normal person. If the particular person is major then for them we have assign the numeric value as 1 and for minor -1 and for the normal people we have assigned the value as 0. By presenting previously recorded inputs to a neural network and then tuning it to produce the desired target outputs. This process is called neural network training.

3.1 Input data

The blood sample of any adult person consist of primarily hemoglobin A (HbA), a small percentage of hemoglobin A2 (HbA2), and trace amounts of fetal hemoglobin (HbF).The carrier of β thalassemia have levels of hemoglobin A2 and F which can be greater than 3.5% and 2% of the total hemoglobin respectively. Determination of concurrently elevated levels of hemoglobin's A2 and F in the blood sample of the person has becomes the most practical means to diagnose the carries of the β thalassemia gene .Methods for the quantization of hemoglobin A2 include electrophoresis and anion exchange column chromatography. Methods for the quantization of hemoglobin F include electrophoresis, alkali denaturation and radial Immune diffusion (RID).High performance liquid chromatography (HPLC), which can be a relatively fast and reproducible methods, has been used for the determination of various hemoglobin's including hemoglobin's A2 and F. The variant is a fully automated HPLC system which can be used to separate and determine area percentage for hemoglobin's A2 and F and to provide qualitative determinations of abnormal hemoglobin's .the most commonly occurring hemoglobin variants include hemoglobin's D, S, C and E.

The range of different Hb variants of β – thalassemia major subjects reported in present study can be seen is as follows:

Table 1.β- Thalassemia major Hb variant

β- thalassemia major (HbA ↓+ HbF ↑ + HbA2)		
Hb variant	**Range %**	**Normal %**
HbA variant	2.0-9.0	>96.0
HbF variant	88.9-99.4	<1.0
HbA2 variant	1.2-4.1	<3.5

Similarly, the range of different Hb variants of β – thalassemia minor subjects reported in present study can be seen is as follows:

Table 2.β-Thalassemia minor Hb variant

β- thalassemia minor (HbA↓+ HbF + HbA2↑)		
Hb variant	**Range %**	**Normal %**
HbA variant	88.1-95.8	>96.0
HbF variant	0.0-1.3	<1.0
HbA2 variant	3.6-6.2	<3.5

Table 3.Candidate Feature for Diagnosis

Feature	**Description**
1	Age
2	Sex
3	Hb gm%
4	P3(Hb variant)
5	A0(Hb variant)
6	F(Hb variant)
7	A2(Hb variant)
8	S(Hb variant)

3.2 The neural network classifier

The next step is to create a neural network (feed forward back propagation network) that will learn to identify the classes. It will differ slightly every time it is run. The random seed or twister is set to avoid this randomness. A 1-hidden layer feed since the neural network starts with random initial weights, the results forward network is created with 11 neurons in the hidden layer[2].

Now the network is ready to be trained. The samples divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy.

Table 4.Parameter setting for Feed forward Neural Network

Parameter	**Values**
Number of input node	8
Number of output node	1
Number of hidden layer	1
Number of neurons in hidden layer	11
Learning rate	0.02
Number of epoch	39

3.3 Testing the classifier

3.3.1 The trained neural network can now be tested with the testing samples. This will give us a sense of how well the network will do when applied to data from the real world. If require the testing can be done with a separate testing set which is created while creating training set.

3.3.2 The network response can now be compared against the desired target response to build the classification matrix.

3.4 Data acquired

Here in this particular work the data collected from the hospital where they are doing treatment of the Thalassemia patients. So we are very thanking full to hospital and all those patients those who have given the personal information for my work.

4. RESULT

The particular Neural Network trained with the help of 100 known input data of Thalassemia patients where we have result for all those data. Now we have tested the Neural Network with the help of 39 data of Thalassemia patients where we do not having the result for that and we got the following observation and the result:

Table 5.Accuracy Assessment of ANN

	Class Type determined from Reference				
Class types determined from classified Data	**#Plot**	**Minor**	**Normal**	**Major**	**Total**
	Minor	14	5	1	20
	Normal	1	13	2	16
	Major	0	0	3	03
	Total	15	18	6	39

Accuracy = ((14+13+3)/39)*100

=76.92%

Diagonal Represent sites classified correctly according to reference data. Off-diagonal were misclassified.

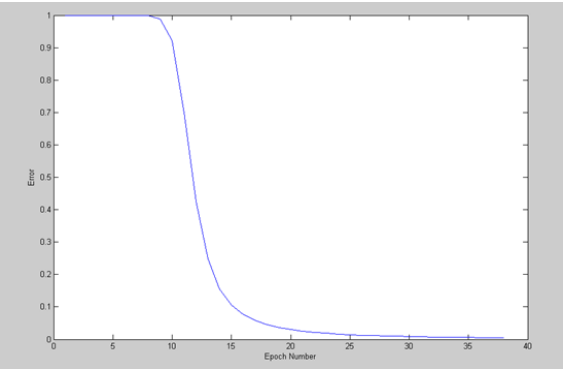

Fig 2: Error Reduction of ANN

CONCLUSION

The Thalassemia major is the very dangerous anemia inside the human and the condition of the thalassemia major patients is very bad they required regular blood transfusion. Transfusions of red blood cells are the main treatment for people who have moderate or severe thalassemia. This treatment gives you healthy red blood cells with normal hemoglobin. During a blood transfusion, a needle is used to insert an intravenous (IV) line into one of your blood vessels. Through this line, you receive healthy blood. The procedure usually takes 1 to 4 hours. Red blood cells live only for about 120 days. So, you may need repeated transfusions to maintain a supply of healthy red blood cells. Therefore the early warning to the Thalassemia minor will avoid the generation of the Thalassemia in the society by avoiding the thalassemia minor female and minor male marriage.

5. FUTURE WORK

There are the numbers of the parameter based on that we can classify the Thalassemia patients. There are the physical as well as the chemical symptom's are there in the patients. By doing more study on the thalassemia we can list more symptoms and those entire symptoms given as the input to the Neural Network and we can classify the patients much and more accurate. Other than Artificial Neural Network we can use other data mining tool e.g. Genetic Algorithm.

8. ACKNOWLEDGMENTS

Apart from the efforts of me, the success of project depends largely on the encouragement and guidelines of many others. I give this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project. I would like to show my greatest appreciation to my internal guide Prof. (Mrs.) Varsha Turkar and Head of Department Prof (Mr.) Santosh Singh, Thakur College of Science and Commerce, Mumbai-400101. I can't say thank you enough for the tremendous support and help. I feel motivated and encouraged every time I met her for weekly reporting. Without her encouragement and guidance this project would not have materialized.

Collectively and individual acknowledgements go to my partner, colleagues from IT Department division for their technical help. I would like to thank my family members and everybody who was important to the successful realization of project, as well as expressing my apology that I could not mention personally one by one.

9. REFERENCES

[1] Geiman, R. M. and W. L. Nolte, "An Expert System for Learning Disability Diagnosis", Proceedings of the IEEE Conference on System Engineering, pp. 363-366 (Aug. 1990).

[2] Hofmeister, A. M. and J. M., Ferrara, Artificial Intelligence Applications in Special Education: How Feasible? Final Report, Logan, Utah State U. (1986).

[3] V Turkar, Polari metric SAR Image Classification by Using Artificial Neural Network.

[4] Puthenveetil G, Scholes J, Carbonell D, Qureshi N, Xia P, Zeng L, Li S, Yu Y, Hiti AL, Yee JK, Malik P: Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector. Blood 2004, 104:3445-3453.

[5] Borgna-Pignatti C, Galanello R"Thalassemia and related disorders: quantitative disorders of hemoglobin synthesis. In Win robe's Clinical Hematology" Volume 42, 11th edition.