Journal of Advanced Research 24 (2020) 183–190

Contents lists available at ScienceDirect

Journal of Advanced Research

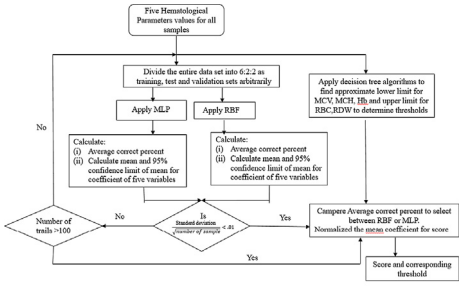
journal homepage: www.elsevier.com/locate/jare

A decision support scheme for beta thalassemia and HbE carrier screening

Reena Das [a], Saikat Datta [b], Anilava Kaviraj [c], Soumendra Nath Sanyal [d], Peter Nielsen [d], Izabela Nielsen [d], Prashant Sharma [a], Tanmay Sanyal [e], Kartick Dey [f], Subrata Saha [d,*]

[a] Department of Hematology, Postgraduate Institute of Medical Education and Research, Chandigarh 160012, India
[b] Department of Clinical Hematology, Anandaloke Hospital, Siliguri 734001, India
[c] Department of Zoology, University of Kalyani, Kalyani 741235, India
[d] Department of Materials and Production, Aalborg University, DK 9220 Aalborg, Denmark
[e] Department of Zoology, Krishnagar Government College, Krishnagar 741101, India
[f] Department of Mathematics, University of Engineering & Management, Kolkata 700160, India

GRAPHICAL ABSTRACT



Computational Scheme

ARTICLE INFO

Article history:
Received 21 March 2020
Revised 6 April 2020
Accepted 11 April 2020

Keywords:
Thalassemia carrier screening
Artificial neural networks
Decision trees

ABSTRACT

The most effective way to combat β-thalassemias is to prevent the birth of children with thalassemia major. Therefore, a cost-effective screening method is essential to identify β-thalassemia traits (BTT) and differentiate normal individuals from carriers. We considered five hematological parameters to formulate two separate scoring mechanisms, one for BTT detection, and another for joint determination of hemoglobin E (HbE) trait and BTT by employing decision trees, Naïve Bayes classifier, and Artificial neural network frameworks on data collected from the Postgraduate Institute of Medical Education and Research, Chandigarh, India. We validated both the scores on two different data sets and found 100% sensitivity of both the scores with their respective threshold values. The results revealed the specificity of the screening scores to be 79.25% and 91.74% for BTT and 58.62% and 78.03% for the joint score of HbE and BTT, respectively. A lower Youden's index was measured for the two scores compared to some existing indices. Therefore, the proposed scores can obviate a large portion of the population from expensive high-performance liquid chromatography (HPLC) analysis during the screening of BTT, and joint determination of BTT and HbE, respectively, thereby saving significant resources and cost currently being utilized for screening purpose.

Peer review under responsibility of Cairo University.

* Corresponding author.
E-mail addresses: subrata.scm@gmail.com, saha@m-tech.aau.dk (S. Saha).

https://doi.org/10.1016/j.jare.2020.04.005
2090-1232/

Introduction

Hemoglobinopathies are a group of inherited hemoglobin (Hb) disorders with abnormal production or structure of the globin molecule due to mutations of globin genes. According to the World Health Organization (WHO) and the thalassemia International Federation, every year over 330,000 babies are added worldwide with Hb disorders. WHO has reported that hemoglobinopathies as a growing health problem in most of the countries [1,2]. The approximate rate of heterozygosity is 13% in Africa, 4% in Asia, and 2% in the United States [2]. In India alone, the estimated number of persons with hemoglobinopathies is 25 million [3]. Although most of the inherited hemoglobin disorders originated from Southeast Asian, Indian, Mediterranean, and Middle-Eastern ethnic groups, currently the entire world is at risk of these disorders due to the large-scale migration [4–7]. Among the various hemoglobin disorders, symptomatic beta-thalassemia is considered to be the commonest autosomal recessive disease worldwide, with 1–5% of the world population being beta-thalassemia trait's (BTTs) [8,9]. Patients with hemoglobin E (HbE) traits and BTT interactions have a significant contribution to morbidity and mortality in India, Bangladesh and Myanmar [10–13]. Though different types of hemoglobinopathies are encountered in India, HbE is most frequently found in the north-eastern regions of India [14,15]. Homozygous α^0-thalassemia causing Hb Bart's hydrops fetalis, homozygous beta-thalassemia, and beta-thalassemia/HbE are important ones, which require attention for prevention and control measures in South East Asia [16]. However, hydrops fetalis due to homozygous alpha zero genotypes are rare and not clinically significant in India [17].

Detection of carriers by screening program is considered to be the most effective way to control symptomatic Hb disorders. The objective of screening programs is to detect potential health risks for themselves or their offspring [18,19]. There is no consensus on the most suitable method for performing such screening programs due to social, cultural, and religious stigma [20–22]. It is important to choose cost-effective and evidence-based approaches for the screening of hemoglobinopathies [23]. In India, the average estimated cost of preventing the birth of 10,000 patients every year by the screening of antenatal women is approximately $90 million. In contrast, the cost of treating these 10,000 patients over an estimated lifespan of 40 years is $975 million and the annual estimated beta-thalassemia major management cost per patient is $2400–3500 [24]. Thus, the cost of prevention is only one-tenth of the treatment costs [25]. Sinha et al [26] predicted that by 2026, the estimated amount of annual blood required for the treatment of Hb disorders in India would increase to 9.24 million units, together with an 86% increase in budgetary requirements which would then account for over 19% of the current National Health Budget, which is alarming. According to Colah and Gorakshakar [27]; Khera et al [28], most initial screening are based on red cell indices, and then samples are subjected to the relatively expensive high-performance liquid chromatography (HPLC) technique [29]. However, the similarity of red cell indices between beta thalassemia trait and iron deficiency can confuse the screening due to low mean corpuscular volume (MCV) and mean corpuscular hemoglobin (MCH) [30]. If the thalassemia screening test is performed for all the individuals having low MCV and MCH, it will cause an over-utilization of expensive HPLC mechanism and will add to the burden of health expenditure.

Nowadays, predictive data mining is extensively used to discover patterns of clinical observation from the perspective of medical diagnosis [31,32]. The researchers successfully employed various techniques such as support vector machine [33], multilayer perceptron (MLP) [34,35] radial basis function (RBF) [35], feed-forward neural network [35], adaptive network-based fuzzy inference system [36], ANN with wavelet transformation [37], fuzzy support vector machine [38], Naïve Bayes (NB) classifier [39], etc to analyze a real-life complex problem and proposed several frameworks in different contexts [40,41]. On the other hand, researchers have developed several optimization techniques such as gradient descent, genetic algorithm [42], dolphin swarm algorithm [43,44], particle swarm optimization techniques [45], Yin-Yang firefly algorithm [46] to optimize a highly complex data analysis framework. However, instead of developing a new algorithm, we focused on some standard techniques to propose a data analytics framework for BTT and HbE screening in this study. In this direction, Amendolia et al [47] investigated the feasibility of two well-known pattern recognition techniques for beta-thalassemia screening. The authors compared the support vector machine and K-nearest neighbor with an MLP. Setsirichoket et al. [48], applied the C4.5 decision tree, NB classifier, and MLP method for thalassemia screening. They concluded that the NB classifier and MLP could efficiently categorize instances. Jahangiri et al. [49] proposed a tree-based method for the differential screening of BTT and iron deficiency anemia (IDA). The authors used a Chi-squared automatic interaction detector (CHAID); an Exhaustive Chi-squared automatic interaction detector; Quick, unbiased, efficient statistical tree (QUEST); and Generalized, unbiased, interaction detection and estimation (GUIDE) for differentiating diagnosis processes between BTT and IDA.

In a thalassemia screening program, a heterogeneous set of samples containing various types of hemoglobinopathies is expected. Therefore, creating a distinction between IDA and BTT, which is the main focus in the existing literature, may not fully serve the cost and resource-saving objective for any government or private organization, especially in a highly populated country like India. Moreover, to the best of our knowledge, the scoring mechanism for the joint determination of BTT and HbE is scanty. The objective of this study is strictly to identify BTT or HbE, even if a small fraction of normal individuals is recommended for further evaluation of the HPLC. And, if a scoring mechanism can provide such assurance, then it can serve as a tangible cost-saving tool for medical practitioners and organizations so that the majority of the population can be competently excluded from performing expensive HPLC approach during a carrier screening program. We used NB classifier, decision trees and employed the simulation of ANN model to develop two robust scoring mechanisms based on the combined impact of routine hematological parameters (MCV, MCH, Red blood cell distribution width (RDW), red blood corpuscles (RBC), and Hb), those can be measured economically through Automated hematology analyzers. It has been documented that several researchers have used some of these five parameters, such as Lafferty et al. [50] and Jiang et al. [51] used only MCH, whereas Old et al. [52] used MCH and MCV, but to make the scoring mechanism robust, we considerd five parameters simultaneously. We compared our results with the existing screening indices such as Mentzer [53], Srivastava [54], Shine & Lal [55], and found the proposed scoring mechanisms have higher sensitivity and lower positive prognostic values. Both the scores proposed in this study were found capable to identify BTT and HbE carriers individually from non-carrier individuals with 100% sensitivities.

Material and methods

Collection of data and diagnostic criteria

Clinical data were collected from the Department of Hematology at the Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh, India, where routine diagnosis for thalassemia and hemoglobinopathies are performed. The data

set consisted of 1076 samples (387 normal individuals, 104 HbE, 293 BTT, 135 IDA, and 157 IDA with BTT). We named it the test data set. We performed the entire data analysis and derived two scoring mechanism separately.

For validation purpose of the proposed scoring scheme, a data set consisting of 252 samples (174 normal individuals, 58 BTT, 14 HbE, 2 Thalassemia major, 1 Thalassemia intermedia, 1 Sickle cell trait, 1 Double heterozygote for HbS and BTT, and 1 Double heterozygote for Hemoglobin D disease (HbD) Punjab and BTT) was also collected from PGIMER, India. We named it the validation data set. The laboratory at PGIMER is under the United Kingdom National External Quality Assessment Service (UK NEQAS) Hematology program.

Besides, a field data set consisting of 240 samples (214 normal individuals, 10 BTT, and 16 HbE) were collected from carrier screening program conducted at Ranaghat, West Bengal, India by the Auxiliary unit of State Thalassemia Control Programmed (STCP), Department of Health and Family Welfare, the Government of West Bengal, India to crosscheck the efficiency of the proposed scores.

The ethical justification was not taken for this data set as only retrospective evaluation of the automated red cell indices was carried out. No additional samples were taken or tests were performed on the samples.

Basic statistical analysis

Statistical analysis of this study was conducted using SPSS 25 (www.ibm.com). We measured preliminary descriptive statistical analysis for the test data set to obtain a generalized overview regarding the relation between the hematological parameters considered in this study.

Score construction

In this study, we employed an NB classifier [48], Decision trees, and ANN framework to derive the scoring schemes. A brief description of each method is presented in the supplementary file. Note that MATLAB 2019a (www.mathworks.com) was used for further analysis. The software was availed through Aalborg University, Denmark. An overview of the computational scheme employed to generate scores is presented in Fig. 1.

We employed both the MLP and RBF techniques to identify the average correct percentage of the classified instances. Simultaneously, we employed decision tree methods to identify the threshold of five parameters. Finally, the outcomes of ANN frameworks and Decision tree methods were jointly used to formulate the equations representing scores for screening. The stepwise explanation for the data analysis scheme is presented in the next section.

Results and discussion

The objective of this study was to rule out non-carrier individuals as much as possible by using a single-cost effective test before initiating the HPLC test. First, we performed preliminary descriptive statistical analysis for five parameters MCV, MCH, RDW, RBC, and Hb; and results are presented in the Supplementary file (Tables S1–S5). Besides, normal ranges of values for five hematological parameters are also presented in the Supplementary file (Table S6). It was observed from descriptive statistical analysis that the mean and median values of Hb, MCV and MCH were higher for the normal individuals compared to BTT and HbE traits. However, the reverse trend was observed for RDW and RBC.

For precise identification of parameters responsible for the identification of BTT and HbE carriers, C4.5 and NB classifiers were employed. Note that IDA samples are considered as normal individual during the process so that the score can be applied in practice for BTT screening purposes in a heterogeneous environment. We separated the data sets into two groups. The first group was used to obtain the significance of the critical parameters accountable for BTT only, whereas the second group was used for BTT and HbE traits, jointly. The results for C4.5 and NB classifiers are presented in Table 1 below.

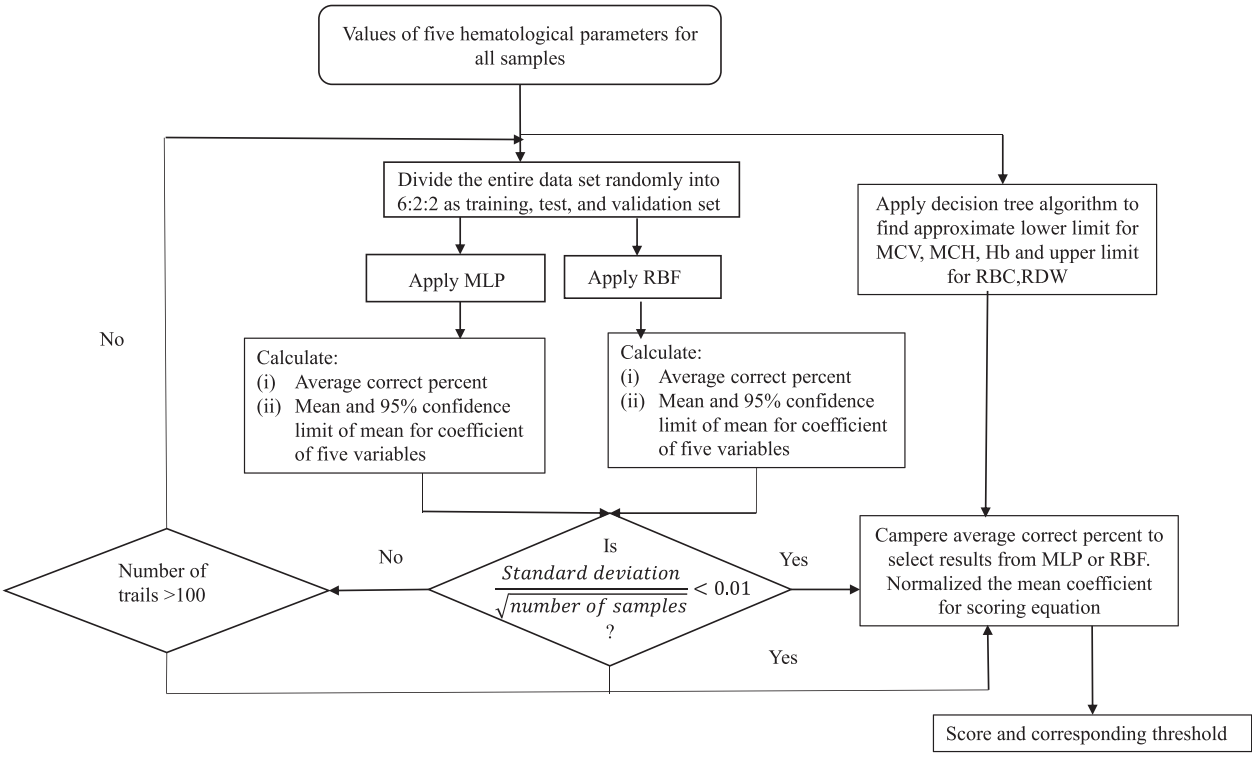


Fig. 1. Data analysis scheme used for developing scores.

Table 1
Correctly classified instances and error details for the C4.5 and NB classifier.

Scenarios	Classifier	Correctly classified instances (%)	Kappa statistics	MAE	RMSE	RAE (%)	RRSE (%)	Precision of NB Classifier
BTT test data set (387 Normal + 293 BTT + 157 IDA & BTT + 135 IDA)	C 4.5	95.27	0.90	0.06	0.21	12.21	41.71	
	NB	93.83	0.87	0.07	0.22	14.48	43.60	RDW-0.17 MCV-0.16 MCH-0.15
HbE and BTT test data set (387 normal +104 HbE + 293 BTT + 157 IDA & BTT + 135 IDA)	C 4.5	90.09	0.61	0.13	0.30	47.54	80.60	
	NB	85.95	0.40	0.18	0.30	66.09	82.11	MCV-0.19 RDW-0.18 MCH-0.15

From Table 1, the following results were obtained:

- MCH, MCV, and RDW appeared to be indicative parameters in C4.5 and NB classifiers
- A higher Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Root Relative Squared Error (RRSE) and lower value of Kappa Statistics were measured for joint HbE and BTT test data set compared to BTT test data set.

Although the NB classifier and C4.5 algorithm are extensively used to analyze clinical data [33], it was however observed, that correctly classified instances were less and RRSE were too high when it was applied for the joint determination of BTT and HbE traits due to the heterogeneous nature of data set. Therefore, we executed both the MLP and RBF techniques which are more robust. In both methods, it was necessary to divide the data set randomly into three sub sets, namely training, test, and holdout sets. The training data were used to find the weights and build the ANN model. The test data set was used to find errors and prevent overtraining. The holdout/validation data were used for the validation of the outcomes. Consequently, the rule of arbitrary division created a significant impact on the calculation of normalized importance for each parameter. To scale down this effect, we executed the simulation process 100 times to measure the average values of normalized importance for each parameter. During the simulation experiment, we observed that the average values of the normalized importance percentage for each parameter were nearly converged with the increasing number of iterations. Based on the data analysis scheme presented in Fig. 1, the stepwise details of scoring mechanism developed for the joint determination of BTT and HbE, we named it *SCS_HbE &BTT,* as presented below:

Step 1: We applied MLP and RBF on the test data set by dividing it randomly into 6:2:2, where five hematological parameters are considered as an independent variable to build ANN model. After 100 iterations the results for a mean of coefficients of relative importance of five hematological parameters are obtained as follows in Table 2:

The results demonstrate the followings:

- The average accuracy of MLP and RBF methodologies was reasonably high compare to the NB classifier and C4.5 algorithm
- MCV and MCH are the most important parameters among the five

Table 2 demonstrates that the average correct percent, in MLP is higher compared to RBF. Consequently, the normalized importance of MLP was used in benchmark scoring for *SCS_HbE&BTT*. Note that the impact of each independent variable can be evaluated in an ANN model by relative importance factors. Therefore, MCV is a major determinant in the perspective of model predictive power compared to the other four.

Table 2
Mean of coefficients of relative importance factors of five hematological parameters.

Hematological Parameters	*SCS_HbE&BTT* score	
	RBF	MLP
Hb	0.4224	0.6222
RDW	0.2351	0.5351
MCH	0.6852	0.5459
MCV	0.9103	0.9103
RBC	0.5077	0.5077
Average correct percentage of prediction = (correct percent of the training set, test set, and holdout set)/3	93.76	95.24

Step 1.1: Determine the approximate value of the threshold to identify the cut-off value for each parameter through decision tree analysis. Note that we focused on the classification to find some pure nodes which are not necessarily to be the immediate leaf and used extensive pruning to identify all the cut-off values. Then, the concept of supremum and infimum values were used to set joint cut-off values that were integrated with normalized importance obtained from MLP. For example, it is found that the influence MCV, MCH and Hb are increasing whereas RDW and RBC are decreasing. Therefore, to determine the threshold cut-off value, the infimum of the first three parameters and supremum of the last two are used to find the threshold for each score. This strict substitution can ensure that all traits are included even if some additional normal individuals are also included in the process for separation.

Step 2: Calculate mean and 95% confidence limit of the mean $\left(\text{mean} \pm 1.96 \times \frac{Standard\ \ deviation}{\sqrt{number\ \ of\ \ sample}}\right)$ for the relative importance coefficient of each parameter. We use the threshold $\frac{Standard\ \ deviation}{\sqrt{number\ \ of\ \ sample}}$ to minimize errors. The upper and lower real limits of the class intervals can be obtained from Table 3.

Note that in most of the existing studies, researchers focused to determine the confidence interval for screening purposes [56]. However, from the perspective of the implementation issue, instead of the interval, it may be easier for the user as well as from the perspective of devise management to use the exact value of the thresholds.

Step 3: By normalizing the coefficient of relative importance factors, we formulate the equation for each score and obtain threshold by substituting the cut-off values from decision tree analysis.

To develop a scoring mechanism for the joint determination of BTT and HbE traits for improving the effectiveness of the screening program, we developed the SCS_*HbE&BTT* score. Based on the normalized importance of MLP, the following scoring mechanism is proposed:

$$SCS\_HbE\&BTT = 0.2916\text{MCV} + 0.1749\text{MCH} - 0.1626\text{RBC} - 0.1714\text{RDW} + 0.1994\text{Hb} \quad (1)$$

Table 3
Mean, S.E., median, and 95% confidence level (CL) of the coefficient of five parameters.

	Hb	RDW	MCH	MCV	RBC
Mean	0.6222	0.5351	0.5459	0.9103	0.5077
Standard Error	0.0457	0.0345	0.0388	0.0255	0.0334
Median	0.7340	0.4999	0.5358	1	0.5249
CL (95.0%)	0.0915	0.0692	0.0776	0.0511	0.0669

From the decision tree analysis, the combined cut-off values were obtained as Hb $\leq$ 11.9, RBC $\geq$ 3.78, MCH $\leq$ 27.7, MCV $\leq$ 85.4, and RDW $\geq$ 12.75. These cut-off values are used in Eq. (1) to obtain the threshold value which is 29.323. Therefore, if the score for a particular sample is greater than 29.323, then that sample can be excluded from further HPLC test.

The SCS_*HbE&BTT* score was applied to two validation sets. 72 samples out of 174 normal individuals were recommended for further HPLC test, the false positive rate for SCS_*HbE&BTT* score was 41.37%. In the second field data set, 47 samples out of 214 normal samples are necessary to recommend for further HPLC test, i.e. the false positive rate is only 21.96%. Most importantly, the score can perfectly determine all the carriers of HbE and BTT for both the data sets.

Similarly, to determine cut-off values precisely for BTT carrier detection, we employed all five types of decision tree methods and drew decision trees for all possible consequences by considering 387 normal, 293 BTT, 157 IDA with BTT and 135 IDA samples. From the analysis of those trees, the combined cut-off values were obtained as Hb$\leq$ 11.7, RBC $\geq$ 4.34, MCH $\leq$ 25.15, MCV $\leq$ 78.75, and RDW $\geq$ 12.25. Note that the cut-off values reported from decision tree analysis in the existing literature from the perspective of BTT screening are summarized in Table 4.

One can observe that five parameters were not simultaneously measured and prioritized in the perspective of BTT screening, and cut-off values identified in the present study are similar to some of the previous studies mentioned in Table 4. Therefore, based on the normalized importance of MLP, the following scoring mechanism is proposed for BTT screening:

$$\text{SCS\_}BTT = 0.2815\text{MCV} + 0.2015\text{MCH} - 0.2641\text{RBC} - 0.1693\text{RDW} + 0.0835\text{Hb} \quad (2)$$

We applied the SCS_*BTT* score on validation data sets. During the validation process, we considered 14 HbE samples as normal samples because SCS_BTT is developed to detect BTT only. Therefore, we had a total of 188 normal samples and the following was found:

- No need for further HPLC if the score of a subject is greater than 24.993
- 39 samples out of 188 normal samples are necessary to recommend for further HPLC tests, i.e. the false positive rate is 20.74%.
- Most importantly, the score can predict all the BTTs, i.e., all the subjects with BTT have a score below 24.993.
- 10 samples out of 14 HbE samples were also recommended for further HPLC tests, although we consider all these 14 samples as normal samples.

The SCS_BTT was also validated on the field data set also. Similarly to the first validation data set 16 HbE samples were considered as normal samples and we have 230 normal samples and the followings were found:

- 19 samples out of 230 normal individuals are necessary to recommend for further HPLC tests, i.e. the false positive rate is 8.26%.
- The score can detect all the samples with BTT because all the BTT samples were having a score of less than 24.993.
- 11 samples out of 16 HbE samples were also recommended for further HPLC tests, although we consider all these 16 samples as normal samples.

To validate the scalability of the above two scoring mechanisms, we compared our results with some commonly practiced indexing mechanisms which are given in Table 5.

Note that, sensitivity (SENS) = $\frac{TP}{TP+FN}$, specificity (SPEC) = $\frac{TN}{TN+FP}$, positive prognostic value (PPV) = $\frac{TP}{TP+FP}$, negative prognostic value (NPV) = $\frac{TN}{TN+FN}$, efficiency (EFF) = $\frac{TP+TN}{TP+TN+FP+FN}$, and Youden's index (YI) = SENS + SPEC $-$ 100, where TP, FP, TN, and FN represents the true positive, false positive, true negative, and false negative, respectively. These measures are used for comparison purposes. Table 5 demonstrates that several indices have been proposed for thalassemia carrier screening, but none has yet been proved to be satisfactory [8]. Therefore, it was necessary to create a robust scoring mechanism. In this study, we considered the joint impact of MCV, MCH, RDW, RBC, and Hb in a single formula. We observed the normalized importance of each of the five parameters is not negligible in Eqs. (1) and (2). This is the reason for obtaining higher Youden's index value as measured in this study for the SCS_*BTT*, compared to other indices mainly developed for BTT screening. The negative prognostic value indicates that the SCS_*BTT* is robust from the perspective of carrier identification without excluding the BTTs.

The decision-support scheme for the application software is presented in Fig. 2., which can be easily implemented on different gadgets like mobile, tablet, phablet, etc. or devices that can imitate intelligent human behavior for ease of application.

Fig. 2 provides a schematic representation of the decision support scheme that can be used for screening purposes. Based on the information of five hematological parameters, a practitioner can use it for the identification of both the BTT and HbE in a screening program.

Over the past three decades, many discriminant formulae have been developed by several researchers, primarily to differentiating thalassemia carriers from patients with IDA [61,62]. Most of them

Table 4
Cut-off values for hematological parameters in some existing literature.

Parameters	Lafferty et al. [50]	Jiang et al. [51]	Old et al. [52]	Rathod et al. [57]	Sahli et al. [58]	Cao et al. [59]	Plengsuree et al. [60]
MCH (picogram)	–	–	<27	<27	<23	<27	<27
MCV (femtoliters)	<72	<80	<79	<76.5	<75	<78	<76
RBC (million/microliter)	–	–	–	>5	>5	–	>5
RDW %	–	–	–	>13.6	>14	–	>14

Table 5
Comparative outcomes of proposed scoring mechanisms with existing indices.

Index	Formula	BTT	Sensitivity	Specificity	PPV	NPV	Efficiency	Youden's Index
Mentzer [53]	$\frac{MCV}{RBC}$	<13	70.31	96.28	86.54	90.50	89.68	66.59
Srivastava [54]	$\frac{MCH}{RBC}$	<3.8	62.50	97.34	88.89	88.40	88.49	59.84
Shine & Lal [55]	$\frac{MCV^2 \times MCH}{100}$	<1530	95.31	79.79	61.62	98.04	83.73	75.10
Jayabose et al.[61]	$\frac{MCV \times RDW}{RBC}$	<220	64.06	90.96	70.69	88.14	84.13	55.02
Sirdah et al. [62]	MCV − RBC − 3Hb	<27	64.06	97.34	89.13	88.83	88.89	61.40
Ehsani et al. [63]	MCV − 10RBC	<15	68.75	96.81	88	90.10	89.68	65.56
SCS_*BTT(PGIMER)*	Eq. (2)	≤24.99	100	79.25	62.13	100	84.52	79.25
SCS_*HbE&BTT(PGIMER)*	Eq. (1)	≤29.323	100	58.62	52	100	71.43	58.62
SCS_*BTT(STCP)*	Eq. (2)	≤24.99	100	91.74	34.48	100	92.08	91.74
SCS_*HbE&BTT(STCP)*	Eq. (1)	≤29.323	100	78.04	35.62	100	80.42	78.04

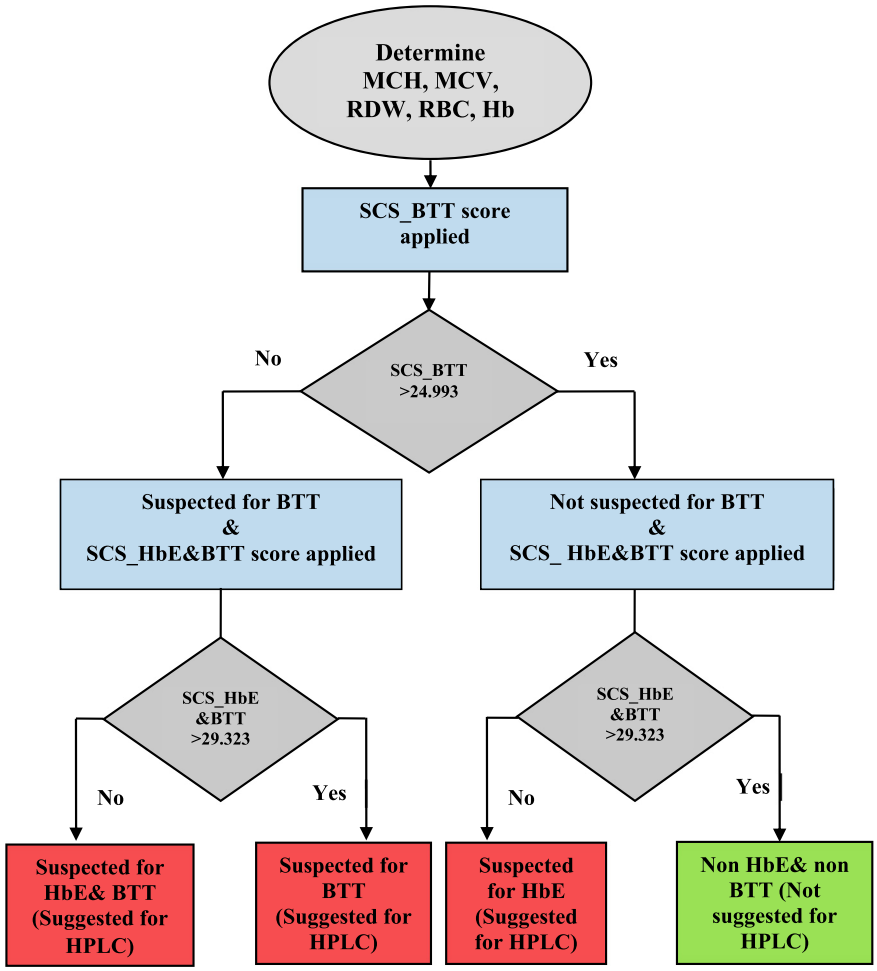


Fig. 2. Decision support scheme for SUSOKA application.

use various combinations of five hematological parameters, but not all [64]. Sometimes these formulae fail to validate the results in some scenarios such as if a sample characterizes thalassemia carriers with concomitant severe IDA. However, initial indications of thalassemia carrier remain important for the practitioners, mainly in countries with limited health-care resources [26,48]. Therefore, the development of a diagnostically useful discriminant formula or scoring mechanism is a priority research direction. It is always challenging to bear accumulated expenses for undertaking BTT screening programs for any organization, especially for government health systems in low- and middle-income countries. Although low values of hematological parameters such as MCV and MCH are generally considered as an indication of BTTs, one subsequently needs to perform HPLC for the quantization of HbA2, HbF and other variants of Hb [65]. At PGIMER, Chandigarh, India the cut-off of ≥4% HbA2 is used to be definite BTTs and values between 3.6 and 3.9% as borderline carriers. The borderline HbA2 cases are advised screening for the partners either after marriage or as a pre-marital screening. In a case where HbE trait is being considered on HPLC, Hb electrophoresis at alkaline pH of 8.6 is performed where the HbA2 cosegregate with HbE. In screening programs, it is envisaged that some additional cases of IDA will also be picked up for performing HPLC. Refereeing all the subjects with reduced MCV and/or MCH for performing HPLC may cause an overutilization of the costly mechanism and corresponding resources. Moreover, existing indices fail to differentiate carriers and

non-carriers perfectly as shown in Table 3. Normalized relative importance obtained from MLP or RBF techniques demonstrates that it is difficult to ignore the impact of parameters such as RDW or Hb. However, most of the indices did not consider the mutual impact of all these parameters. Table 5 demonstrates a higher Youden's Index for both the score, which indicates that the score can be applicable for initial screening purposes effectively compared to some of the existing indices. Although there exists a small number of false-positive results, higher sensitivity for both the scores can lead to satisfactory screening tools.

Conclusion

Hemoglobinopathies are a blood disorder associated with the production of hemoglobin that carries oxygen to cells throughout the body. BTT and HbE are two commonly found variants that may cause abnormal blood clots, pale skin, weakness, enlarged liver, fatigue, and more serious complications. Routine carrier screening is extensively used in this regard. In this study, a novel decision support system was proposed based on the combined impact of MCV, MCH, RDW, Hb, and RBC. The idea is not to miss, even we end up studying more cases for HPLC which may turn out to have a normal HPLC pattern. The false-positive rates of the proposed scoring mechanisms were found to be 20.74% and 41.37%, respectively for validation data set. Most importantly, the scores can predict the true positive rate perfectly. Therefore, a large portion of the population can be excluded at the initial stages of the carrier screening program, which leads to substantial savings in health expenditure. The parameters considered for scoring purposes are determined with a blood test at a reasonable expense. For example, one may solely perform CBC tests at the primary stage of a thalassemia screening program and effectively use the proposed scoring indices. Presently, the HPLC test is at least 10–15 times costlier than the CBC test throughout India [66]. Therefore, the proposed scores can be supportive of the government organization by saving significant expense on thalassemia screening programs and reducing the over utilization of resources.

An application software SUSOKA will be developed for screening purposes after validation of proposed scores for mass utilization. The data analysis framework may also be employed for the identification of disorders such as HbD Punjab trait, HbS trait and other similar variants [67,68]. For any given method with 100% sensitivity may be more theoretical, but it happens due to the impact of supremum and infimum measure considered in the scoring process, but it should be noted that a percentage of normal individuals are also recommended for HPLC, consequently how to reduce false-positive rate would be the next challenge. It should be noted that the normal range of hematological parameters can change country wise, however, by using the model the threshold values can be modified. Although the proposed scoring mechanisms provide us an opportunity to differentiate two major variants of hemoglobinopathies, it needs to be validated with heterogeneous data set collected from various countries for unification and implementation.

Declaration of Competing Interest

The authors declare no competing interests.

R.D. and P.S. had full access to all the data collected from PGIMER, Chandigarh, India. T.S. had full access to all the data collected from STCP, Ranaghat, Department of Health and Family Welfare, Government of West Bengal, India. S.S., S.N.S. took responsibility for the accuracy of the data analysis. S.D. and R.D. acted as a medical mentor and designed the study conception. R.D. and P.S. oversaw laboratory testing and data collection at the PGIMER, Chandigarh, India, and provided critical inputs into the manuscript. S.D., A.K., P.N. and I.N. reviewed the manuscript and advised on results interpretation, modification, and approved it for submission. S. N. S, K.D, R.D, A.K. and S.S wrote the manuscript.

Acknowledgment

TS is thankful to the Head of the Institution, Fulia Sikshaniketan, Nadia, West Bengal, for allowing carrier screening programme in the school campus by STCP, Ranaghat.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2020.04.005.

References

[1] Martinez PA, Angastiniotis M, Eleftheriou A, Gulbis B, Pereira MD, Petrova-Benedict R, et al. Haemoglobinopathies in Europe: health & migration policy perspectives. Orphanet J Rare Dis 2014;9(1):97.
[2] Angastiniotis M, Modell B, Englezos P, Boulyjenkov V. Prevention and control of haemoglobinopathies. Bull World Health Organ 1995;73(3):375.
[3] Saxena A, Phadke SR. Thalassaemia control by carrier screening: The Indian scenario. Curr Sci 2002;83(3):291–5.
[4] Aliyeva G, Asadov C, Mammadova T, Gafarova S, Abdulalimov E. Thalassemia in the laboratory: pearls, pitfalls, and promises. Clin Chem Lab Med 2018;57(2):165–74.
[5] Piel FB, Tatem AJ, Huang Z, Gupta S, Williams TN, Weatherall DJ. Global migration and the changing distribution of sickle haemoglobin: a quantitative study of temporal trends between 1960 and 2000. Lancet Global Health 2014;2(2):e80–9.
[6] El Sayed SM, Abou-Taleb A, Mahmoud HS, Baghdadi H, Maria RA, Ahmed NS, et al. Percutaneous excretion of iron and ferritin (through Al-hijamah) as a novel treatment for iron overload in beta-thalassemia major, hemochromatosis and sideroblastic anemia. Med Hypotheses 2014;83(2):238–46.
[7] Kountouris P, Kousiappa I, Papasavva T, Christopoulos G, Pavlou E, Petrou M, et al. The molecular spectrum and distribution of haemoglobinopathies in Cyprus: a 20-year retrospective study. Sci Rep 2016;6:26371.
[8] Brancaleoni V, Di Pierro E, Motta I, Cappellini MD. Laboratory diagnosis of thalassemia. Int J Lab Hematol 2016;38:32–40.
[9] Vallance H, Ford J. Carrier testing for autosomal-recessive disorders. Crit Rev Clin Lab Sci 2003;40(4):473–97.
[10] Mohanty D, Colah RB, Gorakshakar AC, Patel RZ, Master DC, Mahanta J, et al. Prevalence of β-thalassemia and other haemoglobinopathies in six cities in India: a multicentre study. J Community Genet 2013;4(1):33–42.
[11] Sarker SK, Islam T, Bhuyan GS, Sultana N, Begum MN, Mahmud-Un-Nabi MA, et al. Impaired acylcarnitine profile in transfusion-dependent beta-thalassemia major patients in Bangladesh. J Adv Res 2018;12:55–66.
[12] de Silva S, Fisher CA, Premawardhena A, Lamabadusuriya SP, Peto TE, Perera G, et al. Thalassaemia in Sri Lanka: implications for the future health burden of Asian populations. The Lancet 2000;355(9206):786–91.
[13] Kiran SS, Aithal S, Belagavi CS. Hemoglobin E hemoglobinopathy in an adult from Assam with unusual presentation: A Diagnostic dilemma. J Lab Phys 2016;8(2):116.
[14] Baruah MK, Saikia M, Baruah A. Pattern of hemoglobinopathies and thalassemias in upper Assam region of North Eastern India: high-performance liquid chromatography studies in 9000 patients. Indian J Pathol Microbiol 2014;57(2):236.
[15] Colah R, Italia K, Gorakshakar A. Burden of thalassemia in India: The road map for control. Pediatr Hematol Oncol J 2017;2(4):79–84.
[16] Fucharoen SU, Winichagoon PR, Thonglairoam VA, Siriboon WI, Siritanaratkul NO, Kanokpongsakdi SU, et al. Prenatal diagnosis of thalassemia and hemoglobinopathies in Thailand: experience from 100 pregnancies. Southeast Asian J Tropical Med Public Health 1991;22(1):16–29.
[17] Trehan U, Garewal G, Kaul D, Das R. Alpha thalassemia and alpha gene triplications in Punjabis, with and without beta thalassemia trait. Hematology 2000;6:153–60.
[18] Çil B, Ayyıldız H, Tuncer T. Discrimination of β-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. Med Hypotheses 2020;138:109611.
[19] Mennuti MT. Genetic screening in reproductive health care. Clin Obstet Gynecol 2008;51(1):3–23.
[20] Cousens NE, Gaff CL, Metcalfe SA, Delatycki MB. Carrier screening for beta-thalassaemia: a review of international practice. Eur J Hum Genet 2010;18(10):1077–83.

[21] Bhattacharyya KK, Chatterjee T, Mondal UB. A comprehensive screening program for β-thalassemia and other hemoglobinopathies in the Hooghly District of West Bengal, India, dealing with 21 137 cases. Hemoglobin 2016;40(6):396–9.
[22] Vasudev R, Sawhney V. Thalassemia major and intermedia in Jammu and Kashmir, India: A regional transfusion centre experience. Indian J Hematol Blood Transfus 2014;30(4):297–300.
[23] Heidemann DL, Baker-Genaw K, JosephNA Kuriakose P. Increasing cost sensitivity in the diagnostic evaluation of microcytic anemia. Consultant 2014;54(11):837–40.
[24] Italia KY, Jijina FJ, Merchant R, Panjwani S, Nadkarni AH, Sawant PM, et al. Response to hydroxyurea in β thalassemia major and intermedia: experience in western India. Clin Chim Acta 2009;407(1–2):10–5.
[25] Kantharaj A, Chandrashekar S. Coping with the burden of thalassemia: Aiming for a thalassemia free world. Global J Transfus Med 2018;3(1):1.
[26] Sinha S, Seth T, Colah RB, Bittles AH. Haemoglobinopathies in India: estimates of blood requirements and treatment costs for the decade 2017–2026. J Community Genet 2019;12:1–7.
[27] Colah RB, Gorakshakar A. Control of thalassemia in India. Thalassemia Rep 2014;4(2)..
[28] Khera R, Singh T, Khuana N, Gupta N, Dubey AP. HPLC in characterization of hemoglobin profile in thalassemia syndromes and hemoglobinopathies: a clinicohematological correlation. Indian J Hematol Blood Transfus 2015;31(1):110–5.
[29] Joutovsky A, Hadzi-Nesic J, Nardi MA. HPLC retention time as a diagnostic tool for hemoglobin variants and hemoglobinopathies: a study of 60000 samples in a clinical diagnostic laboratory. Clin Chem 2004;50(10):1736–47.
[30] Stoltzfus RJ. Iron deficiency: global prevalence and consequences. Food Nutr Bull 2003;24:99–103.
[31] Zhu L, Zheng WJ. Informatics, data science, and artificial intelligence. JAMA 2018;320(11):1103–4.
[32] Saniei E, Setayeshi S, Akbari ME, Navid M. Parameter estimation of breast tumour using dynamic neural network from thermal pattern. J Adv Res 2016;7(6):1045–55.
[33] Qiao W, Huang K, Azimi M, Han SA. A novel hybrid prediction model for hourly gas consumption in supply side based on improved whale optimization algorithm and relevance vector machine. IEEE Access 2019;7:88218–30.
[34] Zhou G, Moayedi H, Foon LK. Teaching–learning-based metaheuristic scheme for modifying neural computing in appraising energy performance of building. Eng Comput 2020:1–12.
[35] Shamshirband S, Rabczuk T, Chau KW. A survey of deep learning techniques: application in wind and solar energy resources. IEEE Access 2019;7:164650–66.
[36] Baghban A, Jalali A, Shafiee M, Ahmadi MH, Chau KW. Developing an ANFIS-based swarm concept model for estimating the relative viscosity of nanofluids. Eng Appl Comput Fluid Mech 2019;13(1):26–39.
[37] Qiao W, Tian W, Tian Y, Yang Q, Wang Y, Zhang J. The forecasting of PM2. 5 using a hybrid model based on wavelet transform and an improved deep learning algorithm. IEEE Access 2019;7:142814–25.
[38] Faizollahzadeh Ardabili S, Najafi B, Shamshirband S, Minaei Bidgoli B, Deo RC, Chau KW. Computational intelligence approach for modeling hydrogen production: A review. Eng Appl Comput Fluid Mech 2018;12(1):438–58.
[39] Fotovatikhah F, Herrera M, Shamshirband S, Chau KW, Faizollahzadeh Ardabili S, Piran MJ. Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. Eng Appl Comput Fluid Mech 2018;12(1):411–37.
[40] Chen J, Lu D, Liu W, Fan J, Jiang D, Yi L, et al. Stability study and optimization design of small-spacing two-well (SSTW) salt caverns for natural gas storages. J Energy Storage 2020;27:101131.
[41] Zhang Z, Jiang D, Liu W, Chen J, Li E, Fan J, et al. Study on the mechanism of roof collapse and leakage of horizontal cavern in thinly bedded salt rocks. Environ Earth Sci 2019;78(10):292.
[42] Nabavi-Pelesaraei A, Rafiee S, Mohtasebi SS, Hosseinzadeh-Bandbafha H, Chau KW. Energy consumption enhancement and environmental life cycle assessment in paddy production using optimization techniques. J Clean Prod 2017;162:571–86.
[43] Qiao W, Yang Z. Solving large-scale function optimization problem by using a new metaheuristic algorithm based on quantum dolphin swarm algorithm. IEEE Access 2019;7:138972–89.
[44] Qiao W, Yang Z. Modified dolphin swarm algorithm based on chaotic maps for solving high-dimensional function optimization problems. IEEE Access 2019;7:110472–4086.
[45] Zhou G, Moayedi H, Bahiraei M, Lyu Z. Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings. J Clean Prod 2020;120082.
[46] Wang WC, Xu L, Chau KW, Xu DM. Yin-Yang firefly algorithm based on dimensionally Cauchy mutation. Expert Syst Appl 2020;150:113216.
[47] Amendolia SR, Cossu G, Ganadu ML, Golosio B, Masala GL, Mura GM. A comparative study of k-nearest neighbour, support vector machine and multi-layer perceptron for thalassemia screening. Chemom Intell Lab Syst 2003;69(1–2):13–20.
[48] Setsirichok D, Piroonratana T, Wongseree W, Usavanarong T, Paulkhaolarn N, Kanjanakorn C, et al. Classification of complete blood count and haemoglobin typing data by a C4. 5 decision tree, a naïve Bayes classifier and a multilayer perceptron for thalassemia screening. Biomed Signal Process Control 2012;7(2):202–12.
[49] Jahangiri M, Khodadi E, Rahim F, Saki N, Saki Malehi A. Decision-tree-based methods for differential diagnosis of β-thalassemia trait from iron deficiency anemia. Expert Syst 2017;34(3):e12201.
[50] Lafferty JD, Crowther MA, Ali MA, Levine M. The evaluation of various mathematical RBC indices and their efficacy in discriminating between thalassemic and non-thalassemic microcytosis. Am J Clin Pathol 1996;106(2):201–5.
[51] Jiang F, Chen GL, Li J, Xie XM, Zhou JY, Liao C, et al. Pre gestational thalassemia screening in Mainland China: the first two years of a preventive program. Hemoglobin 2017;41(4–6):248–53.
[52] Old J, Traeger-Synodinos J, Galanello R, Petrou MA, Angastiniotis M. Prevention of thalassaemias and other haemoglobin disorders. Thalassaemia Int Fed Publ 2005;2:113–6.
[53] Mentzer W. Differentiation of iron deficiency from thalassaemia trait. Lancet 1973;1:882.
[54] Srivastava PC, Bevington JM. Iron deficiency and/orthalassaemia trait. The Lancet 1973;301(7807):832.
[55] Shine I, Lal S. A strategy to detect β-thalassaemia minor. The Lancet 1977;309(8013):692–4.
[56] Laengsri V, Shoombuatong W, Adirojananon W, Nantasenamart C, Prachayasittikul V, Nuchnoi P. ThalPred: a web-based prediction tool for discriminating thalassemia trait and iron deficiency anemia. BMC Med Inf Decis Making 2019;19(1):212.
[57] Rathod DA, Kaur A, Patel V, Patel K, Kabrawala R, Patel V, et al. Usefulness of cell counter–based parameters and formulas in detection of β-thalassemia trait in areas of high prevalence. Am J Clin Pathol 2007;128(4):585–9.
[58] Sahli CA, Bibi A, Ouali F, Fredj SH, Dakhlaoui B, Othmani R, et al. Red cell indices: differentiation between β-thalassemia trait and iron deficiency anemia and application to sickle-cell disease and sickle-cell thalassemia. Clin Chem Lab Med 2013;51(11):2115–24.
[59] Cao A, Kan YW. The prevention of thalassemia. Cold Spring Harbor Perspect Med 2013;3(2):a011775.
[60] Plengsuree S, Punyamung M, Yanola J, Nanta S, Jaiping K, Maneewong K, et al. Red cell indices and formulas used in differentiation of β-thalassemia trait from iron deficiency in Thai adults. Hemoglobin 2015;39(4):235–9.
[61] Jayabose S, Giamelli J, LevondogluTugal O, Sandoval C, Ozkaynak F, Visintainer P. # 262 Differentiating iron deficiency anemia from thalassemia minor by using an RDW-based index. J Pediatr Hematol Oncol 1999;21(4):314.
[62] Sirdah M, Tarazi I, Al Najjar E, Al HaddadR. Evaluation of the diagnostic reliability of different RBC indices and formulas in the differentiation of the β-thalassaemia minor from iron deficiency in Palestinian population. Int J Lab Hematol 2008;30(4):324–30.
[63] Ehsani M, Darvish A, Aslani A, Seighali F. A new formula for differentiation of iron deficiency anemia (IDA) and thalassemia trait (TT). Turk J Haematol (Supp) 2005;22:268.
[64] Urrechaga E, Hoffmann JJ. Critical appraisal of discriminant formulas for distinguishing thalassemia from iron deficiency in patients with microcytic anemia. Clin Chem Lab Med (CCLM) 2017;55(10):1582–91.
[65] Urrechaga E. Analytical evaluation of the ADAMS™ A1c HA 8180 thalassemia mode high-pressure liquid chromatography analyser for the measurement of HbA2 and HbF. Int J Lab Hematol 2016;38(6):658–62.
[66] Revised rates of hospital charges all india institute of medical sciences, New Delhi, India. www.aiims.edu/aiims/hosp-serv/hosp-rates/revised-rate-listcopy.htm.
[67] Dolai TK, Bera R, Maji SK, Mandal PK. Profile of hemoglobin D trait in West Bengal, India. Thalassemia Rep 2014;4(1).
[68] Iyer S, Sakhare S, Sengupta C, Velumani A. Hemoglobinopathy in India. Clin Chim Acta 2015;444:229–33.