Contents lists available at ScienceDirect

Computers in Biology and Medicine

A two-layered classifier based on the radial basis function for the screening of thalassaemia

G.L. Masala *, B. Golosio, R. Cutzu, R. Pola

University of Sassari, via Piandanna 4, 07100, Sassari, Italy

ARTICLE INFO

ABSTRACT

The thalassaemias are blood disorders with hereditary transmission. Their distribution is global, with particular incidence in areas affected by malaria. Their diagnosis is mainly based on haematologic and genetic analyses. The aim of this study was to differentiate between persons with the thalassaemia trait and normal subjects by inspecting characteristics of haemochromocytometric data.

The paper proposes an original method that is useful in screening activity for thalassaemia classification. A complete working system with a friendly graphical user interface is presented.

A unique feature of the presented work is the adoption of a two-layered classification system based on Radial basis function, which improves the performance of the system.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Thalassaemia is present all over the world especially in areas affected by malaria in the Mediterranean area (Italy, Greece, Turkey, and Cyprus) and in southeast Asia (India, Vietnam, and Cambodia). It is a genetic disease that causes a reduction in the life span of a red blood cell. The disease is a result of an abnormality in the genes that regulates the formation of haemoglobin (Hb)—a core component of the red blood cell [1,2].

Thalassaemia is an autosomal recessive trait. It occurs therefore in homozygous subjects when both alleles are mutated. The presence of the thalassaemia trait in the heterozygous form does not lead to a pathological condition, so these subjects are commonly considered healthy carriers. Screening of the heterozygous population is fundamental to keep thalassaemic pathology diffusion under control.

In order to make the diagnosis, the blood characteristics must be analysed. A complete blood count is the primary screening test for a laboratory diagnosis of thalassaemia. However, there is still a limitation in the analysis of data due to a large number of possible candidate characteristics. In addition, there are various types of thalassaemia and thalassaemia traits (persons with the thalassaemia trait do not have the disease but inherit genes that cause the disease). As a result, a manual diagnostic process can only be carried out by specialists whose decision is based upon an index of mathematically combined values of blood characteristics [3].

Thalassaemia is present in different forms, the best known of which are called α- or β-thalassaemia, depending on whether the mutated genes are for the α- or for the β-chain of haemoglobin respectively. There are different types of α-thalassaemia resulting from different gene mutations. This study will consider only $\alpha^{3.7}$, α^{NCOI} and α^{α} [4,5], which are typical anomalies in Sardinia (Italy). β-thalassaemia major is the most severe form of thalassaemia. It is characterised by mutations in both copies of the gene coding for the β-chain. In β-thalassaemia minor, one copy of the gene for the β-chain is defective and generally those affected have no symptoms, except in pregnancy, when there is anaemia. [4]

The α- and β-thalassaemia carrier recognition is based on a first-level analysis performed with haemochromocytometric data and a second-level examination (HbA_2 quantification, globin chain synthesis, and genetic analysis) [6,7]. As many of the latter techniques, which are finalised to a secure diagnosis of the genetic defect, are time-consuming and expensive, it would be important to have an automated system for diagnostic support doing mainly the haemochromocytometric data and on the simple HbA_2 quantification.

Thalassaemia classification can generally be formulated into a pattern recognition problem. In this paper a complete diagnostic system is presented based on a two-layered decision module able to distinguish normal patients from α-thalassaemia carriers and β-thalassaemia carriers.

The following sections are organised as follows: previous work on thalassaemia classification are discussed in Section 2, and in Section 3, the description of the diagnostic working station is given. Next, the novel classification model is explained in Section 4. The database and the features used are described in Section 5. Experimental results are presented in Section 6. The discussion of results is presented in Section 7. Finally, the conclusions are drawn in Section 8.

* Corresponding author. Tel.: +39 79 229 486; fax: +39 79 2294 82.
 E-mail address: gilmasala@uniss.it (G.L. Masala).

2. Previous works

Several expert systems have been proposed to detect thalassaemic forms by automated diagnostic tools; early works employed image analysis [8], and statistical [9] and clustering techniques [10]. Later, the implementation protocol shifted to the expert systems, in which both rule-based [11–13] and hybrid neural-network/rule-based systems [14] have been successfully tested in clinical trials. Nonetheless, these tools broadly differentiate between a wide range of blood-related diseases including various types of anaemia. Neural-network-based systems [15], a K-nearest neighbours technique [16] and a support vector machine [16] can differentiate between two types of thalassaemic gene carriers and normal subjects. Further studies have reported the expansion of the tool capability to cover all major types of thalassaemia [17]. The use of a neural network and genetic programming are reported in [3] while a system based on neural networks and a decision tree is proposed in [2].

3. The working model

The following diagram (Fig. 1) represents schematically the diagnostic system presented in this work. The aim was to acquire new cases in the database and make a direct comparison of the automatic results with respect to the medical decision provided after the second-level analysis.

Our system takes as input some haemochromocytometric data and uses two classifiers able to distinguish between normal patients with respect to α-thalassaemia carriers and β-thalassaemia carriers. Each module of this pattern recognition system and the performances are described in the next paragraphs. A friendly graphical user interface is used to manage the system as shown in Fig. 2.

The software is written in C++. The Graphical Users Interface was implemented using the Borland Builder 5 visual library. The program is compatible with all Windows platforms. Actual data are manually typed; we are developing the direct connection between the pc and the haemochromocytometric analyser using a serial standard port.

4. The classifier module

The conceptual boundary between raw input data, feature extraction and proper classification can be somewhat arbitrary. The traditional goal of the feature extractor is to characterise raw data by measurements whose values are very similar for objects in the same category, and very different for objects in different classes. An ideal feature extractor would therefore yield a representation that would
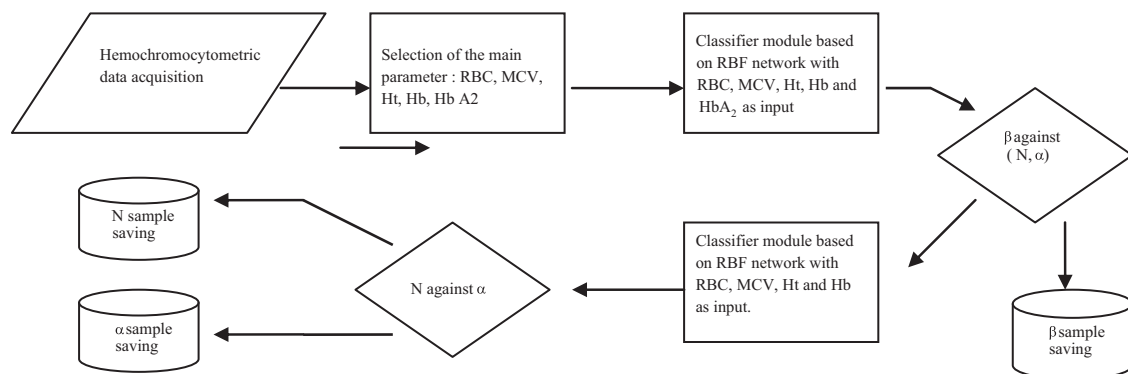

Fig. 1. Block diagram of the actual diagnostic station. After the haemochromocytometric data acquisition, five inputs are provided to the first classifier that discriminates between the β-sample against all (N, α): in this step if a β-sample is found, the algorithm stops and the system saves the result. Otherwise the second RBF classifier decides between a normal or α sample and saves the data.
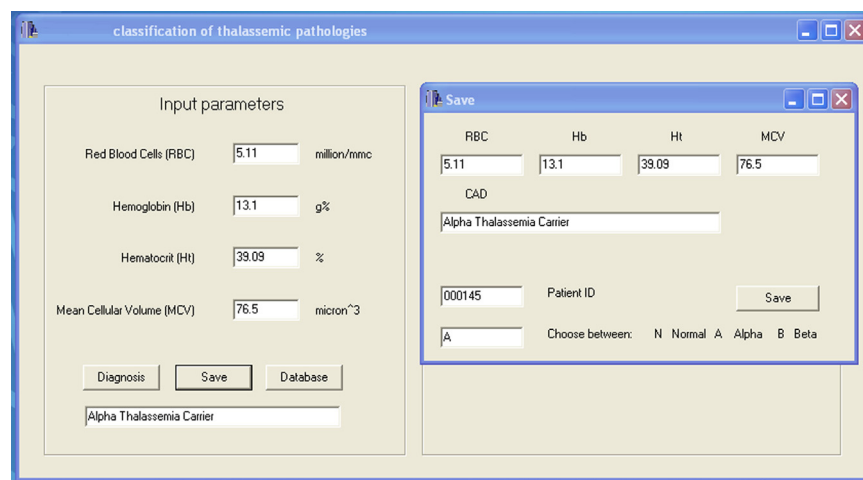

Fig. 2. Graphical User Interface of the actual diagnostic system. In the screenshot it is possible to see the input parameters that are actually manual typed. Pressing button "diagnosis", the results on the bottom box appears. Pressing "save", the window on the right appears that summarises the main data, the results so it is possible to store in the database the results with additional information such as patient ID. The "database" button opens the database mask where it is possible to read all data stored for research purposes.

make the job of the classifier trivial; conversely, an omnipotent classifier would not need the help of a sophisticated feature extractor.

For the purpose of this work we do not need a sophisticated features extractor and we just perform a comparative analysis of some classifiers to obtain the best performance on thalassaemia classification. Previous studies have demonstrated that data reduction based on component analysis does not lead to an improvement in the diagnosis of thalassaemia [15,16] and that data reduction on haemochromocytometric data using component analysis is not suited to our purposes and the features already used are the best, as explained in the database section of this paper.

The limits of the approach in [15,16] are related to the fact that we consider the α-cases as a unique class; as highlighted in the next paragraph, such cases have very different characteristics between them. Furthermore in [15,16], we do not take advantage of the additional feature, HbA$_2$, for the best classification of β-thalassaemia; this examination has now become a relatively economical routine. The system proposed in this work divides the recognition problem into two phases:

- First, recognising and separating only β-samples with respect to all others (normals and α-thalassaemia carriers).
- Second, discriminating several types of α-thalassaemia carriers between them and with respect to normal cases.

For this purpose some different type of classifiers are involved in this work:

i) The Radial Basis Function (RBF) network [18] is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. The architecture is of the feed–forward type without connections between neurons of the same layer. It has a static Gaussian function as the nonlinearity for the hidden layer processing elements. The Gaussian function responds only to a small region of the input space where the Gaussian is centred. The key for a successful implementation of these networks is to find suitable centres for the Gaussian functions. This can be done with a supervised learning, but an unsupervised approach usually produces better results. For this reason we implement RBF networks as a hybrid supervised–unsupervised topology. So the RBF networks are trained by a two-step algorithm. In the first step, the centre vectors of the RBF functions in the hidden layer are chosen. This step can be performed using K-means clustering. Afterwards, a back-propagation step is performed to fine-tune all of the RBF net's parameters.

ii) The Probabilistic Neural Network (PNN) [18] provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers. The probabilistic neural network uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category, or class. The learned patterns can also be combined or weighted with the *a priori* probability, also called the relative frequency, of each category to determine the most likely class for a given input vector.

iii) The K-Nearest Neighbours (KNN) deterministic classifier [19], for which it is necessary to have a training set that is not too small and at a good discriminating distance. KNN performs well in multi-class simultaneous problem solving. There exists an optimal choice for the value of the parameter K, which brings to the best performance of the classifier. This value of K is often approximately close to $N^{1/2}$ where N is the number of training samples. So generally we prefer to test all K values starting from 1 to $N^{1/2}$ with the training set available. Euclidean distances have the characteristics required for a good discrimination.

In the comparison described in this paper, we use the MATLAB 6.5 tool where the classifiers are implemented in the neural networks toolbox and KNN in the Statistical Pattern Recognition Toolbox [20].

As discussed in the next sections, the classifiers are trained on a training set and tested on a validation set (subset of the training set) to determine the optimal configuration.

5. Database

The initial database that was used to train the computer to classify patients consisted of 304 clinical records based on a thalassaemia screening carried out by the Ozieri hospital on public-school students. Several public schools of northern Sardinia were used in the test. Each 8th-grade student (14–15-year-old boys and girls) took part in the screening. Although the records can be considered a random sample, subjects with iron deficiency were excluded from the test, as in their case iron levels in the blood must be normalised before thalassaemia diagnosis can be made [21].

Haemochromocytometric data, HbA$_2$ and genetic determination of the main thalassaemia defects ($\alpha^{3.7}$ and α^{NCOI} variants) were used to make the medical diagnoses. We considered the following haemochromocytometric parameters: Red Blood Cell count (RBC), Haemoglobin (Hb), Haematocrit (Ht), and Mean Corpuscular Volume (MCV). HbA$_2$ was determined to identify β-carriers. 27 subjects had a HbA$_2$ of $\geq 4\%$, while HbA$_2$ was $\leq 3\%$ in the other 277 cases. The first group were diagnosed as being β-carriers by medical analysis. Genetic analysis was used to diagnose the α-carriers.

The features that were considered relevant for the classification are only the values of RBC, Hb, Ht and MCV (without normalisation), plus the HbA$_2$ parameter.

The data from the clinical records were divided into a testing set made of 108 records (which was the gold standard provided by genetic analysis), and a training set made up of 196 clinical records. The data were divided into the same testing and training sets utilised in our previous classification works [15,16]. Table 1 and Figs. 3 and 4 show the overall distribution of the resulting data sets.

From gold standard data, we calculated the statistics only for presenting the phenomenon. The minimum value of RBC is greater than the normal cases for the α carriers, with the exception of cases of $\alpha^{3.7}$-carriers that have similar values to the normal cases.

Statistical analysis shows that the different subclasses of thalassaemia carriers have different distribution of the considered parameters. However, due to the relatively small number of samples available in some of those subclasses it is not possible to discriminate them using such parameters.

In the distributions shown in Fig. 4, the α-thalassaemia carriers are not well separated from normal cases. On the other hand, the

Table 1
Dataset composition.

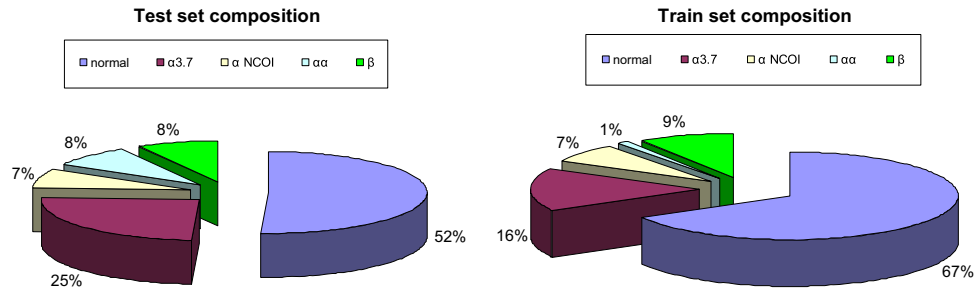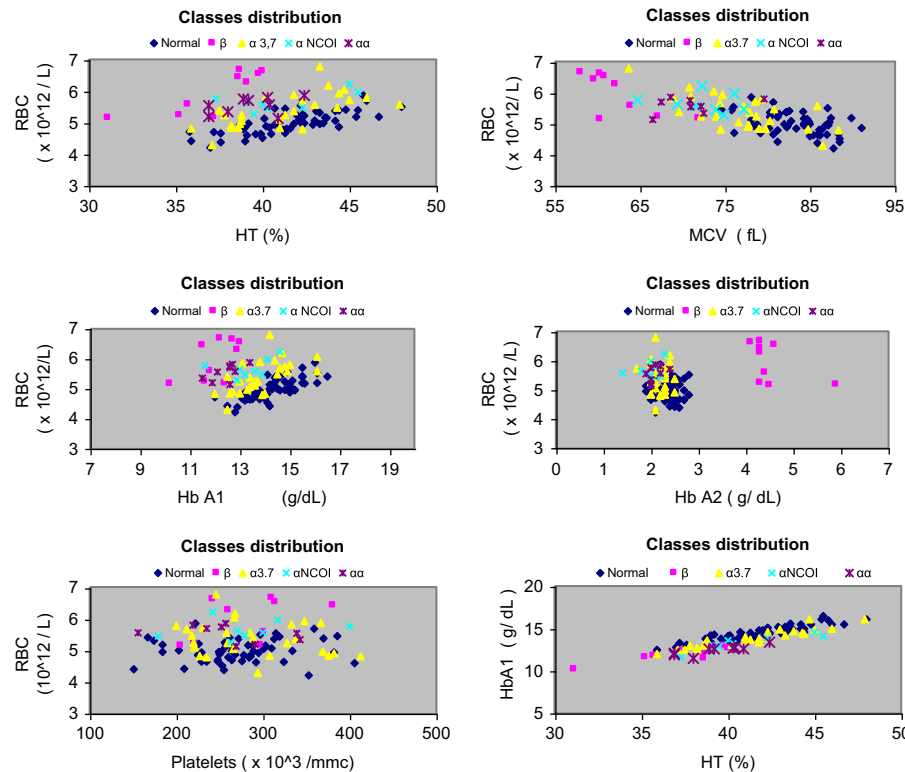| Set	Samples	Normal	$\alpha^{3.7}$	α^{NCOI}	α^{α}	β
Test	108	55	27	8	9	9
Train	196	131	31	14	2	18

Fig. 3. Dataset composition.

Fig. 4. Class distribution of the input parameter.

β-carriers are well separated from normals. The MCV is a highly discriminating feature. In medical practice, a threshold of this parameter is often used to recognise thalassaemia carriers. A threshold value of 77 fL was proposed in [15].

The parameter distributions indicate that the platelets are not useful for the characterisation of thalassaemia carriers. However, the HbA$_2$ provides a clear distinction between β and the group formed from α and normal together cases.

The characteristic parameters used in the next section will therefore be RBC, HbA$_1$, Ht and MCV in agreement with our previous work [15,16], plus the additional HbA$_2$ parameter.

6. Experimental results

The performance of the classification is expressed in terms of sensitivity and specificity. Sensitivity is defined as the ratio between the number of α or β carriers detected and the number of true carriers. The specificity is given by the ratio between the number of normal cases correctly identified with respect to the true normal cases.

The values of sensitivity and specificity are used to build the Receiver Operating Characteristic (ROC) curve, which shows the sensitivity versus false alarms (1−specificity). The value of the area under this curve indicates the accuracy of the system. The area above the curve indicates the error.

The results are reported in all tables with the confidence intervals (lower and upper limits of 95%) evaluated by the Wilson method with correction for continuity [22].
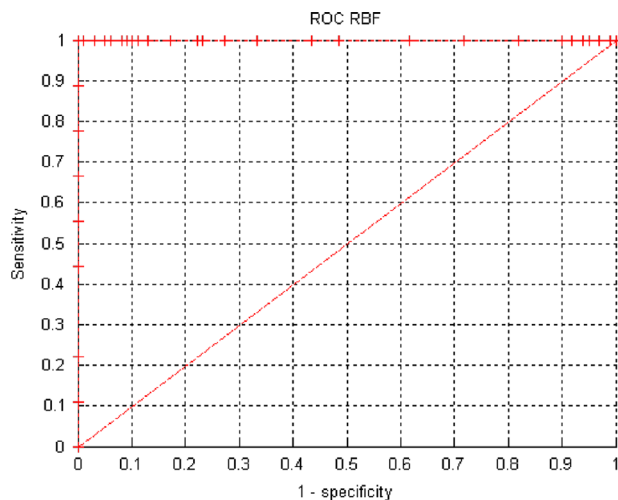
6.1. Preliminary test

We made some comparative tests to see the behaviour of the classifier by varying the input parameters and classes definition. The classifier used in this preliminary test is always the RBF classifier with the best specific configuration found. The training phase of the RBF classifier is made varying spread and using a subset of the training set (with a number of samples equivalent to the test set) to optimise performances. For the calculation of the sensitivity and specificity, the results are divided into two groups, i.e. α/β-carriers and normal patients.

In Table 2, the results with five input parameters and five classes are reported in the first row; using four input parameters and five classes in the second row, the sensitivity and specificity are the same, while working with three classes and five features,

Table 2
RBF classification using different classes and input features. Sensitivity is calculated with respect to α and β while specificity with respect to normals.

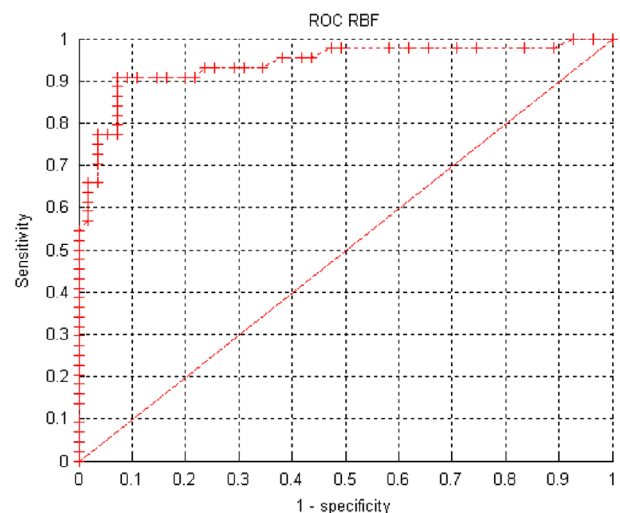| Classes and input features	Area under ROC curve	Specificity		Sensitivity (α and β)	
Normal, $\alpha^{3.7}$, α^{NCOI}, α^{α}, β RBC, MCV, Ht, Hb, HbA_2	0.94	0.91	+0.05 −0.10	0.92	+0.05 −0.11
Normal, $\alpha^{3.7}$, α^{NCOI}, α^{α}, β RBC, MCV, Ht, Hb	0.95	0.91	+0.05 −0.10	0.92	+0.04 −0.11
Normal, generic α, β RBC, MCV, Ht, Hb, HbA_2	0.94	0.87	+0.05 −0.10	0.92	+0.04 −0.11

Table 3
RBF classification for the first step (first layer) using different input parameters; sensitivity is calculated with respect to β samples.

Classes and input features	Area under ROC curve	Specificity		Sensitivity (β)	
β, (Normal+generic α) RBC, MCV, Ht, Hb, HbA_2	1	1	0 −0.4	1	0 −0.30
β, (Normal+generic α) RBC, MCV, Ht, Hb	0.92	0.88	+0.04 −0.08	0.89	+0.09 −0.33

Table 4
RBF classification in the second step (second layer) using the first row a classification on a unique α-class or in the second row calculated with respect to $\alpha^{3.7}$, α^{NCOI} and α^{α} variants. The sensitivity summarises the result for an α generic class.

Classes and input features	Area under ROC curve	Specificity		Sensitivity (α)	
generic α, normal RBC, MCV, Ht, Hb	0.92	0.91	+0.05 −0.13	0.89	+0.05 −0.11
$\alpha^{3.7}$, α^{NCOI}, α^{α}, normal RBC, MCV, Ht, Hb	0.94	0.91	+0.05 −0.13	0.93	+0.04 −0.10

Fig. 5. RBF, classification using 2 classes and 5 features.

the sensitivity does not change, but the specificity decreases with respect to the other rows.

From this brief analysis, we can say that in these classification tests we did not obtain better results than those achieved in previous studies [15,16] even though the parameter HBA_2 gives important information about the β-thalassaemia trait.

6.2. New approach

The new approach to the classification problem is divided into two steps:

- Separating, by RBF classifier, only β-samples with respect to all others (normals and α-thalassaemia carriers).
- Classifying different types of α-thalassaemia carriers among each other and with respect to normal cases.

The classification of β with respect to all the other cases, using RBF classifier, is shown using five input features. So, using the features RBC, Hb, Ht, MCV and HbA_2 and two classes we obtain the ROC curve as shown in Fig. 5, where it can be seen that all β-cases are correctly classified. The training phase of the RBF classifier is made by varying the spread parameter to optimise performances and using a subset of the training set for validation (with a number of samples equivalent to the test set) but removing all β-samples from the data.

These results are obtained by the RBF configuration shown in Table 3 and, just to highlight a comparison, we also show the results with four features (excluding HbA_2).

From Table 3 it is clear that the additional input feature HBA_2 is decisive in the classification in two classes: β-samples against normal–generic α-samples.

Fig. 6. RBF, classification using 4 classes and 4 features.

This result is justified by the biochemical study of the disease. In fact, in this type of thalassaemia, we observe an increase in haemoglobin secondary, in particular HbA_2, because it is not formed with protomer of type β and so it tries to compensate the deficit of HbA_1.

On the basis of results found using RBF and HbA_2 we do not need to compare the performance with other classifiers for this first step.

After the identification of the β cases, we can proceed with the second step of the classification. The discrimination of the normal with respect to all α-carriers was made using four parameters and distinguishing the two classes as normal and α. Using the RBF classifier, we obtain the first row of Table 4. Using RBF and the same four input features but discriminating four classes (normals with respect to $\alpha^{3.7}$, $\alpha^{NCO\,1}$ and α^{α} carriers), we obtain the best ROC in Fig. 6 and the best working point in the second row in Table 4.

Table 4 shows that the distinction of the different types of α-thalassaemias has improved the classification between normal and α-carriers. We do not show the accuracy of each class because it is still inaccurate. The discrimination among different types of α-

carriers could probably be improved if a larger dataset of cases with gold standard genetic analysis was available. This, however, would be beyond the scope of our medical software, through which are decided screening cases by the first-level blood analysis. In addition, we can report that the system has found the main difficulties in classifying $\alpha^{3.7}$-thalassaemia carriers.

For the same four input features and discriminating four classes (to compare results with Table 4 in the same condition), we also tested PNN and KNN classifiers, which provided the results shown in Table 5. In the training phase of these classifiers, we used the same validation set (subset of training set) used for RBF to optimise performances before the test. $K=9$ is used in the best configuration of KNN while for the best PNN configuration we found a spread of 10^{-6}.

We can assert from Table 5 that PNN does not provide better results in the classification with respect to the RBF classifier, although the sensitivity values of KNN were lower than the other classifiers. Even in these tests, it was noticed that the poor accuracy is determined by the incorrect classification of $\alpha^{3.7}$-carriers.

To summarise the performance of different classifiers, we show in Fig. 7 the sensitivity and specificity comparison graph for the three classifiers using four input features and discriminating four classes (normals with respect to $\alpha^{3.7}$, $\alpha^{NCO\ 1}$ and α^{α} carriers).

On the basis of the results, the final system was built using two RBF classifiers, one for each step of the classification. The first of these will have 14 hidden neurons, with a Gaussian spread

value of 25, and one output. The input data will be RBC, MCV, Ht, Hb and HbA$_2$. In the second step, the RBF will have six hidden neurons, with a Gaussian spread of 42.5, and one output. The input features to discriminate normals with respect to all the α-carriers will be RBC, MCV, Ht and Hb. In Fig. 8 we show the RBF architecture for the two layers.

7. Discussion

We have designed a system that improves the performance of previous systems [15,16] by a new classification system in two layers. We made also a comparative analysis using three classifiers: radial basis function, probabilistic neural network and K-nearest neighbours.

The new system is divided into two steps. In the first step, an RBF classifier recognises the carriers of β-thalassaemia. This is made possible by the use of the additional parameter HbA$_2$. This allowed us to detect β-carriers with 100% accuracy. In the second step, a RBF classifier discriminates normals with respect to α-carriers, identifying normal cases with 93% accuracy of and α-carriers with 91% accuracy.

The RBF shows better performance compared to other classifiers. Furthermore, the training of RBF is simple and fast. It also tends to delimit the space of decision in ellipses, reducing the possibility of error between classes very close to each other.

Fig. 9 shows a comparison of the proposed system with existing classification systems. In [1] the system is based on a Multi-Layer Perceptron (MLP) and it consists of three specialised neural networks. Normal individuals are identified with 95% accuracy, while α-carriers with 91% accuracy and β-carriers with 67% accuracy.

The system proposed in [2] is composed of two cascade classifiers both based on a Support Vector Machine (SVM). In the first step, it discriminates normal cases from all carriers, α and β. In the second step it discriminates α with respect to β carriers. With this system, normal cases are identified with 89% accuracy, 93% accuracy of α-carriers and 89% accuracy of β-carriers.

Table 5
Comparison of the best classifiers results on second layer. The sensitivity summarises the result for the α class.

Classifiers	Specificity		Sensitivity (α)	
RBF	0.91	$+0.05$ -0.13	0.93	$+0.04$ -0.10
KNN	0,91	$+0.05$ -0.11	0.80	$+0.09$ -0.14
PNN	0,73	$+0.10$ -0.15	0.89	$+0.05$ -0.11
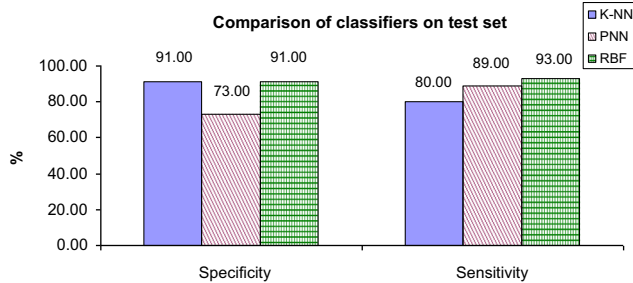
Fig. 7. Comparison of the classifiers used in the paper on test set.

Fig. 9. Comparison of accuracy of our different methods (actual and previous article).
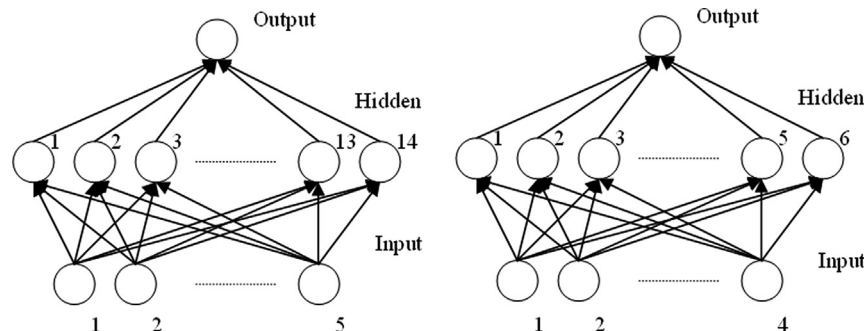

Fig. 8. On the left Radial Basis Function architecture for the first layer and five input parameters MCV, Hb, Ht, HbA$_1$, and HbA$_2$ while on the right Radial Basis Function architecture for the second layer and four input parameters MCV, Hb, Ht, and HbA$_1$.
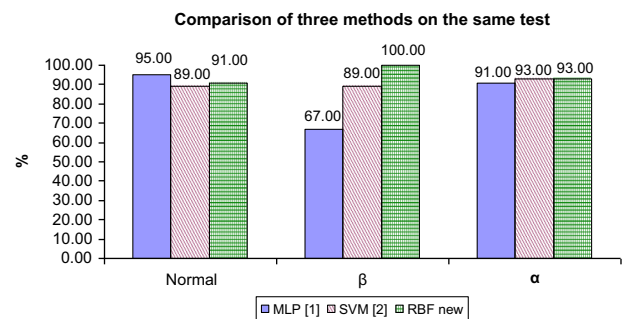
Table 6
Comparison of all methods cited in this paper. The input data in same paper are very different and it is the same for α types, which are present in various forms in the world. The dataset is same only for the first three rows.

Methods	Input data types	Accuracy (%)	Sensitivity (%) (α, β)	Specificity (%) (normals)
Specialized neural networks [15]	Hematomocytometric data	91.0	86.9	95.0
Two layers support vector machine [16]	Hematomocytometric data	90.6	92.3	89.0
Two layers radial basis function	Hematomocytometric data	92.3	93.0	91.0
Neural network and a decision tree [3]	Hematomocytometric data	82.0	–	–
Neural network and a decision tree [2]	Liquid chromatography	93.1	93.1	99.5

Therefore, the new system is an improvement on the previous systems [15,16]. Compared to the previous systems the new system has better performances on β-carriers and comparable performances on the other classes.

The types of α-thalassaemia vary greatly depending on the areas of the world where they are collected. In addition, some authors often use additional features not suitable for low-cost surveys or the first-level analyses. In [3], the authors use a neural network and a decision tree, which is evolved by genetic programming in thalassaemia classification; using 10 classes and 12 input features extracted from red blood cells, reticulocytes and platelets, they obtain an average classification accuracy of 82%. In [2], the authors propose a neural network and a decision tree in thalassaemia screening; their system is based on 13 classes of thalassaemia abnormality and one control class by inspecting the distribution of multiple types of haemoglobin in blood specimens, which are identified via high-performance liquid chromatography; they obtain a sensitivity of 93.1% and a specificity of 99.5%.

In Table 6, we review the main methods presented in the literature. It can be noticed that cited papers uses different input parameters and validation methods.

8. Conclusion

This paper describes a system for the recognition of thalassaemia carriers distinguishing between α, β and healthy cases. A station for screening diagnosis is presented. A classification scheme is also proposed, which improves the existing systems, based on two RBF classifiers placed in two cascade layers. The proposed system is competitive with others in the literature but it is a low-cost system because is trained on data obtained in simple routine analysis.

Summary

The thalassaemias are blood disorders with hereditary transmission. Their distribution is global, with particular incidence in areas affected by malaria. Their diagnosis is mainly based on haematologic and genetic analyses. The aim of the study was to differentiate between persons with the thalassaemia trait and normal subjects by inspecting characteristics of haemochromocytometric data. Thalassaemia is present in different forms, the best known of which are called α- or β-thalassaemia, depending on whether the mutated genes are for the α- or for the β-chain of haemoglobin respectively.

The α- and β-thalassaemia carrier recognition is based on a first-level analysis performed with haemochromocytometric data and a second-level examination (HbA2 quantification, globin chain synthesis, and genetic analysis) [6,7]. Because many of the latter techniques, which are finalised to a secure diagnosis of the genetic defect, are time-consuming and expensive, it would be important to have an automated system for diagnostic support based only on the haemochromocytometric data and the simple HbA2 quantification.

The paper proposes an original method that is useful in screening activity for thalassaemia classification. A complete working system with a friendly graphical user interface is presented.

The classification system proposed in this work divides the recognition problem in two phases:

- First, recognising and separating only β-samples with respect to all others (normals and α-thalassaemia carriers).
- Second, discriminating several types of α-thalassaemia carriers between them and with respect to normal cases.

A unique feature of the presented work is the adoption of a two-layered classification system based on the Radial Basis Function, which improves the performance of the system. A comparative study is made with respect to different types of classifiers (K-nearest neighbours, probabilistic neural networks) on the same dataset and with respect to other systems existing in the literature.

The proposed system is divided into two steps. In the first step, an RBF classifier recognises the carriers of β-thalassaemia. This is made possible by the use of the additional parameter, HbA2, which allowed us to detect β carriers with an accuracy of 100%. In the second step, a RBF classifier discriminates normals with respect to α carriers, identifying normal cases with an accuracy of 93% and α-carriers with an accuracy of 91%.

Conflicts of interest

None declared.

References

[1] D.J. Weatherall, J.B. Clegg, The Thalassaemia Syndromes, 4th Edition, Blackwell Science, Malden, MA, 2001.

[2] T. Piroonratana, W. Wongseree, A. Assawamakin, N. Paulkhaolarn, C. Kanjanakorn, M. Sirikong, W. Thongnoppakhun, C. Limwongse, N. Chaiyaratana, Classification of haemoglobin typing chromatograms by neural networks and decision trees for thalassaemia screening, Chemometrics Intelligent Lab. Syst. 99 (2) (2009) 101–110.

[3] W. Wongseree, N. Chaiyaratana, K. Vichittumaros, P. Winichagoon, S. Fucharoen, Thalassaemia classification by neural networks and genetic programming, Inf. Sci. (Ny). 177 (3) (2007) 771–786.

[4] David L. Nelson, Michael M. Cox (Eds.), Lehninger Principles of Biochemistry, fourth Edition, 2004.

[5] R. Galanello, C. Sollaino, E. Paglietti, et al., Alpha-thalassaemia carrier identification by DNA analysis in the screening for thalassaemia, Am. J. Hematol. 59 (1998) 273–278.

[6] Thalassaemia working party of the British Committee for Standars in Haematology Task Force, guidelines for investigations of the alpha and beta thalassaemia traits, J. Clin. Pathol. 47 (1994) 289–95, .

[7] British Committee for standards in Haematology. Guideline: the laboratory diagnosis of haemaoglobinopathies. Br J Haematol 101, 783-92, 1998.

[8] P.R. Lund, R.D. Barnes, Automated classification of anaemia using image analysis, The Lancet 300 (7775) (1972) 463–464.

[9] R.L. Engle, B.J. Flehinger, S. Allen, R. Friedman, M. Lipkin, B.J. Davis, L.L. Leveridge, HEME: a computer aid to diagnosis of hematologic disease, Bull. N.Y. Acad. Med. 52 (1976) 584–600.

[10] G. Barosi, M. Cazzola, C. Berzuini, S. Quaglini, M. Stefanelli, Classification of anemia on the basis of ferrokinetic parameters, Br. J. Haematol. 61 (1985) 357–370.

[11] S. Quaglini, M. Stefanelli, G. Barosi, A. Berzuini, ANEMIA: an expert consultation system, Comput. Biomed. Res. 19 (1) (1986) 13–27.

[12] S. Quaglini, M. Stefanelli, G. Barosi, A. Berzuini, A performance evaluation of the expert system ANEMIA, Comput. Biomed. Res. 21 (1988) 307–323.
[13] G. Lanzola, M. Stefanelli, G. Barosi, L. Magnani, NEOANEMIA: a knowledge-based system emulating diagnostic reasoning, Comput. Biomed. Res. 23 (1990) 560–582.
[14] N.I. Birndorf, J.O. Pentecost, J.R. Coakley, K.A. Spackman, An expert system to diagnose anemia and report results directly on hematology forms, Comput. Biomed. Res. 29 (1) (1996) 16–26.
[15] S.R. Amendolia, A. Brunetti, P. Carta, G. Cossu, M.L. Ganadu, B. Golosio, G.M. Mura, M.G. Pirastru, A real-time classification system of thalassemic pathologies based on artificial neural networks, Med. Decision Making 22 (1) (2002) 18–26.
[16] S.R. Amendolia, G. Cossu, M.L. Ganadu, B. Golosio, G.L. Masala, G.M. Mura, A comparative study of k-nearest neighbour, support vector machine and multi-layer perceptron for thalassaemia screening, Chemometrics Intelligent Lab. Syst. 69 (1) (2003) 13–20.

[17] S. Fucharoen, P. Winichagoon, Hemoglobinopathies in Southeast Asia: molecular biology and clinical medicine, Hemoglobin 21 (1997) 299–319.
[18] S. Haykin, Neural Networks—A Comprehensive Foundation, second edition, Prentice Hall, 1999.
[19] O. Duda, P.E. Hart, D.G. Stark, Pattern Classification, second edition, A Wiley-Interscience Publication John Wiley & Sons, 2001.
[20] F. Vojtech, H. Václav, Statistical Pattern Recognition Toolbox for Matlab, Center for Machine Perception, Czech Technical University, 2004.
[21] MM Eldibany, KF Totonchi, NJ Joseph, D. Rhone, Usefulness of certain red blood cell indices in diagnosing and differentiating thalassaemia trait from iron-deficiency anemia, Am. J. Clin. Pathol. 111 (1999). (676–82.10,).
[22] EB. Wilson, Probable inference, the law of succession, and statistical inferenceJ. Am. Stata. Assoc. 22 (1927) 209–212.