Computers in Biology and Medicine 43 (2013) 1724–1731

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

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

Article history:
Received 29 April 2013
Accepted 23 August 2013

Keywords:
Radial basis function
Probabilistic neural network
K-Nearest neighbours
α-Thalassaemia
β-Thalassaemia

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.

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 import ant 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).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.08.020

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

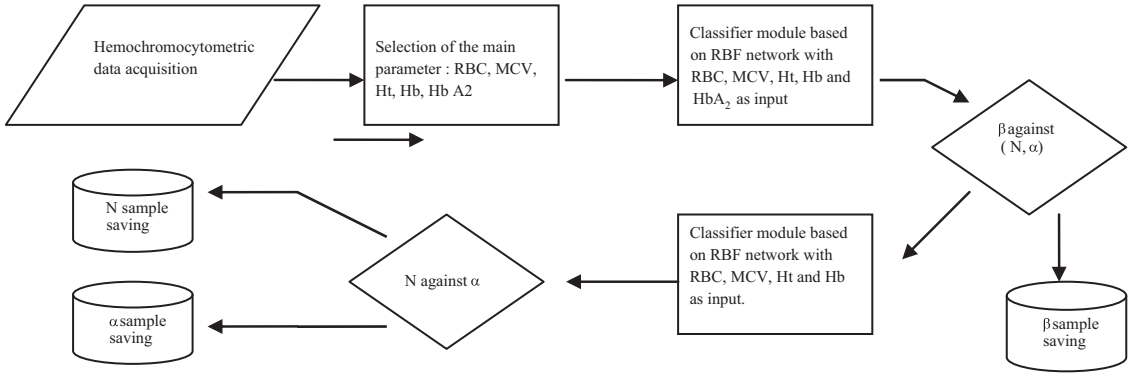

Fig. 1. Block diagram of the actual diagnostic station. After the haemochromocytometric data acquisition, five inputs are provided to the first classifier that discriminates between the β-sample against all (*N*, α): in this step if a β-sample is found, the algorithm stops and the system saves the result.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

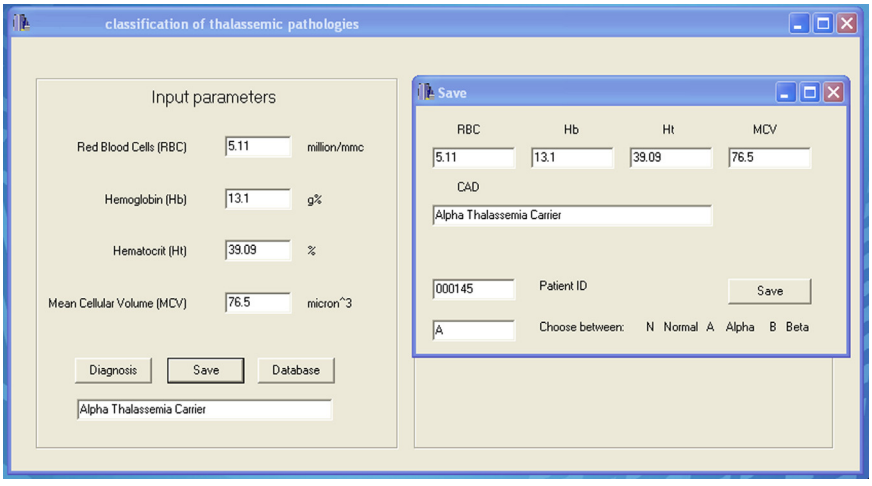

Fig. 2. Graphical User Interface of the actual diagnostic system. In the screenshot it is possible to see the input parameters that are actually manual typed. Pressing button "diagnosis", the results on the bottom box appears. Pressing "save", the window on the right appears that summarises the main 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.