Neural Comput & Applic (2017) 28:2029–2040
DOI 10.1007/s00521-015-2151-1

ORIGINAL ARTICLE

Blood type classification using computer vision and machine learning

Ana Ferraz[1,2] · José Henrique Brito[3] · Vítor Carvalho[1,3] · José Machado[2]

Received: 17 June 2015 / Accepted: 21 December 2015 / Published online: 13 January 2016

Abstract In emergency situations, where time for blood transfusion is reduced, the O negative blood type (the universal donor) is administrated. However, sometimes even the universal donor can cause transfusion reactions that can be fatal to the patient. As commercial systems do not allow fast results and are not suitable for emergency situations, this paper presents the steps considered for the development and validation of a prototype, able to determine blood type compatibilities, even in emergency situations. Thus it is possible, using the developed system, to administer a compatible blood type, since the first blood unit transfused. In order to increase the system's reliability, this prototype uses different approaches to classify blood types, the first of which is based on Decision Trees and the second one based on support vector machines. The features used to evaluate these classifiers are the standard deviation values, histogram, Histogram of Oriented Gradients and fast Fourier transform, computed on different regions of interest. The main characteristics of the presented prototype are small size, lightweight, easy transportation, ease of use, fast results, high reliability and low cost. These features are perfectly suited for emergency scenarios, where the prototype is expected to be used.

✉ José Machado
jmachado@dem.uminho.pt

Ana Ferraz
aps_ferraz@hotmail.com

[1] R&D ALGORITMI Center, School of Engineering, University of Minho, 4800-058 Guimarães, Portugal

[2] MEtRICs Research Center, School of Engineering, University of Minho, 4800-058 Guimarães, Portugal

[3] IPCA-EST, Polytechnic Institute of Cávado and Ave, 4750-810 Barcelos, Portugal

Keywords Blood types · Pre-transfusion tests · Plate test · Image processing · Machine learning

1 Introduction

Before administering a blood transfusion, it is necessary to carry out a wide range of compatibility tests, called pre-transfusion tests [1–6]. However, in certain emergency situations, where time available is short and does not allow the completion of these tests using the current commercial systems available in laboratories, the O negative blood type (universal donor) is administered [7–21]. Nonetheless, sometimes, despite their greater compatibility, this blood type may cause transfusion reactions, which may worsen the state of health or even be fatal to the patient [3–5].

In addition to these existing commercial systems, available in laboratories, there are other manual options [1, 22] which also allow the determination of the blood type. However, these methods (tube and plate tests) always require the intervention of a laboratory technician, being susceptible to human errors or mistakes, either in the execution of the test procedure, or in the reading and interpretation of results [1], [22–26]. Thus, despite having fast results, they are not suitable to emergency situations, due to the possible human errors associated, considering the security that the universal donor provides to whom authorizes the transfusion [23–26]. Furthermore, the universal donor blood is not always enough for all needs and sometimes; there is a stock disruption of this type of blood. Thus, the ideal scenario would be to perform pre-transfusion tests even in emergency situations and always administer a compatible blood, both to eliminate blood incompatibilities, and not to scarce the universal donor blood type.

This paper proposes the methodology for the validation of a developed prototype based on the plate test approach that performs pre-transfusion tests and is appropriate to emergency scenarios. The prototype must have small size to increase portability, in order to allow performing tests in different locations, including ambulances and accident sites. The prototype operates with a customized software application designed to automate the reading and interpretation of results. These results are obtained using image processing techniques. The acquired image is processed after the test, and the blood type is obtained using machine learning algorithms (Decision Trees and support vector machines (SVMs). This paper presents the developed prototype, the image processing techniques used and the classification methods to obtain the blood type results.

This paper is organized into four sections. The second section, Materials and Methods, presents the developed prototype, the image processing techniques used and the classifiers. The third section, Results, shows some experimental results obtained with each of the classifiers. Finally, in the fourth section, Conclusions and Future Work, the main issues of the work are presented, as well as some future work developments.

2 Materials and methods

This section presents the prototype, the image processing techniques used in the developed software application, and the classifiers applied to obtain the blood type test results. The prototype will automate the procedure involved in the plate test which is a manual test. The plate test requires the addition of specific reagents (Anti-A, Anti-B, Anti-AB and Anti-D) and a drop of blood in each container, for ABO and Rh tests. It is important to assure that each reagent is well mixed with the drop of blood for about 30-60 s. After the mixture it is recommended to wait 2 min, so that weak antigens are nor hidden, and after this time it is possible to look for signs of agglutination and interpret the results of the test. The stronger reactions occur in a few seconds, during the mixture, but it is important to wait for the recommended time (2 min) to ensure that weak agglutinations are detected. The whole procedure with the prototype needs about 5 min to achieve results. The prototype has to consider these procedures in the operation to ensure safety and meet the test requirements [24]. Commercial systems available are based on centrifugation to promote the mixture of reagents with the blood, but the times needed to achieve results are higher when compared with the manual tests [7–21]. The prototype presented in this paper is based in a fast test procedure and will be faster than the systems currently available in the market.

The prototype is specially designed to emergency scenarios. The main steps to administering a blood transfusion with the prototype are: first, it is necessary a blood sample to perform the tests with the system; second, using the blood sample collected there are performed the pre-transfusion tests and achieved the test results; and third, it is selected the compatible unit of blood for transfusion (previously tested for infections diseases and all the antigens and antibodies are identified previously); and finally, it is administered the blood to the patient. The unit of blood transfused is from blood donors and so the preoccupation about the possibility of infections is reduced as this blood is previously tracked. Relatively to the possibility of the patients' blood could be contaminated there is not a problem for the system because these contaminations should not interfere with the tests results. In real scenarios, there are not performed tests to infections in the patient before administering a blood transfusion, so the system will be effective even in these situations. In future it will be recommended to have in medical vehicles a refrigerator with blood units' storage, in order to allow remote transfusions. The system could be suited to detect some diseases, like syphilis, typhoid fever, Brucella, tick fever, mononucleosis, bacteria streptococcus, among others, as they are based in the same principle used in the prototype for blood type detection (using specific reagents and a sample of blood). However, it is still necessary to perform tests to confirm the ability of the prototype to achieve safe results for these additional tests.

2.1 Prototype

The prototype, presented in Fig. 1, was modeled using the Autodesk Inventor 2013 software [27, 28], where all the necessary requirements for its operation are indicated.

The prototype shown in Fig. 1 was modeled based on previously developed prototypes [29]. It has two main sections—an upper part and a lower part, (components E, F, G, H, I, J, K, L) and (components A, B, N), respectively. Between the upper part and the lower part there is a rotation axis (C) and a separation plane (D), permitting the replacement of the plate used in each blood type test. In case there is a malfunction in one of the prototype modules, it is possible to replace each specific module separately. The outer cylindrical body (E) along with the part (F) constitutes the chassis of the prototype. The prototype has a part (I), which supports the motor (J) and the blood type test plate (K), which has six containers, where the mixture of blood and reagents is performed. The base (F) contains an Arduino Duemilanove [30] (G), an Arduino Ethernet Shield [31] (G), a battery (H), an encoder (Q) which allows motor speed control (J) and an LDR (*Light Dependent Resistor*) sensor (R) that allows the brightness control of the LEDs *(Light Emitting Diode*) (N).

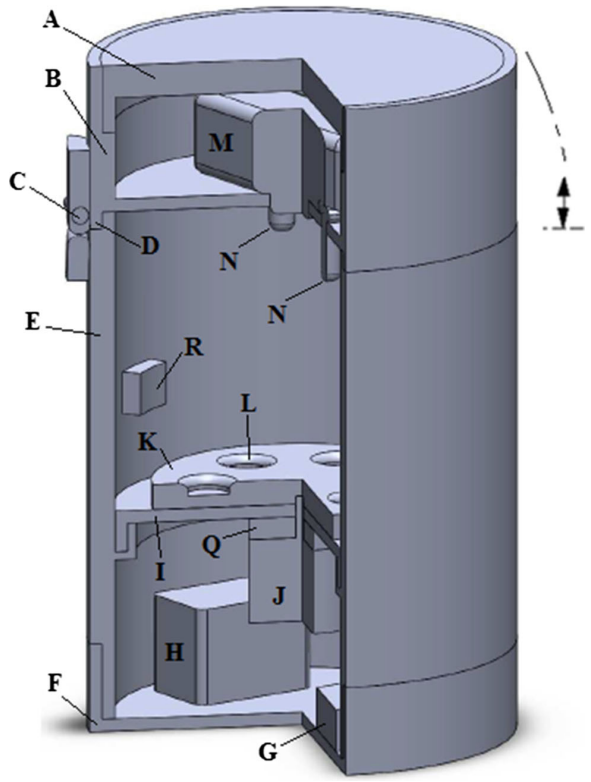


Fig. 1 Functional scheme of the final prototype. A—Top cover; B—Upper support; C—Rotation axis of the upper part; D—Separation plan between top and the bottom parts; E—Cylindrical body; F—Basis where is fitted the outer cylindrical body; G—Both Arduino boards and electronic components; H—Battery; I—Plate test holder; J—DC Motor, Q—Encoder; K—test plate; L—Container for mixing of blood and reagents; M—HD Camera; N—LED for illumination, R—LDR sensor

In the upper part of the prototype (B), there is an HD (*High Definition*) camera (M) that captures images of the performed tests and white LEDs (N) for proper illumination during the image capturing with the camera. The prototype has a top cover (A) that allows easy access to the HD camera (M) and the LEDs (N). Moreover, the body (E) and the base (F) also allow simple access to the inside of the prototype base where most of the required hardware is placed. Aluminum was selected as the material of the outside of the prototype, due to its low density facilitating its portability and mechanical characteristics. The dimensions of the prototype are 160 mm of height and 100 mm diameter,

2.2 Image processing techniques

The developed software application [29] uses a number of image processing steps in the analysis of the image captured by the HD camera. Image processing is widely used in health technological solutions [32–37]. Some of the current systems, available in clinical laboratories, also resort to the use of these techniques to obtain the automatic test results [19, 21].

We believe that the HD camera could be from a smartphone and the software could also be implemented on a smartphone. However, the developed prototype ensures proper illumination conditions for image acquisition which may not be the case if a smartphone was used in a real scenario. Furthermore, for the approach outlined in this paper a B/W camera would be sufficient. Nevertheless, the prototype was fitted with a color camera to enable future research that may take advantage of the use of color images.

The authors' initial studies used a variety of image processing techniques [29, 38–41] to obtain the blood type classification. However, after the construction of the prototype it was possible to optimize the procedure and reduce the number of required image processing techniques allowing faster results and less computational effort. In contrast to the previous work, the proposed solution now only has a few pre-processing steps in the analysis of each blood sample, as shown in Fig. 2.

Observing Fig. 2 it is possible to verify that four (4) image processing steps are needed to obtain the results. Step 1 (color conversion) is applied to the original image Fig. 2-0, extracting the image color planes, converting the image into an *8 bit* image. Then, step 2 performs an image search for a small "white circle", which is the image reference of the position of the first container, using template matching. Afterward, step 3 performs a search for the six larger "white circles", corresponding to the containers where the mixture of blood/reagents is performed, also using template matching. Finally, step 4 quantifies statistical parameters of the image to be analyzed. Step 4 extracts Regions of Interest (ROIs) from the image, one corresponding to each container in the test plate.

A number of features will be extracted from each ROI and will be subsequently fed to the classifier so that a blood type classification may be determined. Different types of ROIs were considered, square ROIs and circular ROIs. The square ROIs (Fig. 2 ROIs 1) were obtained by manually selecting a 40×40 pixel rectangular patch inside the detected container circle. The circular ROIs (Fig. 2 ROIs 2) were obtained by using the area around the central position of the detected containers with a fixed radius. Using square ROIs avoids including the specular reflection of the light coming from the LEDs (visible in Fig. 4). The use of circular ROIs maximizes the use of all available pixels in each container. Both types of ROIs will be used to extract features that will be used by the classifiers. The classifiers will be responsible for determining whether or not the agglutination has occurred in each of the containers.

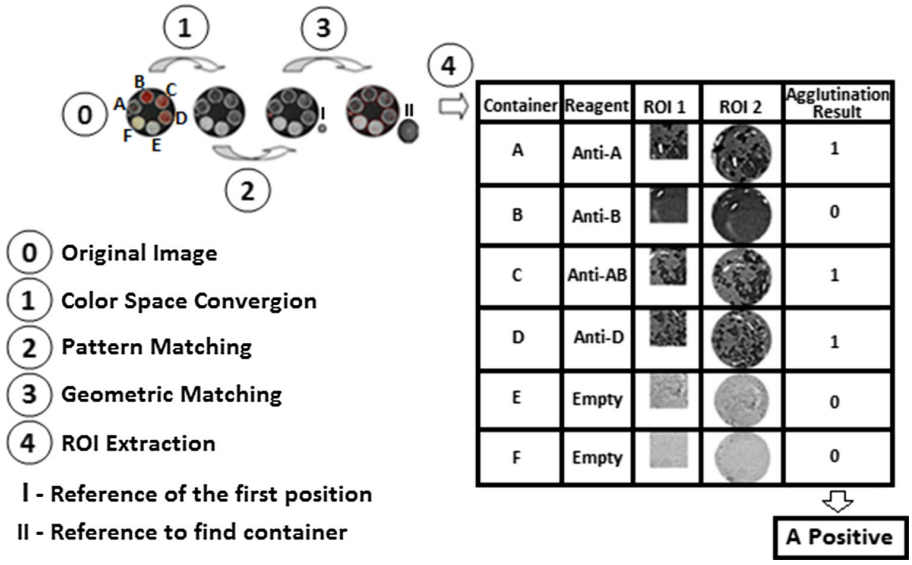


Fig. 2 Image pre-processing steps

From the agglutination detection on each of the containers, it is possible to infer the blood type.

The interpretation of test results and detection of the agglutination may be performed with naked eye, by an experienced laboratory technician, but one would like to remove human intervention from the process to reduce subjectivity and avoid human error. A system capable of automatic detection of agglutination could prevent the occurrence of human error and potentially be safer and more effective.

As can be seen in Fig. 2, when agglutination occurs, the patch of the corresponding container presents some blood clusters, or agglutinates; when the blood has a homogeneous appearance this means that no agglutination has occurred. Thus, observing Fig. 2 it can be concluded that agglutination occurs in container A, which means that the antigen A is identified in the blood; in container B agglutination does not occur, which means that the B antigen is not identified; in container C agglutination has occurred which confirms the presence of A antigen in the blood; in container D agglutination has also occurred, which identifies the Rh(D) antigen. In containers E and F, no tests are performed (Fig. 2-E and F), since only four reagents (anti-A, anti-B, anti-AB and anti-D) are needed for ABO and Rh tests. Thus, these last 2 containers are not considered for analysis, nor used by the classifiers. With this combination of results, a laboratory technician would conclude that the blood type is A positive.

In a previous approach [39], the blood type detection was performed by computing the standard deviation of each square ROI of each container and successively applying a threshold with a value of 16 to each value of the standard deviation, starting from container A, followed by B, C and D. This was implemented with a decision tree with four levels. In each level, the standard deviation of the corresponding container was compared to a threshold and the leaves in the lower levels would contain the classification of the blood type. In this new approach, different types of ROIs were considered, both square and circular (Fig. 2 ROIs 1 and ROIs 2), different features were used, besides the standard deviation and different types of classifiers were used to detect agglutination, which is then used to determine the blood type. The key to this new approach is that the agglutination of the blood samples in each container needs to be reliably detected from the acquired images, for the blood type determination to be successful. The role of the classifiers will be to determine whether or not agglutination is present in a given ROI, corresponding to a given container. From the agglutination classifications, the blood type can be easily determined. In contrast to the previous approach, the determination of blood type is deterministically dependent on the combination of agglutination detections, considering the full set of ROIs, instead of depending directly on the standard deviation values of each ROI. This allows the system to more accurately model the agglutination detection, instead of modeling the blood type directly from the features. This more accurately mirrors the classical laboratory procedure for blood type determination. It is important to highlight that the required image processing computation it is very fast taking between 2 and 5 s and so suitable for emergency scenarios.

2.3 Blood type determination

The occurrence of agglutination identifies the presence of the specific antigen in the blood. For example, in the first

container the presence or absence of the A antigen will be tested; in the second container the B antigen will be tested; the third container will confirm the A or B antigens; and in the fourth container will test the D (Rh) antigen. The agglutination reaction in each container depends on the presence or absence of the corresponding antigen. The combination of agglutination detection results allows a deterministic identification of the blood type, namely blood group ABO and Rh. Figure 3 shows how the blood type may be identified from the agglutination detections. If, for example, the blood sample in Fig. 2 was to be applied to the flowchart in Fig. 3, since agglutination was detected in container A, the decision would be the left branch in level I, no agglutination was found in container B so the right branch of level II would be taken, and agglutination was found in containers C and D, which means that the left branch of both levels III and IV would be taken, leading to the conclusion that the blood type is A positive (A+). The flowchart in Fig. 3 is in accordance with the results interpretation guidelines described in the reagent's protocol [24].

Having extracted the ROIs corresponding to each container from the original image of a blood sample, one needs to compute features that carry enough information for a decision between whether agglutination is present or not. A previous approach simple used the standard deviation value as a feature. This feature was also considered in this new approach, along with other features, namely the gray level histogram, Histogram of Oriented Gradients HoG [42] and fast Fourier transform (FFT). Using the data provided by these features, different types of classifiers were built to determine whether agglutination is present or not. Different machine learning approaches were used to train and test classifiers able to do such classification, based on the features previously mentioned.

Thus to classify the ROIs as agglutinated or not agglutinated, different classifiers were considered, namely classifiers based on Decision Trees and SVMs. In the medical science field, the application of these techniques is widely used [43–49]. These types of classifiers are applied to scenarios of supervised learning where a dataset is available with examples of feature values and their respective class provided as ground truth data. From a new and previously unknown set of feature values, the classifier is able to predict the class which the new example belongs to. The task at hand can be reduced to a supervised learning problem of binary classification of each container image patch, as agglutinated or not agglutinated. There are several classification frameworks, e.g., Decision Trees, SVMs and Neural Networks [50]. For this work the first two were considered, as they have been widely applied in the health domain to aid in the diagnosis of diseases, among others, showing efficient results [51–60]. Each type of classifier has its own advantages and disadvantages [50]. Regarding methods based on decision trees the considered classifiers were Bagging, Random Forests and Boosting, specifically AdaBoost and LPBoost. The used implementations were the ones available in the MATLAB statistics toolbox [61–64]. All the default options were kept for all these classifiers. The used SVM implementation was also the one available in the MATLAB statistic toolbox [65], also keeping the default options.

While other AI algorithms could be considered for the classification task, Decision Trees and SVMs were chosen for this approach as they have less parameters to choose, have readily available implementations and are, in general, more likely to yield the best classification results [66]. However, investigating the use of different classifiers that may be better fitted to the particular classification problem and hence improve classification performance is one of the research efforts that will be pursued in future work.

3 Experiments and results

To evaluate the system's performance, a dataset was collected and stored in a database. Blood samples were collected at *Hospital Professor Doutor Fernando Fonseca,*

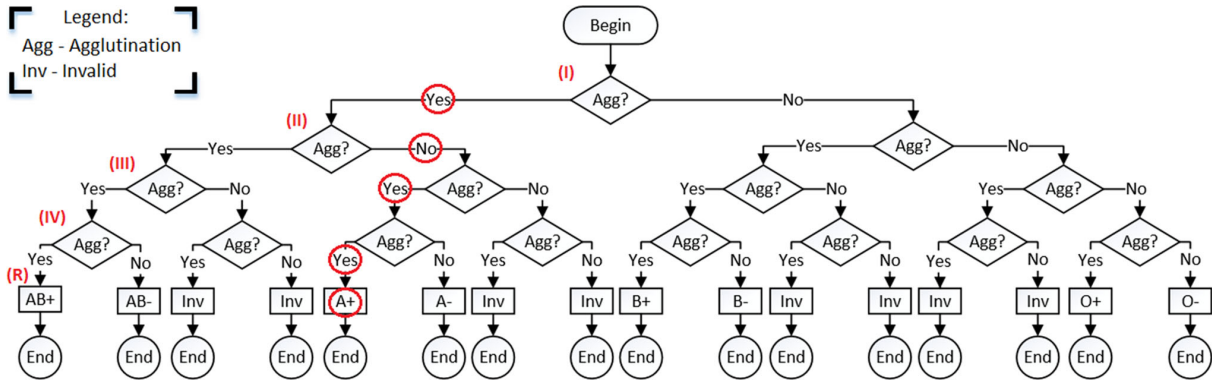


Fig. 3 Flowchart of the classification algorithm developed

EPE, Lisbon, Portugal [66]. Each blood sample was subjected to a test in the prototype. The tests were performed according to the plate test protocol, inserting a drop of each of the reagents in the respective container and then a drop of the blood sample to be analyzed. After mixing the blood with the reagents using the prototype at a rotational speed of 55 rpm during approximately 2 min, the image is captured and pre-processed, by the developed software application. Images captured by the prototype, look like the one in Fig. 4.

Looking at Fig. 4a, a trained laboratory technician would be able to identify agglutination in containers A, C and D, while container B shows no agglutination. This corresponds to an A positive blood type (A+). In Fig. 4b only container D has agglutination which corresponds to an O positive blood type (O+). The pre-processing steps are applied to the acquired images, yielding ROIs (square and circular), as shown in Table 1. The image patches of each container of all the collected blood samples constitute the initial dataset, along with ground truth classification data. In fact, two different datasets are considered, one with square ROIs and the other one with circular ROIs. The ground truth classification for each container (agglutinated/not agglutinated) was provided by an experienced technical analyst. The laboratory technician observed each image looking for agglutination signs and manually classified each container as agglutinated or not agglutinated. Table 1 shows part of the dataset, which corresponds to the test plates shown in Fig. 4.

In Hospital *Professor Doutor Fernando Fonseca EPE*, 41 ABO and Rh tests were performed with blood from different individuals. Each blood test produces 4 images patches corresponding to the four used containers (A, B, C and D), yielding 164 ROIs of each type (square and circular). These 164 manually classified examples were randomly divided into a training set and a test set, with the typical 70 %/30 % split. This means the training set has 124 examples and the test set has 40 examples.

Features were extracted from each ROI and used as input for the classifiers. The evaluated features were: standard deviation, histogram, HoG and FFT. Figure 5 shows a visualization of the features considering square ROIs and circular ROIs, extracted from the image of the test plates shown in Fig. 4. It is interesting to observe that Fig. 5 seem to indicate that the histogram may be a particularly useful feature to detect signs of agglutination, in both types of ROIs. It appears that ROIs with agglutination produce histograms typically with a binomial distribution and values that are more spread out in the pixel value range.

Summarizing the experiments procedure, all combinations of ROI types (square and circular), features (standard deviation, histogram, HoG and FFT) and classifiers (Bagging, Random Forests, LPBoost and AdaBoost) were evaluated. In the following section, results are presented for all these combinations.

3.1 Decision trees

The first experiments considered only square ROIs. The same training set was used to train four types of classifiers based on decision trees: Bagging, Random Forests, LPBoost and AdaBoost. For each type of classifier, four different sets of classifiers were trained, one for each of the features: standard deviation, histogram, HoG and FFT. For each combination of classifier/feature, an increasing number of trees was used for building the classifiers, ranging from 1 to 50 trees. The performance was evaluated using the F1-score [67] which combines precision and recall into one evaluation metric. Higher F1-scores, means we have a better classifiers. Figure 6 shows the performance of the different types of classifiers using the different features for increasing numbers of trees. Figure 6-(I) shows the F1-score of the four types of classifiers trained and tested only with the standard deviation feature, Fig. 6-(II) shows the performance of the classifiers that only use the histogram

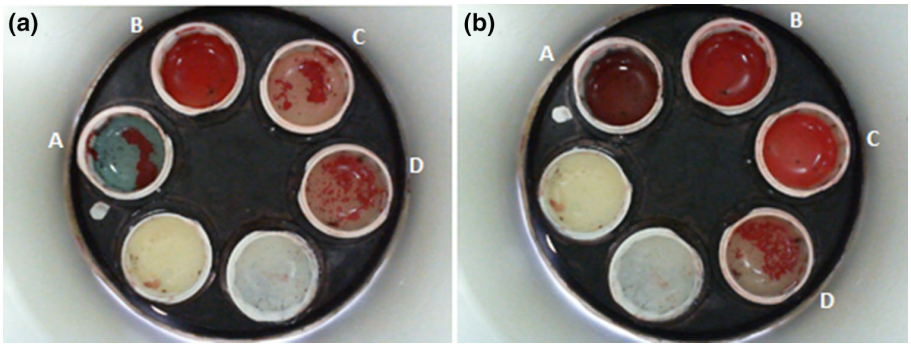
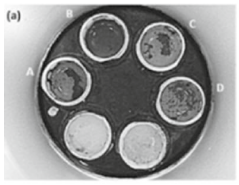
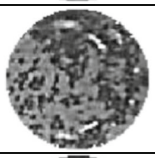
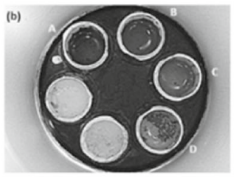
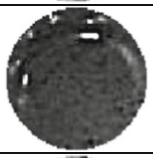
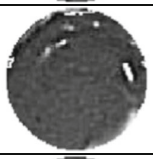
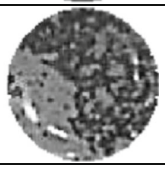


Fig. 4 Captured images with HD camera of two tests performed with the prototype

Table 1 Processed images by prototype's software, square ROIs 1 extracted, circular ROIs 2 extracted, SD 1 achieved with square ROIs, SD 2 achieved with circular ROIs and classification results of agglutination by a technical analyst

Images Processed From Fig. 4	Samples	ROIs 1	ROIs 2	Agglutination Results
	A			**1**
	B			0
	C			**1**
	D			**1**
	A			0
	B			0
	C			0
	D			**1**

SD standard deviation

as a feature, Fig. 6-(III) shows the F1-score of the classifiers that use HoG feature and Fig. 6-(IV) shows the F1-score of the classifiers that use FFT feature.

From Fig. 6-(I), it can be seen that, using the standard deviation, none of the classifiers was able to reach the optimal F1-score of 1, no matter how much the number of trees is increased. AdaBoost gets very close, but not quite, and performance is even degraded when the number of trees is higher than 40. In Fig. 6-(II), the Histogram looks like a promising feature, with reasonable and consistent results for all types of classifiers, but again, none of the classifiers reaches the perfect result. Figure 6-(III) shows that HoG is not at all suitable for the task, no matter what type of classifier is used. Figure 6-(IV) shows that FFT yields very interesting results for all types of classifiers, the best so far for square ROIs, particularly for AdaBoost, as

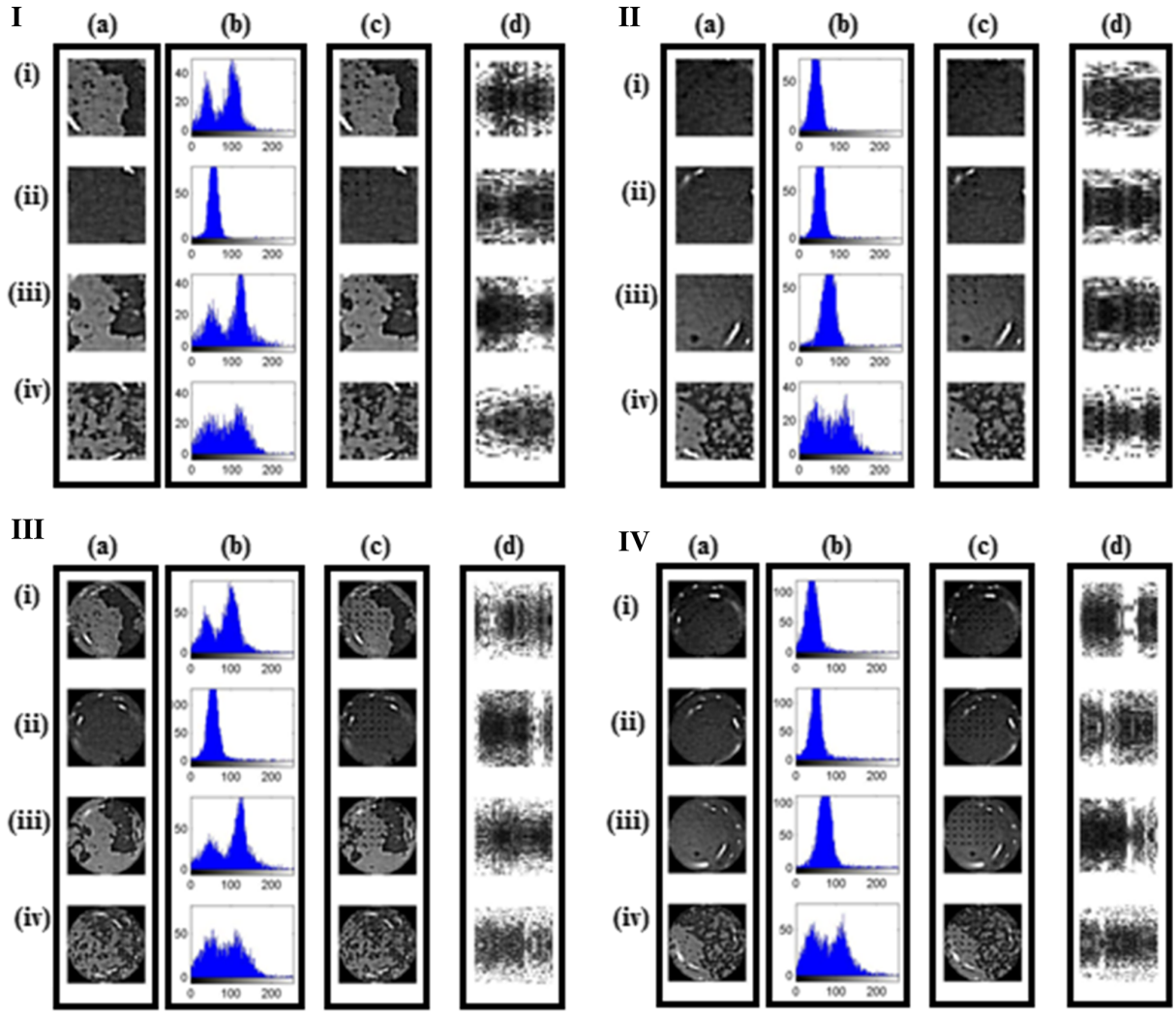


Fig. 5 Results of application of the different features to the ROIs. [*I* and *II*—Square ROIs. *III* and *IV*—Round ROIs. (*a*) Image ROIs. (*b*) Histogram feature applied to the ROIs. (*c*) Hog feature applied to the ROIs. (*d*) FFT feature applied to the ROIs]

long as a reasonable number of trees are used. It can be therefore concluded that the combination of FFT and Adaboost could be a good choice for square ROIs.

The procedure described above was repeated, but this time circular ROIs were considered instead of square ROIs. All combinations of features and classifiers were tried, applying the same approach of using increasing numbers of trees for each combination of feature/classifier, but this time extracting features from circular RIOs. For the features that assume square input patches (HoG and FFT), they were computed on the largest possible square that could be fitted into the circular ROI. Results, using the F1-score as the performance metric, can be seen in Fig. 7.

As Fig. 7 shows, using circular ROIs has an overall positive impact on the detection of agglutination, as more pixels of the container are available for feature extraction. In Fig. 7-(I) using only the standard deviation as a feature, Adaboost continues to be the best choice, in fact producing the best results so far with an F1-score consistently equal to 1; although performance degrades if more than 40 trees are used, but the other classifiers have also improved their results. Also in Fig. 7-(II) most classifiers show better results than in Fig. 6-(II). In Fig. 7-(III) results have improved with respect to Fig. 6-(III), but HoG is still not a viable option. In Fig. 7-(IV) results are in general worse than using square ROIs in Fig. 6-(IV), particularly for LPBoost. In summary, Fig. 7 shows that using circular ROIs seems to help in most cases, standard deviation is a viable option as a feature and that AdaBoost performs better than the other classifiers for most features. The

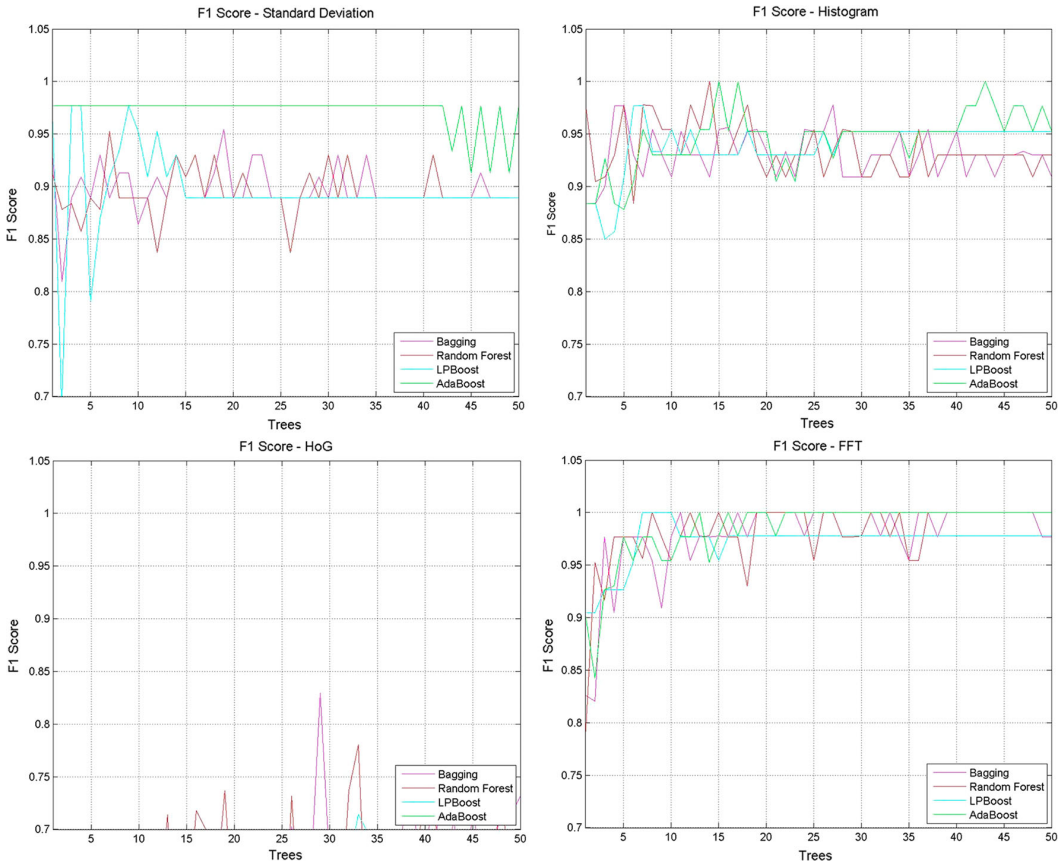


Fig. 6 F1-score results obtained for square ROIs with different functions [64]

performance degradation with high numbers of trees in the classifier structure may be due to the dataset size, which is not very large.

3.2 SVMs

The last experiment was to use linear SVMs as classifiers, instead of classifiers based on decision trees. Both square and circular ROIs were considered and the used features were also standard deviation, histogram, HoG and FFT. Results for the F1-score in this setting are shown in Table 2.

The results in Table 2 show that using circular ROIs continues to help with most features except for the FFT, as in the experiments with Decision Trees, since more information is extracted from the bigger image patches. HoG continues to be a bad choice as a feature. The histogram is the optimal feature for both types of ROIs and the standard deviation produces perfect results when extracted from circular ROIs. Overall, results are better than the ones achieved with Decision Trees.

4 Conclusions and future work

This paper presents a prototype developed for the determination of blood pre-transfusion tests. Once this prototype is able to perform the tests in a short period of time, its main focus is for emergency situations. To fit urgent scenarios, the prototype is portable so it can be used even in remote locations, it is efficient and, also, a low-cost device being suitable for undeveloped countries. With this portable prototype, it will be possible to perform tests in remote places like ambulances, remote accident sites, and in natural disaster areas (tornadoes, earthquakes, tsunamis, wars, among others.). It will be possible to avoid traveling to the laboratory to perform the tests. Moreover, it will also be possible to confirm whether the blood unit sent by the laboratory to the place where the transfusion will be performed, is a unit compatible with the patient.

The blood type classification is based on image processing techniques used in the software application developed to process the initial captured image and give the test results. The test data are stored in a database and may

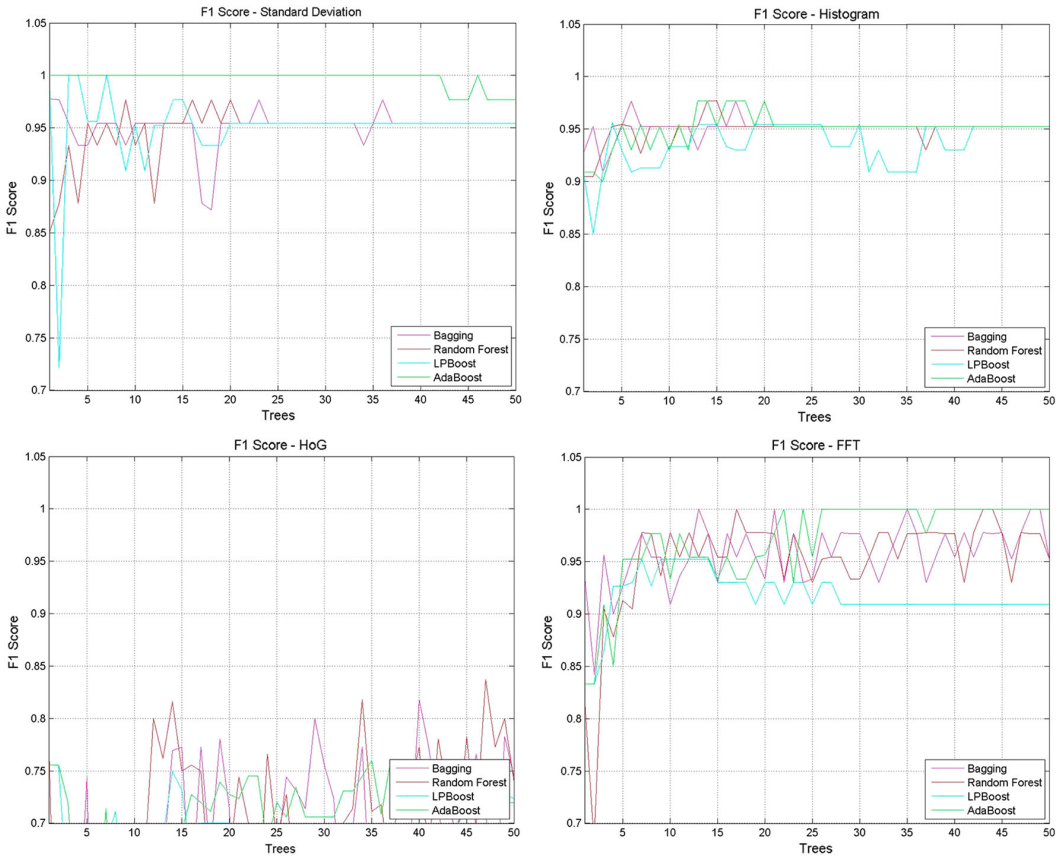


Fig. 7 Round ROIs and the Tree results

Table 2 Results of F1-score with SVMs

F1-score (square ROIs)		F1-score (round ROIs)	
SD	0.9767	SD	1
Histogram	1	Histogram	1
HoG	0.5263	HoG	0.7917
FFT	1	FFT	0.8718

subsequently be accessed and used without loss of information.

The prototype has proven to be effective in detecting agglutination in the blood samples and correctly determining the blood type. Classification was performed using two alternative approaches based on Machine Learning algorithms, Decision Trees and SVMs. Through these different classifiers, it was possible to achieve accurate blood type results with some parameters. With the relatively small dataset used, the achieved accuracy level of the system was perfect for some of the combinations of ROI types, features and classifiers. Relatively to the results obtained with decision trees, AdaBoost is the type of classifier that achieves better results, either with square ROIs, or with circular ROIs. AdaBoost has better results with standard deviation values and with FFT values as features, both with square ROIs and circular ROIs. With circular ROIs, the results using standard deviation are better than with square ROIs, reaching an F1-score value of 1. If FFT is used, the ROI type does not seem to make a significant difference. With SVMs, square ROIs, the histogram and the FFT produce better results, with an F1-score value of 1 in both cases. For circular ROIs, the standard deviation and the histogram allow achieving better results, with an F1-score value of also 1 for both cases. Overall, it may be concluded that circular ROIs will be better for classification because the area that is analyzed is larger and it is possible to use data from the whole container. So, taking into account these observations, the right combination seems to use circular ROIs, the histogram as a feature and a linear SVM as a classifier. Circular ROIs enable the system to extract more information from the

image patches, which is visible in the experimental results. The histogram was the feature that produced the most consistent results for most experiments, especially with SVMs. SVMs yielded the best classification performance, out of all the considered classifiers.

In the future, it is planned to perform more tests in the hospital in order to expand the dataset and make other training and classification with Machine Learning techniques, as some of the experiments seem to indicate that a bigger training set would possibly improve the results. With these future tests, more training data will be obtained, and further experiments will train the classifiers (Decision Trees and SVMs) with the expanded dataset data, and also try using different types of classifiers. The use of color images is also in the plans for future research. It should also be possible to introduce a measure of confidence of classification into the system to express how confident the system is to classify the blood type. This should be possible by checking the margin between a given test example and the separation hyperplane in the SVM classifier. The adaptation of the prototype to detect other diseases is also a possibility.

The obtained results are considered satisfactory up to now. These results reinforce the belief that Artificial Intelligence approaches are useful tools to assist in the classification of results. These techniques will thus be an asset in order to obtain a more efficient and reliable blood type classification, based on the classification approaches presented.

Acknowledgments Authors of this paper want to thank Portuguese Foundation for Science and Technology (FCT) for funding through the PhD scholarship SFRH/BD/81094/2011. This work is funded also by FEDER funds through the "Programa Operacional Factores de Competitividade—COMPETE" and by national funds by FCTFundação para a Ciência e a Tecnologia, project reference PEst-UID/CEC/00319/2013.

Compliance with ethical standards

Conflict of interest The authors declare that they have not conflict of interest.

References

1. Roback JD, Grossman BJ, Harris T, Hillyer CD (2011) Technical manual, 17th edn. American Association of Blood Banks
2. Hammering DM (2012) Modern blood banking and transfusion practices, 6th edn. F.A. Davis Company
3. Rod SR, Tate P, Trent DS (2005) Anatomia and fisiologia, 6th edn. Lusociência, Loures
4. Hoffbrand VA, Pettit EJ, Moss HAP (2004) Fundamentos em Hematologia, 4ª Edição. Artmed, Porto Alegre
5. Caquet R (2004) Guia Prático Climepsi de Análises Clínicas, 1st ed. Climepsi Editores
6. Blood Group Terminology (1935) ISBT—international society of blood transfusion. http://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/blood-group-terminology/
7. U.S. Congress, Office of Technology Assessment, OTA-H-260 (1985) Blood policy and technology. Library of Congress Catalog Card Number 85-601151. U.S. Government Printing Office, Washington, DC
8. Sturgeon P (2001) Automation: its introduction to the field of blood group serology. Immunohematology 17(4):100–105
9. Coakley AW (1981) Handbook of automated analysis. Mercel Dekker, New York
10. Ewing GW (1997) Analytical instrumentation handbook, 2nd edn. Marcel Dekker, New York, p 152
11. The Technicon Auto Analyzer II (2014). http://hydrology1.nmsu.edu/Teaching_Material/SOIL698/Student_Material/AutoAnalyzer/Autodiag.html
12. Garretta M, Gener J, Muller A, Matte C, Moullec J (2000) The groupamatic system for routine immunohematology. Transfusion 15:422–431
13. Zaccarelli GD, Monti G, Malaguti J, Marchesini D, Figliola F, Cagliari G, Basile C (2000) Esperienza di automazione nella determinazione dei gruppi sanguigni. La Trasfusione del Sangue 45:28–31
14. OLYMPUS TECHNOZONE Vol. 67 2006-11 (2011). http://www.olympus-global.com/en/magazine/techzone/vol67_e/page5.cfm
15. Serodia-TPPA Auto (For In Vitro Diagnostic Use). http://www.mastgrp.com/Fuji/IFU/TPPAauto.pdf
16. Formulated for use in Automated System Olympus® PK® Systems"—BLOOD GROUPING REAGENTS (2007). http://www.fda.gov/downloads/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/BloodDonorScreening/BloodGroupingReagent/ucm081307.pdf
17. Immucor. http://immucor.com/site/aum_company_profile.jsp
18. Wittmann G, Frank J, Schram W, Spannagl M (2007) Automation and data processing with the Immucor Galileo® System in a University Blood Bank. Transfus Med Hemother 34:347–352
19. Dada A, Beck D, Schmitz G (2007) Automation and data processing in blood banking using the Ortho AutoVue® Innova System. Transfus Med Hemother 34:341–346
20. Anti-Human Globulin, 2008. http://www.fda.gov/downloads/BiologicsBloodVaccines/BloodProducts/ApprovedProducts/LicensedProductsBLAs/BloodDonor/BloodGroupingReagent/ucm080763.pdf
21. Shin SY, Kwon KC, Koo SH, Park JW, Ko CS, Song JH, Sung JY (2008) Evaluation of two automated instruments for pre-transfusion testing: AutoVueInnova and Techno TwinStation. Korean J Lab Med 3:214–220
22. Roback JD, Combs MR, Grossman BJ, Hillyer CD (2008) Technical manual. AABB, Maryland
23. Murphy MM, Pamphilon DH (2009) Practical transfusion medicine, 3rd edn. Wiley-Blackwell, Oxford
24. Cressiers/Morat (2008) Datasheet of Diamed Diaclon Anti-A, Diaclon Anti-B, Diaclon Anti-AB. http://diamedil.info/diamed/Inserts/100910.pdf
25. Cressiers/Morat Datasheet of Diamed: Diamed-MP Test. Bio-Rad Laboratories, Inc. http://www.bio-rad.com/en-pt/product/diamed-mp-test-b-dvi-dvi-ctl-a1-a2-b
26. Cressier M (2008) Datasheet of diamed-ID micro typing system, MTS gel card
27. Inventor Professional. http://www.autodesk.com/education/free-software/inventor-professional
28. Costa A Autodesk Inventor 2013 - Curso Completo. FCA
29. Ferraz A, Carvalho V, Soares F (2013) A Prototype for blood typing based on image processing. In: SensorDevices 2013: the fourth international conference on sensor device technologies and applications, 2013

30. Arduino Duemilanove. http://arduino.cc/en/Main/arduinoBoardDuemilanove
31. Arduino Ethernet Shield. http://arduino.cc/en/Main/ArduinoEthernetShield
32. Lehmann TM, Gonner C, Spitzer K (1999) Survey: interpolation methods in medical image processing. IEEE Trans Med Imaging 18(11):1049–1075
33. Studholme C, Hill DLG, Hawkes DJ (1999) An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit 32(1):71–86
34. Pluim JPW, Maintz JBA, Viergever MA (2003) Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 22(8):986–1004
35. Boykov YY, Jolly M-P (2001) Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: Proceedings. eighth IEEE international conference on computer vision, 2001. ICCV 2001, pp 105–112
36. Walter T, Klein JC, Massin P, Erginay A (2002) A contribution of image processing to the diagnosis of diabetic retinopathy–detection of exudates in color fundus images of the human retina. IEEE Trans Med Imaging 21(10):1236–1243
37. Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, Leahya RM (2001) Magnetic resonance image tissue classification using a partial volume model. Neuroimage 13(5):856–876
38. Ferraz A, Carvalho V, Brandão P (2010) Automatic determination of human blood types using image processing techniques. In: BIODEVICES 2010 international conference on biomedical electronics and devices
39. Ferraz A, Carvalho V, Soares F, Leão CP (2011) Characterization of blood samples using image processing techniques. Sens Actuators A Phys 172(1):308–314
40. Moreira V, Ferraz A, Carvalho V, Soares F, Machado J (2012) Design of a mechatronic system for human blood typing in emergency situations. In: IEEE international conference on emerging technologies and factory automation, ETFA
41. Bezerra K, Ferraz A, Carvalho V, Machado J, Matos J, Soares F (2012) Advanced design of a mechatronic system for human blood typing. Romanian Rev Precis Mech Opt Mechatron 41:144–150
42. Dalall N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Conference on computer vision and pattern recognition (CVPR), vol 1, 2005, pp 886–893
43. Pauker SG, Kassirer JP (1975) Therapeutic decision making: a cost-benefit analysis. N Engl J Med 293(5):216–221
44. Lin R-H (2009) An intelligent model for liver disease diagnosis. Artif Intell Med 47(1):53–62
45. Sheppard JW, Kaufman MA, Wilmer TJ (2009) IEEE standards for prognostics and health management. Aerosp Electron Syst Mag IEEE 24(9):34–41
46. Lavrac N (1999) Selected techniques for data mining in medicine. Artif Intell Med 16(1):3–23
47. Huanga M-J, Chenb M-Y, Show-Chin L (2007) Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. Expert Syst Appl 32(3):856–867
48. Kononenko I (2001) Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 23(1):89–109
49. Isern D, Sánchez D, Moreno A (2010) Agents applied in health care: a review. Int J Med Inform 79(3):145–166
50. Rocha M, Cortez P, Neves JM Análise Inteligente de Dados, 1st ed. FCA
51. Won Y, Song H-J, Kang TW, Kim J-J, And B-DH, Lee S (2003) Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons. Proteomics 3(12):2310–2316
52. Freitas A, Costa-Pereira A, Brazdil P (2007) Cost-sensitive decision trees applied to medical data. Data Warehous Knowl Discov 4654:303–312
53. Aydin I, Karakose M, Akin E (2011) A multi-objective artificial immune algorithm for parameter optimization in support vector machine. Appl Soft Comput 11(1):120–129
54. Chang C-L, Chen C-H (2009) Applying decision tree and neural network to increase quality of dermatologic diagnosis. Expert Syst Appl 36(2):4035–4041
55. Barakat N, Bradley AP, Barakat MNH (2010) Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans Inf Technol Biomed 14(4):1114–1120
56. Wu T-K, Huang S-C, Meng Y-R (2008) Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities. Expert Syst Appl 34(3):1846–1856
57. Greer BT, Khan J (2004) Diagnostic classification of cancer using DNA microarrays and artificial intelligence. Appl Bioinform Cancer Detect 1020:46–66
58. Polat K, Güneş S (2007) Breast cancer diagnosis using least square support vector machine. Digit Signal Process 17(4):694–701
59. Akay MF (2009) Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 36(2):3240–3247
60. Widodo A, Yang B-S (2007) Support vector machine in machine condition monitoring and fault diagnosis. Mech Syst Signal Process 21(6):2560–2574
61. Statistics Toolbox. http://www.mathworks.com/products/statistics/
62. Matlab Software. The MathWorks, Inc. http://www.mathworks.com/products/matlab/
63. Matlab - Machine Learning. The MathWorks, Inc. http://www.mathworks.com/machine-learning/
64. Classification Trees and Regression Trees. http://www.mathworks.com/help/stats/classification-trees-and-regression-trees-1.html
65. Support Vector Machines. http://www.mathworks.com/help/stats/support-vector-machines.html
66. Hospital Prof. Doutor Fernando Afonseca, EPE. http://www.hff.min-saude.pt/
67. Russel S, Norvig P (2009) Artificial intelligence: a modern approach, 3rd edn. Pearson, London