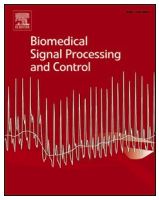
Biomedical Signal Processing and Control 81 (2023) 104489

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Anaemia detection based on sclera and blood vessel colour estimation

Giovanni Dimauro [a,*], Mauro Giuseppe Camporeale [a], Alessandro Dipalma [a], Attilio Guarini [b], Rosalia Maglietta [c]

[a] Department of Computer Science, University of Bari, Bari, Italy
[b] Haematology Dept. National Cancer Institute 'Giovanni Paolo II', Bari, Italy
[c] National Council of Research (CNR), Rome, Italy

ARTICLE INFO

Keywords:
Anaemia detection
Image segmentation
Sclera
Blood vessel

ABSTRACT

Anaemia is a global public health problem with major consequences for human health. Noninvasive methods must be investigated to determine a sick person's anaemic status or conduct screening campaigns, especially in resource-constrained areas of the earth. This study aims to prove that the colour of the sclera and scleral blood vessels extracted from digital images of the eye can be used to check the anaemic status of a person. To date, we have not found in the literature other studies that have attempted this promising approach. We propose a novel pipeline for anaemia estimation consisting of three main contributions: a sclera segmentation algorithm applied to near-taken digital photos of the eye, a vessel extraction algorithm, and a classifier to predict the anaemic status of a person vs normal controls. This study was based on the public dataset Eyes-defy-anaemia, which contains 218 eye pictures taken with a special device that removes any ambient light influence. Very interesting results have been achieved for the sclera segmentation task with good precision (88.53), recall (82.53) and F1 (84.10). The colour features and haemoglobin value appear to be well related allowing us to obtain an F2 score in the anaemia detection task of 86.4% using colour features from the whole sclera and of 83.8% using only vessels' colour features.

1. Introduction

Several diseases show clinical signs on the sclera. The most common is jaundice, which is caused by the accumulation of bilirubin, a breakdown product of red blood cells, in the blood. The yellow-coloured bilirubin gives rise to the characteristic yellow discolouration of the skin of jaundiced patients, but the presence of melanin in the skin makes objective evaluation difficult if based on only a colour evaluation. Meanwhile, the sclera and overlaying conjunctiva are free of melanin and have a high affinity for bilirubin [1]. Mariakakis et al. [2] built a smartphone-based system to detect bilirubin levels through the sclera. Some studies reported alterations of the bulbar conjunctiva vasculature in the presence of sickle cell anaemia [3,4]. Brooks et al. [5] reviewed all syndromes and diseases that have been reported to show blue sclera among the symptoms: the sclera does not actually change colour, but primary collagen defects of the sclera lead to thinning and translucency, which results in a blue to blue–grey appearance.

At the moment the standard clinical method to diagnose Anaemia primarily rely on the invasive determination of blood Hb, which requires venepuncture performed by a nurse or a physician. Frequent blood sampling leads to the patient's discomfort, and the requirement of going to a laboratory or calling a nurse involves considerable expense.

Currently, research is very active in finding new noninvasive methods to detect anaemia [6,7], in fact, it is of great interest to study methods and to design instruments to monitor the concentration of Hb in a non-invasive way, with reduced costs, both in the laboratory and at the patient's home, sometimes even daily. Furthermore, standard laboratory process requires specialized equipment not always available as an example in rural and low-resource settings, where anemia is most prevalent.

Many studies have been conducted to reveal anaemia or regress the haemoglobin concentration level from the palpebral conjunctiva (PC) or other tissues pallor measured on digital images. Some of these methods use special instruments [8], while others use digital photos [9,10]. The use of digital photos taken from a camera or a smartphone has the great advantage of being noninvasive, since it does not require any blood sample; furthermore, if the picture is taken from a smartphone, it can be instantly analysed. However, colours in a picture can be largely

* Corresponding author.
E-mail address: giovanni.dimauro@uniba.it (G. Dimauro).

https://doi.org/10.1016/j.bspc.2022.104489
Received 6 July 2022; Received in revised form 12 November 2022; Accepted 1 December 2022
Available online 9 December 2022

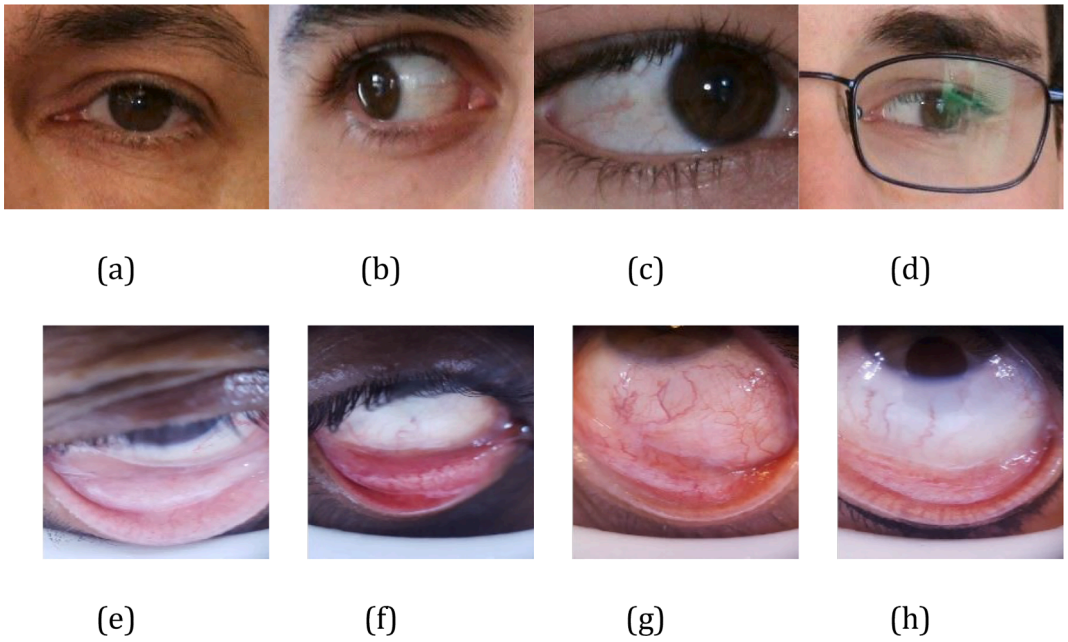


Fig. 1. (a, b, c, d) Typical pictures to test the sclera segmentation algorithm from the UBIRIS dataset. (e, f, g, h) Pictures from the dataset for this work.

influenced by ambient light and various reflections due to the moistness of the eye. Dimauro et al. [6] almost solved the problem of obtaining good-quality photos of the conjunctiva by building a low-cost device that enabled any smartphone camera to take ambient-light-independent pictures.

The analysis of PC to detect anaemia or iron deficiency is not a new outcome; qualitative analysis has been a common practice for physicians for decades [11]. Given the proven correlation between pallor of the conjunctiva and anaemia [12], this work focuses on the study and verification of the hypothesis that the paleness of the sclera and blood vessels in the sclera may also be considered symptoms of anaemia. In particular, a noninvasive pipeline devoted to anaemia diagnosis based on digital images of the eyes is presented here and consists of three main contributions: a sclera segmentation algorithm applied to near-taken digital photos of the eye, a vessel extraction algorithm, and a classifier to predict the anaemic status of a person vs normal controls. Furthermore, specific features are selected and the dataset used in this study, Eyes-defy-anaemia [13], which well captures the sclera. This dataset has been made publicly available recently and is the first public dataset and the only one in the world dedicated to this topic, thus in this paper we exploit a novel dataset, never been used before.

The paper is organized as follows. Section 2 introduces the use of sclera features for other purposes. In Section 3, we describe the methods and dataset for the experiments. Section 4 reports the experimental results. In Section 5, we discuss the results and in Section 6 conclusion are drawn.

2. Related works

To the best of our knowledge, no current study shows that the colour of the sclera and its blood vessels correlates with haemoglobin, while the literature mainly focuses on segmentation related to biometric tasks [14] or other tissues, such as the palpebral conjunctiva [15].

Several studies have tried some invasive and non-invasive approaches to observe the blood flow within the sclera; One is using indocyanine-green injection to the anterior chamber of the eye to visualise the blood flow [16], this study was very unique, but it was invasive to human eye. More recently, optical coherence tomography angiography (OCTA) has been applied to the anterior segment; the OCTA approach can visualise not only retinal blood flow [17] but also anterior ocular circulations (i.e. sclera, iris) [18]. However, the OCTA device results very expensive.

The previously developed sclera segmentation algorithms are calibrated on eye pictures taken in what we can define as “normal” conditions: variable ambient light, camera-eye distance greater than 30 cm, eye opened in a natural position, and variable iris position. As an example, Fig. 1 (a, b, c, d) shows some typical pictures to test the sclera segmentation algorithm from the UBIRIS dataset [19].

State of the art algorithms for sclera recognition rely mainly on the following features:

1. Eye contour. Is delimited by eyelashes and palpebral skin, which are darker than the white sclera;
2. Iris position. The iris is dark, circular, and has well delimited contours. So, it is an object with a standard shape and its colour is darker than the sclera.

The detection of these two features is not an easy task to be performed on the pictures from our dataset, since:

- In many pictures, where the palpebra is stretched, the more vascularized part of the bulbar conjunctiva (which covers the sclera) softly fades into the palpebral conjunctiva, so it becomes harder to define sharp contours following the white scleral colour. An example is given in Fig. 1h;
- Iris positioning cannot be used because in most photos, it is only partially visible, such as in Fig. 1g.

Fig. 1 (e - h) shows some samples on which such features are ineffective.

Despite these complications, we believe that photos such as those included in the Eyes-Defy-Anaemia dataset [13] are more suitable for our study, since they are more focused on the sclera and its vessels. It is indeed easier to extract accurate colour features from higher-resolution pictures, which are taken from a very close distance, show a wide-open eye, and are as light-independent as possible.

(a) (b)

Fig. 2. (a) Acquisition device; (b) moment of acquiring an image.

Table 1
Distributions of anaemia in relation to sex.

	Total	Healthy	Anaemic	Not known
women	86	33	53	0
men	132	93	38	1

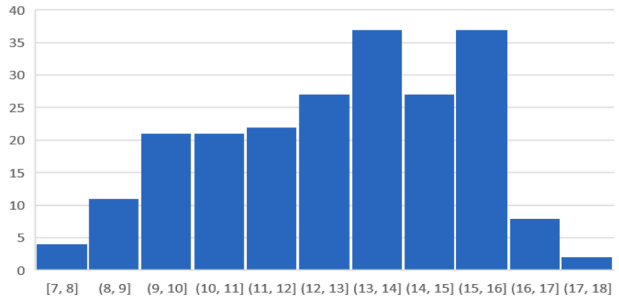


Fig. 3. Distribution of haemoglobin values in the dataset.

3. Materials and methods

3.1. Dataset

The dataset Eyes-defy-anaemia includes eye images, where a large part of the eye area is well exposed, all taken by a single low-cost smartphone to which a particular acquisition device has been mounted. It consists of a macro lens (Aukey PL-M1 25-mm 10x macro lens) applied on an opaque spacer, inside which white LED lights have been inserted (Fig. 2(a)). This tool solves the problem of heterogeneous shooting conditions because it fixes the distance from the eye and standardizes the light conditions by preventing external light from affecting the shot (Fig. 2(b)) [7]. Immediately after taking the photo, the people were subjected to blood sampling to obtain the haemoglobin values measured with standard laboratory equipment.

For each person, the Eyes-defy-anaemia dataset includes photos of the original eyes, only segmented palpebral and bulbar conjunctivae. With the help of medical personnel, we manually segmented the sclerae in the images from the above dataset. The Eyes-defy-anaemia dataset consists of photos of 218 patients organized as shown in Table 1.

The ages of all subjects were 19–88 years with an average of 42 years. The haemoglobin values were 7–17.4 g/dL with an average of 12.7 g/dL and are distributed as shown in Fig. 3.

3.2. Proposed algorithm for sclera segmentation

The first part of our study consisted in the sclera's ROI extraction. In this section, we describe the developed algorithm. As shown in Fig. 4, the algorithm can be split into 4 blocks: picture enhancement (yellow box), thresholding (blue box), k-means clustering with normalized cuts (red box), and iterative refinement (green box).

3.2.1. Picture enhancement

Due to the ocular bulb shape and direction of light, peripheral portions of the sclera are usually darker than the centre. To avoid errors during the k-means clustering phase, the automated Multi-Scale Retinex with Colour Restoration (MSRCR) proposed by Parthasarathy et al. [20] is applied to the eye images. This algorithm, with almost no modification, has been implemented and used to enhance the pictures. It particularly met our needs because it well recovers the colour and contrast.

The next issue to face is the influence of blood vessels: since they are red like the palpebral conjunctiva, they were clustered as part of it in several pictures. This "vessel noise" is removed by applying a simple Gaussian blur filter to the picture. Nevertheless, due to the Gaussian filter, we also lose edge contrast, and this effect of the Gaussian filter is counteracted by subsequent contrast and sharpening manipulation. Fig. 5 shows the full enhancement process. Retinex Filtered picture (Fig. 5b) will be used in the thresholding step as shown in paragraph 3.2.2, while the picture obtained after the exposure and contrast adjustment (Fig. 5d) will be used in the k-means clustering with normalized cuts as shown in paragraph 3.2.3.

3.2.2. Thresholding

Global thresholding was applied to the eye images. The threshold values were determined by statistical analysis using the manually segmented sclerae as ground truth images. The threshold operation was performed on the Retinex Filtered pictures (Fig. 5b).

To improve the thresholding mask precision, morphological operators opening and closing are used. The closing removes small holes, which are mainly caused by blood vessels in our thresholding case.

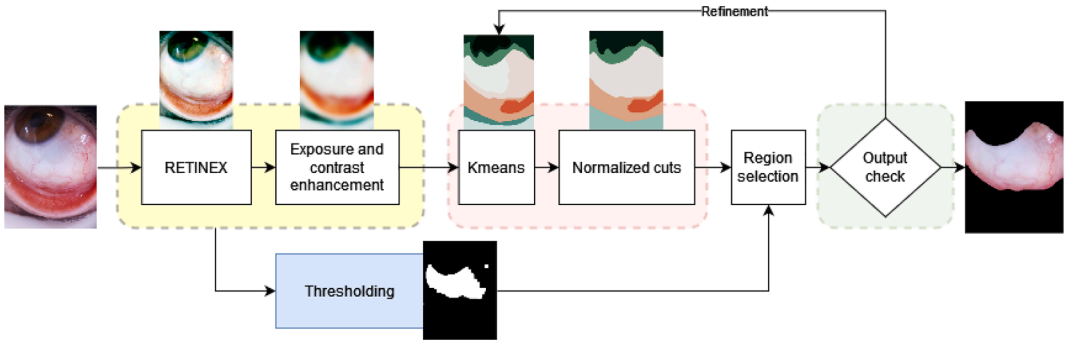


Fig. 4. Algorithm flow diagram.

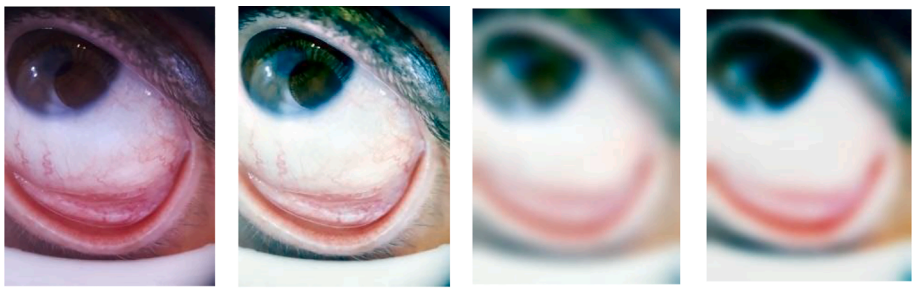


Fig. 5. Full picture enhancement process: original picture (a), retinex-filtered picture (b), picture after the Gaussian filter (c), picture after exposure and contrast adjustment (d).

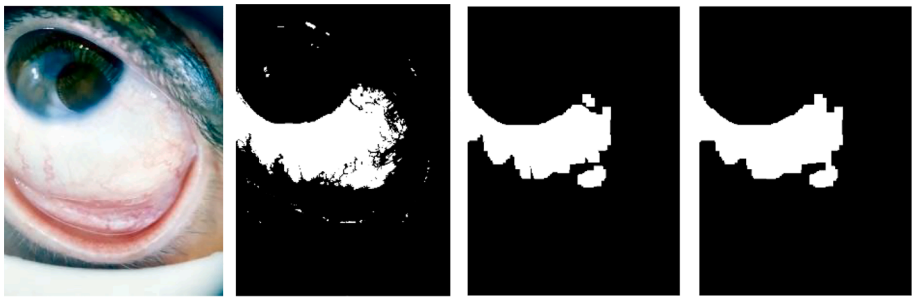


Fig. 6. Thresholding steps: Retinex filtered picture (a), global thresholding (b), opening (c), and closing (d).

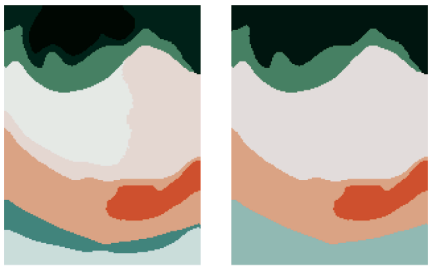


Fig. 7. K-means clustering output (a) and Ncuts refinement (b).

The mask obtained through the thresholding step supports the choice of the correct cluster (see next paragraph); therefore, we are not concerned about whether the output mask corresponds to the entire sclera region, but we are concerned about the mask precision, since the selected cluster must cover the entire sclera.

We observe that having fixed thresholds did not perform well, so we perform this operation through a dynamic procedure: we start with high thresholds on the RGB channels; if the mask results are too small (less than 5 % of the entire picture), we reduce the threshold by a factor of 5 %. The percentages were experimentally selected by analysing the ratios of the sclerae size compared to the entire ground truth picture size.

3.2.3. *K-means clustering and normalized cuts*

The principles from k-means clustering have been applied to image segmentation tasks. Each pixel of the preprocessed image (Fig. 5d) is considered as a vector in a five-dimensional space: x and y coordinates from the matrix, R, G and B channel intensities from colour representation. KMeans as a clustering algorithm is a valuable approach to use local impressions of a scene, but it lacks a global or hierarchical perspective. Thus, we use a grouping algorithm to treat the segmentation task as a graph partitioning problem such as Ncut. The set of superpixels from a five-dimensional feature space is converted into a weighted undirected graph. Fig. 7 shows the clustering output for a picture from the dataset.

Then, the Ncut output is matched with the threshold (Fig. 6d). The region with more pixels included in the threshold will constitute the final mask. Other regions matching the threshold are included if the colour is sufficiently close to the main region. The distance between the two colours is computed as the Euclidean distance in the CIELAB colour space because it minimizes the distance between colour values that the clustering associated with the sclera region. Thus, darker regions of the sclera that match the threshold in small portions are not lost. The final mask is refined by a closing transformation, since very large blood vessels or spots may be excluded because they constitute a large dark region inside the sclera.

3.2.4. *Improving precision*

This algorithm does not always select the better cluster. Clusters that belong to the conjunctiva are sometimes included in the final mask. To manage these cases, we evaluate the colour of the selected region. We reject the segmentation if the standard deviation of the alpha channel is greater than a fixed threshold or L, a* and b* are out of the 25th – 75th percentile range. These thresholds were determined by analysing the ground truths. If the segmentation is rejected, Ncut on the current segmentation is applied again to cut out the incorrectly included clusters. At the end of this process, the iteration that shows the highest green channel average is elected to be the final output. This choice is made because the green channel average is lower if a dark region is included in the final result, and our target is to consider the lightest parts of the

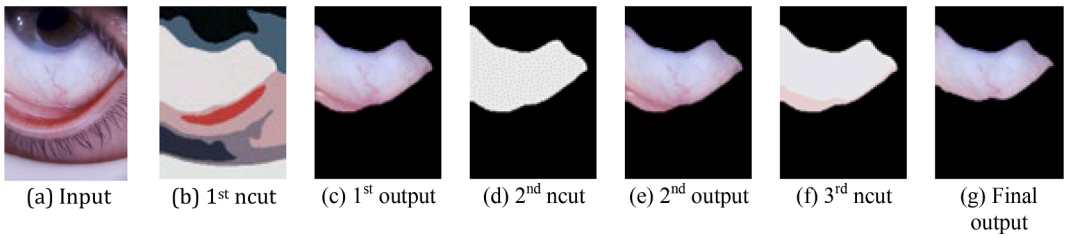


Fig. 8. Refinement process.

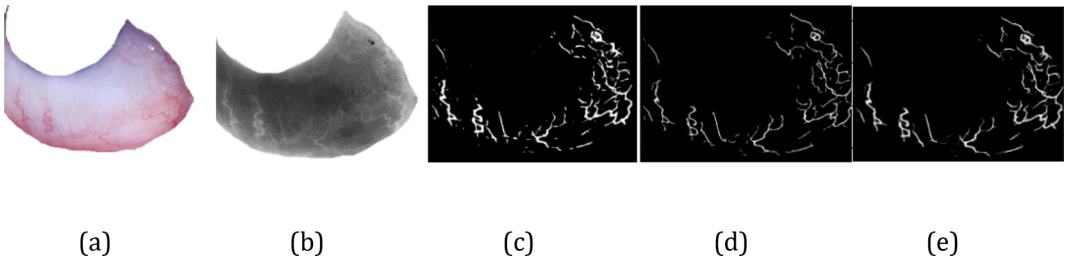


Fig. 9. Original image (a); inverted green channel (b); line detection result (c); vessel centerline tracing result (d); union result (e).

Fig. 10. Isolated segments problem: particular of Fig. 9c (a); same particular from Fig. 9e (b).

picture to better estimate the colour characteristics. Fig. 8 shows the refinement process on a picture.

3.3. Vessel detection

To identify blood vessels in the sclera, both a weighted line detection algorithm and a hidden Markov model (HMM) are used here; the first one is used to identify major vessels inside the sclera, while the latter identifies the centrelines of the vessels, which enables us to also extract thinner vessels that were not captured by the line detector. After the light post-processing, the mask obtained by line detection and HMM are combined to obtain the final vessel mask.

3.3.1. Weighted line detection

We have considered the green channel of the segmented sclera, which shows the greatest contrast between the vessels and the background; in particular, the process was performed on the inverted green channel (Fig. 9b). No pre-processing was necessary, since sharpening operations on the image would have added noise, while noise reduction operations would have made the image less defined. We choose the algorithm presented in [21], which improves the one in [22].

The algorithm [21] works in a window of size W around pixel p and calculates the average intensity to obtain the background intensity in the window; the intensities relating to the pixels identified by the 12 Boolean masks, which represent 12 lines at different angles in the window, are calculated to assign each pixel a weight; this weight is inversely proportional to the distance of the pixel from p. The "winning line" has the greatest weighted average. Nevertheless, one problem is evident: there are many isolated segments that are separated from the others, especially in correspondence with the thinnest vessels (Fig. 10a).

3.3.2. Vessel centerlines

This problem is solved by using a hidden Markov model (HMM) to trace the central lines of the vessels, so it includes the thin vessels that were previously excluded and combines the results of the line detection with those of the HMM. The method that we use takes inspiration from the one proposed in [23], where the HMM is used to detect axons and dendrites in neuronal images. The following operations are performed:

- a thresholding operation eliminates most of the noise in the background;
- The image is divided into blocks of 11 × 11 pixels, and a "seed" for each block is identified as the pixel with the greatest intensity;
- The initial tracing direction of each "seed" is determined by looking for and selecting the point of maximum relative intensity on the perimeter of a circumference of a predetermined size and centred in the "seed".

Once the "seeds" have been defined, you can proceed with the actual tracing of the vessels.

The center lines of the vessels are defined as a sequence $X = \{x_1, x_2, \ldots, x_n\}$ of latent and one-dimensional variables and a related set $Y = \{y_1, y_2, \ldots, y_n\}$ of observation variables thus, the latent variable $x_i = (c_i, r_i)$ denotes the position of the i-th plot point, while yi is the relative observation value denoted by the current and previous plot point. c_i and r_i correspond to the coordinates, relative to column and row, of x_i. It is therefore deduced that each plot point x_i depends on the plot points preceding it, and that the relative variable y_i is dependent on the corresponding plot point and transversely on all preceding ones.

Instead, the HMM assumes that the state of the current latent variable x_i depends only on the state of the one preceding it, so that the observation y_i also depends only on the current and preceding latent variable.

The algorithm ends the tracing from the current "seed" to move on to the next one only when the current tracing point corresponds to a pixel with lower intensity than those in a rectangular region of fixed size (5x7 pixels, as in the original work), which are oriented in the same direction as the last search angle surrounding the current point of tracing.

Once all tracing points have been obtained, they are connected to obtain an image that outlines the central lines of the blood vessels (Fig. 9d). As clearly visible, the problem reported at the end of the

previous paragraph no longer persists (Fig. 10b).

3.3.3. Post-processing

Light post-processing is performed to reduce the noise caused by vessels that are too small or too thin. These are recognized by the line detector and the HMM but appear as small and isolated lines compared to those that are the major vessels, which creates noise in the image. This type of noise can be eliminated by removing those regions with surface area less than $\delta 1/\delta 2$. This denoising process is performed on the image produced by the weighted line detector with $\delta 1 = 40$ pixels and on the product of the tracing of the central lines of the vessels with $\delta 2 = 5$ pixels. Both $\delta 1$ and $\delta 2$ are experimentally found and strictly related to the number of pixels (1067x800) of the images of the Eyes-defy-anaemia dataset. These same thresholds can also be used in a working system that uses the aforementioned low-cost device or a similar one, so they can be used as definitive thresholds. After the noise reduction process has been applied on both images, they can be merged to obtain the final vessel segmentation (Fig. 9e).

Finally, in this type of application, the complete identification of the blood vessels is not necessary, but we can limit the identification of the most evident vessels because they can provide a fairly reliable colour estimate.

3.4. Extracted features

Previous studies underline how haemoglobin is correlated with pallor [12]; then, we extract colour features and not consider other features such as the shape or other bidimensional characteristics. RGB, HSL and CieLAB colour spaces were used in this study. Specifically, for both the entire sclera and extracted vessels, we determined the average:

- R channel value;
- G channel value;
- B channel value;
- L* channel value;
- a* channel value;
- b* channel value;
- H channel value;
- channel S value;
- L channel value;
- R-G channel value (defined as EI, erythema index)

3.5. Classification

In this study, several classification algorithms were applied to detect anaemia through sclera and vessel colour features. To test each method, the implementation provided by the scikit learn library was used [24]. A grid search was performed to try all possible combinations of a classifier among the following:

- SVM with C parameter in range (1, 100) with a step of 20; kernel between “linear”, “rbf” and “poly”; gamma between “scale” and “auto”; degree of 2–5;
- KNN with neighbours in the range of 1–10 and weights between “uniform” and “distance”;
- Random forest with the criterion between “Gini” and “entropy”; max_features between “auto”, “sqrt”, and “log2”; n_estimators in the range of (1, 100) with a step of 20;
- AdaBoost with n_estimators in the range of (1, 100) with a step of 2

4. Results

4.1. Results on sclera segmentation

To evaluate the performance of the algorithm for sclera segmentation, all 218 pictures from the dataset were manually segmented; these segmentations were used as ground truth and compared with the masks obtained through the process described in Section 3. Performance was measured using the precision, recall, F1-score, accuracy and Jaccard metrics, and the results are shown in Table 2.

Table 2
Scleral segmentation algorithm performance scores.

	Precision	Recall	F1	Accuracy	Jaccard
Average	88.53	82.53	84.10	73.58	94.2
Std dev	13.29	12.10	9.39	12.58	4.25

Table 3
Top 3 performing algorithms in SSBC.

Segmentation Model	F1	Precision	Recall
UNet-P	86.8	90.9	83.1
FCN8	85.4	82.0	89.0
RGB-SS-Eye-MS	83.6	91.7	76.9

Of course, these results can vary depending on the tuning of the hyperparameters of the pipeline. The optimal combination, as shown in Table 3, was empirically found by referring to the highest precision with the higher recall, which avoided the inclusion of zones from the palpebral conjunctiva. In fact, we will extract colour features from the sclera, which would be less affected if something is left out for any example because the sclera colour is still the same. Segmented pictures with low recall usually lack some shadowed or thin parts, as shown in Fig. 11. However, even in this case, similar to the case of vessels, the entire sclera does not have to be segmented with the utmost precision; instead, the segmented part is uniform in colour, free from shadows or other noises (cilia, etc.); therefore, it is a good representative for the extraction of colour characteristics.

The algorithms in the literature [25] for the segmentation of the sclera mainly refer to photographs of the sclera at distance, as explained in Section 2, while the images in the dataset here are quite different. It is impractical to test those algorithms on our dataset and vice versa. Although algorithms using different data should not be compared, for illustration purposes, Table 3 shows the algorithms performances for SSBC 2020 [26] and highlights that the proposed algorithm has reasonable performances with respect to state-of-the-art algorithms.

4.2. Results of vessel segmentation

The vessel segmentation algorithm has been qualitatively evaluated by all authors and supported by medical personnel, since the ground truth is not present in the Eyes-Defy-Anaemia dataset and a new manual vessel segmentation is excessively demanding. Furthermore, segmenting all vessels is not possible, since capillaries are difficult to detect, and one works with so few pixels that even an insignificant variation of some dozens of pixels can alter the score of the evaluation metrics. Complete vessel identification is not necessary, since we can limit the identification of the most evident vessels, which can provide a fairly reliable colour estimate. Thus, the vessel segmentation result is considered very satisfying, and some examples are reported in Fig. 12.

4.3. Classification results

After many experiments, we finally submitted the average EI value and average a* value to each classifier among all extracted features because they allow us to obtain superior performances. Consistently, previous studies in the literature reported the good correlation of these features with the haemoglobin concentration level [6,7,12], so we assume that these two features best describe the redness or paleness of the region of interest.

Table 4 and 5 report the best result for each classification algorithm

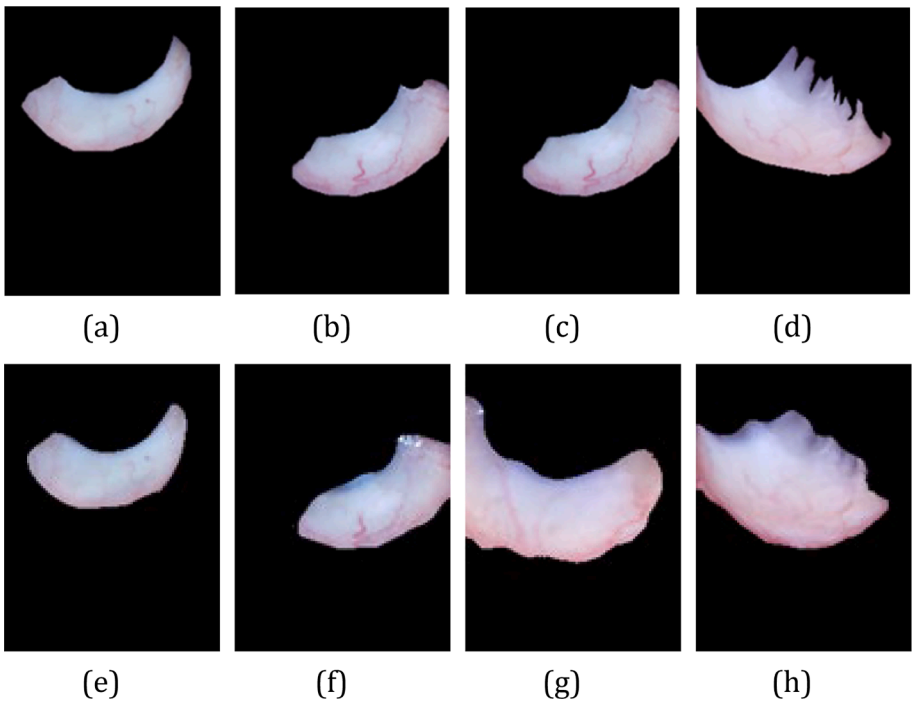


Fig. 11. Output samples. The first row (a, b, c, d) contains the ground truth, and the second row (e, f, g, h) shows the automated segmentation output.

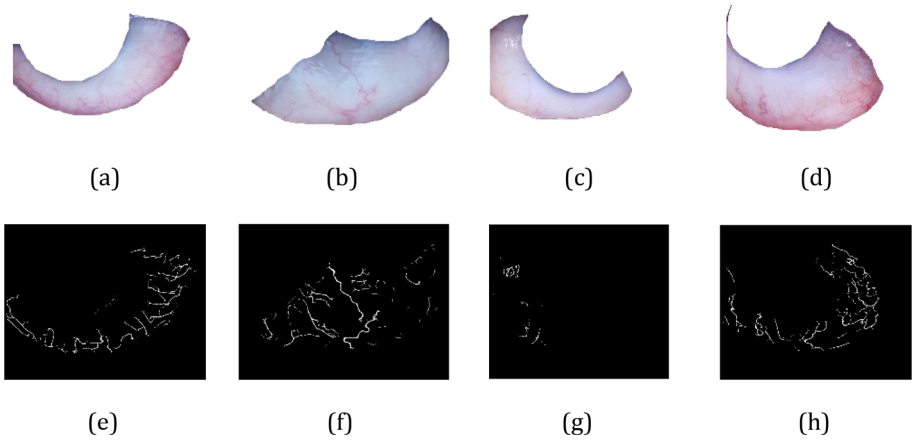


Fig. 12. Output samples. The first row (a, b, c, d) contains the whole sclera, and the second row (e, f, g, h) shows the automated vessels segmentation output.

Table 4
Best result for each classification algorithm on the sclerae validation set.

Features	Algorithm	Parameters	Precision	Recall	F1	F2
EI, a*	Nearest Neighbour	K = 7, uniform weights	87.0	78.0	79.8	78.4
EI, a*	Random Forest	estimators = 20, max features = auto, Gini criterion	78.0	73.0	72.5	72.4
EI, a*	Adaboost	Estimators = 40	83.0	73.0	74.7	73.2
EI, a*	Polynomial SVM	C = 20, degree = 3, gamma = auto	74.6	90.1	80.1	86.4

Table 5
Best result for each classification algorithm on the vessels validation set.

Features	Algorithm	Parameters	Precision	Recall	F1	F2
EI, a*	Nearest Neighbour	K = 7, uniform weights	92.0	77.0	82.6	78.9
EI, a*	Random Forest	estimators = 20, max features = log2, Gini criterion	69.0	72.0	70.0	71.1
EI, a*	Adaboost	Estimators = 2	72.0	72.0	70.3	71.0
EI, a*	Polynomial SVM	C = 60, degree = 4, gamma = auto	57.6	95	71.4	83.8

on the validation set, respectively for sclera and vessels colour features, the metrics are the results of 5-fold cross validation. Although we display many metrics, the F1 and F2 metrics deserve particular interest: because the dataset is unbalanced, F1 is the only reliable metric that can accurately describe the model performance; F2 has the same benefits as F1, but in our case, it has more importance because F2 allows us to give more importance to the recall, so it assigns a higher score to models with

Table 6
Results on the test set for both ROIs.

ROI	Features	Algorithm	Parameters	Precision	Recall	F1	F2
Sclera	EI, a*	Polynomial SVM	C = 20, degree = 3, gamma = auto	74.6	90.1	80.1	86.4
Vessels	EI, a*	Polynomial SVM	C = 60, degree = 4, gamma = auto	57.6	95	71.4	83.8

fewer false negatives. In this specific domain, false negatives are potentially harmful to people's health.

The best configurations of classifier/hyperparameters were as follows:

- for features from the entire sclera, SVM with C = 20, kernel="poly", gamma="auto", degree = 3;
- for features from vessels, SVM with C = 60, kernel="poly", gamma="auto", degree = 4.

These two classifiers were then tested on the test set and obtained the results presented in Table 6.

The better performance of the support vector classifier can be justified by recalling that SVM is generally a good performer even with a small number of samples.

5. Discussion

This study aimed to extract the sclera and scleral blood vessels from near-taken digital photos of the eye and proves that the extracted regions of interest can be used to check the anaemic status of a person. The proposed algorithm for sclera segmentation has comparable performance to state-of-the-art algorithms; its F1 score is 87 %. The good classification results obtained using the average EI value and average a* value as features imply a good correlation between the colour (paleness/redness) of segmented regions (sclera and vessels) and the haemoglobin value.

Several classification algorithms have been tested to estimate the anaemic condition based on the pallor of the tissues considered in this study. The best-performing SVM is a support vector machine with a polynomial kernel. The results were obtained by splitting the haemoglobin values into 2 classes according to the WHO threshold for anaemic/non-anaemic conditions.

The results are sufficient to state that the challenge is feasible, and performance can surely be improved by expanding the training data, especially adding more examples of anaemic patients.

6. Conclusions

Following the implementation and calibration of the segmentation algorithms and the use of a classifier based on a support vector machine, a system was set up to use a photo of a sclera to estimate the presence of anaemia and distinguish whether the patients are safe. Compared to the usual anaemia estimation methods, this method has the following advantages: it is easy and quick to use, is noninvasive and does not discomfort the patient; the device to take eye photos does not require the intervention of specialized personnel; it is not influenced by other pathologies because it is based on the properties relating to the colour of the vessels. In this study, only primary colour features were used, but the selected ROIs contain a lot of other useful information such as vascularization and the vessel length and width. The results of this study may find particular utility in cases where the eyelid conjunctiva is totally absent or very small in size, while a further very interesting investigation will be the estimation that combines features from the entire eye: the sclera, which is investigated in this study, and the entire conjunctiva. Another interesting development could be the adaptation of this work into a smartphone app, thus allowing to make the system more portable and easily available to as many people as possible in the fastest way possible. More extensive clinical trials have been planned with the hospitals of the metropolitan area of Bari (Italy) to certify the actual potential of the approach presented here.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by the University of Bari (Italy) - University Research.

References

[1] T.S. Leung, F. Outlaw, L.W. MacDonald, J. Meek, Jaundice eye color index (jeci): quantifying the yellowness of the sclera in jaundiced neonates with digital photography, Biomed. Opt. Express 10 (2019) 1250–1256.
[2] A. Mariakakis, M.A. Banks, L. Phillipi, L. Yu, J. Taylor, S.N. Patel, Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders, Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol. 1 (2017) 1–26.
[3] A.R. Kent, S.H. Elsing, R.L. Hebert, Conjunctival vasculature in the assessment of anemia, Ophthalmology 107 (2000) 274–277.
[4] A.T. Cheung, J.W. Miller, S.M. Craig, P.L. To, X. Lin, S.L. Samarron, P.C. Chen, T. Zwerdling, T. Wun, C.-S. Li, et al., Comparison of real-time microvascular abnormalities in pediatric and adult sickle cell anemia patients, Am. J. Hematol. 85 (2010) 899–901.
[5] J.K. Brooks, A review of syndromes associated with blue sclera, with inclusion of malformations of the head and neck, Oral. Surg. Oral Med. Oral Pathol. Oral Radiol. 126 (2018) 252–263.
[6] G. Dimauro, D. Caivano, F. Girardi, A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva, IEEE Access 6 (2018) 46968–46975.
[7] G. Dimauro, A. Guarini, D. Caivano, F. Girardi, C. Pasciolla, A. Iacobazzi, Detecting clinical signs of anaemia from digital images of the palpebral conjunctiva, IEEE Access 7 (2019) 113488–113498.
[8] M. Lakshmi, P. Manimegalai, S. Bhavani, Non-invasive haemoglobin measurement among pregnant women using photoplethysmography and machine learning, in: Journal of Physics: Conference Series, volume 1432, IOP Publishing, 2020, p. 012089.
[9] S. Ghosal, D. Das, V. Udutalapally, A.K. Talukder, S. Misra, shemo: Smartphone spectroscopy for blood hemoglobin level monitoring in smart anemia-care, IEEE Sens. J. 21 (2020) 8520–8529.
[10] G. Dimauro, D. Caivano, P. Di Pilato, A. Dipalma, M.G. Camporeale, A systematic mapping study on research in anemia assessment with non-invasive devices, Appl. Sci. 10 (2020) 4804.
[11] R.J. Stoltzfus, A. Edward-Raj, M.L. Dreyfuss, M. Albonico, A. Montresor, M. Dhoj Thapa, K.P. West Jr, H.M. Chwaya, L. Savioli, J. Tielsch, Clinical pallor is useful to detect severe anemia in populations where anemia is prevalent and severe, J. Nutr. 129 (1999) 1675–1681.
[12] T.N. Sheth, N.K. Choudhry, M. Bowes, A.S. Detsky, The relation of conjunctival pallor to the presence of anemia, J. Gen. Intern. Med. 12 (1997) 102–106.
[13] G. Dimauro, R. Maglietta, T. Bai, S. Kasiviswanathan, Eyes-defyanemia, 2022. URL: https://dx.doi.org/10.21227/t5s2-4j73. doi:10. 21227/t5s2-4j73.
[14] Z. Zhou, E.Y. Du, N.L. Thomas, E.J. Delp, A new human identification method: Sclera recognition, IEEE Trans. Syst., Man, Cybernet.-Part A: Syst. Humans 42 (2011) 571–583.
[15] G. Dimauro, L. Baldari, D. Caivano, G. Colucci, F. Girardi, Automatic segmentation of relevant sections of the conjunctiva for noninvasive anemia detection, in: 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), IEEE, 2018, pp. 1–5.
[16] A.S. Huang, A. Camp, B.Y. Xu, R.C. Penteado, R.N. Weinreb, Aqueous angiography: aqueous humor outflow imaging in live human subjects, Ophthalmology 124 (2017) 1249–1251.
[17] S. Kadomoto, Y. Muraoka, A. Uji, R. Tamiya, Y. Oritani, K. Kawai, S. Ooto, T. Murakami, Y. Iida-Miwa, A. Tsujikawa, Nonperfusion area quantification in

branch retinal vein occlusion: a widefield optical coherence tomography angiography study, Retina 41 (2021) 1210–1218.
[18] S. Kadomoto, A. Uji, A. Tsujikawa, Anterior segment optical coherence tomography angiography in a patient with persistent pupillary membrane, JAMA ophthalmology 136 (2018) e182932.
[19] H. Proença, S. Filipe, R. Santos, J. Oliveira, L.A. Alexandre, The ubiris. v2: A database of visible wavelength iris images captured onthe-move and at-a-distance, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2009) 1529–1535.
[20] S. Parthasarathy, P. Sankaran, An automated multi scale retinex with color restoration for image enhancement, in: 2012 National Conference on Communications (NCC), IEEE, 2012, pp. 1–5.
[21] C. Zhou, X. Zhang, H. Chen, A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden markov model, Comput. Methods Programs Biomed. 187 (2020), 105231.
[22] E. Ricci, R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, IEEE Trans. Med. Imaging 26 (2007) 1357–1365.
[23] K.-M. Kim, K. Son, G.T.R. Palmore, Neuron image analyzer: Automated and accurate extraction of neuronal data from low quality images, Sci. Rep. 5 (2015) 1–12.
[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[25] S. Alkassar, W.L. Woo, S.S. Dlay, J.A. Chambers, Robust sclera recognition system with novel sclera segmentation and validation techniques, IEEE Trans. Syst., Man, Cybernet.: Syst. 47 (2015) 474–486.
[26] M. Vitek, A. Das, Y. Pourcenoux, A. Missler, C. Paumier, S. Das, I. De Ghosh, D. R. Lucio, L.A. Zanlorensi, D. Menotti, et al., Ssbc 2020: Sclera segmentation benchmarking competition in the mobile environment, in: 2020 IEEE International Joint Conference on Biometrics (IJCB), IEEE, 2020, pp. 1–10.