An Algorithm to Screen for Preeclampsia Using a Smart Phone

Hemant D. Tagare
Professor, Dept. of Diag. Radiology,
Dept. of Biomedical Engineering
Yale University, New Haven, CT 06520
Email: hemant.tagare@yale.edu

Kara Rood
Maternal Fetal Medicine Fellow
Dept. of Obstetrics and Gynecology
The Ohio State University
Wexner Medical Center
Columbus, OH 43210

Irina A. Buhimschi
Professor, Dept. of Pediatrics
Dir., Center for Perinatal Research
Nationwide Children's Hospital
Columbus, OH 43205

Abstract—**Preeclampsia is one of the leading causes of maternal death in developing and underdeveloped countries. A simple urine test, called the Congo Red Dot Test (CRD Test), was recently proposed to screen for preeclampsia. This paper reports an algorithm for automatically quantifying and interpreting the CRD Test. The algorithm can be easily programmed as an app on a smart phone. When evaluated on a pilot data set, the screening accuracy of the algorithm is high. The CRD Test and the algorithm can be easily deployed in resource-limited countries.**

I. INTRODUCTION

Preeclampsia is a condition that occurs in pregnancy, where the mother develops hypertension (high blood pressure) and proteinuria (excess protein in the urine) [1]. Untreated preeclampsia can cause convulsions, coma, and ultimately, death. Approximately 1 pregnant woman in 20 develops preeclampsia. The WHO estimates that 16% of global maternal deaths (appx. $63,000$ deaths per year) are due to preeclampsia, and that most of these deaths occur in developing or under-developed countries [1], [2].

The only known cure for preeclampsia is delivery, presuming that a reliable diagnosis of severe preeclampsia has been made[1]. Diagnosing preeclampsia requires accurate measurement of blood pressure and proteinuria. Additionally proteinuria from a 24 hr. urine specimen and blood tests are needed to confirm the diagnosis. These measurements are usually made in a laboratory. Due to shortage of expertise, equipment, and laboratories in resource-limited countries, preeclampsia often goes undiagnosed, with dire consequences as noted above.

A simple screening test, which does not need access to a laboratory, and which can be used at point-of-care, would have a significant impact on the morbidity and mortality caused by preeclampsia. A new discovery [3] makes it possible to create such a test. The discovery is that, compared to healthy pregnant women, most preeclamptic women have large amounts of misfolded proteins in their urine. The amount of misfolded protein can be quantified by using a dye called Congo Red.

Different versions of the screening test using Congo Red are available. The version relevant to this report is a dual-dye test where an additional dye called Erioglaucine (which is a commonly used blue-colored food dye) is used along with Congo Red. The procedure, called the *Congo Red Dot Test* (CRD Test), is described in detail in section II. For now, it is sufficient to know that the CRD Test produces concentric red and blue colored regions on paper. The relative areas of the red and blue regions indicate the amount of misfolded protein, thereby making it possible to screen for preeclampsia.

The relative areas of the colored regions can be evaluated by eye, but that judgement is subjective and operator-dependent. An algorithmic calculation of the relative areas can make the test quantitative and objective. Furthermore, if the algorithm is simple enough to be deployed on a smart phone, then the screening can be carried out in the field, without the need for a laboratory or a computer. The goal of this article is to report just such an algorithm and its ability to screen for preeclampsia with pilot CRD Test data.

Currently, the algorithm is implemented as a MATLAB script, and we hope to port it to smart phones in the near future. The algorithm is computationally simple, and we do not anticipate any difficulties in porting it. Indeed, an earlier algorithm for analyzing the CRD Test was successfully ported to the iPhone by us [4] and approved by the iTunes app store. All images used to evaluate the current algorithm are obtained via an iPhone 4. Therefore, the experimental results are indicative of the algorithm's performance when deployed on an iPhone 4.

Congo Red and Erioglaucine are inexpensive and readily available. The CRD Test only requires the two dyes, a plastic dropper, and paper. A kit for the CRD Test is readily assembled, and current estimates are that material costs for the kit are less than $0.04\$$ per kit, making it affordable in resource-limited countries. Smart phones too are becoming more easily available in these countries. Thus, together, the CRD Test and the algorithm are a potential mHealth solution to preeclampsia screening.

The rest of this paper describes the CRD Test, the

---

[1]Mild cases of preeclampsia can be managed until full term with medical care

algorithm, and its evaluation. The paper is organized as follows. Section II describes the Congo Red Dot Test and various factors that influence the quality of its images when they are acquired by a smart phone. Section III describes the algorithm. Section IV contains experimental results evaluating the algorithm. Section V concludes the paper.

II. THE CONGO RED DOT TEST

The CRD Test is based on the principle that Congo Red adheres to misfolded proteins and also to paper, but its adherence to misfolded proteins is stronger than its adherence to paper. Erioglaucine does not adhere to proteins or to paper; it merely mixes uniformly with the urine. Erioglaucine is used to clearly demark the size of a urine drop in the CRD Test.

(a) Diffusion of the drop on paper

(b) CRD Test without misfolded proteins
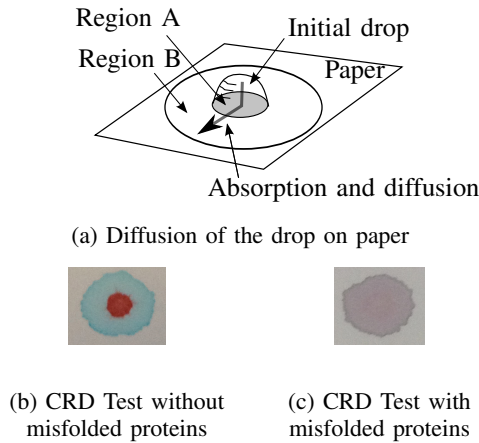
(c) CRD Test with misfolded proteins

Fig. 1. The Congo Red Dot Test of urine with and without misfolded proteins. Note the immediate adherence of Congo Red to paper when there are no misfolded proteins, and the uniform spread of Congo Red when there are misfolded proteins.

Congo Red and Erioglaucine are mixed with urine and a drop of the mixture is placed on paper (see fig. 1a). The drop is absorbed in the paper and it diffuses outwards making a larger circular blot (the large blot is denoted "Region B"in fig. 1a). The diffusion of the Congo Red dye within this blot depends on misfolded proteins:

(1) If there are no, or a relatively few, misfolded proteins in the urine, then Congo Red adheres immediately to the paper in the shape of the initial drop (the "region A" in fig. 1a) and does not diffuse out with the urine. On the other hand, Erioglaucine diffuses evenly with the urine. Thus the final blot on the paper has a small well-defined red central region surrounded by a blue region, as seen in fig. 1b.

(2) If there are many misfolded proteins in the urine, then Congo Red binds to the proteins rather than to the paper. As the initial drop of the urine diffuses, the proteins and the adhered Congo Red diffuse uniformly along with it. As before, Erioglaucine also diffuses uniformly with the urine, so that the final blot has overlapping red and blue regions. This is shown in fig. 1c. Note however, that the example in fig. 1c is an extreme case of severe preeclampsia. There are cases where the spread of Congo Red is intermediate to fig. 1b and c.

Our goal is to automatically analyze images of the CRD Test that are taken with a smart phone camera under non-laboratory conditions. Fig. 1 suggests that this analysis can be carried out by using information in the red and blue channels of the camera. But the intensities in the two channels are influenced by a number of factors, and these factors have to be taken into account during the analysis:

(1) *Ambient illuminant type*: The intensites in the red and blue channels depend on the type of ambient illumination. The intensities for the same CRD Test can be differ substantially depending on whether the ambient illumination is diffuse natural light, incandescent light, or fluorescent light. Fig. 2 shows this effect.

(a) Diffuse natural light
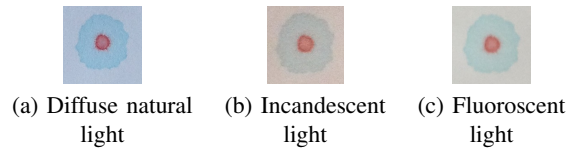
(b) Incandescent light

(c) Fluoroscent light

Fig. 2. The effect of changing the illumination on the same CRD Test blot

(2) *Illumination gradient*: Non-uniform illumination across the CRD Test can cause the intensities to ramp up or down across the image. The CRD Test size is sufficiently small that this phenomenon can be modeled as a linear variation of intensities across the image.

(3) *Type of paper*: The red and blue regions are caused by the dyes diffusing through the paper and the extent of diffusion depends on the type of paper.

We account for the first two factors by pre-processing the image and adaptively adjusting to the illuminant color. We account for the last factor by testing different papers. Preliminary experimentation showed that the type of paper that is used for printing adhesive labels gives a nice separation of the two regions when there are no misfolded proteins. We tested three types of label papers: Avery® laser white mailing labels 5160, Avery® laser/ink jet multipurpose labels 6460, and Avery® laser/ink jet filing labels 5366. The results of this evaluation are reported below in section IV.

III. THE ALGORITHM

The main idea of the algorithm is to segment both the red and blue channels into a background and a foreground region, where the background region corresponds to the paper and the foreground region to the red or blue colored areas. The algorithm has two phases: an initial pre-processing phase, and a subsequent segmentation phase.

The notation used to describe the algorithm is as follows: Each CRD Test image I is defined over a rectangular set of pixels Ω. A pixel in Ω is denoted $u \in \Omega$

and its $x-$ and $y-$ coordinates are u_x and u_y. The red, green, and blue channel intensities of I are I_r, I_g, and I_b respectively. The intensity values of these channels at pixel u are $I_r(u), I_g(u)$ and $I_b(u)$. The algorithm processes only red and blue channels, and when the processing of both channels is identical we simply refer to the channel as I_c where the subscript c indicates color and takes values $c \in \{r, b\}$.

The algorithm assumes that the colored regions are completely contained within the image so that a 1 pixel thick boundary frame at the edge of Ω only contains the white background of the paper. The set of pixels in this background frame is denoted B.

A. Pre-processing

The pre-processing phase eliminates illumination gradients by fitting an affine function to the intensities in the red and blue channel pixels of the background frame B and then subtracting the fitted function from the entire image. The fitting is done by minimizing

$$J_c(\beta_0, \beta_x, \beta_y) = \sum_{u \in B} \left(I_c(u) - \beta_x u_x - \beta_y u_y - \beta_0 \right)^2 \quad (1)$$

with respect to $\beta_0, \beta_x, \beta_y$ separately for $c = r$ and $c = b$. This is, of course, classic regression, and the minimizing $\beta_0, \beta_x, \beta_y$ are easily found in closed form [5]. After minimization, the fitted function is subtracted from the entire image. That is, for $c = r$ and $c = b$, the intensity $I_c(u)$ at every pixel $u \in \Omega$ is set to

$$I_c(u) \leftarrow I_c(u) - \hat{\beta}_x u_x - \hat{\beta}_y u_y - \hat{\beta}_0. \quad (2)$$

In equation (2), $\hat{\beta}_0, \hat{\beta}_x, \hat{\beta}_y$ are the minimizing values of $\beta_0, \beta_x, \beta_y$ for the corresponding channel.

Keeping in mind that the fitting in equation (1) is only over the background frame, whereas the adjustment of intensities in equation (2) is over the entire image, the preprocessing has the effect of setting the entire white background of the image close to 0, and eliminating gradients while maintaining the contrast between the background and the colored regions (the intensity values in the colored regions take negative values).

B. Segmentation

The second phase of the algorithm segments each channel into two parts - one part corresponds to the background and the other to the colored region. Many sophisticated algorithms are available for image segmentation. However, these algorithms are computationally too expensive for a smart phone app, and instead we sought a computationally simple algorithm. After some experimentation, the two stage algorithm reported below proved reliable. Both stages of the algorithm are applied independently to the r and b channels.

In the first stage, k-means clustering is applied to cluster the pixels in the channel into two classes. K-means is initialized by setting the first class mean to

(a) CRD Test Image   (b) Red Ch. segmentation   (c) Blue ch. segmentation

(d) CRD Test Image   (e) Red Ch. segmentation   (f) Blue ch. segmentation
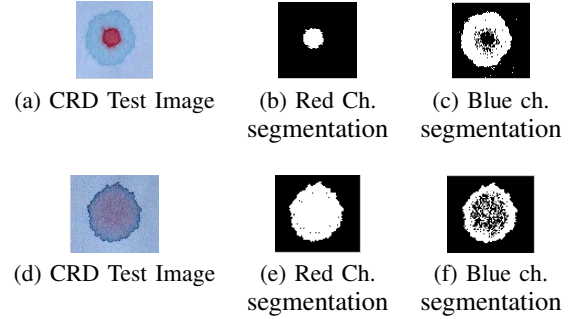
Fig. 3. Segmentations of CRD Test images

the maximum intensity value and the second class mean to the minimum intensity value. Thus, the first class corresponds to the background and the second to the colored region. K-means is iterated 5 times. At the end of this step, a segmentation of each channel into two regions is available.

In the second stage of segmentation, the k-means segmention is improved by modeling the variances as well as the means in each region. That is, we assume that the intensities in the background and the colored region are normally distributed and estimate their means (μ_1, μ_2 respectively for background and colored region) and standard deviations (σ_1, σ_2 respectively) from the k-means segmentation. Then, the entire image is re-segmented using a likelihood test: every pixel u is classified as belonging to the background if

$$-\frac{(I_c(u) - \mu_1)^2}{2\sigma_1^2} - \log \sigma_1 > -\frac{(I_c(u) - \mu_2)^2}{2\sigma_2^2} - \log \sigma_2, \quad (3)$$

else it is classified as belonging to the colored region. The segmentation obtained at the end of this step is taken as the final segmentation. Because the class means are updated in every iteration of the k-means step, the segmentation adjusts to the illuminant color. Fig. 3 shows examples of segmentations given by the algorithm for CRD Test images.

If Ω_r is the set of pixels of the segmented colored red region and Ω_b as the set of pixels of the segmented colored blue region, then the Congo Red area ratio is calculated as $\text{Area}(\Omega_r)/\text{Area}(\Omega_r \cup \Omega_b)$. The area ratio is always between 0 and 1. We expect the CRD Test with no misfolded proteins to have a small ratio and the CRD Test with many misfolded proteins to have a large ratio.

IV. Experimental Results

A. Testing the Paper

The first experiments we report evaluate the three Avery label papers. CRD Tests were carried with three urine samples from healthy pregnant women and three urine samples from preeclamptic women[2]. The CRD Test images were obtained with an iPhone 4 camera

---

[2]All urine samples reported in this section were obtained from a frozen urine sample bank maintained by the author I.A.B.
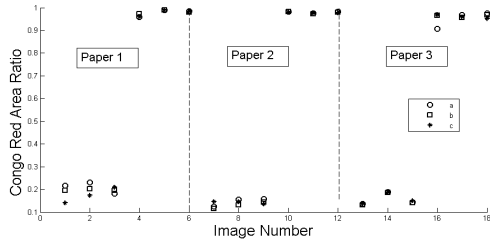
Fig. 4. Testing the paper type. Paper 1=Avery laser white mailing labels 5160, Paper 2= Avery laser/ink jet multipurpose labels 6460, Paper 3 = Avery laser/ink jet filing labels 5366. "a", "b", "c" are three Congo Red Dot Tests of the same urine sample.
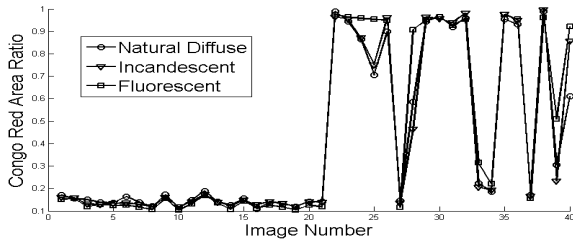

Fig. 5. Performance of CDRT + Algorithm under different illumination conditions

with indoor fluorescent lighting and analyzed with the algorithm reported above. Fig. 4 shows the Congo Red area ratios calculated by the algorithm. The first 6 CRD Tests are from the first paper, the next six are from the next, and so on. The vertical, dashed lines group the CRD Tests paperwise. Within each paper group, the first 3 CRD Tests are from the healthy subjects, and the last 3 from preeclamptic subjects. The entire set of CRD Tests was repeated three times with urine from the same samples, and the repeats are marked as a,b and c in the figure legend.

Several observations can be made about fig. 4: First, the area ratio calculated by the algorithm behaves as expected; it's values are small for healthy subjects and large for preeclamptics. Second, the results are repeatable and consistent. Finally, all three papers perform about equally well.

B. Screening Performance

Next we evaluated the performance of the CRD Test and the algorithm under three different illumination conditions (natural diffuse light, incandescent light, fluorescent light) with 40 urine samples using Avery® laser white mailing labels 5160 (paper 1 in fig. 4). Fig. 5 shows the results. There were 21 healthy samples, and these are plotted as the first 21 data in fig. 5. The remaining 19 were preeclamptic, and are plotted as the last 19 in the figure. As seen in the figure, the Congo Red area ratios for healthy samples are insensitive to illumination, while there is some residual sensitivity in the preeclamptic samples. The Congo Red area ratios for healthy samples

are uniformly small, and except for five cases, the Congo Red area ratios for preeclamptic samples are high. A closer examination of the exceptional fives cases shows that they are only mildly positive cases; their low area ratios reflect this condition.
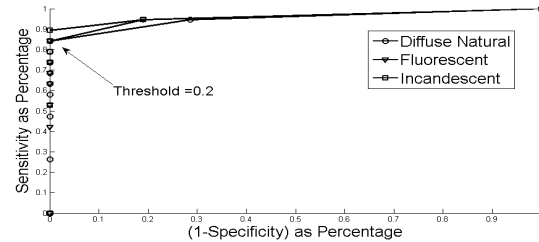

Fig. 6. The specificity-sensitivity curve. The areas under the curve are 0.95 for diffuse, 0.96 for fluorescent and incandescent.

Fig. 5 strongly suggests that a Congo Red area ratio threshold of about 0.2 should separate the normals from most of the preeclamptic CRD Tests. Fig. 6 shows the specificity-sensitivity curve obtained by varying the threshold. The figure suggests that 80% sensitivity can be obtained at almost 100% specificity (at a Congo Red area ratio threshold of 0.2) irrespective of illuminant type.

A larger sample has to be analyzed to confirm these specificity-sensitivity numbers. We plan to do so in the near future. Furthermore, the data in fig. 5 suggests that there may be a need to develop a third category - that of mild pre-eclampsia - in the analysis. This too will be explored with a larger data set in the future.

V. CONCLUSION

The pilot data presented above suggests that this test can easily screen for preeclampsia. The low false positive rate is especially important in resource-limited countries because false positives can over-burden the resources needed for confirmatory testing.

REFERENCES

[1] L. Duley, "The global impact of pre-eclampsia and eclampsia", Semin. Perinatol., 33(3), Jun. 2009, pp. 130-137.

[2] WHO: The World Health Report 2005: Make every mother and child count. Geneva, World Health Organization, 2005.

[3] I. A. Bhumschi et al., "Protein misfolding, congophilia, oligomerization, and defective amyloid processing in preeclampsia", Sci. Transl. Med., 6, July 2014.

[4] T. Westcott, *The Application of Mobile Phone Image Processing to Medical Diagnostics*, Senior Thesis, Yale University 2014.

[5] M. H. Kutner et al., *Applied Linear Statistical Models*, McGraw-Hill, 2005.