Research Article

Regression based myoglobinuria detection from urine image

Md. Imran Khan[*]

Institute of Information and Communication Technology (IICT), Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

Abstract
Now-a-days urine dipstick test is needed to determine the presence of myoglobin in urine and to quantify myoglobin level in urine sophisticated spectrophotometer method is performed. This paper proposes a soft computation based novel regression model that can determine the presence of myoglobin and also can estimate myoglobin level in urine from urine image. Performance analysis of several regression model shows that multivariate nonlinear Hougen regression model provides best fit between estimated myoglobin values with image data. We find that the residuals are near in baseline and the adjusted coefficient of determination (R2) is 0.97 which is very significant. Root mean square error (RMSE) is 19.5 out of 51 urine image samples with 10 error degrees of freedom.

Keywords
Myoglobin, Urine dipstick test, Multivariate nonlinear Hougen regression, Residuals.

1.Introduction
Myoglobinuria is usually the result of rhabdomyolysis or muscle destruction. Any process that interferes with the storage or use of energy by muscle cells can lead to myoglobinuria. The release of myoglobin from muscle cells is often associated with an increase in levels of creatine kinase (CK), aldolase, lactate dehydrogenase (LDH), serum glutamic-pyruvic transaminase (SGPT), and other enzymes [1]. Urine gets its yellow color from a pigment called urochrome but urine color of Myoglobinuria patient is dark red. Myoglobinuria is produced by multiple causes: any condition that accelerates the use or interferes with the availability of oxygen or energy substrates to muscle cells can result in myoglobinuria, as can events that produce direct muscle injury, either mechanical or chemical[2]. Any process which leads to muscle necrosis can lead to myoglobinuria and therefore an increased risk of rhabdomyolysis [3]. Myoglobin is a skeletal muscle protein involved in metabolism and myoglobinaemia usually occurs before a rise in CK in rhabdomyolysis. Any myoglobin that reaches the circulation will be filtered by the kidneys and can lead to acute kidney injury through either direct toxicity or precipitation, or both [4]. This process is facilitated by an acidotic environment and hypovolaemia [5].

Spectrophotometer method is also used to quantifying myoglobin in urine. There is a need for simple, accurate at the same time cost effective solution for evaluation of Mb of all the methods that are available for Myoglobinuria detection, many studies conducted on other alternative methods have found them to be inadequate in addressing the reliability and accuracy of the estimation of Mb in urine, making alternative methods unsuitable. So, we have a requirement to find a suitable easy to-use, economical and simple method for estimation of Mb in urine that can be operated in rural and low-resources settings.

This paper is structured as follows. Section 2 part 1 describes data preparation; part 2 for linear model fit and part 3 shows nonlinear Hougen model fit. Section 3 is for performance analysis. Finally Section 4 is conclusion and future work.

2.Methodology
2.1Data preparation
There is a requirement to extract the color information from an image. It can be done by looking into the picture and analysis it pixel by pixel. Each image consists of a number of pixels. Each pixel is defined by its capacity to hold color information. There are different definitions for this color information. That is, 256 colors, 16-bit color or 24 bit color and so on. 256 colors are defined by 8 bits allocated to represent the information in terms of red,

---

*Author for correspondence

green and blue values. We have uniformly used 24-bit color information, allocating 8 bits for each of the primary colors namely, red, green and blue. From the original picture, the R, G, B values are extracted. This is done by considering the R, G, B values of each pixel in the image. The values are then averaged over the complete image. The standard deviation and variance are also calculated for the image by the actual R, G, and B values of each pixel. For all our analysis, it is better that we have a uniform image. The uniformity of the image is ascertained by the relative values of average of each color and the corresponding variance for each color. If the variance for a color, says Red, is within 10% of the average for the same color, we can declare the image to be uniform and it can be considered for analysis. It can also be appreciated from the images that the color information is almost uniform in the picture. *Figure 1* shows the color difference between healthy person's urine and myoglobinuria affected patient's urine.
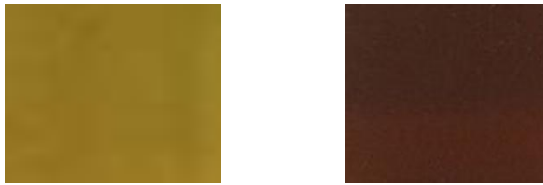
Figure 1 Left picture is healthy person's urine image and right picture is myoglobinuria affected patient's urine image

The method that we use is to create two such images for each person and find the change in R, G and B values of the pictures. It can be seen that the urine image of myoglobinuria affected patient is dark brown but the urine image of healthy person's is dark yellow .The extent of brownness is indicative of the Mb level in Urine. In the present study, about 51 myoglobinuria affected patient's urine samples are taken. The Mb levels are also collected for these samples, to develop a regression model.

$$R=R1-R2 \qquad (1)$$

$$G=G1-G2 \qquad (2)$$

$$B=B1-B2 \qquad (3)$$

$$Mb=function(R, G, B) \qquad (4)$$

Where,
R1, G1, B1=Red, Green and Blue color plane information of every pixel of healthy person's urine

image.
R2, G2, B2= Red, Green and Blue color plane information of every pixel of myoglobinuria affected patient's urine image.

Mb=Myoglobin level in mg/dl.

2.2 Linear model fit
A multivariate linear regression analysis is carried out to find if there is a correlation between the difference in R, G and B values and the Mb value of the urine of the person. The analysis resulted in defining a relationship between Mb value and the primary colors that we have taken. Stepwise linear regression model is carried out so that adding or removing specific term is comparatively easier. Adding nonlinear term not necessary results good fit. We increase power of individual parameter(R, G, B) until the value of Adjusted Determination of coefficient (R^2) becoming lower. It is seen that increasing power 2 to 3 causes a great reduction of R^2 value. Analysis of P value and F statistics indicates the poor relationship of G×B, R×B and B^2 term with Mb values. Removing these three terms we get the relation as:

$$Mb=1-4.0208R-1.1957G+3.7084B+0.079814R^2 +0.078955G^2 - 0.17386RG \qquad (5)$$

2.3 Nonlinear hougen model fit
Hougen model is a nonlinear model where estimated value is a function of the three predictor variables. For a 5-D vector b and 3-D vector x hougen model is defined as:

$$Hougen(b,x) = \frac{b(1)x(2)-x(3)/b(5)}{1+b(2)x(1)+b(3)x(2)+b(4)x(3)} \qquad (6)$$

Where b is the coefficient vector .We use R value as x (1); G value as x (2) and B value as x (3).Finally we get Mb value as hougen (b, x). After simulation we find the value of coefficient. So the Hougen equation for this fit:

$$Mb=N/D \qquad (7)$$
Where;
$$N= 0.58824G – 1.682B \qquad (8)$$
$$D =1- 0.0098948R +0.016647G -0.028256 B \qquad (9)$$

3. Performance analysis
From *Table 1,* it can be said that Nonlinear Hougen model provides best fit between image data with Mb values because of highest value of R^2, adjusted R^2, F statistics and lowest value of RMSE and P value. Here,

EDF= Error Degrees of Freedom
RMS Error=Root Mean Square Error
LT= Linear Terms
NLT= Nonlinear Terms
NHM= Nonlinear Hougen Model

Table 1 Comparison among different regression model

| Model	R^2	Adjusted R^2	p	RMS error	F stat	EDF
R	0.45	0.44	0.232	170	15	21
G	0.10	0.10	0.544	287	8	48
B	0.77	0.67	0.094	49.93	19	36
R,G (LT)	0.51	0.49	0.022	89.9	13	27
G,B (LT)	0.75	0.60	0.004	51.33	19	30
R,B (LT)	0.71	0.65	0.007	55.78	17	29
R,G,B (LT)	0.89	0.88	0.005	44	20	39
R,G,B (NLT)	0.95	0.921	0.001	23.7	23	42
R,G,B (NHM)	0.97	0.923	1.6×10^{-05}	19.5	57.1	45

So it is clear that equation 6 is the best fit for these sets of data value. Equation 5 is better than other model but low R^2 value makes it inferior to equation 6.Further analysis of equation 6 is shown in *Figure 2* to *Figure 6*. *Figure 2* shows 95% confidence bounds data fitted curve.
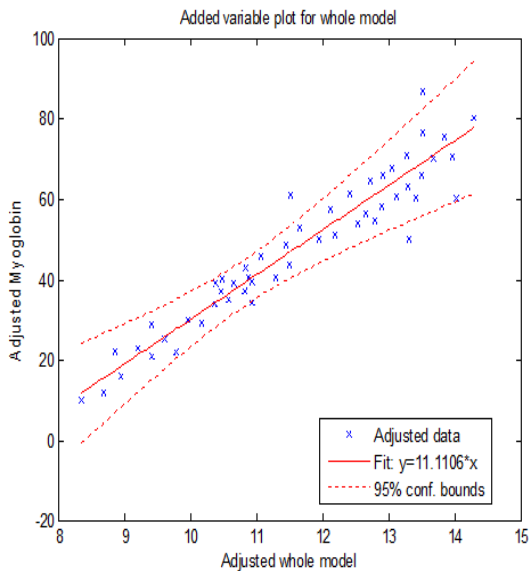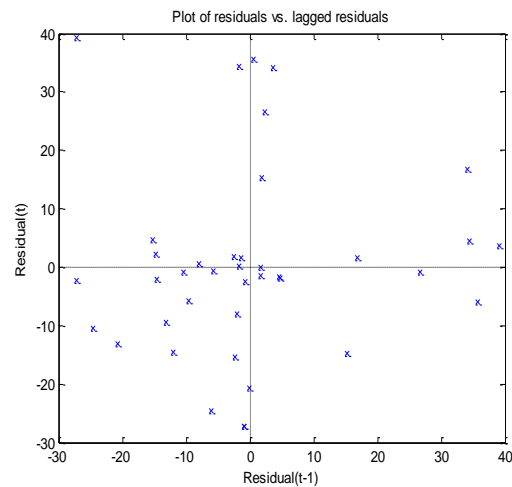

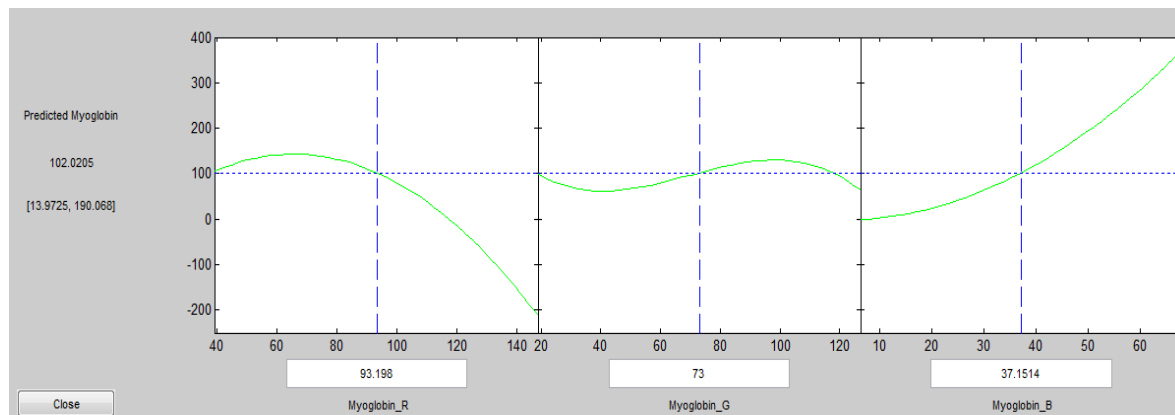**Figure 2** Variable plot for whole model
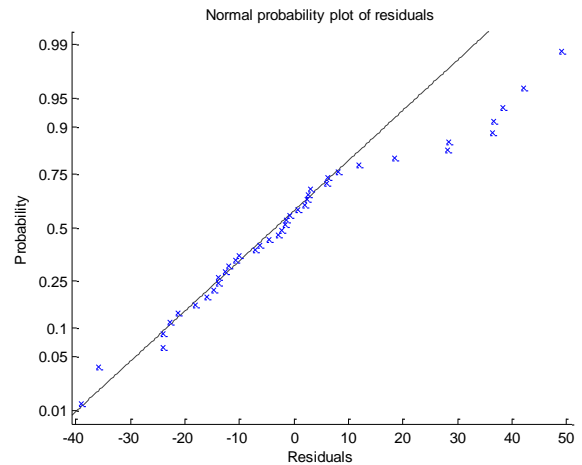
Figure 3 Normal probability plot of residuals
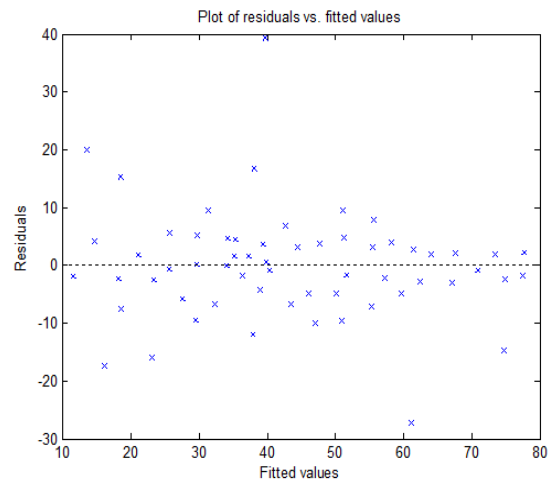
Figure 4 Residuals versus fitted value

Figure 5 Residuals versus lagged residuals

Figure 6 Effect of individual variable(R, G, B) on Mb

Figure 3 clearly indicates that normal probabilty of most of the residuals are below 0.75.The residuals of higher fitted value are comparatively more near in baseline then lower fitted value shown by *Figure 4*, which means the total error is very low at higher fitted value. *Figure 5* is necessary to understand the residuals distributions.Besides the whole model fit the effect of individual variable(R,G,B) on predicted Mb value is shown in *Figure 6*.

4.Conclusion and future work

The diagnostic value of particular tests depends greatly on the accuracy and reliability. Accuracy can be estimated by the comparison of the results obtained by our method with the results of a standard method. Since multivariate regression analysis was the method used to fix the variable values, statistically this method is expected to be more in line with the standard method. It is found that the results help us to simplify the procedure for estimation of Mb in urine of people specifically in rural areas while maintaining accuracy of the most standard method to a greater extent. Once perfected, this scheme of evaluation of Mb in urine can be used extensively as the method is simple, easily portable, and easily operable and does not require services of experienced and skilled technicians. Moreover this method does not need any chemical reagent like traditional dipstick method. Beside this regression model has less computational complexity as a result this method is faster than others. Acute level of Myoglobinuria is detected after suitable thresholding the myoglobin value. The novelty of this method is no microscopic image is required here for myoglobinuria detection.

In future regression model will be developed for making relation among R, G, B value of urine image with Protein content and Hemoglobin content.

Conflicts of interest
The author has no conflicts of interest to declare.

References
[1] http://emedicine.medscape.com/article/982711-overview. Accessed 15 September 2016.
[2] David WS. Myoglobinuria. Neurologic Clinics. 2000; 18(1):215-43.
[3] Bosch X, Poch E, Grau JM. Rhabdomyolysis and acute kidney injury. New England Journal of Medicine. 2009; 361(1):62-72.
[4] Perazella MA, Walter LA, Catenacci MH. Resident grand rounds: rhabdomyolysis. Hospital Physician. 2008; 44(1):25.
[5] Beetham R. Biochemical investigation of suspected rhabdomyolysis. Annals of Clinical Biochemistry. 2000; 37(5):581-7.

Md. Imran Khan is a postgraduate student of Bangladesh University of Engineering and Technology, Dhaka, Bangladesh. He has completed his B.Sc. in Electronics and Comminication Engineering degree from Khulna University of Engineering and Technology, Khulna, Bangladesh. He has 14 publications in reputed journal and international conference proceedings.
Email: imrankhankuet@gmail.com