Central Asian Journal of Medicine

ANALYSIS OF HEMOGLOBIN LEVELS USING IMAGE-BASED METRICS: A MULTIVARIATE APPROACH

Guzal A. Dadaboyeva

English for Specific Purposes teacher, Department of uzbek and foreign languages,
Tashkent Medical Academy, Tashkent, Uzbekistan
E-mail: gozalina_0606777@mail.ru

ABSTRACT

This study investigates the relationship between hemoglobin (Hb) levels and image analysis metrics derived from blood sample images. Utilizing a dataset of 425 observations, we employed descriptive statistics, graphical analysis, confidence intervals, hypothesis testing, and multivariate regression to explore the impact of various image-based features on hemoglobin levels. Our findings reveal that blue channels in the image analysis have the most significant influence on hemoglobin levels, with a strong negative correlation. The study provides insights into the potential use of image-based metrics for hemoglobin analysis and highlights the importance of considering color channels in medical diagnostics.

Key words: Hemoglobin, image analysis, multivariate regression, color channels, medical diagnostics.

INTRODUCTION

Hemoglobin, a critical protein in red blood cells, is responsible for oxygen transport throughout the body. Accurate measurement and understanding of hemoglobin levels are essential for diagnosing and managing various medical conditions. Traditional methods of hemoglobin measurement often require invasive procedures, but recent advancements in imaging technology offer a non-invasive alternative. This study aims to explore the relationship between hemoglobin levels and image-based metrics extracted from blood sample images, focusing on the influence of color channels and other image characteristics.

Materials and Methods

The dataset used in this study comprises 425 observations, each containing hemoglobin values and corresponding image analysis metrics. The image metrics include mean values of blue, green, and red channels, as well as hue, saturation, and brightness levels in the middle portion of the image. The following methods were employed for data analysis:

Dataset:

This dataset contains hemoglobin (Hb) values along with corresponding image analysis metrics extracted from images of blood samples. The data was collected using a custom imaging system capable of capturing various parameters related to hemoglobin concentration and image characteristics. Dataset contains 425 observations.

Research goal:

To study the impact of various features on hemoglobin levels, a dataset containing relevant information can be instrumental in gaining insights and drawing conclusions. Hemoglobin, a protein found in red blood cells, plays a crucial role in transporting oxygen throughout the body. Understanding the factors that influence hemoglobin levels can provide valuable information for medical research and healthcare practices.

Variables description:

Variable	**Description**
hbvalues	Hemoglobin values measured in the blood sample
model	Model or device used for hemoglobin measurement
mid_b_mean, mid_g_mean, mid_r_mean	Values of blue, green, and red channels in the middle portion of the image.
mid_h_mean, mid_s_mean, mid_v_mean	Values of hue, saturation, and value (brightness) in the middle portion of the image.

Methods:

- Descriptive statistics
- Graphical analysis
- Confidence intervals construction
- Hypothesis testing
- Bivariate analysis
- Multivariate analysis

Descriptive statistics:

Hemoglobine:

Parameter	Value
Mean	119.4
Standard deviation	31.5
Minimum	13
Maximum	170
Median	127

This data set appears to represent a distribution of values, likely a sample of some measurements or observations. The mean (average) value is 119.4, and the standard deviation is 31.5, indicating the spread of the data around the mean.

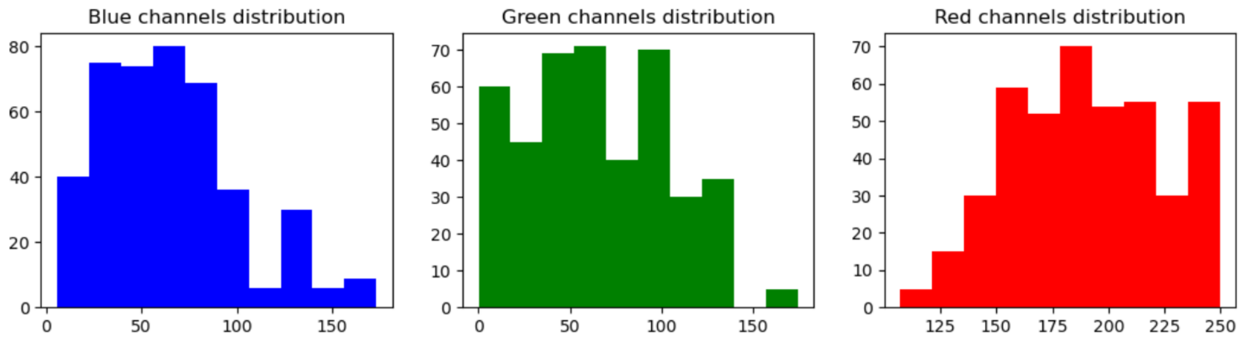
The minimum value in the data set is 13, and the maximum value is 170, showing the range of the data. The median, which is the middle value when the data is arranged in ascending order, is 127.

Parameter	Blue channels	Green channels	Red channels
Mean	65,4	65,3	189,7
Standard deviation	36,6	39,2	33,8
Minimum	5,7	0,1	107,2
Maximum	173,1	174,6	249,5
Median	64,3	61,8	191,3

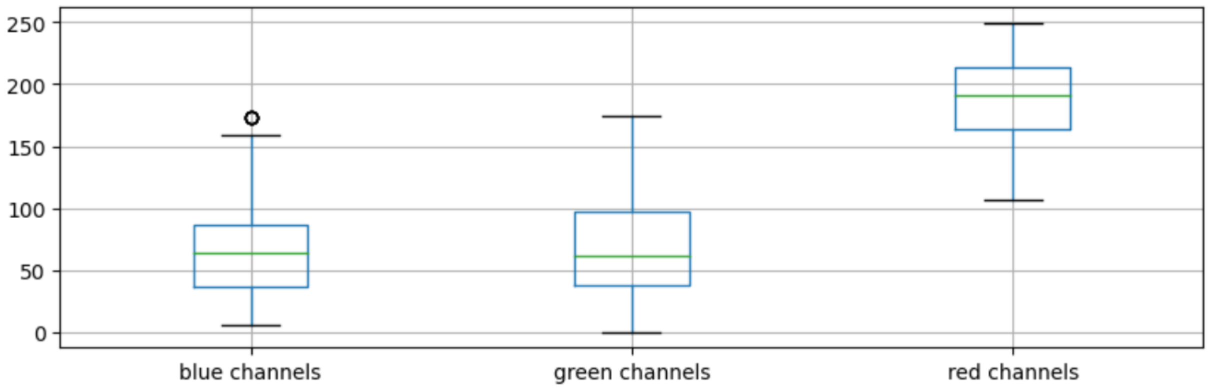
It is worth noting that the blue and green channel values have a median that is smaller than the mean. This indicates that the data has a right-skewed distribution, meaning that more than half of the objects are smaller than the mean.

In contrast, the red channel values have a median that is larger than the mean. This indicates that the data has a left-skewed distribution, meaning that more than half of the objects are larger than the mean.

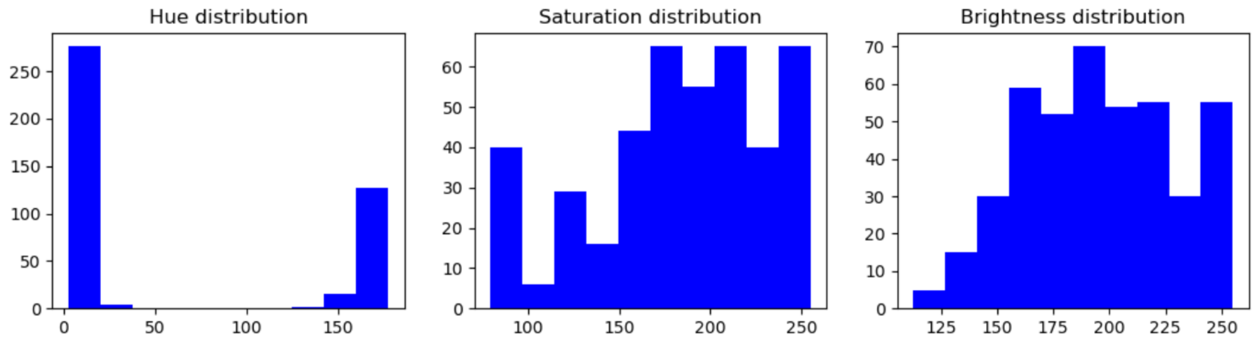
These data sets likely represent the intensity levels or values of the blue, green, and red channels in hemoglobin images.



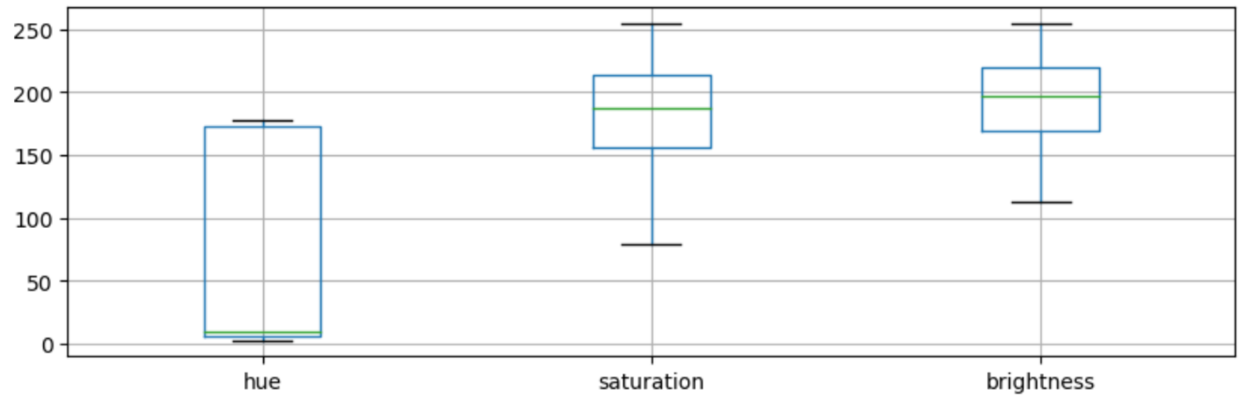
The graphs above show histograms of the distribution of these three variables. As we can see, the distribution is not visually normal.



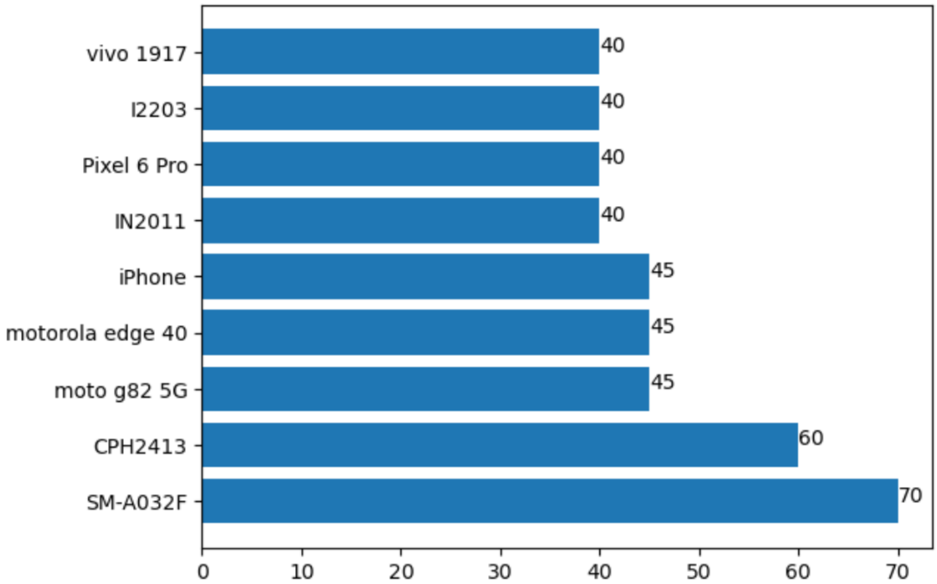
The graphs above show boxplots for the three variables. They are depicted according to the concept of interquartile range. It is obvious that there is a single outlier in the blue channel variable. It is above than Q3 + 1.5 * IQR.



The graphs above show the distribution of hue, saturation and brightness levels. It is obvious that most of the values of objects in the variable «Hue» are concentrated near 0. A smaller part has values greater than 150. The other two variables take values over the entire interval, but their distribution cannot be called normal.



The graphs above show boxplots for these three variables. It is clear that there are no outliers among them.



The bar graph shows how many times hemoglobin was measured with each device. Most often, this was done using SM-A032F, CPH2413 (70, 60 times respectively).

Confidence intervals:

$$x \pm z\alpha/_2 * \frac{s}{\sqrt{n}}$$

Let's construct a 95% confidence interval for all quantitative variables.

variable	Left bound	Right bound
Hemoglobin	116,44	122,43
blue channels	61,96	68,92

green channels	61,54	69,00
red channels	186,44	192,87
hue	55,95	70,80
saturation	178,89	188,03
brightness	191,93	198,34

Hypothesis testing:

Let's test the hypothesis that the hemoglobin level is equal to a certain value. The alternative hypothesis will be that the indicator is less than this value on 5% significance level.

$$H_0: \mu = 120$$

$$H_a: \mu < 120$$

$$z_{stat} = \frac{\mu - mean}{st.dev./\sqrt{N}} = \frac{120 - 119.4}{31.5/\sqrt{425}} \approx 0.39$$

$$z_{crit} = -1.645$$

$Z_{stat} > Z_{crit}$, so we do not have enough evidence to reject the null hypothesis and accept the alternative hypothesis that the average hemoglobin level is less than 120.

Let's test the hypothesis that the number of blue channels is equal to the number of green channels on 5% significance level

$$H_0: \mu_b = \mu_g$$

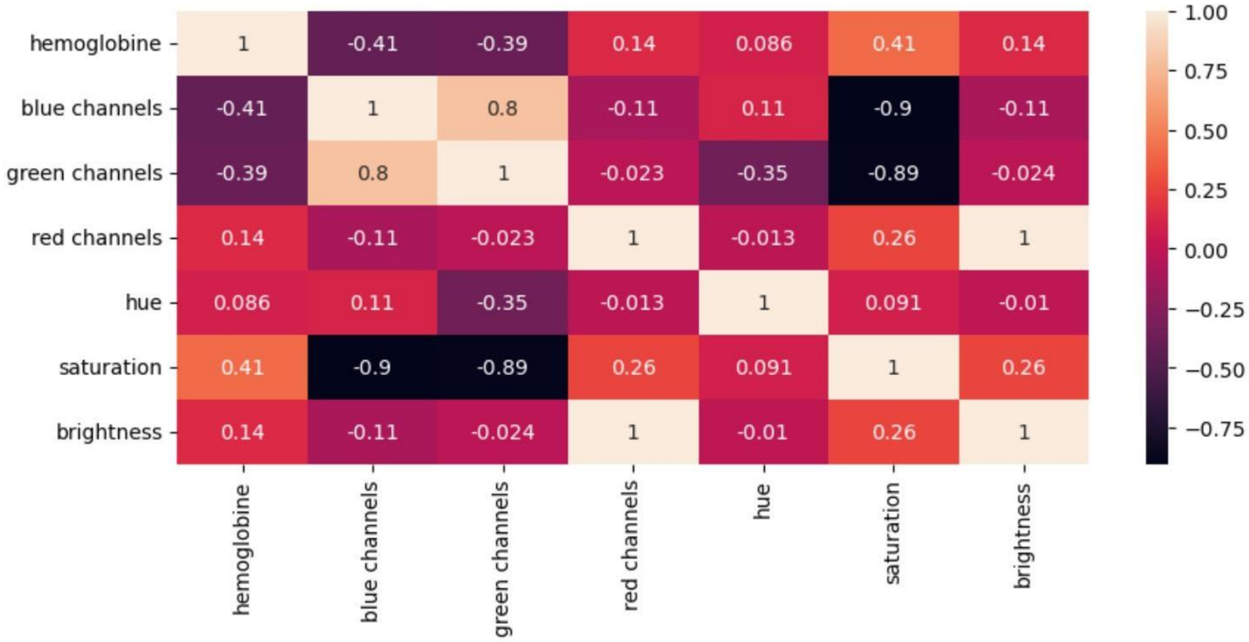
$$H_a: \mu_b \neq \mu_g$$

$$z_{stat} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(x)}{n} + \frac{D(Y)}{m}}} = \frac{65.4 - 65.3}{\sqrt{\frac{36.6^2}{425} + \frac{39.2^2}{425}}} \approx 0.038$$

$$z_{crit} = -1.96; 1.96$$

-1.96 < 0.038 < 1.96, so we do not have enough evidence to reject the null hypothesis and accept the alternative hypothesis that the number of blue channels does not equal to the number of green channels.

Correlation matrix:



The correlation matrix shows the correlation coefficients between all variables. Note the strong positive correlation between the variables blue channels and green channels (0.8), as well as the strong negative correlation between the variables saturation and blue channels (-0.9), saturation and green channels (-0.89). All other correlations are either average or weak.

Bivariate analysis:

A linear regression needs to be performed to understand the relationship between hemoglobin and the number of blue channels.

	coef	std err	t	P>|t|	[0.025	0.975]
const	142.6926	2.861	49.882	0.000	137.070	148.315
blue channels	-0.3554	0.038	-9.311	0.000	-0.430	-0.280

The model is:

$$\hat{Y} = 142.69 - 0.3554 * X_1$$

The obtained result indicates that each subsequent blue channel will reduce hemoglobin by 0.3554.

We also note that, judging by the p-value, which turned out to be equal to 0, this variable is significant at 1% and 5% significance levels.

R-squared:	0.170
Adj. R-squared:	0.168
F-statistic:	86.70
Prob (F-statistic):	6.96e-19
Log-Likelihood:	-2029.4
No. Observations:	425
Df Residuals:	423
AIC:	4063
BIC:	4071

R-squared of 0.17 indicates that the blue channel variable can explain 17% of the variance in the hemoglobin variable.

Multivariate analysis:

	coef	std err	t	P>|t|	[0.025	0.975]
const	139.7134	26.040	5.365	0.000	88.528	190.899
blue channels	-0.3954	0.154	-2.572	0.010	-0.698	-0.093
green channels	-0.0993	0.121	-0.823	0.411	-0.337	0.138
red channels	-0.8303	2.815	-0.295	0.768	-6.363	4.702
hue	0.0429	0.029	1.497	0.135	-0.013	0.099
saturation	-0.1024	0.129	-0.794	0.428	-0.356	0.151
brightness	0.9512	2.836	0.335	0.737	-4.623	6.525

The model is:

$$\hat{Y} = 139.7 - 0.3954X_1 - 0.0993X_2 - 0.8303X_3 - 0.0429X_4 - 0.1024X_5 + 0.9512X_6$$

The coefficients show how much hemoglobin will change if a certain variable is increased by 1, all other things being equal.

For example, with an increase in X1 by 1, hemoglobin will decrease by 0.3954, with an increase in X2 by 1, hemoglobin will decrease by 0.0993, and so on.

It should be noted that the only statistically significant variable is the number of blue channels. All other variables are insignificant at the 1% and 5% significance levels.

R-squared:	0.198
Adj. R-squared:	0.187
F-statistic:	17.22
Prob (F-statistic):	8.02e-18
Log-Likelihood:	-2022.1
No. Observations:	425
Df Residuals:	418
AIC:	4058
BIC:	4087
Df Model:	6

R-squared of 0.198 indicates that the blue channel variable can explain 19.8% of the variance in the hemoglobin variable.

Conclusion:

In conclusion, our research on hemoglobin analysis has provided valuable insights into the factors influencing hemoglobin levels. By determining the average values of hemoglobin within our dataset, we have established a baseline for comparison and further analysis. Furthermore, our findings indicate that blue channels have the most significant influence on hemoglobin levels.

These results highlight the importance of considering diverse factors, including color channels, in understanding hemoglobin levels. Future research could delve deeper into the specific mechanisms through which blue channels impact hemoglobin levels and explore potential applications in healthcare and medical diagnostics.

REFERENCES

1. Das, M., & AKhurshid, A. (2020). Image processing to quantitate hemoglobin level for diagnostic support. Helix, 10(1), 228–234. https://doi.org/10.29042/2020-10-1-228-234

2. Chen, Y., Hu, X., Zhu, Y., Liu, X., & Yi, B. (2024). Real-time non-invasive hemoglobin prediction using deep learning-enabled smartphone imaging. BMC Medical Informatics and Decision Making, 24(1). https://doi.org/10.1186/s12911-024-02585-1

3. Kato, S., Chagi, K., Takagi, Y., Hidaka, M., Inoue, S., Sekiguchi, M., Adachi, N., Sato, K., Kawai, H., & Kato, M. (2024). Machine/deep learning-assisted hemoglobin level prediction using palpebral conjunctival images. British Journal of Haematology. https://doi.org/10.1111/bjh.19621

4. https://www.who.int/data/#collection