



Quantitative Approach for Anemia Detection Using Regression Analysis

Vinit P. Kharkar^{1*}, Ajay P. Thakare²

¹Assistant Professor, Department of Electronics & Telecommunication Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera-Amravati, MH, India.

²Professor, Department of Electronics & Telecommunication Engineering,
Sipna College of Engineering and Technology, Amravati, MH, India
Emails: vinitkharkar27@gmail.com; apthakare40@rediffmail.com

Abstract

Anemia, generally termed as deficiency of hemoglobin or red blood cells in the blood is significant global health concern for the population in underdeveloped as well as in developing nations specially, for children and young women in rural areas. This paper proposes a quantitative approach for anemia detection by regression analysis technique which predicts hemoglobin level in the blood. To achieve this, the image dataset of microscopic blood sample is collected from 70 individuals. The data collection requires proper procedure as it plays vital part in system implementation. The statistical feature utilizing mean pixel intensity values from the red, green, and blue color planes of the images are given as input to the regression model. For the proposed system, we have employed multiple regression analysis model using machine learning approach with both three and four regression coefficients to establish relation between features obtained from blood samples and the hemoglobin level in the blood to achieve the specified task of anemia detection in an individual. Performance analysis show promising results for the proposed system with coefficient of determination (R^2) and root mean square error (RMSE) found out be 0.923 and 1.682 respectively. Overall, this paper presents valuable system for anemia detection based on hemoglobin estimation which can be implemented in areas with limited medical resources and gives another supportive technological solution for current healthcare problems.

Received: August 02, 2023 Revised: November 25, 2023 Accepted: April: 13 2024

Keywords: Anemia; Hemoglobin; Microscopy; Regression Analysis.

1. Introduction

Anemia poses a significant global health risk, particularly affecting populations in underdeveloped and developing countries, notably children and young women in rural regions. It is health condition is typified by a deficiency of hemoglobin (HB) or red blood cells (RBC) in the blood, which reduces the amount of oxygen that reaches the tissues and organs. Both the general socioeconomic growth of communities and countries, as well as the health, happiness, and productivity of people, can be significantly impacted by anemia. Worldwide, 42% of children under age of 5 and 40% of pregnant women are projected to be anemic, according to the World Health Organization (WHO). The World Health Assembly has supported global nutrition targets, one of which is a 50% decrease in anemia among women of reproductive age by 2025[1,2]. The results of the National Family Health Survey-5 (NFHS-5) show that the widespread presence of anemia is rising in most Indian states among women and children [3]. Identifying hemoglobin content in the bloodstream is key factor for identifying the anemic condition in an individual. The reference ranges for hemoglobin level in the blood for anemia detection as per WHO guidelines are as shown in table 1.

Table 1: Hemoglobin ranges for Anemia Detection (in g/dl)

| Subjects | Anemia Level | | | |
|------------------|--------------|------------|-------------|------------|
| | Severe | Moderate | Mild | Non-anemic |
| Children | | | | |
| 6-59 Months | < 7.0 | 7.0 - 9.9 | 10.0 – 10.9 | > 11.0 |
| 5-11 Years | < 7.0 | 8.0 - 10.9 | 11.0 – 11.4 | > 11.5 |
| 12-14 Years | < 7.0 | 8.0 - 10.9 | 11.0 – 11.9 | > 12.0 |
| Women | | | | |
| Non-Pregnant | < 7.0 | 8.0 - 10.9 | 11.0 – 11.9 | > 12.0 |
| Pregnant | < 7.0 | 7.0 - 9.9 | 10.0 – 10.9 | > 11.0 |
| Men | | | | |
| 15 Years & Above | < 7.0 | 8.0 - 10.9 | 11.0 – 12.9 | > 13.0 |

One of the most often used diagnostic tests for the identification of anemia is hemoglobin investigation. There are different laboratory methods such as cyanmethemoglobin method which also considered as gold standard method, hemocrit method, complete blood count methods are available in the diagnostic field for hemoglobin detection. since, they are laboratory methods they require clinical setup, expert technician as well as fair amount of time for getting results. This paper proposes a new, simple-to-use method for measuring hemoglobin levels in blood to detect anemia based on regression analysis [4].

Hemoglobin levels are one of the most vital features of an anemia diagnosis, and regression analysis is one of the important tools for detection of anemia since it helps researchers to study the complicated relations of various factors. Regression analysis is very important in anemia detection because it helps to establish the relationship between an independent variable (for instance, features obtained from blood samples) and a dependent variable (for example, hemoglobin levels). The key function of regression analysis is its capacity to create the dependence between the extracted features from the microscopic blood sample images and the actual hemoglobin levels. Moreover, regression analysis enables the development of a predictive model which in turn can generalize well to a new, unseen data and thereby increases the system's ability to make reliable predictions on heterogeneous patient populations.

Considering this, we have developed predictive models to estimate the hemoglobin levels for anemic condition identification which are based on the input features of digital microscopic images of blood sample smears with the help of regression analysis. The research paper further explains the related literature work in the area of anemia identification and screening using latest technologies. The subsequent section highlights the details of various stages of system development using regression analysis along with relevant mathematical analysis. The in-dept performance analysis is carried out in last section which depicts the results related with experimentation. The research paper concludes with a conclusion and future implications for the suggested approach.

2. Related Work

The orthodox methodology of diagnosing anemia is dependent on the analysis of several lab tests used for defining and distinguishing various forms of anemia. The result of the Complete Blood Count (CBC) is comprised of RBC indices including but not limited to hemoglobin mass, hematocrit, and mean corpuscular volume (MCV). By applying the Peripheral Blood Smear (PBS) examination, the assessment will be done concerning RBC morphology, which includes microcytosis, macrocytosis, and hypochromia. Reticulocyte count is an assay for recycling RBCs into circulation which is an indicator of bone marrow activity and provides a difference in distinguishing amongst the various causes of anemia. Iron studies comprehensively assaying iron indices: i.e., serum iron, total iron binding capacity (TIBC), and serum ferritin, enable the proper diagnosis of anemia of deficiency of iron. These pathology methods collectively provide a comprehensive approach to anemia detection but has some limitation such as cost and accessibility of certain tests may limit their widespread use, especially in resource-constrained settings. The use of certain methods including cyanmethemoglobin test technique involves dealing with hazardous chemical and needs highly skilled

technician to operate the tests. Considering this, the development of modern technologies is essential to be able to find a solution to the problem with anemia testing, and therefore make it more useful for social purposes [5,6].

The different researchers have put forward various approaches for anemia detection which involves both minimally invasive and non-invasive methods. Anemia modelling is about to forecast anemia in a population using biomedical variables by application of multiple linear regression analysis that is aimed by the researchers Murat Sari and Arshed A. Ahmad [7]. The study is based on a data set of 539 subjects which was obtained from the blood laboratories. Many factors from the biomedical domain, comprising eight blood levels, sex, and age, are analyzed for diagnosing the type of anemia. The paper examines the mathematical model developed via multiple regression analysis and demonstrates its superiority over linear deep learning techniques. Introduce blood variables, sex, and age into the regression equation, which makes it a better predictor of anemia types than deep learning methods.

Researchers also employed a minimally invasive technique to count red blood cells to confirm anemia and leukemia [8]. The offered method makes use of microscopic pictures from the ALL-IDB datasets, which are one of the most widely used datasets for automated recognition, counting, and categorizing blood cells by algorithm. The authors have employed the RBC counting and grading method for the analysis of anemia patient. Efforts in this paper are made to classify RBCs and their types including the normal ones and such associated to anemias like the Target Cells and the Elliptocytes through the SVM algorithm. The data of local medical institutes was used to conduct research.

Classification of RBCs from peripheral blood smear images is another option of implementing a diagnosis of anemia and other related disorders [9]. Authors have collected database of the PBS images from patients of medical institute and available online on internet. These PBS images from database have been preprocessed using an adaptive histogram algorithm to ensure perpetual contrast and better illumination purpose. These images are then processed with machine learning algorithm with help texture features and gray level co-occurrence matrix (GLCM) to obtain the classification of blood cell into red blood cell and white blood cell. This classification provides information about red blood cell which is important component related which hemoglobin in the blood and can be useful for anemia screening in patient.

Putut Dewantoro proposed a method based on PPG signals correlating with pulse strength measured non-invasively from the patient's fingertips [10]. The system comprises digitization, hemoglobin level measurement, and an Android app. It utilizes LEDs and a photodiode for signal acquisition, employing OP-AMP MPC 6002 and ATmega328PP-AU for processing. Hemoglobin levels are estimated through light component analysis and linear regression. Data is stored internally and transmitted to a smartphone via BLE 4.0 for monitoring. Thirty subjects were tested by both noninvasive and invasive methods. The root means square error, maximum difference between the two methods and standard deviation are 1.56 gm/dL, 3.7 gm/dL and 0.245 gm/dL respectively for given method.

The researcher Akshaya Krishan et. al. has proposed digital image processing-based method for hemoglobin identification using invasive approach [11]. The primary idea behind the digitalization of hemoglobin predict is to employ fundamental MATLAB image processing algorithms to obtain rapid results. The author's first methodology is histogram analysis, which failed to produce adequate findings. The intensity concentrations in the first approach used by the authors did not change with the image orientation, so it cannot be differentiated as a discrete method to find the hemoglobin range. They subsequently switched to studying the pixel region, which is intensity levels according to RGB scaling. The proposed second method based on Pixel region using MATLAB provides more accurate result as compared to first one. on the trials of two main methods, the authors concluded that out of two methods, the RGB colour scale method was appropriate to desired results. They determined that the range of hemoglobin levels occurs at a certain RGB color level.

While there have been a great number of empirical and theoretical works revolving around important and cutting-edge topics, the researcher community mainly relied on models that are single-variable-based in their studies. Thus, the present study explains the type of anemia employing a substantially larger group of observational variables and gives more credible results. What makes multiple regression analysis a favorite modeling technique among researchers is that it serves to address numerous issues ranging from anemia to others. Hence, it is applied in the present biomedical problem.

3. Methodology

This section gives the brief idea about the system presented in this research article. This system development section establishes the context and significance of developing an anemia detection system using regression analysis with machine learning approach.

3.1 System Architecture

The block diagram in the figure 2 below demonstrates the flow of data and processes with the emphasis on the data collection, preprocessing, and regression analysis model implementation. This part of the article highlights the role of the system in accurate identification of anemia using computational techniques.

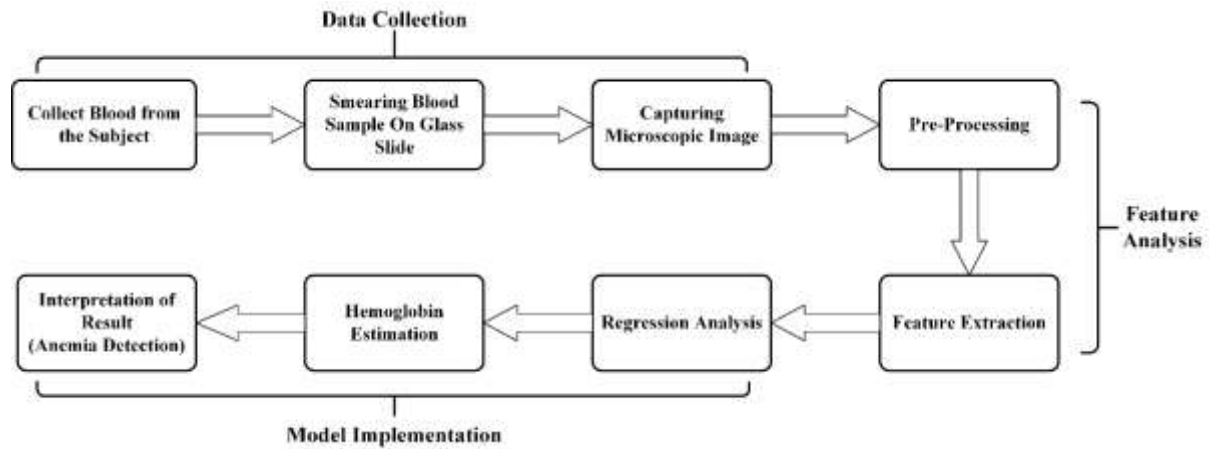


Figure 1: Block diagram of Anemia Detection System

The architecture of the system for anemia detection involves 8 distinct sequential stages. At first, blood samples are taken from patients, later, the smears of the samples will be made on the glass slide and then microscopic images of the samples will be taken for analysis. These images are subjected to preprocessing to enhance the quality and improve the image and then features that are relevant to anemia are extracted. Consequently, the features are used in regression analysis to predict anemia presence and severity. The hemoglobin level is then estimated based on the regression analysis outcomes. The last step is the results' interpretation where it is usually done by using a reference range of hemoglobin concentration and having it compared with diagnostic thresholds. This proposed approach involves integrating sample acquisition with the image processing, feature extraction, regression analysis, hemoglobin estimation, and result interpretation to be able to accurately diagnose the anemia which are discussed in detail further sections.

3.2 Dataset Collection

In this section, process relate to database is discussed in detail. For this, blood samples are collected from individuals across various age groups to study hemoglobin levels. A digital image of each blood sample is captured to extract color information, which is then utilized to estimate the hemoglobin levels in the blood [12].

Creating a microscopic picture of a blood smear sample by being careful about different issues assists in having an errorless and insightful outcome. Above else, samples preparation should be emphasized, with this, care taken is needed when blood is spread on slides, proper fixation applied, and staining techniques used to improve cellular contrast. Choice of proper microscope settings like magnification of the objective lens and modification of brightness, contrast, and focus levels is important and could be a vital factor in the imaging process. Fidelity and accuracy can be decreased in the imaging capture if we do not have the microscope and digital camera system compatibility and calibration.

Additionally, laboratory settings should be monitored to ensure that the environmental factors are kept constant to prevent fluctuations, such as temperature, humidity, and lighting conditions. Camera specifications, for example what resolution would be best, and the sensitivity would be important to consider in order to produce a clear picture. Moreover, retouching image capture settings by adjusting things such as exposure time and aperture has a positive effect on the quality of the images while minimizing noise and artifacts at the same time. Due to the attention to focus and depth of field the techniques certainly guarantees that all the cellular layers related to blood smear be in focus, especially if the blood is thick or uneven.

The glass slide is prepared by smearing volume of 10 µlitre (about 1-2 drops) blood sample on glass slide and is covered with glass cover slip so that it should be uniformly spreaded. Smearred glass slide of a blood sample is then placed under microscope to capture images of blood sample. The images of blood sample are captured with high resolution camera mounted on a microscope with 10x magnification. In this implementation, we are using images having pixel size 1024 x1280 pixel with .bmp format. Several file formats based on different compression techniques are designed and recommended to store image data after acquisition. Most widely used image file formats are .jpg, .png, .bmp, .gif, etc. Out of these, .bmp is a lossless file format in which no data after acquisition can be lost or compressed. This file format has the capability of storing data pertaining to colors

without distortion. Because of this reason, it is mostly use for medical image analysis purpose and file size of the image with .bmp extension is also large as compared with others Some of the images from collected database of blood samples are as shown:

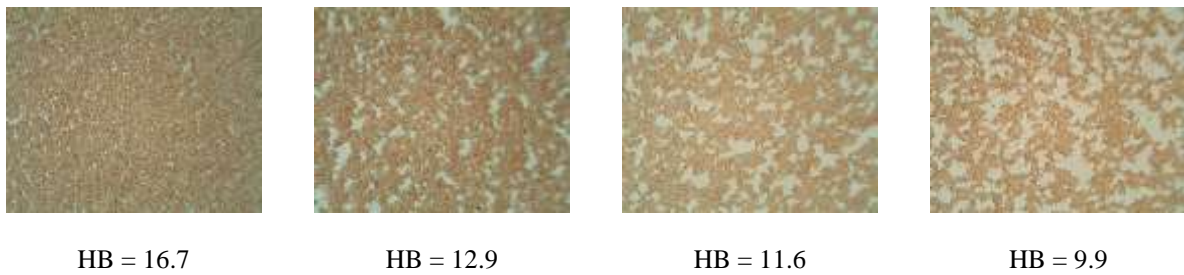


Figure 2: Blood sample images with corresponding HB value.

3.3 Feature Extraction

Hemoglobin is found within red blood cells and allows the red blood cells to bind to oxygen molecules Each red blood cell constitutes about one third portion of hemoglobin by volume. The red blood cell makes up the significant bottom part of the human blood. Blood is typically depicted as red in color when oxygenated and a darker red or even blue when deoxygenated. This alteration in color occurs because of hemoglobin, which is a protein present in red blood cells, responsible for transporting oxygen in the body [13]. Consequently, it can be concluded that the color of blood provides a reliable indication of the available hemoglobin levels. Traditionally, it is common practice to observe the color of blood using direct vision or chemical agents to estimate the amount of Hb in blood. These methods employ human perception to match colors. As a result, the color of the blood can influence the determination of Hb. Our approach is to replace subjective human perception regarding color with quantitative machine vision. Therefore, we opted to devise a method for transferring the color data of each blood sample into a digital image file. With reference to this, the chromic variation for hemoglobin is as shown in figure 3 below.

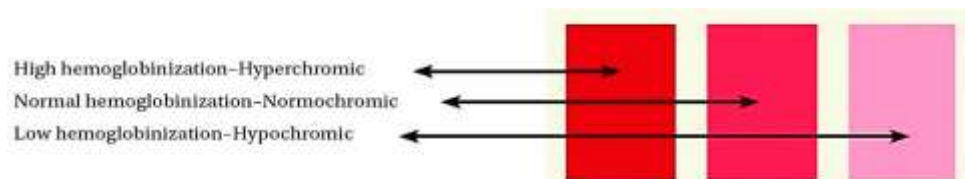


Figure 3: Chromic variation of blood for Hemoglobin (colors are indicative purpose only)

Since the hemoglobin estimation is directly based on colour of blood sample as discussed earlier, so the pixel intensity value is the primary information, and it is the most popular and important feature which can be for mathematical modeling. The statistical feature used is mean values of pixel intensities from red, green and blue plane from RGB colour space of a digital image as feature vector [14] and provided as an input to for regression analysis as discussed in next section.

3.4 Mathematical Modelling

Multiple regression analysis is a method of statistical analysis that examines the relationship between a single dependent variable and multiple independent variables. It aims to utilize known values of independent variables to forecast the value of the dependent variable. It explores the linkages between numerous independent factors and a dependent variable, allowing for a better understanding of inter-variable relationships and predictive modeling [15]. It aids in determining which elements are strongly connected with the outcome, allowing for a better knowledge and prediction of disease development, treatment efficacy, and other clinical outcomes using data retrieved from medical imaging. Interpreting the results of multiple regression involves assessing the significance of each coefficient with the help of measures such as R-squared can indicate the overall fit of the model. So, mathematical modelling is employed to Construct a regression model that predicts hemoglobin level for anemia status based on the selected independent variables [16].

In this proposed approach of anemia detection using multiple regression analysis, the primary goal is to determine the best set of regression co-efficient, b_0 , b_1 , b_2 , and b_3 , such that the model predicts the dependent variable i.e., Hb as accurately as possible from the features of image of blood sample i.e. dependent variables. In this implementation, we are using multiple regression analysis to calculate hemoglobin in blood using mean of R, G, and B planes which are given as input for regression analysis.

The details procedure of multiple regression analysis with three regressions co-efficient is as follows:

For a regression model with three regression coefficients (b_0 , b_1 , & b_2), the procedure involves calculating these coefficients using equations derived from the sum of products of the dependent variable (Hb) and each independent variable (x , y , z). The equations are as follows:

$$Hb = b_0 * x + b_1 * y + b_2 * z \quad (1)$$

The regression co-efficient b_0 , b_1 and b_2 are calculated using equations

$$\sum Hb * x = b_0 \sum x^2 + b_1 \sum x * y + b_2 \sum x * z \quad (2)$$

$$\sum Hb * y = b_0 \sum x * y + b_1 \sum y^2 + b_2 \sum y * z \quad (3)$$

$$\sum Hb * z = b_0 \sum x * z + b_1 \sum z * y + b_2 \sum z^2 \quad (4)$$

here, b_0 , b_1 , b_2 : Regression co-efficient.

x , y , z : features of RGB color space.

For a regression model with four regression coefficients (b_0 , b_1 , b_2 , and b_3), the procedure is similar, but an additional term (b_3) is added to the regression equation. The equations for calculating the coefficients are as follows:

$$Hb = b_0 * x + b_1 * y + b_2 * z + b_3 \quad (5)$$

The regression co-efficient b_0 , b_1 , b_2 and b_3 are calculated using equations

$$\sum Hb * x = b_0 \sum x^2 + b_1 \sum x * y + b_2 \sum x * z + b_3 \sum x \quad (6)$$

$$\sum Hb * y = b_0 \sum x * y + b_1 \sum y^2 + b_2 \sum y * z + b_3 \sum y \quad (7)$$

$$\sum Hb * z = b_0 \sum x * z + b_1 \sum z * y + b_2 \sum z^2 + b_3 \sum z \quad (8)$$

$$\sum Hb = b_0 \sum z + b_1 \sum y + b_2 \sum z + b_3 * n \quad (9)$$

here, b_0 , b_1 , b_2 , b_3 : Regression co-efficient.

x , y , z : features of RGB color space

n : total no. of observations.

Then, the regression coefficients can be formed and consequently a regression model will be obtained with RGB plane's mean values as input feature values. The model can be used to explicate hemoglobin levels in blood making use of these potential features.

Generally, the multiple regression analysis is the computation of regression coefficients using the equations obtained from the sum of products of the dependent variable and each of the independent variable. These coefficients are used to develop a regression model, which is used to predict the dependent variable from independent variables. This technology can be used to estimate hemoglobin concentration in blood from its reflectance spectrum in RGB color space.

4. Results

In this section, we present the results of our study on estimating hemoglobin levels for anemia detection using dataset collected from individuals across diverse age groups as discussed in earlier section. The data obtained from the digital images were subjected to rigorous analysis to determine effectiveness of the proposed approach.

To analyze the performance of the proposed system, dataset is collected from 70 subjects typically encompasses individuals from diverse demographic backgrounds, ranging from various age groups, genders, and socio-economic statuses. This rigorous endeavor involves the meticulous scrutiny and interpretation of multifaceted

data points derived from the activities, responses, and interactions of the subjects within the experimental framework. Initially the values for regression co-efficient are calculated using the mathematical modelling as explained in the earlier part. The results for regressor coefficient are listed below in the table:

Table 2: Regression Coefficient Analysis

| No. of Regressor | Values of Regression Co-efficient | | | |
|----------------------|-----------------------------------|---------|--------|---------|
| | b_0 | b_1 | b_2 | b_3 |
| 3 - Regressor | -0.3 | 1.16 | -0.82 | - |
| 4 - Regressor | 0.1706 | -0.2036 | 0.0005 | 18.9179 |

The experiment employs a 70 samples dataset which is stratified into a training and testing data. The training set is used to build and train the model, and the test set measuring the model's performance. With 60 samples comprising the 85% of the whole dataset, the task of the model is made easier to learn from maximum data to ascertain the patterns and relations. Besides, the 10 last samples (15%) are kept for the purpose of testing and verifying the model's performance. During the training phase model parameters are calibrated with training data to minimize the deviation between actual observed values and forecast results. The model is further analyzed through evaluation process using the testing set, employing performance metrics like Mean Square Error, Pearson's correlation coefficient, co-efficient of determination to determine its generalization capability to unseen and new data.

In our investigation, we employed multiple regression analysis to explore the relationships between the RGB (Red, Green, Blue) plane and their corresponding errors, focusing on both sets of regression coefficients. These coefficients are the ones obtained from regression equations that contain three regression coefficients and those from regression equations that contain four regression coefficients. Feature vectors were computed based upon the means of the RGB plane, with the value of hemoglobin forming the output variable of interest. In here the mean values for the RGB plane were designated as independent variables prior to the HB value which is the dependent one. The results are shown in table 3.

Table 3: Results of Multiple Regression Analysis

| SN | HB | Output | | Error | | Anemia Status |
|----|------|-------------|-------------|-------------|-------------|---------------|
| | | 3-Regressor | 4-Regressor | 3-Regressor | 4-Regressor | |
| 1 | 13.9 | 12.81 | 14.08 | 1.09 | 0.18 | Non-Anemic |
| 2 | 13.8 | 14.89 | 13.86 | 1.09 | 0.06 | Non-Anemic |
| 3 | 13.2 | 13.83 | 13.95 | 0.63 | 0.75 | Non-Anemic |
| 4 | 11.7 | 12.76 | 12.98 | 1.06 | 1.28 | Non-Anemic |
| 5 | 11.3 | 15.14 | 13.21 | 3.84 | 1.91 | Non-Anemic |
| 6 | 11 | 14.39 | 12.24 | 3.39 | 1.24 | Non-Anemic |
| 7 | 10.8 | 13.48 | 12.49 | 2.68 | 1.69 | Anemic |
| 8 | 10.5 | 13.52 | 12.60 | 3.02 | 2.10 | Anemic |
| 9 | 9.3 | 13.61 | 11.20 | 4.31 | 1.90 | Anemic |
| 10 | 7.8 | 12.01 | 10.96 | 4.21 | 3.16 | Anemic |

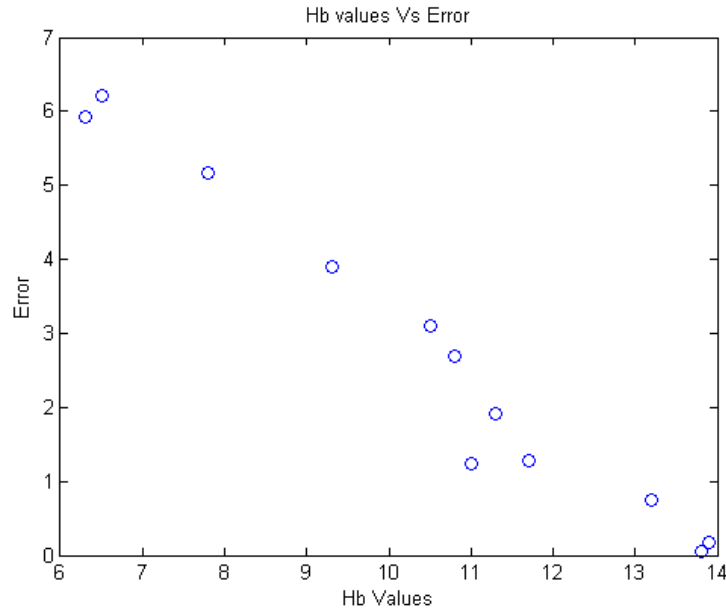


Figure 4: HB value versus Error for 4-regression co-efficient.

5. Discussion

In multiple regression analysis, R^2 , and R should be considered as main statistics for checking how good the model works as predictor of hemoglobin while RMSE as main statistics for the accuracy of the model. The Co-efficient of Determination, R^2 spanning from 0 to 1, the indicator of how much different independent variables included in the regression analysis explain the variability of hemoglobin levels in the participants. Thus, a larger R^2 value denotes that model conforms the model well with the analysis data that it can predict level of hemoglobin concentration well. The correlation coefficient hypothesis (R) presents the strength and contains the direction of the linear relation between hemoglobin and the predictor variables. A value near 1 or -1 means a strong positive or a negative relationship essentially, these numbers extrapolate the proportion of changes between the predictors and the hemoglobin level. RMSE estimators show the level of difference between measured and predicted hemoglobin values, revealing the precise predictive efficiency of the algorithm. Minimality of error in estimation of hemoglobin levels is reflected when lower RMSE values are there which in turn reflects a good predictive performance.

In hemoglobin estimation using multiple regression, these statistical parameters collectively aid in assessing the model's performance, understanding the relationships between variables, and determining the accuracy of predicted hemoglobin levels. The statistical parameter obtained for proposed multiple regression model is as follows.

Table 4: Statical Parameter Analysis

| SN | Statistical Parameter | 3-Regressor | 4-Regressor |
|----|---|-------------|-------------|
| 1 | Co-efficient of Determination (R^2) | 0.154 | 0.923 |
| 2 | Pearson's correlation coefficient (R) | 0.392 | 0.961 |
| 3 | Root Mean Square Error (RMSE) | 2.876 | 1.682 |

For the proposed methodology, it can be seen that statistical parameter for 4-regressor model is better than 3-regressor model which shows that the number of observations i.e. no. of blood samples plays a significant role in the accuracy. So, it evident that the number of observations should be more for implementation purpose. The results of statistical analysis in anemia detection have significant clinical implications. Accurate predictive models can assist healthcare professionals in estimating Hb levels non-invasively, aiding in timely diagnosis and treatment decisions. Additionally, statistical analysis can contribute to a deeper understanding of anemia's

pathophysiology and its correlation with image-derived features, potentially leading to advancements in patient care sets. Once the hemoglobin values are estimated using proposed approach, anemic status of an individual can be determined by comparing obtained hemoglobin value with reference hemoglobin values for anemia as depicted in table 1. Such a system with minimum requirement is extremely beneficial for the rural tribal areas with insufficient diagnostic facilities for anemia detection. This methodology proposed in this system provides a possible breakthrough for improved healthcare facility in area with low medical resource availability, thus contributing to the social cause of the nation.

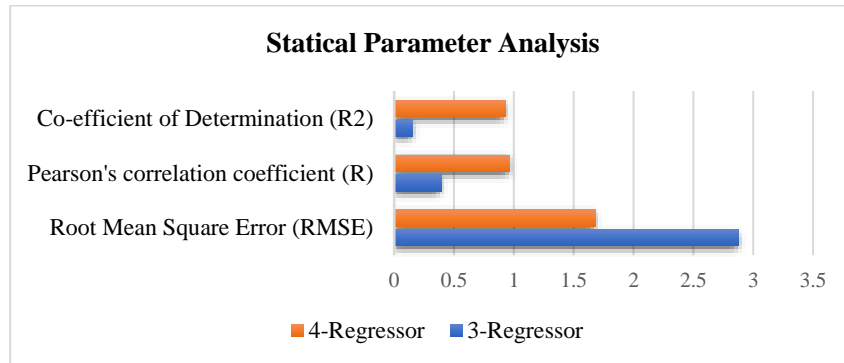


Figure 5: Statistical Parameter Analysis

6. Conclusion

As anemia - a disorder related with hemoglobin level in the blood is emerging as a major social concern for the developing countries like India. The research work presented in the paper helps to tackle this alarming situation by providing useful method for anemia detection with the help of hemoglobin estimation. The paper explains data collection process in depth which is the main factor. The system uses multiple regression analysis approach to establish the relation between HB Value and extracted feature from microscopic images of blood smear. The multiple regression is carried out for 3-regressor and 4-regression coefficient. The performance evaluation shows that multiple regression analysis model with 4-coefficient shows the promising results for the proposed approach. The values of co-efficient of determination (R^2) and root mean square error (RMSE) confirms the effectivity of 4-regressor model over the 3-regressor. Thus, the number of observations plays a significant role in the implementation of the system. Once the Hemoglobin value is predicted with the help of proposed system the anemic status of the individual can be determined from the WHO defined hemoglobin ranges for anemia. Thus, this paper gives a simple to use, efficient method for anemia detection using hemoglobin estimation in low resource setting areas in developing country like India. Also, in this digital age of technological advancements, using cutting edge technology to provide innovative solutions to traditional medical practices is advantageous for both therapeutic and diagnostic purposes. These medical innovations using technology aim to support current healthcare systems in making it faster, better, and more accurate assessments to enhance patient care, rather than replacing the existing healthcare system.

References

- [1] World Health Organization (WHO). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System. Geneva: World Health Organization; 2011. Report No.: WHO/NMH/NHD/MNM/11.1. Available from: <http://www.who.int/vmnis/indicators/haemoglobin.pdf> Last Accessed 28 February 2024.
- [2] Safiri, S., Kolahi, A. A., Noori, M., et al. Burden of anemia and its underlying causes in 204 countries and territories, 1990–2019: Results from the Global Burden of Disease Study 2019. *Journal of Hematology & Oncology*, 14 (185). (2021) 10.1186/s13045-021-01202-2.
- [3] Government of India, Ministry of Health and Family Welfare, National Family Health Survey - 5, 2019-21. Available at: http://rchiips.org/nfhs/NFHS-5_FCTS/India.pdf Last accessed 10 March 2024.
- [4] Whitehead R. D., Mei, Z., Mapango, C., & Jefferds, M. E. D., "Methods and analyzers for hemoglobin measurement in clinical laboratories and field settings", *Annals of the New York Academy of Sciences*, 1450(1), 147-171 (2021). 10.1111/nyas.14124

- [5] An, R., Huang, Y., Man, Y., Valentine, R. W., Kucukal, E., Goreke, U., Sekyonda, Z., Piccone, C., Owusu-Ansah, A., Ahuja, S., Little, J. A., & Gurkan, U. A. Emerging point-of-care technologies for anemia detection. *Lab Chip*, 21(10), 1843-1865 (2021) DOI: 10.1039/d0lc01235a.
- [6] Srivastava, T., Negandhi H., Neogi, S. B., Sharma, J., Saxena, R. Methods for Hemoglobin Estimation: A Review of What Works. *Journal of Hematology Transfusion*, 2(3), 1028. (2014). DOI: 10.47739/2333-6684/1028.
- [7] Murat Sari, Arshed A. Ahmad, Anemia Modelling Using the Multiple Regression Analysis, *International Journal of Analysis and Applications*. 17(5), 838-849 (2019).
- [8] A. A. Aslani, M. Zolfaghari and H. Sajedi. Automatic Counting Red Blood Cells in the Microscopic Images by EndPoints Method and Circular Hough Transform. In: 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea, pp. 1-5 (2022) DOI: 10.1109/IMCOM53663.2022.9721754.
- [9] K. T. Navya, K. Prasad and B. M. K. Singh, Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques, In: 2nd Global Conference for Advancement in Technology. pp. 1-4. (2021) DOI: 10.1109/GCAT52182.2021.9587524
- [10] P. Dewantoro, C. E. Gandana, R. O. R. H. Zakaria and Y. S. Irawan. Development of Smartphone based Non-Invasive Hemoglobin Measurement In: 2018 International Symposium on Electronics and Smart Devices (ISESD). pp. 1-6. (2018) doi: 10.1109/ISESD.2018.8605489.
- [11] Krishnan, A., Srikanth, A., Robin, S. K. B., & Kulkarni, S. Digitized estimation of haemoglobin using image processing. *International Journal of Engineering Research and General Science*. 5(3), 71-76. (2017)
- [12] Nithya, R., & Nirmala, K. Detection of Anaemia using Image Processing Techniques from microscopy blood smear images. *Journal of Physics: Conference Series*, 2318(1), 012043. (2022). DOI: 10.1088/1742-6596/2318/1/012043
- [13] D. P. Lokwani. Red Blood Cells. In: *The ABC of CBC: Interpretation of complete Blood Count & Histograms*. 1st ed. Jaypee Brothers Medical Publishers Ltd, New Delhi, India, pp 8-9. (2013).
- [14] Nixon, M. S., & Aguado, A. S. *Feature Extraction and Image Processing for Computer Vision* (4th ed.). Academic Press. (2020).
- [15] Poon E., Feng C., Univariate and Multiple Regression Analyses in Medical Research. *Biometrical Letters*. 60(1), 65-76. (2023) DOI:10.2478/bile-2023-0005
- [16] Mukhopadhyay K, Singh R, Dalai CK, Ahmed SN, Banerjee K. Understanding the application of multiple linear regression model in health care research using simulation with computer generated data. *Journal of The West Bengal University of Health Sciences*. 1(2), 44-52. (2020).
- [17] Acharya, S., et al. (2020). Non-Invasive Estimation of Hemoglobin Using a Multi-Model Stacking Regressor. *IEEE Journal of Biomedical and Health Informatics*, 24(6), 1717-1726. 10.1109/JBHI.2019.2954553.
- [18] Kharkar, V. P. and Thakare, A. P. A Comprehensive Review of Emerging Technologies for Anemia Detection. 8th International Conference on Signal Processing and Communication (ICSC), Noida, India, pp. 230-235. (2022) DOI: 10.1109/ICSC56524.2022.10009310