Proceedings of ICBioMed 2025 Symposium: Computational Modelling and Simulation for Biology and Medicine
DOI: 10.54254/2753-8818/2025.LD26658

Statistical Modeling on Visual Detection of Anemia

Yiyang Shen

College of Biological Science, University of California Davis, Davis, USA
shen3519004721@gmail.com

Abstract. Anemia, a prevalent blood disorder, disproportionately impacts vulnerable populations globally. Traditionally, its diagnosis relies on invasive blood tests that measure hemoglobin levels, necessitating laboratory equipment and skilled personnel. This requirement often restricts access in resource-limited settings. This has spurred interest in non-invasive diagnostic methods, especially those using imaging technologies. This review focuses on image-based methods and the role of statistical modeling. Visual signals from images, such as pixel intensity values, are transformed into predictors and used in models like logistic regression (LR), support vector machines (SVM), decision trees (DT), and k-nearest neighbors (kNN). Among them, kNN is often the most precise and interpretable, especially in studies with small sample sizes. Still, challenges remain. Variations in image quality, lighting, and preprocessing can affect model performance, and visual features may not detect all anemia types. Further work is needed to improve standardization and expand validation. Overall, statistical models help transform image data into meaningful clinical insight and support the development of accessible, non-invasive anemia diagnostics.

Keywords: Anemia, hemoglobin, non-invasive, image analysis, statistical modeling

1. Introduction

Anemia is a blood disorder that impacts the global population. It has a high prevalence among females, individuals with increasing age, and those in low to middle-income countries [1]. Detected by decreased hemoglobin (Hb) levels, anemia reduces the blood's oxygen-carrying capacity, causing symptoms like pale skin and shortness of breath. Often untreated and unrecognized in resource-limited settings, effective diagnostic methods are crucial for improving detection. According to the World health Organization (WHO), anemia impacts over 25% of the global population [2-3]. It is defined as hemoglobin levels below 13.0g/dL for males, 12.0 g/dL for non-pregnant females, and 11.0 g/dL for pregnant females [4]. The current reliable standard for diagnosing anemia involves invasive blood testing, which measures hemoglobin concentrations and red blood cell indices [5]. However, they all rely on access to expensive laboratory equipment, specialized personnel, and resources that are not universally available. These barriers cause a significant challenge in some low income communities.

The continuously increasing trend of anemia prevalence has led to an increasing focus on alternative and non-invasive diagnostic strategies that are more accessible [6]. The visual assessment of physiological signs, such as pallor in the conjunctiva, skin, or nail beds, is a traditional clinical

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.LD26658

technique widely used by clinicians as a screening tool to detect anemia. Although this technique is easy to conduct and cost-effective, it is highly subjective and relies heavily on the clinician's personal experience and intuition. This makes it hard to standardize and can lead to inconsistent results. However, advances in digital imaging and computational analysis provide a potentially promising alternative method [7]. These modern technologies allow clinicians and researchers to quantify visual features, for example, pixel intensity and color distribution, to support more objective and reliable anemia screening.

This review examines both traditional diagnostic methods and image-based methods for anemia detection, emphasizing the statistical methods underlying their development and evaluation. It also highlights the role of quantitative image analysis, feature extraction, and predictive modeling in advancing non-invasive detection. By identifying the strengths and limitations of current image-based statistical methods, this article aims to inform future research on developing more accurate and standardized diagnostic tools.

2. Overview of anemia

Various factors can lead to anemia, such as iron deficiency, chronic blood loss, vitamin deficiency, long term illness such as cancer, or genetic disorders such as sickle cell disease [8]. Different Type of anemia has their unique pathophysiological mechanisms, which cause different symptoms and make detection more difficult. Anemia could be classified based on the red blood cell morphology, such as microcytic, normocytic, macrocytic, or by etiology, such as iron deficiency anemia, hemolytic anemia, etc.

Anemia has symptoms that are highly similar to other illnesses and may be mistakenly diagnosed as other conditions. For example, dizziness, headache, shortness of breath, and paleness of skin [9]. Anemia impacts vulnerable populations differently. Certain populations, such as reproductive women, infants, children, and the elderly, are more likely to develop anemia due to their physiological, nutritional, or socioeconomic demands [10]. Anemia causes significant consequences for different individuals. In females of reproductive age, it increases the risk of preterm birth, maternal mortality, and low birth weight [8]. These clinical measurements are well-established outcomes that are used to define anemia and serve as reliable targets for statistical models. A promising and accurate diagnosis of anemia is crucial for clinical practice. The traditional diagnostic approaches to confirm the presence of anemia are based on invasive blood tests and standard laboratory diagnostics. It is completed by measuring hemoglobin (Hb), red blood cell count, mean corpuscular volume (MCV), blood film analysis, or Hb electrophoresis [11]. The most common assessment method is to measure the Hb concentration, because it is inexpensive and widely available compared to other methods [12]. While reliable and widely used, traditional diagnostic methods for anemia require laboratory equipment and trained personnel for invasive procedures, limiting accessibility in low-resource settings and potentially causing underdiagnosis. Invasive procedures can also induce anxiety and pose infection risks, especially for vulnerable populations. This has spurred interest in developing non-invasive diagnostic tools. From a statistical standpoint, traditional approaches rely not only on fixed clinical thresholds but also benefit from incorporating various predictors through statistical modeling. These models can enhance predictive accuracy, quantify uncertainty associated with different predictors, and individualize anemia risk assessment.

DOI: 10.54254/2753-8818/2025.LD26658

3. Methods

3.1. Emerging image-based diagnostics

Non-invasive approaches focus on providing more patient-friendly and cost-effective diagnostic methods for screening vulnerable populations. The methods eliminate the need for well-trained blood draws and expensive laboratory equipment, which makes diagnosis more accessible and affordable. It requires external lighting and various techniques, including imaging-based techniques (Digital Image Processing), mobile-based applications, video analysis, and spectral imaging. For example, digital photos of the conjunctiva, palm, or nail beds are often analyzed using RGB pixel values [13]. Those techniques allow for potential remote and scalable diagnosis, which is suitable for low-resource areas.

Further, statistical techniques were essential for image calibration, feature extraction, and diagnostic classification. In many recent correlated studies, statistical methods such as regression models, color classification algorithms, and supervised classifiers were applied to reduce error and improve prediction accuracy and transform the visual information into diagnostic insight. However, visual detection methods still face challenges in standardization, acquisition conditions, and ensuring device calibration.

3.2. Image acquisition and analysis

Hemoglobin has unique absorption properties that absorb light in the green spectrum with wavelengths between 490 nm and 570 nm and reflect light in the red spectrum with wavelengths between 630 nm and 760 nm (red) [6]. The unique characteristics of the hemoglobin absorption of light wavelengths allow us to observe physiological changes in RGB (red, green, blue) pixel intensities in response to variation in hemoglobin concentration.

In many studies, images were captured using standard smartphone cameras targeting regions of interest (ROI) such as the conjunctiva, nail beds, or skin. These regions have minimal pigmentation, which provides better contrast for color-based analysis and makes detection of pallor easier. Among the ROIs, the conjunctiva is often selected as the most preferred site because it has lower variability in light reflection and highly consistent vascularization. It improves the reliability of pixel-based measurements.

From a biostatistical perspective, capturing the image is the first step of the quantitative transformation of visual signals into digital pixel data that determines the anemia status. The digital images can be viewed as a collection of pixel values, which can then be turned into red, green, and blue intensity components. These pixel values are the raw variables, which allow researchers to standardize the visual assessments and reduce observer bias. The structured numerical values form variables that can be aggregated or transformed to construct structured predictors. These predictors are used to quantify the associations with anemia status.

3.3. Feature extraction and preprocessing

After digital image acquisition, feature extraction is crucial for transforming captured pixel-based indicators into interpretable predictors in statistical models. Feature extraction converts unstructured raw pixel-level data into quantifiable variables, capturing patterns associated with anemia.

In many studies, descriptive statistics such as the mean, standard deviation, or median are used to summarize the color data distribution of RGB pixel intensities within the ROI [6, 13]. Some

DOI: 10.54254/2753-8818/2025.LD26658

researchers also estimate ratios between multiple color channels, such as the ratio of red to green, which captures subtle differences in pallor related to hemoglobin levels, to enhance feature discrimination. To reduce the influence of lighting variability, CIELAB, transformations into alternative color spaces are often applied. These techniques are intended to approximate human vision, help to create more robust and meaningful features that improve the performance of statistical models.

3.4. Statistical modeling and prediction

Statistical modeling is the major tool for connecting visual indicators to anemia status after feature extraction from digital images. These models are used to predict continuous hemoglobin concentrations and classify anemic status. Various statistical and machine learning models have been used for development, including Logistic Regression (LR), Support Vector Machines (SVM), k-Nearest Neighbor (kNN), and Decision Trees (DT). Logistic regression is commonly used for binary classification, but it tends to have higher precision with relatively small datasets. SVM and DT have been used to capture nonlinear decision thresholds. However, they can be sensitive to parameter settings.

The k-Nearest Neighbor (kNN) algorithm has the most precise and interpretable classifiers among all. As reported in recent comparative analyses, kNN models outperform other classifiers in terms of precision and accuracy [13]. It is a simple and powerful non-parametric approach that classifies samples based on similarity to neighboring data points [14]. In the context of image-based anemia classification, kNN stands out because it operates on small-sized datasets effectively with the least bias. Its interpretability makes it suitable for clinical applications.

In the study by Dimauro, the researchers use the kNN algorithm as the primary statistical classifier to predict anemia based on the features extracted from smartphone images of the conjunctiva [6]. They used the CIELAB color space to compute Lightness (L), Green to Red Ratio (a*), and Blue to Yellow Ratio (b*) values, which are used as input variables. The algorithm classified new samples based on the majority class among their three closest training samples, specifically setting k to 3. The value of parameter k significantly influences the performance of the kNN algorithm. Smaller k values are more sensitive to noise, while larger k values have more stable classification at the cost of reduced sensitivity. The choice of $k = 3$ reflects a balance between sensitivity to local variation and resistance to noise.

4. Discussion

4.1. Limitations of statistical modeling and existing challenges

Statistical modeling plays a central role in transforming visual indicators into predictive values. Models such as LR, SVM, DT, and kNN are used to estimate hemoglobin levels. Among the techniques explored, the k-Nearest Neighbors (kNN) algorithm has the highest precision and is most interpretable.

However, the performance of these models is limited by several factors. Traditional models, such as logistic regression, may oversimplify non-linear feature interactions. SVM and DT are sensitive to parameter settings and might overfit when training data is limited. kNN requires careful selection of the parameter k. Inconsistent image acquisition, variations in lighting, and unstandardized preprocessing can introduce bias and undermine model generalizability. Statistical predictions may lack reliability across the population without calibration.

DOI: 10.54254/2753-8818/2025.LD26658

4.2. Future direction

Several important areas require further development to enhance clinical applicability. Most current models are developed based on iron-deficiency anemia, which often present with visible pallor. However, other forms such as anemia of chronic disease may not show velar visual features. Future research should explore how statistical models perform across different anemia and adjust feature extraction accordingly. Future exploration should focus on integrating image-based anemia detection tools into mobile health platforms, so it can be widely used for low-resource settings. Lastly, variability in lighting conditions and image quality can introduce bias and reduce model accuracy. To ensure consistency across studies and applications, establishing calibration tools is essential.

5. Conclusion

There are various types of anemia resulting from different underlying causes. The visual feature assessment techniques show strong promise for examining paleness in the conjunctiva to identify anemia caused by iron deficiency. However, there are also possibilities that they might be unreliable for detecting other forms of anemia, such as hemolytic anemia or anemia of chronic disease may not present with visible pallor. This limits the effectiveness of image-based diagnostics in these cases. The integration of statistical modeling with visual data has significantly enhanced the potential for non-invasive screening. Methods such as logistic regression, support vector machines, and especially the kNN algorithm have demonstrated encouraging accuracy and sensitivity in classifying anemia status based on visual image features. Accounting for diverse anemia etiologies and validating them across broader populations will be essential. Standardization of image acquisition protocols improves model generalizability. Continued development of a robust statistical framework will help establish visual diagnostics as a reliable clinical tool. Statistical models do not merely assist in prediction; they serve as the foundation for translating raw image data into meaningful clinical insights.

However, this review is limited in scope and depth due to its focus primarily on supervised learning techniques applied to image-based anemia detection. It does not cover all possible statistical methods, including deep learning models. The comparative analysis is based on selected literature and may not fully represent the diversity of datasets. The absence of shared datasets complicates direct performance comparisons across studies.

Future research should explore the integration of deep learning techniques with traditional statistical approaches to enhance feature extraction and classification performance. In addition, establishing, large, publicly available image datasets with standardized acquisition tools will be essential for reproducible research.

References

[1] Gandhi, S. J., Hagans, I., Nathan, K., Hunter, K., & Roy, S. (2017). Prevalence, comorbidity and investigation of anemia in the primary care office. Journal of Clinical Medicine Research, 9, 970–980. https://doi.org/10.14740/jocmr3221w.

[2] Dimauro, G., Camporeale, M. G., Dipalma, A., Guarini, A., & Maglietta, R. (2023). Anaemia detection based on sclera and blood vessel colour estimation. Biomedical Signal Processing and Control, 81, 104489. https://doi.org/10.10/j.bspc.2022.104489.

[3] Owais, A., Merritt, C., Lee, C., & Bhutta, Z. A. (2021). Anemia among women of reproductive age: An overview of global burden, trends, determinants, and drivers of progress in low- and middle-income countries. Nutrients, 13(8), 2745. doi: 10.3390/nu13082745.

DOI: 10.54254/2753-8818/2025.LD26658

[4] World Health Organization. (2011). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. World Health Organization. https: //iris.who.int/handle/10665/85839
[5] Kumar, R. D., Guruprasad, S., Kansara, K., Rao, K. N. R., Mohan, M. R., & Reddy, M. R., et al. (2021). A novel noninvasive hemoglobin sensing device for anemia screening. IEEE Sensors Journal, 21(13), 15318-15329. 10.1109/JSEN.2021.3070971
[6] Dimauro, G., Caivano, D., & Girardi, F. (2018). A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva. IEEE Access, 6, 46968–46975.
[7] Appiahene, P., Arthur, E. J., Korankye, S., Afrifa, S., Asare, J. W., & Donkoh, E. T. (2023). Detection of anemia using conjunctival images: A smartphone application approach. Medicine in Novel Technology and Devices, 100237.
[8] Chaparro, C. M., & Suchdev, P. S. (2019). Anemia epidemiology, pathophysiology, and etiology in low-and middle-income countries. Annals of the New York Academy of Sciences, 1450(1), 15-31.
[9] Dejene, B. E., Abuhay, T. M., & Bogale, D. S. (2022). Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. BMC Medical Informatics and Decision Making, 22(1), 247.
[10] Nosratnejad, S., Barfar, E., Hosseini, H., Barooti, E., & Rashidian, A. (2014). Cost-effectiveness of anemia screening in vulnerable groups: A systematic review. International Journal of Preventive Medicine, 5(7), 813-819. PMID: 25104991; PMCID: PMC4124557.
[11] Balarajan, Y., Ramakrishnan, U., Özaltin, E., et al. (2011). Anaemia in low-income and middle-income countries. The Lancet, 378, 2123–2135.
[12] Centers for Disease Control and Prevention (CDC). (1989). CDC criteria for anemia in children and childbearing-aged women. MMWR Morbidity and Mortality Weekly Report, 38(22), 400-404. PMID: 2542755.
[13] Dimauro, G., De Ruvo, S., Di Terlizzi, F., Ruggieri, A., Volpe, V., Colizzi, L., & Girardi, F. (2020). Estimate of Anemia with New Non-Invasive Systems—A Moment of Reflection. Electronics, 9(5), 780. https: //doi.org/10.3390/electronics9050780
[14] Wang, H., Düntsch, I., Gediga, G., & Guo, G. (2005). Nearest neighbours without k. In Monitoring, Security, and Rescue Techniques in Multiagent Systems (Advances in Soft Computing, vol 28, pp. 179–189). Springer, Berlin, Heidelberg. https: //doi.org/10.1007/3-540-32370-8_12