Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 5
May 2025

Automatic Anemia Detection for Early Identification and Categorization using Machine Learning

Suhasini S Goilkar [1], Shashikant S Goilkar [2]

[1] EXTC Department, Finolex Academy of Management and Technology, Ratnagiri, Maharashtra, India.
Email: spkashid@gmail.com

[2] Mechanical Department. Finolex Academy of Management and Technology, Ratnagiri, Maharashtra, India.
Email: ssgoilkar@gmail.com

Abstract

Anemia, characterized by a decrease in red blood cells or hemoglobin levels, is a prevalent global health issue affecting millions of individuals. Prompt and accurate diagnosis of Anemia is crucial for effective treatment and prevention of allied complications. This research work involves developing a digital image processing technique to analyse digital images of blood smears provided by patients. The algorithm will identify and measure red blood cell morphological abnormalities such as variations in size, shape, and color indicates different types of Anemia. To use machine learning methods for feature extraction and classification, the system will be trained using a vast collection of annotated blood smear images. Additionally, clinical data including patient demographics, medical histories, and laboratory test results will be integrated to enhance the algorithm's diagnostic accuracy and predictive capabilities. The proposed method aims to provide healthcare professionals with a reliable, cost-effective, and non-invasive tool for early detection and classification of Anemia. This initiative has the potential to significantly enhance patient outcomes and alleviate the burden of Anemia related complications on healthcare systems worldwide through the utilization of advancements in digital imaging and machine learning technologies.

Keywords: Blood samples, Red Blood Cell (RBC), Peripheral Blood Smear (PBS), Anemia, Neural Networks.

1. Introduction

Anemia is a prevalent and persistent global health issue that affects individuals of all ages. It is considered as to decrease the quantity or quality of red blood cells. The accurate and timely identification of Anemia is of outmost importance [1]. However, current diagnostic approaches often fall short in terms of accuracy, hindering effective intervention. This research implementation aims to address this demanding need by developing Anemia detection Blood Smear Analysis. This research focuses on identifying subtle morphological cues within red blood cells by carefully examining blood smears. Anemia is a condition characterized by a deficiency of iron in Hemoglobin and red blood cells. It affects individuals across all stages of life, with pregnant women and young children being particularly susceptible. Globally, Anemia impacts approximately 1.62 billion people, accounting for around 24.8% of the population, making it the second leading cause of illness worldwide according to the World Health Organization (WHO). To meet the urgent need for Anemia identification, the analysis of peripheral blood smear (PBS) images has long been utilized as a crucial diagnostic tool. However, manual analysis is time-consuming, prone to errors, and requires specialized expertise. The automation of PBS analysis has led to a successful field of research, with many teams exploring cutting-edge methods that leverage image processing tools [3, 4]. This study begins with a comprehensive review of techniques used to analyse RBC features in PBS images, with a primary focus on their application in Anemia identification. Certain risk factors associated with blood cancers include chronic infections, gender, age, nutritional deficiencies, and family history. Chronic infections, such as HIV/AIDS, can lead to Anemia by affecting red-blood cell production. Infants, teenagers, and age-appropriate women are more vulnerable due to their growth, development, and monthly blood loss. Nutritional deficiencies, especially in iron, can be a risk factor for vegetarians or those with poor eating habits. Additionally, genetic factors in family history may play a role in conditions like thalassemia or sickle cell Anemia [5, 6].

2. Literature Survey

2.1 Methodology Behind Image Processing

There have been several successful advancements in image processing within the healthcare industry and other sectors. Image processing involves performing standard procedures on an image to generate a higher quality picture. It is a method used to convert an image into a digital format. There are three main steps involved: utilizing optical image analysis, using a scanner to import images, and implementing management techniques such as compression, image enhancement, and pattern recognition in images like satellite photos. Image processing is sometimes mistaken for image editing to achieve a desired level of attractiveness or to support a particular claim [9,10]. For instance, in diagnosing Anemia, image processing is used to identify abnormalities in red blood cell images. Common segmentation methods include clustering and marker-controlled watershed approaches. Machine learning, particularly Support Vector Machines (SVM), is employed to classify different types of Anemia and their subtypes. Automated image segmentation precision is enhanced by techniques such as Otsu's method. This comprehensive approach improves the efficiency and accuracy of Anemia diagnosis, allowing for more personalized patient care. Various classification methods were compared for identifying sickle cell Anemia, including KNN, SVM, and ELM. Researchers utilized geometric features, Sobel edge detection, random walk, Fuzzy C-means clustering, and morphological filters to identify red blood cells with sickle cell formation. Reported accuracy rates ranged from 85% to 95%. Sickle cell anemia affects red blood cells, which play a crucial role in oxygen transport throughout the body. Early detection of this condition is essential to prolong life expectancy. To achieve this, an image processing method is being developed to detect sickle cells in blood samples, involving noise reduction, grayscale conversion, and noise filtering. The main objective of the study is to accurately segment high-resolution whole-slide blood cells based on their semantic content using image processing techniques. The primary focus is on creating a reliable method for precise identification and segmentation of various cell types or structures in these large images. The project will address the challenges associated with handling extensive and complex material, emphasizing the importance of feature extraction in the image processing pipeline to capture essential components such as cell shapes, colors, and textures. In addition to its technical aspects, the project underscores the significance of image processing in biological applications, particularly in the precise semantic segmentation of blood cell images to support medical research and diagnostics. The project aims to use image processing algorithms to identify anemia from microscopy blood smear pictures. The goal is to develop an accurate image processing system that can detect anemia by analysing detailed information about blood cell morphology. This involves enhancing and preprocessing images, extracting features like cell size and red blood cell count, and utilizing machine learning techniques for diagnosis and classification. The project also focuses on recognizing and differentiating red and white blood cells in microscopic images for research and medical diagnostics. Image processing techniques are used to improve image quality, reduce noise, and prepare for cell detection. Segmentation algorithms isolate individual cells, while feature extraction techniques retrieve attributes like size, shape, and colour [11,12]. The study primarily focuses on identifying abnormal red blood cells and diagnosing specific forms of anemia using image processing methods and machine learning, particularly Support Vector Machines (SVM). The project aims to develop a system that can automatically detect anemia by distinguishing between normal and abnormal red blood cells in microscopic blood samples. This involves capturing and analysing images of blood cells, enhancing and pre-processing them using image processing techniques. The research focuses on white blood cell counting in blood smear images using different colour segmentation methods. Scholars from University Tun Hussein Onn Malaysia are exploring various approaches to accurately identify and count white blood cells in digital blood smear images. The objective is to improve the accuracy and efficiency of white blood cell counting, which is important for diagnosing medical conditions, by experimenting with different colour segmentation techniques. The findings could potentially lead to the development of automated and more reliable systems for analysing blood smear images, benefiting the field of medicine [13,14].

2.2 Methodology behind Machine Learning Technology

The research [14] uses machine learning with the YOLO algorithm to count and identify blood cells. It achieves

96.09% accuracy for RBCs and 86.89% accuracy for WBCs using 364 annotated images from the BCCD dataset. The YOLO model is adjusted with three outputs and threshold levels determined by average absolute error. However, this method may result in multiple platelets counts, so the researchers also utilize KNN and IOU techniques. The investigation [15] focuses on machine learning for predicting anemia. It covers important aspects such as anemia prediction, dataset selection, algorithm choice, feature engineering, training with historical data, model effectiveness and accuracy assessment, and potential clinical applications in mammography and early anemia detection. This research highlights the importance of technology in healthcare and shows how machine learning can be used to predict and detect disorders like anemia early on. The study examines the use of machine learning techniques in categorizing blood cell images for computer-aided diagnostics. Various methods and algorithms, such as decision trees and neural networks, are evaluated and summarized. The study aims to detect blood disorders using traditional machine learning methods and create a system capable of identifying conditions like anemia and leukaemia. Feature engineering is used to extract valuable information from the datasets. Machine learning models are crucial in diagnosing and categorizing diseases and identifying blood abnormalities. The research focuses on developing a dual-objective machine learning model. The objective is to create a framework for calculating blood hemoglobin levels, which is important for diagnosing anemia and other blood disorders. It also aims to categorize anemia based on hemoglobin levels and potentially other clinical data. Machine learning methods will be used to predict hemoglobin levels and classify individuals into anemia categories. The research [11,16] analyzes how traditional machine learning techniques are used to classify white blood cells in blood smear images. It explores the effectiveness of models like support vector machines, decision trees, and k-nearest neighbours for WBC classification. These techniques are commonly used in medical diagnostics for feature extraction, dataset selection, and image preprocessing. The performance of these models will be assessed and compared using metrics like accuracy and sensitivity.

2.3 Methodology behind Deep Learning Technology

Deep learning has significantly improved vision tasks like object tracking, image categorization, and pattern recognition by using convolutional neural networks (CNNs). CNN-based methods have advanced the diagnosis of anemia by automating the identification of blood cell types in a sample. By training a deep learning model on annotated blood sample images, semantic segmentation allows for accurate categorization of platelets, RBCs, and WBCs. This automated process enhances the accuracy and efficiency of blood cell counting, potentially advancing medical research. Through deep learning methodologies, precise models are developed for detecting different blood cell types, utilizing sophisticated architectures like CNNs. These algorithms have been trained on labeled blood cell images to understand their intricate characteristics. In their paper, Alzubaidi et al. developed deep learning models with 99.54% accuracy for diagnosing sickle cell anemia [17]. They used data from erythrocytes IDB, ALL-IDB, and other sources to create three CNN models, which were then used to train a multi-class SVM with an accuracy range of 98-99%. Sickle cell disease can lead to severe arterial blockages with life-threatening consequences. The study introduces an innovative deep learning technique for automatically detecting and counting cells in microscopic blood images. The authors detail the development and training of their deep learning model, highlighting its superior performance in accuracy and efficiency compared to other methods. The empirical findings showcase the success of this approach and its potential to transform medical image processing. This research emphasizes the benefits of deep learning in streamlining labor-intensive processes in the medical field. Deep learning is being used in a study to automate the identification and separation of red and white blood cells in small blood images. Neural network topologies analyze complex image patterns, with a region proposal technique applying deep learning algorithms to identify regions of interest likely to contain blood cells. This automated method is crucial for accurate cell quantification in medical diagnosis and research. Another study focuses on using deep learning to automatically identify and detect sickle cells in blood samples, emphasizing the model's accuracy in recognizing various blood cells. Deep learning's ability to recognize intricate patterns enables precise sickle cell identification, with techniques explored for training the system to distinguish between healthy and sickle cells in microscopic organism images. The study introduces a new method that combines Deep Neural Networks (DNNs) and Locality Sensitive Hashing (LSH) to create artificial blood smears. LSH is used to identify similar patches from real blood smear images, and DNNs are used to enhance these artificial images and replicate the characteristics of real blood smears. This approach addresses the lack of labelled data in medical image analysis, particularly in haematology. Experimental findings validate the effectiveness of this method in producing realistic artificial blood smears. These synthetic images can be used to train machine learning algorithms for various haematological tasks. The technique has potential applications in medical research, education, and the development of diagnostic tools. It also improves existing datasets and facilitates the design and evaluation of blood smear analysis algorithms. Overall, this methodology is a promising contribution to medical image synthesis and analysis, utilizing advancements in hashing techniques and deep learning architectures [18,19].

3. Adopted Methodology

Blood smear analysis is crucial for identifying anemia, a condition marked by a deficiency in red blood cells. The process involves collecting blood samples and examining them under a digital microscope. Preprocessing ensures the samples are ready for analysis. Classification methods, including image processing, machine learning, and deep learning algorithms are then used to classify different types and levels of anemia. Deep learning algorithms improve the accuracy of classification by identifying complex patterns in the data. This process flow is shown in fig.1 as block diagram of adopted methodology.

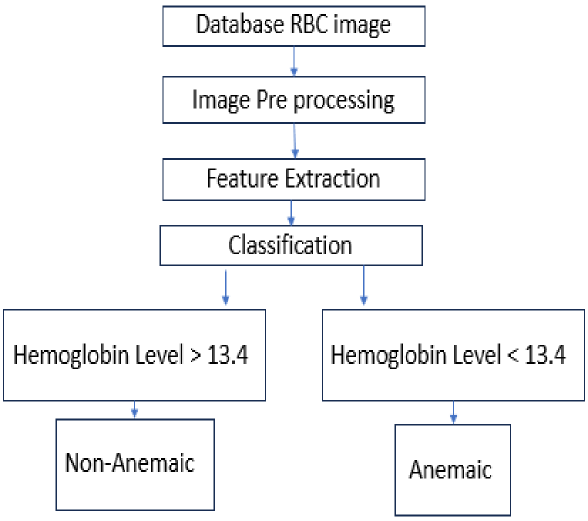


Figure 1. Block Diagram of Adopted Methodology

The process of identifying anemia through blood smear analysis involves several steps. Initially the blood samples undergo preprocessing to prepare them for further analysis. Then, feature extraction is performed based on the characteristics of red blood cells (RBCs), including their size, color, and shape. This information provides valuable insights into the morphology and properties of RBCs, which can help identify different types of anemia.

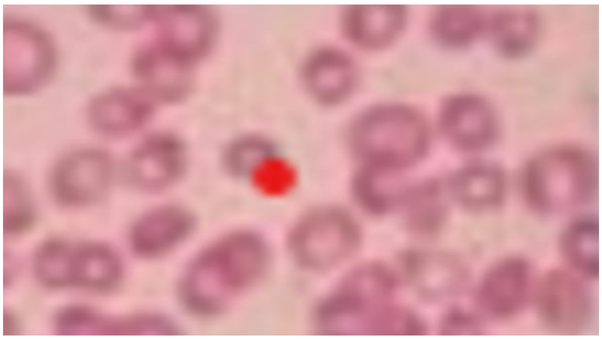
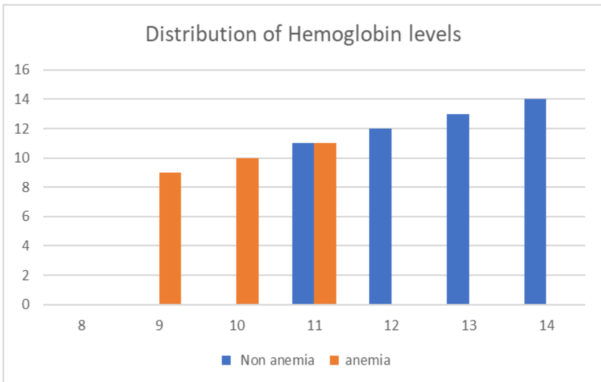


Figure 2. Sample RBC Database image

In the feature extraction process the dataset used for analysis consists of 1500 samples, each with six attributes: gender, hemoglobin levels, mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and result. These attributes collectively offer valuable information about various haematological parameters. By conducting a comprehensive analysis of the dataset, researchers can identify patterns and correlations, which can aid in the diagnosis and understanding of haematological conditions across different genders and hemoglobin levels. This knowledge is crucial for timely intervention and effective management of anemia, a prevalent blood disorder. Figure 2 illustrates the distribution of hemoglobin levels based on the results of anemia diagnosis.



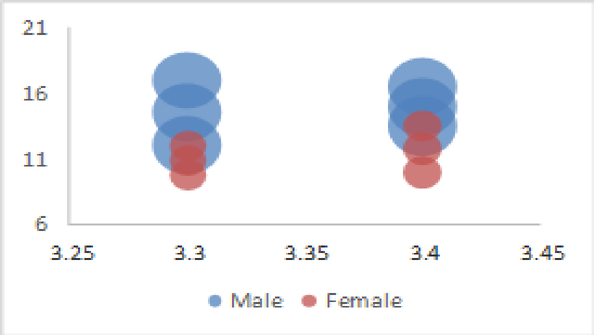


Figure 3. Distribution of hemoglobin levels

4. Results and Discussion

A machine learning algorithm has been applied to analyzes data on hemoglobin and mean corpuscular volume (MCV), providing valuable insights into the detection of anemia and its subtypes in both males and females. The specified ranges of hemoglobin levels for males (12–15 gm/dL) and females (11.5–14.5 gm/dL) played a crucial role in diagnosing anemia. It is important to note that individuals with hemoglobin levels above or within the designated range were categorized as non-anemic, while those with levels below the threshold were classified as anemic. The inclusion of MCV measurements (80.0-95.0fL for males and 82.0-98.0fL for females) further enhanced the classification of anemia. By examining the MCV values, it became possible to differentiate between microlytic, normocytic. and macrolytic anemia. MCV values below the established range indicated microlytic anemia, whereas values exceeding the range indicated macrolytic anemia.

For performance analysis precision, recall and F1score is used. In Table 1. performance analysis of proposed method is compared with previous five implemented techniques.

Table 1. Distribution of hemoglobin levels

Reference	Precision	Recall	F1-Score
Lee et.al	70.10	86.7	77.51
Hortinela et.al	83.3	87.5	85.34
Jaiswal et.al	86.3	91.3	85.94
Chy et.al	90	94.3	92.09
Alsheref et.al	94.61	97.59	96.07
Proposed Method	96.09	98.01	98.00

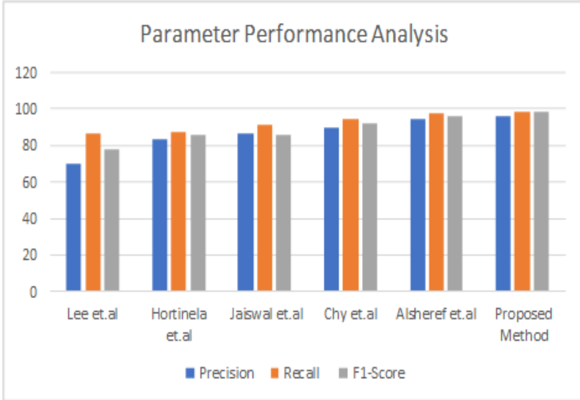


Figure 4. Parameter performance analysis

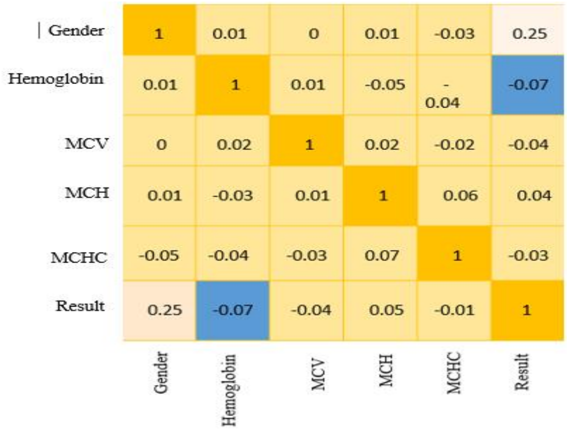


Figure 5. Parameter Analysis in Matrix form

5. Conclusion and Future Scope

A proposed solution to the global health challenge of anemia involves the use of automated image processing. This initiative aims to provide healthcare professionals with a reliable, cost-effective, and non-invasive tool for early detection and classification of anemia through the identification of red blood cell abnormalities using digital imaging and machine learning advancements. The algorithm's accuracy is further enhanced by incorporating clinical data. Successful implementation of this approach could significantly improve patient outcomes and alleviate the burden of anemia-related issues on healthcare systems worldwide. This implementation work highlights the potential of interdisciplinary collaborations in data science, technology, and medicine to tackle complex health problems and improve public health outcomes.

References

[1] Lee, Shin-Jye, Pei-Yun Chen, and Jeng-Wei Lin. "Complete blood cell detection and counting based on deep neural networks." Applied Sciences 12, no. 16 (2022): 8140-46.

[2] Al-Dulaimi, and Teba Makki. "Blood Cell Microscopic Image classification in Computer Aided Diagnosis Using Machine Learning: " Iraqi Journal for Computer Science and Mathematics 4, no. 2 (2023): 43-55.

[3] Nithya, R., and K. Nirmala. "Detection of Anemia using Image Processing Techniques from microscopy blood smear images." In Journal of Physics: Conference Series, vol. 2318, no.1,p. 012043. IOP Publishing, 2022.

[4] Alzubaidi, Laith, Mohammed A. Fadhel, Omran AlShamma, Jinglan Zhang, and Ye Duan. "Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis." Electronics 9,no. 3 (2020): 427.

[5] M. M. Alam and M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells", Healthcare technology letters, vol. 6, no. 4, pp. 103-108, 2019.

[6] Chy, Tajkia Saima, and Mohammad Anisur Rahaman. "A comparative analysis by KNN, SVM & ELM classification to detect sickle cell anemia." In 2019 International conference on robotics, electrical and signal processing techniques (ICREST), pp. 455-459. IEEE, 2019.

[7] Shahzad, Muhammad, Arif Iqbal Umar, Muazzam A. Khan, Syed Hamad Shirazi, Zakir Khan, and Waqas Yousaf. "Robust method for semantic segmentation of whole-slide blood cell microscopic images." Computational and mathematical methods in medicine 2020.

[8] Di Ruberto, Cecilia, Andrea Loddo, and Lorenzo Putzu. "A region proposal approach for cells detection and counting from microscopic blood images." In Image Analysis and Processing–ICIAP 2019: 20th International Conference, Trento, Italy, September 9–13, 2019.

[9] Di Ruberto, Cecilia, Andrea Loddo, and Lorenzo Putzu. "Detection of red and white blood cells from microscopic blood images using a region proposal approach." Computers in biology and medicine 116 (2020): 103530.

[10] Alsheref, Fahad Kamal, and Wael Hassan Gomaa. "Blood diseases detection using lassical machine learning algorithms." International Journal of Advanced Computer Science and Applications 10, no. 7 (2019).

[11] De Haan, Kevin, Hatice Ceylan Koydemir, Yair Rivenson, Derek Tseng, Elizabeth Van Dyne, Lissette Bakic, Doruk Karinca et al. "Automated screening of sickle cells using a smartphone-based microscope and deep learning." NPJ digital medicine 3, no. 1 (2020):76.

[12] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, doi:10.1109/INCET 49848.2020. 9154130.. India,2020, pp.1-7,

[13] Al-Qudah, Rabiah, and Ching Y. Suen. "Synthetic blood smears generation using locality sensitive hashing and deep neural networks." IEEE Access 8 (2020): 102530-102539.

[14] Safuan, Syadia Nabilah Mohd, Mohd Razali Md Tomari, and Wan Nurshazwani Wan Zakaria. "White blood cell (WBC) counting analysis in blood smear images using various color segmentation methods." Measurement 116 (2018): 543-555.

[15] Alam, Mohammad Mahmudul, and Mohammad TariqulIslam. "Machine learning approach of automatic identification and counting of blood cells." Healthcare technology letters 6, no. 4 (2019): 103-108

[16] Jaiswal, Manish, Anima Srivastava, and Tanveer J. Siddiqui. "Machine learning algorithms for anemia disease prediction." In Recent Trends in Communication, Computing, and Electronics: Select Proceedings of IC3E 2018, pp. 463-469. Springer Singapore, 2019.

[17] Rahadi, Irwan, Meechoke Choodoung, and Arunsri Choodoung. "Red blood cells and white blood cells detection by image processing." In Journal of Physics: Conference Series, vol. 1539, no. 1, p. 012025.

[18] A. Khouani, M. El Habib Daho, S. A. Mahmoudi, M. A. Chikh and B. Benzineb, "Automated recognition of white blood cells using deep learning", Biomedical Engineering Letters, vol. 10, no. 3, pp. 359-367, 2020.

[19] Hortinela, Carlos C., Jessie R. Balbin, Janette C. Fausto, Paul Daniel C.Divina, and John Philip T. Felices. "Identification of abnormal red blood cells and diagnosing specific types of anemia using image processing and support vector machine." In 2019 (HNICEM), pp. 1-6. IEEE, 2019

[20] L. U. Novia, J. R. T. Alipo-on, F. I. F. Escobar, M. J. "White Blood Cell Classification of Porcine Blood Smear Images", In IAPR Workshop on Artificial Neural Networks in Pattern Recognition, pp. 156-168, 2023.