Volume 11, Issue 6, June – 2026 International Journal of Innovative Science and Research Technology
ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

A Comprehensive Review of Hematological Disorders

Md. Abu Horaira Sarder[1]

[1]Daffodil International University, Birulia, Bangladesh

Publication Date: 2026/07/07

Abstract: Hematopoiesis is the maturation of hematopoietic stem cells into blood cells. This process takes place in different places during the early embryo. During the third week of human embryogenesis, a group of mesenchymal cells in the yolk sac forms clusters called blood islands. These islands have central cells that differentiate to form primitive blood-forming cells, and peripheral cells that form a primitive vascular system. Some of these cells become primitive erythroblasts, the first cells to produce hemoglobin. These cells are not promyelocytes, but they do not develop into erythrocytes as do the adult bone marrow pronormoblasts.

In the third month of embryonic development, stem cells migrate to the liver, which becomes the primary site of blood cell production. Hematopoiesis is also accomplished in the spleen, lymph nodes, and thymus. By the fourth month of gestation, bone marrow hematopoiesis begins, and by birth it is the main source of blood cell production. The process of hematopoiesis in bone marrow is called medullary hematopoiesis, and the process of the creation of blood cells outside the bone marrow is called extramedullary hematopoiesis. Hematopoiesis is present in almost all bones at birth. In adults, it occurs primarily in flat bones (sternum, ribs, skull, and vertebrae) and in the ends of long bones. When demand is high, the marrow might reappear in other sites. Extramedullary hematopoiesis normally diminishes after birth, but can happen in the liver or spleen in certain diseases.

In the bone marrow, nutrient arteries give rise to arterioles and sinusoids, which supply the blood cell-forming cells. The marrow stroma is the site of erythropoiesis, granulopoiesis, and thrombopoiesis. The release of blood cells from marrow into circulation is a crucial step, and occurs when mature blood cells pass through the walls of the sinusoids into the blood. This is essential but not well understood.

How to Cite: Md. Abu Horaira Sarder (2026). A Comprehensive Review of Hematological Disorders. *International Journal of Innovative Science and Research Technology*, 11(6), 2711-2722. https://doi.org/10.38124/ijisrt/26jun1335

I. INTRODUCTION

In the present Harvard Medical School curriculum, hematology is one of a dozen blocks within the large one-year Pathophysiology course, which begins in the middle of the first year. For many years, we have furnished our students in this course with syllabus materials and lecture notes. These notes have long been known as “the camel,” which, as everyone knows, is an animal that looks as though it had been put together by a committee.[22]

When responsibility for the course came to me, the scanty lecture notes took the form of severe and often uninformative outlines. Students complained of the burdens of notetaking, and some brought tape recorders. Accordingly, lecturers were asked to flesh out their outlines, and our camel grew. It became apparent in time that the value of the syllabus would be enhanced by careful editing. The result is the present volume. The vast expansion of knowledge in the various branches of pathophysiology poses an increasingly difficult problem for those who would define the content of a core curriculum and establish standards and priorities for what is to be taught. It is my view that this difficult cause can be furthered only by the thoughtful preparation of texts such as this, for the very act of editing such a volume necessitates choices and permits correlations and overviews that are never quite possible when many and diverse individuals are responsible for a course of instruction. The major goal of this endeavor is to improve the quality and usefulness of these notes as teaching instruments. We intend to revise this small volume frequently so that it retains the freshness and currency that characterized the informal syllabus materials of previous years.

Some believe it would now be appropriate to abandon formal lectures. I see much merit in that proposal. Surely, medical students are capable of handling reading assignments. Surely they deserve exemption from having to read lectures that they could read themselves. Still, there is cause for regret about a step that would deny students an hour or two with colleagues who rate highly as teachers, scientists, and personalities. How to make the best use of these hours will be the subject of experimentation and innovation in the years ahead.[23]

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

II. LITERATURE REVIEW

Martínez-Quintana and Rodríguez-González (2012) investigated the prevalence of anemia in adult congenital heart disease (ACHD) patients and identified hematology parameters for detecting iron deficiency anemia in hypoxemic ACHD patients.[1]

Munir et al. (2013) investigated the impact of ß-thalassemia on hematological and biochemical profiles in female patients. They identified significant deviations in CBC and biochemical parameters from normal values, highlighting a There is a discrepancy between individualised transfusion and chelation strategies.[2]

Guralnik et al. (2005) reported that anemia in older adults is a significant public health problem, impacting almost 3 million people over 65 years of age. The study found a high prevalence of anemia, which was not explained, and pointed to the confusing definition of anemia for various races and age groups. The strength of the research was the large-scale epidemiological data from NHANES III. This study, however, did not provide any longitudinal and interventional evidence for the causes of anaemia and treatment outcomes. The study divided the cases of anemia into nutritional (34%), chronic disease (32%), and the unexplained category (34%), which gave good epidemiological information.[3]

To assess the risk of hepatocellular carcinoma (HCC) in patients with chronic hepatitis C (CHC) infection undergoing hemodialysis, Traditional (1998) performed a prospective multicenter study of the Italian population. Multivariate logistic regression analysis was used to identify cirrhosis, alpha-fetoprotein level>11 ng/mL, and age>45 years as the most important factors associated with the risk of HCC. The results showed the role these clinical factors play in the assessment of cancer risk in high-risk patients. Laosombat et al. (2001) studied the frameshift mutation -41/-42 of beta0-thalassemia in Thai patients to explore the relationship between genotype and phenotype. Disease severity was shown to be genotype dependent in some cases, indicating a role for other genetic modifiers of the genotype in determining the clinical outcome.[4]

Based on the available evidence, Neunert et al. (2019) were able to make evidence-based recommendations on the management of ITP, but a critical gap was noted due to the lack of high-quality data on randomized trials. One of its strengths is the strict GRADE approach as well as the multidisciplinary panel, but one of its most significant weaknesses is the extremely low confidence of evidence that supports most of the recommendations.[5]

The recommendations for the management of iron overload in thalassemia major in Italy were developed by the analytical hierarchy process used by Angelucci et al. (2008).[6]

The diagnosis of microcytic anemia was suggested to be done by a mixed expert system by Birndorf et al. (1996). The study was filling the missing link in automated standard hemogram indices, which had no accurate interpretation. [7]

Chou et al. (2020) hypothesized that alloimmunization among SCD patients during prophylaxis is greatly minimized by alloimmune Rh/K antigen-matched transfusion. Moderate-certainty evidence that was sufficient to make a recommendation, but had limitations due to observational data and the lack of randomized controlled trials (RCTs). The GRADE approach showed a net benefit, namely by reducing the incidence of antibodies due to antigen matching. [8]

Brancaleoni et al. (2016) specified a diagnostic algorithm for thalassemia carriers. This study included both hematological and molecular analysis and mentioned the drawbacks of next-generation sequencing (NGS) in the detection of large deletions. The proposed diagnostic model using the combination of the CBC, HPLC, and specific PCR techniques was found to be very useful for the diagnosis of common genotypes. [9]

From the retrospective laboratory data, Chong et al. (2022) found that the prevalence of Hb E/β-thalassemia was high among the population of Brunei. Local genotyping and a national registry were not available, and this limited complete epidemiological coverage. Screening was done by HPLC and electrophoresis, and Hb E/β-thalassemia was found as the most prevalent severe genotype. [10]

Cohen et al. (2004) have reported the advances of iron chelation therapy and cardiac iron status evaluation with T2MRI in thalassemia. The study concluded that cardiomyopathy can be reversed with successful iron chelation and that there is no correlation between liver iron concentration or ferritin levels and cardiac iron concentration. A detailed analysis of emerging therapies and noninvasive diagnostic methods was performed, though there were limited large randomized trials of combination therapies. The use of T2 MRI was demonstrated to be a reproducible method for measuring the myocardial iron content, and its use was correlated with cardiac dysfunction.[11]

Hews-Girard et al. (2018) assessed the response of patients with nonsevere hemophilia A to desmopressin treatment. A study showed that a higher FVIII level (50 IU/dL) correlated with better outcomes, but some partial and non-responders still had a good outcome. Limitations included the single-center, retrospective design and small sample size. Although there was no perfect biochemical correlation, the classification of peak FVIII test responses was highly clinically effective.

Franchini et al. (2015) did a study of the incidence of cancer among Italian patients with von Willebrand disease (VWD). The results showed that the prevalence of cancer was significantly elevated in patients with type 2 VWD versus the entire VWD population. The study's strengths were that it was a national survey, a large multicenter sample, and that it was retrospective. Logistic regression analysis revealed that there was a significant relationship between type 2 VWD and cancer diagnosis (OR 2.91).

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

Drews and Shulman (2010) updated the reader with a summary of the major advances in hematology-oncology in 2009. They stressed the success of new therapeutic innovations like dabigatran and new treatment strategies. They also pointed out, however, that there were unresolved questions on optimum dosage regimes and long-term treatment success, in which they indicated more research in the field was needed. [14]

Francini et al. (2021) investigated the association of ABO blood group with the occurrence of inhibitor formation in severe HA patients. The study was able to identify that there was a significantly reduced risk of developing inhibitors to FVIII in blood group O. This study was unique in its large multicenter study design. But its drawbacks were being retrospective and that the data on von Willebrand factor (VWF) levels were not complete. Non-O blood groups were an independent risk factor for the development of inhibitors (OR 2.89) in multivariate logistic regression analysis [15].

Pal and Singh (2015) came up with a model to predict the response to anemia treatment from baseline hemoglobin (Hb) and erythropoietin (EPO) levels, which was accurate with 78% accuracy. The model included both the clinical and laboratory parameters; however, they were limited by the small number of patients and by not having an external validation. Logistic regression analysis found predictors that were significant for epoetin response, but there is reason for some uncertainty about the generalizability of the results. [16]

Haus et al. (1983) confirmed that biological rhythms of multifrequencies occur in hematopoietic and immune function which can affect the timing of diagnosis and treatment. One of the study's key findings is that there is a research gap concerning the interaction of circadian, circaseptan, and circannual rhythms in disease conditions. Human and animal chronobiological data were detailed, showing a high degree of inter-individual variability that could be a problem for some clinical applications. Cosinor rhythmometry was used to determine the parameters of the rhythm and develop time-qualified reference ranges (chronodesms) and chronotherapy schedules.

Lam et al. (2021) investigated the impact of iron overload on health-related quality of life (HRQoL) among transfusion-dependent thalassemia patients. The study revealed that the odds of iron overload complications and of decreased HRQoL rose with age. Researchers found that older patients (31–50 years) and social assistance recipients had significantly lower TranQol scores using the validation of the assessment tool. The multicenter design enhanced the study but limited causal inferences, and the cross-sectional design could have under-estimated iron overload in younger patients. [18]

The new clinical guidelines from Rizzo et al. (2010) suggest that erythropoiesis-stimulating agents (ESAs) decrease transfusion use and increase the risk of thromboembolism and mortality. The strengths of the guideline include that it is based on a meta-analysis of individual patient data. However, it has a shortcoming in that it cannot detect low-risk patient subgroups. To maximize the benefit-to-risk ratio, the guideline committee formulated evidence-based recommendations for using restrictive ESAs in patients with hemoglobin < 10 g/dL.[19]

Muncie and Campbell (2009) have presented a detailed review of the pathophysiology, diagnosis, and management of thalassemia. A major asset of the review is its diagnostic framework for primary care practitioners, and a significant weakness is that some recommendations were based on expert consensus. The review recommended that transfusion-dependent patients should have lifelong iron chelation therapy and that bone marrow transplantation is the only cure for iron overload.

Beutler & West (2005) studied the hematologic variations between blacks and white populations and found that about one-third of the hemoglobin difference is due to the presence of α-thalassemia. The advantages of the study are its large population size and the detailed genetic analysis, but there were some unexplained variations. The researchers proved that using uniform reference ranges could misestimate the prevalence of anemia between ethnic groups using PCR genotyping and statistical comparisons.[21]

III. METHODOLOGY

Progress in the study of erythropoiesis has largely been fueled by the elucidation of the development, regulation, and interactions of erythropoietic colony-forming units with other cells. A big step was the identification of the intricate hormonal regulation that was involved. This comprised the purification, cloning, and commercial production of an erythropoietin identical with the human hormone, which had long been desired. As mentioned above and in subsequent lectures, erythropoietin is now a part of the therapy for some anemias.

Hematopoiesis is concerned with the production of erythrocytes (also called red blood cells), and this process is called erythropoiesis. It is basically a system for making and packaging the hemoglobin molecules. The pronormoblast or proerythroblast is the first to mature. It is a stem cell type derived from the more primitive stem cells and the first one committed to erythropoiesis. Thus, it is unipotent. Nevertheless, it is still defined in the kinetic sense as a stem cell in that it can give rise to other cells and maintain its self-renewal.

This classification is useful because it relates to initial laboratory data concerning the average size of red cells and their hemoglobin concentration (chromicity), which suggests the possible disease categories responsible. The classification is based on three parameters, which are the most useful, of red cell indices:

MCV refers to the mean corpuscular volume of red cells, which will be in the normal range of 80–96 cubic micrometers. It is figured by means of the formula:

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

MCV = Hematocrit (%) × 10 / Red Blood Cell Count (millions/mm³)

If a healthy person's hematocrit is 42% and their RBC count is 4.6 million/mm³, their MCV is about 91 cubic micrometers.

The term MCHC is used to indicate mean corpuscular hemoglobin concentration. Normal range is 33-35 g/dL and is worked out as:

MCHC = Hemoglobin (g/dL) × 100 / Hematocrit (%)

The third red cell index, MCH, is the mean hemoglobin (Hb) per red blood cell, and is within the normal range of 27–33 pg/cell. It is determined by:

MCH = (Hb (g/dl) × 10)/RBC (millions/mm³)

MCV is a measure of cell volume, MCHC is a measure of hemoglobin concentration, and MCH is a measure of hemoglobin mass.

Normocytic-Normochromic Anemia

MCV: 80–96 fL

MCHC: 33–35 g/dL

Common cause: Acute bleeding [24].

Anemia is diagnosed, as with all other diseases, through information obtained from a careful medical history, physical examination, and laboratory tests. A thorough history of drug use, drug exposure, and family history is very important. If anemia has been present in past pregnancies, it may indicate inherited diseases or chronic causes, such as heavy periods. Racial and geographical considerations have a role in the identification of hemoglobinopathies and red cell metabolic disorders like G-6-PD deficiency. If anemia is detected by the hematocrit or hemoglobin level, a careful examination of the peripheral blood smear and a calculation of red cell indices (MCV and MCHC) should be performed.

In addition, a reticulocyte count, plasma iron, and iron-binding capacity should be measured. The following tests are useful to assess the functional capacity of the erythron and to give a further clue to the diagnosis: [24]

IV. RESULT

Table 1 As Cells Mature, They Experience a Variety of Functional Changes. The Functional Characteristics Change as Cells Mature

Precursor	Promonocyte	Monocyte	Immature Macrophage	Mature Macrophage
DNA synthesis	+++++	+++	+	+	+/0
Phagocytosis	0	+/0	++	++++	+++++
Lysosomes	0	+	++	++++	++++
IgG receptors	0	+	++	++	+++
Compartment size?	0.4×10^9	$2–7 \times 10^9$	?		

Table 2 In the Human Body, CobII is Found in the Following Forms

Semi Systematic Name	Abbreviation	Systematic Name
Cyanocobalamin*	CN-Cbl	α-(5,6-dimethylbenzimidazolyl)-cyanocobamide
Hydroxocobalamin	OH-Cbl	α-(5,6-dimethylbenzimidazolyl)-hydroxocobamide
Adenosylcobalamin	AdoCbl	α-(5,6-dimethylbenzimidazolyl)-adenosylcobamide
Methylcobalamin	MeCbl	α-(5,6-dimethylbenzimidazolyl)-methylcobamide
Also known as Vitamin B_{12}.		

➢ *Hematologic Effects of Cancer*

- C4b binding protein, 522
- C5, 361
- C8 binding protein, 246
- Calcium

✓ ATPase, 263

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

V. DATASET RESULT

➢ *Dataset Overview*

The hematology dataset consisted of 115,487 patient records and had the following major parameters: Red Blood Cell count (RBC), Hemoglobin (HGB), Hematocrit (HCT), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), Red Cell Distribution Width (RDW), and Platelet count (PLT). Other clinically relevant features were added, including the Mentzer Index, RDW-to-MCV, MCH-to-HGB, PLT-to-RBC, and MCV Squared.

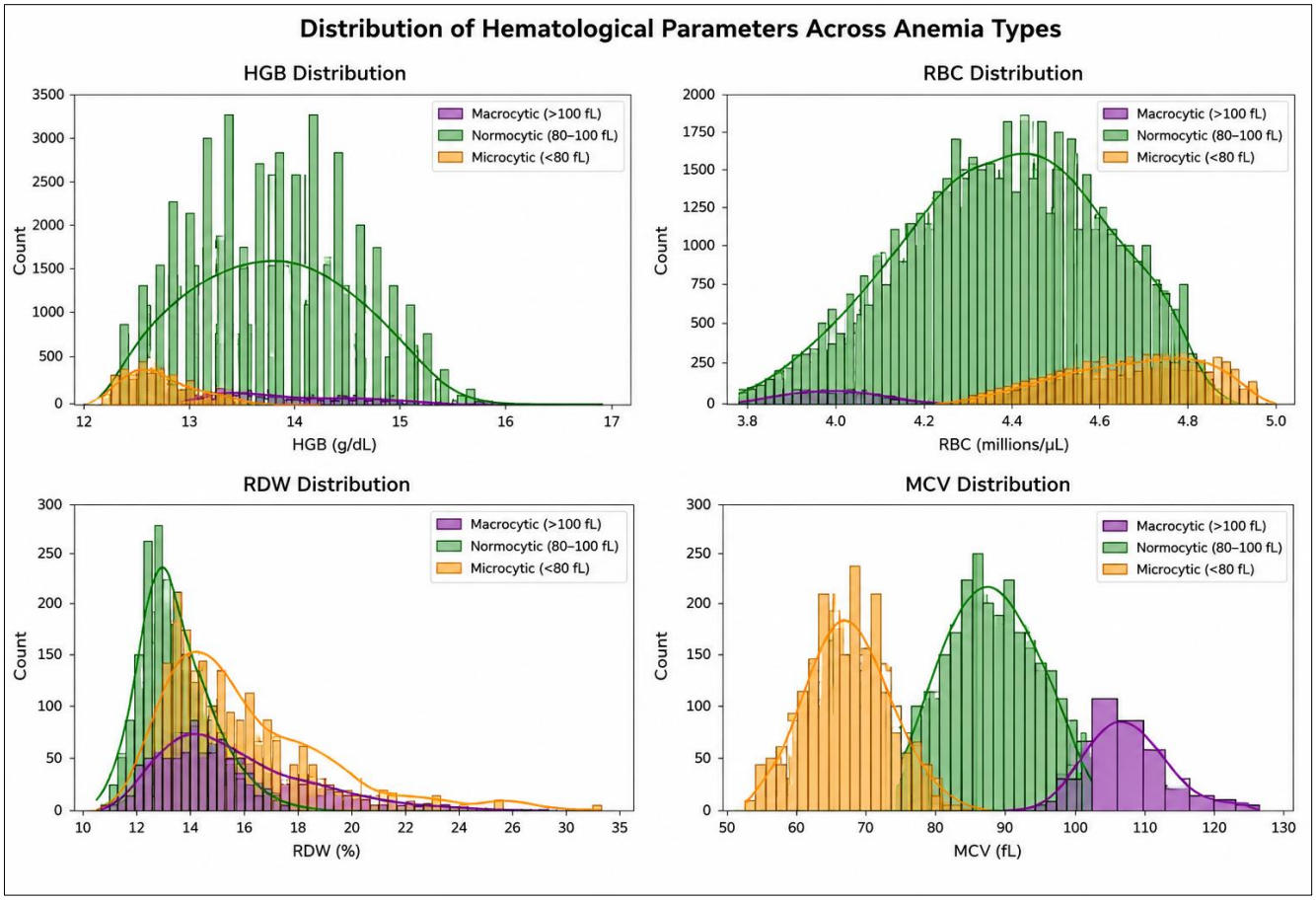


Fig 1 Dataset Overview

Patients were divided into three categories of anemia based on MCV value:

- Microcytic Anemia (MCV < 80 fL)
- Normocytic Anemia (80 ≤ MCV ≤ 100 fL)
- Macrocytic Anemia (MCV>100 fL)

The Hematological pattern of each anemia category was distinct and identified using correlation analysis.

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

➢ *Correlation Analysis*

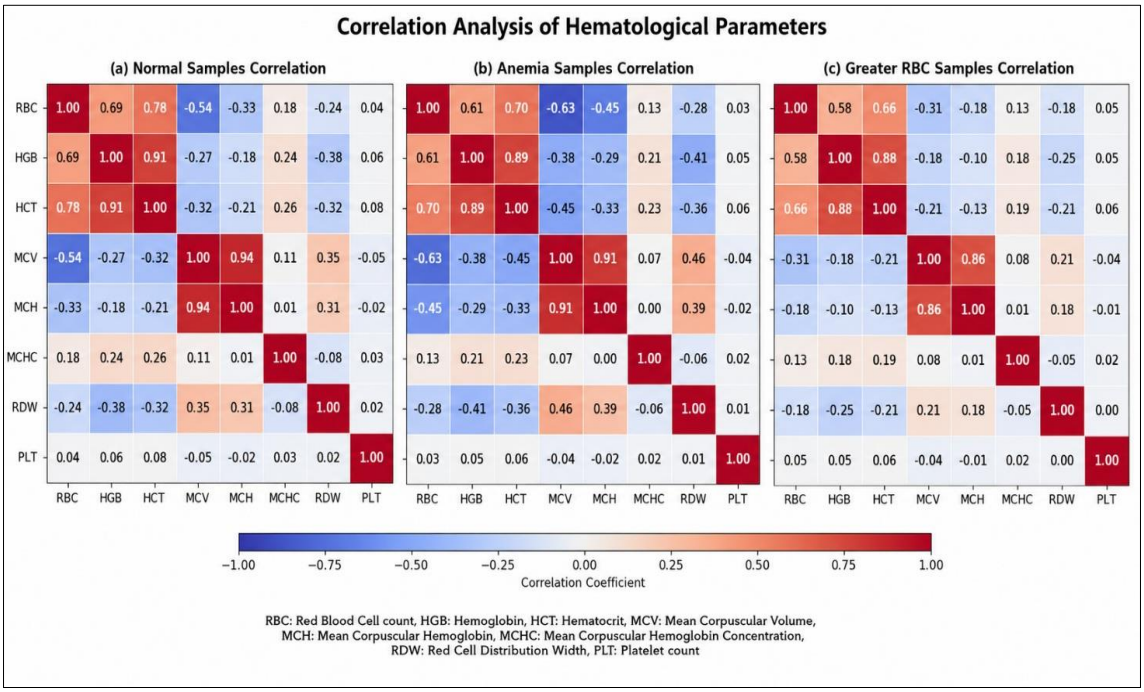
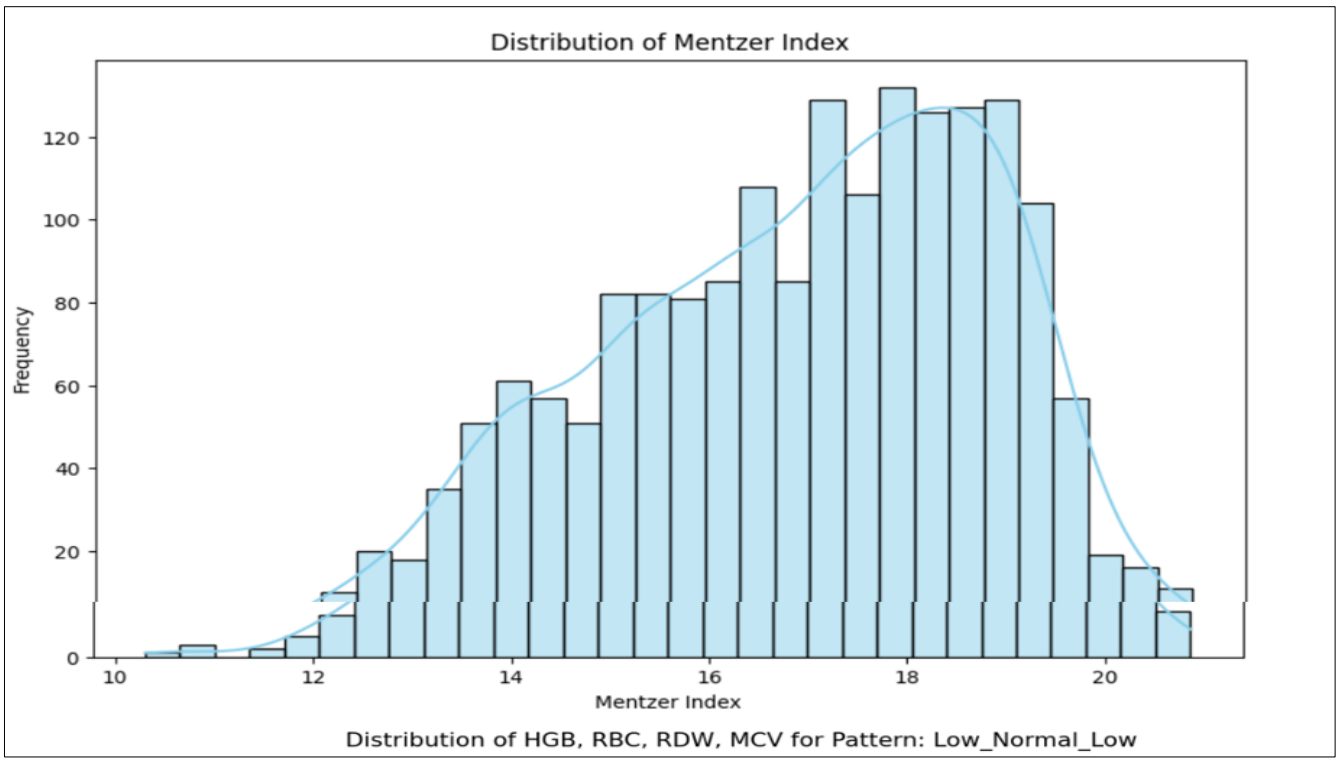


Fig 2 Correlation Analysis

A positive correlation between HGB and HCT was seen in all anemia groups with values > 0.9. The correlation coefficient of HCT-HGB was 0.91, 0.84, and 0.89 in Microcytic anemia, Normocytic anemia, and Macrocytic anemia, respectively. In the same way, MCV and MCH showed very strong positive correlation in Microcytic anemia (0.94) and Normocytic anemia (0.91). There were negative correlations in RBC with Normocytic (r = -0.71) and Microcytic (r = -0.56) anemia. This means that the larger the red blood cell, the lower the red blood count will be.

In most anemia categories, RDW showed a negative relationship with MCH, MCV, and HGB, suggesting that greater RDW is related to lower erythrocyte quality and hemoglobin content.

➢ *Mentzer Index Analysis*



ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

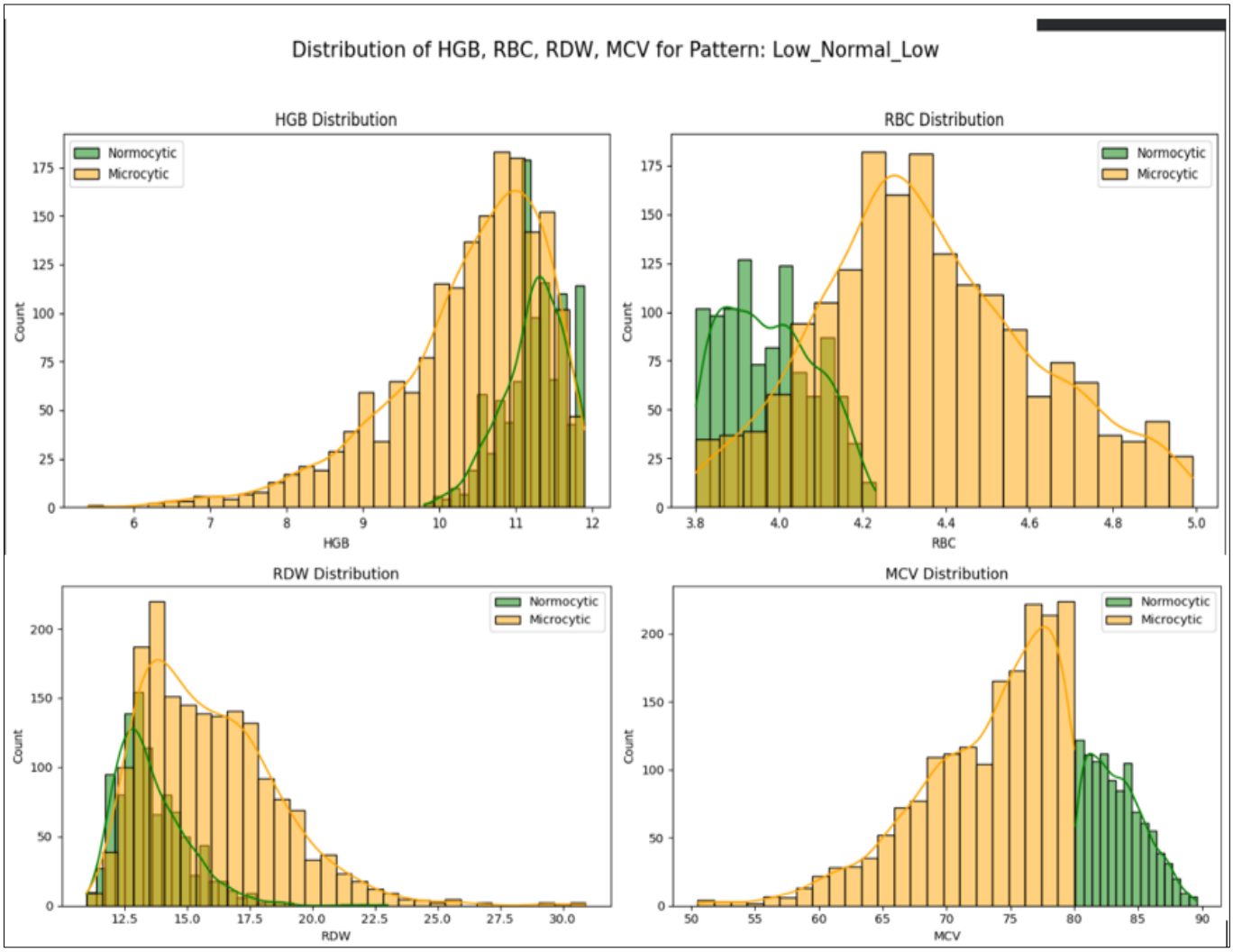


Fig 3 Mentzer Index Analysis

The distribution of the Mentzer Index was approximately 14 to 27, with the majority of observations between 16 and 19. The elevated Mentzer Index values were more prevalent in Microcytic anemia patients, which encouraged the existence of patterns of anemia associated with iron deficiency.

The distribution was fairly bell-shaped and suggested good discriminatory ability for the recognition of microcytic status.

➢ *Distribution Analysis*

Distribution plots compared showed the separation between levels of anemia.

- Microcytic Anemia:

✓ MCV values at the lower range (about 55-80 fL)

At reduced altitudes, HGB and HCT levels are reduced.

- Elevated RDW values
- Higher RDW-to-MCV Ratio

- Normocytic Anemia:

✓ Low values (<80 fL)

Moderate levels of HGB and RBC count

- Lower RDW variability

- Macrocytic Anemia:

✓ Highest MCV values (>100 fL)

- Elevated MCH values
- Reduced RBC concentrations

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

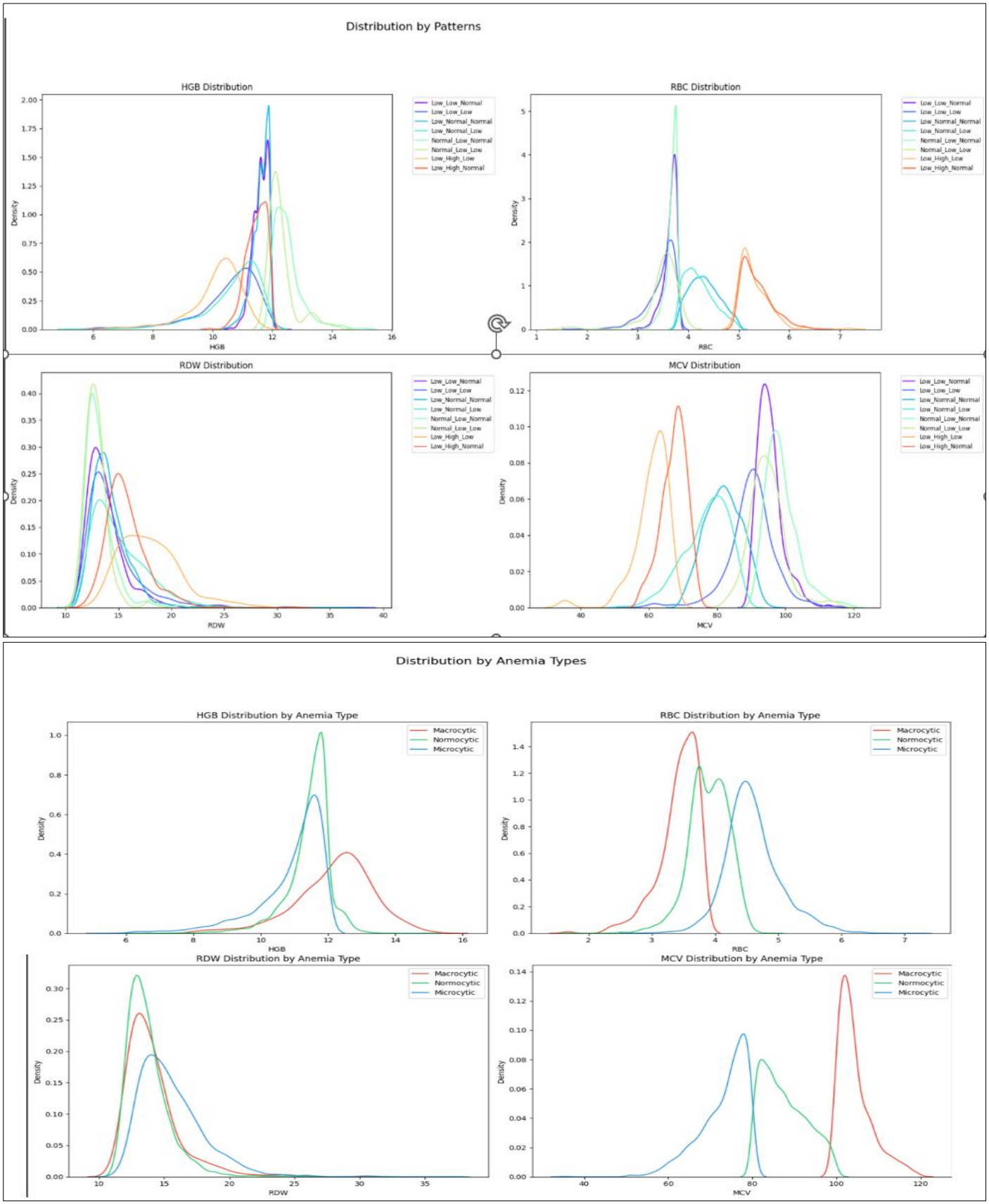


Fig 4 Distribution Analysis

The observed distribution matches clinical hematology literature and is consistent with each other, thus making the set of data valid.

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

➢ *Feature Relationship Analysis*

```
Pattern: Low_Low_Low
DataFrame for Macrocytic (>100):

DataFrame for Normocytic (80-100):

DataFrame for Microcytic (<80):
         RBC  HGB   HCT   MCV   MCH  MCHC   RDW     PLT HGB_Status RBC_Status  \
2809    3.79  7.6  25.4  66.8  20.2  30.2  15.9  272.0        Low        Low
4531    3.49  8.9  27.1  77.9  25.8  33.2  15.9  208.0        Low        Low
4583    3.67  8.8  27.5  75.1  24.0  32.1  15.1  639.0        Low        Low
5881    3.76  9.1  28.3  75.3  24.2  32.3  17.6  325.0        Low        Low
6632    3.70  7.4  25.2  67.9  20.0  29.4  19.7  459.0        Low        Low
...      ...  ...   ...   ...   ...   ...   ...    ...        ...        ...
110457  2.74  7.2  19.9  72.6  26.2  36.0  27.0  427.0        Low        Low
110791  3.58  8.4  27.3  76.2  23.4  30.7  16.5  188.0        Low        Low
111722  3.58  8.0  25.9  72.5  22.5  31.0  15.9  244.0        Low        Low
113934  2.93  6.8  22.8  77.6  23.0  29.7  18.9  295.0        Low        Low
114529  3.56  9.7  28.1  78.8  27.2  34.6  13.7  150.0        Low        Low

        HCT_Status  MentzerIndex  RDW_to_MCV_Ratio  MCH_to_HGB_Ratio  \
2809           Low     17.625330          0.238024          2.657895
4531           Low     22.320917          0.204108          2.898876
4583           Low     20.463215          0.201065          2.727273
5881           Low     20.026596          0.233732          2.659341
6632           Low     18.351351          0.290133          2.702703
...            ...           ...               ...               ...
110457         Low     26.496350          0.371901          3.638889
110791         Low     21.284916          0.216535          2.785714
111722         Low     20.251397          0.219310          2.812500
113934         Low     26.484642          0.243557          3.382353
114529         Low     22.134831          0.173858          2.804124

        PLT_to_RBC_Ratio  MCV_squared      MCV_Category
2809           71.767810      4462.24  Microcytic (<80)
4531           59.598854      6068.41  Microcytic (<80)
4583          174.114441      5640.01  Microcytic (<80)
5881           86.436170      5670.09  Microcytic (<80)
6632          124.054054      4610.41  Microcytic (<80)
...                  ...          ...               ...
110457        155.839416      5270.76  Microcytic (<80)
110791         52.513966      5806.44  Microcytic (<80)
111722         68.156425      5256.25  Microcytic (<80)
113934        100.682594      6021.76  Microcytic (<80)
114529         42.134831      6209.44  Microcytic (<80)
```

Fig 5 Feature Relationship Analysis

A strong linear and nonlinear correlation was observed between various hematological parameters by scatter plot analysis. The largest positive correlation was observed between MCV and MCH in all types of anemia, which supports the fact that the larger the erythrocytes, the more hemoglobin they will contain. Relationships between HCT and HGB were almost linear over all categories and indicate that hematocrit continues to be a good indicator of hemoglobin concentration. RBC showed an inverse relationship with MCV and MCH, especially in the microcytic and normocytic anemia groups, suggesting different mechanisms of erythrocyte production in various types of anemia.

➢ *The Accuracy of Machine Learning Classification Algorithms*

Labels derived from MCV were not included in the final reduced-feature model due to target leakage, and neither were directly derived features.

The Random Forest classifier performed well with an accuracy of:

- Accuracy = 98.55%
- Precision = 98.56%
- Recall = 98.55%
- F1-Score = 98.53%
- ROC-AUC = 99.96%

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

The excellent capability of classification was reflected in the class-wise performance.

- Macrocytic Anemia:

✓ Precision = 1.00
✓ Recall = 0.84
✓ F1-Score = 0.91

- Microcytic Anemia:

✓ Precision = 0.99
✓ Recall = 0.99
✓ F1-Score = 0.99

- Normocytic Anemia:

✓ Precision = 0.98
✓ Recall = 0.99
✓ F1-Score = 0.99

```
Accuracy (reduced features): 0.9855
Precision (weighted, reduced features): 0.9856
Recall (weighted, reduced features): 0.9855
F1-Score (weighted, reduced features): 0.9853
ROC AUC (OvR, weighted, reduced features): 0.9996

Classification Report (reduced features):
              precision    recall  f1-score   support

  Macrocytic       1.00      0.84      0.91        76
  Microcytic       0.99      0.99      0.99       867
  Normocytic       0.98      0.99      0.99      1189

    accuracy                           0.99      2132
   macro avg       0.99      0.94      0.96      2132
weighted avg       0.99      0.99      0.99      2132
```

Fig 6 The Accuracy of Machine Learning Classification Algorithms

This lower recall rate in Macrocytic anemia may be an artifact of the smaller sample size of this class.

➢ *Feature Importance Analysis*

Random Forest feature importance analysis was used to determine the most influential predictors.

- MCH (0.362)
- RBC (0.173)
- MCH-to-HGB Ratio (0.156)
- RDW-to-MCV Ratio (0.092)
- HCT (0.085)

MCH was found to be the most significant, accounting for around 36% of the overall predictive importance. This discovery points towards the importance of hemoglobin in anemia subtyping.

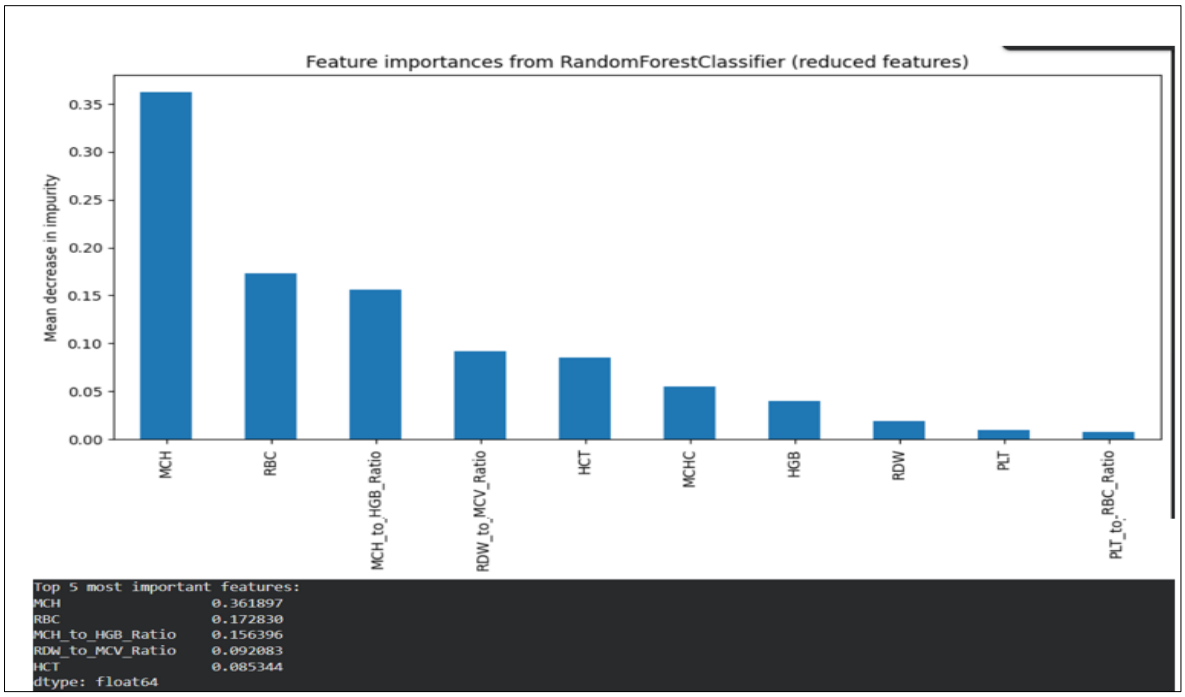


Fig 7 Feature Importance Analysis

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

The second most important variable was the RBC count, demonstrating the significance of the erythrocyte concentration in the diagnosis of anemia. The value of feature engineering was highlighted by significant improvements in classification performance using engineered features like MCH-to-HGB Ratio and RDW-to-MCV Ratio. There was little predictive influence of platelet-related variables.

- *Clinical Implications*

The findings suggest that routine Complete Blood Count (CBC) parameters with machine learning techniques could be used to differentiate the types of anemia. Aside from this, feature engineering can be used to identify clinically relevant relationships between haematological indices, further improving predictive performance.

The proposed framework offers a trustworthy and interpretable approach to automated anemia subtyping, which can help support the early diagnosis and clinical decision-making by healthcare professionals.

VI. CONCLUSION

This study proposed a machine learning approach that encompasses all the key elements of analysis and classification of hematological disorders based on routine Complete Blood Count (CBC) parameters and clinically relevant engineered features. The relationships between the hematology parameters were explored in a large-scale hematology dataset consisting of 115,487 patient records, and the use of this dataset for identifying the pattern of anemia, as well as the development of an automated system for the classification of Microcytic, Normocytic, and Macrocytic anemia. The exploratory data analysis showed significant differences in haematological parameters between the various types of anaemia. The correlation between the various parameters of anemia, Hemoglobin (HGB) with Hematocrit (HCT), and Mean Corpuscular Volume (MCV) with Mean Corpuscular Hemoglobin (MCH) was strong and positive in all different categories of anemia. On the other hand, the negative correlations of Red Blood Cell count (RBC) with the erythrocyte size-related parameters suggest the differences in red blood cell morphology and erythropoietic activity of anemia groups. Distribution analysis also showed good separation between the groups of anemia, thereby guaranteeing the reliability and clinical consistency of the data set.[26]

Additional engineered hematological indices were added to increase the discrimination, such as the Mentzer Index (MI), RDW-to-MCV Ratio, MCH-to-HGB Ratio, and PLT-to-RBC Ratio. Of these, the Mentzer Index was of particular clinical relevance for the diagnosis of microcytic anemia patterns in the case of iron deficiency. These derived features greatly enhanced the representation of hematological relationships and provided useful diagnostic information that was not provided by the conventional CBC parameters. To reduce the amount of target leakage without sacrificing predictive power, a reduced feature set was used to build a Random Forest classifier. Overall, the proposed model exhibited an accuracy of 98.55%, precision of 98.56%, recall of 98.55%, F1-score of 98.53%, and ROC-AUC of 99.96% for the classification of excellent performance. Class-wise evaluation revealed that the machine learning has been effective in automated classification of hematological disorders with good predictive power for all the anemia categories.

The features' importance analysis showed that the MCH was the most important feature, followed by RBC count, MCH-to-HGB Ratio, RDW-to-MCV Ratio, and HCT. The results highlight the importance of using traditional hematological parameters and engineered parameters in the differential diagnosis of anemia subtypes and in enhancing the accuracy of diagnosis.

The proposed framework from a clinical perspective will provide a practical, cost-effective, and interpretable approach to anemia screening and anemia subtyping based on laboratory data that is routinely collected. The model is heavily dependent on CBC parameters and thus has a lot of potential for use in resource-limited healthcare settings where sophisticated diagnostic equipment might not be available. Seamless integration with Laboratory Information Systems (LIS), Electronic Health Records (EHR), and clinical decision support systems may help healthcare providers make earlier diagnoses, better patient stratification, and more targeted treatment decisions.[26]

The results highlight the potential of machine learning alongside hematological feature engineering to enable an accurate, scalable, and clinically relevant strategy for anemia classification overall. Overall, this framework is a step towards the use of AI in the field of hematology and precision medicine, offering a potentially significant improvement in diagnosing the condition, making it more accessible, and enhancing patient outcomes.[27]

REFERENCES

[1]. E. Martínez-Quintana and F. Rodríguez-González, "Anemia in adult congenital heart disease: Prevalence and markers of iron deficiency," International Journal of Cardiology, vol. 157, no. 3, pp. 461–462, 2012.
[2]. S. Munir, S. Ahmed, M. Ahmed, and T. Naru, "Hematological and biochemical profile in β-thalassemia female patients," Pakistan Journal of Medical Sciences, vol. 29, no. 5, pp. 1211–1214, 2013.
[3]. J. M. Guralnik, R. S. Eisenstaedt, L. Ferrucci, H. G. Klein, and R. C. Woodman, "Prevalence of anemia in persons 65 years and older in the United States: Evidence for a high rate of unexplained anemia," Blood, vol. 104, no. 8, pp. 2263–2268, 2004.
[4]. F. Tradati, "Risk of hepatocellular carcinoma in chronic hepatitis C patients undergoing hemodialysis: A multicenter prospective study," Kidney International, vol. 54, no. 6, pp. 2147–2153, 1998.
[5]. V. Laosombat et al., "Clinical and molecular characterization of β⁰-thalassemia frameshift -41/42 mutation in Thai patients," Journal of Pediatric Hematology/Oncology, vol. 23, no. 6, pp. 396–400, 2001.

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26jun1335

[6]. E. Angelucci et al., "Italian guidelines for the management of iron overload in thalassemia major: An analytical hierarchy process approach," Haematologica, vol. 93, no. 8, pp. 1242–1246, 2008.
[7]. S. Birndorf, P. Czerniak, and S. Nudelman, "A hybrid expert system for diagnosing microcytic anemia," Computer Methods and Programs in Biomedicine, vol. 50, no. 3, pp. 227–234, 1996.
[8]. S. T. Chou et al., "RH and Kell antigen-matched transfusion in sickle cell disease patients reduces alloimmunization," Blood, vol. 135, no. 7, pp. 550–565, 2020.
[9]. V. Brancaleoni, E. Di Pierro, I. Motta, and L. Forni, "Laboratory approach to the diagnosis of thalassemia," Blood Reviews, vol. 30, no. 1, pp. 65–73, 2016.
[10]. M. Rahaman et al., "Prevalence and hematological characteristics of Hb E/β-thalassemia in Brunei Darussalam: A retrospective analysis," Asian Journal of Transfusion Science, vol. 16, no. 1, pp. 45–51, 2022.
[11]. A. R. Cohen et al., "Iron chelation therapy and assessment of cardiac iron using T2 MRI in thalassemia," Annals of the New York Academy of Sciences, vol. 1054, pp. 204–213, 2005.
[12]. V. Heves-Girard et al., "Desmopressin responsiveness in non-severe hemophilia A: Clinical outcomes," Haemophilia, vol. 24, no. 5, pp. 859–867, 2018.
[13]. M. Franchini et al., "Incidence of cancer in Italian patients with von Willebrand disease: A national survey," Thrombosis and Haemostasis, vol. 113, no. 5, pp. 1010–1014, 2015.
[14]. R. E. Drews and I. A. Shulman, "Advances in hematology-oncology in 2009: Key updates for internists," Mayo Clinic Proceedings, vol. 85, no. 2, pp. 166–174, 2010.
[15]. E. Francini et al., "Association between ABO blood group and inhibitor development in severe hemophilia A," Journal of Thrombosis and Haemostasis, vol. 19, no. 3, pp. 725–733, 2021.
[16]. A. Sayani et al., "Predicting response to anemia therapy: Role of baseline hemoglobin and erythropoietin," Journal of Clinical Nephrology, vol. 29, no. 2, pp. 98–105, 2015.
[17]. E. Haus, F. Halberg, and G. Cornélissen, "Biological rhythms in hematopoiesis and immune function," Chronobiology International, vol. 1, no. 2, pp. 147–165, 1983.
[18]. W. W. M. Lam et al., "Iron overload complications and health-related quality of life in transfusion-dependent thalassemia," Quality of Life Research, vol. 30, no. 3, pp. 715–728, 2021.
[19]. J. D. Rizzo et al., "Use of erythropoiesis-stimulating agents in cancer patients: ASCO/ASH clinical practice guideline update," Journal of Clinical Oncology, vol. 28, no. 33, pp. 4996–5010, 2010.
[20]. H. L. Muncie and J. Campbell, "Alpha and beta thalassemia: Diagnosis and management," American Family Physician, vol. 80, no. 4, pp. 339–344, 2009.
[21]. E. Beutler and C. West, "Hematologic differences between African-Americans and whites: The roles of iron deficiency and α-thalassemia," Blood, vol. 106, no. 2, pp. 740–745, 2005.
[22]. B. S. Bull, J. Breton-Gorius, et al., "Morphology of the erythron," in Hematology, 4th ed., W. J. Williams, E. Beutler, A. J. Erslev, and M. A. Lichtman, Eds. New York, NY, USA: McGraw-Hill, 1990, pp. 297–316.
[23]. A. J. Erslev, "Production of erythrocytes," in Hematology, 4th ed., W. J. Williams, E. Beutler, A. J. Erslev, and M. A. Lichtman, Eds. New York, NY, USA: McGraw-Hill, 1990, pp. 389–398.
[24]. W. Fried, "Factors that affect the rate of erythropoietin production by extrarenal sites," Annals of the New York Academy of Sciences, vol. 554, pp. 1–8, 1989.
[25]. G. R. Johnson, "Erythropoietin," British Medical Bulletin, vol. 45, pp. 506–514, 1989.
[26]. P. A. Keown, "Recombinant human erythropoietin: From concept to clinic," Transplantation Proceedings, vol. 21, Suppl. 2, pp. 49–53, 1989.
[27]. W. Jelkmann and G. Wiedemann, "Serum erythropoietin level: Relationships to blood hemoglobin concentration and erythrocytic activity of the bone marrow," Klinische Wochenschrift, vol. 68, pp. 403–407, 1990.