JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Multilabel Classification of Anemia Types Using Machine Learning: A Comparative Analysis of Problem Transformation Techniques and Classifiers.

Dr. Bhavinkumar A. Patel[1] and Prof. Dr. Ajay Parikh[2]

[1] Dept. of Computer Science, Gujarat Vidyapith, Ahmedabad, Gujarat, India
[2] Professor, Dept. of Computer Science, Gujarat Vidyapith, Ahmedabad, Gujarat, India

Abstract. A blood count test (CBC) is said to be the primary and important test for a complete body health checkup, which plays a significant role in the primary diagnosis of bacterial (Neutrophils ↑) or viral (Lymphocyte ↑) infection, cancer, anemia, vitamin and mineral deficiency. Anemia is a sign of decreased blood flow in the body. The CBC parameter pattern is very complex so it is difficult to predict different types of anemia. In our research, a Multilabel classification model has been developed to predict anemia and its eight types IDA (Iron Deficiency Anemia), VitaminB12, Aplastic, SickleCell, FDA (Folate Deficiency Anemia), ACD (Anemia of Chronic Disease), Hemolytic, and Thalassemia. A Multilabel classification model has been developed using machine learning problem transformation's three techniques: binary relevance, classifier chain, and label power set with six different classifiers (J48, Naïve Bayes, Logistic Regression, Random Forest, and SVM). Results provide interesting insights from classified data J48 algorithm in the classifier chain technique is proved to be the best compared to other methods.

Keywords: Anemia, Healthcare, Multi-Label Classification, Machine Learning.

1 Introduction

CBC or full blood count report plays an important role in the complete checkup of the body's health. It is useful for diagnosing anemia, infections, cancer, vitamin, and mineral deficiency. Anemia is a common problem among people around the world. Due to which there is a possibility of causing serious illness in the body. So early diagnosis is very important. Which can cause worldwide epidemics and mortality [5]. Anemia and its 8 types are classified based on CBC. It indicates a type of anemia caused by abnormal CBC parameters. Many variations and complexities have been observed in the pattern of CBC parameters. Modern approaches to health care the prognosis of anemia is based on the basic blood test parameters of the clinical laboratory, thus it is very important to prevent anemia at an early stage. Which can improve their health and education level to prevent the risk of serious diseases.
A proposed model has been predicted IDA (Iron Deficiency Anemia), VitaminB12, Aplastic, SickleCell, FDA (Folate Deficiency Anemia), ACD (Anemia of Chronic Disease), Hemolytic, and Thalassemia using a multi-label classification problem transformation method by machine learning. Various classifiers that come beneath this approach are Binary Relevance (BR), Classifier Chain (CC), and Label Powerset (LP), etc.

1.1. Types of Anemia

Anemia is the low level of hemoglobin (Hb) and red blood cells (RBC) in the human body. Here below Fig.1 shows the types of anemia.

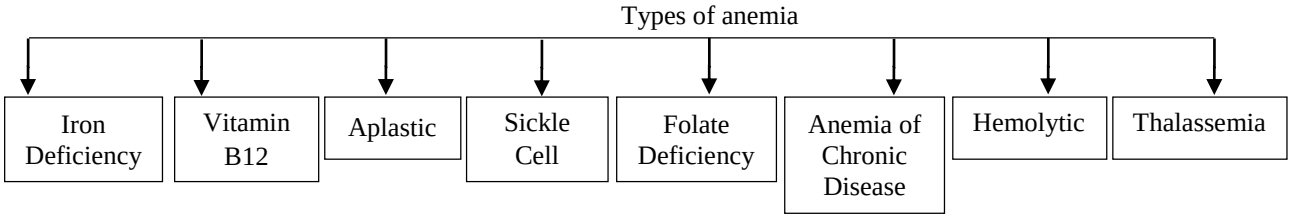


Fig. 1. Types of anemia

1.1.1. Iron deficiency

Iron deficiency anemia which is caused due to a low amount of iron in the body is called iron deficiency anemia. Iron is a major and essential component of hemoglobin for its proper functioning. Long-term blood loss is the main cause of iron deficiency in the body. The fundamental parameters of the CBC report help identify iron deficiency anemia, including low HB, low PCV, microcytic MCV, and low MCHC.
Case-1
Status of IDA when anemia is present: When anemia is present, clinical parameters of iron deficiency anemia include low HB and RBC levels.
Case-2
IDA when anemia is not present: Iron deficiency anemia when anemia is not present the condition of clinical parameters is observed only in case of low hemoglobin level. Which can be called stage 3, i.e. latent iron deficiency anemia
In stage 3, anemia (decreased hemoglobin level) is present but the red blood cell count remains normal.

1.1.2. Vitamin B12

Vitamin B12 deficiency occurs when the body has insufficient levels of vitamin B-12. This is an important vitamin for the production of red blood cells and the healthy functioning of the nervous system.
Vitamin B12 has low levels of both Hb and RBC parameters, with the clinical parameters of anemia being observed. The parameters of CBC (HB-Low, PCV-Low, MCV-Macrocytic, and MCHC-High) are also taken into account for its prediction. When there is no anemia but there is vitamin B-12, only HB is found to be below.

1.1.3. Aplastic

Aplastic anemia is an autoimmune disease in which the body fails to produce a sufficient number of blood cells. It has the lowest proportion of major CBC parameters like RBC, WBC, and Platelet. The parameters of CBC (HB-Low, RBC-Low, WBC-Low, PCV-Low, MCV-Macrocytic, MCH-High, and Platelet-Low) are also taken into account for its prediction [10].

1.1.4. Sickle cell

A mutation in the genetic pathway that produces hemoglobin causes the formation of HbS (sickle cell) instead of normal HbA hemoglobin.
Blood cells with HbS are sickle-shaped instead of the normal shape. These sickle-shaped cells break down quickly.
Early diagnosis of this disease is possible from birth.
Signs of the disease are seen according to the level of HbS in the body.
Treatment of this disease includes folic acid medication, Hydroxurea, chronic transfusion therapy, and periodic blood transfusions.
Sickle Cell Anemia is predicted based on the parameters of the CBC (HB-Low, RBC-Low, WBC-High, PCV-Low, MCV-Macrocytic, MCH-High, and Platelet-High) [11].

1.1.5. Folate Deficiency Anemia (FDA)

Like vitamin B-12, folic acid deficiency also causes large blood cells (granulocytosis) and large oval blood cells and more fragmented white blood cells. The total blood cell size is also increased and the percentage of fibrous cells is low; but the serum level of vitamin B-12 is normal and the level of folate in blood cells and serum is low. Folic acid is useful in the production of methionine and thymidylate and therefore is essential for the synthesis of deoxyribonucleic acid. Its deficiency causes defects in the regeneration of the nucleus of newly formed cells.

1.1.6. Anemia of Chronic Disease (ACD)

Anemia of chronic disease is multifactorial. The diagnosis usually requires the presence of a chronic inflammatory condition such as infection, autoimmune disease, kidney disease, or cancer. Microcytic or normocytic anemia with a low reticulocyte count characterizes it. Serum iron and transferrin values are typically low, while serum ferritin values may be normal or elevated. Treatment aims to reverse the underlying disorder and, in some cases, administer erythropoietin.

1.1.7. Hemolytic

A group of disorders in which circulating blood cells are continuously or episodically destroyed before their full lifespan is called hematopoietic aplastic anemia. In such cases, the bone marrow produces eight times more blood cells, and aplastic anemia occurs when the rate of blood cell destruction exceeds the rate of new blood cell formation. When hemolysis occurs or blood is lost from the body, the total blood cell count decreases and is more than 3% per week. Therefore, if it is low, it indicates a decrease in blood cell production. When hemolysis occurs, the percentage and number of fibrous cells increase, but if there is a concomitant folic acid deficiency, they do not increase. It is not possible to determine if there is a sudden, massive loss of blood, but if there is an increased or decreased percentage of reticulocytes and a decreasing total blood cell count, it indicates a hemolytic disorder.

1.1.8. Thalassemia.

Due to some genetic reasons, the blood cells break down quickly due to the lack of hemoglobin, which is responsible for this anemia. How quickly the blood cells break down due to genetic problems. Thalassemia major, intermedia, minor is determined based on that. Since this anemia is genetic, it comes to children from the genes of the parents. This disease can be diagnosed before birth. The treatment of this disease depends on how quickly the blood cells break down.
Thalassemia major: Regular blood transfusions are required from time to time and iron is given.
Thalassemia intermedia: Regular blood transfusions are required after a certain age.
Thalassemia minor: Blood transfusions are not required in most cases.
If both parents have thalassemia, the chances of their child having thalassemia major increase. Therefore, if thalassemia testing is done before marriage, the chances of having children with thalassemia major decrease.

Table 1. Differences between the laboratory parameters of the types of anemia

CBC Parameter	IDA	VitaminB12	Aplastic	SickleCell	FDA	ACD	Hemolytic	Thalassemia
HB	⬇	⬇	⬇	⬇	⬇	⬇	⬇	⬇
RBC	⬇	⬇	⬇	⬇	⬇	⬇	⬇	⬇
WBC	ordinary	ordinary	⬇	⬆	⬇	⬇	ordinary	⬆
PCV	⬇	⬇	⬇	⬇	⬇	⬇	⬇	⬇
MCV	⬇ Microcytic	⬆ Macrocytic	⬆ Macrocytic	⬆ Macrocytic	⬆ Macrocytic	Normocytic	⬆ Macrocytic	⬇ Microcytic
MCH	ordinary	ordinary	⬆	⬆	⬆	ordinary	ordinary	⬇
Platelet	⬆	ordinary	⬇	⬆	⬇	ordinary	⬆	ordinary
MCHC	⬇	⬆	ordinary	ordinary	ordinary	ordinary	ordinary	⬇

The effect of abnormal conditions of different parameters of CBC on the types of anemia is shown in the above table.

2. Materials and Methods

2.1. Data Set

For anemia multi-label classification, the CBC report of the newly admitted students of Gujarat Vidyapith for higher education from the last five years i.e. 2014 to 2018 has been used as a dataset. In total 2190 samples were taken. In addition, CBC samples of people aged 21 to 32 years have been taken from a total of 5 clinical pathology laboratories, 2 from Ahmedabad district, and 3 from Gandhinagar district. It includes a total of 40938 sample training datasets and 768 sample testing datasets. There are many disparities due to different testing equipment. So it can be said that not all the parameters of the CBC report are important for the prognosis of anemia. That can be said based on our previous research paper [5][18]. Only 20 main parameters have been taken for it. Which plays an important role in the diagnosis of anemia.

Table 2. Multi Label Anemia Classification Dataset Attributes

Sr.no	Attribute	Description	Sr.no	Attribute	Description
1.	Age	Age Group 21-32	11.	Platelet	Male < 13500 Female < 157000
2.	Sex	Male Female	12.	Anemia_Status (Target) Lable-1	**Male** HB<13.5, RBC<4.3 **Female** HB<12, RBC<3.5
3.	HB	Male <13.5 Female < 12	13.	IDA (Target) Lable-2	**Low** HB,RBC,PCV,MCHC MCV=Microcytic
4.	RBC	Male < 4.3 Female < 3.5	14.	VitaminB12 (Target) Lable-3	**Low** HB, PCV, **High** MCHC MCV=Macrocytic
5.	PCV (HCT)	Male < 41 Female < 36	15.	Aplastic (Target) Lable-4	**Low** HB, RBC, WBC, PCV, Platelet **High** MCH MCV=Macrocytic
6.	MCV	Microcytic < 80 Normocytic <80 >100 Macrocytic >100	16.	Sickle Cell (Target) Lable-5	**Low** HB, RBC, PCV **High** WBC, MCH, Platelet MCV=Macrocytic
7.	MCH	Normal < 25.4 > 34.6 Abnormal ↓ <25.4 ↑ >34.6	17	FDA (Target) Lable-6	**Low** HB, RBC, PCV **High** RDW, MCV=Macrocytic
8.	MCHC	Normal <31 >36 Abnormal ↓<31 ↑>36	18	ACD (Target) Lable-7	**Low** HB, RBC, PCV MCV=Normocytic
9.	RDW	Normal <11 >15 Abnormal ↓<11 ↑>15	19	Hemolytic (Target) Lable-8	**Low** HB, RBC, PCV **High** RDW, Platelet MCV=Macrocytic
10.	WBC	Normal <3400 >9600 Abnormal ↓<3400↑>9600	20	Thalassemia (Target) Lable-9	**Low** HB, RBC, PCV MCV= Microcytic **High** WBC

Multi-label classification of anemia dataset using a total of 20 parameters as above.

2.2. Multi-Label Classification

Multi-label classification methods are closely related to modern real-world applications, including disease diagnosis and genre classification.

Table 3. Multi-Label Classification.

IDA	VitaminB12	Aplastic	Sickle Cell	FDA	ACD	Hemolytic	Thalassemia
No	Yes	Yes	No	Yes	No	Yes	Yes

The above table.3. Shows five out of eight positive labels.

Where we have more than one target variable, these are known as multi-label classification problems.

Multi-label classification techniques are divided into two main categories. Problem transformation methods convert multi-label datasets into one or more single-label datasets so that the classification problem can be solved by single-label classification. Direct multi-label data is handled by the algorithm adaptation method.

Therefore multi-label classification is evaluated for solving the problem by problem transformation method.

2.2.1. Problem Transformation

The main purpose of the approach is to convert the actual Multilabel into a set of single-label classification. Its function is not directly dependent on the classification system used and is therefore independent. In this research, cluster-related algorithms are used, which have many problem transformation methods.
In this research, three methods have been used related to the problem transformation category. Binary Relevance, which transforms a multi-label problem into several single-label problems. Categorical chains are called binary compatibility corrections. It concatenates each label one by one using the previous target variable as the input space. Label power set, which converts a multi-label problem into a single-label problem. Where X is independent variable and Y is target variable.

Table 4. Multi Label Classification Labels.

X	Y1 (Anemic)	Y2 (IDA)	Y3 (VitB12)	Y4 (APA)	Y5 (SickleCell)	Y6 (FDA)	Y7 (ACD)	Y8 Hemolytic	Y9 Thalassemia
X1	1	0	0	0	0	0	1	1	0
X2	1	1	0	0	0	0	0	1	1
X3	1	0	1	1	0	1	0	1	1
X4	0	0	1	0	1	0	0	0	0
X5	0	1	1	0	1	0	0	0	1

In the above table.4. X1, X2, X3…Xn are the independent variables and Y1 (Anemic), Y2 (IDA), Y3 (VitB12), Y4 (APA), Y5 (SickleCell), Y6 (FDA), Y7 (ACD), Y8 Hemolytic, Y9 Thalassemia are the target variables. Where Y1 is anemic or non-anemic, Y2 is an iron deficiency, Y3 is a vitamin B12 deficiency, Y4 is aplastic, and Y5 is a sickle cell, Y6 is Folate Deficiency Anemia (FDA), Y7 is Anemia of Chronic Disease (ACD), Y8 is Hemolytic, and Y9 is Thalassemia. Where 0 is for negative and 1 is for positive.

2.2.1.1. Binary Relevance

BR is one of the most widely used transformation methods. BR treats each label's prediction as an independent binary classification function. Thus, BR creates M binary classifier, for each different label L (where M = L). For the classification of new patterns, BR Labels out Li attachment which is a positive prediction by M classifiers. The main drawback of this method is the fact that it assumes that the labels assigned to the example are independent, ignoring the possible relationships between the possible labels. This is a simple technique that essentially treats each label as a separate classification problem.

Table 5. Binary Relevance each Label Classification

X	Y1
X1	1
X2	1
X3	1
X4	0
X5	0

X	Y2
X1	0
X2	1
X3	0
X4	0
X5	1

X	Y3
X1	0
X2	0
X3	1
X4	1
X5	1

X	Y4
X1	0
X2	0
X3	1
X4	0
X5	0

X	Y5
X1	0
X2	0
X3	0
X4	1
X5	1

X	Y6
X1	1
X2	0
X3	1
X4	0
X5	0

X	Y7
X1	1
X2	0
X3	0
X4	0
X5	0

X	Y8
X1	1
X2	1
X3	1
X4	0
X5	0

X	Y9
X1	0
X2	1
X3	1
X4	0
X5	1

The binary relevance problem divides into 9 different single class classification problems as shown in the above table.5.
Experimental results have been evaluated using different evaluation measures, with example-based measures (hemming loss, accuracy), label-based measures (micro F1 and macro F1), and ranking-based measures (one-error, coverage, and Average accuracy) is detected.

The below table compares the multi-label classification evaluation model based on its different criteria using different classifiers with the help of the binary relevance (BR) method.

Table 6. A. Binary Relevance Model Evaluation Comparison.

Classifier	J48	SVM	Decision Table	Random Forest	Logistic R.	Naïve Bayes
Accuracy ↑	1	0.999	0.997	0.997	0.996	0.9
Hamming Loss ↓	0	0	0.001	0.001	0.001	0.031
Micro-F1 ↑	1	0.999	0.998	0.998	0.998	0.935
Macro-F1 ↑ Averaged by example	0.936	0.935	0.934	0.933	0.933	0.88
Macro-F1 ↑ Averaged by label	1	0.999	0.999	0.997	0.994	0.948
One Error ↓	0.064	0.065	0.064	0.064	0.064	0.068
AvgPrecision ↑	1	0.999	1	1	1	0.996

B. Label wise Accuracy Comparison.

Label	J48	SVM	Decision Table	Random Forest	Logistic R.	Naïve Bayes
Anemic	1	1	1	0.999	1	0.990
IDA	1	1	1	1	1	0.982
VitB12	1	0.999	1	1	0.996	0.985
APA	1	0.999	0.999	0.999	1	0.989
SickleCell	1	1	1	1	1	0.976
FDA	1	0.999	0.996	0.998	0.999	0.904
ACD	1	1	1	0.999	0.998	1
Hemolytic	1	0.999	0.999	0.998	1	0.947
Thalassemia	1	1	1	1	1	0.950

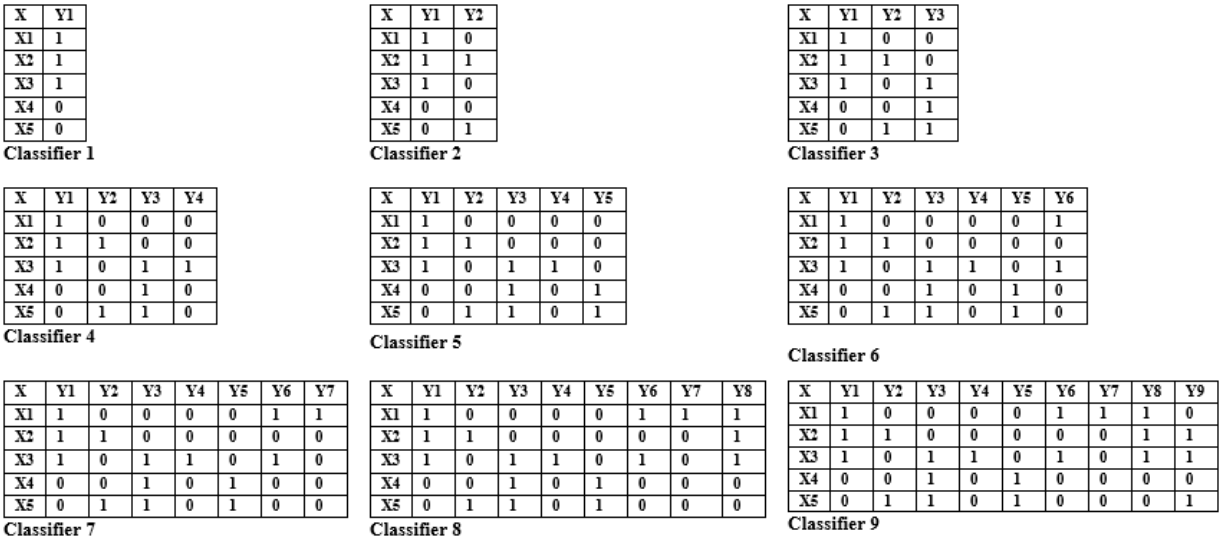
The above table.6. A. shows model evaluation comparison with 6 different evaluation metrics and table.6. B. shows an example base evaluation of 9 different labels using the binary relevance method of problem transformation of multi-label classification. The best results achieved by the J48 classifier in each step are strong. Which is statistically significant.

2.2.1.2. Classifier Chains

The classifier chains method based on the BR method eliminates the shortcomings of the BR and performs a high forecast, retaining the important advantage of the BR, a very important short-term complexity [19]. CC provides a normal problem transformation method that has previously achieved BR efficiency and competes with higher accuracy of more computational complex methods [8].

In this method, the first classifier is trained only on the input data, and then each subsequent classifier is trained on the input space and all previous classifiers in the chain.

Here X is taken as the input space and Y as the label.



X	Y1
X1	1
X2	1
X3	1
X4	0
X5	0

Classifier 1

X	Y1	Y2
X1	1	0
X2	1	1
X3	1	0
X4	0	0
X5	0	1

Classifier 2

X	Y1	Y2	Y3
X1	1	0	0
X2	1	1	0
X3	1	0	1
X4	0	0	1
X5	0	1	1

Classifier 3

X	Y1	Y2	Y3	Y4
X1	1	0	0	0
X2	1	1	0	0
X3	1	0	1	1
X4	0	0	1	0
X5	0	1	1	0

Classifier 4

X	Y1	Y2	Y3	Y4	Y5
X1	1	0	0	0	0
X2	1	1	0	0	0
X3	1	0	1	1	0
X4	0	0	1	0	1
X5	0	1	1	0	1

Classifier 5

X	Y1	Y2	Y3	Y4	Y5	Y6
X1	1	0	0	0	0	1
X2	1	1	0	0	0	0
X3	1	0	1	1	0	1
X4	0	0	1	0	1	0
X5	0	1	1	0	1	0

Classifier 6

X	Y1	Y2	Y3	Y4	Y5	Y6	Y7
X1	1	0	0	0	0	1	1
X2	1	1	0	0	0	0	0
X3	1	0	1	1	0	1	0
X4	0	0	1	0	1	0	0
X5	0	1	1	0	1	0	0

Classifier 7

X	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8
X1	1	0	0	0	0	1	1	1
X2	1	1	0	0	0	0	0	1
X3	1	0	1	1	0	1	0	1
X4	0	0	1	0	1	0	0	0
X5	0	1	1	0	1	0	0	0

Classifier 8

X	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9
X1	1	0	0	0	0	1	1	1	0
X2	1	1	0	0	0	0	0	1	1
X3	1	0	1	1	0	1	0	1	1
X4	0	0	1	0	1	0	0	0	0
X5	0	1	1	0	1	0	0	0	1

Classifier 9

Fig.2. Multi Label Chain Classifier each Labels chain.

Both this method and the binary relevance method are similar, the only difference is that they form a chain to maintain the relationship of the label.

The below table compares the multi-label classification evaluation model based on its different criteria using different classifiers using the classifier chains (CC) method.

Table 7. A. Chain Classifier Model Evaluation Comparison.

Classifier	J48	SVM	Decision Table	Random Forest	Logistic R.	Naïve Bayes
Accuracy ↑	1	0.999	1	1	1	0.85
Hamming Loss ↓	0	0	0	0	0	0.049
Micro-F1 ↑	1	0.999	1	1	1	0.902
Macro-F1 ↑ Averaged by example	0.936	0.935	0.936	0.936	0.936	0.861

Macro-F1 ↑ Averaged by label	1	0.999	1	1	1	0.908
One-Error ↓	0.064	0.065	0.064	0.064	0.064	0.142
AvgPrecision ↑	1	0.999	1	1	1	0.9

B. Label wise Accuracy Comparison.

Label	J48	SVM	Decision Table	Random Forest	Logistic R.	Naïve Bayes
Anemic	1	1	1	1	1	0.978
IDA	1	1	1	1	1	0.969
VitB12	1	0.999	1	1	1	0.989
APA	1	0.999	1	1	1	0.971
SickleCell	1	1	1	1	1	0.960
FDA	1	0.999	1	1	1	0.951
ACD	1	1	1	1	1	0.997
Hemolytic	1	0.999	1	1	1	0.806
Thalassemia	1	1	1	1	1	0.936

The above table7. A. shows model evaluation comparison with 6 different evaluation metrics and table.7. B. shows an example base evaluation of 9 different labels using the Classifier Chains method of Problem Transformation. The best results achieved by the J48 classifier in each step are strong. Which is statistically significant.

2.2.1.3. Label Power Set

In the problem transformation approach, the label power set method creates a new class for each connection of the label and then uses a multiclass classification approach to solve the problem. Its disadvantage is that this approach leads to a fatal increase in the number of classes, as a result, many generated classes have very low labeled patterns that lead to over-exploitation. And the relationship between BR labels is not taken into account. In BR this defect is eliminated by the label power set, also known as label cardinality [8].

Moreover, using all unique existing subsets (specific subsets of labels) of multi-labels in the training pattern using class target values.

Here X is taken as the input space and Y as the label.

Table 8. Label Powerset each label Classification

X	Y1 (Anemic)	Y2 (IDA)	Y3 (VitB12)	Y4 (APA)	Y5 (SickleCell)	Y6 (FDA)	Y7 (ACD)	Y8 Hemolytic	Y9 Thalassemia
X1	1	0	0	0	0	0	1	1	0
X2	1	1	0	0	0	0	0	1	1
X3	0	0	1	0	1	0	0	0	0
X4	1	0	0	0	0	0	1	1	0
X5	1	1	0	0	0	0	0	1	1
X6	0	0	1	0	1	0	0	0	0

The above table.8.shows the same labels in X1 and X4 and the same set of labels in X3 and X6. Therefore, the label power set transforms this problem into a multi-class problem as described below.

Table 9. Label Powerset each label Classification

X	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y1	1	2	3	1	4	3	5	6	3

The label power set problem as shown in the above table.9. has given each possible label combination a unique class.

Table 10. Label Power Set Model Evaluation Comparison.

Classifier	J48	SVM	Decision Table	Random Forest	Logistic R.	Naïve Bayes
Accuracy ↑	1	0.996	0.94	0.999	1	0.981
Hamming Loss ↓	0	0.001	0.021	0	0	0.005
Micro-F1 ↑	1	0.997	0.955	1	1	0.991
Macro-F1 ↑	0.936	0.933	0.883	0.936	0.936	0.927

Averaged by example						
Macro-F1 ↑ Averaged by label	1	0.997	0.919	0.999	1	0.985
One-Error ↓	0.064	0.066	0.116	0.064	0.064	0.08
AvgPrecision ↑	1	0.997	0.961	1	1	0.989

B. Label wise Accuracy Comparison.

Label	J48	SVM	Decision Table	Random Forest	Logistic R.	Naïve Bayes
Anemic	1	0.999	0.994	1	1	0.999
IDA	1	1	0.986	1	1	0.999
VitB12	1	0.999	0.974	1	1	0.994
APA	1	0.997	0.963	1	1	0.988
SickleCell	1	0.996	0.998	1	1	0.990
FDA	1	0.999	0.949	1	1	0.997
ACD	1	1	0.989	1	1	0.998
Hemolytic	1	0.998	0.967	1	1	0.996
Thalassemia	1	1	0.989	1	1	0.998

In this table.10. A. shows model evaluation comparison with 6 different evaluation metrics and table.10. B. the outcome acquires by each Multilabel classification method are classified on the basis of six supervised algorithms. The results are presented using different evaluation measures. The best results achieved by the J48 classifier in each step are strong. Which is statistically significant.

In the proposed work, the hamming-loss and accuracy of the example-based steps as well as in label-based measures, average precision, and one-error are used to appraise the operation of the machine learning classifier.

3. Result and Discussion

The dataset contains 40938 samples in the training set and 768 samples in the test set where each component has 9 class labels. A multi-label classification has been performed on the anemia dataset using the Meka-release-1.9.6 tool. Problem transformation methods are used to build a model and evaluate their performance using measures such as average precision, one error, accuracy, and hamming loss.

The operation of multi-label classifiers with various base classifiers, including J48, SVM, Random Forest, Logistic Regression, Decision Tree, and Naive Bayes is present in the previous table.

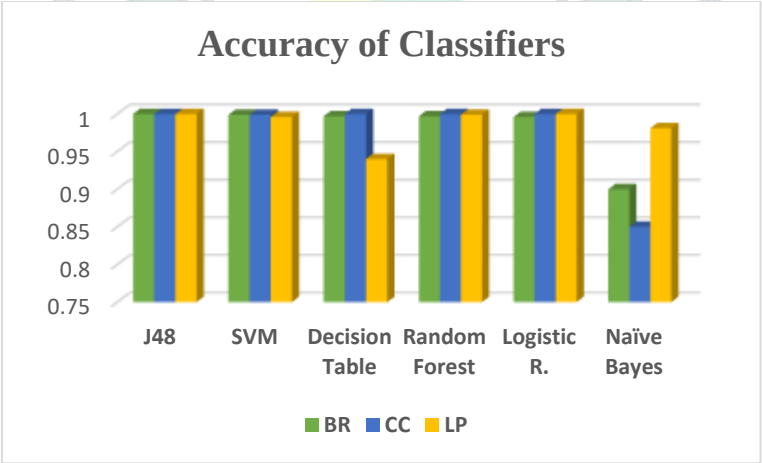


Fig.3. Accuracy of classifiers difference.

Above Fig.3. Shows Classifier Chains with J48 classifier Depending on the method, the accuracy of the machine learning classifier is more than any other classifier.

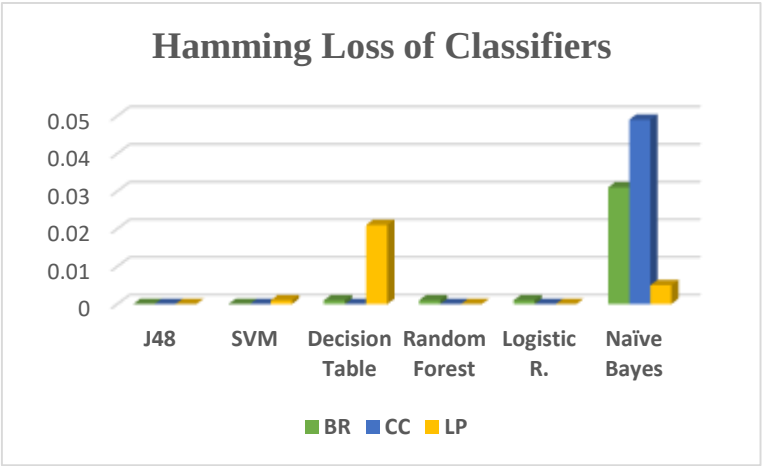


Fig.4. Hamming- Loss classifiers difference.

Superior classification performance is indicated by a Hamming loss value of zero or lower. From Fig.4, it can be seen that the J48 is showing the good performance of the CC based model with base classifier in which the hamming loss value is found below.

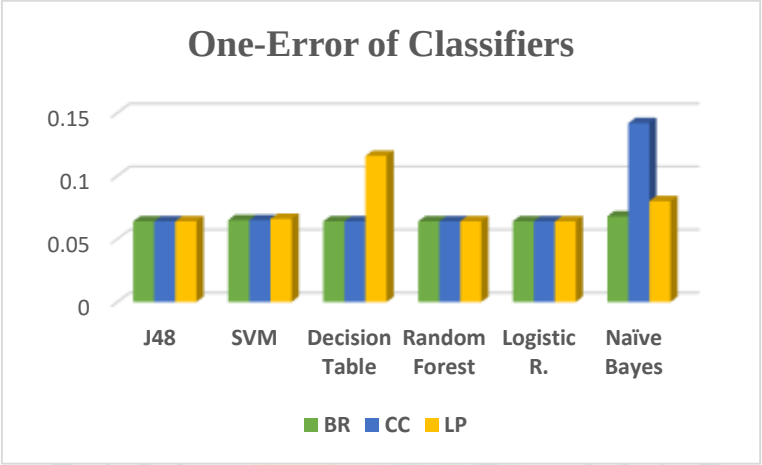


Fig.5. The difference between a One-Error of classifiers.

Like the Hamming Loss, a reduced one-error value signifies a well-performing classifier. As shown in Fig.5, better performance of both BR and CC-based classification is observed.

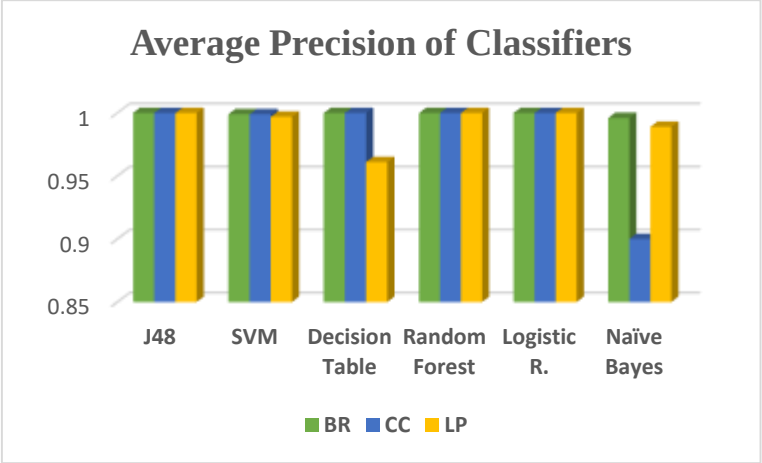


Fig.6. The difference between an Average Precision of classifiers.

Overall results in terms of accuracy and other multi-label classification criteria suggest that BR And CC-based J48 classifiers best performance. Classification provides significantly higher performance. Therefore, multi-label classification can work very well in anemia-type classification tasks in the CBC test.

The results indicate that the chain classification method using J48 classification provides a better classification of anemia type than other methods. An analogical analysis of multi-label classifiers based on different operational steps is presented in the diagrams.

4. Conclusions and Future work

Thus, the conclusion of the early diagnosis of anemia is very important. Through this severe disease can be avoided. An Individual's health can be corrected and the dropout proportion can be decreased.

The eight different types of anemia classification are also useful for predicting many serious diseases. IDA, Vitamin B12, Aplastic, Sickle Cell, FDA (Folate Deficiency Anemia), ACD (Anemia of Chronic Disease), Hemolytic, and Thalassemia anemia have been classified using classification algorithms. This work has been carried out with problem transformation-based methods like Binary Relevance, Classifier Chains, and Label Powerset. The analysis is done using the classification function. The result of the multi-label classification shows that J48 classifier in CC-based classifications offer significantly better performance than other classifications.

As research opportunities and future work, the task of classifying different types of vitamin and mineral deficiency, infections, inflammations, and cancer could be expanded using other machine learning algorithms to improve performance. Chronic diseases and other serious diseases can be detected by other future parameters of CBC.

References:

1. Justice Williams Asare, Peter Appiahene, Emmanuel Timmy Donkoh, Giovanni Dimauro.: Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images. John Wiley & Sons Ltd. (2023).
2. Justice Williams Asare, Peter Appiahene, Emmanuel Timmy Donkoh.: Detection of anaemia using medical images: A comparative study of machine learning algorithms – A systematic literature review. Elsevier, Science Direct (2023).
3. Patel, Bhavin., Parikh, Ajay.: Anemia Multilabel Classification based on Problem Transformation methods. Proceedings Of Second International Conference on Computing, Communications, and Cyber-Security IC4S 2020 Springer Nature Singapore Pte. Ltd (Springer LNNS 2021).
4. Sasikala, Rasitha Banu, Thgani Babiker, Pushpalatha.: A Role of Data Mining Techniques to Predict Anemia Disease. International Journal of Computer Applications (0975 – 8887) Volume 174 – No. 20, February 2021
5. Patel, Bhavin., Parikh, Ajay.: Impact Analysis of the Complete Blood Count Parameter using Naive Bayes. IEEE 5th International Conference on Inventive Computation Technologies (ICICT 2020).
6. Christine, A. Hughes.: Anemia Assessment. Springer Nature Switzerland AG (2019)
7. Patel, Bhavin., Parikh, Ajay.: Search for an essential parameter & technique for effective prediction of disease using Machine Learning Technique., JETIR, Volume 5, Issue 10, October (2018)
8. M, Pushpaa, S, Karpagavalli.: Multi-label Classification: Problem Transformation methods in Tamil Phoneme classification. Elsevier, Science Direct, 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22- 24 August (2017)
9. Jain, Shubham.: Solving Multi-Label Classification problems (Case studies included). Analyticsvidhya. (2017)
10. Pooja Tukaram Dalvi, Nagaraj Vernekar.: Anemia Detection using Ensemble Learning Techniques and Statistical Models. IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India
11. Raed, Alazaidah., Fadi Thabtah., Qasem, Al-Radaideh.: A Multi-Label Classification Approach Based on Correlations Among Labels. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 2 (2015)
12. Andres Felipe Giraldo-Forero., Jorge Alberto Jaramillo-Garzon., Cesar German Castellanos-Domınguez.: Evaluation of Example-BasedMeasures for Multi-label Classification Performance. F. Ortuno and I. Rojas (Eds.): IWBBIO 2015, Part I, LNCS 9043, pp. 557–564, 2015. Springer International Publishing Switzerland (2015)
13. Ling, Li., Huawen Liu., Zongjie, Ma., Yuchang, Mo., Zhengjie, Duan., Jiaqing Zhou, Jianmin, Zhao.: Multi-label Feature Selection via Information Gain. X. Luo, J.X. Yu, and Z. Li (Eds.): ADMA 2014, LNAI 8933, pp. 345–355, 2014. Springer International Publishing Switzerland (2014)
14. Agarwal, Archana M.: Diagnostic approach to Anemia., University of Utah School of Medicine, Dept.of Pathology, ARUP Laboratories (2014)
15. Read, Jesse.: Multi-label Classification. Universidad Carlos III de Madrid. Department of Signal Theory and Communications Madrid, Spain (2013)
16. Newton Spolaor, Everton Alvares Cherman, Maria Carolina Monard, Huei Diana Lee.: A Comparison of Multi-label Feature Selection Methods using the Problem Transformation Approach. Elsevier, Science Direct, Electronic Notes in Theoretical Computer Science 292 (2013) 135-151.
17. Purvi Prajapati, Amit Thakkar, Amit Ganatra, A Survey and Current Research Challenges in Multi-Label Classification Methods. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March (2012).
18. Santos, Araken M., Canuto, Anne M P., Neto Antonino Feitosa.: A Comparative Analysis of Classification Methods to Multi-label Tasks in Different Application Domains. IJCISIM, ISSN 2150-7988 Volume 3 pp. 218-227 (2011)
19. Jesse, Read., Bernhard, Pfahringer., Geoff Holmes., Eibe Frank.: Classifier chains for multi-label classification. Mach Learn (2011) 85:333–359, Springer (2011)
20. Grigorios, Tsoumakas., Ioannis, Katakis., Ioannis, Vlahavas.: Mining Multi-label Data. O. Maimon, L. Rokach (eds.), Data Mining and Knowledge Discovery Handbook, 2nd ed., Springer Science Business Media, LLC (2010)
21. Carvalho, Andre de., Alex Freitas.: A Tutorial on Multi-label Classification Techniques. researchgate, In book: Foundations of Computational Intelligence Volume 5 (2009)
22. Grigorios, Tsoumakas., Ioannis, Katakis.: Multi-Label Classification: An Overview. IGI PUBLISHING. (2007)