

Search for an essential parameter & technique for effective prediction of disease using Machine Learning Technique.

¹Bhavin Patel, ²Prof.Ajay Parikh, ¹ResearchScholar, Gujarat Vidyapith, Ahmedabad, Gujarat, India, ² Prof & Head, Dept. of Computer Sci., Gujarat Vidyapith, Ahmedabad, Gujarat, India.

Abstract:

Anemia is an indicator of reduced red blood cell (RBC) and hemoglobin (HB). It affects nearly 1.62 billion individuals worldwide and 51% of women aged 15-49 are suffering through anemia in India. To predict the different types of anemia at an early stage of human life is important. Much research has been done in the past and still, researchers are doing research on the kind of anemia prediction problem using machine learning an alternate way. This paper is the crux of the study I did on the research done on a similar problem across the world. I did the investigation related to the issue of anemia some of the institutes of higher education in Gujarat. I identified the research gap and local relevance of the anemia prediction issue. Study leads me to the conclusion that there is a scope of improvement in the issue of anemia prediction using machine learning algorithm.

Index Terms:

Anemia, RBC, HB, CBC, human life, Machine Learning, Prediction algorithms, healthcare.

I. INTRODUCTION:

Anemia leads to increased mortality and other complex diseases. In Greek, the meaning of Anemia is “without blood” Based on study one-third of the world's population is suffering from anemia. A program called “Mission Shakti” has been started by the Health Department of Gujarat Government. Considering the role of anemia in the state of Gujarat, it is found that in women 54.9% is 21.7% in males and 62.6% in children. Therefore, the prediagnosis revealed by the thorough investigation of this sign should be supported by laboratory Parameter according to the underlying pathological process.

Complete blood count (CBC) is primary, first and foremost laboratory test suggests by a doctor in an abnormal condition of the people. A complete blood count test indicates many diseases and infections in our body such as autoimmune disease, vitamin and mineral deficiencies, Anemia, Thalassemia, leukemia, etc.

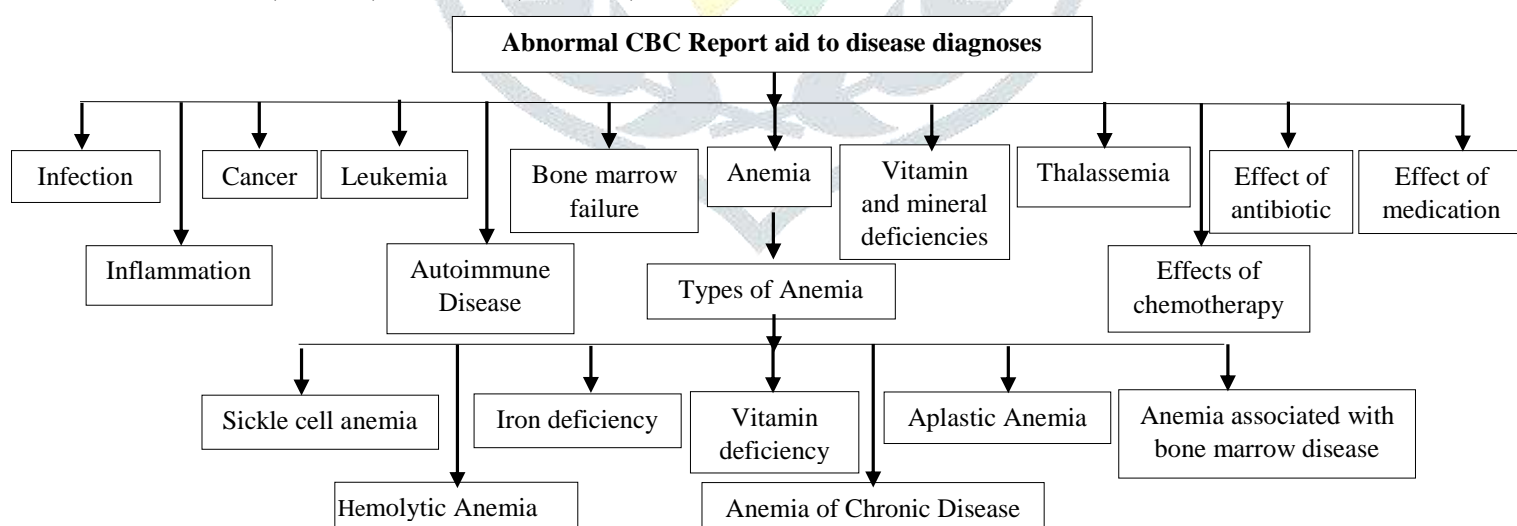


Figure-1 Abnormal CBC Report Diagnose Disease from that specific shows Anemia.

The research paper has been concluded in the last 20 years of research papers. That has been a lot of research on earlier anemia. In which sickle cell anemia, iron deficiency anemia, and Anemia of chronic disease have been observed. But it has not been researched on such type of anemia, aplastic anemia, Anemia associated with bone marrow disease.

If there is a higher education in Gujarat, an issue of anemia is being looked at, as per the year of 2018 about 3000 students of the Gujarat Vidyapith University have been studied and found that out of 1013(33.76%) students, there is anemia. Due to which the dropout ratio is more in the students. On the basis of my studies, needed to develop the Anemia Prediction model.

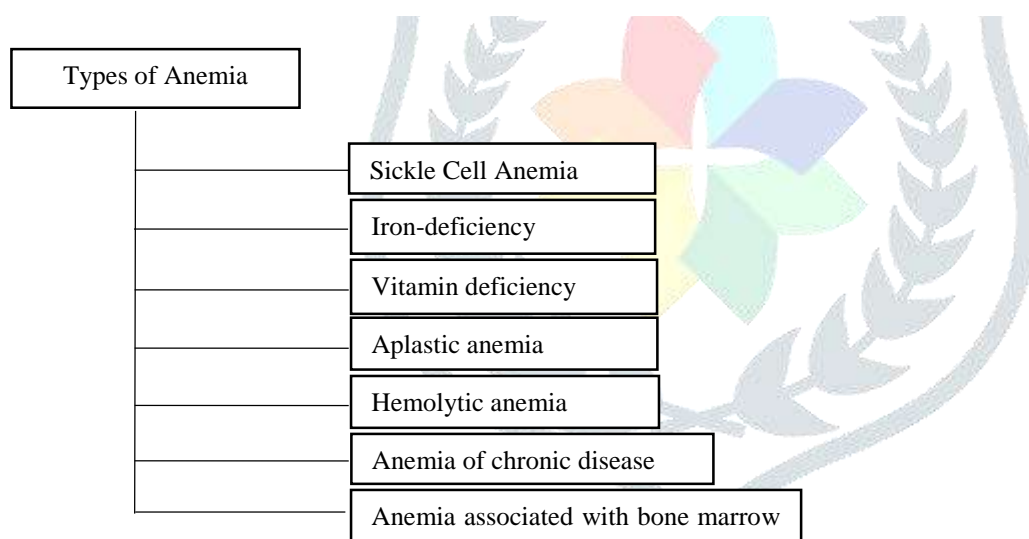
II. RELATED WORKS:

In anemia prediction there are various machine learning algorithms are used to develop 2 models for smart blood analytics (SBA HEM 181, and SBA HEM 61) with various attributes for 43 categories of disease prediction. In this related research paper 10 CBC Parameter (Age, HB, RBC, WBC, Platelet, P.C.V (HCT), MCH, Neutrophils %, Lymphocyte %, and Lymphocyte count) are used for 8233 cases with the help of different algorithms such as SVM, Naive Bayesian Classifier, Decision Tree, random forest etc. From that algorithms, the random forest gives the best accuracy. They have done two clinical test one for six hematological specialists and another one is for eight non-hematological internal medicine specialist. Prediction model gives the best accuracy by Gregor Gunčar1 et al. [1]. There is another related study has been done regarding the prevalence of anemia in indoor patients of PDU Hospital Rajkot. In this study researcher studied 4775 cases with 13 parameters such as (HB, RBC, WBC, Platelet, P.C.V (HCT), M.C.V, MCH, MCHC, Lymphocyte, Polymorphs, Neutrophils, Eosinophils, Monocytes), out of them find out the various types of anemia such as microcytic hypochromic, normochromic normocytic, and macrocytic hyperchromic. After that, they concluded most cases are found microcytic hypochromic anemia such as iron deficiency anemia in the age group 21-40 years Mital J Gamit et al. [6]. In addition, some other researchers find out five types of anemia predictive model using rule-based data mining classification algorithms such as Naive Bayes, Neural Network (Multilayer Perception), J48, SMO through WEKA analysis tool performance are measured with F-measure, accuracy, true positive rate, precision, false positive rate, recall etc. This research paper 41 samples are used with 7 Parameter like (Age, Gender, MCV, P.C.V (HCT), HGB, MCHC, and RDW) by Manal Abdullah, Salma Al-Asmari et al. [7].

III. FINDINGS & PROPOSED WORK:

During this research clinical laboratory blood test report 15 Parameter (Age, Gender, HB, RBC, WBC, Platelet, P.C.V (HCT), M.C.V, MCH, MCHC, RDW, Lymphocyte, Neutrophils, Eosinophils, Monocytes) are using with 10000 sampling. As per our knowledge and study, none of the previous work focused on all these types of anemia prediction using machine learning technique.

Types of anemia:

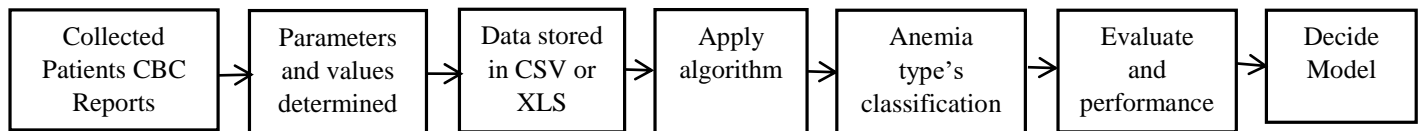


Classification of anemia:

Morphological Classification of anemia :		
Microcytic (MCV < 80 fl)	Normocytic (MCV 80 -100 fl)	Macrocytic (MCV > 100 fl)
Iron deficiency Anemia	Haemolytic (Reticulocyte ↑)	Megaloblastic
Sideroblastic Anemia	Sickle cell anemia	Vitamin B12 Deficiency
	Autoimmune disease (SLE)	Vitamin B9 Deficiency (Folate)
	Non Haemolytic (Reticulocyte ↓)	Non Megaloblastic
	Aplastic anemia	Liver Disease
	Anemia of Chronic Disease	Alcohol use disorder
	Kidney Disease	Metabolism disorder

Parameter of anemia dataset:

Parameter	Adult Normal Range	Category
Age	>12	mild, moderate, severe
Gender	Male/Female	
Hbg (Hemoglobin)	Male-14.0 - 17.4 g/dl Female -12.0 - 16.0 g/dl	
RBC (erythrocyte)	Male - 4.5 - 5.5 × 10 ⁶ /μL Female - 4.0 - 5.0 × 10 ⁶ /μL	
WBC (leukocyte)	4.5 - 5.5 × 10 ³ /μL	
Platelet Count (thrombocyte)	150-450 × 10 ³ /μL	
Blood Indices		
P.C.V (HCT) Hematocrit	37-47 %	Normal
MCV	82-92 fl	
	< 80	Microcytic
	80-100	Normocytic
	>100	Macrocytic
MCH	27-31 pg.(pico.gm)	
MCHC	< 32	Hypochromic
	32-36 %	Normochromic
RDW	11.5-14.0 %	Normal
	>14.0	High
Differential Count		
Neutrophils	40-70 %	
Lymphocytes	20-40 %	
Eosinophils	1-4 %	
Monocytes	2-6 %	



Proposed System Workflow

IV. DATA COLLECTION METHODS, TOOLS AND TECHNIQUES:

The methodology to be adopted in the present study has been described here through the choice of area of research, population and sampling technique, types of data and tools for data collection, a procedure of data analysis and time schedule for the overall research work.

As per the literature review, there are various classification machine learning algorithms used like Logistic Regression, Random Forest, SVM (Support Vector Machine), Naive Bayes, Decision Tree, comparison of all these algorithms along with basic blood test Parameter from those algorithms show a best performance and accuracy of Logistic Regression.

V. CONCLUSION:

As per Literature Review study find out there are health care industry use machine learning techniques which can help for disease prediction and analysis of healthcare data. Machine Learning Approach is better than a traditional approach for prediction of disease. We see, with a possibility of disease developed various prediction model and cost-effective analysis to the patients. Use of various machine learning algorithms such as Logistic Regression, Random Forest, Naive Bayes classifier, Decision Tree, Support Vector Machine, Principal Component Analysis and classification methods and applications for prediction model on a particular disease with a combination of different blood Parameter. From that logistical regression algorithm gives best accuracy based on literature review. It helps to patients for alerting and risk management decision support tools, targeted at improving patients' safety and health care quality. With the need to reduce health care costs and the movement towards personalized health care. Machine learning technology helpful in health care service that can make better decisions and prediction of disease diagnosis and provide supportive treatment options to doctor's, which leads to the overall improvement of health care services. A predictive model is very helpful to practitioner, physicians or doctors to predict a possibility of disease development.

Words: Hemoglobin (American) abbreviated **Hb** or **Hgb**.

REFERENCES:

- 1). Gregor Gunčar¹, Matjaž Kukar¹, Mateja Notar¹, Miran Brvar², Peter Černelč³, Manca Notar¹ & Marko Notar¹, (2017). An application of machine learning to hematological diagnosis, *Scientific Reports* | (2018) 8:411 | DOI:10.1038/s41598-017-18564-8.
- 2). Anand Borad., (2018). Disease Prediction, Machine Learning, and Healthcare – DZone AI.
- 3). Ebru Dünder Yenilmez and Abdullah Tuli., (2018). Laboratory Approach to Anemia, DOI: 10.5772/intechopen.70359, ISBN 978-953-51-3775-7, Print ISBN978-953-51-37740, <https://www.intechopen.com/books/current-topics-in-anemia>
- 4). Anand Borad., (2017). Healthcare and Machine Learning: The Future with Possibility. <https://www.einfochips.com/blog/healthcare-and-machine-learning-the-future-with-possibilities/>
- 5). B.Dhomse Kanchan, IT department, MET'S BKC IOE, Nasik, India. M. Mahale Kishor, IT department, MET'S BKC IOE, Nasik, India. (2017). Study of machine learning algorithms for special disease prediction using principal of component analysis. *IEEE Xplore*: 26 June 2017. <https://ieeexplore.ieee.org/document/7955260>
- 6). Mital J Gamit, Hemlata S Talwelkar, (2017). Survey of different type of anemia. *Indian Journal of Medical Science and Public Health*, 2017; Vol 6, Issue 3, Pages 493–496.
- 7). Manal Abdullah, Salma Al-Asmari, (2016). Anemia type's prediction based on data mining classification algorithms. ISBN 978-1-138-02972-9. Research Gate.