Predictive Modelling of Three Subtypes of Leukemia using Complete Blood Count Reports: A Case Study of Pakistan

By

IQRA TAHREEM

Master of Science in Bioinformatics

Fall 2018-MS BI-3-00000277534

Supervised by:

Dr. Zamir Hussain

Research Centre for Modelling and Simulation (RCMS)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan.

April 2021

Dedication

I dedicate this dissertation to my beloved parents, husband and brothers.

Certificate of Originality

I hereby declare that the results presented in this research work titled as "Predictive Modelling of Three Subtypes of Leukemia using Complete Blood Count Reports: A Case Study of Pakistan" are generated by myself. Moreover, none of its contents are plagiarized nor set forth for any kind of evaluation or higher education purposes. I have acknowledged/referenced all the literary content used for support in this research work.

IQRA TAHREEM

(Fall 2018-MS BI-3- 00000277534)

Acknowledgment

I would first like to thank ALLAH Almighty for blessing me with health and strength to accomplish this project. Secondly, I want to express my deepest gratitude to my respected supervisor and Co-supervisor, Dr. Zamir Hussain and Dr. Mehak Rafiq for their consistent support, guidance, and immense knowledge throughout this project. This work would not have been possible without their motivation, enthusiasm, and tolerance. Undoubtedly, their commitment to the research work will always have been an inspiration for me.

I am indebted to National University of Sciences and Technology (NUST) and Research Centre for Modelling and Simulation (RCMS) for providence of research lab and quality environment. I whole-heartedly appreciate the support of Principal Dr. Muizuddin Shami and my Guidance Examination Committee (GEC): Dr. Ishrat Jabeen, Dr. Rehan Zafar Paracha, and Dr. Yasmin Badshah. I would like to pay my special regards to the research group specially to my fellows Azka Iqbal, Ayesha Shabbir, and Hira Qureshi for their kind support.

Last but not the least, I would like to thank my parents and my brothers for supporting me spiritually throughout my life and in the pursuit of this project.

Contents

List of Abbreviations

AML	Acute Myeloid Leukemia
CML	Chronic Myeloid Leukemia
ALL	Acute Lymphocytic Leukemia
CLL	Chronic Lymphocytic Leukemia
CP	Chronic Phase
AP	Accelerated Phase
BC	Blast Crisis
WHO	World Health Organization
CBC	Complete Blood Count
MRD	Minimal Residual Disease
FISH	Fluorescence in situ hybridization
PCR	Polymerase Chain Reaction
CSF	Cerebrospinal Fluid
ANN	Artificial Neural Network
SVM	Support Vector Machine
PCA	Principal Component Analysis
PPCA	Probabilistic Principal Component Analysis
LDH	Lactate Dehydrogenase
GMDH	Group Method of Data Handling
ESR	Erythrocyte Sedimentation Rate
CPD	Cell Population Data
DT	Decision Tree
RF	Random Forest
SGD	Stochastic Gradient Descent
ANOVA	Analysis of Variance
IDA	Iron Deficiency Anemia
PIMS	Pakistan Institute of Medical Sciences.

ASAB	Atta Ur Rahman School of Applied Biosciences
KRL	Khan Research Laboratories
WBC	White Blood Cell
RBC	Red Blood Cell
Hb	Haemoglobin
HCT/PCV	Haematocrit
MCV	Mean Corpuscular Volume
MCH	Mean Corpuscular Haemoglobin
MCHC	Mean Corpuscular Haemoglobin Concentration
PLT	Platelet Count
ANC	Absolute Neutrophil Count
LYM	Lymphocyte Count
BASO	Basophil Count
EO	Eosinophil Count
MO	Monocyte Count
EM	Expected Maximization
SPSS	Statistical Software for Social Sciences
SD	Standard Deviation
MLR	Multinomial Logistic Regression
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
CC	Correlation Coefficient
OR	Odds Ratio
LRT	Likelihood Ratio Test

List of Figures

List of Tables

Abstract

Leukemia is a malignancy of white blood cells (WBC's) arises from hematopoietic stem cells. A common, essential, initial, and normal examination test which may indicate the presence of leukemia and its subtypes is Complete Blood Count (CBC). A CBC report provides useful information of different characteristics of blood cells that can be used for differential diagnosis. This study is designed to analysis different characteristics of CBC reports to develop predictive models for the screening of suspected patients of leukemia and its subtypes. In this study, primary data set of 302 CBC reports is collected from eight different hospitals of Rawalpindi and Islamabad regions. Out of these 302 CBC reports 67 are normal (non-leukemic), 123 are Acute Myeloid Leukemia (AML), 79 are Chronic Myeloid Leukemia (CML) and 18 are Acute Lymphocytic Leukemia (ALL). A CBC report usually consists of 21 different characteristics/variables of blood picture of a person. Out of these 21 variables, 15 variables are selected for the analysis by dropping information of percentages of various variables to avoid duplication. Comparative analysis has been used to validate statistically significant differences between the numerical estimates of means with respect to four categories of all selected variables. The results show that Mean Corpuscular Haemoglobin (MCH) is the only variable having statistically insignificant difference between the means of normal, AML, CML and ALL. To check the existence of linear relationship between variables, correlation analysis is performed. This analysis also helps in the identification of multicollinearity problem for the development of logistic regression models. For the development of Multinomial Logistic Regression (MLR) model, five different combinations of methods for inclusion of relevant variables in the model or exclusion of irrelevant variables from the model. These are backward elimination method using Wald's criteria, selection of variables using odds ratios (OR), selection of variables from combination of dropping insignificant variables simultaneously and Wald's test, selection of variables from combination of dropping insignificant variables simultaneously and OR and selection of variables from combination of Wald test and OR. Final selection of any variable is done based on the criteria that it is successfully shortlisted in at least three methods of selection.

Therefore, four variables have been identified namely haemoglobin, neutrophil count, monocyte count and gender being appropriate variables for development of multinomial logistic regression model. The performance of the developed model is checked through different measures like accuracy, sensitivity, specificity, and precision. The results show that in case of Normal vs AML the accuracy is 86 %, sensitivity is 86%, specificity is 85% and precision is 91%. For Normal vs CML, accuracy is 88%, sensitivity is 91%, specificity is 85% and precision is 87%. For Normal vs ALL, accuracy is 88%, sensitivity is 100%, specificity is 85% and precision is 64%. These results show that the developed models can be used with confidence for the subjective screening of disease, i.e leukemia or its subtypes. A notable point is that the proposed model is not intended to be used as replacement of the formal diagnostic tests of leukemia like bone marrow biopsy, flow cytometry, etc. It facilitates basic technical support for screening of patients using data driven models. Therefore, a combination of subjective and objective assessment can improve the quality of diagnosis of leukemia or its subtypes at early stages.

INTRODUCTION

Leukemia is a malignancy of white blood cells (WBC's) arises from hematopoietic stem cells, where the normal cell divisions and proliferations are deregulated by the genetic mutations. The affected leukemic cells when damaged, do not go through normal cell apoptosis, thus accumulating and overcrowding the normal blood cells [1]. Due to wide range of WBC's in the human body, leukemia is totally different from other cancers in the range of cases. Any person in any age can be affected to it. Leukemia is not considered "metastatic", because it does not form tumours, however it forms dangerous accumulations in the brain, spleen and lymph nodes [2]. Few details of the classification of leukemia are:

1.1 Subtypes of Leukemia:

Classification of leukemia is usually based on clinical behaviour (acute leukemia or chronic leukemia) and the affected hematopoietic stem cells (myeloid leukemia or lymphoid leukemia). Figure 1.1 shows the details of subtypes of leukemia. The occurrence, medical appearance, and survival, etc. is different with respect to subtypes. The four primary diagnostic types and their brief descriptions are mentioned below [3], [4]:

1. Acute lymphocytic leukemia (ALL)
2. Acute myeloid leukemia (AML)
3. Chronic lymphocytic leukemia (CLL)
4. Chronic myeloid leukemia (CML)

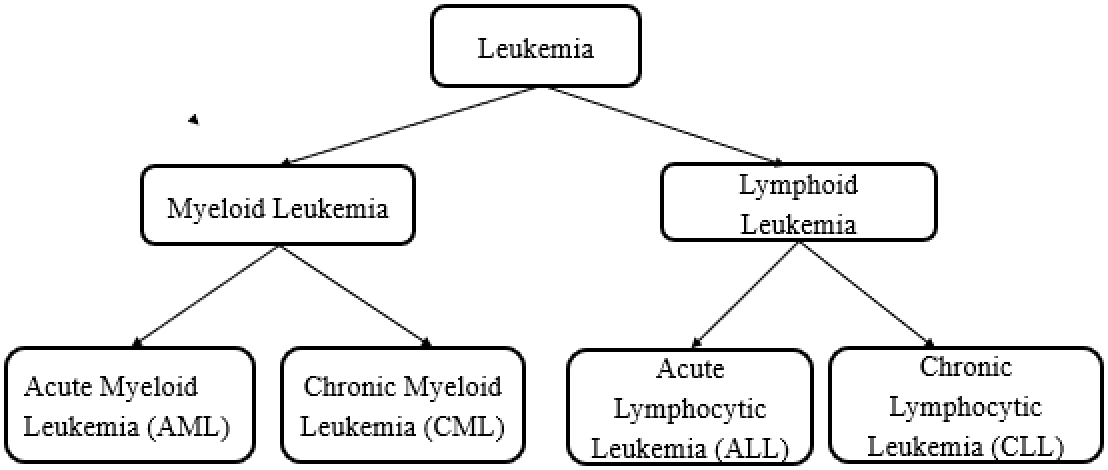

Leukemia
Myeloid Leukemia
Lymphoid Leukemia
Acute Myeloid Leukemia (AML)
Chronic Myeloid Leukemia (CML)
Acute Lymphocytic Leukemia (ALL)
Chronic Lymphocytic Leukemia (CLL)

Figure 1.1: Details of Subtypes of Leukemia

1.1.1 Acute Myeloid Leukemia:

The rise in the number of myeloid cells in the bone marrow and halt in their growth causes AML which normally results in the insufficiency of hematopoietic cells and leads toward anaemia, granulocytopenia or thrombocytopenia with or without leukocytosis [5]. AML is common in adults and the median age of identification is around 65 years or older [6] , [7], [8].

Clinical Symptoms of AML:

The clinical symptoms of AML include bone and joint pain. Moreover, about 50% of patients are observed with large spleen [9].

Diagnostic Symptoms of AML:

Standard adult body have 4,000 to 10,000 WBC’s per microliter but the patient suffering from AML has greater or lower number of WBC’s along with the abnormal increment in the myeloid cells (granulocytes and monocytes) [10], [11]. Studies also showed that in case of AML, WBC’s are increased from the normal range of 10,000 per microliter.

1.1.2 Chronic Myeloid Leukemia:

Chronic myeloid leukemia accounts for 15% of leukemia cases and it is a rare cancer [12]. It is a malicious hematopoietic stem cells disorder that results not only in the increment of myeloid cells but also platelets and erythroid cells in the cellular components of blood and marked myeloid hyperplasia in the bone marrow [13]. CML is usually detected between the age of 35-45 years [9]. The male to female ratio is usually 1.2 to 1.7 [14].

CML development is divided into three phases: chronic phase (CP), accelerated phase (AP) and blast crisis (BC) [12]. The staging of disease depends on the ratio of immature blast cells in the blood and in bone marrow. Majority of the CML cases are detected in chronic phase (CP) [15].

In CML-CP phase patient have few or no symptoms of the disease and it can be controlled successfully with ordinary treatment because in this case less than 10% of blast is present in the blood [16]. Patients when move from CP to AP phase of CML have 10-19% of blasts, and there occurs a decline in platelets and red blood cells, variations in WBC's, an increment in blast cells, and inflammation of the spleen [17]. World Health Organization (WHO) defines that CML-blast phase consists of patients having at-least 20% blasts, while the BC phase is different from the AP in that 30% or more blast cells originate in the blood cells or bone marrow. This causes swelling of the liver along with the symptoms of earlier phases [17]. The BC phase is usually lethal [18].

Clinical Symptoms of CML:

Clinical symptoms of CML are fatigue, weight loss, liver , bleeding due to the dysfunction of platelets and spleen enlargement [9].

Diagnostic Symptoms of CML:

In CML the amount of WBC's surpasses 250,000 per microliter [19] and the amount of platelets are usually decreased, normal or increased from 150,000-450,000 per microliter

[20]. This research also shows that most of the CML cases has platelet counts less than the normal range.

1.1.3 Acute Lymphocytic Leukemia:

ALL is known as childhood leukemia. It is consisting of 80% of overall cases [21]. The age statistics of ALL in research data also shows that it is a childhood leukemia. ALL is characterized by the uncontrollable and irregular production of lymphoid precursor cells known as lymphoblasts in the bone marrow with blocked development [22]. In Pakistan the median age of ALL diagnosis is 6 years [23], [24].

Clinical Symptoms of ALL:

The clinical symptoms of ALL are fatigue, fever, vomiting, pale skin and loss of appetite [9].

Diagnostic Symptoms of ALL:

Patients with ALL have WBC's greater than 10,000 per microliter to 50,000 per microliter with 30% lymphoblast in the bone marrow and platelets are less than 150,000 per microliter [25], [26].

1.1.4 Chronic Lymphocytic Leukemia:

Clonal proliferation and accumulation of B lymphocytes in the bone marrow and lymphoid tissues leads toward CLL. It is also linked with cellular and humoral immune response [27]. The usual incidence age of CLL is 60-80 years [28].

Clinical Symptoms of CLL:

The symptoms of CLL are fatigue, shortness of breath, gums and nose bleeding [9].

Diagnostic Symptoms of CLL:

The normal range of lymphocytes is 1000-4800 per microliter but detection of CLL needs the existence of at-least 5000 per microliter B lymphocytes in the peripheral blood [29].

1.2 Risk Factors of Subtypes of Leukemia:

Leukemia is highly associated with bulky doses of different chemicals such as benzene which is used in the manufacturing of paints and plastics. Its occupational and environmental exposure is a well-known aspect of leukemia in adults , especially AML [30], [31]. Exposure to radiation, contaminations with particular viruses (e.g., human lymphotropic virus, Epstein-Barr virus, etc), contact to electromagnetic fields and cigarette smoking are also the major causes of leukemia [32]. Exposure to household pesticides in utero before birth and in the initial three years of lifespan has been related with high chance of childhood ALL [31]. Later in life, hematopoietic stem cells malignancy is also a reason for development of different subtypes of leukemia [33].

1.3 Incidence of Leukemia Subtypes across the World:

Leukemia contributes 30% of childhood cancers [34]. It accounts for some 300,000 new cases every year (2.8% of all new cancer cases) and 222,000 fatalities. The high death rate (74%) mirrors late or miss diagnosis of leukemia in many regions of the world, where the facilities of treatment are not accessible [35].

In Western countries it has been estimated that the most frequent type of leukemia is CLL with almost 30% of all cases [36]. CML characterizes 20% of cases [37] while AML represents approximately 25% of the cases [38].

1.4 Incidence of Leukemia Subtypes in Pakistan:

In Pakistan, the incidence of AML is around 12% under the age of 10 years, 28% between ages 10-15 years and 80-90% in adults while ALL is a childhood leukemia [39]. CLL is the least common and accounts for about 5% of all leukemia cases. However, the chances of having CML are thrice relative to CLL [40], [4].

1.5 Subjective Screening of Leukemia:

For a preventive measure against Leukemia, specialists carried out various screening tests to examine possible health condition or illness in someone who does not yet have signs or symptoms. Early detection helps to minimize the risk of infection and maximizes the chance of effective treatment. Screening tests are simple and cheap. These include physical examination, health history and Complete Blood Count (CBC) report of a person.

1.5.1 Physical examination and health history:

Health history of a person examined by the doctor indicates the signs, risk factors and all the medical conditions the person had experienced in the past. The specialist taking a health history, will ask questions about an individual's history of: symptoms that recommend leukemia, high radiation contact, hereditary disorders, such as Down syndrome, Fanconi anaemia or Bloom syndrome, chemicals exposures, former chemotherapy of blood diseases and viral contaminations [41].

1.5.2 Complete blood count (CBC) Report:

The essential, initial, and normal examination test which may indicate this disorder is complete blood count (CBC). A CBC calculates the quantity and condition of white blood cells (WBC), red blood cells (RBC), haemoglobin (Hb), haematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC) and platelets (PLT) present in the blood. CBC test also gives information about different types of WBC's which are neutrophils, lymphocytes, monocytes, eosinophils, and basophils. Leukemia and other infections may cause an excessive number of blood cells. Immature blood cells also known as blast or leukemic cells are usually not grasped in the blood, and specialists will presume leukemia if irregular blood cells occur. CBC deviations are essential laboratory findings in the diagnosis of subtypes of leukemia, and it is difficult to detect leukemia patients without CBC aberrations [42], [43]. Table 1.1 shows the details of a usual CBC report in Pakistan with their reference ranges.

In Table 1.1 units are abbreviated as:

Litre = L

Grams per decilitre = g/dL

Femtolitre = f/L

Picograms = Pg

Microlitre = u/L

Table 1.1: Details of a usual CBC report [44].

Sr. No.	Blood Components	Reference Ranges	Unit
1	Age	-	-
2	Gender	-	-
3	White Blood Cells	4 -10	×10^9/L
4	Red Blood Cells	3.8 - 4.8	×10^12/L
5	Haemoglobin	12.5 - 14.5	g/dL
6	Haematocrit		%
7	Mean Corpuscular Volume	80 - 95	f/L
8	Mean Corpuscular Haemoglobin	27 - 32	Pg
9	Mean Corpuscular Haemoglobin Concentration	31.5 - 34.5	g/dL
10	Platelet Count	150 - 400	×10^3/L
11	Neutrophil Counts	2 - 7	×10^3/L
12	Lymphocyte Counts	1 -3	u/L
13	Basophil Counts	0.02 - 0.1	u/L
14	Eosinophil Counts	0.02 - 0.5	u/L
15	Monocyte Counts	0.2 - 1	u/L
16	Neutrophil Percentage	40% - 80%	%
17	Lymphocyte Percentage	20% - 40%	%
18	Basophil Percentage	0.5% - 1%	%
19	Eosinophil Percentage	1% - 6%	%
20	Monocyte Percentage	2% - 10%	%
21	Reticulocyte Percentage	0.5% - 1.5%	%

1.6 Diagnostic Tests for Leukemia:

Multiple tests are carried out by the specialists for the diagnosis of Leukemia and its subtypes. These tests include but are not limited to blood chemistry tests, cytochemistry, immunophenotyping, flow cytometry, cytogenetic, molecular studies, lumbar punctures and bone marrow biopsies [4]. Few details of these tests are provided below.

1.6.1 Blood Chemistry Test:

Measurement of certain chemicals in the blood is done by blood chemistry test. This test helps the specialists to find the abnormalities occur in liver and kidney due the spread of leukemic infectious cells [45].

1.6.2 Cytochemistry:

Cytochemistry utilizes stains or dyes to detect components and structures of tissues in blood or bone marrow cells. Specific microscopic stains are attracted to specific substances present in some sorts of leukemia blasts. Microscope is used to see the staining results. Cytochemistry aids doctors to identify the type of cells that are present [46].

1.6.3 Immunophenotyping:

Immunophenotyping proteins identification in tissues or cells is done by a very specific antigen-antibody reaction. Monoclonal antibodies are marked with specific fluorescent or enzyme label that binds only to specific antigens (proteins). This allows doctors to see the blast cells [47].

1.6.4 Flow Cytometry:

Flow cytometry is used in sorting and classification of cells by the help of fluorescent labels their surface. It allows doctors to view many antibodies at the same time and collect data rapidly from thousands of cells in a single sample and helps to describe unique characteristics of blasts. These features can help specialists in treatment of leukemia using minimal residual disease (MRD) [48].

1.6.5 Cytogenetics:

Cytogenetics is the examination of chromosomal cells, including their number, size, shape and arrangement. Some main chromosomal aberrations of the cells can be observed under microscope. But to observe DNA changes a deeper analysis is done by fluorescence in situ hybridization (FISH) and polymerase chain reaction (PCR). FISH is used to find the genetic aberrations in the leukemic blast cells. PCR is used to make multiple copies of a specific gene segment and then tested in the laboratory. DNA mutations, inversions or deletions that are associated with different types of leukemia is find by PCR. Different subtypes of leukemia are diagnosed by PCR [49], [50].

1.6.6 Bone Marrow Biopsy:

In this process, cells are detached from the bone marrow and tested in laboratory. The report obtained from the lab will confirm the presence or absence of leukemic cells in the sample. A positive report can be helpful in identification of the subtype of leukemia [51].

1.6.7 Lumber Punctures:

In lumber puncture process, a small amount of cerebrospinal fluid (CSF) from the space around the spine is removed and observed under a microscope. The process is done to see if malignancy has spread to the spinal fluid [52].

These diagnostic tests are painful, time consuming and highly expensive such as the sample collection procedure of bone marrow biopsy procedure takes 10-20 minutes and its report duration is two to three weeks [53]. This is a highly painful procedure as a person feels pain for about a week [53]. To overcome this pain doctor may recommend medicines such as ibuprofen [53]. After bone marrow biopsy a person may experience extreme bleeding and fever [54]. The cost of bone marrow biopsy is around 6000-8000 rupees [54]. In a developing country like Pakistan, people cannot afford the price of these tests and in such cases, this disease remains undiagnosed. As compared to these expensive diagnostic tests CBC test is the simplest and cheap test as its cost is about 650-700 rupees. CBC test takes just a few minutes and it may take a few hours to a day for the results to be available.

The aim of a diagnostic test is to assess the presence (or absence) of the disease in symptomatic or screen-positive individuals as a basis for treatment decisions (confirmatory test). The factors of time, money and painful procedures are common causes of late or no diagnosis of Leukemia because of affordability, etc. Moreover, the subjectivity factor in the examination of CBC reports may produce false positive results. Therefore, this study is designed to provide data driven models for the detection of Leukemia and its subtypes using all or significant characteristics of a CBC report.

1.7 Problem Statement:

Screening of leukemia is usually practiced through subjective assessment of variations in different characteristics of a CBC report. Hence, assessment varies from practitioner to practitioner and there is a high chance of miss / no diagnosis.

Proposed Solution:

Development of the objective data driven models using Multinomial Logistic Regression to support subjective assessment of a physician. Hence, this support will help in improving accuracy and reliability in terms of prediction of leukemia and its subtypes.

1.8 Objectives:

The main objectives of this study are:

- Analyses of general trends and tendencies of various characteristics of CBC by comparing Leukemic subtypes cases and non-Leukemic (normal) cases.
- Development of a predictive model based on significant characteristics of CBC reports for the screening of Leukemic subtypes cases or non-Leukemic cases.

LITERATURE REVIEW

2.1 Background of Study:

Statistical analysis enables a researcher to draw meaningful conclusions from a study in which data are collected through observation, survey, or experimentation. The success of a medical study however depends to a great extent, on adequate statistical analysis of the data originating from such a study [55]. Prediction models using logistic regression can help healthcare professionals in making clinical decision to diagnose and predict the outcome [56].

2.1.1 Image Based Analysis:

Leukemia develops in the bone marrow and greatly affects the making of proper blood cells. Hence, its early diagnosis is very important for human living. Various studies have focused on the detection of leukemia and its subtypes from the microscopic images, as the analysis and segmentation of images are very important to find the abnormalities present in the blood cells. This section uses image analysis techniques for the development of machine learning models.

The study of Abedy *et al*, 2019 used computational methods to detect ALL by analysing blood cells and its components automatically from microscopic images. This analysis involved classification of cells and blast counting. Publicly available ALL-IDB dataset was used to predict leukemia from microscopic images of human blood cells. To detect the exact shape of lymphocytes Canny edge detector and noise reduction operators were used. When the exact shapes were detected, Principal Component Analysis (PCA) was applied on them which reduces the dimensions of data without losing any important information and also reduced the computational cost. After dimension reduction, classification was done by logistic regression. The validation of results was done by using n-fold cross-validation method. The accuracy of the obtained model was 96% [57].

Bhattacharjee *et al*, 2012 designed an automatic method to detect the blast cells of ALL and AML from human microscopic blood cell images. 40 images of ALL and 40 images of AML were used in this study. The constructed method was consisting of four steps that is pre-processing, de-noising, enhancement section, threshold selection and segmentation of the cells through microscopic cells images. The noise reduction was done by Principle Component Analysis (PCA) which uses an orthogonal transformation for the complete de-correlation of centralized matrix. Colour space conversion and morphological filtering based on pixel intensities was performed in contrast enhancement step. Segmentation of blast cells based on threshold value obtained from Edge sensitive Variational Thresholding technique. For counting the number of existing blast cells in the images Connected Component Analysis technique was used. The evaluation depend on comparison of number of blast cells perceived by manual count and those found by the selective thresholding based automated method [58].

Markiewicz *et al*, 2005 performed a study based on system that identify the AML blast cells. The recognition process was based on bone marrow aspirate image. The database used in this process was consist of 17 different classes of blood cells in which 16 classes belonged to different abnormal types such as basophilic erythroblast, neutrophilic myelocyte, neutrophilic metamyelocyte, neutrophilic band, segmented neutrophils, polychromatic erythroblast, ortochromatic erythroblast, mesoblast, promyelocyte, proerythroblast, segmented eosinophils, prolymphocyte, lymphocyte, plasmocyte, promegaloblast and erythropoesis while the 17th class was consist of the cells deprived of nucleus etc., and was denoted as heterogenic class. Support Vector Machine (SVM) was used as the classifier to recognize the AML blast cells and exploits the features of the image of the blood cells linked to the texture, geometry, histograms and statistical features which were mean, variance, skewness and kurtosis of the image of the whole cell [59].

Shafique *et al*, 2018 designed a computer-aided diagnostic technique to detect ALL diagnosis. The diagnosis technique was based on four steps which were pre-processing, segmentation, feature extraction, and classification. In the pre-processing step the quality of image was enhanced by removing the noise for proper segmentation and classification, this process was done by linear contrast stretching technique. Segmentation of white blood cells was done through K-means clustering, which is a semi supervised learning technique that is used when the data is not labelled. Different feature selection techniques were used in this study such as PCA technique was used to reduce the features to avoid any redundancy. Genetic Algorithm was also used to select important features. PPCA (Probabilistic Principal Component Analysis) technique also gave better performance for features reduction. Classification was done by SVM which efficiently classify the normal and blast cells [60].

Several studies are available with reference to predictive modelling for the detection of Leukemia Subtypes by using microscopic images however they have not used numerical dataset based on CBC reports.

2.1.2 Complete Blood Count (CBC) Based Analysis:

CBC is the simplest and the primary blood test used to detect different blood diseases. There are few published studies using CBC test for a laboratory detection of leukemia and its subtypes.

Fathi *et al,* 2020 performed a study to investigate the use of neuro-fizzy for the detection of acute leukemia in children based on complete blood count test. The data was collected from Tehran Children’s Medical Centre, Iran. The data was consisting of 346 samples in which 172 were ALL and 74 were AML. In the collected data 110 were normal while 243 were patients. The important features included in the study were haemoglobin (Hb), red blood cells (RBC), white blood Cells (WBC), platelets (Plt), mean corpuscular volume (MCV) (the average volume of red cells), mean corpuscular haemoglobin (MCH), lactate dehydrogenase (LDH), erythrocyte sedimentation rate (ESR) and Uric acid. Their study used Principal Component Analysis (PCA), neuro-fizzy and Group method of data handling (GMDH) for the detection of children with Acute Myeloid Leukemia and Acute Lymphocytic Leukemia disease [61].

Syed-Abdul *et al*, 2020 performed a study in Keokuk University Medical Centre (KUMC), South Korea for screening haematological malignancies using Cell Population Data (CPD). The data of 882 was collected in which 457 with hematologic malignancy and 425 with hematologic non-malignancy were used for the assessment. The total data was collected from February 2019 to March 2019. In their study seven machine learning models were used. These models were Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Random Forests (RF), Decision Tree (DT), Linear Regression model, Logistic Regression and Artificial Neural Networks (ANN). For the performance evaluation of machine learning models, the stratified 10-fold cross-validation was used. Their result showed that high ratio of malignancy was found in males with 277 cases as compared to females with 180 cases. Myeloid leukemia had the highest percentage (20.07%) with 177 cases, in which 167 cases were belonged to Acute Myeloid leukemia. The diagnostic ability of ANN was best among all the machine learning algorithms. ANN classifier achieved the highest accuracy of 98.7% [62].

Rathee *et al*, 2014 performed a study to find out the geographic pattern of leukemia subtypes all over Haryana state of India. The study was consisting of 650 blood samples of leukemia patients investigated during 2008-2015 in Haryana. Standard laboratory procedures were used to find blast cell percentage, indices of red blood cell and white blood cell, platelets count and the quantity of haemoglobin. Leishman stain was used to find out the morphology of blast cells in the blood sample of all blood cancer patients. 20% blast criteria were used to detect leukemia and then ‘Sudan Black B’ was used to differentiate AML and ALL. Analysis of Variance (ANOVA) was used to find the interaction of factors (such as age/gender/subtype) affecting leukemia patients. Data on leukemia patients was examined and then subjected to ANOVA. The major outcome of the study were 33.8% patients were affected with AML, 39% patients with CML, 17.2% patients with ALL and 10% with CLL. There were 71.4% and 62.6% male patients affected with chronic and acute leukemia while 28.6% female patients were affected with acute leukemia and 37.4% female patients were affected with chronic leukemia. Among four major type of leukemia, 58% male patients and 42% female patients were observed with ALL, 65% male patients and 35% female patients were detected with AML, 69% male patients and 31% female patients were diagnosed with CML and CLL was observed in 80% male patients and 20% female patients. The male to female ratio in the study was 2:1 [63].

Moussavi *et al*, 2014 performed a descriptive study in Shohada Tajrish Hospital, Iran. Their study included 97 cases included one-month old to fourteen-year-old children of Acute Lymphocytic Leukemia. CBC reports were used to detect ALL. CBC abnormal findings such as blast counts, neutropenia, leucocytosis, thrombocytopenia, and pancytopenia were gathered. The collected data was analysed by SPSS software. Their study showed that large number of WBC in patients was due to the increased number of lymphocytes in blood [64].

Munir *et al*, 2019 performed a descriptive study in Khyber Teaching Hospital Peshawar, from January 2015 to July 2017 with the total cases of 117. Their study included the cases of Chronic and Acute Leukemia's by Nonprobability purposive sampling technique. 8 Patients were those whose aspirates were insufficient, and they were excluded from the study. Remaining 109 cases were included in the study and complete blood counts on these cases were done by Sysmex analyser. CBC findings were recorded, and results were drawn. Mean and standard deviation were used for quantitative data which, while frequency and percentages were used for qualitative data. In their study 61 cases were males and 48 cases were females. Male to female ratio was 1.27 :1. Mean age of sample study was 49 ± 19 years. Changes in blood counts were increased TLC (Total Leukocyte Count) in 52% cases of ALL, 66.6% cases of AML, 87.5% cases of CML, and 66.6% cases of CLL. The low haemoglobin level was observed in 82% cases of ALL, 97.4% cases of AML, 87.5% cases of CML, and 100% cases of CLL. The low platelets count was observed in 88% ALL, 92.3% cases of AML, and 58% cases CLL, but high in CML as it was consisting 62.5% cases. The outcome of their study was that Anaemia, high white blood cell count and thrombocytopenia were observed in all leukemia's, except chronic myeloid leukemia where platelet count was high than the normal range [6].

Naeem *et al*, 2017 conducted a study in Pathology Department of King Edward Medical University, Lahore. For this purpose, CBC was performed on 77 cases of Acute Myeloid Leukemia. CBC was done by automated blood cell counters. The CBC data was assembled and analysed by SPSS software. The purpose of their study was to find the demographic and clinical features of various subtypes of acute leukemia. Descriptive statistics was done on the blood counts and the mean of Haemoglobin, Platelets and TLC were calculated. Their study also showed the male predominance with male to female ratio of 1.5:1 [65].

Khan *et al,* 2016 performed descriptive a study to calculate the frequency of subtypes of Leukemia. For this purpose, the CBC data of 200 patients were collected from Ayyub Teaching Hospital, Abbottabad. Mean and Standard deviation were used for quantitative variables while frequency and percentages were used to explain categorical variables. Their study showed that the occurrence of acute leukemia was higher than chronic leukemia. In their study 16% of patients had acute myeloid leukemia and 32% patients were with acute lymphocytic leukemia. On the other hand 11% patients had chronic myeloid leukemia and only 3% had chronic lymphocytic leukemia [66].

Farzana *et al,* 2016 performed a descriptive study to examine the haematological parameters in acute myeloid leukemia patients. The data of 107 patients were collected from National Institute of Bone Diseases, Karachi. The parameters examined from the CBC were Haemoglobin, Total Leucocyte Count, Platelet count and Blast count. Majority of the patients had less percentage of Haemoglobin and greater number of WBC's. In their study male to female ratio was 1.4:1 [67].

2.1.3 Data Mining Techniques:

This section provides literature using CBC reports. As in medical science, data mining techniques have been used CBC tests to diagnose different blood diseases such as anaemia and thalassemia.

Alshami *et al,* 2012 investigated the existence of thalassemia and its subtypes by the help of data mining classifiers. The dataset used in the study was consist of 46920 samples. The study was depending on CBC having feature such as age, gender, red blood cells, haemoglobin and platelets. Three data mining classifiers used in this investigation were Decision tree, Naïve Bayes and Artificial Neural Network (ANN). These classifiers were used to differentiate between thalassemia traits patients- with its different levels-: the patient who suffer from other blood diseases, iron deficiency patients and normal persons. The results showed that ANN classifier was the most significant classifier to differentiate between the subtypes of thalassemia and other blood diseases [68].

Abdullah *et al*, 2017 performed the study on anaemia which is one of the most common blood diseases. This study investigated the five most common types of anaemia. The dataset consists of the CBC test results of the patients. The undesirable variables were eliminated, and the filtered data was then implemented on different classification algorithms such as Naïve Bayes, Multilayer Perceptron, J48 and SMO using WEKA data-mining tool. From Numerous experiments it was proved that J48 decision tree algorithm gave the best possible classification of anaemia subtypes. J48 decision tree algorithm gave the best results with accuracy, precision, recall, True Positive rate, False Positive rate and F-measure [69].

Hasani *et al*, 2017 illustrated the detection of three types of anaemia namely iron deficiency anaemia (IDA), β-thalassemia trait and α-thalassemia trait (cis and trans). The detection of these three types were difficult because of their nature and homogeneity in characteristics. The research was done to provide a model to correctly diagnose anaemia types. To this end, the simple CBC test was used to identify and differentiate between these forms of anaemia in Weka software instead of some other tests. For this purpose, five classification algorithms and a vote algorithm (hybrid algorithm) were used to obtain the highest accuracy and the minimum mean absolute error. The performance of those five algorithms were compared with the performance of vote algorithm. The results of this study indicated that vote algorithm increases the diagnosis accuracy and decreases error rate in comparison with the single classifiers [43].

2.1.4 Prevalence of Leukemia Subtypes:

Pakistan is a developing country and there is no cancer registry programs to keep a track related to the prevalence and incidence of leukemia, for this purpose studies were designed in Khyber Pakhtunkhwa, Lahore, and its nearby regions.

Nasim *et al,* 2013 performed a survey analysis to investigate the prevalence of leukemia subtypes in Lahore and its nearby regions such as Kasur, Hasilpur and Dipalpur. The data were collected from Lahore General Hospital during the period of two years from June 2010 to June 2012 and was consist of 45 patients who were diagnosed with leukemia. Sudan Black B was used to stain the peripheral blood smears. Blood counts and bone marrow biopsy were performed. The results showed that 80% of the patients were observed with acute leukemia in which 49% patients had ALL and 31% had AML while 20% patients were observed with chronic leukemia in which 16% had CML and 2% had CLL. They also performed age and gender-based distribution which showed that 57% males and 43% females were diagnosed with AML, 59% males and 41% females were diagnosed with ALL, 43% males and 57% females patient were observed with CML and only one patient was observed with CLL [70].

Ahmad *et al,* 2019 designed a study to find out the prevalence of leukemia subtypes in Khyber Pakhtunkhwa, Pakistan during the period of January 2015 to December 2016. The data of 400 admitted patients at Institute of Radiotherapy and Nuclear Medicine Peshawar were investigated. The result showed that acute leukemia was dominant than the chronic leukemia, as 80% patients were observed with acute leukemia and 20% were observed with chronic leukemia. 49.5% patients were diagnosed with ALL while 31.5% were diagnosed with AML. ALL was more prevalent than AML. 10% patients were detected with CML while 9.25% were detected with CLL. The prevalence of leukemia was dominant in males (64.5%) as compared to females (35.5%) and the male to female ratio was 1.8:1 [71].

2.1.5 Gaps in the Literature:

Majority of the studies are focusing predictive modelling using microscopic images for the detection or diagnosis of Leukemia or its subtypes while undermining the strength of models based on numerical data.

In Pakistan, limited literature is available for descriptive and inferential analysis using different variables of CBC reports for the objective screening of Leukemia or its subtypes.

METHODOLOGY

Statistical procedures carry out a study which include planning, designing, data collection, data analysis, conclude significant description and reporting of the research outcomes. Statistical analysis provides meaning to the meaningless numbers and bring life to a lifeless data. The precision of results and interpretations depend on the use of proper statistical tests [72].

The emphasis of this study is to analyses significant characteristics of CBC reports for the development of a predictive model. This model will be useful for the initial screening of Leukemia Subtypes. A primary data consisting of about 302 CBC reports has been collected from different hospitals of Rawalpindi and Islamabad regions. Table 3.1 shows the details related to CBC reports.

Table 3.1: Details related to CBC reports.

S. No.	**Source of Information / Abbreviations**	**Frequency**				**Total**
		AML	**CML**	**ALL**	**Normal**	
1.	Fauji Foundation Hospital	82	50	12	00	144
2.	Pakistan Institute of Medical Sciences (PIMS)	14	10	02	00	26
3.	SHIFA International Hospital	04	17	00	00	21
4.	Atta-Ur-Rahman School of Applied Biosciences Diagnostic Lab (ASAB)	00	08	04	15	27
5.	Khan Research Laboratories (KRL) G-9/1	02	00	00	22	24
6.	Maroof International Hospital	00	00	00	11	11
7.	Quaid-e-Azam International Hospital	24	00	00	20	44
8.	Excel Labs	05	00	00	00	5
9.	**Grand Total**	**131**	**85**	**18**	**68**	**302 CBC Reports**

In this study both the quantitative and qualitative data is used for the analysis. Qualitative data is non-numerical and descriptive in nature. This data is collected in the form of words and sentences [72]. In this research the qualitative variable is gender. The data that show some quantity through mathematical value is known as quantitative data. This data is numerical in nature [73]. The quantitative data are age, White Blood Cells, Red Blood Cells, Haemoglobin, Haematocrit, Mean Corpuscular Volume, Mean Corpuscular Haemoglobin, Mean Corpuscular Haemoglobin Concentration, Platelet Count, Neutrophil

Counts, Lymphocyte Counts, Basophil Counts, Eosinophil Counts and Monocyte Counts. Detail of variables and their short description are shown in Table 3.2.

Table 3.2: Variables and their short description[44].

S. No.	Variables	Abbrev -iations	Description
1	Age	NA	In Years
2	Gender	M/F	M = Male, F = Female
3	White Blood Cells	WBC	WBCs are also known as leukocytes. These are the immune system cells and helps in protecting the body from infections and external attackers such as viruses, bacteria's, and other pathogens.
4	Red Blood Cells	RBC	RBCs are also known as erythrocytes. These cells circulate throughout the body and transfer oxygen to the body tissues. The stem cells in the bone marrow form these cells.
5	Haemoglobin	Hb	Haemoglobin is the protein that carries oxygen found inside all RBCs. It gives red color to the RBCs. It transports carbon dioxide from tissues and organs back to the lungs.
6	Haematocrit	PCV	In CBC test haematocrit calculates the blood fraction that is composed of RBCs. Its value is set as a percentage of red blood cells in a volume of blood.
7	Mean Corpuscular Volume	MCV	MCV measures the size of red blood cells.
8	Mean Corpuscular Haemoglobin	MCH	The MCH calculates the haemoglobin content of each red blood cell.
9	Mean Corpuscular Haemoglobin Concentration	MCHC	MCHC shows the quantity of haemoglobin in per unit volume of red blood cell.
10	Platelet Count	PLT	Platelets are also known as thrombocytes. They are the smallest type of blood cells. When bleeding happens, these cells helps in clotting as

			they swell, bundle together, and form a sticky mass to halt bleeding.
11	Neutrophil Counts	ANC	Neutrophils are rich type of WBCs and constitute 65% of the leukocytes. They protect body from infections and consume infectious agents.
12	Lymphocyte Counts	LYM	Lymphocytes consist of 25% of the leukocytes. They are divided into two cells B cells and T cells. These cells start different forms of immune response by producing different antibodies.
13	Basophil Counts	BASO	Basophils cells constitute 1% of the leukocytes. They are the form of WBCs and cause immunological reaction to parasites.
14	Eosinophil Counts	EO	Eosinophils constitute 4% of the leukocytes. These are the type of white blood cells which fight against viral infections and allergies.
15	Monocyte Counts	MO	Monocytes constitute 6 % of the leukocytes. These are the type of white blood cells and the largest leukocytes. They provide immediate protection by engulfing and digesting the infectious agents.

3.1 Data Pre-processing:

Data pre-processing or data screening is the process to prepare the data for further statistical analysis [74]. Screening includes the checking of missing values, errors or omission in the data and checking the feasibility of the variables for further analysis. It makes data valid for testing.

Missing data poses many issues. These includes inefficient prediction, complication in the study's research, reduction in the statistical power, and sample representation. All these issues may lead toward the invalid assumptions [75].

3.1.1 Dropping of Cases:

Since the data is gathered from various sources; therefore, first data completeness has been checked. On inspection, there were few missing observations within the dataset. This problem is tackled in two parts. Firstly, the cases having more than 60% percent missing values and variables are removed. All the zero present in the data are considered as missing values. The variable reticulocyte has 67% missing values, so it is removed from the analyses. Out of 302 cases, 15 cases are omitted, while 287 cases are further analysed. In addition, the remaining missing values are calculated using the statistical method, Expected Maximization (EM), using the Statistical Software for Social Sciences (SPSS). Table 3.3 shows the percentage of missing values in the variables.

Table 3.3: Percentage of missing values in variables.

S. No.	Variables	Percentage of missing values
1	Basophil Count	29
2	Basophil Percentage	29
3	Eosinophil Count	6
4	Eosinophil Percentage	5
5	Monocyte Count	4
6	Monocyte Percentage	4
7	Neutrophil Count	4
8	Neutrophil Percentage	2
9	Reticulocyte Percentage	67

3.2 Estimation of Missing Values:

Missing data can lead to a serious impact on quantitative research. It can lead to a biased estimate of parameters, loss of information, reduced statistical power, increment in standard errors, and reduced generalizability of outcomes [76]. There are variety of techniques to manage the missing data which are Listwise or case deletion, Pairwise deletion, Mean substitution, Regression imputation, Maximum likelihood, Expectation-Maximization, Multiple imputation [75]. In this study Expected-Maximization (EM) method is used to estimate missing values. Figure 3.1 shows the steps of this method.

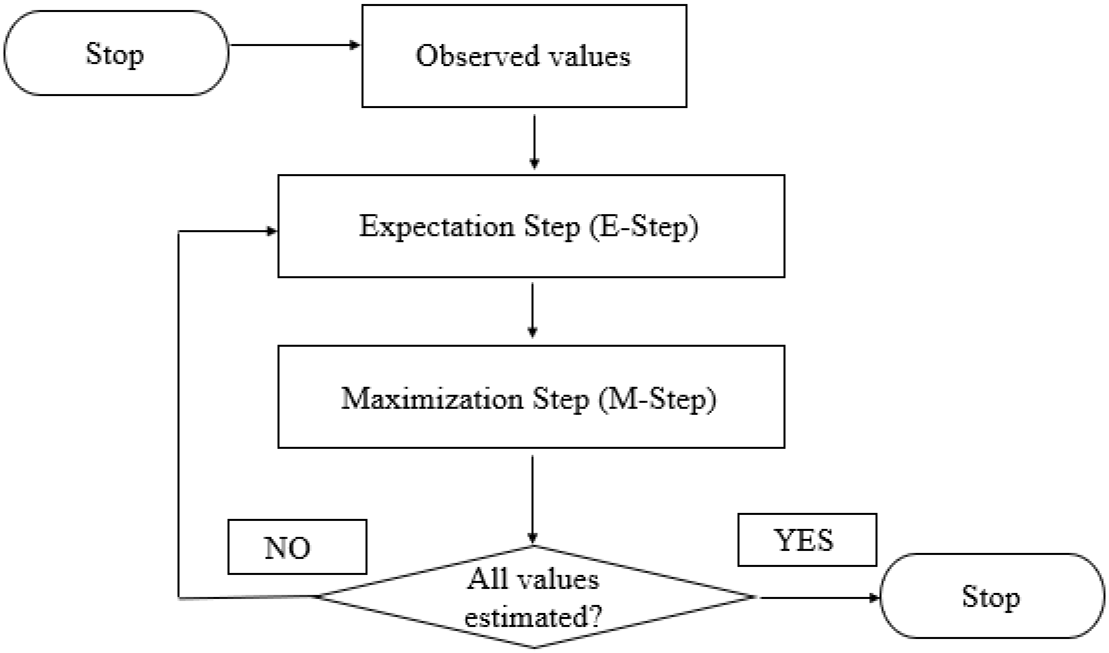


Figure 3.1: Steps of Expectation-Maximization method

3.2.1 Expected-Maximization:

Expectation-Maximization (EM) is a method of maximum likelihood that can be used to construct a new data set where all missing values are imputed with values determined by the methods of maximum likelihood [75]. This algorithm works in two steps: An E-step or Expectation step and the M-step or Maximization step [77]. This method starts with the step of expectation, during which the parameters such as variances, covariances, and means are calculated, possibly using the deletion of the list. Those estimates are then used to construct an equation of regression to estimate the missing data. The step of maximization uses certain equations to fill in the missing details because the missing values are not directly filled in. For the new parameters, the expectation step is then repeated, where the new regression equations are calculated to "fill in" the missing data. Expectations and maximizations are repeated until the system stabilizes, when the covariance matrix for the subsequent iteration is practically the same as for the preceding iteration [75].

3.3 Bias Variable:

To perform this study first, the blood cell count is replaced with percentages (as we believe that these variables were carrying a similar type of information) and performed the modelling. The results were insignificant in terms of predictive ability to discriminate between the normal and disease case. Therefore, these variables were replaced and in second stage we used counts instead of percentages. In this study percentages of blood cells such as neutrophil percentage, lymphocyte percentage, eosinophil percentage, basophil percentage and monocyte percentage are dropped from the analysis because these variables have less significant influence on leukemia and its subtypes. This study uses absolute counts of the blood cells.

3.4 Variable Selection:

There are generally 21 variables in the CBC reports. The percentages of blood cells are dropped and 15 variables Age, Gender, White Blood Cells, Red Blood Cells, Haemoglobin, Haematocrit, Mean Corpuscular Volume, Mean Corpuscular Haemoglobin, Mean Corpuscular Haemoglobin Concentration, Platelet Count, Neutrophils Counts, Lymphocytes Counts, Basophil Counts, Eosinophil Counts, Monocytes Counts are included in this study.

3.5 Descriptive Analyses:

Descriptive statistics is the discipline that quantitatively describe the major properties of collected information. Descriptive analysis gives summary of data in the form of mean, median, mode, minimum, maximum, skewness and kurtosis [72]. The measure of central tendency used in this research is mean and the measure of dispersion used is standard deviation.

3.6 Inferential Statistics:

In this study, one-way analysis of variance (ANOVA) has been used to check whether there exists statistically significant difference in means of four categories (Normal, AML, CML, ALL) with respect to each characteristics of a CBC report.

3.7 Coefficient of Correlation:

Correlation coefficient (r) calculates the intensity and direction of linear relationship between the sets of continuous variables. The Pearson Correlation is a parametric measure [78].

The range of correlation coefficient is from -1 to 1. In correlation coefficient, the direction of relationship is mentioned by sign, while the degree of the correlation (how close it is to -1 or +1) specifies the power of the relationship [78]. In correlation coefficient -1 shows perfect negative linear relationship. 0 shows no relationship while +1 shows perfect positive linear relationship [78].

3.8 Predictive Modelling:

Predictive modelling assist healthcare practitioners and patients in making clinical decisions [79]. The objective of an exact prediction model is to deliver categorization of patient risk in order to facilitate personalized clinical decision taking with the aim of improving patient results and quality of care [79].

3.8.1 Regression Analysis:

For the analysis of medical data, regression analysis is an important statistical tool. It allows relationships between multiple factors to be defined and characterized. It also helps prognostically important risk factors to be defined and risk scores to be determined for individual prognosis [80].

3.8.2 Logistic Regression:

Logistic regression is a statistical model that uses a logistic function to model a binary dependent variable in its basic form, although there are several more complex extensions [81]. The logistic regression model is a representative of the supervised classification algorithm family. Building block principles of logistic regression can aid deep learning when constructing neural networks [82].

Logistic Regression can be regarded as a basic regression extension and can model only a dichotomous variable that typically describes an event's occurrence or non-occurrence. Logistic Regression helps to find the possibility of a new case belonging to a particular class [82].

Based on individual characteristics, the logistic regression technique models the chance of an outcome. As the chance is a ratio, what is going to be modelled is the chance logarithm given by [83]:

$$\boldsymbol{l = \ log_b \frac{p}{1-p} = \ \beta_0 + \ \beta_1 x_1 \ + \beta_2 \, x_2 + \beta_3 x_3 \ + \beta_4 x_4 \ \ + \cdots + \beta_n x_n + e_i}$$

In this equation *l* is the log-odds, *b* is the base of the algorithm, p shows the probability of an event e.g., diseased and 1 – p indicates the normal. β_i, are the regression coefficients linked with the reference group, x_i are explanatory variables or predictor and e_i is the error term.

3.8.3 Types of Logistic Regression:

There are in general two types of logistic regression based upon the nature of the dependent variable which is qualitative or categorical in nature [81].

3.8.3.1 Binary Logistic Regression:

In binary logistic regression, the dependent variable has only two possible outcomes. These outcomes may be labelled as “0” and “1”.

3.8.3.2 Multinomial or Ordinal Logistic Regression:

In multinomial logistic regression, the dependent variable has at least three possible outcomes. If there is an order in multiple categories, then it is known as ordinal logistic regression.

3.8.4 Assumptions of Logistic Regression:

Following are the major assumptions associated to the estimation of logistic regression modelling [84], [85]:

1- There is no requirement of linear relationship between dependent and independent variables.
2- Usually there is no need of normal behaviour of error term (residuals) of the model.
3- Homoscedasticity is not mandatory in logistic regression model.
4- It assumes that the observations should be independent of each other.
5- The dependent variable is not calculated on an interval or ratio scale.
6- It is desirable that among the independent variables there is minimal to no multicollinearity.
7- To predict correctly, logistic regression typically needs a broad sample size.
8- The two-class logistic regression assumes that the dependent variable is binary, and the ordered logistic regression includes the order of the dependent variable.

3.8.5 Multinomial Logistic Regression:

Multinomial Logistic Regression (MLR) is a supervised learning technique to conduct when there are more than two nominal or unordered categories in the dependent variable [86]. It is the extension of binary logistic regression and uses maximum likelihood estimation to assess the possibility of categorical membership. In this study, Multinomial Logistic Regression is applied because there are 4 categories of dependent variable i.e. Normal, AML, CML and ALL.

3.9 Model Evaluation:

In this study, the data is tested with:

- True Positive (TP) as diseased cases are correctly predicted as diseased.
- False Positive (FP) as normal cases that are incorrectly predicted as diseased.
- True Negative (TN) as real normal cases that are correctly predicted as normal.
- False Negative (FN) as diseased cases that are incorrectly identified as normal [87].

The 2x2 matrix is shown below:

	Observed Positive (1)	**Observed Negative (0)**
Predicted Positive (1)	True Positive (TP)	False Positive (FP)
Predicted Negative (0)	False Negative (FN)	True Negative (TN)

Four important measures of model assessment include the classification accuracy, sensitivity or true positive rate, specificity or true negative rate, and precision or positive prediction value (PPV). Details and formulas of these measures are as follows:

1. **Classification Accuracy:**

Accuracy estimates the right sample ratio and is one of the most intuitive and fundamental output metrics for any model [62].

$$P = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Sensitivity:**

The ability of a test to correctly identify a person as ′diseased′ is known as sensitivity[88].

$$P_p = \frac{TP}{TP + FN}$$

3. **Specificity:**

The specificity of a test refers to its ability to accurately identify a person as disease-free[88].

$$P_n = \frac{TN}{TN + FP}$$

4. **Precision or Positive Predicted Value (PPV):**

Precision is the percentage of patients with a positive test who actually have the disease[88].

$$PPV = \frac{TP}{TP + FP}$$

RESULTS and DISCUSSION

The focus of this study is to develop a predictive modelling for the screening of leukemia and its subtypes using numerical estimates of CBC reports. For this purpose, the data of 302 subjects has been collected from different hospitals of twin cities (Islamabad and Rawalpindi). These hospitals are Fauji Foundation, Pakistan Institute of Medical Sciences (PIMS), SHIFA International hospital, Diagnostic Lab of ASAB, Khan Research Laboratories (KRL), Maroof International hospital, Quaid-e-Azam International hospital, and Excel Lab of Shifa International Hospital.

CBC report usually consists of 21 different characteristics of a subject. In the report both the frequency and percentages are available for few of the characteristics/ variables such as Basophil, Eosinophil, Monocytes, Lymphocytes and Neutrophil. In first attempt of our analyses, we have used percentages instead of counts believing that they hold more meaningful information. However, the estimates of the model are statistically insignificant in term of predictive ability to discriminate between normal and disease cases. Therefore, in second attempt those percentages have been dropped and replaced by their respective counts. By doing so, this time, the results of the model are showed statistical significance in terms of appropriates of the choice of independent variables in the model.

15 variables have been selected or short listed for the analysis. These variables are Age, Gender, WBC, RBC, Haemoglobin, PCV, MCV, MCH, MCHC, Platelet Count, Neutrophil count, Lymphocyte count, Basophil count, Eosinophil count and Monocytes Count.

4.1 Descriptive Analyses:

Descriptive analyses such as mean and standard deviation is calculated for subtype 0,1,2, and 3.

Table 4.1: Descriptive Measures of Different Variables for Normal and Subtypes of Disease (AML, CML, ALL)

Sr. No.	Variables	Subtypes4 Coding	N	Mean	SD
1	Age	0	67	39.25	19.49
		1	123	33.95	20.49
		2	79	45.42	17.18
		3	18	15.56	16.02
2	WBC	0	67	8.22	3.18
		1	123	14.94	27.16
		2	79	100.2	136.0
		3	18	21.34	41.85
3	RBC	0	67	4.48	0.60
		1	123	3.22	0.76
		2	79	3.58	0.94
		3	18	3.38	0.59
4	Haemoglobin	0	67	12.99	1.75
		1	123	9.43	2.07
		2	79	10.13	2.34
		3	18	9.617	2.03
5	Haematocrit	0	67	38.83	4.84
		1	123	27.08	6.19
		2	79	31.22	7.08
		3	18	29.70	4.04
6	MCV	0	67	86.15	6.56
		1	123	84.46	7.48
		2	79	87.68	9.40
		3	18	86.77	8.11
7	MCH	0	67	28.85	2.89
		1	123	29.48	2.77
		2	79	28.94	3.43
		3	18	29.17	3.86
8	MCHC	0	67	33.34	1.36
		1	123	34.88	1.68
		2	79	32.91	2.03
		3	18	33.24	2.86
9	Platelet Count	0	67	251.75	54.74
		1	123	150.5	157.5
		2	79	200.1	158.8
		3	18	150.1	84.5
10	Neutrophil Count	0	67	4.59	0.87
		1	123	19.39	81.28
		2	79	79.8	112.4
		3	18	11.26	20.00

Table 4.1 Continued……					
11	Lymphocyte Count	0	67	2.38	0.94
		1	123	5.60	16.41
		2	79	10.80	15.30
		3	18	18.29	38.40
12	Basophil Count	0	67	0.24	0.24
		1	123	0.34	1.19
		2	79	1.93	3.00
		3	18	0.26	0.29
13	Eosinophil Count	0	67	0.23	0.17
		1	123	0.40	0.71
		2	79	1.90	3.49
		3	18	0.18	0.28
14	Monocyte Count	0	67	0.48	0.46
		1	123	2.61	4.92
		2	79	12.33	17.71
		3	18	1.50	3.746

Here: 0 = Normal**, 1** = AML, **2** = CML, **3 =** ALL, **n** = Number of observations and **SD** = Standard Deviation

Literature in Chapter 2 shows that ALL is a childhood leukemia and the descriptive analyses of age in Table 4.1 also shows the same as the average age of ALL is 15.2. As leukemia is the cancer of White Blood Cells so with respect to WBC there is an increase in the mean of blood count of CML. When WBC increases there occur decrease in the RBC. The mean Red Blood Cell Counts for AML, CML and ALL is lower than the normal. When RBC counts decreases it also effects hemoglobin and hematocrit. The Table 4.1 shows that the average hemoglobin and the average hematocrit of three subtypes are lower than the average of normal. The average MCV, MCH and MCHC of the three subtypes are almost similar to the average of their normal.

The average platelet counts of CML, AML and ALL is lower than the mean of normal. As the WBC's are divided into five types which are neutrophils, lymphocytes, monocytes, eosinophils, and basophils. The average neutrophil counts of all the three subtypes are greater than the mean of normal and all the means are significantly different from each other. The mean lymphocyte count of ALL is very high as compared to normal. The average of basophil counts of all the three subtypes are higher than the mean basophil counts of normal. The mean of CML is significantly different from the normal mean. The average eosinophil counts of AML and CML is higher than the mean eosinophil counts of normal. The mean of ALL is lower than the normal mean. The average of monocyte counts of the three subtypes are greater than the average of normal. The average monocyte count of CML is very high from the average monocyte count of normal.

4.2 Comparing Means Through ANOVA:

The results of descriptive analysis in Table 4.1 are showing variations in the values of mean for Normal vs Three subtypes of leukemia. Therefore, there is a need to validate statistically that whether there are statistically significant differences between the numerical estimates of means with respect to 4 categories for all the 14 variables or not?

4.2.1 Comparing Means:

Table 4.2: Results of ANOVA

S. No.	VARIABLES	F -VALUE	P-VALUE
1	Age	13.96	0.00
2	WBC	26.24	0.00
3	RBC	38.56	0.00
4	Haemoglobin	44.75	0.00
5	Haematocrit	54.57	0.00
6	MCV	2.78	0.04
7	MCH	0.82	0.48
8	MCHC	22.77	0.00
9	Platelet Count	8.56	0.00
10	Neutrophil Count	13.55	0.00
11	Lymphocyte Count	6.31	0.00
12	Basophil Count	16.33	0.00
13	Eosinophil Count	13.33	0.00
14	Monocyte Count	22.17	0.00

Hypothesis:

The hypothesis for ANOVA is:

$$\boldsymbol{H_O: \mu_{Normal} = \mu_{AML} = \mu_{CML} = \mu_{ALL}}$$

$$\boldsymbol{H_1: \mu_{Normal} \neq \mu_{AML} \neq \mu_{CML} \neq \mu_{ALL}}$$

$$\propto = \mathbf{0.05}$$

The p-value is compared to α, which can be set at different levels. If α = 0.05, then a p score less than 0.05 indicates statistically significant differences, a p scores greater than 0.05 means that there is no statistical difference [89].

The Table 4.2 shows that out of 14 variables only MCH has the p-value of 0.48 which is greater than alpha, and has insignificant difference between the means of Normal, AML, CML and ALL while all other variables show statistically significant result.

4.3 Correlation Analysis:

Table 4.3: Correlation Matrix of Fourteen Numeric Variables

Correlations															
		Age	**WBC**	**RBC**	**Hb**	**HCT**	**MCV**	**MCH**	**MCHC**	**PLT Ct**	**ANC**	**LC**	**BC**	**EC**	**MC**
Age	CC	1	0.15	0.09	0.05	0.09	0.06	-0.07	-0.23	0.00	0.22	-0.01	0.18	0.15	0.17
	(p-value)		0.00	0.09	0.31	0.11	0.27	0.23	0.00	0.91	0.00	0.74	0.00	0.01	0.00
WBC	CC	0.15	1	-0.29	-0.27	-0.25	0.19	0.08	-0.14	0.12	0.81	0.31	0.84	0.67	0.70
	(p-value)	0.00		0.00	0.00	0.00	0.00	0.14	0.01	0.02	0.00	0.00	0.00	0.00	0.00
RBC	CC	0.09	-0.29	1	0.86	0.92	-0.28	-0.38	-0.22	0.40	-0.27	-0.16	-0.22	-0.23	-0.27
	(p-value)	0.09	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hb	CC	0.05	-0.27	0.86	1	0.94	0.07	0.06	0.01	0.41	-0.24	-0.18	-0.20	-0.19	-0.24
	(p-value)	0.31	0.00	0.00		0.00	0.18	0.31	0.81	0.00	0.00	0.00	0.00	0.00	0.00
HCT	CC	0.09	-0.25	0.92	0.94	1	0.03	-0.10	-0.23	0.40	-0.24	-0.13	-0.18	-0.21	-0.21
	(p-value)	0.11	0.00	0.00	0.00		0.60	0.07	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MCV	CC	0.06	0.19	-0.28	0.07	0.03	1	0.83	0.00	-0.00	0.17	0.14	0.17	0.14	0.24
	(p-value)	0.27	0.00	0.00	0.18	0.60		0.00	0.98	0.92	0.00	0.01	0.00	0.01	0.00
MCH	CC	-0.07	0.08	-0.38	0.06	-0.10	0.83	1	0.50	-0.05	0.15	-0.03	0.12	0.15	0.10
	(p-value)	0.23	0.14	0.00	0.31	0.07	0.00		0.00	0.33	0.00	0.55	0.03	0.00	0.06
MCHC	CC	-0.23	-0.14	-0.22	0.01	-0.23	0.00	0.50	1	-0.07	-0.00	-0.26	-0.07	0.03	-0.16
	(p-value)	0.00	0.01	0.00	0.81	0.00	0.98	0.00		0.20	0.87	0.00	0.22	0.57	0.00
PLT Ct	CC	0.00	0.12	0.40	0.41	0.40	-0.00	-0.05	-0.07	1	0.12	-0.13	0.10	0.27	0.01
	(p-value)	0.91	0.02	0.00	0.00	0.00	0.92	0.33	0.20		0.03	0.02	0.08	0.00	0.80
ANC	CC	0.22	0.81	-0.27	-0.24	-0.24	0.17	0.15	-0.00	0.12	1	0.09	0.87	0.63	0.59
	(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.87	0.03		0.10	0.00	0.00	0.00
LC	CC	-0.01	0.31	-0.16	-0.18	-0.13	0.14	-0.03	-0.26	-0.13	0.09	1	0.12	0.08	0.25
	(p-value)	0.74	0.00	0.00	0.00	0.02	0.01	0.55	0.00	0.02	0.10		0.03	0.13	0.00
BC	CC	0.18	0.84	-0.22	-0.20	-0.18	0.17	0.12	-0.07	0.10	0.87	0.12	1	0.52	0.70
	(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.22	0.08	0.00	0.03		0.00	0.00
EC	CC	0.15	0.67	-0.23	-0.19	-0.21	0.14	0.15	0.03	0.27	0.63	0.08	0.52	1	0.27
	(p-value)	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.57	0.00	0.00	0.13	0.00		0.00
MC	CC	0.17	0.70	-0.27	-0.24	-0.21	0.24	0.10	-0.16	0.01	0.59	0.25	0.70	0.27	1
	(p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.80	0.00	0.00	0.00	0.00	

Here: **CC** = Correlation Coefficient, **PLT Ct** = Platelet Count, **ANC** = Absolute Neutrophil Count **LC** = Lymphocyte Count, **BC** = Basophil Count, **EC** = Eosinophil Count and **MC** = Monocyte Count

For the inquiry of existence of linear relationship between variables, correlation analysis has been performed. This analysis will also help in the identification of multicollinearity problem for the development of logistic regression models stated in the assumption no 6 of section 3.8.4.

Following is the procedure for testing the significance of correlation coefficient.

Hypothesis:

The hypothesis for correlation analysis is:

$$\boldsymbol{H_O : \rho = 0}$$

$$\boldsymbol{H_1 : \rho \neq 0}$$

Level of significance $\propto = \mathbf{0.05}$

4.3.1 Age:

Age has statistically significant correlation with 6 variables. These variables are WBC, MCHC, Neutrophil Count, Basophil Count, Eosinophil Count and Monocyte Count. It has weak and statistically insignificant correlation with 7 variables. These variables are RBC, Haemoglobin, Haematocrit, MCV, MCH, Platelet Count and Lymphocyte Count.

4.3.2 WBC:

WBC has statistically significant correlation with 12 variables. These variables are Age, RBC, Haemoglobin, Haematocrit, MCV, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. WBC has weak and statistically insignificant correlation with 1 variable which is MCH. WBC has negative correlation with RBC, Haemoglobin, Haematocrit and MCHC.

4.3.3 RBC:

RBC has statistically significant correlation with 12 variables. These variables are WBC, Haemoglobin, Haematocrit, MCV, MCH, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. Variable Age has weak and statistically insignificant correlation with RBC.

4.3.4 Haemoglobin:

Haemoglobin has statistically significant correlation with 9 variables. These variables are WBC, RBC, Haematocrit, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. It has weak and statistically insignificant correlation with 4 variables which are Age, MCV, MCH and MCHC.

4.3.5 Haematocrit:

Haematocrit has statistically significant correlation with 10 variables. These variables are WBC, RBC, Haemoglobin, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. Haematocrit has weak and statistically insignificant correlation with 3 variables which are Age, MCV and MCH.

4.3.6 MCV:

MCV has statistically significant correlation with 8 variables. These variables are WBC, RBC, MCH, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. It has weak and statistically insignificant correlation with 5 variables. These variables are Age, Haemoglobin, Haematocrit, MCHC and Platelet Count.

4.3.7 MCH:

MCH has statistically significant correlation with 6 variables and these variables are RBC, MCV, MCHC, Neutrophil Count, Basophil Count and Eosinophil Count. MCH has weak and statistically insignificant correlation with 7 variables. These variables are Age, WBC, Haemoglobin, Haematocrit, Platelet Count, Lymphocyte Count and Monocyte Count.

4.3.8 MCHC:

MCHC has statistically significant correlation with 7 variables. These variables are Age, WBC, RBC, MCH, Haematocrit, Lymphocyte Count and Monocyte Count. It has weak and insignificant correlation with 6 variables. These variables are Haemoglobin, MCV, Platelet Count, Neutrophil Count, Basophil Count and Eosinophil Count.

4.3.9 Platelet Count:

Platelet Count has statistically significant correlation with 7 variables. These variables are WBC, RBC, Haemoglobin, Haematocrit, Neutrophil Count, Eosinophil Count and Lymphocyte Count. It has weak and insignificant correlation with 6 variables. These variables are Age, MCV, MCH, MCHC, Basophil Count and Monocyte Count.

4.3.10 Neutrophil Count:

Neutrophil Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, MCH, MCV, Haemoglobin, Haematocrit, Platelet Count, Basophil Count, Eosinophil Count and Monocyte Count. It has weak and insignificant correlation with 2 variables. These variables are MCHC and Lymphocyte Count.

4.3.11 Lymphocyte Count:

Lymphocyte Count has statistically significant correlation with 9 variables. These variables are WBC, RBC, Haematocrit, Haemoglobin, MCV, MCHC, Platelet Count, Basophil Count and Monocyte Count. It has weak and insignificant correlation with 4 variables. These variables are Age, MCH, Neutrophil Count and Eosinophil Count.

4.3.12 Basophil Count:

Basophil Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, Haematocrit, Haemoglobin, MCV, MCH, Neutrophil Count, Lymphocyte Count, Eosinophil Count, Monocyte Count. It has weak and insignificant correlation with 2 variables. These variables are MCHC and Platelet Count.

4.3.13 Eosinophil Count:

Eosinophil Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, Haematocrit, Haemoglobin, MCV, MCH, Neutrophil Count, Platelet Count, Basophil Count and Monocyte Count. It has weak and insignificant correlation with 2 variables. These variables are MCHC, and Lymphocyte Count.

4.3.14 Monocyte Count:

Monocyte Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, Haematocrit, Haemoglobin, MCV, MCHC, Neutrophil Count, Lymphocyte Count Basophil Count and Eosinophil Count. It has weak and insignificant correlation with 2 variables. These variables are MCH and Platelet Count.

4.4 Development of Multinomial Logistic Regression:

4.4.1 Variables Selection using Backward Elimination Method:

In backward selection criteria, we start with the model having all the independent variables. Then dropping insignificant variables one after another based on their rate of insignificance. Corresponding p-value of the Wald test has been used for exclusion of insignificant variables [90].

A brief of the procedure i.e., dropping of variables at every step is provided below.

Step 1: Dropping MCH

The p-value of Wald test shows that MCH has highest statistically insignificant relation with reference to subtypes 1, 2 and 3. According to subtype 1, the p-value is 0.766 and for subtype 2 its p- value is 0.720. For subtype 3 the P-value is 0.816.

The results of Likelihood Ratio Test show that MCH has Chi-square value of 1.208 with p-value 0.751. Therefore, we are unable to reject the hypothesis that the effect of this parameter in the model is zero.

Another important consideration is that correlation matrix shows MCH has strongly positive significant correlation with 6 variables namely RBC, MCV, MCHC, Neutrophil Count, Basophil Count and Eosinophil Count. It has weak and statistically insignificant correlation with 7 variables which are Age, WBC, Haemoglobin, Haematocrit, Platelet Count, Lymphocyte Counts and Monocyte Count. Therefore, this variable may cause problem of multicollinearity and effect on the assessment measure of model.

Similar principles have been used for dropping of rest of the variables step by step. To avoid repetition only statistical details have been provided for the rest of the steps.

Step 2: Dropping Platelet Count

The p-value of Wald test for subtype 1, 2 and 3 is 0.914, 0.548 and 0.343, respectively.

For Likelihood Ratio Test Chi-square value is 2.170 with p-value 0.538.

It has strongly positive and significant correlation with 7 variables namely WBC, RBC, Haemoglobin, Haematocrit, Neutrophil Count, Lymphocyte Count, and Eosinophil Count. Whereas weak and insignificant correlation with 6 variables which are Age, MCV, MCH, MCHC, Basophil Count and Monocyte Count.

Step 3: Dropping Eosinophil Count

The p-value of Wald test for subtype 1, 2 and 3 is 0.290, 0.176 and 0.565, respectively.

For Likelihood Ratio Test Chi-square value of 5.156 with p-value 0.161.

It has strong positive and significant correlation with 11 variables such as Age, WBC, RBC, Haemoglobin, Haematocrit, MCV, MCH, Platelet Count, Neutrophil Count, Basophil Count, and Monocyte Count, while it has weak and insignificant correlation with 2 variables which are MCHC and Lymphocyte Count.

Step 4: Dropping MCV

The p-value of Wald test for subtype 1, 2 and 3 is 0.055, 0.099 and 0.328, respectively.

For Likelihood Ratio Test Chi-square value is 4.276 with p-value 0.233.

It has strong positive and significant correlation with 8 variables namely WBC, RBC, MCH, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. It has weak and insignificant correlation with 5 variables which are Age, Haemoglobin, Haematocrit, MCHC and Platelet Count.

Step 5: Dropping Haematocrit

The p-value of Wald test for subtype 1, 2 and 3 is 0.513, 0.093 and 0.156, respectively.

For Likelihood Ratio Test Chi-square value is 16.726 with p-value 0.001.

It has strong positive and significant correlation with 10 variables which are WBC, RBC, Haemoglobin, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. Haematocrit has weak and insignificant correlation with 3 variables which are Age, MCV and MCH.

Step 6: Dropping RBC

The p-value of Wald test for subtype 1, 2 and 3 is 0.770, 0.987 and 0.043, respectively.

For Likelihood Ratio Test Chi-square value is 5.198 with p-value 0.158.

It has strong positive and significant correlation with 12 variables which are WBC, Haemoglobin, Haematocrit, MCV, MCH, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. RBC has weak and insignificant correlation with 1 variable which is age.

Step 7: Dropping Lymphocyte Counts

The p-value of Wald test for subtype 1, 2 and 3 is 0.059, 0.217 and 0.005, respectively.

For Likelihood Ratio Test Chi-square value is 21.219 with p-value 0.000.

It has strong positive and significant correlation with 9 variables. These variables are WBC, RBC, Haemoglobin, Haematocrit, MCV, MCHC, Platelet Count, Basophil Count and Monocyte Count. It has weak and insignificant correlation with 4 variables which are Age, MCH, Neutrophil Count and Eosinophil Count.

Step 8: Dropping WBC

The p-value of Wald test for subtype 1, 2 and 3 is 0.099, 0.736 and 0.987, respectively.

For Likelihood Ratio Test Chi-square value is 15.798 with p-value 0.001.

It has strong positive and significant correlation with 12 variables which are Age, RBC, Haemoglobin, Haematocrit, MCV, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. WBC has weak and insignificant correlation with 1 variable which is MCH.

Step 9: Dropping MCHC

The p-value of Wald test for subtype 1, 2 and 3 is 0.000, 0.034 and 0.653, respectively.

For Likelihood Ratio Test Chi-square value is 75.005 with p-value 0.000.

It has strong positive and significant correlation with 8 variables namely Age, WBC, RBC, Haematocrit, MCHC, MCH, Lymphocyte Count, and Monocyte Count. MCHC has weak and insignificant correlation with 6 variables. These variables are Haemoglobin, MCV, Platelet Count, Neutrophil Count, Basophil Count, Eosinophil Count.

Step 10: Dropping AGE

The p-value of Wald test for subtype 1, 2 and 3 is 0.470, 0.010 and 0.012, respectively.

For Likelihood Ratio Test Chi-square value is 27.361 with p-value 0.000.

It has strong positive and significant correlation with 6 variables. These variables are WBC, MCHC, Neutrophil Count, Basophil Count, Eosinophil Count and Monocyte Count. It has weak and insignificant correlation with 7 variables. These variables are RBC, Haemoglobin, Haematocrit, MCV, MCH, Platelet Count and Lymphocyte Count.

Set of Statistically Significant Variables:

Following the backward elimination procedure and dropping insignificant variables we left with five variables showing statistically significant results. Details of these variables, their coefficients, significant values, and odds ratio are present in Table 4.4.

Table 4.4: Set of statistically significant variables obtained from backward elimination method.

Subtypes	Variables	B	Wald	Df	p-value	Exp(B)
1	Hemoglobin	-1.11	52.64	1	0.00	0.32
	Neutrophil Count	0.19	17.02	1	0.00	1.21
	Basophil Count	-3.84	13.49	1	0.00	0.02
	Monocyte Count	1.51	8.92	1	0.00	4.53
	[Gender=F]	-2.55	16.00	1	0.00	0.07
	[Gender=M]	0^b	.	0	.	.
2	Hemoglobin	-0.75	25.18	1	0.00	0.46
	Neutrophil Count	0.17	13.59	1	0.00	1.18
	Basophil Count	-2.30	5.08	1	0.02	0.10
	Monocyte Count	1.61	10.15	1	0.00	5.02
	[Gender=F]	-2.66	16.68	1	0.00	0.07
	[Gender=M]	0^b	.	0	.	.
3	Hemoglobin	-1.06	32.16	1	0.00	0.34
	Neutrophil Count	0..18	13.57	1	0.00	1.19
	Basophil Count	-3.39	7.15	1	0.00	0.03
	Monocyte Count	1.47	8.19	1	0.00	4.35
	[Gender=F]	-3.44	19.19	1	0.00	0.03
	[Gender=M]	0^b	.	0	.	.

4.4.2 Variables Selection using Odds Ratio / Exp(B):

This section provides details of the procedure of selection of variables using odds ratio. Odds ratio (OR) is used to find out the occurrence of the consequence of interest. It is also used to assess if a single exposure is a risk factor for a specific outcome, and to compare the magnitude of different risk factors for that outcome [91].

$$Odds = \frac{P(diseased)}{P(Normal)}$$

OR = 1 shows exposure does not affect outcome probabilities.

OR > 1 shows exposure associated with greater chances of outcome.

OR < 1 shows exposure associated with lower chances of outcome.

Starting with the variable having OR greater than 3 in at least two subtypes are dropped step by step.

Step 1: Dropping Monocyte Count:

The values of OR of Monocyte Count for subtype 1, 2 and 3 is 7.162, 7.372 and 3.695, respectively. Therefore, it has been dropped and we re-run the model for the rest of the variables.

Step 2: Dropping MCHC

The OR of MCHC for subtype 1, 2 and 3 is 4.597, 5.539 and 1.916, respectively.

Step 3: Dropping RBC

The OR of RBC for subtype 1, 2 and 3 is 5.273, 3.698 and 2.930, respectively

Step 4: Dropping Eosinophil Count

The OR of Eosinophil Count for subtype 1, 2 and 3 is 3.073, 4.344 and 1.108, respectively.

Selected Variables Based on OR:

Following the OR criteria and dropping variables with large OR step by step, we left with 11 variables showing logical range OR. Details of these variables, their coefficients, significant values, and OR are present in Table 4.5.

Table 4.5: Set of variables obtained from Odds Ratio.

Subtypes	Variables	B	Wald	df	p-value	Exp(B)
1	Age	-0.00	0.08	1	0.76	0.99
	WBC	-0.05	5.94	1	0.01	0.94
	Hemoglobin	-0.13	0.02	1	0.87	0.87
	Hematocrit	-0.41	2.13	1	0.14	0.66
	MCV	-0.24	4.07	1	0.04	0.78
	MCH	0.75	4.54	1	0.03	2.13
	Platelet Count	0.00	0.00	1	0.94	1.00
	Neutrophil Count	0.20	13.81	1	0.00	1.22
	Lymphocyte Count	0.12	1.28	1	0.25	1.12
	Basophil Count	-2.35	6.05	1	0.01	0.09
	[Gender=F]	-0.13	0.03	1	0.85	0.87
	[Gender=M]	0^b	.	0	.	.
2	Age	0.02	4.42	1	0.03	1.02
	WBC	-0.00	0.02	1	0.88	0.99
	Hemoglobin	-1.11	1.86	1	0.17	0.33
	Hematocrit	0.12	0.19	1	0.65	1.13
	MCV	-0.10	0.73	1	0.39	0.90
	MCH	0.43	1.51	1	0.21	1.55
	Platelet Count	-0.00	0.61	1	0.43	0.99
	Neutrophil Count	0.16	10.02	1	0.00	1.18
	Lymphocyte Count	0.07	0.46	1	0.49	1.07
	Basophil Count	-0.59	1.03	1	0.30	0.55
	[Gender=F]	-1.31	3.17	1	0.07	0.26
	[Gender=M]	0^b	.	0	.	.
	Age	-0.11	8.22	1	0.00	0.89
	WBC	-0.08	5.12	1	0.02	0.91
	Hemoglobin	-1.10	1.53	1	0.21	0.33

3	Hematocrit	0.00	0.00	1	0.99	1.00
	MCV	-0.02	0.02	1	0.86	0.97
	MCH	0.28	0.40	1	0.52	1.33
	Platelet Count	0.00	0.48	1	0.48	1.00
	Neutrophil Count	0.20	12.47	1	0.00	1.22
	Lymphocyte Count	0.17	2.33	1	0.12	1.18
	Basophil Count	-1.48	0.95	1	0.32	0.22
	[Gender=F]	-2.00	3.99	1	0.04	0.13
	[Gender=M]	0[b]	.	0	.	.

4.4.3 Selection of Variables using a Combination of Dropping Insignificant Variables Simultaneously and Wald's Criteria:

Step 1: Dropping Insignificant Variables Simultaneously

This section provides information about dropping the insignificant variables simultaneously based on Likelihood Ratio Test. The Chi-square statistics is the difference in 2-log-liklehoods between the final model and the reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all the parameters of that effect are zero. These insignificant variables are:

- RBC
- MCV
- MCH
- Platelet Count
- Basophil Count
- Eosinophil Count

Table 4.6: Set of variables dropped simultaneously based on Likelihood Ratio Test.

Likelihood Ratio Tests				
Variables	**Model Fitting Criteria**	**Likelihood Ratio Tests**		
	-2 Log Likelihood of Reduced Model	**Chi-Square**	**df**	**p-value**
RBC	279.00	4.21	3	0.24
MCV	276.85	2.06	3	0.56
MCH	276.00	1.20	3	0.75
Platelet Count	277.00	2.21	3	0.52
Basophil Count	280.52	5.73	3	0.12
Eosinophil Count	280.08	5.29	3	0.15

After dropping the above-mentioned insignificant variables simultaneously, the rest of insignificant variables are dropped step by step based on p-value of Wald test.

Step 2: Dropping Haematocrit

For haematocrit, the corresponding p-value of Wald test for subtypes 1, 2 and 3 are 0.640, 0.129 and 0.407, respectively.

Step 3: Dropping Age

The p-value of Wald test for subtype 1, 2 and 3 is 0.188, 0.006 and 0.011, respectively.

Step 4: Dropping Gender

The p-value of Wald test for subtype 1, 2 and 3 is 0.165, 0.013 and 0.001, respectively.

Step 5: Dropping Lymphocyte Count

The p-value of Wald test for subtype 1, 2 and 3 is 0.036, 0.110 and 0.013, respectively.

Step 6: Dropping WBC

The p-value of Wald test for subtype 1, 2 and 3 is 0.192, 0.959 and 0.950, respectively.

Set of Statistically Significant Variables:

Following the procedure of variables selections using a combination of dropping insignificant variables simultaneously and Wald's criteria we left with 4 variables showing statistically significant results. Details of these variables, their coefficients, significant values, and OR are present in Table 4.7.

Table 4.7: Set of statistically significant variables obtained from a combination of dropping insignificant variables simultaneously and Wald's criteria.

Subtypes	Variables	B	Wald	df	p-value	Exp(B)
1	Hemoglobin	-1.19	46.28	1	0.00	0.30
	MCHC	1.10	39.64	1	0.00	3.00
	Neutrophil Count	0.16	12.56	1	0.00	1.17
	Monocyte Count	0.96	5.27	1	0.02	2.62
2	Hemoglobin	-0.79	20.60	1	0.00	0.45
	MCHC	0.41	6.12	1	0.01	1.51
	Neutrophil Count	0.16	13.61	1	0.00	1.18
	Monocyte Count	1.03	6.11	1	0.01	2.82
3	Hemoglobin	-1.12	29.85	1	0.00	0.32
	MCHC	0.53	7.45	1	0.00	1.71
	Neutrophil Count	0.15	9.90	1	0.00	1.16
	Monocyte Count	0.87	4.17	1	0.04	2.40

4.4.4 Selection of Variables using a Combination of Dropping Insignificant Variables Simultaneously and OR:

Step 1: Dropping Insignificant Variables Simultaneously:

This section provides information about dropping the insignificant variables simultaneously based on Likelihood Ratio Test. The Chi-square statistics is the difference in 2-log-liklehoods between the final model and the reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all the parameters of that effect are zero. These insignificant variables are:

- RBC
- MCV
- MCH
- Platelet Count
- Basophil Count
- Eosinophil Count

Table 4.8: Set of variables dropped simultaneously based on Likelihood Ratio Test

Likelihood Ratio Tests				
Variables	**Model Fitting Criteria**	**Likelihood Ratio Tests**		
	-2 Log Likelihood of Reduced Model	**Chi-Square**	**df**	**p-value**
RBC	279.00	4.21	3	0.24
MCV	276.85	2.06	3	0.56
MCH	276.00	1.20	3	0.75
Platelet Count	277.00	2.21	3	0.52
Basophil Count	280.52	5.73	3	0.12
Eosinophil Count	280.08	5.29	3	0.15

After dropping the above-mentioned insignificant variables simultaneously, the rest of variables are dropped step by step based on illogical OR/ Exp(B).

Step 2: Dropping Monocyte Count

The values of OR of Monocyte Count for subtype 1, 2 and 3 is 7.162, 7.372 and 3.695, respectively.

Step 3: Dropping MCHC

The values of OR of MCHC for subtype 1, 2 and 3 is 4.597, 5.539 and 1.916, respectively.

Set of Selected Variables:

Following the procedure of variables selections using a combination of dropping insignificant variables simultaneously and dropping variables with large OR step by step, we left with 7 variables showing logical range OR. Details of these variables, their coefficients, significant values, and OR are present in Table 4.9.

Table 4.9: Set of variables obtained from a combination of dropping insignificant variables simultaneously and OR.

Subtypes	Variables	B	Wald	df	p-value	Exp(B)
1	Age	-0.00	0.12	1	0.72	0.99
	WBC	-0.06	6.67	1	0.01	0.94
	Haemoglobin	1.20	6.74	1	0.00	3.34
	Haematocrit	-0.85	24.99	1	0.00	0.42
	Neutrophil Count	0.15	12.91	1	0.00	1.17
	Lymphocyte Count	0.12	1.81	1	0.17	1.13
	[Gender=F]	-0.31	0.24	1	0.61	0.72
	[Gender=M]	0^b	.	0	.	.
2	Age	0.03	5.69	1	0.01	1.03
	WBC	-0.01	0.29	1	0.58	0.98
	Haemoglobin	-0.29	0.45	1	0.50	0.74
	Haematocrit	-0.17	1.23	1	0.26	0.83
	Neutrophil Count	0.15	11.67	1	0.00	1.16
	Lymphocyte Count	0.08	0.98	1	0.32	1.09
	[Gender=F]	-1.53	5.24	1	0.02	0.21
	[Gender=M]	0^b	.	0	.	.
	Age	-0.10	8.97	1	0.00	0.90
	WBC	-0.09	6.56	1	0.01	0.90
	Haemoglobin	-0.08	0.02	1	0.86	0.92
	Haematocrit	-0.31	3.40	1	0.06	0.72

3	Neutrophil Count	0.17	15.03	1	0.00	1.19
	Lymphocyte Count	0.18	3.87	1	0.04	1.20
	[Gender=F]	-2.03	5.03	1	0.02	0.13
	[Gender=M]	0^b	.	0	.	.

4.4.5 Selection of Variables using a Combination of Wald Test and OR / Exp(B):

This section provides details of dropping variables step by step based on p-value of Wald test and OR/ Exp(B).

Step 1: Dropping MCH

The p-value of Wald test for subtype 1, 2 and 3 are 0.766, 0.720 and 0.816, respectively.

Step 2: Dropping Platelet Count

The p-value of Wald for subtype 1, 2 and 3 are 0.914, 0.548 and 0.343, respectively.

Step 3: Dropping Eosinophil Count

The p-value of Wald test for subtype 1, 2 and 3 are 0.290, 0.176 and 0.565, respectively.

The values of OR of Eosinophil count for subtype 1, 2 and 3 are 3.148, 4.398 and 0.297, respectively.

Step 4: Dropping MCV:

The p-value of Wald test for subtype 1, 2 and 3 are 0.055, 0.094 and 0.328, respectively. MCV has logical OR but it is dropped based on p-value of Wald test.

Step 5: Dropping Haematocrit

The p-value of Wald test for subtype 1, 2 and 3 are 0.513, 0.093 and 0.156, respectively. Haematocrit has logical OR, but it is dropped based on p-value of Wald test.

Step 6: Dropping RBC

The p-value of Wald test for subtype 1, 2 and 3 are 0.770, 0.987 and 0.043, respectively. RBC has logical OR, but it is dropped based on p-value of Wald test.

Step 7: Dropping Lymphocyte Counts:

The p-value of Wald test for subtype 1, 2 and 3 are 0.059, 0.217 and 0.005, respectively. Lymphocyte count has logical OR, but it is dropped based on p-value of Wald test.

Step 8: Dropping WBC

The p-value of Wald test for subtype 1, 2 and 3 are 0.099, 0.736 and 0.934, respectively.

WBC has logical OR, but it is dropped based on p-value of Wald test.

Step 9: Dropping MCHC:

The p-value of Wald test for subtype 1, 2 and 3 are 0.000, 0.034 and 0.653, respectively.

MCHC has logical OR, but it is dropped based on p-value of Wald test.

Step 10: Dropping AGE:

The p-value of Wald test for subtype 1, 2 and 3 are 0.470, 0.010 and 0.012, respectively.

Age has logical OR, but it is dropped based on p-value of Wald test.

Set of Statistically Significant Variables:

Following the procedure of variables selections using a combination of Wald test and OR step by step, we left with 5 variables. Details of these variables, their coefficients, significant values, and OR are present in Table 4.10.

Table 4.10: Set of variables obtained from using combination of Wald test and OR.

Subtypes	Variables	B	Wald	df	p-value	Exp(B)
1	Hemoglobin	-1.11	52.64	1	0.00	0.32
	Neutrophil Count	0.19	17.02	1	0.00	1.21
	Basophil Count	-3.84	13.49	1	0.00	0.02
	Monocyte Count	1.51	8.92	1	0.00	4.53
	[Gender=F]	-2.55	16.00	1	0.00	0.07
	[Gender=M]	0^b	.	0	.	.
2	Hemoglobin	-0.75	25.18	1	0.00	0.46
	Neutrophil Count	0.17	13.59	1	0.00	1.18
	Basophil Count	-2.30	5.08	1	0.02	0.10
	Monocyte Count	1.61	10.15	1	0.00	5.02
	[Gender=F]	-2.66	16.68	1	0.00	0.07
	[Gender=M]	0^b	.	0	.	.
3	Hemoglobin	-1.06	32.16	1	0.00	0.34
	Neutrophil Count	0.18	13.57	1	0.00	1.19
	Basophil Count	-3.39	7.15	1	0.00	0.03
	Monocyte Count	1.47	8.19	1	0.00	4.35
	[Gender=F]	-3.44	19.19	1	0.00	0.03
	[Gender=M]	0^b	.	0	.	.

4.4.6 Summary of Selection of Variables:

Five different combination of methods have been used for the selection of appropriate variables to be used as independent variables in logistic regression modelling. Table 4.11 shows presence or absence of different variables in the final selection using various methods. Final selection of any variables is done based on the criteria that they are successfully shortlisted in at least three methods of selection. Therefore, we finally left with four variables namely:

1- Haemoglobin
2- Neutrophil Count
3- Monocyte Count
4- Gender

Table 4.11: Methods summary

Sr. No	Variables	1st Method	2nd Method	3rd Method	4th Method	5th Method	Selected Variables
		Wald	OR	LRT and Wald	LRT and OR	Wald + OR	
1	**Gender**	✓	✓	×	✓	✓	**4/5**
2	Age	×	✓	×	✓	×	2/5
3	WBC	×	✓	×	✓	×	2/5
4	RBC	×	×	×	×	×	0/5
5	**Hemoglobin**	✓	✓	✓	✓	✓	**5/5**
6	Hematocrit	×	✓	×	✓	×	2/5
7	MCV	×	✓	×	×	×	1/5
8	MCH	×	✓	×	×	×	1/5
9	MCHC	×	×	✓	×	×	1/5
10	Platelet Count	×	✓	×	×	×	1/5
11	**Neutrophil Count**	✓	✓	✓	✓	✓	**5/5**
12	Lymphocyte Count	×	✓	×	✓	×	2/5
13	Basophil Count	✓	×	×	×	✓	2/5
14	Eosinophil Count	×	×	×	×	×	0/5
15	**Monocyte Count**	✓	×	✓	×	✓	**3/5**

OR = Odds Ratio

LRT = Likelihood Ratio Test

4.4.7 Logistic Regression Modelling Using Successful Variables:

Table 4.12 provide the final selected variables. Details of these variables, their coefficients, significant values, and the values of OR.

Table 4.12: Set of final selected variables.

Subtypes	Variables	B	Wald	df	p-value	Exp(B)
1	Intercept	12.32	43.85	1	0.00	
	Hemoglobin	-1.05	50.81	1	0.00	0.34
	Neutrophil Count	0.14	10.57	1	0.00	1.15
	Monocyte Count	1.36	8.15	1	0.00	3.92
	[Gender=F]	-2.04	13.91	1	0.00	0.13
	[Gender=M]	0^b	.	0	.	.
2	Intercept	8.48	20.26	1	0.00	
	Hemoglobin	-0.77	27.14	1	0.00	0.46
	Neutrophil Count	0.14	11.17	1	0.00	1.15
	Monocyte Count	1.46	9.33	1	0.00	4.31
	[Gender=F]	-2.17	14.66	1	0.00	0.11
	[Gender=M]	0^b	.	0	.	.
3	Intercept	10.71	25.26	1	0.00	
	Hemoglobin	-1.02	32.74	1	0.00	0.35
	Neutrophil Count	0.13	9.05	1	0.00	1.14
	Monocyte Count	1.32	7.37	1	0.00	3.75
	[Gender=F]	-2.94	16.76	1	0.00	0.05
	[Gender=M]	0^b	.	0	.	.

Table 4.12 shows that in subtype 1 hemoglobin has 66% less chance in the disease. Neutrophil count has mild effect in the disease. Monocytes count has three times more effect in the disease and gender has female effect. In subtype 2 hemoglobin has 54% less chance in the disease. Neutrophil count has mild effect in the disease. Monocytes count has four times more effect in the disease and gender has female effect. In subtype 3 hemoglobin has 65% less chance in the disease. Neutrophil count has mild effect in the disease. Monocytes count has three times more effect in the disease and gender has female effect.

4.4.8 Model Equations (Eq):

The model equations for subtypes 1, 2 and 3 are mentioned below:

4.4.8.1 Equation for Subtype 1:

$$log\frac{p}{1-p} = -1.05 * Hemoglobin + 0.14 * Neutrophil\ Count + 1.36 * Monocyte\ Count - 2.04 * Gender$$

Eq (4.1)

Eq (4.1) shows that for subtype 1, Hemoglobin has negative effect, Neutrophil count has positive effect, Monocyte count has also positive effect while gender has negative effect.

4.4.8.2 Equation for Subtype 2:

$$log\frac{p}{1-p} = -0.77 * Hemoglobin + 0.14 * Neutrophil\ Count + 1.46 * Monocyte\ Count - 2.17 * Gender$$

Eq (4.2)

Eq (4.2) shows that for subtype 2, Hemoglobin has negative effect, Neutrophil count and Monocyte count has also positive effect while gender has negative effect.

4.4.8.3 Equation for Subtype 3:

$$log\frac{p}{1-p} = -1.02 * Hemoglobin + 0.13 * Neutrophil\ Count + 1.32 * Monocyte\ Count - 2.94 * Gender$$

Eq (4.3)

Eq (4.3) shows that for subtype 3, Haemoglobin has negative effect, Neutrophil count has positive effect, Monocyte count has also positive effect and gender has negative effect.

4.5 Model Evaluation:

4.5.1 Normal vs AML:

In case of AML, out of 123 cases 16 cases are predicted as normal while 107 are predicted as diseased.

True Positive: Diseased people correctly identified as diseased. TP = 107

False Negative: Diseased people incorrectly identified as normal. FN = 16

In case of normal, out of 67 cases 10 cases are predicted as diseased while 57 cases are predicted as normal.

True Negative: Normal cases correctly identified as normal. TN = 57

False Positive: Normal cases incorrectly identified as diseased cases. FP = 10

So, in Normal vs AML case the 2 x 2 matrix is:

	Observed Positive	**Observed Negative**
Predicted Positive	TP = 107	FP = 10
Predicted Negative	FN = 16	TN = 57

4.5.1.1 Classification Accuracy:

$$P = \frac{TP + TN}{TP + TN + FP + FN} \qquad Eq\ (4.4)$$

$$P = \frac{107 + 57}{107 + 57 + 10 + 16}$$

$$P = \frac{164}{190}$$

$$P = 0.86$$

In terms of percentage the accuracy is **86%.**

4.5.1.2 Sensitivity:

Sensitivity is the accuracy of positive prediction or the true positive rate.

The formula for calculating sensitivity is:

$$P_p = \frac{TP}{TP + FN} \qquad Eq\ (4.5)$$

$$P_p = \frac{107}{107 + 16}$$

$$P_p = \frac{107}{123}$$

$$P_p = 0.86$$

In terms of percentage the sensitivity is **86%.**

4.5.1.3 Specificity:

Specificity is the accuracy of negative prediction or true negative rate.

The formula for calculating specificity is:

$$P_n = \frac{TN}{TN + FP} \qquad Eq\ (4.6)$$

$$P_n = \frac{57}{57 + 10}$$

$$P_n = \frac{57}{67}$$

$$P_n = 0.85$$

In terms of percentage the specificity is **85%.**

4.5.1.4 Precision or Positive Predicted Value (PPV):

Precision is the hit rate.

The formula for precision is:

$$PPV = \frac{TP}{TP + FP} \qquad Eq\ (4.7)$$

$$PPV = \frac{107}{107 + 10}$$

$$PPV = \frac{107}{117}$$

$$PPV = 0.91$$

Precision percentage is **91%.**

4.5.2 Normal vs CML:

In case of CML, out of 79 cases 7 cases are predicted as normal while 72 are predicted as diseased.

True Positive: Diseased people correctly identified as diseased. TP = 72

False Negative: Diseased people incorrectly identified as normal. FN = 7

In case of normal, out of 67 cases 10 cases are predicted as diseased while 57 cases are predicted as normal.

True Negative: Normal cases correctly identified as normal. TN = 57

False Positive: Normal cases incorrectly identified as diseased cases. FP = 10

So, in Normal vs CML case the 2 x 2 matrix is:

	Observed Positive	**Observed Negative**
Predicted Positive	TP = 72	FP = 10
Predicted Negative	FN = 7	TN = 57

4.5.2.1 Classification Accuracy:

$$P = \frac{TP + TN}{TP + TN + FP + FN} \qquad Eq\ (4.8)$$

$$P = \frac{72 + 57}{72 + 57 + 10 + 7}$$

$$P = \frac{129}{146}$$

$$P = 0.88$$

In terms of percentage the accuracy is **88%.**

4.5.2.2 Sensitivity:

The true positive rate is:

$$P_p = \frac{TP}{TP + FN} \qquad Eq\ (4.9)$$

$$P_p = \frac{72}{72 + 7}$$

$$P_p = \frac{72}{79}$$

$$P_p = 0.91$$

The percentage for sensitivity is **91%.**

4.5.2.3 Specificity:

The true negative rate is:

$$P_n = \frac{TN}{TN + FP} \qquad Eq\ (4.10)$$

$$P_n = \frac{57}{57 + 10}$$

$$P_n = \frac{57}{67}$$

$$P_n = 0.85$$

The percentage of specificity is **85%.**

4.5.2.4 Precision or Positive Predicted Value (PPV):

Positive predicted value is calculated as:

$$PPV = \frac{TP}{TP + FP} \qquad Eq\ (4.11)$$

$$PPV = \frac{72}{72 + 10}$$

$$PPV = \frac{72}{82}$$

$$PPV = 0.87$$

The percentage of precision is **87%.**

4.5.3 Normal vs ALL:

In case of ALL, out of 18 cases no case is predicted as normal.

True Positive: Diseased people correctly identified as diseased. TP = 18

False Negative: Diseased people incorrectly identified as normal. FN = 0

In case of normal, out of 67 cases 10 cases are predicted as diseased while 57 cases are predicted as normal.

True Negative: Normal cases correctly identified as normal. TN = 57

False Positive: Normal cases incorrectly identified as diseased cases. FP = 10

So, in Normal vs ALL case the 2 x 2 matrix is:

	Observed Positive	**Observed Negative**
Predicted Positive	TP = 18	FP = 10
Predicted Negative	FN = 0	TN = 57

4.5.3.1 Classification Accuracy:

$$P = \frac{TP + TN}{TP + TN + FP + FN} \qquad Eq\ (4.12)$$

$$P = \frac{18 + 57}{18 + 57 + 10 + 0}$$

$$P = \frac{75}{85}$$

$$P = 0.88$$

In terms of percentage the accuracy is **88%.**

4.5.3.2 Sensitivity:

The true positive rate is:

$$P_p = \frac{TP}{TP + FN} \qquad Eq\ (4.13)$$

$$P_p = \frac{18}{18 + 0}$$

$$P_p = \frac{18}{18}$$

$$P_p = 1$$

The percentage of sensitivity is **100%.**

4.5.3.3 Specificity:

The true negative rate is:

$$P_n = \frac{TN}{TN + FP} \qquad Eq\ (4.14)$$

$$P_n = \frac{57}{57 + 10}$$

$$P_n = \frac{57}{67}$$

$$P_n = 0.85$$

The percentage of specificity is **85%.**

4.5.3.4 Precision or Positive Predicted Value (PPV):

Positive predicted value is calculated as:

$$PPV = \frac{TP}{TP + FP} \qquad Eq\ (4.15)$$

$$PPV = \frac{18}{18 + 10}$$

$$PPV = \frac{18}{28}$$

$$PPV = 0.64$$

The percentage of precision is **64%.**

4.6 Performance Evaluation Summary:

The performance evaluation summary in terms of percentage is shown in table 4.13.

Table 4.13: Summary of model evaluation.

S. No.	**Models**	**Accuracy Percentage**	**Sensitivity Percentage**	**Specificity Percentage**	**Precision Percentage**
1	Normal vs AML	86	86	85	91
2	Normal vs CML	88	91	85	87
3	Normal vs ALL	88	100	85	64

CONCLUSIONS

One of the main objectives of this research is to analyses the general trends and tendencies of various characteristics of CBC reports by comparing Leukemic subtypes cases and non-Leukemic (normal) cases. Another objective is to develop a predictive model based on significant characteristics of CBC reports for the screening of Leukemic subtypes cases or non-Leukemic (normal) cases.

Few of the major conclusions are described below:

I. Out of 21 variables in CBC report, 15 variables are selected for the analysis by dropping the information of percentages of various variables to avoid duplication.

II. Descriptive analysis shows variations in the values of mean for Normal vs Three subtypes of leukemia.

III. Comparative analysis shows that only MCH has statistically insignificant difference between the means of normal, AML, CML and ALL.

IV. For the development of MLR model, five different combination of methods have been used for the selection of appropriate variables to be used as independent variables in logistic regression modelling. Final selected variables based on these methods are haemoglobin, neutrophil count, monocyte count and gender.

V. The assessment analysis shows that in case of Normal vs AML the accuracy is 86%, sensitivity is 86%, specificity is 85% and precision is 91%. For Normal vs CML the accuracy is 88%, sensitivity is 91%, specificity is 85% and precision is 87%. For Normal vs ALL the accuracy is 88%, sensitivity is 100%, specificity is 85% and precision is 64%.

These findings suggest that the developed model can be trusted for the subjective screening of disease, i.e., leukemia or its subtypes. It is worth noting that the proposed model is not meant to take the place of traditional leukemia diagnosis tests such as bone marrow biopsy, lumber puncture, flow cytometry, and so on. It provides basic technical support for the objective screening of patients using data driven models. Therefore, a combination of subjective and objective assessment can improve the quality of diagnosis of leukemia or its subtypes at early stage.

Limitations of the Study:

This study has following limitations:

I. There was class difference between the three subtypes of leukemia. In our study only 18 cases of ALL were present as compared to AML and CML. As AML has 123 cases and CML has 79 cases.

II. In this study there is no validation through external data.

Future Recommendations:

The future recommendations related to this research are:

a) More data will be collected for the analysis.
b) The class imbalance between the data will be removed by adding more data.
c) Cluster analysis will be performed between the variables of the CBC report.
d) Other machine learning models will be used for the predictive modelling of the disease, i.e., leukemia and its subtypes.

REFERENCES

[1] W.-L. Hsu *et al.*, “The Incidence of Leukemia, Lymphoma and Multiple Myeloma among Atomic Bomb Survivors: 1950–2001,” *Radiat. Res.*, vol. 179, no. 3, p. 361, 2013.

[2] M. Trendowski, “The inherent metastasis of leukaemia and its exploitation by sonodynamic therapy,” *Crit. Rev. Oncol. Hematol.*, vol. 94, no. 2, pp. 149–163, 2015.

[3] J. Wu, J. E. Fantasia, and R. Kaplan, “Oral Manifestations of Acute Myelomonocytic Leukemia: A Case Report and Review of the Classification of Leukemias,” *J. Periodontol.*, vol. 73, no. 6, pp. 664–668, 2002.

[4] A. S. Davis, A. J. Viera, and M. D. Mead, “Leukemia: An overview for primary care,” *Am. Fam. Physician*, vol. 89, no. 9, pp. 731–738, 2014.

[5] A. Hakeem *et al.*, “Clinical presentation and outcomes of drug-eluting stent-associated coronary aneurysms,” *EuroIntervention*, vol. 7, no. 4, pp. 487–496, 2011.

[6] A. H. Munir and M. I. Khan, “Pattern of basic hematological parameters in acute and chronic leukemias,” *J. Med. Sci.*, vol. 27, no. 2, pp. 125–129, 2019.

[7] N. F. Grigoropoulos, R. Petter, M. B. Van’T Veer, M. A. Scott, and G. A. Follows, “Leukaemia update. Part 1: Diagnosis and management,” *BMJ*, vol. 346, no. 7902, pp. 29–32, 2013.

[8] A. M. Almeida and F. Ramos, “Acute myeloid leukemia in the older adults,” *Leuk. Res. Reports*, vol. 6, pp. 1–7, 2016.

[9] S. J. Mishra and A. P. Deshmukh, “Detection of Leukemia Using Matlab,” *Int. J. Adv. Res. Electron. Commun. Eng.*, vol. 4, no. 2, pp. 2–6, 2015.

[10] H. Döhner, D. J. Weisdorf, and C. D. Bloomfield, “Acute myeloid leukemia,” *N. Engl. J. Med.*, vol. 373, no. 12, pp. 1136–1152, 2015.

[11] C. Murdoch, M. Muthana, S. B. Coffelt, and C. E. Lewis, “The role of myeloid cells in the promotion of tumour angiogenesis,” *Nat. Rev. Cancer*, vol. 8, no. 8, pp. 618–631, 2008.

[12] I. Hirji *et al.*, “Chronic myeloid leukemia (CML): Association of treatment satisfaction, negative medication experience and treatment restrictions with health outcomes, from the patient’s perspective,” *Health Qual. Life Outcomes*, vol. 11, no. 1, pp. 1–11, 2013.

[13] R. Bhatia, “Chronic Myeloid Leukemia,” *Hematol. Basic Princ. Pract.*, pp. 1055–1070, 2018.

[14] M. Höglund, F. Sandin, and B. Simonsson, “Epidemiology of chronic myeloid leukaemia: an update,” *Ann. Hematol.*, vol. 94, no. 2, pp. 241–247, 2015.

[15] M. Höglund *et al.*, “Tyrosine kinase inhibitor usage, treatment outcome, and prognostic scores in CML: Report from the population-based Swedish CML registry,” *Blood*, vol. 122, no. 7, pp. 1284–1292, 2013.

[16] K. Dalziel, A. Round, K. Stein, R. Garside, and A. Price, “Effectiveness and cost-effectiveness of imatinib for treatment of chronic myeloid leukaemia,” *Heal. Technol. assesment*, vol. 8, no. 28, pp. 1–131, 2004.

[17] J. W. Vardiman, N. L. Harris, and R. D. Brunning, “The World Health Organization (WHO) classification of the myeloid neoplasms,” *Blood*, vol. 100, no. 7, pp. 2292–2302, 2002.

[18] M. Bonifacio, F. Stagno, L. Scaffidi, M. Krampera, and F. Di Raimondo, “Management of chronic myeloid leukemia in advanced phase,” *Front. Oncol.*, vol. 9, no. OCT, pp. 1–18, 2019.

[19] D. G. Savage, R. M. Szydlo, and J. M. Goldman, “Clinical features at diagnosis in 430 patients with chronic myeloid leukaemia seen at a referral centre over a 16-year period,” *Br. J. Haematol.*, vol. 96, no. 1, pp. 111–116, 1997.

[20] R. Ebrahem, B. Ahmed, S. Kadhem, and Q. Truong, “Chronic Myeloid Leukemia: A Case of Extreme Thrombocytosis Causing Syncope and Myocardial Infarction,” *Cureus*, vol. 8, no. 2, pp. 8–11, 2016.

[21] T. Terwilliger and M. Abdul-Hay, “Acute lymphoblastic leukemia: a comprehensive review and 2017 update,” *Blood Cancer J.*, vol. 7, no. 6, p. e577, 2017.

[22] R. Snodgrass, L. T. Nguyen, M. Guo, M. Vaska, C. Naugler, and F. Rashid-Kolvear, “Incidence of acute lymphocytic leukemia in Calgary, Alberta, Canada: A retrospective cohort study,” *BMC Res. Notes*, vol. 11, no. 1, pp. 9–12, 2018.

[23] Z. Fadoo *et al.*, “Clinical Features and Induction Outcome of Childhood Acute LymphoblasticLeukemia in a Lower/Middle Income Population: A Multi-Institutional Report FromPakistan,” *Pediatr. Blood Cancer*, vol. 62, no. 10, pp. 1700–1708, 2015.

[24] N. Yasmeen and S. Ashraf, “Childhood acute lymphoblastic leukaemia; epidemiology and clinicopathological features,” *J. Pak. Med. Assoc.*, vol. 59, no. 3, pp. 150–153, 2009.

[25] F. Ceppi, G. Cazzaniga, A. Colombini, A. Biondi, and V. Conter, "Risk factors for relapse in childhood acute lymphoblastic leukemia: Prediction and prevention," *Expert Rev. Hematol.*, vol. 8, no. 1, pp. 57–70, 2015.

[26] X. . Luo, Z. . Ke, L. . Huang, X. . Guan, Y. . Zhang, and X. . Zhang, "High-Risk Childhood Acute Lymphoblastic Leukemia in China: Factors Influencing the Treatment and Outcome," *Pediatr. Blood Cancer*, vol. 52, no. 2, pp. 191–195, 2009.

[27] D. Szczepanek *et al.*, "Central nervous involvement by chronic lymphocytic leukaemia," *Neurol. Neurochir. Pol.*, vol. 52, no. 2, pp. 228–234, 2018.

[28] A. Ferrajoli, "Treatment of younger patients with chronic lymphocytic leukemia.," *Hematology Am. Soc. Hematol. Educ. Program*, vol. 2010, pp. 82–89, 2010.

[29] M. Hallek, "Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment," *Am. J. Hematol.*, vol. 94, no. 11, pp. 1266–1287, 2019.

[30] A. Khalade, M. S. Jaakkola, E. Pukkala, and J. J. K. Jaakkola, "Exposure to benzene at work and the risk of leukemia: A systematic review and meta-analysis," *Environ. Heal. A Glob. Access Sci. Source*, vol. 9, no. 1, pp. 1–8, 2010.

[31] P. Buffler, M. Kwan, P. Reynolds, and K. Urayama, "Environmental and Genetic Risk Factors for Childhood Leukemia: Appraising the Evidence," *Cancer Invest.*, vol. 23, no. 1, pp. 60–75, 2005.

[32] O. Raaschou-Nielsen, U. A. Hvidtfeldt, N. Roswall, O. Hertel, A. H. Poulsen, and M. Sørensen, "Ambient benzene at the residence and risk for subtypes of childhood leukemia, lymphoma and CNS tumor," *Int. J. Cancer*, vol. 143, no. 6, pp. 1367–1373, 2018.

[33] L. Diller, "Adult primary care after childhood acute lymphoblastic leukemia," *N. Engl. J. Med.*, vol. 365, no. 15, pp. 1417–1424, 2011.

[34] S. E. Puumala, J. A. Ross, R. Aplenc, and L. G. Spector, "Epidemiology of childhood acute myeloid leukemia," *Pediatr. Blood Cancer*, vol. 60, no. 5, pp. 728–733, 2013.

[35] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global Cancer Statistics , 2002," *CA. Cancer J. Clin.*, vol. 55, no. 2, pp. 74–108, 2005.

[36] R. Ruchlemer and A. Polliack, "Geography, ethnicity and 'roots' in chronic lymphocytic leukemia.," *Leuk. Lymphoma*, vol. 54, no. 6, pp. 1142–1150, 2013.

[37] R. Frazer, A. E. Irvine, and M. F. McMullin, "Chronic myeloid leukaemia in the 21st century," *Ulster Med. J.*, vol. 76, no. 1, pp. 8–17, 2007.

[38] N. Maksimovic *et al.*, “Incidence and mortality patterns of acute myeloid leukemia in Belgrade, Serbia (1999-2013),” *Med.*, vol. 54, no. 1, pp. 1–8, 2018.

[39] G. N. Kakepoto, I. A. Burney, S. Zaki, S. N. Adil, and M. Khurshid, “Long-term outcomes of Acute Myeloid Leukemia in adults in Pakistan,” *J. Pak. Med. Assoc.*, vol. 52, no. 10, pp. 482–486, 2002.

[40] A. Khalid, M. Zahid, A. Rehman, Z. U. Ahmad, S. Qazi, and Z. Aziz, “Clinicoepidemiological features of adult leukemias in Pakistan.,” *J. Pak. Med. Assoc.*, vol. 47, no. 4, pp. 119–122, 1997.

[41] J. L. Davis and J. F. Murray, “History and Physical Examination,” *Murray Nadel’s Textb. Respir. Med.*, p. 263, 2016.

[42] M. Kato *et al.*, “Case series of pediatric acute leukemia without a peripheral blood abnormality, detected by magnetic resonance imaging,” *Int. J. Hematol.*, vol. 93, no. 6, pp. 787–790, 2011.

[43] M. Hasani and A. Hanani, “Automated Diagnosis of Iron Deficiency Anemia and Thalassemia by Data Mining Techniques,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 4, pp. 326–331, 2017.

[44] L. Dean, *Blood Groups and Red Cell Antigens*, vol. 2, no. Bethesda, Md, USA. 2005.

[45] B. S. Lee *et al.*, “Fully integrated lab-on-a-disc for simultaneous analysis of biochemistry and immunoassay from whole blood,” *Lab Chip*, vol. 11, no. 1, pp. 70–78, 2011.

[46] A. Ahuja *et al.*, “Comparison of Immunohistochemistry, Cytochemistry, and Flow Cytometry in AML for Myeloperoxidase Detection,” *Indian J. Hematol. Blood Transfus.*, vol. 34, no. 2, pp. 233–239, 2018.

[47] E. . Akanni and A. Palini, “Immunophenotyping of Peripheral Blood and Bone Marrow Cells by Flow Cytometry,” *Ejifcc*, vol. 17, no. 1, p. 17, 2006.

[48] S. Lindström and H. Andersson-Svahn, “Flow Cytometry and Microscopy as Means of Studying Single Cells: A Short Introductional Overview,” *Single-Cell Anal.*, vol. 853, pp. 13–15, 2012.

[49] B. J. Bain, “Routine and specialised techniques in the diagnosis of haematological neoplasms,” *J. Clin. Pathol.*, vol. 48, no. 6, pp. 501–508, 1995.

[50] M. Riegel, “Human molecular cytogenetics: From cells to nucleotides,” *Genet. Mol. Biol.*, vol. 37, no. 1, pp. 194–209, 2014.

[51] B. J. Bain, “Bone marrow aspiration,” *J. Clin. Pathol.*, vol. 54, no. 9, pp. 657–663, 2001.

[52] M. L. Schreiber, “Lumbar Puncture,” *Medsurg Nurs.*, vol. 28, no. 6, pp. 402–404, 2019.

[53] J. R. Berenson *et al.*, “Using a Powered Bone Marrow Biopsy System Results in Shorter Procedures, Causes Less Residual Pain to Adult Patients, and Yields Larger Specimens,” *Diagn. Pathol.*, vol. 6, no. 1, p. 23, 2011.

[54] N. Hjortholm, E. Jaddini, K. Hałaburda, and E. Snarski, “Strategies of pain reduction during the bone marrow biopsy,” *Ann. Hematol.*, vol. 92, no. 2, pp. 145–149, 2013.

[55] R. A. Oster and F. T. Enders, “The Importance of Statistical Competencies for Medical Research Learners,” *J. Stat. Educ.*, vol. 26, no. 2, pp. 137–142, 2018.

[56] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew, “Artificial intelligence in medicine,” *Ann. R. Coll. Surg. Engl.*, vol. 86, no. 5, pp. 334–338, 2004.

[57] H. Abedy, F. Ahmed, M. N. Qaisar Bhuiyan, M. Islam, M. N. Ali, and M. Shamsujjoha, “Leukemia Prediction from Microscopic Images of Human Blood Cell Using HOG Feature Descriptor and Logistic Regression,” *Int. Conf. ICT Knowl. Eng.*, vol. 2018-Novem, pp. 7–12, 2019.

[58] R. Bhattacharjee and M. Chakraborty, “LPG-PCA algorithm and selective thresholding based automated method: ALL & AML blast cells detection and counting,” *Proc. 2012 Int. Conf. Commun. Devices Intell. Syst. CODIS 2012*, vol. 9, pp. 109–112, 2012.

[59] T. Markiewicz, S. Osowski, B. Marianska, and L. Moszczyński, “Automatic recognition of the blood cells of myelogenous leukemia using SVM,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 4, pp. 2496–2501, 2005.

[60] S. Shafique and S. Tehsin, “Computer-Aided Diagnosis of Acute Lymphoblastic Leukaemia,” *Comput. Math. Methods Med.*, vol. 2018, 2018.

[61] E. Fathi, M. J. Rezaee, R. Tavakkoli-Moghaddam, A. Alizadeh, and A. Montazer, “Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning,” *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, vol. 234, no. 10, pp. 1051–1069, 2020.

[62] S. Syed-Abdul *et al.*, “Artificial Intelligence based Models for Screening of Hematologic Malignancies using Cell Population Data,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020.

[63] R. Rathee, M. Vashist, A. Kumar, and S. Singh, "Incidence of acute and chronic forms of leukemia in Haryana," *Int. J. Pharm. Pharm. Sci.*, vol. 6, no. 2, pp. 323–325, 2014.

[64] F. Moussavi, S. Hosseini, S. Saket, and H. Derakhshanfar, "the First Cbc in Diagnosis of Childhood Acute Lymphoblastic Leukemia," *Int. J. Med. Investig.*, vol. 3, no. 1, pp. 0–0, 2014.

[65] R. Naeem, S. Naeem, A. Sharif, H. Rafique, and A. Naveed, "Acute Myeloid Leukemia; Demographic Features and Frequency of Various Subtypes in Adult Age Group," *Prof. Med. J.*, vol. 28, no. 2, pp. 298–301, 2016.

[66] T. M. Khan, "Pattern Of Leukaemia Patients Admitted In Ayub Teaching Hospital Abbottabad," *J. Ayub Med. Coll. Abbottabad*, vol. 28, no. 2, pp. 298–301, 2016.

[67] C. Farzana and S. Tahir Sultan, "Clinical and Hematological Profile of Acute Myeloid Leukemia (AML) Patients of Sindh," *J. Hematol. Thromboembolic Dis.*, vol. 04, no. 02, 2016.

[68] I. H. Alshami and A. M. Alhalees, "Automated Diagnosis of Thalassemia Based on DataMining Classifiers," *Int. Conf. Informatics Appl.*, no. June, pp. 440–445, 2012.

[69] M. Abdullah and S. Al-Asmari, "Anemia types prediction based on data mining classification algorithms," *Commun. Manag. Inf. Technol. - Proc. Int. Conf. Commun. Manag. Inf. Technol. ICCMIT 2016*, no. November, pp. 615–621, 2017.

[70] N. Nasim, K. Malik, N. A. Malik, S. Mobeen, S. Awan, and N. Mazhar, "Investigation on the Prevalence of Leukaemia At a Tertiary Care Hospital , Lahore," *Biomedica*, vol. 29, no. 1, pp. 19–22, 2013.

[71] S. Ahmad *et al.*, "Prevalence of acute and chronic forms of leukemia in various regions of Khyber Pakhtunkhwa, Pakistan: Needs much more to be done!," *Bangladesh J. Med. Sci.*, vol. 18, no. 2, pp. 222–227, 2019.

[72] Z. Ali and S. B. Bhaskar, "Basic statistical tools in research and data analysis," *Indian J. Anaesth.*, vol. 60, no. 9, pp. 662–669, 2016.

[73] P. Mishra, C. M. Pandey, U. Singh, and A. Gupta, "Scales of measurement and presentation of statistical data," *Ann. Card. Anaesth.*, vol. 21, no. 4, pp. 419–422, 2018.

[74] L. Abdulwahab, Z. M. Dahalin, and M. B. Galadima, "Data screening and preliminary analysis of the determinants of user acceptance of telecentre," *J. Inf. Syst. New Paradig.*, vol. 1, no. 1, pp. 11–23, 2011.

[75] H. Kang, “The prevention and handling of the missing data,” *Korean J. Anesthesiol.*, vol. 64, no. 5, pp. 402–406, 2013.

[76] Y. Dong and C. Y. J. Peng, “Principled missing data methods for researchers,” *Springerplus*, vol. 2, no. 1, p. 222, 2013.

[77] G. Luczynska, F. Pena-Pereira, M. Tobiszewski, and J. Namiesnik, “Expectation-maximization model for substitution of missing values characterizing greenness of organic solvents,” *Molecules*, vol. 23, no. 6, pp. 1–9, 2018.

[78] H. Akoglu, “User’s guide to correlation coefficients,” *Turkish J. Emerg. Med.*, vol. 18, no. 3, pp. 91–93, 2018.

[79] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, “Developing prediction models for clinical use using logistic regression: An overview,” *J. Thorac. Dis.*, vol. 11, no. Suppl 4, pp. S574–S584, 2019.

[80] A. Schneider, G. Hommel, and M. Blettner, “Linear regression analysis: part 14 of a series on evaluation of scientific publications,” *Dtsch. Arztebl. Int.*, vol. 107, no. 44, pp. 776–782, 2010.

[81] J. Tolles and W. J. Meurer, “Logistic regression: Relating patient characteristics to outcomes,” *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 5, pp. 533–534, 2016.

[82] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, “Comparing different supervised machine learning algorithms for disease prediction,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–16, 2019.

[83] S. Sperandei, “Understanding logistic regression analysis,” *Biochem. Medica*, vol. 24, no. 1, pp. 12–18, 2014.

[84] J. C. Stoltzfus, “Logistic regression: A brief primer,” *Acad. Emerg. Med.*, vol. 18, no. 10, pp. 1099–1104, 2011.

[85] N. Šarlija, A. Bilandžic, and M. Stanic, “Logistic regression modelling: Procedures and pitfalls in developing and interpreting prediction models,” *Croat. Oper. Res. Rev.*, vol. 8, no. 2, pp. 631–652, 2017.

[86] A. Bayaga, “Multinomial Logistic Regression: Usage and Application in Risk Analysis,” *J. Appl. Quant. methods*, vol. 5, no. 2, pp. 288–297, 2010.

[87] F. Huang, S. Wang, and C. C. Chan, “Predicting disease by using data mining based on healthcare information system,” *Proc. - 2012 IEEE Int. Conf. Granul. Comput. GrC 2012*, pp. 191–194, 2012.

[88] R. Trevethan, “Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice,” *Front. Public Heal.*, vol. 5, no. November, pp. 1–7, 2017.

[89] G. Byrne, “A statistical primer: Understanding descriptive and inferential statistics,” *Evid. Based Libr. Inf. Pract.*, vol. 2, no. 1, pp. 32–47, 2007.

[90] D. W. Gudicha, V. D. Schmittmann, and J. K. Vermunt, “Statistical power of likelihood ratio and Wald tests in latent class models with covariates,” *Behav. Res. Methods*, vol. 49, no. 5, pp. 1824–1837, 2017.

[91] M. Szumilas, “Explaining Odds Ratios,” *J. Can. Acad. child Adolesc. psychiatry*, vol. 19, no. 3, p. 227, 2010.

APPENDIX

ANOVA Tables:

ONE-WAY ANOVA

NORMAL_Age, AML_Age, CML_Age, ALL_Age:

Source	DF	SS	MS	F	P
Factor	3	15348	5116	13.96	0.00
Error	283	103689	366		
Total	286	119037			

NORMAL_WBC, AML_WBC, CML_WBC, ALL_WBC:

Source	DF	SS	MS	F	P
Factor	3	434887	144962	26.24	0.00
Error	283	1563234	5524		
Total	286	1998121			

NORMAL_RBC, AML_RBC, CML_RBC, ALL_RBC:

Source	DF	SS	MS	F	P
Factor	3	70.479	23.493	38.56	0.00
Error	283	172.415	0.609		
Total	286	242.894			

NORMAL_Haemoglobin, AML_Haemoglobin, CML_Haemoglobin, ALL_Haemoglobin:

Source	DF	SS	MS	F	P
Factor	3	582.58	194.19	44.75	0.00
Error	283	1228.17	4.34		
Total	286	1810.75			

NORMAL_Haematocrit, AML_Haematocrit, CML_Haematocrit, ALL_Haematocrit:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	6029.7	2009.9	54.57	0.00
Error	283	10422.6	36.8		
Total	286	16452.3			

NORMAL_MCV, AML_MCV, CML_MCV, ALL_MCV:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	522.4	174.1	2.78	0.04
Error	283	17694.7	62.5		
Total	286	18217.1			

NORMAL_MCH, AML_MCH, CML_MCH, ALL_MCH:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	23.19	7.73	0.82	0.48
Error	283	2666.30	9.42		
Total	286	2689.48			

NORMAL_MCHC, AML_MCHC, CML_MCHC, ALL_MCHC:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	224.75	74.92	22.77	0.00
Error	283	931.16	3.29		
Total	286	1155.90			

NORMAL_Platetelet Count, AML_Platelet Count, CML_Platelet Count, ALL_Platelet Count:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	482156	160719	8.56	0.00
Error	283	5310918	18766		
Total	286	5793074			

NORMAL_Neutrophil Count, AML_Neutrophil Count, CML_Neutrophil Count, ALL_Neutrophil Count:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	258474	86158	13.55	0.00
Error	283	1798843	6356		
Total	286	2057316			

NORMAL_lymph_Count, AML_lymph_Count, CML_lymph_Count, ALL_lymph_Count:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	5102	1701	6.31	0.00
Error	283	76257	269		
Total	286	81360			

NORMAL_Basophil Count, AML_Basophil Count, CML_Basophil Count, ALL_Basophil Count:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	152.66	50.89	16.33	0.00
Error	283	881.85	3.12		
Total	286	1034.51			

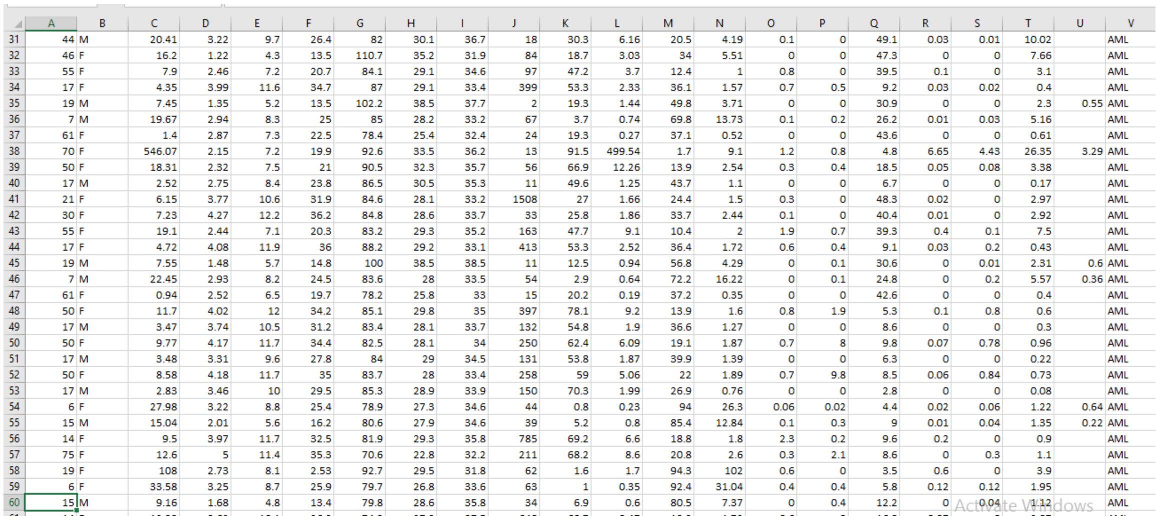
NORMAL_Eosinophil Count, AML_Eosinophil Count, CML_Eosinophil Count, ALL_Eosinophil Count:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	143.72	47.91	13.33	0.00
Error	283	1017.22	3.59		
Total	286	1160.95			

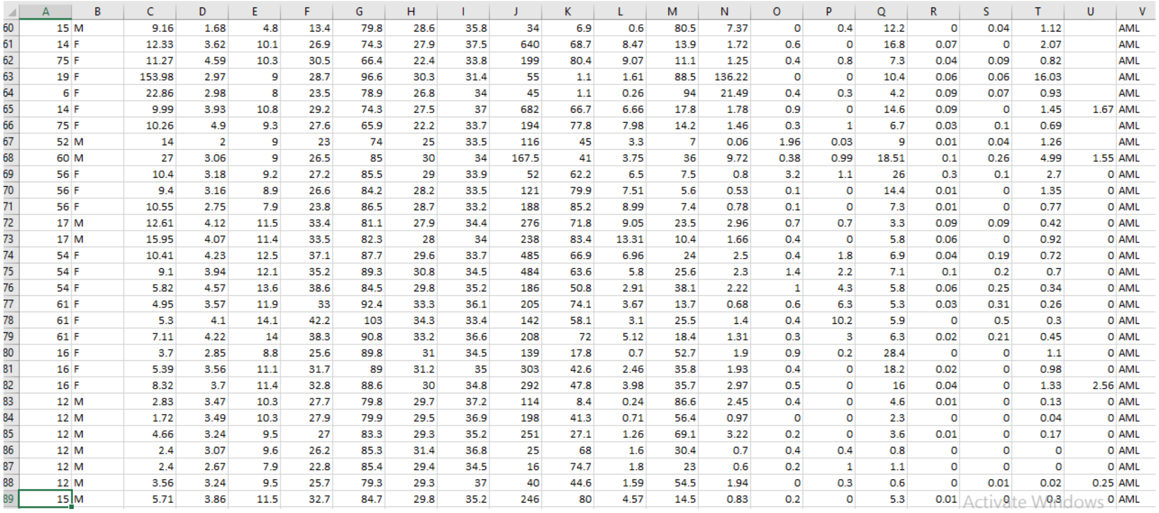
NORMAL_Monocyte Count, AML_Monocyte Count, CML_Monocyte Count, ALL_Monocyte Count:

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor	3	6504.4	2168.1	22.17	0.00
Error	283	27673.3	97.8		
Total	286	34177.7			

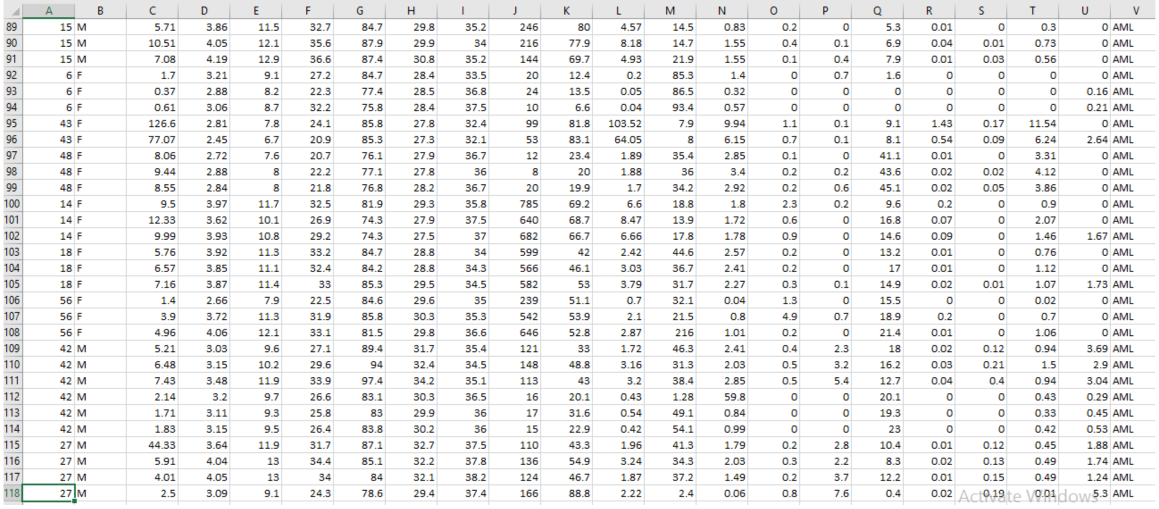
DATASET:



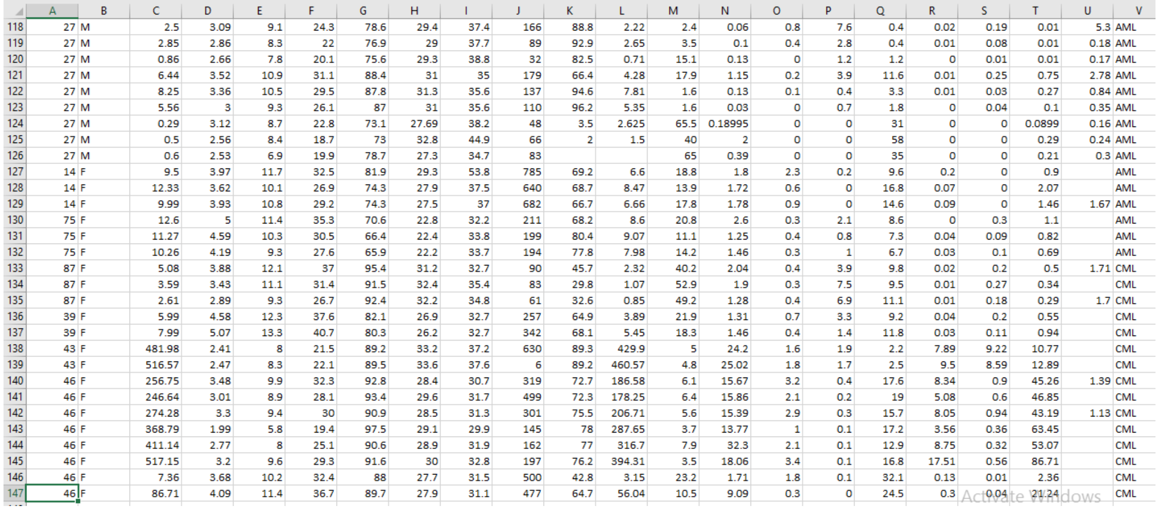
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
31	44	M	20.41	3.22	9.7	26.4	82	30.1	36.7	18	30.3	6.16	20.5	4.19	0.1	0	49.1	0.03	0.01	10.02		AML
32	46	F	16.2	1.22	4.3	13.5	110.7	35.2	31.9	84	18.7	3.03	34	5.51	0	0	47.3	0	0	7.66		AML
33	55	F	7.9	2.46	7.2	20.7	84.1	29.1	34.6	97	47.2	3.7	12.4	1	0.8	0	39.5	0.1	0	3.1		AML
34	17	F	4.35	3.99	11.6	34.7	87	29.1	33.4	399	53.3	2.33	36.1	1.57	0.7	0.5	9.2	0.03	0.02	0.4		AML
35	19	M	7.45	1.35	5.2	13.5	102.2	38.5	37.7	2	19.3	1.44	49.8	3.71	0	0	30.9	0	0	2.3	0.55	AML
36	7	M	19.67	2.94	8.3	25	85	28.2	33.2	67	3.7	0.74	69.8	13.73	0.1	0.2	26.2	0.01	0.03	5.16		AML
37	61	F	1.4	2.87	7.3	22.5	78.4	25.4	32.4	24	19.3	0.27	37.1	0.52	0	0	43.6	0	0	0.61		AML
38	70	F	546.07	2.15	7.2	19.9	92.6	33.5	36.2	13	91.5	499.54	1.7	9.1	1.2	0.8	4.8	6.65	4.43	26.35	3.29	AML
39	50	F	18.31	2.32	7.5	21	90.5	32.3	35.7	56	66.9	12.26	13.9	2.54	0.3	0.4	18.5	0.05	0.08	3.38		AML
40	17	M	2.52	2.75	8.4	23.8	86.5	30.5	35.3	11	49.6	1.25	43.7	1.1	0	0	6.7	0	0	0.17		AML
41	21	F	6.15	3.77	10.6	31.9	84.6	28.1	33.2	1508	27	1.66	24.4	1.5	0.3	0	48.3	0.02	0	2.97		AML
42	30	F	7.23	4.27	12.2	36.2	84.8	28.6	33.7	33	25.8	1.86	33.7	2.44	0.1	0	40.4	0.01	0	2.92		AML
43	55	F	19.1	2.44	7.1	20.3	83.2	29.3	35.2	163	47.7	9.1	10.4	2	1.9	0.7	39.3	0.4	0.1	7.5		AML
44	17	F	4.72	4.08	11.9	36	88.2	29.2	33.1	413	53.3	2.52	36.4	1.72	0.6	0.4	9.1	0.03	0.2	0.43		AML
45	19	M	7.55	1.48	5.7	14.8	100	38.5	38.5	11	12.5	0.94	56.8	4.29	0	0.1	30.6	0	0.01	2.31	0.6	AML
46	7	M	22.45	2.93	8.2	24.5	83.6	28	33.5	54	2.9	0.64	72.2	16.22	0	0.1	24.8	0	0.2	5.57	0.36	AML
47	61	F	0.94	2.52	6.5	19.7	78.2	25.8	33	15	20.2	0.19	37.2	0.35	0	0	42.6	0	0	0.4		AML
48	50	F	11.7	4.02	12	34.2	85.1	29.8	35	397	78.1	9.2	13.9	1.6	0.8	1.9	5.3	0.1	0.8	0.6		AML
49	17	M	3.47	3.74	10.5	31.2	83.4	28.1	33.7	132	54.8	1.9	36.6	1.27	0	0	8.6	0	0	0.3		AML
50	50	F	9.77	4.17	11.7	34.4	82.5	28.1	34	250	62.4	6.09	19.1	1.87	0.7	8	9.8	0.07	0.78	0.96		AML
51	17	M	3.48	3.31	9.6	27.8	84	29	34.5	131	53.8	1.87	39.9	1.39	0	0	6.3	0	0	0.22		AML
52	50	F	8.58	4.18	11.7	35	83.7	28	33.4	258	59	5.06	22	1.89	0.7	9.8	8.5	0.06	0.84	0.73		AML
53	17	M	2.83	3.46	10	29.5	85.3	28.9	33.9	150	70.3	1.99	26.9	0.76	0	0	2.8	0	0	0.08		AML
54	6	F	27.98	3.22	8.8	25.4	78.9	27.3	34.6	44	0.8	0.23	94	26.3	0.06	0.02	4.4	0.02	0.06	1.22	0.64	AML
55	15	M	15.04	2.01	5.6	16.2	80.6	27.9	34.6	39	5.2	0.8	85.4	12.84	0.1	0.3	9	0.01	0.04	1.35	0.22	AML
56	14	F	9.5	3.97	11.7	32.5	81.9	29.3	35.8	785	69.2	6.6	18.8	1.8	2.3	0.2	9.6	0.2	0	0.9		AML
57	75	F	12.6	5	11.4	35.3	70.6	22.8	32.2	211	68.2	8.6	20.8	2.6	0.3	2.1	8.6	0	0.3	1.1		AML
58	19	F	108	2.73	8.1	2.53	92.7	29.5	31.8	62	1.6	1.7	94.3	102	0.6	0	3.5	0.6	0	3.9		AML
59	6	F	33.58	3.25	8.7	25.9	79.7	26.8	33.6	63	1	0.35	92.4	31.04	0.4	0.4	5.8	0.12	0.12	1.95		AML
60	15	M	9.16	1.68	4.8	13.4	79.8	28.6	35.8	34	6.9	0.6	80.5	7.37	0	0.4	12.2	0	0.04	1.12		AML



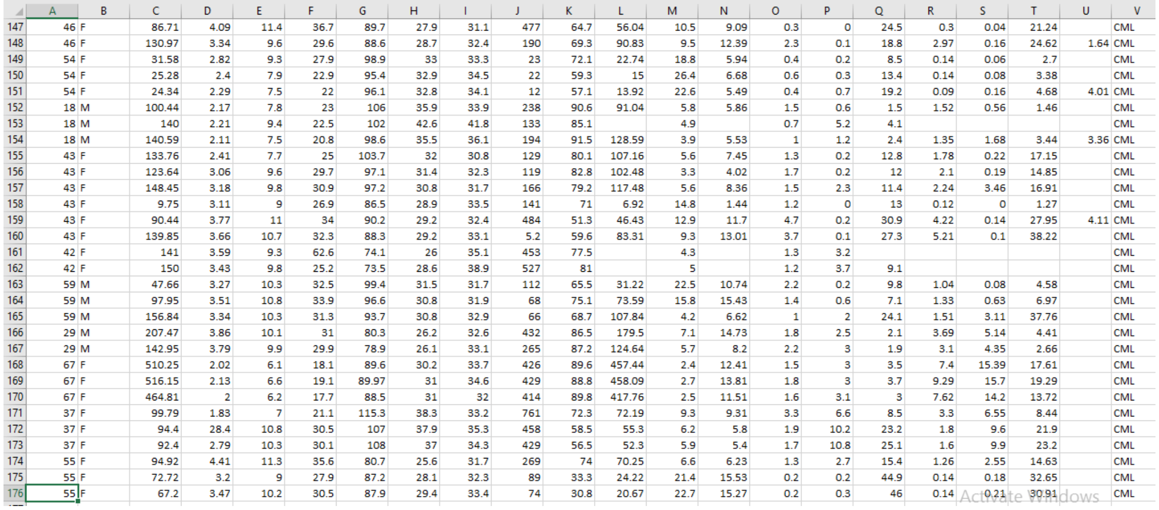
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
60	15	M	9.16	1.68	4.8	13.4	79.8	28.6	35.8	34	6.9	0.6	80.5	7.37	0	0.4	12.2	0	0.04	1.12		AML
61	14	F	12.33	3.62	10.1	26.9	74.3	27.9	37.5	640	68.7	8.47	13.9	1.72	0.6	0	16.8	0.07	0	2.07		AML
62	75	F	11.27	4.59	10.3	30.5	66.4	22.4	33.8	199	80.4	9.07	11.1	1.25	0.4	0.8	7.3	0.04	0.09	0.82		AML
63	19	F	153.98	2.97	9	28.7	96.6	30.3	31.4	55	1.1	1.61	88.5	136.22	0	0	10.4	0.06	0.06	16.03		AML
64	6	F	22.86	2.98	8	23.5	78.9	26.8	34	45	1.1	0.26	94	21.49	0.4	0.3	4.2	0.09	0.07	0.93		AML
65	14	F	9.99	3.93	10.8	29.2	74.3	27.5	37	682	66.7	6.66	17.8	1.78	0.9	0	14.6	0.09	0	1.45	1.67	AML
66	75	F	10.26	4.9	9.3	27.6	65.9	22.2	33.7	194	77.8	7.98	14.2	1.46	0.3	1	6.7	0.03	0.1	0.69		AML
67	52	M	14	2	9	23	74	25	33.5	116	45	3.3	7	0.06	1.96	0.03	9	0.01	0.04	1.26		AML
68	60	M	27	3.06	9	26.5	85	30	34	167.5	41	3.75	36	9.72	0.38	0.99	18.51	0.1	0.26	4.99	1.55	AML
69	56	F	10.4	3.18	9.2	27.2	85.5	29	33.9	52	62.2	6.5	7.5	0.8	3.2	1.1	26	0.3	0.1	2.7	0	AML
70	56	F	9.4	3.16	8.9	26.6	84.2	28.2	33.5	121	79.9	7.51	5.6	0.53	0.1	0	14.4	0.01	0	1.35	0	AML
71	56	F	10.55	2.75	7.9	23.8	86.5	28.7	33.2	188	85.2	8.99	7.4	0.78	0.1	0	7.3	0.01	0	0.77	0	AML
72	17	M	12.61	4.12	11.5	33.4	81.1	27.9	34.4	276	71.8	9.05	23.5	2.96	0.7	0.7	3.3	0.09	0.09	0.42	0	AML
73	17	M	15.95	4.07	11.4	33.5	82.3	28	34	238	83.4	13.31	10.4	1.66	0.4	0	5.8	0.06	0	0.92	0	AML
74	54	F	10.41	4.23	12.5	37.1	87.7	29.6	33.7	485	66.9	6.96	24	2.5	0.4	1.8	6.9	0.04	0.19	0.72	0	AML
75	54	F	9.1	3.94	12.1	35.2	89.3	30.8	34.5	484	63.6	5.8	25.6	2.3	1.4	2.2	7.1	0.1	0.2	0.7	0	AML
76	54	F	5.82	4.57	13.6	38.6	84.5	29.8	35.2	186	50.8	2.91	38.1	2.22	1	4.3	5.8	0.06	0.25	0.34	0	AML
77	61	F	4.95	3.57	11.9	33	92.4	33.3	36.1	205	74.1	3.67	13.7	0.68	0.6	6.3	5.3	0.03	0.31	0.26	0	AML
78	61	F	5.3	4.1	14.1	42.2	103	34.3	33.4	142	58.1	3.1	25.5	1.4	0.4	10.2	5.9	0	0.5	0.3	0	AML
79	61	F	7.11	4.22	14	38.3	90.8	33.2	36.6	208	72	5.12	18.4	1.31	0.3	3	6.3	0.02	0.21	0.45	0	AML
80	16	F	3.7	2.85	8.8	25.6	89.8	31	34.5	139	17.8	0.7	52.7	1.9	0.9	0.2	28.4	0	0	1.1	0	AML
81	16	F	5.39	3.56	11.1	31.7	89	31.2	35	303	42.6	2.46	35.8	1.93	0.4	0	18.2	0.02	0	0.98	0	AML
82	16	F	8.32	3.7	11.4	32.8	88.6	30	34.8	292	47.8	3.98	35.7	2.97	0.5	0	16	0.04	0	1.33	2.56	AML
83	12	M	2.83	3.47	10.3	27.7	79.8	29.7	37.2	114	8.4	0.24	86.6	2.45	0.4	0	4.6	0.01	0	0.13	0	AML
84	12	M	1.72	3.49	10.3	27.9	79.9	29.5	36.9	198	41.3	0.71	56.4	0.97	0	0	2.3	0	0	0.04	0	AML
85	12	M	4.66	3.24	9.5	27	83.3	29.3	35.2	251	27.1	1.26	69.1	3.22	0.2	0	3.6	0.01	0	0.17	0	AML
86	12	M	2.4	3.07	9.6	26.2	85.3	31.4	36.8	25	68	1.6	30.4	0.7	0.4	0.4	0.8	0	0	0	0	AML
87	12	M	2.4	2.67	7.9	22.8	85.4	29.4	34.5	16	74.7	1.8	23	0.6	0.2	1	1.1	0	0	0	0	AML
88	12	M	3.56	3.24	9.5	25.7	79.3	29.3	37	40	44.6	1.59	54.5	1.94	0	0.3	0.6	0	0.01	0.02	0.25	AML
89	15	M	5.71	3.86	11.5	32.7	84.7	29.8	35.2	246	80	4.57	14.5	0.83	0.2	0	5.3	0.01	0	0.3	0	AML



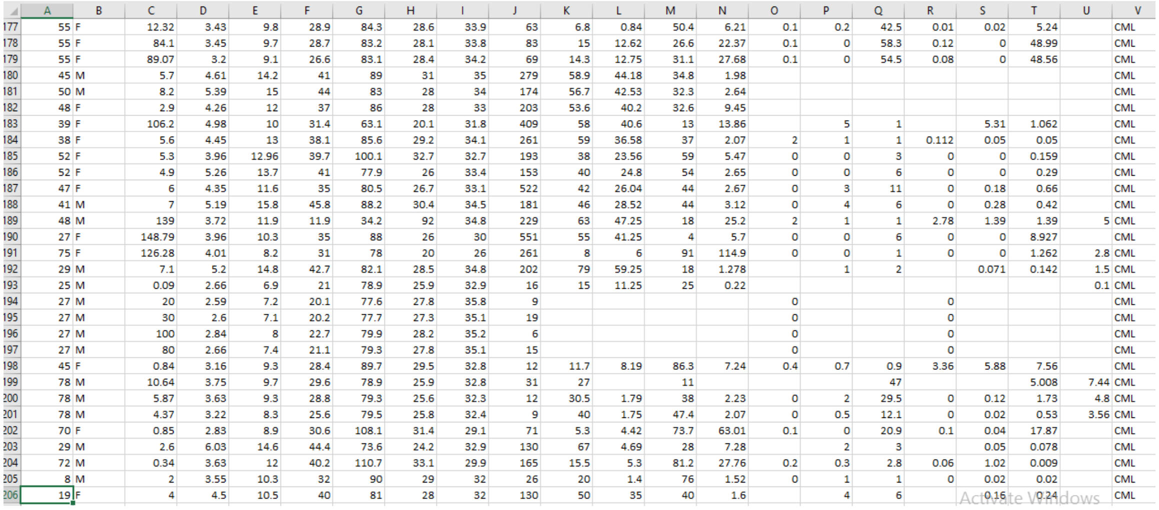
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
89	15	M	5.71	3.86	11.5	32.7	84.7	29.8	35.2	246	80	4.57	14.5	0.83	0.2	0	5.3	0.01	0	0.3	0	AML
90	15	M	10.51	4.05	12.1	35.6	87.9	29.9	34	216	77.9	8.18	14.7	1.55	0.4	0.1	6.9	0.04	0.01	0.73	0	AML
91	15	M	7.08	4.19	12.9	36.6	87.4	30.8	35.2	144	69.7	4.93	21.9	1.55	0.1	0.4	7.9	0.01	0.03	0.56	0	AML
92	6	F	1.7	3.21	9.1	27.2	84.7	28.4	33.5	20	12.4	0.2	85.3	1.4	0	0.7	1.6	0	0	0	0	AML
93	6	F	0.37	2.88	8.2	22.3	77.4	28.5	36.8	24	13.5	0.05	86.5	0.32	0	0	0	0	0	0	0.16	AML
94	6	F	0.61	3.06	8.7	32.2	75.8	28.4	37.5	10	6.6	0.04	93.4	0.57	0	0	0	0	0	0	0.21	AML
95	43	F	126.6	2.81	7.8	24.1	85.8	27.8	32.4	99	81.8	103.52	7.9	9.94	1.1	0.1	9.1	1.43	0.17	11.54	0	AML
96	43	F	77.07	2.45	6.7	20.9	85.3	27.3	32.1	53	83.1	64.05	8	6.15	0.7	0.1	8.1	0.54	0.09	6.24	2.64	AML
97	48	F	8.06	2.72	7.6	20.7	76.1	27.9	36.7	12	23.4	1.89	35.4	2.85	0.1	0	41.1	0.01	0	3.31	0	AML
98	48	F	9.44	2.88	8	22.2	77.1	27.8	36	8	20	1.88	36	3.4	0.2	0.2	43.6	0.02	0.02	4.12	0	AML
99	48	F	8.55	2.84	8	21.8	76.8	28.2	36.7	20	19.9	1.7	34.2	2.92	0.2	0.6	45.1	0.02	0.05	3.86	0	AML
100	14	F	9.5	3.97	11.7	32.5	81.9	29.3	35.8	785	69.2	6.6	18.8	1.8	2.3	0.2	9.6	0.2	0	0.9	0	AML
101	14	F	12.33	3.62	10.1	26.9	74.3	27.9	37.5	640	68.7	8.47	13.9	1.72	0.6	0	16.8	0.07	0	2.07	0	AML
102	14	F	9.99	3.93	10.8	29.2	74.3	27.5	37	682	66.7	6.66	17.8	1.78	0.9	0	14.6	0.09	0	1.46	1.67	AML
103	18	F	5.76	3.92	11.3	33.2	84.7	28.8	34	599	42	2.42	44.6	2.57	0.2	0	13.2	0.01	0	0.76	0	AML
104	18	F	6.57	3.85	11.1	32.4	84.2	28.8	34.3	566	46.1	3.03	36.7	2.41	0.2	0	17	0.01	0	1.12	0	AML
105	18	F	7.16	3.87	11.4	33	85.3	29.5	34.5	582	53	3.79	31.7	2.27	0.3	0.1	14.9	0.02	0.01	1.07	1.73	AML
106	56	F	1.4	2.66	7.9	22.5	84.6	29.6	35	239	51.1	0.7	32.1	0.04	1.3	0	15.5	0	0	0.02	0	AML
107	56	F	3.9	3.72	11.3	31.9	85.8	30.3	35.3	542	53.9	2.1	21.5	0.8	4.9	0.7	18.9	0.2	0	0.7	0	AML
108	56	F	4.96	4.06	12.1	33.1	81.5	29.8	36.6	646	52.8	2.87	216	1.01	0.2	0	21.4	0.01	0	1.06	0	AML
109	42	M	5.21	3.03	9.6	27.1	89.4	31.7	35.4	121	33	1.72	46.3	2.41	0.4	2.3	18	0.02	0.12	0.94	3.69	AML
110	42	M	6.48	3.15	10.2	29.6	94	32.4	34.5	148	48.8	3.16	31.3	2.03	0.5	3.2	16.2	0.03	0.21	1.5	2.9	AML
111	42	M	7.43	3.48	11.9	33.9	97.4	34.2	35.1	113	43	3.2	38.4	2.85	0.5	5.4	12.7	0.04	0.4	0.94	3.04	AML
112	42	M	2.14	3.2	9.7	26.6	83.1	30.3	36.5	16	20.1	0.43	1.28	59.8	0	0	20.1	0	0	0.43	0.29	AML
113	42	M	1.71	3.11	9.3	25.8	83	29.9	36	17	31.6	0.54	49.1	0.84	0	0	19.3	0	0	0.33	0.45	AML
114	42	M	1.83	3.15	9.5	26.4	83.8	30.2	36	15	22.9	0.42	54.1	0.99	0	0	23	0	0	0.42	0.53	AML
115	27	M	44.33	3.64	11.9	31.7	87.1	32.7	37.5	110	43.3	1.96	41.3	1.79	0.2	2.8	10.4	0.01	0.12	0.45	1.88	AML
116	27	M	5.91	4.04	13	34.4	85.1	32.2	37.8	136	54.9	3.24	34.3	2.03	0.3	2.2	8.3	0.02	0.13	0.49	1.74	AML
117	27	M	4.01	4.05	13	34	84	32.1	38.2	124	46.7	1.87	37.2	1.49	0.2	3.7	12.2	0.01	0.15	0.49	1.24	AML
118	27	M	2.5	3.09	9.1	24.3	78.6	29.4	37.4	166	88.8	2.22	2.4	0.06	0.8	7.6	0.4	0.02	0.19	0.01	5.3	AML



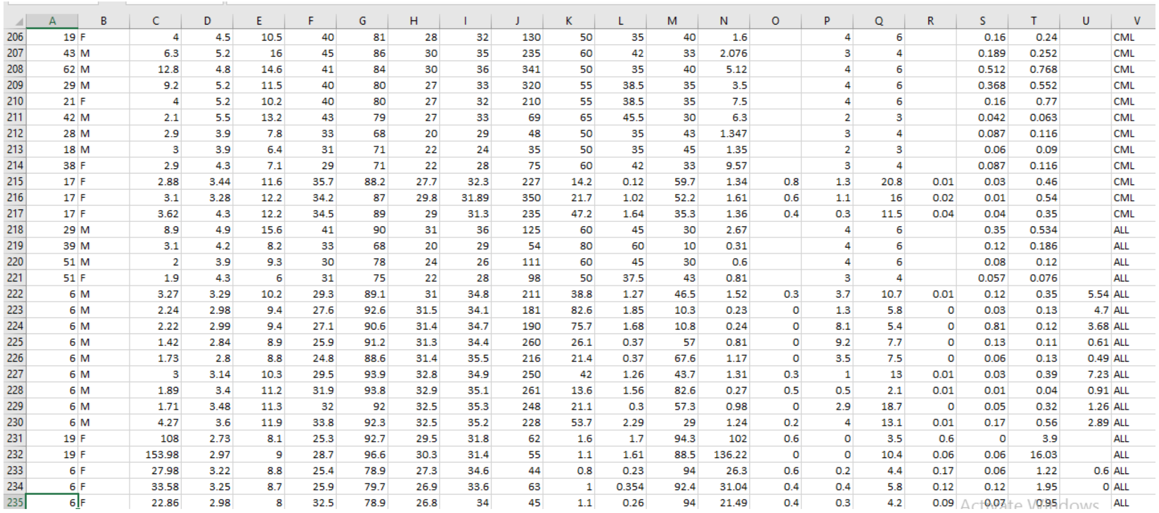
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
118	27	M	2.5	3.09	9.1	24.3	78.6	29.4	37.4	166	88.8	2.22	2.4	0.06	0.8	7.6	0.4	0.02	0.19	0.01	5.3	AML
119	27	M	2.85	2.86	8.3	22	76.9	29	37.7	89	92.9	2.65	3.5	0.1	0.4	2.8	0.4	0.01	0.08	0.01	0.18	AML
120	27	M	0.86	2.66	7.8	20.1	75.6	29.3	38.8	32	82.5	0.71	15.1	0.13	0	1.2	1.2	0	0.01	0.01	0.17	AML
121	27	M	6.44	3.52	10.9	31.1	88.4	31	35	179	66.4	4.28	17.9	1.15	0.2	3.9	11.6	0.01	0.25	0.75	2.78	AML
122	27	M	8.25	3.36	10.5	29.5	87.8	31.3	35.6	137	94.6	7.81	1.6	0.13	0.1	0.4	3.3	0.01	0.03	0.27	0.84	AML
123	27	M	5.56	3	9.3	26.1	87	31	35.6	110	96.2	5.35	1.6	0.03	0	0.7	1.8	0	0.04	0.1	0.35	AML
124	27	M	0.29	3.12	8.7	22.8	73.1	27.69	38.2	48	3.5	2.625	65.5	0.18995	0	0	31	0	0	0.0899	0.16	AML
125	27	M	0.5	2.56	8.4	18.7	73	32.8	44.9	66	2	1.5	40	2	0	0	58	0	0	0.29	0.24	AML
126	27	M	0.6	2.53	6.9	19.9	78.7	27.3	34.7	83			65	0.39	0	0	35	0	0	0.21	0.3	AML
127	14	F	9.5	3.97	11.7	32.5	81.9	29.3	53.8	785	69.2	6.6	18.8	1.8	2.3	0.2	9.6	0.2	0	0.9		AML
128	14	F	12.33	3.62	10.1	26.9	74.3	27.9	37.5	640	68.7	8.47	13.9	1.72	0.6	0	16.8	0.07	0	2.07		AML
129	14	F	9.99	3.93	10.8	29.2	74.3	27.5	37	682	66.7	6.66	17.8	1.78	0.9	0	14.6	0.09	0	1.46	1.67	AML
130	75	F	12.6	5	11.4	35.3	70.6	22.8	32.2	211	68.2	8.6	20.8	2.6	0.3	2.1	8.6	0	0.3	1.1		AML
131	75	F	11.27	4.59	10.3	30.5	66.4	22.4	33.8	199	80.4	9.07	11.1	1.25	0.4	0.8	7.3	0.04	0.09	0.82		AML
132	75	F	10.26	4.19	9.3	27.6	65.9	22.2	33.7	194	77.8	7.98	14.2	1.46	0.3	1	6.7	0.03	0.1	0.69		AML
133	87	F	5.08	3.88	12.1	37	95.4	31.2	32.7	90	45.7	2.32	40.2	2.04	0.4	3.9	9.8	0.02	0.2	0.5	1.71	CML
134	87	F	3.59	3.43	11.1	31.4	91.5	32.4	35.4	83	29.8	1.07	52.9	1.9	0.3	7.5	9.5	0.01	0.27	0.34		CML
135	87	F	2.61	2.89	9.3	26.7	92.4	32.2	34.8	61	32.6	0.85	49.2	1.28	0.4	6.9	11.1	0.01	0.18	0.29	1.7	CML
136	39	F	5.99	4.58	12.3	37.6	82.1	26.9	32.7	257	64.9	3.89	21.9	1.31	0.7	3.3	9.2	0.04	0.2	0.55		CML
137	39	F	7.99	5.07	13.3	40.7	80.3	26.2	32.7	342	68.1	5.45	18.3	1.46	0.4	1.4	11.8	0.03	0.11	0.94		CML
138	43	F	481.98	2.41	8	21.5	89.2	33.2	37.2	630	89.3	429.9	5	24.2	1.6	1.9	2.2	7.89	9.22	10.77		CML
139	43	F	516.57	2.47	8.3	22.1	89.5	33.6	37.6	6	89.2	460.57	4.8	25.02	1.8	1.7	2.5	9.5	8.59	12.89		CML
140	46	F	256.75	3.48	9.9	32.3	92.8	28.4	30.7	319	72.7	186.58	6.1	15.67	3.2	0.4	17.6	8.34	0.9	45.26	1.39	CML
141	46	F	246.64	3.01	8.9	28.1	93.4	29.6	31.7	499	72.3	178.25	6.4	15.86	2.1	0.2	19	5.08	0.6	46.85		CML
142	46	F	274.28	3.3	9.4	30	90.9	28.5	31.3	301	75.5	206.71	5.6	15.39	2.9	0.3	15.7	8.05	0.94	43.19	1.13	CML
143	46	F	368.79	1.99	5.8	19.4	97.5	29.1	29.9	145	78	287.65	3.7	13.77	1	0.1	17.2	3.56	0.36	63.45		CML
144	46	F	411.14	2.77	8	25.1	90.6	28.9	31.9	162	77	316.7	7.9	32.3	2.1	0.1	12.9	8.75	0.32	53.07		CML
145	46	F	517.15	3.2	9.6	29.3	91.6	30	32.8	197	76.2	394.31	3.5	18.06	3.4	0.1	16.8	17.51	0.56	86.71		CML
146	46	F	7.36	3.68	10.2	32.4	88	27.7	31.5	500	42.8	3.15	23.2	1.71	1.8	0.1	32.1	0.13	0.01	2.36		CML
147	46	F	86.71	4.09	11.4	36.7	89.7	27.9	31.1	477	64.7	56.04	10.5	9.09	0.3	0	24.5	0.3	0.04	21.24		CML



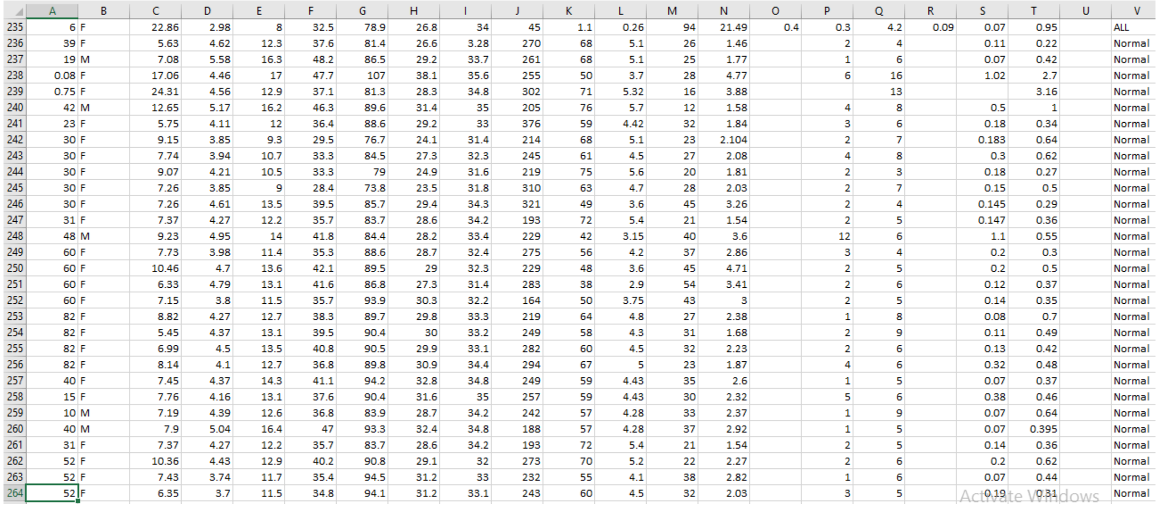
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
147	46	F	86.71	4.09	11.4	36.7	89.7	27.9	31.1	477	64.7	56.04	10.5	9.09	0.3	0	24.5	0.3	0.04	21.24		CML
148	46	F	130.97	3.34	9.6	29.6	88.6	28.7	32.4	190	69.3	90.83	9.5	12.39	2.3	0.1	18.8	2.97	0.16	24.62	1.64	CML
149	54	F	31.58	2.82	9.3	27.9	98.9	33	33.3	23	72.1	22.74	18.8	5.94	0.4	0.2	8.5	0.14	0.06	2.7		CML
150	54	F	25.28	2.4	7.9	22.9	95.4	32.9	34.5	22	59.3	15	26.4	6.68	0.6	0.3	13.4	0.14	0.08	3.38		CML
151	54	F	24.34	2.29	7.5	22	96.1	32.8	34.1	12	57.1	13.92	22.6	5.49	0.4	0.7	19.2	0.09	0.16	4.68	4.01	CML
152	18	M	100.44	2.17	7.8	23	106	35.9	33.9	238	90.6	91.04	5.8	5.86	1.5	0.6	1.5	1.52	0.56	1.46		CML
153	18	M	140	2.21	9.4	22.5	102	42.6	41.8	133	85.1		4.9		0.7	5.2	4.1					CML
154	18	M	140.59	2.11	7.5	20.8	98.6	35.5	36.1	194	91.5	128.59	3.9	5.53	1	1.2	2.4	1.35	1.68	3.44	3.36	CML
155	43	F	133.76	2.41	7.7	25	103.7	32	30.8	129	80.1	107.16	5.6	7.45	1.3	0.2	12.8	1.78	0.22	17.15		CML
156	43	F	123.64	3.06	9.6	29.7	97.1	31.4	32.3	119	82.8	102.48	3.3	4.02	1.7	0.2	12	2.1	0.19	14.85		CML
157	43	F	148.45	3.18	9.8	30.9	97.2	30.8	31.7	166	79.2	117.48	5.6	8.36	1.5	2.3	11.4	2.24	3.46	16.91		CML
158	43	F	9.75	3.11	9	26.9	86.5	28.9	33.5	141	71	6.92	14.8	1.44	1.2	0	13	0.12	0	1.27		CML
159	43	F	90.44	3.77	11	34	90.2	29.2	32.4	484	51.3	46.43	12.9	11.7	4.7	0.2	30.9	4.22	0.14	27.95	4.11	CML
160	43	F	139.85	3.66	10.7	32.3	88.3	29.2	33.1	5.2	59.6	83.31	9.3	13.01	3.7	0.1	27.3	5.21	0.1	38.22		CML
161	42	F	141	3.59	9.3	62.6	74.1	26	35.1	453	77.5		4.3		1.3	3.2						CML
162	42	F	150	3.43	9.8	25.2	73.5	28.6	38.9	527	81		5		1.2	3.7	9.1					CML
163	59	M	47.66	3.27	10.3	32.5	99.4	31.5	31.7	112	65.5	31.22	22.5	10.74	2.2	0.2	9.8	1.04	0.08	4.58		CML
164	59	M	97.95	3.51	10.8	33.9	96.6	30.8	31.9	68	75.1	73.59	15.8	15.43	1.4	0.6	7.1	1.33	0.63	6.97		CML
165	59	M	156.84	3.34	10.3	31.3	93.7	30.8	32.9	66	68.7	107.84	4.2	6.62	1	2	24.1	1.51	3.11	37.76		CML
166	29	M	207.47	3.86	10.1	31	80.3	26.2	32.6	432	86.5	179.5	7.1	14.73	1.8	2.5	2.1	3.69	5.14	4.41		CML
167	29	M	142.95	3.79	9.9	29.9	78.9	26.1	33.1	265	87.2	124.64	5.7	8.2	2.2	3	1.9	3.1	4.35	2.66		CML
168	67	F	510.25	2.02	6.1	18.1	89.6	30.2	33.7	426	89.6	457.44	2.4	12.41	1.5	3	3.5	7.4	15.39	17.61		CML
169	67	F	516.15	2.13	6.6	19.1	89.97	31	34.6	429	88.8	458.09	2.7	13.81	1.8	3	3.7	9.29	15.7	19.29		CML
170	67	F	464.81	2	6.2	17.7	88.5	31	32	414	89.8	417.76	2.5	11.51	1.6	3.1	3	7.62	14.2	13.72		CML
171	37	F	99.79	1.83	7	21.1	115.3	38.3	33.2	761	72.3	72.19	9.3	9.31	3.3	6.6	8.5	3.3	6.55	8.44		CML
172	37	F	94.4	28.4	10.8	30.5	107	37.9	35.3	458	58.5	55.3	6.2	5.8	1.9	10.2	23.2	1.8	9.6	21.9		CML
173	37	F	92.4	2.79	10.3	30.1	108	37	34.3	429	56.5	52.3	5.9	5.4	1.7	10.8	25.1	1.6	9.9	23.2		CML
174	55	F	94.92	4.41	11.3	35.6	80.7	25.6	31.7	269	74	70.25	6.6	6.23	1.3	2.7	15.4	1.26	2.55	14.63		CML
175	55	F	72.72	3.2	9	27.9	87.2	28.1	32.3	89	33.3	24.22	21.4	15.53	0.2	0.2	44.9	0.14	0.18	32.65		CML
176	55	F	67.2	3.47	10.2	30.5	87.9	29.4	33.4	74	30.8	20.67	22.7	15.27	0.2	0.3	46	0.14	0.21	30.91		CML



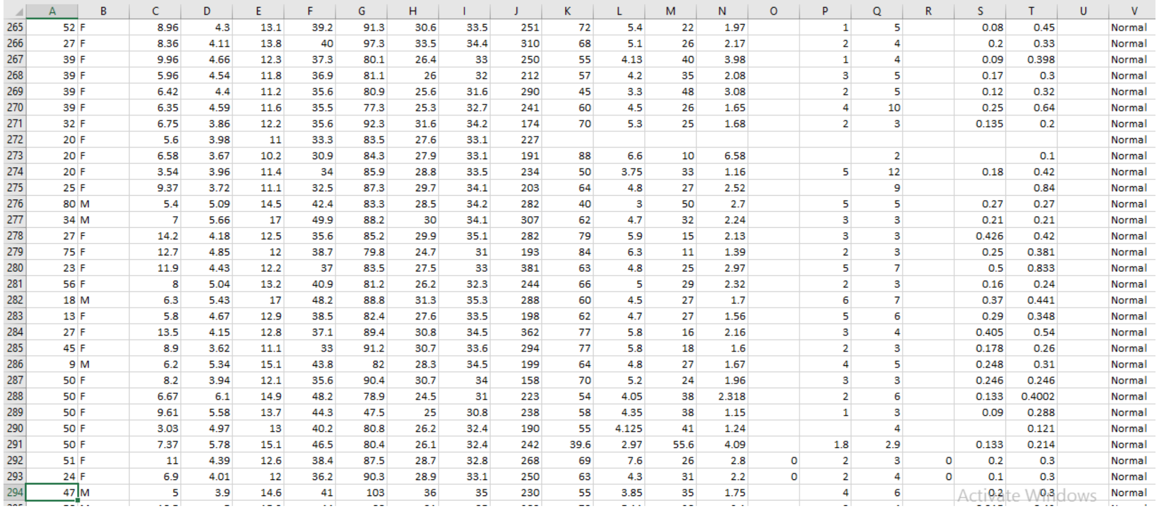
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
177	55	F	12.32	3.43	9.8	28.9	84.3	28.6	33.9	63	6.8	0.84	50.4	6.21	0.1	0.2	42.5	0.01	0.02	5.24		CML
178	55	F	84.1	3.45	9.7	28.7	83.2	28.1	33.8	83	15	12.62	26.6	22.37	0.1	0	58.3	0.12	0	48.99		CML
179	55	F	89.07	3.2	9.1	26.6	83.1	28.4	34.2	69	14.3	12.75	31.1	27.68	0.1	0	54.5	0.08	0	48.56		CML
180	45	M	5.7	4.61	14.2	41	89	31	35	279	58.9	44.18	34.8	1.98								CML
181	50	M	8.2	5.39	15	44	83	28	34	174	56.7	42.53	32.3	2.64								CML
182	48	F	2.9	4.26	12	37	86	28	33	203	53.6	40.2	32.6	9.45								CML
183	39	F	106.2	4.98	10	31.4	63.1	20.1	31.8	409	58	40.6	13	13.86		5	1		5.31	1.062		CML
184	38	F	5.6	4.45	13	38.1	85.6	29.2	34.1	261	59	36.58	37	2.07	2	1	1	0.112	0.05	0.05		CML
185	52	F	5.3	3.96	12.96	39.7	100.1	32.7	32.7	193	38	23.56	59	5.47	0	0	3	0	0	0.159		CML
186	52	F	4.9	5.26	13.7	41	77.9	26	33.4	153	40	24.8	54	2.65	0	0	6	0	0	0.29		CML
187	47	F	6	4.35	11.6	35	80.5	26.7	33.1	522	42	26.04	44	2.67	0	3	11	0	0.18	0.66		CML
188	41	M	7	5.19	15.8	45.8	88.2	30.4	34.5	181	46	28.52	44	3.12	0	4	6	0	0.28	0.42		CML
189	48	M	139	3.72	11.9	11.9	34.2	92	34.8	229	63	47.25	18	25.2	2	1	1	2.78	1.39	1.39	5	CML
190	27	F	148.79	3.96	10.3	35	88	26	30	551	55	41.25	4	5.7	0	0	6	0	0	8.927		CML
191	75	F	126.28	4.01	8.2	31	78	20	26	261	8	6	91	114.9	0	0	1	0	0	1.262	2.8	CML
192	29	M	7.1	5.2	14.8	42.7	82.1	28.5	34.8	202	79	59.25	18	1.278		1	2		0.071	0.142	1.5	CML
193	25	M	0.09	2.66	6.9	21	78.9	25.9	32.9	16	15	11.25	25	0.22							0.1	CML
194	27	M	20	2.59	7.2	20.1	77.6	27.8	35.8	9					0			0				CML
195	27	M	30	2.6	7.1	20.2	77.7	27.3	35.1	19					0			0				CML
196	27	M	100	2.84	8	22.7	79.9	28.2	35.2	6					0			0				CML
197	27	M	80	2.66	7.4	21.1	79.3	27.8	35.1	15					0			0				CML
198	45	F	0.84	3.16	9.3	28.4	89.7	29.5	32.8	12	11.7	8.19	86.3	7.24	0.4	0.7	0.9	3.36	5.88	7.56		CML
199	78	M	10.64	3.75	9.7	29.6	78.9	25.9	32.8	31	27		11				47			5.008	7.44	CML
200	78	M	5.87	3.63	9.3	28.8	79.3	25.6	32.3	12	30.5	1.79	38	2.23	0	2	29.5	0	0.12	1.73	4.8	CML
201	78	M	4.37	3.22	8.3	25.6	79.5	25.8	32.4	9	40	1.75	47.4	2.07	0	0.5	12.1	0	0.02	0.53	3.56	CML
202	70	F	0.85	2.83	8.9	30.6	108.1	31.4	29.1	71	5.3	4.42	73.7	63.01	0.1	0	20.9	0.1	0.04	17.87		CML
203	29	M	2.6	6.03	14.6	44.4	73.6	24.2	32.9	130	67	4.69	28	7.28		2	3		0.05	0.078		CML
204	72	M	0.34	3.63	12	40.2	110.7	33.1	29.9	165	15.5	5.3	81.2	27.76	0.2	0.3	2.8	0.06	1.02	0.009		CML
205	8	M	2	3.55	10.3	32	90	29	32	26	20	1.4	76	1.52	0	1	1	0	0.02	0.02		CML
206	19	F	4	4.5	10.5	40	81	28	32	130	50	35	40	1.6		4	6		0.16	0.24		CML



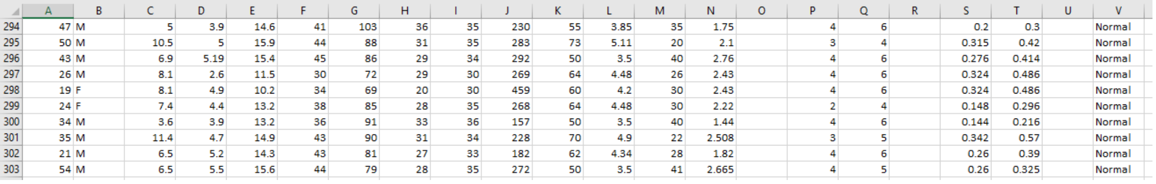
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
206	19	F	4	4.5	10.5	40	81	28	32	130	50	35	40	1.6		4	6		0.16	0.24		CML
207	43	M	6.3	5.2	16	45	86	30	35	235	60	42	33	2.076		3	4		0.189	0.252		CML
208	62	M	12.8	4.8	14.6	41	84	30	36	341	50	35	40	5.12		4	6		0.512	0.768		CML
209	29	M	9.2	5.2	11.5	40	80	27	33	320	55	38.5	35	3.5		4	6		0.368	0.552		CML
210	21	F	4	5.2	10.2	40	80	27	32	210	55	38.5	35	7.5		4	6		0.16	0.77		CML
211	42	M	2.1	5.5	13.2	43	79	27	33	69	65	45.5	30	6.3		2	3		0.042	0.063		CML
212	28	M	2.9	3.9	7.8	33	68	20	29	48	50	35	43	1.347		3	4		0.087	0.116		CML
213	18	M	3	3.9	6.4	31	71	22	24	35	50	35	45	1.35		2	3		0.06	0.09		CML
214	38	F	2.9	4.3	7.1	29	71	22	28	75	60	42	33	9.57		3	4		0.087	0.116		CML
215	17	F	2.88	3.44	11.6	35.7	88.2	27.7	32.3	227	14.2	0.12	59.7	1.34	0.8	1.3	20.8	0.01	0.03	0.46		CML
216	17	F	3.1	3.28	12.2	34.2	87	29.8	31.89	350	21.7	1.02	52.2	1.61	0.6	1.1	16	0.02	0.01	0.54		CML
217	17	F	3.62	4.3	12.2	34.5	89	29	31.3	235	47.2	1.64	35.3	1.36	0.4	0.3	11.5	0.04	0.04	0.35		CML
218	29	M	8.9	4.9	15.6	41	90	31	36	125	60	45	30	2.67		4	6		0.35	0.534		ALL
219	39	M	3.1	4.2	8.2	33	68	20	29	54	80	60	10	0.31		4	6		0.12	0.186		ALL
220	51	M	2	3.9	9.3	30	78	24	26	111	60	45	30	0.6		4	6		0.08	0.12		ALL
221	51	F	1.9	4.3	6	31	75	22	28	98	50	37.5	43	0.81		3	4		0.057	0.076		ALL
222	6	M	3.27	3.29	10.2	29.3	89.1	31	34.8	211	38.8	1.27	46.5	1.52	0.3	3.7	10.7	0.01	0.12	0.35	5.54	ALL
223	6	M	2.24	2.98	9.4	27.6	92.6	31.5	34.1	181	82.6	1.85	10.3	0.23	0	1.3	5.8	0	0.03	0.13	4.7	ALL
224	6	M	2.22	2.99	9.4	27.1	90.6	31.4	34.7	190	75.7	1.68	10.8	0.24	0	8.1	5.4	0	0.81	0.12	3.68	ALL
225	6	M	1.42	2.84	8.9	25.9	91.2	31.3	34.4	260	26.1	0.37	57	0.81	0	9.2	7.7	0	0.13	0.11	0.61	ALL
226	6	M	1.73	2.8	8.8	24.8	88.6	31.4	35.5	216	21.4	0.37	67.6	1.17	0	3.5	7.5	0	0.06	0.13	0.49	ALL
227	6	M	3	3.14	10.3	29.5	93.9	32.8	34.9	250	42	1.26	43.7	1.31	0.3	1	13	0.01	0.03	0.39	7.23	ALL
228	6	M	1.89	3.4	11.2	31.9	93.8	32.9	35.1	261	13.6	1.56	82.6	0.27	0.5	0.5	2.1	0.01	0.01	0.04	0.91	ALL
229	6	M	1.71	3.48	11.3	32	92	32.5	35.3	248	21.1	0.3	57.3	0.98	0	2.9	18.7	0	0.05	0.32	1.26	ALL
230	6	M	4.27	3.6	11.9	33.8	92.3	32.5	35.2	228	53.7	2.29	29	1.24	0.2	4	13.1	0.01	0.17	0.56	2.89	ALL
231	19	F	108	2.73	8.1	25.3	92.7	29.5	31.8	62	1.6	1.7	94.3	102	0.6	0	3.5	0.6	0	3.9		ALL
232	19	F	153.98	2.97	9	28.7	96.6	30.3	31.4	55	1.1	1.61	88.5	136.22	0	0	10.4	0.06	0.06	16.03		ALL
233	6	F	27.98	3.22	8.8	25.4	78.9	27.3	34.6	44	0.8	0.23	94	26.3	0.6	0.2	4.4	0.17	0.06	1.22	0.6	ALL
234	6	F	33.58	3.25	8.7	25.9	79.7	26.9	33.6	63	1	0.354	92.4	31.04	0.4	0.4	5.8	0.12	0.12	1.95	0	ALL
235	6	F	22.86	2.98	8	32.5	78.9	26.8	34	45	1.1	0.26	94	21.49	0.4	0.3	4.2	0.09	0.07	0.95		ALL



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
235	6	F	22.86	2.98	8	32.5	78.9	26.8	34	45	1.1	0.26	94	21.49	0.4	0.3	4.2	0.09	0.07	0.95		ALL
236	39	F	5.63	4.62	12.3	37.6	81.4	26.6	3.28	270	68	5.1	26	1.46		2	4		0.11	0.22		Normal
237	19	M	7.08	5.58	16.3	48.2	86.5	29.2	33.7	261	68	5.1	25	1.77		1	6		0.07	0.42		Normal
238	0.08	F	17.06	4.46	17	47.7	107	38.1	35.6	255	50	3.7	28	4.77		6	16		1.02	2.7		Normal
239	0.75	F	24.31	4.56	12.9	37.1	81.3	28.3	34.8	302	71	5.32	16	3.88			13			3.16		Normal
240	42	M	12.65	5.17	16.2	46.3	89.6	31.4	35	205	76	5.7	12	1.58		4	8		0.5	1		Normal
241	23	F	5.75	4.11	12	36.4	88.6	29.2	33	376	59	4.42	32	1.84		3	6		0.18	0.34		Normal
242	30	F	9.15	3.85	9.3	29.5	76.7	24.1	31.4	214	68	5.1	23	2.104		2	7		0.183	0.64		Normal
243	30	F	7.74	3.94	10.7	33.3	84.5	27.3	32.3	245	61	4.5	27	2.08		4	8		0.3	0.62		Normal
244	30	F	9.07	4.21	10.5	33.3	79	24.9	31.6	219	75	5.6	20	1.81		2	3		0.18	0.27		Normal
245	30	F	7.26	3.85	9	28.4	73.8	23.5	31.8	310	63	4.7	28	2.03		2	7		0.15	0.5		Normal
246	30	F	7.26	4.61	13.5	39.5	85.7	29.4	34.3	321	49	3.6	45	3.26		2	4		0.145	0.29		Normal
247	31	F	7.37	4.27	12.2	35.7	83.7	28.6	34.2	193	72	5.4	21	1.54		2	5		0.147	0.36		Normal
248	48	M	9.23	4.95	14	41.8	84.4	28.2	33.4	229	42	3.15	40	3.6		12	6		1.1	0.55		Normal
249	60	F	7.73	3.98	11.4	35.3	88.6	28.7	32.4	275	56	4.2	37	2.86		3	4		0.2	0.3		Normal
250	60	F	10.46	4.7	13.6	42.1	89.5	29	32.3	229	48	3.6	45	4.71		2	5		0.2	0.5		Normal
251	60	F	6.33	4.79	13.1	41.6	86.8	27.3	31.4	283	38	2.9	54	3.41		2	6		0.12	0.37		Normal
252	60	F	7.15	3.8	11.5	35.7	93.9	30.3	32.2	164	50	3.75	43	3		2	5		0.14	0.35		Normal
253	82	F	8.82	4.27	12.7	38.3	89.7	29.8	33.3	219	64	4.8	27	2.38		1	8		0.08	0.7		Normal
254	82	F	5.45	4.37	13.1	39.5	90.4	30	33.2	249	58	4.3	31	1.68		2	9		0.11	0.49		Normal
255	82	F	6.99	4.5	13.5	40.8	90.5	29.9	33.1	282	60	4.5	32	2.23		2	6		0.13	0.42		Normal
256	82	F	8.14	4.1	12.7	36.8	89.8	30.9	34.4	294	67	5	23	1.87		4	6		0.32	0.48		Normal
257	40	F	7.45	4.37	14.3	41.1	94.2	32.8	34.8	249	59	4.43	35	2.6		1	5		0.07	0.37		Normal
258	15	F	7.76	4.16	13.1	37.6	90.4	31.6	35	257	59	4.43	30	2.32		5	6		0.38	0.46		Normal
259	10	M	7.19	4.39	12.6	36.8	83.9	28.7	34.2	242	57	4.28	33	2.37		1	9		0.07	0.64		Normal
260	40	M	7.9	5.04	16.4	47	93.3	32.4	34.8	188	57	4.28	37	2.92		1	5		0.07	0.395		Normal
261	31	F	7.37	4.27	12.2	35.7	83.7	28.6	34.2	193	72	5.4	21	1.54		2	5		0.14	0.36		Normal
262	52	F	10.36	4.43	12.9	40.2	90.8	29.1	32	273	70	5.2	22	2.27		2	6		0.2	0.62		Normal
263	52	F	7.43	3.74	11.7	35.4	94.5	31.2	33	232	55	4.1	38	2.82		1	6		0.07	0.44		Normal
264	52	F	6.35	3.7	11.5	34.8	94.1	31.2	33.1	243	60	4.5	32	2.03		3	5		0.19	0.31		Normal



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
265	52	F	8.96	4.3	13.1	39.2	91.3	30.6	33.5	251	72	5.4	22	1.97		1	5		0.08	0.45		Normal
266	27	F	8.36	4.11	13.8	40	97.3	33.5	34.4	310	68	5.1	26	2.17		2	4		0.2	0.33		Normal
267	39	F	9.96	4.66	12.3	37.3	80.1	26.4	33	250	55	4.13	40	3.98		1	4		0.09	0.398		Normal
268	39	F	5.96	4.54	11.8	36.9	81.1	26	32	212	57	4.2	35	2.08		3	5		0.17	0.3		Normal
269	39	F	6.42	4.4	11.2	35.6	80.9	25.6	31.6	290	45	3.3	48	3.08		2	5		0.12	0.32		Normal
270	39	F	6.35	4.59	11.6	35.5	77.3	25.3	32.7	241	60	4.5	26	1.65		4	10		0.25	0.64		Normal
271	32	F	6.75	3.86	12.2	35.6	92.3	31.6	34.2	174	70	5.3	25	1.68		2	3		0.135	0.2		Normal
272	20	F	5.6	3.98	11	33.3	83.5	27.6	33.1	227												Normal
273	20	F	6.58	3.67	10.2	30.9	84.3	27.9	33.1	191	88	6.6	10	6.58			2			0.1		Normal
274	20	F	3.54	3.96	11.4	34	85.9	28.8	33.5	234	50	3.75	33	1.16		5	12		0.18	0.42		Normal
275	25	F	9.37	3.72	11.1	32.5	87.3	29.7	34.1	203	64	4.8	27	2.52			9			0.84		Normal
276	80	M	5.4	5.09	14.5	42.4	83.3	28.5	34.2	282	40	3	50	2.7		5	5		0.27	0.27		Normal
277	34	M	7	5.66	17	49.9	88.2	30	34.1	307	62	4.7	32	2.24		3	3		0.21	0.21		Normal
278	27	F	14.2	4.18	12.5	35.6	85.2	29.9	35.1	282	79	5.9	15	2.13		3	3		0.426	0.42		Normal
279	75	F	12.7	4.85	12	38.7	79.8	24.7	31	193	84	6.3	11	1.39		2	3		0.25	0.381		Normal
280	23	F	11.9	4.43	12.2	37	83.5	27.5	33	381	63	4.8	25	2.97		5	7		0.5	0.833		Normal
281	56	F	8	5.04	13.2	40.9	81.2	26.2	32.3	244	66	5	29	2.32		2	3		0.16	0.24		Normal
282	18	M	6.3	5.43	17	48.2	88.8	31.3	35.3	288	60	4.5	27	1.7		6	7		0.37	0.441		Normal
283	13	F	5.8	4.67	12.9	38.5	82.4	27.6	33.5	198	62	4.7	27	1.56		5	6		0.29	0.348		Normal
284	27	F	13.5	4.15	12.8	37.1	89.4	30.8	34.5	362	77	5.8	16	2.16		3	4		0.405	0.54		Normal
285	45	F	8.9	3.62	11.1	33	91.2	30.7	33.6	294	77	5.8	18	1.6		2	3		0.178	0.26		Normal
286	9	M	6.2	5.34	15.1	43.8	82	28.3	34.5	199	64	4.8	27	1.67		4	5		0.248	0.31		Normal
287	50	F	8.2	3.94	12.1	35.6	90.4	30.7	34	158	70	5.2	24	1.96		3	3		0.246	0.246		Normal
288	50	F	6.67	6.1	14.9	48.2	78.9	24.5	31	223	54	4.05	38	2.318		2	6		0.133	0.4002		Normal
289	50	F	9.61	5.58	13.7	44.3	47.5	25	30.8	238	58	4.35	38	1.15		1	3		0.09	0.288		Normal
290	50	F	3.03	4.97	13	40.2	80.8	26.2	32.4	190	55	4.125	41	1.24			4			0.121		Normal
291	50	F	7.37	5.78	15.1	46.5	80.4	26.1	32.4	242	39.6	2.97	55.6	4.09		1.8	2.9		0.133	0.214		Normal
292	51	F	11	4.39	12.6	38.4	87.5	28.7	32.8	268	69	7.6	26	2.8	0	2	3	0	0.2	0.3		Normal
293	24	F	6.9	4.01	12	36.2	90.3	28.9	33.1	250	63	4.3	31	2.2	0	2	4	0	0.1	0.3		Normal
294	47	M	5	3.9	14.6	41	103	36	35	230	55	3.85	35	1.75		4	6		0.2	0.3		Normal



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
294	47	M	5	3.9	14.6	41	103	36	35	230	55	3.85	35	1.75		4	6		0.2	0.3		Normal
295	50	M	10.5	5	15.9	44	88	31	35	283	73	5.11	20	2.1		3	4		0.315	0.42		Normal
296	43	M	6.9	5.19	15.4	45	86	29	34	292	50	3.5	40	2.76		4	6		0.276	0.414		Normal
297	26	M	8.1	2.6	11.5	30	72	29	30	269	64	4.48	26	2.43		4	6		0.324	0.486		Normal
298	19	F	8.1	4.9	10.2	34	69	20	30	459	60	4.2	30	2.43		4	6		0.324	0.486		Normal
299	24	F	7.4	4.4	13.2	38	85	28	35	268	64	4.48	30	2.22		2	4		0.148	0.296		Normal
300	34	M	3.6	3.9	13.2	36	91	33	36	157	50	3.5	40	1.44		4	6		0.144	0.216		Normal
301	35	M	11.4	4.7	14.9	43	90	31	34	228	70	4.9	22	2.508		3	5		0.342	0.57		Normal
302	21	M	6.5	5.2	14.3	43	81	27	33	182	62	4.34	28	1.82		4	6		0.26	0.39		Normal
303	54	M	6.5	5.5	15.6	44	79	28	35	272	50	3.5	41	2.665		4	5		0.26	0.325		Normal