Modelling of Variables of Leukemia by Analyzing Complete Blood Count Reports of Normal and Disease Cases: A Case Study of Pakistan

By

Azka Iqbal

Master of Science in Bioinformatics

Fall 2018-MS BI-3 00000278101

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 parents.

Certificate of Originality

I hereby declare that the results presented in this research work titled as “Modelling of Variables of Leukemia by Analyzing Complete Blood Count Reports of Normal and Disease Cases: 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.

____

Azka Iqbal

(Fall 2018-MS BI-3-00000278101)

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 Hussin and Dr. Mehak Rafiq for their consistent support, guidance, and immense knowledge throughout this project. This work would not have been possible without your 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 Center for Modelling and Simulation (RCMS) for providing research faculites. I whole-heartedly appreciate the support of Principal and my Guidance Examination Committee (GEC): Dr. Muizuddin Shami, Dr. Ishrat Jabeen, Dr. Rehan Zafar Paracha, and Dr. Maria Shabbir. I would like to pay my special regards to the research group especially to my fellows Iqra Tahreem, Hira Qureshi and Ayesha Shabbir.

Last but not the least; I would like to thank my parents and my brother for supporting me.

Table of Contents

List of Abbreviations

WHO	World Health Organization
CBC	Complete Blood Count
MRD	Minimal Residual Disease
FISH	Fluorescence in situ hybridization
PCR	Polymerase Chain Reaction
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
WBC	White Blood Cell
RBC	Red Blood Cell
Hb	Haemoglobin
HCT	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
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
AML	Acute Myeloid Leukemia
CML	Chronic Myeloid Leukemia
ALL	Acute Lymphocytic Leukemia
CLL	Chronic Lymphocytic Leukemia
SGD	Stochastic Gradient Descent
PIMS	Pakistan Institute of Medical Sciences.
ASAB	Atta Ur Rahman School of Applied Biosciences
KRL	Khan Research Laboratories
SD	Standard Deviation
CC	Correlation Coefficient
OR	Odds Ratio
LRT	Likelihood Ratio Test

List of Figures

List of Tables

Abstract

Leukemia is one of the fatal diseases that originates in the bone marrow and causes abnormal proliferation of White Blood Cells (WBC), Red Blood Cells (RBC) and Platelets. A basic and usual investigation/screening test which may signify leukemia disease is CBC report. A CBC report measures the parameters and features of almost all different types of cells present in the blood. Current investigation procedure of leukemia using a CBC report is usually subjective. Thus, varies from practitioner to practitioner; hence, having high risk of mis/no diagnosis. Therefore, there is a need to develop objective data driven models for the prediction of leukemia. This study is designed to develop predictive models using logistic regression based on significant variables of a CBC report for screening of leukemia. Primary data of 302 CBC reports is collected from eight hospitals of Rawalpindi and Islamabad (twin cities of Pakistan). In these reports, 235 are disease/leukemic cases and 67 are normal/non-leukemic cases.

The analysis consists of three sections. Section I deals with pre-processing of the variables. A CBC report usually consists of 21 variables namely Age, Gender, White Blood Cell count (WBC), Red Blood Cell count (RBC), Hemoglobin (Hb), Haematocrit (HCT/PCV), Mean Corpuscular Volume (MCV), Mean Corpuscular Haemoglobin (MCH), Mean Corpuscular Haemoglobin Concentration (MCHC), Platelet Count (PLT), Neutrophil Count (Neut), Lymphocytes count (LYM), Basophil Count (BASO), Eosinophil Count (Eo), Monocytes Count (Mo) , Neutrophil Percentage, Lymphocytes Percentage, Basophil Percentage, Eosinophil Percentage, Monocytes Percentage and Reticulocytes percentage (RT). In pre-processing step, variables with high percentage of missing values have been dropped like the variable “Reticulocytes percentage” having 67.33% missing values. Overall, for any variable, all values with entry “zero (0)” are considered as missing values. In case any entry is missing in pair of values of the variables, complete entry is deleted. Therefore, a total of 15 cases have been deleted and 287 cases or entries have been used for further analysis. A CBC report includes duplicate information of few variables in terms of their counts as well as percentage, for instance, “Neutrophil”.

To avoid this duplication, variables having information of percentages have been dropped and final set of 15 variables have been selected for further analysis. These short-listed variables are Age, Gender, WBC, RBC, Haemoglobin, Haematocrit, MCV, MCH, MCHC, Platelet Count, Neutrophil Count, Lymphocyte count, Basophil Count, Eosinophil Count, Monocytes Count.

Section II provides results of independent sample t-test to compare means of Normal vs Disease cases for the 14 quantitative variables. Results show that 11 variables have significant difference between means of normal and disease cases while 3 variables Age, MCV and MCH are showing insignificant difference.

A bivariate correlation analysis has been performed to check the existence of multicollinearity in variables. Results show that variables have strong significant correlations between them. Therefore, inclusion of all the variables in the development of binary logistic regression is not appropriate and can introduce problem of multicollinearity.

Section III deals with the development of binary logistic regression model. Seven different methods of model development namely Enter Method, Forward Stepwise Selection and Backward Stepwise Elimination (using Conditional, Likelihood Ratio, Wald’s criteria) have been used in the study. Features/variables selection has been done using Wald's criteria (p-value) and the odds ratios. The results of different combinations of model specification show that 5 variables Gender, Hemoglobin, MCHC, Neutrophil Count and Monocyte Count are statistically and biologically significant for the screening of leukemic patients using CBC report. For the binary logistic model based on these 5 variables; the accuracy, sensitivity, specificity, and precision are about 92%, 94%, 86% and 95% respectively. The results of the study are useful for the physicians in decision making for the screening of leukemia using estimates of different characteristics/variables of a CBC report. A combination of objective and subjective judgment will improve accuracy and precision in early diagnosis or screening of leukemia using a common/cheaper test.

Introduction

Leukemia is one of the fatal diseases. Its morbidity and mortality costs increasing globally[1] ,[2]. Leukemia is form of cancer that originates in the bone marrow and causes abnormal proliferation of white blood cells, red blood cells and platelets which results in various infections including anaemia and bleeding, etc[3].

In leukemia, the normal differentiation pathway is blocked at a stage of differentiation at which the cells continue to proliferate, and most cells do not move on to terminal differentiation[4]. The number of new cells in normal pathway are produced by asymmetric division of tissue stem cells. This division is equal to the number of cells that terminally differentiate as an outcome, the total number of cells in the tissue remains constant. While in leukemia some of the daughter cells do not differentiate into non-dividing cells as they retain the capacity to divide and die, which results in increase in the number of cancer cells with time [4].

Figure 1.1 shows the classification of leukemia occurrence, if maturation detention happens at an early stage of differentiation, in the myeloid stem cell, the result will be an acute undifferentiated leukemia. This acute leukemia can be acute myeloid leukemia or acute lymphocyte leukemia. If arrest occurs at a later stage of differentiation (myelocyte stage), the result will be a more chronic differentiated leukemia. This chronic leukemia can be chronic myeloid leukemia or chronic lymphocytic leukemia [4].

Studies showed that its early detection is very crucial for human living as it massively affects the production of appropriate blood cells [5]. Early diagnosis of leukemia generally increases the chances for successful treatment by focusing on detecting symptomatic patients as early as possible. Delays in accessing cancer care are common with late-stage presentation, particularly in lower resource settings and vulnerable populations. The consequences of delayed or inaccessible cancer care are lower likelihood of survival, greater morbidity of treatment and higher costs of care, resulting in

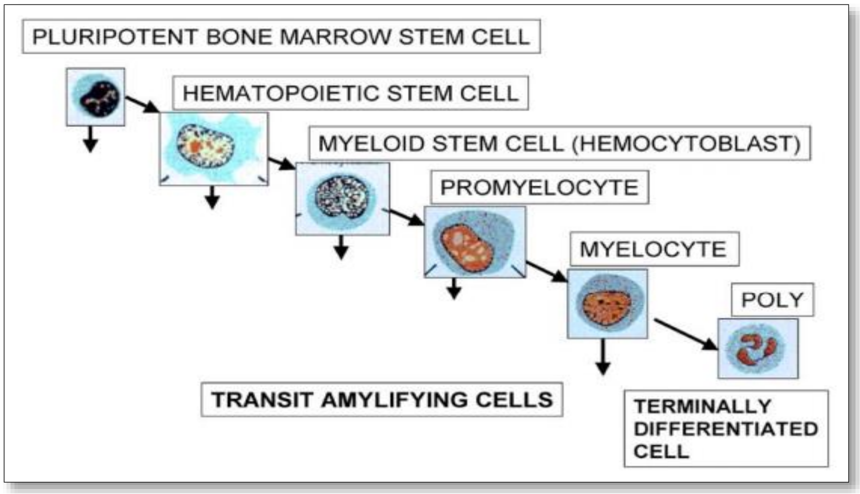

PLURIPOTENT BONE MARROW STEM CELL
HEMATOPOIETIC STEM CELL
MYELOID STEM CELL (HEMOCYTOBLAST)
PROMYELOCYTE
MYELOCYTE
POLY
TRANSIT AMYLIFYING CELLS
TERMINALLY DIFFERENTIATED CELL

Figure 1.1**: The normal renewal of blood cells that rapidly replaced through proliferation of stem and progenitor cells in the bone marrow whereas in disease it is different for Myeloid and Lymphoid Leukemia* ***[1].

deaths and disability from cancer. Early detection through analysis improves outcomes by providing care at the earliest possible stage and is therefore an important public health strategy in all settings[5].

1.1 Subtypes of leukemia:

Leukemia is further classified into four subtypes:

I. Acute Myeloid Leukemia (AML)

Insufficiency of hematopoietic cells normally results in AML and then leads toward anaemia, thrombocytopenia [5].

II. Acute Lymphoid Leukemia (ALL)

The abnormal production of lymphoid precursor cells known as lymphoblasts in the bone marrow with blocked development leads toward ALL [6]

III. Chronic Myeloid Leukemia (CML)

CLL is characterized by clonal proliferation and accumulation of B lymphocytes in the bone marrow and lymphoid tissues.

IV. Chronic Lymphoid Leukemia (CLL)

If there 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 leads toward CML [7].

1.2 Symptoms of leukemia:

The clinical symptoms of leukemia include bone and joint pain, fatigue, fever, vomiting, pale skin and loss of appetite, weight loss, liver, bleeding due to the dysfunction of platelets and spleen enlargement, shortness of breath, gums and nose bleeding[8].

1.3 Risk factors:

The growing rate of leukemia is a result of several factors, including population growth and aging, and the changing prevalence of some social and economic developments such as radiation exposure (from the explosion of atomic bomb or working in the atomic plant), infection of various viruses (human lymphotropic virus, Epstein-Barr virus, etc.) and interaction with electromagnetic fields [9]. Some of the social risk factors of leukemia are obesity and smoking as cigarettes has major cancer-causing agents[10],[11] . A person has two to four-fold increased risk among those with a first degree relative (the parent, the kid, or the sibling) suffered from leukemia[11]. Hematopoietic stem cell malignancy is also a cause for the occurrence of leukemia later in life[12].

1.4 Worldwide Statistics:

Leukemia ranks at 10th position in World cancer statistics. The prevalence of leukemia in the globally over is 4.20% and contributes to about 25% of childhood cancers. More than 300,000 new cases of leukemia (2.8% of all new cancer cases) are diagnosed annually worldwide[13].

1.5 Pakistan Statistics:

Public in Pakistan have also been tormented by leukemia. Its incidence is increasing gradually as reported in several studies [14], [15], [16] from which approximately 58 % are males and 42 % are females[1], [17].In Northern areas of Pakistan, leukemias is reported as the second common cancer and is commonly develop in the male population [15], [16].

1.6 Subjective Screening of leukemia:

Screening applies to the medical protocol for the risk of disease in the healthy population which may assist, or treatment of the detected condition based on subsequent diagnostic testing or procedures [18]. For prevention effort against leukemia, clinicians carried out various screening procedures to a person who has no signs or symptoms yet. Early diagnosis decreases the risk of infection and maximizes the probability of successful treatment[5]. Leukemia, though, is typically tested by bone marrow biopsy, immunophenotyping, blood chemistry testing, etc. Process of sample collection of these tests are painful, costly and time consuming. On the other hand, CBC is the simplest screening test.

1.6.1 Physical Examinations and Health History:

Doctors see the patients' health history like he/she has any genetic syndromes, such as Down syndrome, Fanconi anemia or viral infection or any blood disorders. Beside it they also squared essential signs to see if patient have a fever and rapid heartbeat or skin for bruising and paleness etc. [19].

1.6.2 Complete Blood Counts (CBC):

CBC is a common test suggested by doctors to detect, or monitor any disease and disorders affecting blood cells, such as anemia, infections, inflammation, bleeding disorders, or cancer. CBC report measures different parameters and features, including the number of cells and the physical properties of some of the cells. A standard CBC test includes Red blood cells (RBC) counts, Hemoglobin (HB), Haematocrit (HCT), Mean corpuscular volume (MCV), Mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), White blood cells (WBC) counts and counts of the five types of white blood cells (neutrophils, lymphocytes, monocytes, eosinophils, and basophils) and Platelets (PLT). Further examination using different tests is recommended by the doctor if there are any deregulation in any or few characteristics of the CBC report [20].Table 1.1 shows the details of CBC reports in Pakistan with their reference ranges.

***Table 1.1**: Details of a usual CBC Report* [21]**.**

Sr. #	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%	%

Litre = L, Grams per decilitre = g/dL, Femtolitre = f/L, Picograms = Pg, Microlitre = u/L

1.7 Diagnoses of Leukemia:

There exist various medical tests for the diagnose of Leukemia. These include blood chemistry tests, cytochemistry, immunophenotyping, flow cytometry, cytogenetic and molecular studies, lumbar punctures, and bone marrow biopsies and also many other test [3]. Short description of these tests is provided below table 1.2.

***Table 1.2**: Short Description of Diagnose Tests.*

Sr.#	Tests	Description
1.	Blood Chemistry Tests	Blood chemistry tests measure certain synthetic/chemical compounds in the blood.[19].
2.	Cytochemistry	Cytochemistry utilizes stains or dyes, to distinguish tissue structures and segments in blood or bone marrow cells. [22].
3.	Immunophenotyping	Immunophenotyping is the study of proteins expressed by cells. It can be used to determine the type or subtype of leukemia.[23].
4.	Flow Cytometry	Flow cytometry is a technique used to sort and classify cells using fluorescent labels on their surface [23].
5.	Lumbar puncture	A lumbar puncture, or spinal tap, removes a small amount of cerebrospinal fluid (CSF) from the space around the spine to look at it under a microscope. CSF is a fluid that surrounds the brain and spinal cord. Lumbar punctures are performed to determine if the cancer has spread to the spinal fluid[24].
6.	Bone marrow aspiration and biopsy	During bone marrow aspiration and biopsy, cells are removed from the bone marrow so that they can be examined in the laboratory. The report from the lab confirms whether the sample contains leukemia cells and if so, the type of leukemia[19].

These diagnostic tests are painful, time taking and costly such as bone marrow biopsy procedure takes 10-20 minutes for sample collection and its report duration is two to three weeks [25]. This is a painful procedure as a person feels pain for about a week. To manage this pain doctor may recommend medicines such as ibuprofen. After bone marrow biopsy a person may experience excessive bleeding, fever, and drainage. The cost of bone marrow biopsy is around 6000-8000 rupees [26]. 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.

1.8 Problem Statement:

Few or all CBC report variables have been used for the screening of various blood-related diseases, such as anaemia, infections, inflammation, bleeding disorders, or cancer. Literature also showed that machine learning algorithms have been used based on images for the detection of Leukemia. Very few studies, however, have used numerical values of various CBC report characteristics for informative and/or inferential analyses relevant to the screening of potential Leukemia patients.

Propose Solution:

The research aim is to analyze and model important CBC report variables. The outcomes of this research would be beneficial in the initial screening of suspected patients of leukemia. As screening tests are not considered diagnostic but are used to identify a subset of the population who should have additional testing to determine the presence or absence of disease as a result, the proposed model is not and cannot be used to diagnose or treat leukemia; however, it merely offers basic technical assistance for screening patients using numerical data derived from Pakistani CBC data.

1.9 Objectives:

Based on the above-mentioned details, main objectives of the study are

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

Literature Review

The success of a medical study however depends to a great extent, on the proper statistical analysis of the data originating study [27]. As statistical analysis enables a researcher to draw meaningful conclusions from a study in which data are collected through observation, survey, or experimentation. Predictive models help healthcare professionals in making clinical decisions. These techniques are capable of analyzing complex medical data. Their potential to exploit meaningful relationship within a data set can be used in the diagnosis, treatment and predicting outcome in many clinical scenarios [28].

Medical Data mining includes analyzing a relationships and patterns within the medical data that would provide useful knowledge [29]. To enhance the medical diagnosis, data mining techniques have been applied to several medical domains[30], including prediction of survival of breast cancer[31]. The feature selection is one of the important and widely used techniques for pre-processing data mining applications in medicine. Sarojini and Ramaraj used the feature selection technique to find an optimal feature subset for Type II diabetes database. Symmetrical Uncertainty Attribute Set Evaluator and Fast Correlation-Based Filter (FCBF) was utilized to excrete the subset and feature reduction was up to 62.5% [32].

Jatoi et al, in 2018 performed a study to discover the core-relationship between Anemia and Thalassemia from CBC test. CBC is the basic and usual screening test which may signify blood related disease. For this purpose, the data was collected from Liaqat Medical University Hospital Jamshoro, Pakistan. The data was based on 400 patients in which 16% were males, 53% were children and 29% were females. Naive Bayes technique was used to find out the correlation between these two diseases. The result showed that the patients suffering from anemia had Iron deficiency because of low mean corpuscular volume (MCV), Mean Corpuscular Hemoglobin (MCH) and low hemoglobin (HB); however, Vitamin B12 deficiency occurred in them due to high MCV, high MCH and low HB. The patients diagnosed with thalassemia had low MCV, low MCH and high or normal red blood cells [20].

Alshami et al. (2012) used data mining classifiers to analyze the presence of thalassemia. There were 46920 samples in the dataset used in the analysis. The study relied on CBC results that included information such as age, ethnicity, red blood cells, haemoglobin, and platelets. Decision tree, Nave Bayes, and Artificial Neural Network were the three data mining classifiers used in this study (ANN). These classifiers were used to distinguish between patients with thalassemia traits (of various levels), patients with other blood deficiencies, patients with iron deficiency, and healthy people. The results showed that ANN classifier was the most significant classifier to differentiate between the thalassemia and other blood diseases [33]

Munir *et al*, 2019 performed a descriptive study in Khyber Teaching Hospital, considering data from January 2015 to July 2017. To determine the basic pattern of hematological parameters in leukemia total data collected about 109 cases out of 117. The data of 8 patients, whose aspirates were insufficient, were excluded from the study. Their study included the cases of Chronic and Acute Leukemia by Nonprobability purposive sampling. For the remaining 109 cases, complete blood counts were done by Sysmex analyser. CBC findings were recorded and used for analysis. Mean and standard deviation were used for quantitative data, while frequency and percentages were used for qualitative data. Result shows increase in total leucocytes counts (TLC) and low hemoglobin as well platelets count patients. Thrombocytopenia and Anemia was also observed in cases. In their study, 61 cases were males, and 48 cases were females[34].

Fathi et al., 2020 conducted an analysis to see whether neuro-fizzy could be used to diagnose acute leukemia in children using a full blood count test. The information was gathered at the Tehran Children's Medical Centre in Iran. There were 346 samples in all, with 172 being ALL and 74 being AML. The data represents 110 normal and 243 leukemic cases. Haemoglobin (Hb), red blood cells (RBC), white blood Cells (WBC), platelets (Plt), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), lactate dehydrogenase (LDH), erythrocyte sedimentation rate (ESR) and Uric acid were included in this study. For the diagnosis of children with Acute Leukemia disease, they used

Principal Component Analysis (PCA), neuro-fizzy, and Group System of Data Handling (GMDH) [35].

Syed Abdul et al, in 2020 performed a study on the cell population data (CPD) to screen hematological malignancies by using Machine Learning algorithms. The data collection was completed At Konkuk University Medical Center, Seoul with total of 882 cases. In those cases 457 were hematological malignancies and 425 hematological non-malignancies cases. Seven machine learning models were used in the study: Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Random Forests (RF), Decision Tree (DT), Linear Regression, Logistic Regression, and Artificial Neural Networks (ANN). Stratified 10-fold cross validation was used to assess the efficiency of models, and metrics such as precision, accuracy, recall, and area under curve (AUC) were used for evaluation of model. Findings shows that the ANN model had superior performance relative to other ML models. The highest precision, precision, recall, and AUC±Standard Deviation was obtained by the diagnostic capacity of ANN as 82.8%, 82.8%, 84.9%, and 93.5% ±2.6. According to this study, the CPD-based ANN algorithm appeared to be an important aid for clinical laboratory screening of hematologic malignancies. Results also show that in haematological malignancies male ratio was higher than the female and myeloid leukemia was predicted with high percentage 20.7% [2].

Rathee et al, 2014 performed a study on 650 blood samples of leukemia patients in Haryana state of India. Percentage of blast cell, red blood cell indices, white blood cell indices, number of platelets and the amount of hemoglobin was determined according to standard laboratory procedures. Diagnosis of leukemia was done by 20% blast criteria and then ‘Sudan Black B’ was used to distinguish between leukemia. Analysis of variance (ANOVA) was used for interactions of factors (like age/gender/subtype) affecting leukemia patients. Data on leukemia patients was analyzed and then subjected to ANOVA. Acute and chronic leukemia accounts 71.4% and 62.6% male patients while 28.6% and 37.4% female patients, respectively. [36].

Pakistan had also been affected by leukemia, however there is limited data on the prevalence of various forms of leukemia. At present, there are no cancer registry programs in Pakistan which can keep a track and notify regarding the prevalence and incidence of various types of cancers including leukemia's. This is of utmost importance as Pakistan is a developing country and cancer is becoming a serious health issue in country. In Asia Pacific countries leukemia being the leading cause of death [2].

Ahmad et al, 2019 studied the prevalence of different types of Leukemia in Khyber Pakhtunkhwa, Pakistan during the period of January 2015 to December 2016. A sample of 400 patients were used taken from Institute of Radiotherapy and Nuclear Medicine (IRNUM) Peshawar. They observed that acute leukemia is more prevalent then chronic leukemia. About 80.75% were acute patients and 19.25% were chronic patients. In these patients, 258 were male and 152 were female. Study shows that the disease is male predominated [2].

Nighat et al, in 2013 conducted a survey analysis to investigate the prevalence of different types of leukemia. The study area was Lahore General Hospital which receives a variety of patients from various cities in the surroundings of Lahore like Kasur, Hasilpur, Dipalpur, etc. A sample size of 45 patients was collected during the period of two years from June 2010 to June 2012. Sudan Black B was used to stain the peripheral blood smears. CBC test and bone marrow biopsy were performed on dataset. The results showed that 80% of the patients were observed with acute leukemia while 20% patients were observed with chronic leukemia. The study reveals that the ratio of acute and chronic cases is 4:1 with 58% males and 42% females[1].

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 [37].

In a study by Hossain et al, 2018, computational methods were used for the detection of Leukemia by analyzing the blood cells and its components from microscopic images. This analysis involves cell classification and blast counting. Generally, healthy or infected blood cell features were extracted from an image dataset based on the morphological analysis, this process was done by Principle Component Analysis (PCA) which reduces the data dimensions without losing any valuable information. Then, the Logistic Regression classifier was employed to predict the condition of a blood cell and classify it accordingly [5].

R Bhattacharjee et al, 2012, study shows that an automated method was designed to detect Acute Lymphocytic Leukemia (ALL) and Acute Myeloid Leukemia (AML) blast cells from human microscopic blood cell images. In this research the noise reduction was done by Principle Component Analysis (PCA) which uses an orthogonal transformation to completely de-correlate the centralized matrix. Morphological operations and Connected Component Analysis were used to count the number of blast cells present in the images. The performance evaluation was carried out in terms of accuracy based on comparison of number of blast cells detected by manual count and those detected by the selective thresholding based automated method [38].

Research Gap:

Numerous model-based studies have been published using microscopic images for the prediction and detection of Leukemia. However, fewer studies have used numerical estimate of CBC report.

In Pakistan, there is a scarcity of literature on informative and inferential analysis using various variables from CBC studies for objective leukemia screening.

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 test [39].

The focus of this research is to analyze important variables of CBC reports for predictive model development. This model would be useful for initial leukemia screening. The variables used in the binary logistic regression, the source of the data, sample size, and the tools used for analysis are all discussed in this chapter. This chapter also addresses model estimation and validation. The data have been analyzed by IBM SPSS (originally, Statistical Package for the Social Sciences).

In the analysis, both quantitative and qualitative data are analyzed. Quantitative data are measures of values or counts and are expressed as numbers. Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code [40]. 14 numeric variables Age, White Blood Cells, Red Blood Cells, Hemoglobin, Hematocrit, Mean Corpuscular Volume, Mean Corpuscular Hemoglobin, Mean Corpuscular Hemoglobin Concentration Platelet Count Neutrophils Counts, Lymphocytes Counts, Basophil Counts, Eosinophil Counts, Monocytes Counts are included in the study. Categorical variables Gender is included the research. Detail of variables and its function is mentioned in Table 3.2.

The methodology includes measures related to pre-processing of the data, selecting of variables for the development of predictive model by using binary logistics regression, assessment analysis of the model, etc. Few details of these steps are provided below.

***Table 3.1**: Variables and their Functions* [41].

Sr.#	Variables	Abbreviations	Definitions
1	Gender	M/F	Male and Female
2	Age		In years
3	White Blood Cells	WBC	WBC make up about one percent of the cells in the blood by number. To recognize and target pathogens, such as invading bacteria, viruses, and other foreign species, WBC is mainly involved in the immune response.
4	Red Blood Cells	RBC	RBC are specialized cells that circulate across the body providing oxygen to cells; they are produced in the bone marrow from stem cells.
5	Hemoglobin	Hb	In red blood cells, hemoglobin is an essential protein which carries oxygen from the lungs to all parts of our body. In the CBC test, the quantity of hemoglobin is determined in the blood.
6	Hematocrit	HCT/PCV	The sum of space (volume) taken up by red blood cells in the blood in the CBC Examination. A proportion of red blood cells in the amount of blood is given as the meaning.
7	Mean Corpuscular Volume	MCV	Measures the size of RBC
8	Mean Corpuscular Hemoglobin	MCH	Calculates the hemoglobin content of each RBC
9	Mean Corpuscular Hemoglobin Concentration	MCHC	Quantity of per unit volume in RBC.
10	Platelet Count	PLT	Platelets are not cells in truth, but small cell fragments. By concentrating at the site of an injury, sticking to the lining of the wounded blood vessel, and providing a platform on which blood coagulation can occur, they facilitate the blood clotting mechanism (or coagulation).
11	Neutrophils Counts	Neut	Neutrophil is a WBC type that first responds to bacterial infection and participates in a smaller inflammatory phase.
12	Lymphocytes Counts	LYM	They produce the antibodies in body.
13	Basophil Counts	BASO	They are predominantly responsible for histamine release's antigenic and allergic reaction, inducing inflammation.
14	Eosinophil Counts	EO	They react against infections by parasites.
15	Monocytes Counts	MO	They are large devorating cells that consume endocytosis into foreign cells, dead cells, and waste cells.

3.1 Data Preprocessing:

In a statistical analysis, the first and the foremost step is pre-processing/screening of the available data for analyses. Checking feasibility of the data in terms of completeness, adequacy, and recording, etc. Data screening is usually performed to ensure that the information is adequate, reliable, and valid for testing causal theory [42]. Missing data poses many issues. First, the lack of data decreases predictive strength, which relates to the likelihood that if it is wrong the study will reject the null hypothesis. Second, missing values reduces the statistical power and can give biased results. Thirdly, this could complicate the study's research. All these issues may lead toward the invalid assumptions [43].

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. In second phases, this problem has been managed. Firstly, there were excluded cases that had more than 90% missing values and variables. All the zero are considered as missing values and Reticulocytes percentage variable were removed. In total, 15 cases were omitted, while 287 cases were analyzed further. Secondly, using the statistical form, Expected Maximization (EM), using the Statistical Package for Social Sciences (SPSS), the remaining missing values were estimated. Detail of missing variable is mentioned in figure 3.1 which were estimated.

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 [44], [45]. There are variety of techniques to manage the missing data which are Listwise or case deletion, Pairwise deletion, Mean substitution, Maximum likelihood, Expectation-Maximization, Multiple imputation. In this study Expected-Maximization (EM) method is used to estimate missing values.

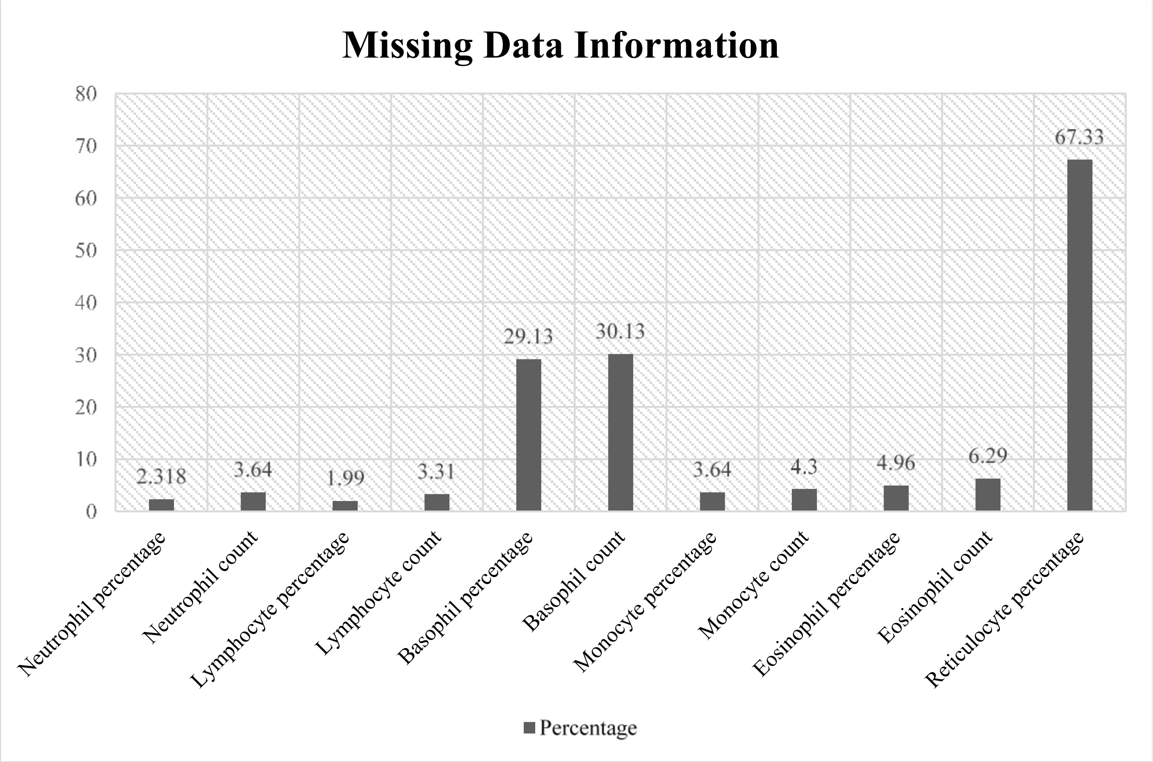


Figure 3.1: *Missing Variable Information in terms of percentage. Reticulocyte percentage variable shows the highest missing values percentage 67.33% and Lymphocyte percentage variable shows the lowest missing values percentage 1.99%*

3.2 Expected Maximization:

Expectation-Maximization (EM) is a maximum likelihood approach for creating a new data set in which all incomplete values are imputed with values calculated by maximum likelihood methods[44]. For the partially missing data, this method assumes a distribution, and bases inferences on the probability under that distribution. There is an E step and a M step in each iteration. According to observed values and current estimates, the E stage determines the conditional expectation of the "missing" data. The "missing" data is then substituted for these expectations. In the M step, the parameters' maximum probability estimates are computed as when the missing data had been filled in. "Missing" is enclosed in quotation marks so it does not explicitly fill in the missing values. Instead, functions in the log-likelihood are used for them[45].

3.3 Descriptive statistics:

Descriptive or summary statistics are used to describe the essential data characteristics of single or set of variables. These include concise summaries of the sample and measurements. Descriptive Statistics are used in a simple manner to provide quantitative explanations. It very important because if we simply presented our raw data, it would be hard to visualize what the data was showing, especially if there was a lot of it. There can be dozens of measurements in a sample analysis, or a massive number of respondents will be evaluated by each measure [46]. Common groups of descriptive statistics include measures of location, dispersion, and shape. These measures are useful to describe the trends and tendencies of the distribution of uni-variate information. To describe the trends and tendencies of each quantitative variable of CBC report, the study has used mean, standard deviation.

We have 14 variables on which descriptive analysis was performed. These variables were Age, WBC, RBC, Hemoglobin, PCV, MCV, MCH, MCHC, Platelet Count, Neutrophil count, Lymphocyte count, Basophil count, Eosinophil count, Monocytes Count, Reticulocytes percentage. For descriptive analysis we find mean, standard deviation.

3.3.1 Measures of Central tendency:

Central-tendency assessments are the most fundamental and the most detailed definition of the characteristics of a population [46].

3.3.2 Measures of Dispersion:

For a univariate information, dispersion refers to the distribution of values across the main trend. The Standard Deviation is a more common and useful approximation of dispersion for quantitative continuous normal random variables [46].

3.3.3 Independent Sample t-test:

Before proceeding for the development of logistic regression model based on significant characteristics of CBC reports, it is important to compare the distribution of variables of CBC report with respect to leukemic and non-leukemic (normal). Since the comparison of quantitative variables is based on two categories only. Therefore, independent sample t-test is an appropriate choice. The methodology for the application of t-test for comparing means of two populations include formulation of hypotheses, deciding a level of significance, calculation of test statistic and its corresponding p-value to decide in favors or against the null hypothesis.

To assess whether there is statistical evidence that the related population measures are substantially different, the Independent Samples t-test measures the means of two independent populations [47].For independent sample t-test hypothesis for each variable is:

$$H_0: \mu_{Normal} = \mu_{Disease}$$

$$H_1: \mu_{Normal} \neq \mu_{Disease}$$

$$\alpha = 0.05$$

For the application of independent sample t-test, a critical assumption is the verification of equality of variances of the two populations. A procedure based on F-distribution has been adopted for which the formulation of hypotheses is:

$$H_0: \sigma^2{}_{Normal} = \sigma^2{}_{Disease}$$

$$H_1: \sigma^2{}_{Normal} \neq \sigma^2{}_{Disease}$$

$$\alpha = 0.05$$

3.4 Coefficient of Correlation:

A coefficient of correlation, denoted by r, measures the intensity and direction of linear relationship between pairs of continuous variables [48].

Correlation coefficient range is -1 to +1. -1 shows that perfectly negative linear relationship exists between two variables and +1 shows that perfectly positive linear relationship exists between variables [48][49].Hypothesis for the correlation is show in equation below.

$$H_O : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$\alpha = 0.05$$

3.5 Model Development:

Regression analysis is normally used to classify the relationship between a dependent variable usually denoted by T and one or more set of independent variables (usually denoted by X's. If the outcome of the dependent variable is estimated or described using two or more independent variables, this is known as multiple regression. In this study diseased (leukemic cases) is dependent variable, while 14 variables are independent variables. Predicting the effect or impact of individual developments, trends and potential values and evaluating the intensity of various predictors involves the regression analysis [49].

3.5.1 Logistic Regression:

Logistic regression is a statistical model which utilizes a logistic equation in its simplest form to model a binary dependent variable. In other words, the relationship between one or more independent and dependent variable can be evaluated by logistic estimation of probability [50]. The logistic regression model is a member of the supervised classification algorithm family, and its building block concepts can support deep learning when building neural networks [51]. Logistic regression refers to any single field that contains population or categorical response variables such as biodiversity, fisheries, forestry, epidemiology, plant biology and environmental health [52].

Many statisticians assume that logistic regression is more accurate, optimal for modelling conditions, and competes with discriminatory analysis. Since logistic regression does not presume that the independent variables are normally distributed, this is assumed by discriminant analysis [53].

A logistic regression modelled the probability of a response dependent on individual characteristics. As probability is a ratio, what modeled the logarithm of the chance given by equation [54].

$$l = log_b \frac{(p)}{(1-p)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \cdots + \beta_n x_n + e_i \qquad (eq\ 3.1)$$

Here p indicates the probability of an event of disease and 1-p indicates the event of normal case and β are the regression coefficients associated with the reference group and the x independent variables whereas e_i is the error term.

The association of a dichotomous dependent variable and predictors using odds ratios is quantified by logistic regression. The odds ratio is that the chances of an occurrence are divided by the chances of the event not occurring. The odds ratio for this analysis is the likely occurrence of a disease divided by the likelihood of a disease not occurring.

Odds are calculated **u**sing the formula.

$$Odd = \frac{p\ (Disease)}{p\ (Normal)} = \frac{P\ (Disease)}{1-p(Normal)} \qquad (eq\ 3.2)$$

Where, the probability of success (Disease) and is the probability of failure (Normal). The odds ratio has a starting value zero but has no upper bound. A lower value means that in all conditions the situation is not likely to prevail, and a higher value indicates a high probability of belonging. The greater the gap between the odds ratio and one, the better the partnership [55].

The Wald statistics is another measure for the value of individual coefficients for logistic regression. If the test p-value is less than 0,05, the null hypothesis is dismissed (significance level). A coefficient with a Wald p-value below 0.05 means that the variable is significant in the model [55].

3.5.2 Logistic Regression Assumptions:

Logistic regression is very different from linear regression. Firstly, logistic regression does not involve a linear relationship between dependent and independent variables; secondly, error terms (residuals) do not usually need to be distributed normally; thirdly, homoscedasticity is not required; and in logistic regression, the dependent variable is not evaluated on an interval or ratio scale [56].

3.5.2.1 Structure of Outcome:

The important assumptions are the required outcome variable configuration, which suggests that binary logistic regression needs the binary variable to be binary in the form of 0 or 1, and ordinal logistic regression needs the ordinal variable to be dependent [49], [56].

3.5.2.2 Observation:

The repetitive measurements or paired data don't include as an observation[49], [56].

3.5.2.3 Multicollinearity:

Logistic regression demands that of all the independent variables there be little or no multicollinearity. This implies that the independent variables should not be correlated with each other too positively[49], [56].

3.5.2.4 Independent Variables and Log Odds:

While this analysis does not require that the dependent and independent variables be related linearly, it demands that the log odds be related linearly to the independent variables [49], [56].

3.5.2.5 Sample Size:

There should be large sample size and there must be at least 10 samples with the least frequent result for every independent variable in the model[49], [56].

3.5.3 Stepwise Logistic Regression:

A systematic technique for statistically sequence of events separating variables for inclusion or exclusion from a model is known as stepwise logistic regression. There are essentially two types of incremental logistic regression: forward selection and backward exclusion [57].

When there are multiple independent variables, stepwise logistic regression is often used, and it employs a series of likelihood ratio tests, conditional tests, and Wald tests to assess the variables should be used or excluded from the model. It should be stressed that no single-size model is appropriate for any situation and so two or three models must be added to the same sample for comparative purposes. Thus, it is important.

A null or primitive model (consisting only of the constant β_0) is used to start the forward selection process. Backward elimination method begins from a complete model (one that includes all feasible explanatory variables) and eliminates irrelevant variables. [58]

3.5.3.1 Stepwise Forward Selection (Conditional):

In this type of stepwise selection strategy begins with a null model, and then includes several variables one by one with the addition of meaning tests of the new variables in the evaluated model with the ranking.

The variable with the highest significant score statistic is added first, and the process is repeated until no significant variables remain outside the model. The meaning cut-off is p-value = 0.05.

Following the addition of each variable, the computer checks to see whether any variables can be omitted. The probability of the frequency ratio statistic of conditional parameter projections is used to evaluate variables for exclusion from the model [58].

3.5.3.2 Stepwise Forward Selection (Likelihood Ratio):

This is a step-by-step selection strategy with a null model and then include more variables one at a time test the importance of new variables to the model evaluated with a score statistic.

The variable with the most significant score statistic is added to the model first and this process is continued until there is no significant variable left outside the model. The cut-off for significance is p-value = 0.05 [58].

The likelihood ratio statistic of conditional parameter estimates is used to evaluate variables for exclusion from the model. This entails comparing the current model to the model after the variables has been removed. If removing a variable result in a better-fitting model, it is removed; otherwise, the variable is retained in the model [58].

3.5.3.3 Stepwise Forward Selection (Wald):

This approach begins step by step with a null model and is evaluated using the statistics of the score. The probability of the Wald statistic is used for dropping the variables. Variable with a large wald value is discarded [58].

3.5.3.4 Backward Elimination (Conditional):

This process is progressively selected, starting with a complete model (including all variables) and excluding the variables from the model with the probability of the likelihood ratio statistic of estimated conditional parameter [58].

3.5.3.5 Backward Elimination (Likelihood Ratio):

It is a step-by-step selection process that begins with a complete model and variables are omitted by the probability of the likelihood ratio dependent on partial maximum probability estimates. The existing model is then compared with the model where the element is deleted. If the removal of the variable results in a stronger model, the variable is otherwise excluded from the model [58].

3.5.3.6 Backward Elimination (Wald):

This is a step-by-step selection process beginning with an entire model and the minor variables are omitted based on the Wald statistical likelihood. Variable with a large wald value is discarded [58].

3.5.3.7 Enter Method:

The enter method is a technique for variable selection that involves including all variables at a single step. Then one by one variables are dropped on the bases of Significance or insignificance of a variable using Wald test.

3.5.4 Types of Logistic Regression:

Based on the categorical variable / dependent variable nature, there are 3 types of logistic regression. When categorical variable /response has two possible outcome it falls under the binary logistic regression. If three or more categorical variable and order of outcomes does not matter it is known as multinomial logistic regression if order of outcomes matter it is known as ordinal logistic regression [50].

3.5.4.1 Binary Logistic Regression:

A type of regression analysis, where the dependent variable is a dummy variable, is binary logistic regression. It is a variation of ordinary linear regression that is used where a dichotomous variable is the response variable and constant, categorical, or both are independent variables [59].

3.6 Model Evaluation:

In this study, the data is tested with.

- True Positive (TP) as diseased cases are correctly identified as diseased.
- False Positive (FP) as normal cases that are incorrectly identified to be diseased.
- True Negative (TN) as real normal cases that are correctly identified as normal.
- False Negative (FN) as diseased cases that are incorrectly identified as normal [18]. The 2*2 matrix is shown below:

	Actually Positive	**Actually Negative**
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Four important measures of model assessment include sensitivity, specificity, accuracy, and precision. Details and formulas of these measures are as follows:

- The outcomes of the precision output measurements tested denote the proportion of positive cases or TP predicted.

$$Senstivity \qquad P_P = \frac{TP}{TP + FN} \qquad (eq\ 3.4)$$

- Recall corresponds to sensitivity which recognizes all good cases or TP rates in medical terminology.

$$\boldsymbol{S}pecificity \qquad P_n = \frac{TN}{FN + FP} \qquad (eq\ 3.5)$$

- Accuracy estimates the right sample ratio and is one of the most intuitive and fundamental output metrics for any model [2].

$$Accuracy \qquad P = \frac{TP + TN}{TP + TN + FP + FN} \qquad (eq\ 3.6)$$

- Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives.

$$Precision \qquad PPV = \frac{TP}{TP + FP} \qquad (eq\ 3.7)$$

Results and Discussion

The results of the study are presented in this chapter. Data Collection, Comparative analysis of Normal vs Diseased case and Correlation matrix is explained in chapter. Different methods of model fitting were used for fitting multivariable binary logistic regression to predict the suspected patients of leukemia. The seven methods of model fitting were the Enter method, forward selection, and backward stepwise elimination method (Conditional, Likelihood Ratio and Wald respectively). A comparison of the models to determine the best model was also conducted.

Data collection is the procedure of collecting and evaluating information on variables of interest, in an efficient way, which allows an individual to response the stated research questions, hypotheses testing and evaluation of the results [60], [61].

A primary data of about 302 CBC reports has been collected from various hospitals of Islamabad and Rawalpindi. Detail information about CBC reports sources is mentioned in Table 4.1.

To perform this study first, the blood cells count is replaced with percentages (as we believe that these variables were carrying 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.

***Table 4.1**: Source of Information about CBC reports.*

Sr. No.	Source of information	Frequency of CBC reports	Sample Size
1.	Fauji Foundation	144-Disease 0-Normal	144
2.	Pakistan Institute of Medical Sciences (PIMS)	26-Disease 0-Normal	26
3.	Shifa International Hospital	21-Disease 0-Normal	21
4.	Atta Ur Rahman School of Applied Biosciences Diagnostic Lab (ASAB)	12-Disease 15-Normal	27
5.	Khan Research laboratories (KRL)	02-Disease 22-Normal	24
6.	Maroof International	0-Disease 11-Normal	11
7.	Quaid-e-Azam International	24-Disease 20-Normal	44
8.	Excel Labs	0-Disease 05-Normal	5
9.	Grand Total	234 -Disease 68 -Normal	302

4.1 Comparative Analysis:

Mean, standard deviation and comparative analyses of location and dispersion characteristics of each variable with respect to normal and disease using t-test and F-test respectively is provided in Table 4.2.

Out of 21, 14 variables were selected, on which descriptive analysis was performed. These variables were Age, WBC, RBC, Haemoglobin, PCV, MCV, MCH, MCHC, Platelet Count, Neutrophil count, Lymphocyte count, Basophil count, Eosinophil count, Monocytes Count. For descriptive analysis we find Mean, Standard Deviation.

Chapter 2 literature review say that leukemia is the cancer of White Blood Cells so with respect to WBC there is an increase in the mean of blood count of disease. When WBC increases there occur decrease in the RBC. The mean Red Blood Cell Counts for disease is lower than the normal. When RBC counts decreases it also effects hemoglobin and hematocrit.

WBC, RBC, Haemoglobin, Haematocrits, MCHC, Platelet Count, Neutrophil count, Lymphocyte count, Basophil count, Eosinophil count, Monocytes Count variables statistics shows that there is a significant difference between the mean of normal and disease cases.

***Table 4.2**: Comparative Analysis between Normal vs Diseased*

Sr.#	Variable	n	Mean	t Value	p Value	SD	F Test	p Value
1.	Age	N=67	N=39.25	-0.94	0.35	N=19.49	2.34	0.13
		D=220	D=36.56			D=20.67		
2.	WBC	N=67	N=8.23	5.98	0.00	N=03.18	39.76	0.00
		D=220	D=46.07			D=93.73		
3.	RBC	N=67	N=4.49	-11.96	0.00	N=0.61	8.32	0.00
		D=220	D=3.37			D=0.84		
4.	Haemoglobin	N=67	N=12.99	-12.67	0.00	N=1.76	3.90	0.05
		D=220	D=09.70			D=2.19		
5.	Haematocrit	N=67	N=38.83	-13.53	0.00	N=4.84	8.66	0.00
		D=220	D=28.79			D=6.66		
6.	MCV	N=67	N=86.15	-0.35	0.72	N=6.56	3.09	0.08
		D=220	D=85.81			D=8.38		
7.	MCH	N=67	N=28.86	0.96	0.34	N=2.90	0.15	0.70
		D=220	D=29.27			D=3.12		
8.	MCHC	N=67	N=33.34	3.19	0.00	N=1.36	7.65	0.01
		D=220	D=34.05			D=2.15		
9.	Platelet Count	N=67	N=251.75	-6.74	0.00	N=54.74	45.43	0.00
		D=220	D=168.29			D=154.65		
10.	Neutrophil Count	N=67	N=04.60	5.57	0.00	N=0.88	27.33	0.00
		D=220	D=40.44			D=95.36		
11.	lymphocyte Count	N=67	N=02.37	4.75	0.00	N=0.94	18.71	0.00
		D=220	D=08.51			D=19.04		
12.	Basophil Count	N=67	N=0.25	4.48	0.00	N=0.24	20.34	0.00
		D=220	D=0.91			D=2.15		
13.	Eosinophil Count	N=67	N=0.23	4.46	0.00	N=0.18	20.69	0.00
		D=220	D=0.92			D=2.28		
14.	Monocyte Count	N=67	N=0.48	6.71	0.00	N=0.46	39.08	0.00
		D=220	D=6.01			D=12.20		

n= number of observations, SD= Standard Deviation,N=Normal case, D= Diseased cases

The p value of t test show that we have an evidence to reject the null hypothesis and able to accept the alternative hypothesis.

Age, MCV and MCH statistics shows that there is insignificant difference between the mean of normal and disease cases. The p value of t test show that we have an evidence to accept the null hypothesis and able to reject the alternative hypothesis.

Eleven out of fourteen variables showed a mean difference which is statistically significant at 5% level of significance. These eleven variables are WBC, RBC, Haemoglobin, Haematocrit, MCHC, Platelet Count, Neutrophil Count, lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count Three variables namely Age, MCV and MCH showed a statistically insignificant difference between means of normal and disease cases. These results will also help in identification of relevant variables for inclusion in the development of logistic regression model. The binary coding for normal and disease case is ‘0’ and ‘1’ respectively.

4.2 Correlation Matrix:

Hypothesis for the correlation is show in equation below.

$$H_O : \rho = 0$$

$$H_1 : \rho \neq 0$$

As to attain objective, to develop a model using these variables as independent variables. Therefore, they should have a high correlation with the dependent variable and less or no correlation with the independent variables stated in the assumption of section 3.5.2.3. Results of correlation show that variables have strong significant correlations between them. Therefore, inclusion of all the variables in the development of binary logistic regression is not appropriate and can introduce problem of multicollinearity.

***Table 4.3**: Correlation Matrix:*

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	

CC = Correlation Coefficient, PLT Ct = Platelet Count, ANC = Absolute Neutrophil Count

LC = Lymphocyte Count, BC = Basophil Count, EC = Eosinophil Count, MC = Monocyte Count

4.2.1 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 insignificant correlation with 7 variables. These variables are RBC, Haemoglobin, Haematocrit, MCV, MCH, Platelet Count and Lymphocyte Count.

4.2.2 WBC:

WBC has statistically significant correlation with 12 variables. These variables are Age, RBC, Hemoglobin, Hematocrit, 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.

4.2.3 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 insignificant correlation with RBC.

4.2.4 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 insignificant correlation with 4 variables which are Age, MCV, MCH and MCHC.

4.2.5 PCV:

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

4.2.6 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 insignificant correlation with 5 variables. These variables are Age, Hemoglobin, Hematocrit, MCHC and Platelet Count.

4.2.7 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 insignificant correlation with 7 variables. These variables are Age, WBC, Hemoglobin, Hematocrit, Platelet Count, Lymphocyte Count and Monocyte Count.

4.2.8 MCHC:

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

4.2.9 Platelet Counts:

Platelet Count has statistically significant correlation with 7 variables. These variables are WBC, RBC, Hemoglobin, Hematocrit, 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

4.2.10 Neutrophil Counts:

Neutrophil Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, MCH, MCV, Hemoglobin, Hematocrit, 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.

4.2.11 Lymphocyte Count:

Lymphocyte Count has statistically significant correlation with 9 variables. These variables are WBC, RBC, Hematocrit, Hemoglobin, 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

4.2.12 Basophil Count:

Basophil Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, Hematocrit, Hemoglobin, 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.

4.2.13 Eosinophil Count:

Eosinophil Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, Hematocrit, Hemoglobin, 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.

4.2.14 Monocyte Counts:

Monocyte Count has statistically significant correlation with 11 variables. These variables are Age, WBC, RBC, Hematocrit, Hemoglobin, 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.

4.3 Development of Logistic Regression Model:

There are three main methods to develop the model using SPSS software.

- Enter method.
- Forward stepwise method

- Backward stepwise method.

Forward and backward methods are further divided into 3 types. Conditional, Likelihood Ratio and Wald, respectively.

4.3.1 Enter Method:

Enter method means in which all the variables in a block are entered in a single move. Then one by one variables are dropped on the bases of Significance or insignificance of a variable using Wald test.

Table 4.4 shows the value of coefficient, Wald value, Wald significant vale (p value) and odd ratio. One by one drop the following variables having insignificant values.

1. **Dropping Platelet count:** The p-value of Wald test is **0.870**.
   It has strongly positive and significant correlation with 7 variables. Highest positive relation is with WBC (0.02). Weak and insignificant correlation with 6 variables. Highly insignificant relation is with MCV (0.92).
2. **Dropping Eosinophil count:** The P-value of Wald test is **0.247**.
   It has strong positive and significant correlation with 11 variables. Highest positive relation is with MCV and age (0.01) respectively. Weak and insignificant correlation with 2 variables which are MCHC and Lymphocyte Count.
3. **Dropping WBC:** The P-value of Wald test is **0.094**.
   It has strong positive and significant correlation with 12 variables. WBC has weak and insignificant correlation with 1 variable which is MCH.
4. **Dropping Lymphocyte Count**: The P-value of Wald test is **0.375**.
   It has strong positive and significant correlation with 9 variables. Highest positive relation is with Haematocrit (0.02). It has weak and insignificant correlation with 4 variables and highly insignificant relation is with Age (0.74).

***Table 4.4**: Results of Binary Logistic Regression including all available independent.*

Sr. #	Variables	B	Wald	df	Sig.	Exp(B)
1.	Gender (1)	2.783	10.676	1	0.001	16.169
2.	Age	-0.028	3.824	1	0.051	0.973
3.	WBC	0.062	2.754	1	0.097	1.064
4.	RBC	5.256	4.399	1	0.036	191.794
5.	Hemoglobin	4.059	3.907	1	0.048	57.898
6.	Hematocrit	-1.571	3.681	1	0.055	0.208
7.	MCV	0.278	0.770	1	0.380	1.320
8.	MCH	-0.049	0.004	1	0.951	0.952
9.	MCHC	-1.822	8.261	1	0.004	0.162
10.	Platelet Count	0.000	0.029	1	0.865	1.000
11.	Neutrophil Count	-0.273	15.317	1	0.000	0.761
12.	Lymphocyte Count	-0.089	2.899	1	0.089	0.915
13.	Basophil Count	3.714	3.654	1	0.056	41.010
14.	Eosinophil Count	-1.241	1.422	1	0.233	0.289
15.	Monocyte Count	-1.936	10.010	1	0.002	0.144

5. **Dropping MCV**: The P-value of Wald test is **0.110**.

   It has strong positive and significant correlation with 8 variables. Highest positive relation is with Lymphocyte Count (0.01). It has weak and insignificant correlation with 5 and highly insignificant relation is with MCHC (0.98).

6. **Dropping RBC:** The P-value of Wald test is **0.294.**

   It has strong positive and significant correlation with 12 variables. RBC has weak and insignificant correlation with 1 variable which is age (0.09).

7. **Dropping Haematocrit:** The P-value of Wald test is **0.152.**

   It has strong positive and significant correlation with 10 variables. Haematocrit has weak and insignificant correlation with 3 variables which are Age, MCV and MCH.

8. **Dropping Age:** The P-value of Wald test is **0.52.**

   It has strong positive and significant correlation with 6 variables. It has weak and insignificant correlation with 7 variables. Highly insignificant relation is with Platelet count (0.91).

9. **Dropping Basophil count:** The P-value of Wald test **0.004,** shows that basophil count has statistically significant relation but has logical odds ratio is **27.505.**

***Table 4.5**: Final Results of Binary Logistic Regression using Enter method.*

Sr.#	Variables	B	Wald	df	Sig.	Exp(B)
1.	Gender (1)	1.716	8.095	1	0.004	5.560
2.	Haemoglobin	1.080	40.658	1	0.000	2.946
3.	MCHC	-0.628	15.826	1	0.000	.534
4.	Neutrophil Count	-0.169	10.551	1	0.001	.844
5.	Monocyte Count	-1.246	7.011	1	0.008	.288

The fitted model using the enter method is in Table 4.5.

$$log\frac{p}{1-p} = 1.716x_1 + 1.080x_2 - 0.628x_3 - 0.169x_4 - 1.246x_5 + e_i \qquad (eq\ 4.1)$$

Selected variable in enter method are x1 is Gender (1), x2 is Haemoglobin, x3 is MCHC, x4 is Neutrophil Count and x5 is Monocyte Count.

***Table 4.6**: Binary Logistic Regression Forward Stepwise method*

Sr.#	Variables	B	Wald	df	Sig.	Exp(B)
1.	Gender (1)	1.421	5.601	1	0.018	4.140
2.	Hemoglobin	1.184	46.103	1	0.000	3.268
3.	MCHC	-0.623	16.907	1	0.000	0.536
4.	Neutrophil Count	-0.153	11.065	1	0.001	0.858

4.3.2 Stepwise Forward Selection Criteria:

Forward Conditional Stepwise method, Forward Likelihood Ratio Stepwise and Forward Wald Stepwise methods show the same results.

This section provides details of the development of logistic regression using stepwise forward selection method. The process includes 5 steps to provide a suitable model. This criterion has selected “Hematocrit” as the most significant variable at first step, followed by "Gender" at step 2, then followed by "Neutrophil Count" at step 3, then followed by "MCHC" at step 4, then followed by "Hemoglobin" at the last step. Details of estimates of coefficients, their significance using Wald's method and odds ratios of the final model are provided in Table 4.6.

***Table 4.7**: Binary Logistic Regression Backward Stepwise method*

Sr.#	Variables	B	Wald	df	Sig.	Exp(B)
1.	Gender (1)	2.821	11.320	1	0.001	16.789
2.	Age	-0.029	4.775	1	0.029	0.971
3.	WBC	0.052	2.800	1	0.094	1.053
4.	RBC	5.197	4.046	1	0.044	180.784
5.	Hemoglobin	4.144	5.787	1	0.016	63.045
6.	Hematocrit	-1.584	4.981	1	0.026	0.205
7.	MCV	0.256	3.281	1	0.070	1.292
8.	MCHC	-1.903	10.025	1	0.002	0.149
9.	Neutrophil Count	-0.272	16.287	1	0.000	0.762
10.	Lymphocyte Count	-0.095	3.333	1	0.068	0.910
11.	Basophil Count	3.392	4.187	1	0.041	29.716
12.	Monocyte Count	-2.028	10.309	1	0.001	0.132

4.3.3 Stepwise Backward Elimination Criteria:

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[62].

Backward stepwise method using conditional, likelihood ratio and Wald's criteria shows the same results. Variables entered in step are Gender, Age, WBC, RBC, Haemoglobin, Haematocrit, MCV, MCH, MCHC, Platelet Count, Neutrophil Count, Lymphocyte Count, Basophil Count, Eosinophil Count and Monocyte Count. Details of estimates of coefficients, their significance using Wald's method and odds ratios of the final model are provided in Table 4.7.

***Table 4.8**: Comparison of all methods*

Sr. #	Variables	Enter	Likelihood		Wald		Conditional		Selected
			Forward	Backward	Forward	Backward	Forward	Backward	Variables
1.	Gender (1)	✓	✓	✓	✓	✓	✗	✓	6/7
2.	Age	✗	✗	✓	✗	✓	✓	✓	4/7
3.	WBC	✗	✗	✓	✗	✓	✓	✓	4/7
4.	RBC	✗	✗	✓	✗	✓	✓	✓	4/7
5.	Hemoglobin	✓	✓	✓	✓	✓	✗	✓	6/7
6.	Hematocrit	✗	✗	✓	✗	✗	✗	✓	3/7
7.	MCV	✗	✗	✓	✗	✓	✓	✓	4/7
8.	MCHC	✓	✓	✗	✓	✓	✗	✓	5/7
9.	MCH	✗	✗	✓	✗	✗	✓	✗	2/7
10.	Platelet Count	✗	✗	✗	✗	✗	✓	✗	1/7
11.	Neutrophil Count	✓	✓	✓	✓	✓	✓	✓	7/7
12.	Lymphocyte Count	✗	✗	✓	✗	✓	✓	✓	4/7
13.	Basophil Count	✗	✗	✓	✗	✓	✓	✓	4/7
14.	Eosinophil Count	✗	✗	✗	✗	✓	✓	✗	1/7
15.	Monocyte Count	✓	✗	✓	✗	✓	✓	✓	5/7

4.4 Summary of Logistic Regression Variable Selection Methods:

This study has used 7 different methods for the selection of significant variables in the development of binary logistic regression. Based on the provided details, a summary of the inclusion or exclusion of variables in summarized in Table 4.8. The table shows the set of statistically significant variables which are finally selected after the comparison of different methods. These variables are following.

1. Gender (1)
2. Hemoglobin
3. MCHC
4. Neutrophil Count
5. Monocyte Count

***Table 4.9**: Estimates, Significance, and Odds ratios of Variables of the Final Binary Logistic Model.*

Sr.#	**Variables**	**B**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% Confidence Interval**	
							Lower	Upper
1.	Gender (1)	1.989	12.402	1	0.000	7.311	2.416	22.123
2.	Hemoglobin	1.091	42.415	1	0.000	2.976	2.143	4.132
3.	Neutrophil Count	-0.161	10.802	1	0.000	0.851	0.773	0.937
4.	MCHC	-0.394	38.982	1	0.001	0.674	0.596	0.763
5.	Monocyte Count	-1.254	6.813	1	0.009	0.285	0.111	0.732

4.5 Binary Model Equation:

$$log \frac{p}{1-p} = 1.989\ (Gender) + 1.091\ (Hemoglobin) - 0.394(\ Neutrophil\ Count) - 0.161\ (MCHC) - 1.254(Monocyte\ Count) \qquad (eq\ 4.2)$$

The coefficient of Gender as shown in Table 4.9 was 1.989, this implies that $exp\ (1.989) \approx (0.000)$. Thus, a unit increase in male gender leads to an increase about 7.3% in the odds of increase in chance of occurs of disease. Thus, Male has more chance of occurs of disease as compared to the female.

The coefficient of Haemoglobin as shown in Table 4.9 was 1.091, this implies that $exp\ (1.091) \approx (0.000)$. Thus, a unit increase in haemoglobin leads to an increase about 2.98% in the odds of increase in chance of occurs of disease.

The coefficient of Neutrophil Count as shown in Table 4.9 was -0.161, this implies that $exp\ (-0.161) \approx (0.000)$. Thus, a unit increase in Neutrophil Count leads to a declined about 0.85% in the odds of increase in chance of occurs of disease. Thus, high value of Neutrophil Count is associated with in decease in chance of occurs of disease.

The coefficient of MCHC as shown in Table 4.9 was -0.394, this implies that $exp\ (-0.394) \approx (0.001)$. Thus, a unit increase in MCHC leads to a decline about 0.67% in the odds of increase in chance of occurs of disease. Thus, high value of MCHC is associated with in decease in chance of occurs disease.

The coefficient of Monocytes Count as shown in Table 4.9 was -1.254, this implies that $exp\ (-1.254) \approx (0.009)$. Thus, a unit increase in monocytes leads to a decrease about 0.28% in the odds of increase in chance of occurs of disease. Thus, high value of monocytes count is associated with the decrease in disease occurs.

The literature also gives the evidence of the presence of these variables in previous studies. Such as gender, hemoglobin and total leucocyte counts (neutrophil count and monocyte count) are also present in studies conducted by Rathee *et al*, 2014, Naeem *et al*, 2017 and Munir *et al*, 2019.

***Table 4.10**: Model Evaluation Statistics*

Case	Statistic	Statistic
Diseased (1)	TP=207	FN=13
Normal (0)	TN=57	FP=10

4.6 Model Evaluation:

Statistics shows that 13 cases of diseased people are incorrectly identified as normal out of 220 cases. These cases are knowns as False Negatives cases. While 207 cases are correctly identified as diseased case which are known as True Positive. 10 cases are identified incorrectly as diseased where they had to be predicted as normal cases. Out of 67 case 57 cases are predicted correctly as normal case.

4.6.1 Classification Accuracy:

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

$$P = \frac{207 + 57}{207 + 57 + 10 + 13}$$

$$P = \frac{264}{287}$$

$$P = 0.9198$$

In terms of percentage the accuracy is 92**%.**

4.6.2 Sensitivity:

Sensitivity is the accuracy of positive prediction or the true positive rate. The formula for calculating sensitivity is:

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

$$P_P = \frac{207}{207 + 13}$$

$$P_P = 0.94$$

In terms of percentage the sensitivity is 94**%.**

4.6.3 Specificity:

Specificity is the accuracy of negative prediction or true negative rate. The formula for calculating specificity is:

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

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

$$P_n = 0.86$$

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

4.6.4 Precision or Positive Predicted Value (PPV):

Precision is the hit rate. The formula for precision is:

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

$$PPV = \frac{207}{207 + 10}$$

$$PPV = \frac{207}{217}$$

$$PPV = 0.95$$

In terms of percentage the precision is 95**%.**

Conclusions and Recommendations

The main objective of the research is to develop a model by using Binary Logistic Regression for screening of suspected patients of Leukemia on the base of significant variables of CBC reports. Another important aim of study is to provides an analysis of the general trend and tendencies of variables considered in CBC reports by comparing means of Leukemic and non-Leukemic cases. To achieve the objective different steps are followed the major finding of steps are mentioned below:

I. CBC contain about 21 variables; out of them 15 variables are selected for the analysis by dropping information of percentages of various variables to avoid duplication.

II. Descriptive analysis shows variations in the values of mean for non-Leukemic (Normal) vs Leukemic (Disease) cases.

III. Comparative analysis by applied t-test results shows that three variables Age, MCV and MCH has statistically insignificant difference between the means of non-Leukemic (Normal) vs Leukemic (Disease) cases.

IV. To select a suitable variable to be used as independent variables in development of Binary Logistic regression model seven different combination of methods have been used. These methods are Enter method, Forward stepwise selection, and Backward stepwise elimination.

V. Final selected variables based on these methods are MCHC, haemoglobin, neutrophil count, monocyte count and gender.

VI. The performance evaluation shows that in case of Normal vs Disease the accuracy is 92%, sensitivity is 94%, specificity is 86% and precision is 95%.

The results of research showed that the established model can be used for the subjective screening of cancer, namely leukemia. The proposed model does not intend the conventional diagnostic tests for leukemia such as bone marrow biopsy, flow cytometry, etc. It provides basic technical support for the objective screening of patients using data driven models.

Therefore, the accuracy of leukemia diagnosis at an early stage can be improved by combining subjective and quantitative evaluation.

Limitations of the Study:

Every study has some limitations, below some of the limitations are mentioned about this study:

I. Class imbalance: Between the case of Normal and Disease there was a class distinction. Just 67 cases of normal relative to the case of disease were found in our sample. Instead of disease, there are 235 cases.

II. Objective Assessment: Only quantitative evaluation of cancer, i.e., leukemia, is available from this model.

III. Validation: No analysis by external evidence is available in this report

Future Recommendations:

The future suggestions for this study are:

I. Additional analytical evidence will be gathered.

II. Data would be entered to solve the problem of class imbalance.

III. The study machine learning and cluster analysis will be carried out between the CBC report variables.

References

[1] 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.

[2] N. U. R. Shujaat Ahmad , Kifayatullah , Kiramat Ali Shah , Haya Hussain, Anwar Ul Haq, Abid Ullah , Asaf Khan, "Prevalence of Acute and Chronic Forms of Leukemia in Various Regions of Khyber Pakhtunkhwa, Pakistan: Needs Much More to be done!," vol. 3154, no. 01, pp. 18–27, 2016.

[3] 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.

[4] S. Sell, "Leukemia: Stem cells, maturation arrest, and differentiation therapy," *Stem Cell Rev.*, vol. 1, no. 3, pp. 197–206, 2005, doi: 10.1385/SCR:1:3:197.

[5] 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–November, pp. 7–12, 2019, doi: 10.1109/ICTKE.2018.8612303.

[6] 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, doi: 10.1186/s13104-018-3225-9.

[7] M. Führer, "Palliative Care in Hematopoietic Stem Cell Transplantation," *Palliat. Care Pediatr. Oncol.*, vol. Springer, no. Cham, pp. 103–117, 2018.

[8] 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.

[9] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.

[10] 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, doi: 10.1186/1476-069X-9-31.

[11] 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, doi: 10.1081/cnv-200046402.

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

[13] 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, doi: 10.3322/canjclin.55.2.74.

[14] C. R. Preethi, "Clinico-hematological study of acutemyeloid leukemias," *J. Clin. Diagnostic Res.*, vol. 8, no. 4, pp. 14–17, 2014, doi: 10.7860/JCDR/2014/7854.4298.

[15] A. M. Chang, F., Shamsi, T. S., & Waryah, "Clinical and hematological profile of acute myeloid leukemia (AML) patients of Sindh.," *J. Hematol. Thromboembolic Dis.*, 2016.

[16] A. Andleeb Masood, K. M., Hussain, M., Ali, W., Riaz, M., Zafar Alauddin, M. A., Masood, M., & Shahid, "Thirty Years Cancer Incidence Data For Lahore, Pakistan: Trends And Patterns 1984-2014," *Asian Pacific J. Cancer Prev. APJCP*, vol. 19, no. 3, pp. 709–717, 2018.

[17] R. A. CARTWRIGHT, R.A., CARTWRIGHT, “Epidemiology. In: Leukaemia,” *Blackwell Sci. Publ. Oxford*, pp. 3–33, 1992.

[18] 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, doi: 10.1038/s41598-020-61247-0.

[19] M. B. Fischbach, F. T., & Dunning, “A manual of laboratory and diagnostic tests.,” 2009.

[20] S. Jatoi, M. A. Panhwar, M. S. Memon, and J. A. Baloch, “Mining Complete Blood Count Reports For Disease Discovery,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 18, no. 1, pp. 121–127, 2018.

[21] L. Dean, “Blood Groups and Red Cell Antigens,” *ABO blood Gr.*, p. Chapter 5, 2005, [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK2267.

[22] B. D. Cheson *et al.*, “Revised Recommendations of the International Working Group for diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in acute myeloid leukemia,” *J. Clin. Oncol.*, vol. 21, no. 24, pp. 4642–4649, 2003, doi: 10.1200/JCO.2003.04.036.

[23] T. Deguchi, “Immunophenotype of Pediatric ALL.,” *Pediatr. Acute Lymphoblastic Leuk. Springer, Singapore.*, pp. 29–36, 2020.

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

[25] 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, pp. 4–9, 2011, doi: 10.1186/1746-1596-6-23.

[26] 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, doi: 10.1007/s00277-012-1641-9.

[27] 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, doi: 10.1080/10691898.2018.1484674.

[28] 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, doi: 10.1308/147870804290.

[29] D. V. J. Phd, “online.” .

[30] M. Kononenko, I., & Kukar, “Machine learning for medical diagnosis.,” *Proc. CADAM*, vol. 95, 1995.

[31] D. Delen, G. Walker, and A. Kadam, “Predicting breast cancer survivability: A comparison of three data mining methods,” *Artif. Intell. Med.*, vol. 34, no. 2, pp. 113–127, 2005, doi: 10.1016/j.artmed.2004.07.002.

[32] S. Balakrishnan and R. Narayanaswamy, “Feature selection using fcbf in type II diabetes databases,” *Int. Conf. IT to Celebr. S. Charmonman’s 72nd Birthday, March 2009, Thail.*, no. March, pp. 1–8, 2009, [Online]. Available: http://ijcim.th.org/SpecialEditions/v17nSP1/pdf/02_50_Online_Sarojini.pdf%0Ahttp://www.ijcim.th.org/SpecialEditions/v17nSP1/pdf/02_50_Online_Sarojini.pdf.

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

[34] 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.

[35] 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, doi: 10.1177/0954411920938567.

[36] 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.

[37] 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. 24, no. 09, pp. 1302–1305, 2017, doi: 10.17957/tpmj/17.3942.

[38] 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, doi: 10.1109/CODIS.2012.6422148.

[39] S. B. Ali, Z., & Bhaskar, “Basic statistical tools in research and data analysis.,” *Indian J. Anaesth.*, vol. 60, no. 9, p. 662, 2016.

[40] T. C. F. M. D. &. C. J. W. Guetterman, “Integrating Quantitative And Qualitative Results In Health Science Mixed Methods Research Through Joint Displays.,” *Ann. Fam. Med.*, vol. 13, no. 6, pp. 554–561, 2015.

[41] C. Schaller, J., Gerber, S., Kaempfer, U., Lejon, S., & Trachsel, *Human blood plasma proteins: structure and function.* 2008.

[42] J. C. S. A. E. R. &. H. K. Van Den Broeck, “Data Cleaning: Detecting, Diagnosing, And Editing Data Abnormalities,” *Plos Med*, vol. 2, no. 10, p. E267, 2005.

[43] H. Kang, “The prevention and handling of the missing data,” *Korean J. Anesthesiol.*,

vol. 64, no. 5, pp. 402–406, 2013, doi: 10.4097/kjae.2013.64.5.402.

[44] Y. Dong and C. Y. J. Peng, "Principled missing data methods for researchers," *Springerplus*, vol. 2, no. 1, pp. 1–17, 2013, doi: 10.1186/2193-1801-2-222.

[45] C. B. Do and S. Batzoglou, "What is the expectation maximization algorithm?," *Nat. Biotechnol.*, vol. 26, no. 8, pp. 897–899, 2008, doi: 10.1038/nbt1406.

[46] W. M. &. D. J. P. Trochim, "Research Methods Knowledge Base.2001.," *[Online]. Available: Https://Conjointly.Com/Kb/Descriptive-Statistics/.* .

[47] K. S. University, "SPSS tutorials: independent samples t-test.," *Libraries.*, 2017.

[48] K. Yeager, "Libguides: SPSS Tutorials: Pearson Correlation," *Kent. Edu*, 2019.

[49] T. ANAMIKA, "What is Logistic Regression?," *careerfoundry*, 2020. https://careerfoundry.com/en/blog/data-analytics/what-is-logistic-regression/.

[50] J. Tolles and W. J. Meurer, "Logistic Regression Relating Patient Characteristics to Outcomes JAMA Guide to Statistics and Methods," *JAMA August*, vol. 2, no. 5, p. 533, 2016, [Online]. Available: http://jama.jamanetwork.com/.

[51] J. Brownlee, *Master Machine Learning Algorithms: discover how they work and implement them from scratch.* Machine Learning Mastery, 2016.

[52] X. Liu, "TRACE : Tennessee Research and Creative Exchange A Statistical Analysis of Key Factors Influencing the Location of Biomass-using Facilities," 2009.

[53] M. Mcgee, "Logistic Regression," *Key Top. Clin. Res.*, pp. 152–154, 2002.

[54] S. Sperandei., "Understanding Logistic Regression Analysis," *Biochem. Medica*, vol. 24, pp. 12–18, 2014.

[55] H. Muchabaiwa, “Logistic Regression To Determine Significant Factors Associated With Share Price Change,” *Dr. Diss.*, 2013.

[56] D. Schreiber-Gregory and K. Bader, “Logistic and Linear Regression Assumptions: Violation Recognition and Control,” *Midwest SAS User Gr.*, pp. 1–21, 2018, [Online]. Available: https://www.lexjansen.com/wuss/2018/130_Final_Paper_PDF.pdf.

[57] M. H. N. C. J. N. J. &. L. W. Kutner, “Applied Linear Statistical Models,” *Bost. Mcgraw-Hill Irwin.*, vol. 5, p. 2005.

[58] S. K. M. H. &. R. S. Sarkar, “Model Selection In Logistic Regression And Performance Of Its Predictive Ability,” *Aust. J. Basic Appl. Sci.*, vol. 4, no. 12, pp. 5813–5822, 2010.

[59] H. Midi, S. K. Sarkar, and S. Rana, “Collinearity diagnostics of binary logistic regression model,” *J. Interdiscip. Math.*, vol. 13, no. 3, pp. 253–267, 2010, doi: 10.1080/09720502.2010.10700699.

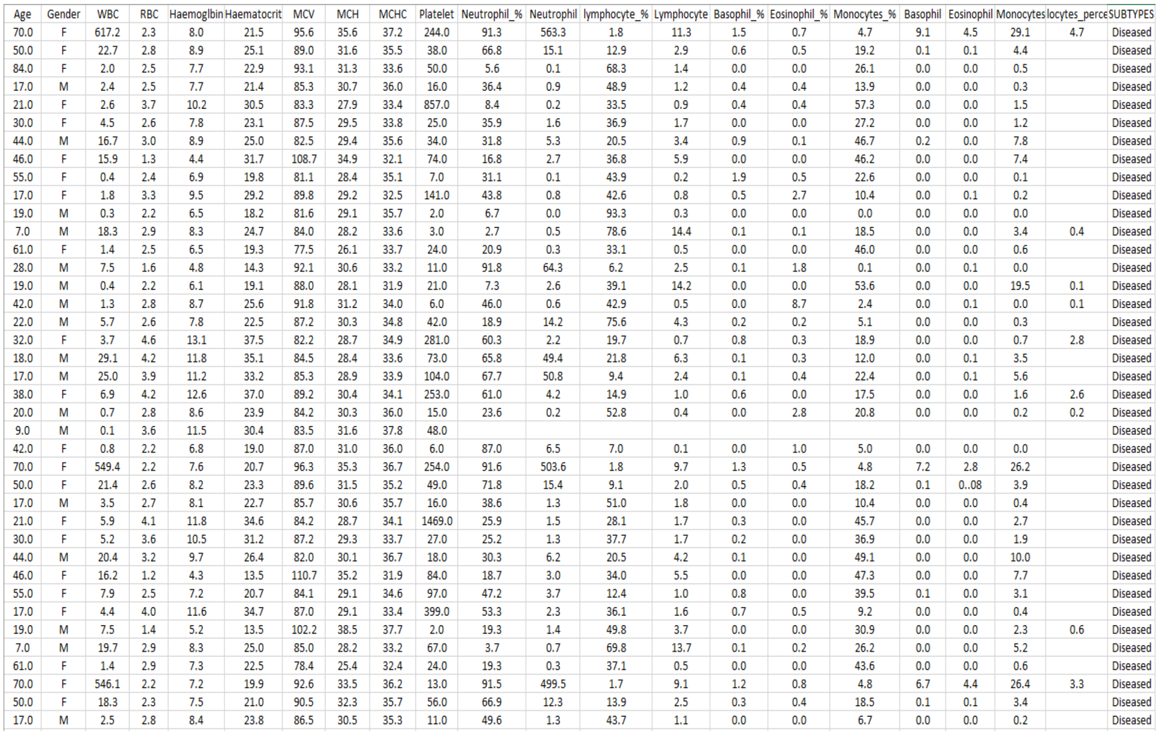
[60] S. M. S. Kabir, “Methods Of Data Collection,” *Basic Guidel. Res.*, vol. 14, no. 2, pp. 201–275, 2016.

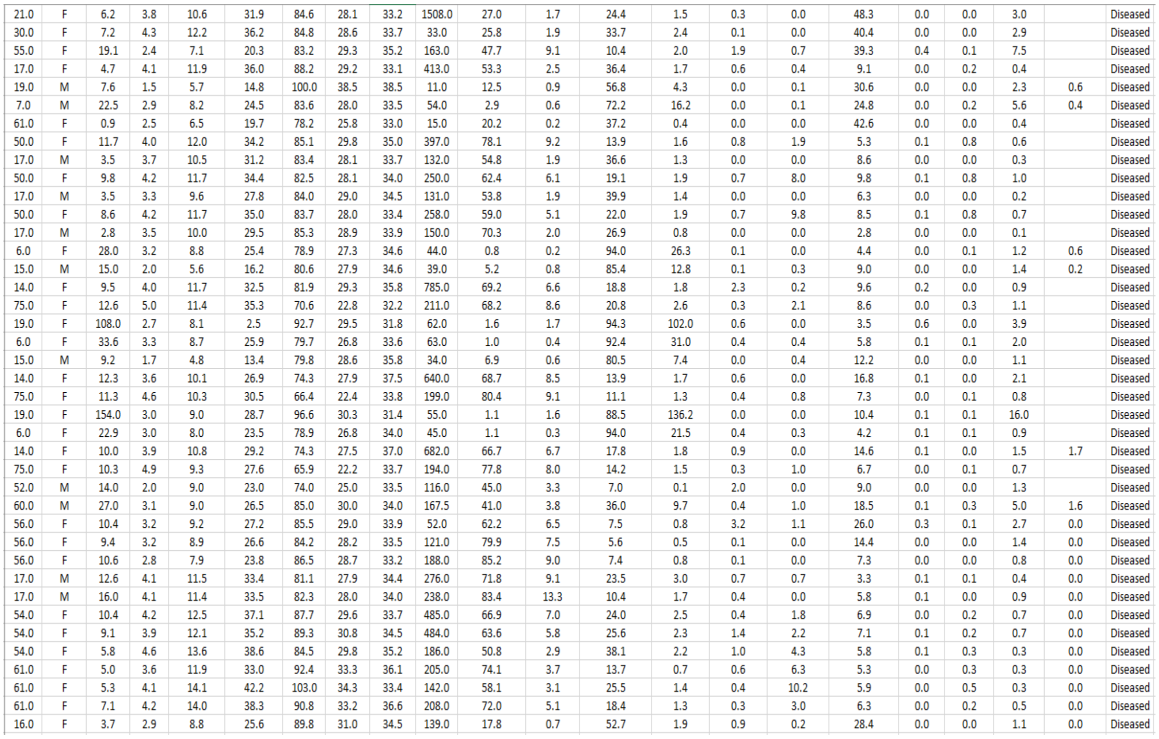
[61] R. Morley, “Data Collection And Management For Research Studies,” *Arch. Dis. Child.*, vol. 73, no. 4, p. 364, 1995.

[62] V. D. S. A. J. K. V. D. W. Gudicha, “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.

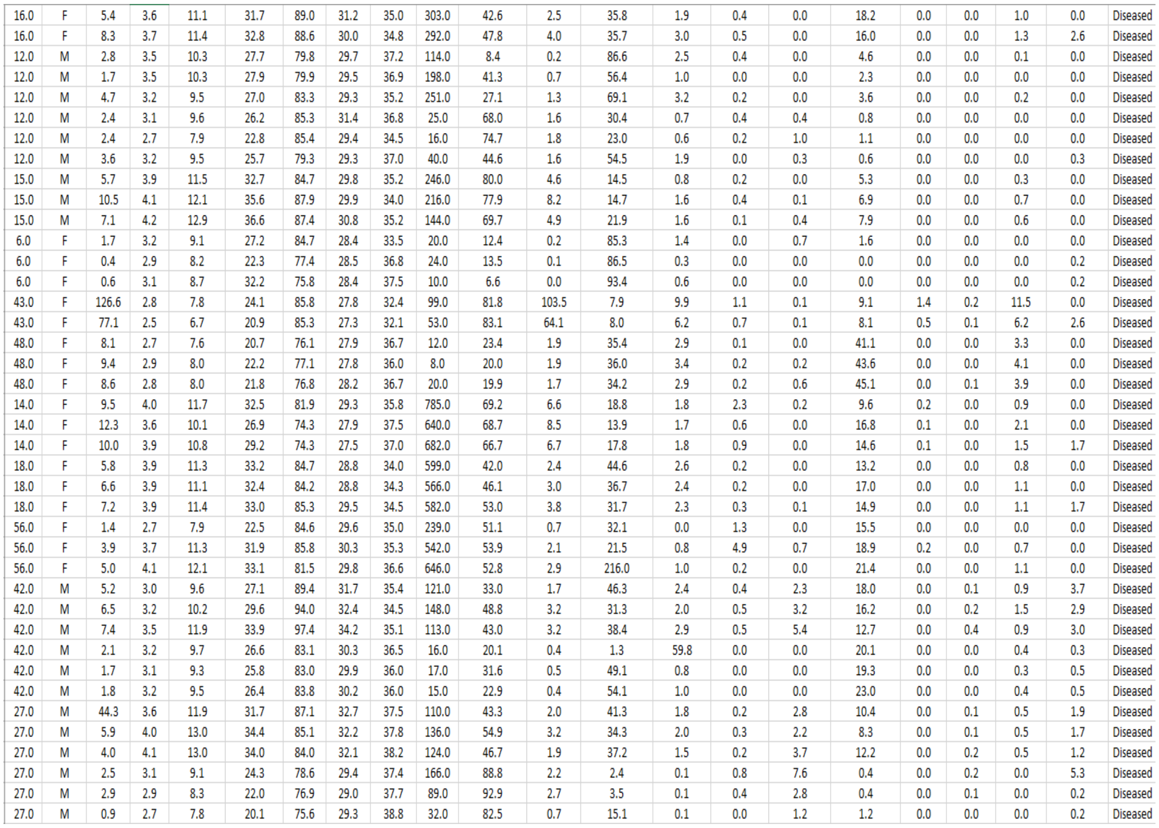
Appendix

7.1 Dataset

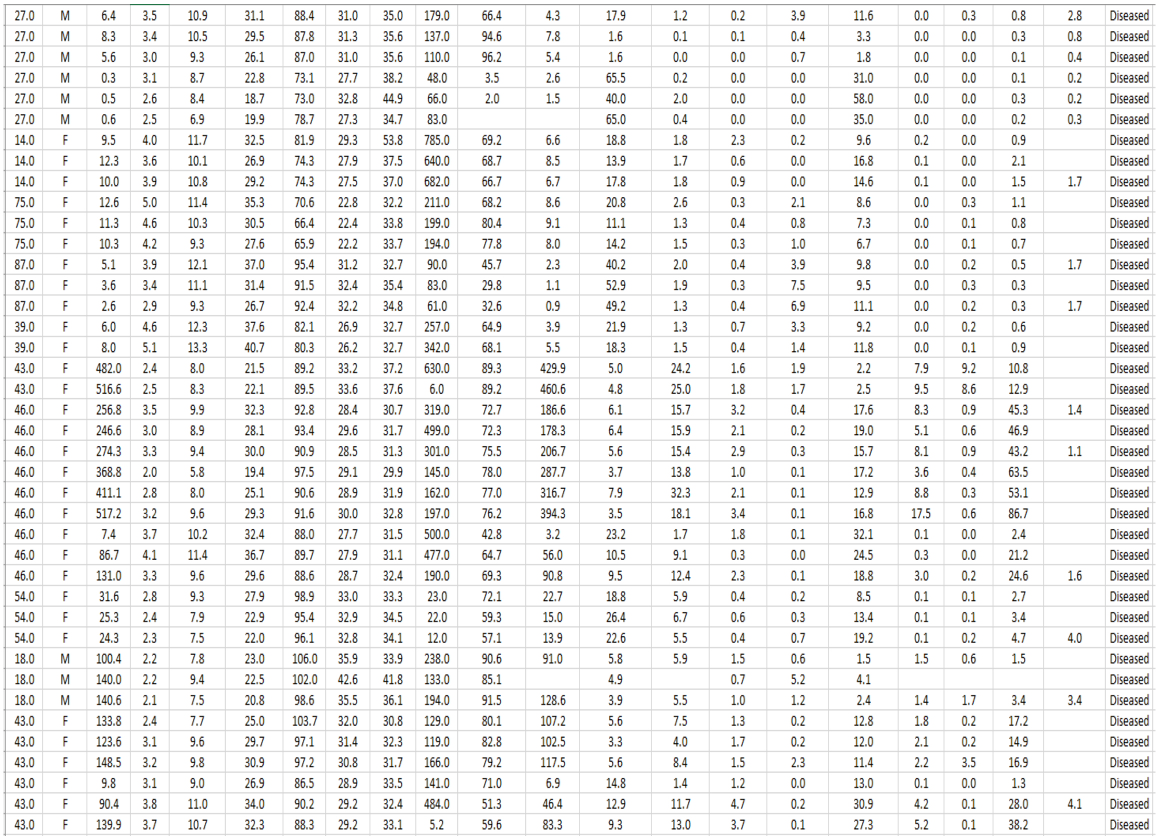




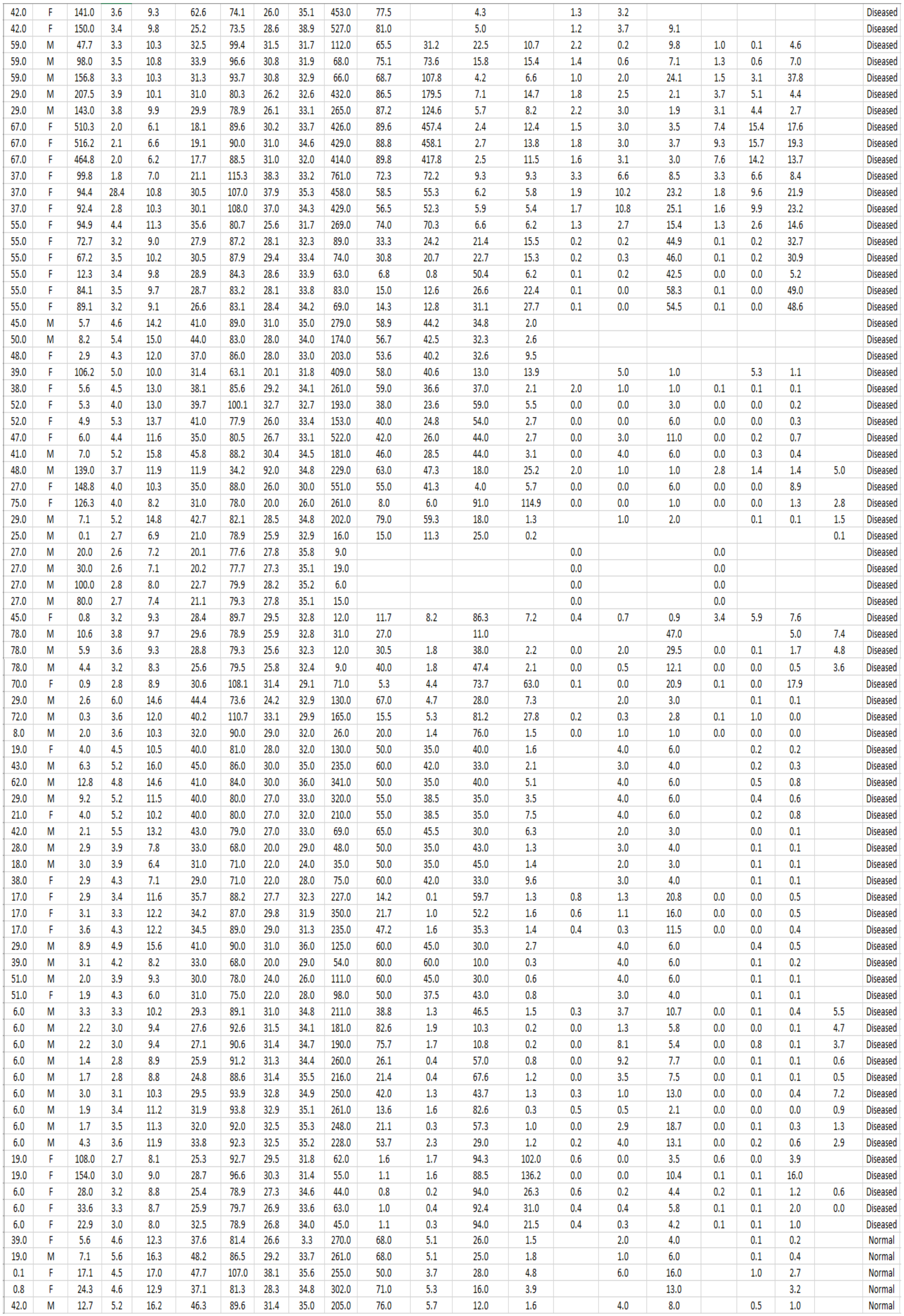
Age	Gender	WBC	RBC	Haemoglbin	Haematocrit	MCV	MCH	MCHC	Platelet	Neutrophil_%	Neutrophil	lymphocyte_%	Lymphocyte	Basophil_%	Eosinophil_%	Monocytes_%	Basophil	Eosinophil	Monocytes	locytes_perce	SUBTYPES
70.0	F	617.2	2.3	8.0	21.5	95.6	35.6	37.2	244.0	91.3	563.3	1.8	11.3	1.5	0.7	4.7	9.1	4.5	29.1	4.7	Diseased
50.0	F	22.7	2.8	8.9	25.1	89.0	31.6	35.5	38.0	66.8	15.1	12.9	2.9	0.6	0.5	19.2	0.1	0.1	4.4		Diseased
84.0	F	2.0	2.5	7.7	22.9	93.1	31.3	33.6	50.0	5.6	0.1	68.3	1.4	0.0	0.0	26.1	0.0	0.0	0.5		Diseased
17.0	M	2.4	2.5	7.7	21.4	85.3	30.7	36.0	16.0	36.4	0.9	48.9	1.2	0.4	0.4	13.9	0.0	0.0	0.3		Diseased
21.0	F	2.6	3.7	10.2	30.5	83.3	27.9	33.4	857.0	8.4	0.2	33.5	0.9	0.4	0.4	57.3	0.0	0.0	1.5		Diseased
30.0	F	4.5	2.6	7.8	23.1	87.5	29.5	33.8	25.0	35.9	1.6	36.9	1.7	0.0	0.0	27.2	0.0	0.0	1.2		Diseased
44.0	M	16.7	3.0	8.9	25.0	82.5	29.4	35.6	34.0	31.8	5.3	20.5	3.4	0.9	0.1	46.7	0.2	0.0	7.8		Diseased
46.0	F	15.9	1.3	4.4	31.7	108.7	34.9	32.1	74.0	16.8	2.7	36.8	5.9	0.0	0.0	46.2	0.0	0.0	7.4		Diseased
55.0	F	0.4	2.4	6.9	19.8	81.1	28.4	35.1	7.0	31.1	0.1	43.9	0.2	1.9	0.5	22.6	0.0	0.0	0.1		Diseased
17.0	F	1.8	3.3	9.5	29.2	89.8	29.2	32.5	141.0	43.8	0.8	42.6	0.8	0.5	2.7	10.4	0.0	0.1	0.2		Diseased
19.0	M	0.3	2.2	6.5	18.2	81.6	29.1	35.7	2.0	6.7	0.0	93.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0		Diseased
7.0	M	18.3	2.9	8.3	24.7	84.0	28.2	33.6	3.0	2.7	0.5	78.6	14.4	0.1	0.1	18.5	0.0	0.0	3.4	0.4	Diseased
61.0	F	1.4	2.5	6.5	19.3	77.5	26.1	33.7	24.0	20.9	0.3	33.1	0.5	0.0	0.0	46.0	0.0	0.0	0.6		Diseased
28.0	M	7.5	1.6	4.8	14.3	92.1	30.6	33.2	11.0	91.8	64.3	6.2	2.5	0.1	1.8	0.1	0.0	0.1	0.0		Diseased
19.0	M	0.4	2.2	6.1	19.1	88.0	28.1	31.9	21.0	7.3	2.6	39.1	14.2	0.0	0.0	53.6	0.0	0.0	19.5	0.1	Diseased
42.0	M	1.3	2.8	8.7	25.6	91.8	31.2	34.0	6.0	46.0	0.6	42.9	0.5	0.0	8.7	2.4	0.0	0.1	0.0	0.1	Diseased
22.0	M	5.7	2.6	7.8	22.5	87.2	30.3	34.8	42.0	18.9	14.2	75.6	4.3	0.2	0.2	5.1	0.0	0.0	0.3		Diseased
32.0	F	3.7	4.6	13.1	37.5	82.2	28.7	34.9	281.0	60.3	2.2	19.7	0.7	0.8	0.3	18.9	0.0	0.0	0.7	2.8	Diseased
18.0	M	29.1	4.2	11.8	35.1	84.5	28.4	33.6	73.0	65.8	49.4	21.8	6.3	0.1	0.3	12.0	0.0	0.1	3.5		Diseased
17.0	M	25.0	3.9	11.2	33.2	85.3	28.9	33.9	104.0	67.7	50.8	9.4	2.4	0.1	0.4	22.4	0.0	0.1	5.6		Diseased
38.0	F	6.9	4.2	12.6	37.0	89.2	30.4	34.1	253.0	61.0	4.2	14.9	1.0	0.6	0.0	17.5	0.0	0.0	1.6	2.6	Diseased
20.0	M	0.7	2.8	8.6	23.9	84.2	30.3	36.0	15.0	23.6	0.2	52.8	0.4	0.0	2.8	20.8	0.0	0.0	0.2	0.2	Diseased
9.0	M	0.1	3.6	11.5	30.4	83.5	31.6	37.8	48.0												Diseased
42.0	F	0.8	2.2	6.8	19.0	87.0	31.0	36.0	6.0	87.0	6.5	7.0	0.1	0.0	1.0	5.0	0.0	0.0	0.0		Diseased
70.0	F	549.4	2.2	7.6	20.7	96.3	35.3	36.7	254.0	91.6	503.6	1.8	9.7	1.3	0.5	4.8	7.2	2.8	26.2		Diseased
50.0	F	21.4	2.6	8.2	23.3	89.6	31.5	35.2	49.0	71.8	15.4	9.1	2.0	0.5	0.4	18.2	0.1	0..08	3.9		Diseased
17.0	M	3.5	2.7	8.1	22.7	85.7	30.6	35.7	16.0	38.6	1.3	51.0	1.8	0.0	0.0	10.4	0.0	0.0	0.4		Diseased
21.0	F	5.9	4.1	11.8	34.6	84.2	28.7	34.1	1469.0	25.9	1.5	28.1	1.7	0.3	0.0	45.7	0.0	0.0	2.7		Diseased
30.0	F	5.2	3.6	10.5	31.2	87.2	29.3	33.7	27.0	25.2	1.3	37.7	1.7	0.2	0.0	36.9	0.0	0.0	1.9		Diseased
44.0	M	20.4	3.2	9.7	26.4	82.0	30.1	36.7	18.0	30.3	6.2	20.5	4.2	0.1	0.0	49.1	0.0	0.0	10.0		Diseased
46.0	F	16.2	1.2	4.3	13.5	110.7	35.2	31.9	84.0	18.7	3.0	34.0	5.5	0.0	0.0	47.3	0.0	0.0	7.7		Diseased
55.0	F	7.9	2.5	7.2	20.7	84.1	29.1	34.6	97.0	47.2	3.7	12.4	1.0	0.8	0.0	39.5	0.1	0.0	3.1		Diseased
17.0	F	4.4	4.0	11.6	34.7	87.0	29.1	33.4	399.0	53.3	2.3	36.1	1.6	0.7	0.5	9.2	0.0	0.0	0.4		Diseased
19.0	M	7.5	1.4	5.2	13.5	102.2	38.5	37.7	2.0	19.3	1.4	49.8	3.7	0.0	0.0	30.9	0.0	0.0	2.3	0.6	Diseased
7.0	M	19.7	2.9	8.3	25.0	85.0	28.2	33.2	67.0	3.7	0.7	69.8	13.7	0.1	0.2	26.2	0.0	0.0	5.2		Diseased
61.0	F	1.4	2.9	7.3	22.5	78.4	25.4	32.4	24.0	19.3	0.3	37.1	0.5	0.0	0.0	43.6	0.0	0.0	0.6		Diseased
70.0	F	546.1	2.2	7.2	19.9	92.6	33.5	36.2	13.0	91.5	499.5	1.7	9.1	1.2	0.8	4.8	6.7	4.4	26.4	3.3	Diseased
50.0	F	18.3	2.3	7.5	21.0	90.5	32.3	35.7	56.0	66.9	12.3	13.9	2.5	0.3	0.4	18.5	0.1	0.1	3.4		Diseased
17.0	M	2.5	2.8	8.4	23.8	86.5	30.5	35.3	11.0	49.6	1.3	43.7	1.1	0.0	0.0	6.7	0.0	0.0	0.2		Diseased
21.0	F	6.2	3.8	10.6	31.9	84.6	28.1	33.2	1508.0	27.0	1.7	24.4	1.5	0.3	0.0	48.3	0.0	0.0	3.0		Diseased
30.0	F	7.2	4.3	12.2	36.2	84.8	28.6	33.7	33.0	25.8	1.9	33.7	2.4	0.1	0.0	40.4	0.0	0.0	2.9		Diseased
55.0	F	19.1	2.4	7.1	20.3	83.2	29.3	35.2	163.0	47.7	9.1	10.4	2.0	1.9	0.7	39.3	0.4	0.1	7.5		Diseased
17.0	F	4.7	4.1	11.9	36.0	88.2	29.2	33.1	413.0	53.3	2.5	36.4	1.7	0.6	0.4	9.1	0.0	0.2	0.4		Diseased
19.0	M	7.6	1.5	5.7	14.8	100.0	38.5	38.5	11.0	12.5	0.9	56.8	4.3	0.0	0.1	30.6	0.0	0.0	2.3	0.6	Diseased
7.0	M	22.5	2.9	8.2	24.5	83.6	28.0	33.5	54.0	2.9	0.6	72.2	16.2	0.0	0.1	24.8	0.0	0.2	5.6	0.4	Diseased
61.0	F	0.9	2.5	6.5	19.7	78.2	25.8	33.0	15.0	20.2	0.2	37.2	0.4	0.0	0.0	42.6	0.0	0.0	0.4		Diseased
50.0	F	11.7	4.0	12.0	34.2	85.1	29.8	35.0	397.0	78.1	9.2	13.9	1.6	0.8	1.9	5.3	0.1	0.8	0.6		Diseased
17.0	M	3.5	3.7	10.5	31.2	83.4	28.1	33.7	132.0	54.8	1.9	36.6	1.3	0.0	0.0	8.6	0.0	0.0	0.3		Diseased
50.0	F	9.8	4.2	11.7	34.4	82.5	28.1	34.0	250.0	62.4	6.1	19.1	1.9	0.7	8.0	9.8	0.1	0.8	1.0		Diseased
17.0	M	3.5	3.3	9.6	27.8	84.0	29.0	34.5	131.0	53.8	1.9	39.9	1.4	0.0	0.0	6.3	0.0	0.0	0.2		Diseased
50.0	F	8.6	4.2	11.7	35.0	83.7	28.0	33.4	258.0	59.0	5.1	22.0	1.9	0.7	9.8	8.5	0.1	0.8	0.7		Diseased
17.0	M	2.8	3.5	10.0	29.5	85.3	28.9	33.9	150.0	70.3	2.0	26.9	0.8	0.0	0.0	2.8	0.0	0.0	0.1		Diseased
6.0	F	28.0	3.2	8.8	25.4	78.9	27.3	34.6	44.0	0.8	0.2	94.0	26.3	0.1	0.0	4.4	0.0	0.1	1.2	0.6	Diseased
15.0	M	15.0	2.0	5.6	16.2	80.6	27.9	34.6	39.0	5.2	0.8	85.4	12.8	0.1	0.3	9.0	0.0	0.0	1.4	0.2	Diseased
14.0	F	9.5	4.0	11.7	32.5	81.9	29.3	35.8	785.0	69.2	6.6	18.8	1.8	2.3	0.2	9.6	0.2	0.0	0.9		Diseased
75.0	F	12.6	5.0	11.4	35.3	70.6	22.8	32.2	211.0	68.2	8.6	20.8	2.6	0.3	2.1	8.6	0.0	0.3	1.1		Diseased
19.0	F	108.0	2.7	8.1	2.5	92.7	29.5	31.8	62.0	1.6	1.7	94.3	102.0	0.6	0.0	3.5	0.6	0.0	3.9		Diseased
6.0	F	33.6	3.3	8.7	25.9	79.7	26.8	33.6	63.0	1.0	0.4	92.4	31.0	0.4	0.4	5.8	0.1	0.1	2.0		Diseased
15.0	M	9.2	1.7	4.8	13.4	79.8	28.6	35.8	34.0	6.9	0.6	80.5	7.4	0.0	0.4	12.2	0.0	0.0	1.1		Diseased
14.0	F	12.3	3.6	10.1	26.9	74.3	27.9	37.5	640.0	68.7	8.5	13.9	1.7	0.6	0.0	16.8	0.1	0.0	2.1		Diseased
75.0	F	11.3	4.6	10.3	30.5	66.4	22.4	33.8	199.0	80.4	9.1	11.1	1.3	0.4	0.8	7.3	0.0	0.1	0.8		Diseased
19.0	F	154.0	3.0	9.0	28.7	96.6	30.3	31.4	55.0	1.1	1.6	88.5	136.2	0.0	0.0	10.4	0.1	0.1	16.0		Diseased
6.0	F	22.9	3.0	8.0	23.5	78.9	26.8	34.0	45.0	1.1	0.3	94.0	21.5	0.4	0.3	4.2	0.1	0.1	0.9		Diseased
14.0	F	10.0	3.9	10.8	29.2	74.3	27.5	37.0	682.0	66.7	6.7	17.8	1.8	0.9	0.0	14.6	0.1	0.0	1.5	1.7	Diseased
75.0	F	10.3	4.9	9.3	27.6	65.9	22.2	33.7	194.0	77.8	8.0	14.2	1.5	0.3	1.0	6.7	0.0	0.1	0.7		Diseased
52.0	M	14.0	2.0	9.0	23.0	74.0	25.0	33.5	116.0	45.0	3.3	7.0	0.1	2.0	0.0	9.0	0.0	0.0	1.3		Diseased
60.0	M	27.0	3.1	9.0	26.5	85.0	30.0	34.0	167.5	41.0	3.8	36.0	9.7	0.4	1.0	18.5	0.1	0.3	5.0	1.6	Diseased
56.0	F	10.4	3.2	9.2	27.2	85.5	29.0	33.9	52.0	62.2	6.5	7.5	0.8	3.2	1.1	26.0	0.3	0.1	2.7	0.0	Diseased
56.0	F	9.4	3.2	8.9	26.6	84.2	28.2	33.5	121.0	79.9	7.5	5.6	0.5	0.1	0.0	14.4	0.0	0.0	1.4	0.0	Diseased
56.0	F	10.6	2.8	7.9	23.8	86.5	28.7	33.2	188.0	85.2	9.0	7.4	0.8	0.1	0.0	7.3	0.0	0.0	0.8	0.0	Diseased
17.0	M	12.6	4.1	11.5	33.4	81.1	27.9	34.4	276.0	71.8	9.1	23.5	3.0	0.7	0.7	3.3	0.1	0.1	0.4	0.0	Diseased
17.0	M	16.0	4.1	11.4	33.5	82.3	28.0	34.0	238.0	83.4	13.3	10.4	1.7	0.4	0.0	5.8	0.1	0.0	0.9	0.0	Diseased
54.0	F	10.4	4.2	12.5	37.1	87.7	29.6	33.7	485.0	66.9	7.0	24.0	2.5	0.4	1.8	6.9	0.0	0.2	0.7	0.0	Diseased
54.0	F	9.1	3.9	12.1	35.2	89.3	30.8	34.5	484.0	63.6	5.8	25.6	2.3	1.4	2.2	7.1	0.1	0.2	0.7	0.0	Diseased
54.0	F	5.8	4.6	13.6	38.6	84.5	29.8	35.2	186.0	50.8	2.9	38.1	2.2	1.0	4.3	5.8	0.1	0.3	0.3	0.0	Diseased
61.0	F	5.0	3.6	11.9	33.0	92.4	33.3	36.1	205.0	74.1	3.7	13.7	0.7	0.6	6.3	5.3	0.0	0.3	0.3	0.0	Diseased
61.0	F	5.3	4.1	14.1	42.2	103.0	34.3	33.4	142.0	58.1	3.1	25.5	1.4	0.4	10.2	5.9	0.0	0.5	0.3	0.0	Diseased
61.0	F	7.1	4.2	14.0	38.3	90.8	33.2	36.6	208.0	72.0	5.1	18.4	1.3	0.3	3.0	6.3	0.0	0.2	0.5	0.0	Diseased
16.0	F	3.7	2.9	8.8	25.6	89.8	31.0	34.5	139.0	17.8	0.7	52.7	1.9	0.9	0.2	28.4	0.0	0.0	1.1	0.0	Diseased



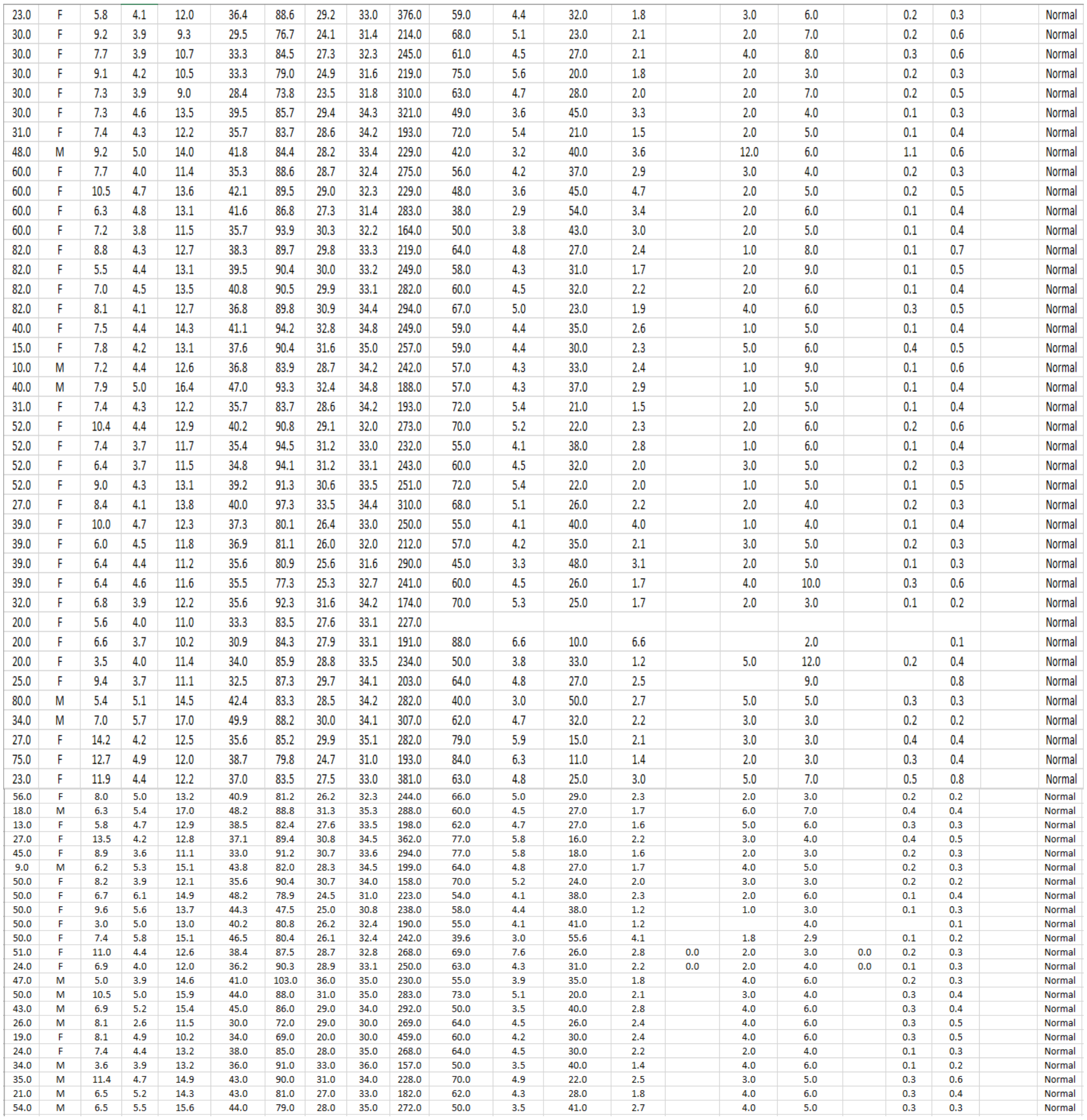
16.0	F	5.4	3.6	11.1	31.7	89.0	31.2	35.0	303.0	42.6	2.5	35.8	1.9	0.4	0.0	18.2	0.0	0.0	1.0	0.0	Diseased
16.0	F	8.3	3.7	11.4	32.8	88.6	30.0	34.8	292.0	47.8	4.0	35.7	3.0	0.5	0.0	16.0	0.0	0.0	1.3	2.6	Diseased
12.0	M	2.8	3.5	10.3	27.7	79.8	29.7	37.2	114.0	8.4	0.2	86.6	2.5	0.4	0.0	4.6	0.0	0.0	0.1	0.0	Diseased
12.0	M	1.7	3.5	10.3	27.9	79.9	29.5	36.9	198.0	41.3	0.7	56.4	1.0	0.0	0.0	2.3	0.0	0.0	0.0	0.0	Diseased
12.0	M	4.7	3.2	9.5	27.0	83.3	29.3	35.2	251.0	27.1	1.3	69.1	3.2	0.2	0.0	3.6	0.0	0.0	0.2	0.0	Diseased
12.0	M	2.4	3.1	9.6	26.2	85.3	31.4	36.8	25.0	68.0	1.6	30.4	0.7	0.4	0.4	0.8	0.0	0.0	0.0	0.0	Diseased
12.0	M	2.4	2.7	7.9	22.8	85.4	29.4	34.5	16.0	74.7	1.8	23.0	0.6	0.2	1.0	1.1	0.0	0.0	0.0	0.0	Diseased
12.0	M	3.6	3.2	9.5	25.7	79.3	29.3	37.0	40.0	44.6	1.6	54.5	1.9	0.0	0.3	0.6	0.0	0.0	0.0	0.3	Diseased
15.0	M	5.7	3.9	11.5	32.7	84.7	29.8	35.2	246.0	80.0	4.6	14.5	0.8	0.2	0.0	5.3	0.0	0.0	0.3	0.0	Diseased
15.0	M	10.5	4.1	12.1	35.6	87.9	29.9	34.0	216.0	77.9	8.2	14.7	1.6	0.4	0.1	6.9	0.0	0.0	0.7	0.0	Diseased
15.0	M	7.1	4.2	12.9	36.6	87.4	30.8	35.2	144.0	69.7	4.9	21.9	1.6	0.1	0.4	7.9	0.0	0.0	0.6	0.0	Diseased
6.0	F	1.7	3.2	9.1	27.2	84.7	28.4	33.5	20.0	12.4	0.2	85.3	1.4	0.0	0.7	1.6	0.0	0.0	0.0	0.0	Diseased
6.0	F	0.4	2.9	8.2	22.3	77.4	28.5	36.8	24.0	13.5	0.1	86.5	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.2	Diseased
6.0	F	0.6	3.1	8.7	32.2	75.8	28.4	37.5	10.0	6.6	0.0	93.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.2	Diseased
43.0	F	126.6	2.8	7.8	24.1	85.8	27.8	32.4	99.0	81.8	103.5	7.9	9.9	1.1	0.1	9.1	1.4	0.2	11.5	0.0	Diseased
43.0	F	77.1	2.5	6.7	20.9	85.3	27.3	32.1	53.0	83.1	64.1	8.0	6.2	0.7	0.1	8.1	0.5	0.1	6.2	2.6	Diseased
48.0	F	8.1	2.7	7.6	20.7	76.1	27.9	36.7	12.0	23.4	1.9	35.4	2.9	0.1	0.0	41.1	0.0	0.0	3.3	0.0	Diseased
48.0	F	9.4	2.9	8.0	22.2	77.1	27.8	36.0	8.0	20.0	1.9	36.0	3.4	0.2	0.2	43.6	0.0	0.0	4.1	0.0	Diseased
48.0	F	8.6	2.8	8.0	21.8	76.8	28.2	36.7	20.0	19.9	1.7	34.2	2.9	0.2	0.6	45.1	0.0	0.1	3.9	0.0	Diseased
14.0	F	9.5	4.0	11.7	32.5	81.9	29.3	35.8	785.0	69.2	6.6	18.8	1.8	2.3	0.2	9.6	0.2	0.0	0.9	0.0	Diseased
14.0	F	12.3	3.6	10.1	26.9	74.3	27.9	37.5	640.0	68.7	8.5	13.9	1.7	0.6	0.0	16.8	0.1	0.0	2.1	0.0	Diseased
14.0	F	10.0	3.9	10.8	29.2	74.3	27.5	37.0	682.0	66.7	6.7	17.8	1.8	0.9	0.0	14.6	0.1	0.0	1.5	1.7	Diseased
18.0	F	5.8	3.9	11.3	33.2	84.7	28.8	34.0	599.0	42.0	2.4	44.6	2.6	0.2	0.0	13.2	0.0	0.0	0.8	0.0	Diseased
18.0	F	6.6	3.9	11.1	32.4	84.2	28.8	34.3	566.0	46.1	3.0	36.7	2.4	0.2	0.0	17.0	0.0	0.0	1.1	0.0	Diseased
18.0	F	7.2	3.9	11.4	33.0	85.3	29.5	34.5	582.0	53.0	3.8	31.7	2.3	0.3	0.1	14.9	0.0	0.0	1.1	1.7	Diseased
56.0	F	1.4	2.7	7.9	22.5	84.6	29.6	35.0	239.0	51.1	0.7	32.1	0.0	1.3	0.0	15.5	0.0	0.0	0.0	0.0	Diseased
56.0	F	3.9	3.7	11.3	31.9	85.8	30.3	35.3	542.0	53.9	2.1	21.5	0.8	4.9	0.7	18.9	0.2	0.0	0.7	0.0	Diseased
56.0	F	5.0	4.1	12.1	33.1	81.5	29.8	36.6	646.0	52.8	2.9	216.0	1.0	0.2	0.0	21.4	0.0	0.0	1.1	0.0	Diseased
42.0	M	5.2	3.0	9.6	27.1	89.4	31.7	35.4	121.0	33.0	1.7	46.3	2.4	0.4	2.3	18.0	0.0	0.1	0.9	3.7	Diseased
42.0	M	6.5	3.2	10.2	29.6	94.0	32.4	34.5	148.0	48.8	3.2	31.3	2.0	0.5	3.2	16.2	0.0	0.2	1.5	2.9	Diseased
42.0	M	7.4	3.5	11.9	33.9	97.4	34.2	35.1	113.0	43.0	3.2	38.4	2.9	0.5	5.4	12.7	0.0	0.4	0.9	3.0	Diseased
42.0	M	2.1	3.2	9.7	26.6	83.1	30.3	36.5	16.0	20.1	0.4	1.3	59.8	0.0	0.0	20.1	0.0	0.0	0.4	0.3	Diseased
42.0	M	1.7	3.1	9.3	25.8	83.0	29.9	36.0	17.0	31.6	0.5	49.1	0.8	0.0	0.0	19.3	0.0	0.0	0.3	0.5	Diseased
42.0	M	1.8	3.2	9.5	26.4	83.8	30.2	36.0	15.0	22.9	0.4	54.1	1.0	0.0	0.0	23.0	0.0	0.0	0.4	0.5	Diseased
27.0	M	44.3	3.6	11.9	31.7	87.1	32.7	37.5	110.0	43.3	2.0	41.3	1.8	0.2	2.8	10.4	0.0	0.1	0.5	1.9	Diseased
27.0	M	5.9	4.0	13.0	34.4	85.1	32.2	37.8	136.0	54.9	3.2	34.3	2.0	0.3	2.2	8.3	0.0	0.1	0.5	1.7	Diseased
27.0	M	4.0	4.1	13.0	34.0	84.0	32.1	38.2	124.0	46.7	1.9	37.2	1.5	0.2	3.7	12.2	0.0	0.2	0.5	1.2	Diseased
27.0	M	2.5	3.1	9.1	24.3	78.6	29.4	37.4	166.0	88.8	2.2	2.4	0.1	0.8	7.6	0.4	0.0	0.2	0.0	5.3	Diseased
27.0	M	2.9	2.9	8.3	22.0	76.9	29.0	37.7	89.0	92.9	2.7	3.5	0.1	0.4	2.8	0.4	0.0	0.1	0.0	0.2	Diseased
27.0	M	0.9	2.7	7.8	20.1	75.6	29.3	38.8	32.0	82.5	0.7	15.1	0.1	0.0	1.2	1.2	0.0	0.0	0.0	0.2	Diseased



27.0	M	6.4	3.5	10.9	31.1	88.4	31.0	35.0	179.0	66.4	4.3	17.9	1.2	0.2	3.9	11.6	0.0	0.3	0.8	2.8	Diseased
27.0	M	8.3	3.4	10.5	29.5	87.8	31.3	35.6	137.0	94.6	7.8	1.6	0.1	0.1	0.4	3.3	0.0	0.0	0.3	0.8	Diseased
27.0	M	5.6	3.0	9.3	26.1	87.0	31.0	35.6	110.0	96.2	5.4	1.6	0.0	0.0	0.7	1.8	0.0	0.0	0.1	0.4	Diseased
27.0	M	0.3	3.1	8.7	22.8	73.1	27.7	38.2	48.0	3.5	2.6	65.5	0.2	0.0	0.0	31.0	0.0	0.0	0.1	0.2	Diseased
27.0	M	0.5	2.6	8.4	18.7	73.0	32.8	44.9	66.0	2.0	1.5	40.0	2.0	0.0	0.0	58.0	0.0	0.0	0.3	0.2	Diseased
27.0	M	0.6	2.5	6.9	19.9	78.7	27.3	34.7	83.0			65.0	0.4	0.0	0.0	35.0	0.0	0.0	0.2	0.3	Diseased
14.0	F	9.5	4.0	11.7	32.5	81.9	29.3	53.8	785.0	69.2	6.6	18.8	1.8	2.3	0.2	9.6	0.2	0.0	0.9		Diseased
14.0	F	12.3	3.6	10.1	26.9	74.3	27.9	37.5	640.0	68.7	8.5	13.9	1.7	0.6	0.0	16.8	0.1	0.0	2.1		Diseased
14.0	F	10.0	3.9	10.8	29.2	74.3	27.5	37.0	682.0	66.7	6.7	17.8	1.8	0.9	0.0	14.6	0.1	0.0	1.5	1.7	Diseased
75.0	F	12.6	5.0	11.4	35.3	70.6	22.8	32.2	211.0	68.2	8.6	20.8	2.6	0.3	2.1	8.6	0.0	0.3	1.1		Diseased
75.0	F	11.3	4.6	10.3	30.5	66.4	22.4	33.8	199.0	80.4	9.1	11.1	1.3	0.4	0.8	7.3	0.0	0.1	0.8		Diseased
75.0	F	10.3	4.2	9.3	27.6	65.9	22.2	33.7	194.0	77.8	8.0	14.2	1.5	0.3	1.0	6.7	0.0	0.1	0.7		Diseased
87.0	F	5.1	3.9	12.1	37.0	95.4	31.2	32.7	90.0	45.7	2.3	40.2	2.0	0.4	3.9	9.8	0.0	0.2	0.5	1.7	Diseased
87.0	F	3.6	3.4	11.1	31.4	91.5	32.4	35.4	83.0	29.8	1.1	52.9	1.9	0.3	7.5	9.5	0.0	0.3	0.3		Diseased
87.0	F	2.6	2.9	9.3	26.7	92.4	32.2	34.8	61.0	32.6	0.9	49.2	1.3	0.4	6.9	11.1	0.0	0.2	0.3	1.7	Diseased
39.0	F	6.0	4.6	12.3	37.6	82.1	26.9	32.7	257.0	64.9	3.9	21.9	1.3	0.7	3.3	9.2	0.0	0.2	0.6		Diseased
39.0	F	8.0	5.1	13.3	40.7	80.3	26.2	32.7	342.0	68.1	5.5	18.3	1.5	0.4	1.4	11.8	0.0	0.1	0.9		Diseased
43.0	F	482.0	2.4	8.0	21.5	89.2	33.2	37.2	630.0	89.3	429.9	5.0	24.2	1.6	1.9	2.2	7.9	9.2	10.8		Diseased
43.0	F	516.6	2.5	8.3	22.1	89.5	33.6	37.6	6.0	89.2	460.6	4.8	25.0	1.8	1.7	2.5	9.5	8.6	12.9		Diseased
46.0	F	256.8	3.5	9.9	32.3	92.8	28.4	30.7	319.0	72.7	186.6	6.1	15.7	3.2	0.4	17.6	8.3	0.9	45.3	1.4	Diseased
46.0	F	246.6	3.0	8.9	28.1	93.4	29.6	31.7	499.0	72.3	178.3	6.4	15.9	2.1	0.2	19.0	5.1	0.6	46.9		Diseased
46.0	F	274.3	3.3	9.4	30.0	90.9	28.5	31.3	301.0	75.5	206.7	5.6	15.4	2.9	0.3	15.7	8.1	0.9	43.2	1.1	Diseased
46.0	F	368.8	2.0	5.8	19.4	97.5	29.1	29.9	145.0	78.0	287.7	3.7	13.8	1.0	0.1	17.2	3.6	0.4	63.5		Diseased
46.0	F	411.1	2.8	8.0	25.1	90.6	28.9	31.9	162.0	77.0	316.7	7.9	32.3	2.1	0.1	12.9	8.8	0.3	53.1		Diseased
46.0	F	517.2	3.2	9.6	29.3	91.6	30.0	32.8	197.0	76.2	394.3	3.5	18.1	3.4	0.1	16.8	17.5	0.6	86.7		Diseased
46.0	F	7.4	3.7	10.2	32.4	88.0	27.7	31.5	500.0	42.8	3.2	23.2	1.7	1.8	0.1	32.1	0.1	0.0	2.4		Diseased
46.0	F	86.7	4.1	11.4	36.7	89.7	27.9	31.1	477.0	64.7	56.0	10.5	9.1	0.3	0.0	24.5	0.3	0.0	21.2		Diseased
46.0	F	131.0	3.3	9.6	29.6	88.6	28.7	32.4	190.0	69.3	90.8	9.5	12.4	2.3	0.1	18.8	3.0	0.2	24.6	1.6	Diseased
54.0	F	31.6	2.8	9.3	27.9	98.9	33.0	33.3	23.0	72.1	22.7	18.8	5.9	0.4	0.2	8.5	0.1	0.1	2.7		Diseased
54.0	F	25.3	2.4	7.9	22.9	95.4	32.9	34.5	22.0	59.3	15.0	26.4	6.7	0.6	0.3	13.4	0.1	0.1	3.4		Diseased
54.0	F	24.3	2.3	7.5	22.0	96.1	32.8	34.1	12.0	57.1	13.9	22.6	5.5	0.4	0.7	19.2	0.1	0.2	4.7	4.0	Diseased
18.0	M	100.4	2.2	7.8	23.0	106.0	35.9	33.9	238.0	90.6	91.0	5.8	5.9	1.5	0.6	1.5	1.5	0.6	1.5		Diseased
18.0	M	140.0	2.2	9.4	22.5	102.0	42.6	41.8	133.0	85.1		4.9		0.7	5.2	4.1					Diseased
18.0	M	140.6	2.1	7.5	20.8	98.6	35.5	36.1	194.0	91.5	128.6	3.9	5.5	1.0	1.2	2.4	1.4	1.7	3.4	3.4	Diseased
43.0	F	133.8	2.4	7.7	25.0	103.7	32.0	30.8	129.0	80.1	107.2	5.6	7.5	1.3	0.2	12.8	1.8	0.2	17.2		Diseased
43.0	F	123.6	3.1	9.6	29.7	97.1	31.4	32.3	119.0	82.8	102.5	3.3	4.0	1.7	0.2	12.0	2.1	0.2	14.9		Diseased
43.0	F	148.5	3.2	9.8	30.9	97.2	30.8	31.7	166.0	79.2	117.5	5.6	8.4	1.5	2.3	11.4	2.2	3.5	16.9		Diseased
43.0	F	9.8	3.1	9.0	26.9	86.5	28.9	33.5	141.0	71.0	6.9	14.8	1.4	1.2	0.0	13.0	0.1	0.0	1.3		Diseased
43.0	F	90.4	3.8	11.0	34.0	90.2	29.2	32.4	484.0	51.3	46.4	12.9	11.7	4.7	0.2	30.9	4.2	0.1	28.0	4.1	Diseased
43.0	F	139.9	3.7	10.7	32.3	88.3	29.2	33.1	5.2	59.6	83.3	9.3	13.0	3.7	0.1	27.3	5.2	0.1	38.2		Diseased



42.0	F	141.0	3.6	9.3	62.6	74.1	26.0	35.1	453.0	77.5		4.3		1.3	3.2						Diseased
42.0	F	150.0	3.4	9.8	25.2	73.5	28.6	38.9	527.0	81.0		5.0		1.2	3.7	9.1					Diseased
59.0	M	47.7	3.3	10.3	32.5	99.4	31.5	31.7	112.0	65.5	31.2	22.5	10.7	2.2	0.2	9.8	1.0	0.1	4.6		Diseased
59.0	M	98.0	3.5	10.8	33.9	96.6	30.8	31.9	68.0	75.1	73.6	15.8	15.4	1.4	0.6	7.1	1.3	0.6	7.0		Diseased
59.0	M	156.8	3.3	10.3	31.3	93.7	30.8	32.9	66.0	68.7	107.8	4.2	6.6	1.0	2.0	24.1	1.5	3.1	37.8		Diseased
29.0	M	207.5	3.9	10.1	31.0	80.3	26.2	32.6	432.0	86.5	179.5	7.1	14.7	1.8	2.5	2.1	3.7	5.1	4.4		Diseased
29.0	M	143.0	3.8	9.9	29.9	78.9	26.1	33.1	265.0	87.2	124.6	5.7	8.2	2.2	3.0	1.9	3.1	4.4	2.7		Diseased
67.0	F	510.3	2.0	6.1	18.1	89.6	30.2	33.7	426.0	89.6	457.4	2.4	12.4	1.5	3.0	3.5	7.4	15.4	17.6		Diseased
67.0	F	516.2	2.1	6.6	19.1	90.0	31.0	34.6	429.0	88.8	458.1	2.7	13.8	1.8	3.0	3.7	9.3	15.7	19.3		Diseased
67.0	F	464.8	2.0	6.2	17.7	88.5	31.0	32.0	414.0	89.8	417.8	2.5	11.5	1.6	3.1	3.0	7.6	14.2	13.7		Diseased
37.0	F	99.8	1.8	7.0	21.1	115.3	38.3	33.2	761.0	72.3	72.2	9.3	9.3	3.3	6.6	8.5	3.3	6.6	8.4		Diseased
37.0	F	94.4	28.4	10.8	30.5	107.0	37.9	35.3	458.0	58.5	55.3	6.2	5.8	1.9	10.2	23.2	1.8	9.6	21.9		Diseased
37.0	F	92.4	2.8	10.3	30.1	108.0	37.0	34.3	429.0	56.5	52.3	5.9	5.4	1.7	10.8	25.1	1.6	9.9	23.2		Diseased
55.0	F	94.9	4.4	11.3	35.6	80.7	25.6	31.7	269.0	74.0	70.3	6.6	6.2	1.3	2.7	15.4	1.3	2.6	14.6		Diseased
55.0	F	72.7	3.2	9.0	27.9	87.2	28.1	32.3	89.0	33.3	24.2	21.4	15.5	0.2	0.2	44.9	0.1	0.2	32.7		Diseased
55.0	F	67.2	3.5	10.2	30.5	87.9	29.4	33.4	74.0	30.8	20.7	22.7	15.3	0.2	0.3	46.0	0.1	0.2	30.9		Diseased
55.0	F	12.3	3.4	9.8	28.9	84.3	28.6	33.9	63.0	6.8	0.8	50.4	6.2	0.1	0.2	42.5	0.0	0.0	5.2		Diseased
55.0	F	84.1	3.5	9.7	28.7	83.2	28.1	33.8	83.0	15.0	12.6	26.6	22.4	0.1	0.0	58.3	0.1	0.0	49.0		Diseased
55.0	F	89.1	3.2	9.1	26.6	83.1	28.4	34.2	69.0	14.3	12.8	31.1	27.7	0.1	0.0	54.5	0.1	0.0	48.6		Diseased
45.0	M	5.7	4.6	14.2	41.0	89.0	31.0	35.0	279.0	58.9	44.2	34.8	2.0								Diseased
50.0	M	8.2	5.4	15.0	44.0	83.0	28.0	34.0	174.0	56.7	42.5	32.3	2.6								Diseased
48.0	F	2.9	4.3	12.0	37.0	86.0	28.0	33.0	203.0	53.6	40.2	32.6	9.5								Diseased
39.0	F	106.2	5.0	10.0	31.4	63.1	20.1	31.8	409.0	58.0	40.6	13.0	13.9		5.0	1.0		5.3	1.1		Diseased
38.0	F	5.6	4.5	13.0	38.1	85.6	29.2	34.1	261.0	59.0	36.6	37.0	2.1	2.0	1.0	1.0	0.1	0.1	0.1		Diseased
52.0	F	5.3	4.0	13.0	39.7	100.1	32.7	32.7	193.0	38.0	23.6	59.0	5.5	0.0	0.0	3.0	0.0	0.0	0.2		Diseased
52.0	F	4.9	5.3	13.7	41.0	77.9	26.0	33.4	153.0	40.0	24.8	54.0	2.7	0.0	0.0	6.0	0.0	0.0	0.3		Diseased
47.0	F	6.0	4.4	11.6	35.0	80.5	26.7	33.1	522.0	42.0	26.0	44.0	2.7	0.0	3.0	11.0	0.0	0.2	0.7		Diseased
41.0	M	7.0	5.2	15.8	45.8	88.2	30.4	34.5	181.0	46.0	28.5	44.0	3.1	0.0	4.0	6.0	0.0	0.3	0.4		Diseased
48.0	M	139.0	3.7	11.9	11.9	34.2	92.0	34.8	229.0	63.0	47.3	18.0	25.2	2.0	1.0	1.0	2.8	1.4	1.4	5.0	Diseased
27.0	F	148.8	4.0	10.3	35.0	88.0	26.0	30.0	551.0	55.0	41.3	4.0	5.7	0.0	0.0	6.0	0.0	0.0	8.9		Diseased
75.0	F	126.3	4.0	8.2	31.0	78.0	20.0	26.0	261.0	8.0	6.0	91.0	114.9	0.0	0.0	1.0	0.0	0.0	1.3	2.8	Diseased
29.0	M	7.1	5.2	14.8	42.7	82.1	28.5	34.8	202.0	79.0	59.3	18.0	1.3		1.0	2.0		0.1	0.1	1.5	Diseased
25.0	M	0.1	2.7	6.9	21.0	78.9	25.9	32.9	16.0	15.0	11.3	25.0	0.2							0.1	Diseased
27.0	M	20.0	2.6	7.2	20.1	77.6	27.8	35.8	9.0					0.0			0.0				Diseased
27.0	M	30.0	2.6	7.1	20.2	77.7	27.3	35.1	19.0					0.0			0.0				Diseased
27.0	M	100.0	2.8	8.0	22.7	79.9	28.2	35.2	6.0					0.0			0.0				Diseased
27.0	M	80.0	2.7	7.4	21.1	79.3	27.8	35.1	15.0					0.0			0.0				Diseased
45.0	F	0.8	3.2	9.3	28.4	89.7	29.5	32.8	12.0	11.7	8.2	86.3	7.2	0.4	0.7	0.9	3.4	5.9	7.6		Diseased
78.0	M	10.6	3.8	9.7	29.6	78.9	25.9	32.8	31.0	27.0		11.0				47.0			5.0	7.4	Diseased
78.0	M	5.9	3.6	9.3	28.8	79.3	25.6	32.3	12.0	30.5	1.8	38.0	2.2	0.0	2.0	29.5	0.0	0.1	1.7	4.8	Diseased
78.0	M	4.4	3.2	8.3	25.6	79.5	25.8	32.4	9.0	40.0	1.8	47.4	2.1	0.0	0.5	12.1	0.0	0.0	0.5	3.6	Diseased
70.0	F	0.9	2.8	8.9	30.6	108.1	31.4	29.1	71.0	5.3	4.4	73.7	63.0	0.1	0.0	20.9	0.1	0.0	17.9		Diseased
29.0	M	2.6	6.0	14.6	44.4	73.6	24.2	32.9	130.0	67.0	4.7	28.0	7.3		2.0	3.0		0.1	0.1		Diseased
72.0	M	0.3	3.6	12.0	40.2	110.7	33.1	29.9	165.0	15.5	5.3	81.2	27.8	0.2	0.3	2.8	0.1	1.0	0.0		Diseased
8.0	M	2.0	3.6	10.3	32.0	90.0	29.0	32.0	26.0	20.0	1.4	76.0	1.5	0.0	1.0	1.0	0.0	0.0	0.0		Diseased
19.0	F	4.0	4.5	10.5	40.0	81.0	28.0	32.0	130.0	50.0	35.0	40.0	1.6		4.0	6.0		0.2	0.2		Diseased
43.0	M	6.3	5.2	16.0	45.0	86.0	30.0	35.0	235.0	60.0	42.0	33.0	2.1		3.0	4.0		0.2	0.3		Diseased
62.0	M	12.8	4.8	14.6	41.0	84.0	30.0	36.0	341.0	50.0	35.0	40.0	5.1		4.0	6.0		0.5	0.8		Diseased
29.0	M	9.2	5.2	11.5	40.0	80.0	27.0	33.0	320.0	55.0	38.5	35.0	3.5		4.0	6.0		0.4	0.6		Diseased
21.0	F	4.0	5.2	10.2	40.0	80.0	27.0	32.0	210.0	55.0	38.5	35.0	7.5		4.0	6.0		0.2	0.8		Diseased
42.0	M	2.1	5.5	13.2	43.0	79.0	27.0	33.0	69.0	65.0	45.5	30.0	6.3		2.0	3.0		0.0	0.1		Diseased
28.0	M	2.9	3.9	7.8	33.0	68.0	20.0	29.0	48.0	50.0	35.0	43.0	1.3		3.0	4.0		0.1	0.1		Diseased
18.0	M	3.0	3.9	6.4	31.0	71.0	22.0	24.0	35.0	50.0	35.0	45.0	1.4		2.0	3.0		0.1	0.1		Diseased
38.0	F	2.9	4.3	7.1	29.0	71.0	22.0	28.0	75.0	60.0	42.0	33.0	9.6		3.0	4.0		0.1	0.1		Diseased
17.0	F	2.9	3.4	11.6	35.7	88.2	27.7	32.3	227.0	14.2	0.1	59.7	1.3	0.8	1.3	20.8	0.0	0.0	0.5		Diseased
17.0	F	3.1	3.3	12.2	34.2	87.0	29.8	31.9	350.0	21.7	1.0	52.2	1.6	0.6	1.1	16.0	0.0	0.0	0.5		Diseased
17.0	F	3.6	4.3	12.2	34.5	89.0	29.0	31.3	235.0	47.2	1.6	35.3	1.4	0.4	0.3	11.5	0.0	0.0	0.4		Diseased
29.0	M	8.9	4.9	15.6	41.0	90.0	31.0	36.0	125.0	60.0	45.0	30.0	2.7		4.0	6.0		0.4	0.5		Diseased
39.0	M	3.1	4.2	8.2	33.0	68.0	20.0	29.0	54.0	80.0	60.0	10.0	0.3		4.0	6.0		0.1	0.2		Diseased
51.0	M	2.0	3.9	9.3	30.0	78.0	24.0	26.0	111.0	60.0	45.0	30.0	0.6		4.0	6.0		0.1	0.1		Diseased
51.0	F	1.9	4.3	6.0	31.0	75.0	22.0	28.0	98.0	50.0	37.5	43.0	0.8		3.0	4.0		0.1	0.1		Diseased
6.0	M	3.3	3.3	10.2	29.3	89.1	31.0	34.8	211.0	38.8	1.3	46.5	1.5	0.3	3.7	10.7	0.0	0.1	0.4	5.5	Diseased
6.0	M	2.2	3.0	9.4	27.6	92.6	31.5	34.1	181.0	82.6	1.9	10.3	0.2	0.0	1.3	5.8	0.0	0.0	0.1	4.7	Diseased
6.0	M	2.2	3.0	9.4	27.1	90.6	31.4	34.7	190.0	75.7	1.7	10.8	0.2	0.0	8.1	5.4	0.0	0.8	0.1	3.7	Diseased
6.0	M	1.4	2.8	8.9	25.9	91.2	31.3	34.4	260.0	26.1	0.4	57.0	0.8	0.0	9.2	7.7	0.0	0.1	0.1	0.6	Diseased
6.0	M	1.7	2.8	8.8	24.8	88.6	31.4	35.5	216.0	21.4	0.4	67.6	1.2	0.0	3.5	7.5	0.0	0.1	0.1	0.5	Diseased
6.0	M	3.0	3.1	10.3	29.5	93.9	32.8	34.9	250.0	42.0	1.3	43.7	1.3	0.3	1.0	13.0	0.0	0.0	0.4	7.2	Diseased
6.0	M	1.9	3.4	11.2	31.9	93.8	32.9	35.1	261.0	13.6	1.6	82.6	0.3	0.5	0.5	2.1	0.0	0.0	0.0	0.9	Diseased
6.0	M	1.7	3.5	11.3	32.0	92.0	32.5	35.3	248.0	21.1	0.3	57.3	1.0	0.0	2.9	18.7	0.0	0.1	0.3	1.3	Diseased
6.0	M	4.3	3.6	11.9	33.8	92.3	32.5	35.2	228.0	53.7	2.3	29.0	1.2	0.2	4.0	13.1	0.0	0.2	0.6	2.9	Diseased
19.0	F	108.0	2.7	8.1	25.3	92.7	29.5	31.8	62.0	1.6	1.7	94.3	102.0	0.6	0.0	3.5	0.6	0.0	3.9		Diseased
19.0	F	154.0	3.0	9.0	28.7	96.6	30.3	31.4	55.0	1.1	1.6	88.5	136.2	0.0	0.0	10.4	0.1	0.1	16.0		Diseased
6.0	F	28.0	3.2	8.8	25.4	78.9	27.3	34.6	44.0	0.8	0.2	94.0	26.3	0.6	0.2	4.4	0.2	0.1	1.2	0.6	Diseased
6.0	F	33.6	3.3	8.7	25.9	79.7	26.9	33.6	63.0	1.0	0.4	92.4	31.0	0.4	0.4	5.8	0.1	0.1	2.0	0.0	Diseased
6.0	F	22.9	3.0	8.0	32.5	78.9	26.8	34.0	45.0	1.1	0.3	94.0	21.5	0.4	0.3	4.2	0.1	0.1	1.0		Diseased
39.0	F	5.6	4.6	12.3	37.6	81.4	26.6	3.3	270.0	68.0	5.1	26.0	1.5		2.0	4.0		0.1	0.2		Normal
19.0	M	7.1	5.6	16.3	48.2	86.5	29.2	33.7	261.0	68.0	5.1	25.0	1.8		1.0	6.0		0.1	0.4		Normal
0.1	F	17.1	4.5	17.0	47.7	107.0	38.1	35.6	255.0	50.0	3.7	28.0	4.8		6.0	16.0		1.0	2.7		Normal
0.8	F	24.3	4.6	12.9	37.1	81.3	28.3	34.8	302.0	71.0	5.3	16.0	3.9			13.0			3.2		Normal
42.0	M	12.7	5.2	16.2	46.3	89.6	31.4	35.0	205.0	76.0	5.7	12.0	1.6		4.0	8.0		0.5	1.0		Normal



23.0	F	5.8	4.1	12.0	36.4	88.6	29.2	33.0	376.0	59.0	4.4	32.0	1.8		3.0	6.0		0.2	0.3		Normal
30.0	F	9.2	3.9	9.3	29.5	76.7	24.1	31.4	214.0	68.0	5.1	23.0	2.1		2.0	7.0		0.2	0.6		Normal
30.0	F	7.7	3.9	10.7	33.3	84.5	27.3	32.3	245.0	61.0	4.5	27.0	2.1		4.0	8.0		0.3	0.6		Normal
30.0	F	9.1	4.2	10.5	33.3	79.0	24.9	31.6	219.0	75.0	5.6	20.0	1.8		2.0	3.0		0.2	0.3		Normal
30.0	F	7.3	3.9	9.0	28.4	73.8	23.5	31.8	310.0	63.0	4.7	28.0	2.0		2.0	7.0		0.2	0.5		Normal
30.0	F	7.3	4.6	13.5	39.5	85.7	29.4	34.3	321.0	49.0	3.6	45.0	3.3		2.0	4.0		0.1	0.3		Normal
31.0	F	7.4	4.3	12.2	35.7	83.7	28.6	34.2	193.0	72.0	5.4	21.0	1.5		2.0	5.0		0.1	0.4		Normal
48.0	M	9.2	5.0	14.0	41.8	84.4	28.2	33.4	229.0	42.0	3.2	40.0	3.6		12.0	6.0		1.1	0.6		Normal
60.0	F	7.7	4.0	11.4	35.3	88.6	28.7	32.4	275.0	56.0	4.2	37.0	2.9		3.0	4.0		0.2	0.3		Normal
60.0	F	10.5	4.7	13.6	42.1	89.5	29.0	32.3	229.0	48.0	3.6	45.0	4.7		2.0	5.0		0.2	0.5		Normal
60.0	F	6.3	4.8	13.1	41.6	86.8	27.3	31.4	283.0	38.0	2.9	54.0	3.4		2.0	6.0		0.1	0.4		Normal
60.0	F	7.2	3.8	11.5	35.7	93.9	30.3	32.2	164.0	50.0	3.8	43.0	3.0		2.0	5.0		0.1	0.4		Normal
82.0	F	8.8	4.3	12.7	38.3	89.7	29.8	33.3	219.0	64.0	4.8	27.0	2.4		1.0	8.0		0.1	0.7		Normal
82.0	F	5.5	4.4	13.1	39.5	90.4	30.0	33.2	249.0	58.0	4.3	31.0	1.7		2.0	9.0		0.1	0.5		Normal
82.0	F	7.0	4.5	13.5	40.8	90.5	29.9	33.1	282.0	60.0	4.5	32.0	2.2		2.0	6.0		0.1	0.4		Normal
82.0	F	8.1	4.1	12.7	36.8	89.8	30.9	34.4	294.0	67.0	5.0	23.0	1.9		4.0	6.0		0.3	0.5		Normal
40.0	F	7.5	4.4	14.3	41.1	94.2	32.8	34.8	249.0	59.0	4.4	35.0	2.6		1.0	5.0		0.1	0.4		Normal
15.0	F	7.8	4.2	13.1	37.6	90.4	31.6	35.0	257.0	59.0	4.4	30.0	2.3		5.0	6.0		0.4	0.5		Normal
10.0	M	7.2	4.4	12.6	36.8	83.9	28.7	34.2	242.0	57.0	4.3	33.0	2.4		1.0	9.0		0.1	0.6		Normal
40.0	M	7.9	5.0	16.4	47.0	93.3	32.4	34.8	188.0	57.0	4.3	37.0	2.9		1.0	5.0		0.1	0.4		Normal
31.0	F	7.4	4.3	12.2	35.7	83.7	28.6	34.2	193.0	72.0	5.4	21.0	1.5		2.0	5.0		0.1	0.4		Normal
52.0	F	10.4	4.4	12.9	40.2	90.8	29.1	32.0	273.0	70.0	5.2	22.0	2.3		2.0	6.0		0.2	0.6		Normal
52.0	F	7.4	3.7	11.7	35.4	94.5	31.2	33.0	232.0	55.0	4.1	38.0	2.8		1.0	6.0		0.1	0.4		Normal
52.0	F	6.4	3.7	11.5	34.8	94.1	31.2	33.1	243.0	60.0	4.5	32.0	2.0		3.0	5.0		0.2	0.3		Normal
52.0	F	9.0	4.3	13.1	39.2	91.3	30.6	33.5	251.0	72.0	5.4	22.0	2.0		1.0	5.0		0.1	0.5		Normal
27.0	F	8.4	4.1	13.8	40.0	97.3	33.5	34.4	310.0	68.0	5.1	26.0	2.2		2.0	4.0		0.2	0.3		Normal
39.0	F	10.0	4.7	12.3	37.3	80.1	26.4	33.0	250.0	55.0	4.1	40.0	4.0		1.0	4.0		0.1	0.4		Normal
39.0	F	6.0	4.5	11.8	36.9	81.1	26.0	32.0	212.0	57.0	4.2	35.0	2.1		3.0	5.0		0.2	0.3		Normal
39.0	F	6.4	4.4	11.2	35.6	80.9	25.6	31.6	290.0	45.0	3.3	48.0	3.1		2.0	5.0		0.1	0.3		Normal
39.0	F	6.4	4.6	11.6	35.5	77.3	25.3	32.7	241.0	60.0	4.5	26.0	1.7		4.0	10.0		0.3	0.6		Normal
32.0	F	6.8	3.9	12.2	35.6	92.3	31.6	34.2	174.0	70.0	5.3	25.0	1.7		2.0	3.0		0.1	0.2		Normal
20.0	F	5.6	4.0	11.0	33.3	83.5	27.6	33.1	227.0												Normal
20.0	F	6.6	3.7	10.2	30.9	84.3	27.9	33.1	191.0	88.0	6.6	10.0	6.6			2.0			0.1		Normal
20.0	F	3.5	4.0	11.4	34.0	85.9	28.8	33.5	234.0	50.0	3.8	33.0	1.2		5.0	12.0		0.2	0.4		Normal
25.0	F	9.4	3.7	11.1	32.5	87.3	29.7	34.1	203.0	64.0	4.8	27.0	2.5			9.0			0.8		Normal
80.0	M	5.4	5.1	14.5	42.4	83.3	28.5	34.2	282.0	40.0	3.0	50.0	2.7		5.0	5.0		0.3	0.3		Normal
34.0	M	7.0	5.7	17.0	49.9	88.2	30.0	34.1	307.0	62.0	4.7	32.0	2.2		3.0	3.0		0.2	0.2		Normal
27.0	F	14.2	4.2	12.5	35.6	85.2	29.9	35.1	282.0	79.0	5.9	15.0	2.1		3.0	3.0		0.4	0.4		Normal
75.0	F	12.7	4.9	12.0	38.7	79.8	24.7	31.0	193.0	84.0	6.3	11.0	1.4		2.0	3.0		0.3	0.4		Normal
23.0	F	11.9	4.4	12.2	37.0	83.5	27.5	33.0	381.0	63.0	4.8	25.0	3.0		5.0	7.0		0.5	0.8		Normal
56.0	F	8.0	5.0	13.2	40.9	81.2	26.2	32.3	244.0	66.0	5.0	29.0	2.3		2.0	3.0		0.2	0.2		Normal
18.0	M	6.3	5.4	17.0	48.2	88.8	31.3	35.3	288.0	60.0	4.5	27.0	1.7		6.0	7.0		0.4	0.4		Normal
13.0	F	5.8	4.7	12.9	38.5	82.4	27.6	33.5	198.0	62.0	4.7	27.0	1.6		5.0	6.0		0.3	0.3		Normal
27.0	F	13.5	4.2	12.8	37.1	89.4	30.8	34.5	362.0	77.0	5.8	16.0	2.2		3.0	4.0		0.4	0.5		Normal
45.0	F	8.9	3.6	11.1	33.0	91.2	30.7	33.6	294.0	77.0	5.8	18.0	1.6		2.0	3.0		0.2	0.3		Normal
9.0	M	6.2	5.3	15.1	43.8	82.0	28.3	34.5	199.0	64.0	4.8	27.0	1.7		4.0	5.0		0.2	0.3		Normal
50.0	F	8.2	3.9	12.1	35.6	90.4	30.7	34.0	158.0	70.0	5.2	24.0	2.0		3.0	3.0		0.2	0.2		Normal
50.0	F	6.7	6.1	14.9	48.2	78.9	24.5	31.0	223.0	54.0	4.1	38.0	2.3		2.0	6.0		0.1	0.4		Normal
50.0	F	9.6	5.6	13.7	44.3	47.5	25.0	30.8	238.0	58.0	4.4	38.0	1.2		1.0	3.0		0.1	0.3		Normal
50.0	F	3.0	5.0	13.0	40.2	80.8	26.2	32.4	190.0	55.0	4.1	41.0	1.2			4.0			0.1		Normal
50.0	F	7.4	5.8	15.1	46.5	80.4	26.1	32.4	242.0	39.6	3.0	55.6	4.1		1.8	2.9		0.1	0.2		Normal
51.0	F	11.0	4.4	12.6	38.4	87.5	28.7	32.8	268.0	69.0	7.6	26.0	2.8	0.0	2.0	3.0	0.0	0.2	0.3		Normal
24.0	F	6.9	4.0	12.0	36.2	90.3	28.9	33.1	250.0	63.0	4.3	31.0	2.2	0.0	2.0	4.0	0.0	0.1	0.3		Normal
47.0	M	5.0	3.9	14.6	41.0	103.0	36.0	35.0	230.0	55.0	3.9	35.0	1.8		4.0	6.0		0.2	0.3		Normal
50.0	M	10.5	5.0	15.9	44.0	88.0	31.0	35.0	283.0	73.0	5.1	20.0	2.1		3.0	4.0		0.3	0.4		Normal
43.0	M	6.9	5.2	15.4	45.0	86.0	29.0	34.0	292.0	50.0	3.5	40.0	2.8		4.0	6.0		0.3	0.4		Normal
26.0	M	8.1	2.6	11.5	30.0	72.0	29.0	30.0	269.0	64.0	4.5	26.0	2.4		4.0	6.0		0.3	0.5		Normal
19.0	F	8.1	4.9	10.2	34.0	69.0	20.0	30.0	459.0	60.0	4.2	30.0	2.4		4.0	6.0		0.3	0.5		Normal
24.0	F	7.4	4.4	13.2	38.0	85.0	28.0	35.0	268.0	64.0	4.5	30.0	2.2		2.0	4.0		0.1	0.3		Normal
34.0	M	3.6	3.9	13.2	36.0	91.0	33.0	36.0	157.0	50.0	3.5	40.0	1.4		4.0	6.0		0.1	0.2		Normal
35.0	M	11.4	4.7	14.9	43.0	90.0	31.0	34.0	228.0	70.0	4.9	22.0	2.5		3.0	5.0		0.3	0.6		Normal
21.0	M	6.5	5.2	14.3	43.0	81.0	27.0	33.0	182.0	62.0	4.3	28.0	1.8		4.0	6.0		0.3	0.4		Normal
54.0	M	6.5	5.5	15.6	44.0	79.0	28.0	35.0	272.0	50.0	3.5	41.0	2.7		4.0	5.0		0.3	0.3		Normal