UNIVERSITY OFGAZİANTEP
GRADUATE SCHOOL OF
NATURAL & APPLIED SCIENCES

IMPROVEMENT OF CLASSIFICATION PERFORMANCE FOR TRIPLE SERUM SCREENING TEST USING DATA MINING APPROACHES

M.Sc. THESIS
IN
INDUSTRIAL ENGINEERING

BY
MEMET MERHAD AY
DECEMBER 2017

DECEMBER 2017
M.Sc. in Industrial Enginee ring
MEMET MERHAD AY

Improvement of Classification Performance for Triple Serum Screening Test Using Data Mining Approaches

M.Sc. Thesis
In
Industrial Engineering
University of Gaziantep

Supervisor
Assist. Prof. Dr. Alptekin DURMUŞOĞLU

By
Memet Merhad AY
December 2017

REPUBLIC OF TURKEY

UNIVERSITY OF GAZIANTEP

GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES

INDUSTRIAL ENGINEERING DEPARTMENT

Name of thesis: Improvement of Classification Performance for Triple Screening Test Using Data Mining Approaches

Name of student: Memet Merhad AY

Exam Date: 28.12.2017

Approval of the Graduate School of the Natural and Applied Sciences

Prof. Dr. Ahmet Necmeddin YAZICI

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master Science.

Prof. Dr. Serap ULUSAM SEÇKİNER

Head of the Department

This is to certify that we have read this thesis and that in our consensus/~~majority~~ opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master Science.

Assist. Prof. Dr. Alptekin DURMUŞOĞLU

Supervisor

Examining Committee Member:	Signature
Assist. Prof. Dr. Alptekin DURMUŞOĞLU	.....
Assist. Prof. Dr. Yunus EROĞLU	.....
Assist. Prof. Dr. Zeynep Didem Unutmaz DURMUŞOĞLU	.....

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original of this work.

Memet Merhad AY

ABSTRACT

IMPROVEMENT OF CLASSIFICATION PERFORMANCE FOR TRIPLE TEST USING DATA MINING APPROACHES

AY, Memet Merhad
M.Sc. in Industrial Engineering
Supervisor: Assist. Prof. Dr. Alptekin DURMUŞOĞLU
December 2017
122 pages

The triple test is a screening test used to calculate the probability of a pregnant woman having a fetus that has an aneuploidy. AFP (Alpha-Fetoprotein), hCG (Human Chorionic Gonadotropin), and uE3 (Unconjugated Estriol) values of pregnant women are computed and compared with the similar records where the outputs (healthy baby or having a disease) are actually known. Bayes theorem is combined with a prior probability derived from maternal age at expected date of delivery is used to calculate the likelihood ratio of a fetus to have diseases like Down syndrome. Current approaches to the calculation of likelihood are known to produce high bias. In this paper, a data mining analysis has been performed to find the best model that is capable of explaining the likelihood of a fetus to have an aneuploidy. 81 triple test records of actually completed pregnancies have been analyzed. 76 of the 81 singleton pregnancies were detected unaffected and 5 of them associated with Down syndrome. The number of 5 pregnancies were increased to 50 pregnancies with the over-sampling technique SMOTE (Synthetic Minority Over-sampling Technique). The Multilayer Perceptron model provided the least false positive rate (13%) and the best detection rate (94%) among several modeling alternatives with the proposed approach. It has been seen that performance of the triple screening test has been significantly improved when compared to the conventional risk assessment.

Key Words: Triple Screening Test, Data Mining Models, Down Syndrome

ÖZET

VERİ MADENCİLİĞİ KULLANARAK ÜÇLÜ TESTİN SINIFLANDIRMA PERFORMANSININ İYİLEŞTİRİLMESİ

AY, Memet Merhad
Yüksek Lisans Tezi, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Alptekin DURMUŞOĞLU
Aralık 2017
122 sayfa

Üçlü tarama testleri gebelik döneminde fetüste herhangi bir kromozomal bozukluk olup olmadığının olasılığını hesaplamak için kullanılmaktadır. AFP (Alfa-fetoprotein), hCG (Human Koryonik Gonadotropin) ve uE3 (Serbest Estriol) değerine sahip gebelerin verileri, gebelik bitimindeki sağlıklı ya da kromozomal bozukluğa sahip doğum yapmış gebelerin verilerine göre değerlendirildi ve kıyaslandı. Fetüsün Down sendromu gibi kromozomal bozukluklara sahip olma ihtimalini hesaplamak için Bayes teoremi, gebenin beklenen doğum anındaki yaşından oluşturulmuş olasılık ile birleştirilerek kullanılır. Günümüzde kullanılan olasılık hesaplama yaklaşımları yüksek derecede peşin hükümlü yani belli bir yöne eğilimliler. Bu çalışmada kromozomal bozukluklara sahip fetüsleri daha doğru olasılık oranları ile tespit edebilecek model araştırıldı. 81 tekil gebeliğe sahip gebelerin üçlü tarama testi verileri analiz edildi. 81 tekil gebeliğin 76'sında kromozal bir bozukluğa rastlanmamış olup 5 tanesinde ise Down sendromu tespit edilmiştir. Down sendromlu bebeğe sahip olan 5 gebe, örneklem tekniği olan SMOTE (Synthetic Minority Over Sampling) ile 50 gebeye arttırıldı. Birçok modelleme denemeleri arasında Multilayer Perceptron modeli en az yanlış pozitif oranına (%13) ve en iyi tespit oranına (%94) sahip oldu. Geleneksel risk değerlendirmeleri kıyaslandığında, üçlü test performansının önemli derecede geliştirildiği görülmektedir.

Anahtar Kelimeler: Üçlü Tarama Testi, Veri Madenciliği, Down Sendromu

To My Family…

ACKNOWLEDGEMENTS

On the accomplishment of this thesis, I would like to express my gratitude to all those people who provided me help and encouragement to complete my research. I would like to thank to my supervisor who is the first and foremost person. I am indebted to my respectable supervisor **Assist. Prof. Dr. Alptekin DURMUŞOĞLU** who does not withhold his paramount patient and support evaluation of outcomes of my thesis subjected as "**Improvement of Classification Performance for Triple Screening Test Using Data Mining Approaches"**.

I would like to thank all staff in Sahinbey Training and Research Hospital for their supports and help during my study. Special thanks to the head of the department of obstetrics & gynaecology **Assoc. Prof. Dr. Mete Gürol UĞUR,** head of the biochemistry laboratory **Prof. Dr. Mehmet TARAKÇIOĞLU,** and the staff of the medical and molecular genetics laboratory, **Fatma TEYMUR,** for their supports.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. I am forever indebted to my parents for giving me the opportunities and experiences that have made me who I am. They selflessly encouraged me to explore new directions in life and seek my own destiny. This journey would not have been possible if not for them, and I dedicate this milestone to them.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

AFP	Alpha-Fetoprotein
AIC	Akaike Information Criterion
BDeu	Bayesian Dirichlet Equivalence
BPD	Biparietal Diameter
CRL	Crown Rump Length
CSV	Comma Separated Value
CVS	Chorionic Villus Sampling
DAGs	Directed Acyclic Graphs
DMQL	Data Mining Query Language
DNA	Deoxyribonucleic Acid
DR	Detection Rate
FASTER	First and Second Trimester Evaluation of Risk
FISH	Fluorescence in Situ Hybridization
FLDA	Fisher Linear Discriminant Analysis
FP	False Positives
FPR	False Positive Rate
GUI	Graphical User interface
hCG	Human Chorionic Gonadotropin
ID3	Iterative Dichotomiser 3
IPS	Integrated Prenatal Screen
k-NN	K Nearest Neighbors
LL	Log Likelihood
LR	Likelihood Ratio
MCC	Matthews Correlation Coefficient
MDL	Minimum Description Length
MLE	Maximum Likelihood Estimation
MLP	Multilayer Perceptron
MoM	Multiples of the Median

NT	Nuchal Translucency
OAPR	Odds of being affected given a positive result
ONTD	Open Neural Tube Defects
PAPP-A	Pregnancy Associated Plasma Protein A
PRC	Precision-Recall Curve
PUK	Pearson VII Universal Function Kernel
QP	Quadratic Programming
RBF	Radial Basis Function
RCOG	Royal College of Obstetrics and Gynecology
ROC	Receiver Operating Characteristic
SLP	Single Layer Perceptron
SVM	Support Vector Machine
SMO	Sequential Minimal Optimization
SMOTE	Synthetic Minority Over-sampling Technique
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SURUSS	Serum Urine and Ultrasound Screening Study
TN	True Negatives
uE3	Unconjugated Estriol
UK	United Kingdom
XOR	Exclusive Or

CHAPTER 1

INTRODUCTION

The number of data, which is stored in the digital environment, increases rapidly with the developing technology nowadays. Analysis of data and transformation of data into knowledge is an indispensable activity for the future. Data mining can be explained as extracting information from a large amount of data [1]. Data mining aims to forecast the future and acquire useful information by making it easy for people to understand difficult data stacks through computer programs. This process can be summarized as shown in Figure 1.1. It primarily begins with the identifying the goal and selecting the target data that is specific data set. After that, cleaning and integrating data are the next steps. Thirdly, the data are converted into the appropriate format for mining by performing data summaries or aggregations. Selecting data, cleaning and integrating data and steps of the transformation data are parts of preprocessing phase. After the data preprocessing is finished, next step is mining data. There are different techniques such as classification, clustering, and regression for extracting data patterns. Data patterns may represent hidden knowledge. As a consequence, useful and future-oriented knowledge is obtained.

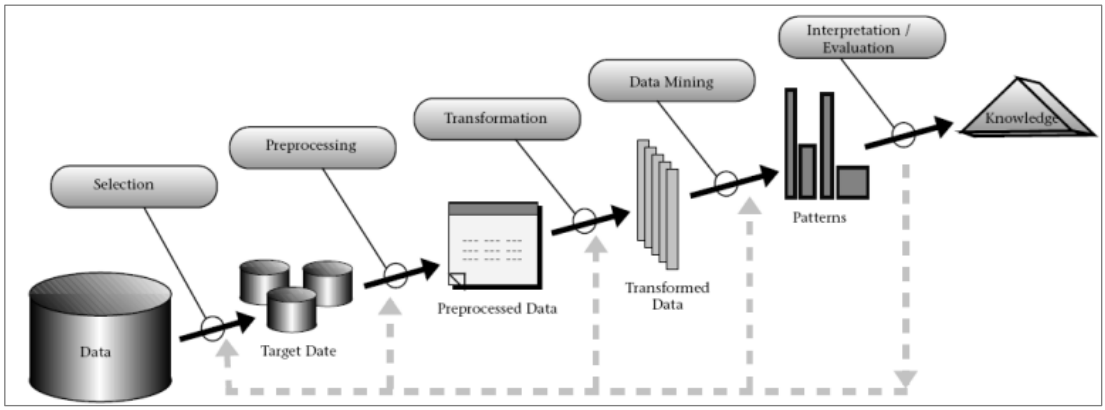


Figure 1.1 Knowledge Discovery Process [2].

1.1 Data Mining Architecture

Data mining consists of various parts [3]. Figure 1.2 shows the architecture of data mining.

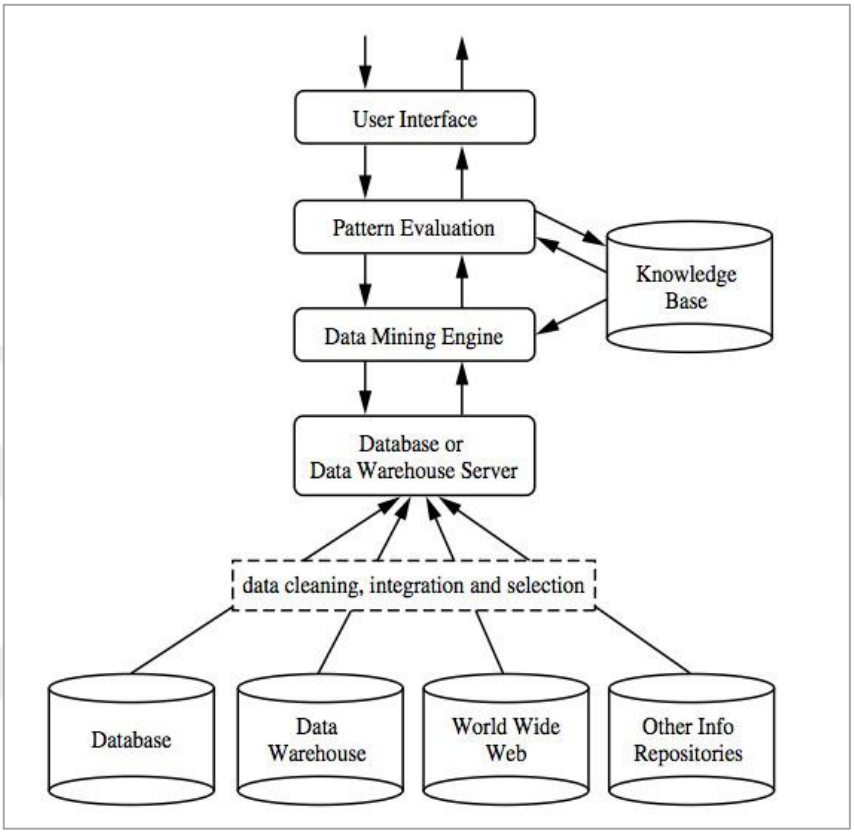


Figure 1.2 Data Mining Architecture [3]

Some data sources such as the data warehouse, the internet, and other sources can cover the data sets after the data cleaning, integration, and selection processes. The database is accountable to get related data, place on user's request for data mining. Knowledge base is information that is used to guide research or to calculate the relevance of emerging forms. Data mining engine is required for data mining structure and preferably contains a set of practical units for tasks for instance characterization, association and classification based on mining principles for data access, prediction, cluster analysis and evolutionary query languages (DMQL). The model calculation module collaborates with the data mining parts for yielding the research to determine attracting models [2]. A communication is provided between the users and the data mining system. It permits client for collaborating with the

approach by highlighting a data mining query or task. It permits user for scanning the data store or the data architecture, envision a models in the distinct shapes, and comment mined models [2].

1.2 Data Mining Techniques

Data mining methods are generally divided into three types [2]. Figure 1.3 shows the basic data mining techniques:

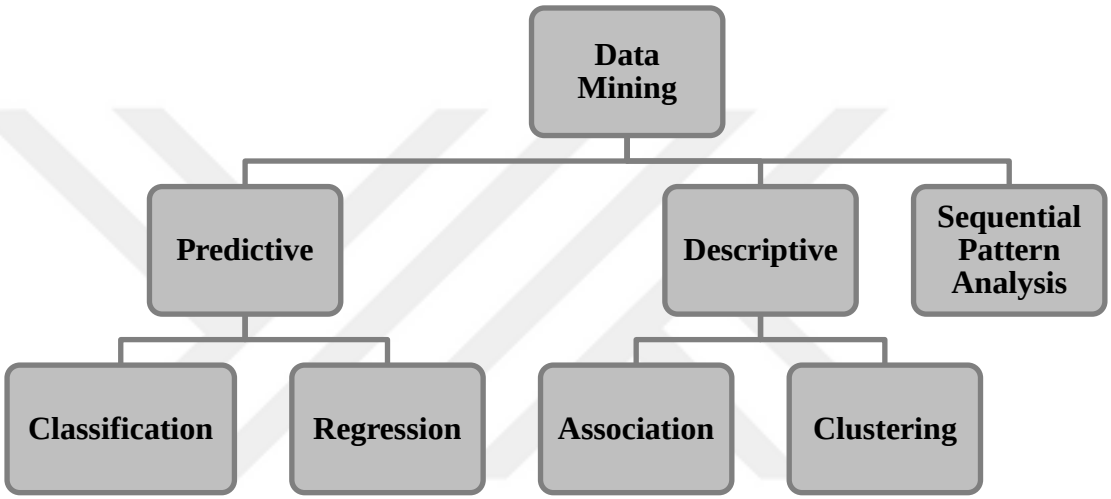


Figure 1.3 Some basic data mining techniques

1.2.1 Predictive techniques

It uses the values of several variables to estimate or procure the future values of the other variables. For instance, the average temperature of the climate can be used to estimate the temperature of the day.

Classification: Fundamentally, each item use classification to categorize them in a set of the data into one of an identified previous set of classes, and sets. Mathematical techniques are used by the classification method like statistics, linear programming, decision trees, and neural network. An example of the classification is shown in Figure 1.4.

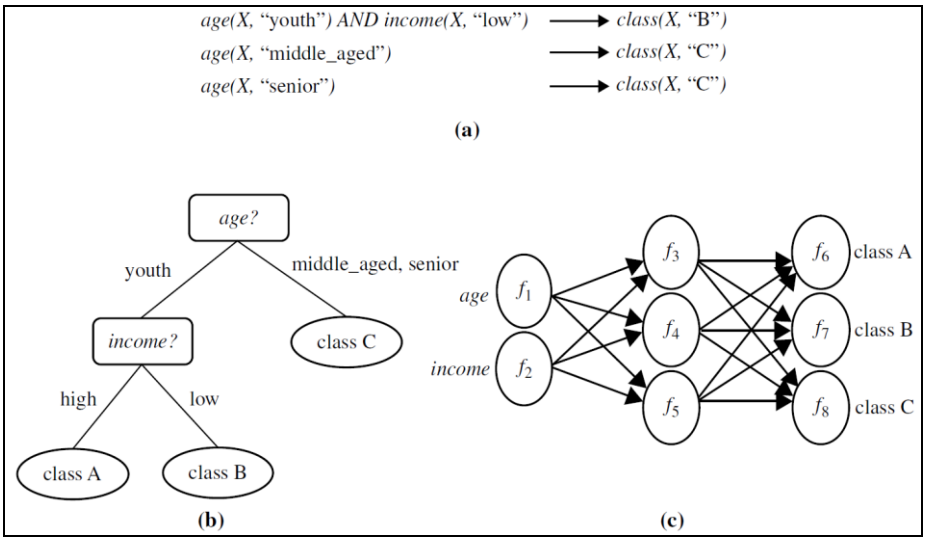


Figure 1.4 A classification models (a) If-Then rules, (b) a decision tree, (c) a neural network [4].

Regression: This analysis is the most frequently used statistical methodology for numerical estimation, though other techniques exist as well. Besides, regression includes the identification of distribution trends that are being founded on the existing data [4].

1.2.2 Descriptive techniques

It describes the typical features of the data in the database.

Association: This method is commonly used in data mining. In this connection, the shape is explored based on an association among objects in the similar operation. That is the cause why the relationship method is also known as relation method. The relationship method is applied in the market basket investigation for recognizing a set of goods that customers typically buy mutually. Suppliers are using association method to study customer's buying behaviors. Based on old sale data, sellers can find that customers continuously buy crisps when customers buy beers, and so customers may place beers and crisps together to save time for the client and rise trades.

Clustering: It creates a meaningful and practical set of items with similar properties using an automated technique. The classes and places objects in every class are described by clustering technique, clustering method assigns objects in default classes when in classification methods.

1.2.3 Sequential Pattern Analysis

Sequential pattern analysis searches to find out and characterize same forms, standard procedures and trends in transactional data during a business cycle. Companies can recognize a set of products, which clients purchase together dissimilar periods in a year, with old transaction data in selling. Then companies may apply this info to advise to clients purchase it with improved contracts based on customers' buying rate in the previous time.

1.3 Data Mining in Medicine

Nowadays, data mining is used in many sectors such as marketing, retail, finance, banking, transportation, medicine, governments, scientific analysis, insurance, etc. The role of data mining is also important in medicine. Figure 1.5 shows the relationship between medicine and data mining. In early diagnostics, correlating certain symptoms to each other increases the accuracy of the diagnosis. Most of the data in the medicine are laboratory results. Therefore, the diagnosis of many diseases depends on the understanding of the values in the laboratory results. Detection of disease with blood test outputs and status of fetus health during the pregnancy can be identified via data mining. During the pregnancy, the status of a baby who has Down syndrome or not is estimated by screening tests which are based on data mining techniques.

Figure 1.5 Relationship between medicine and data mining

Down syndrome is a serious genetic disorder affecting the quality of life of a person. This chromosomal abnormality is seen on average every thousand births [5]. Some Down syndrome pregnancies are terminated with miscarriage. The births with Down syndrome bring dangerous health problems. For example, persons who have Down syndrome cannot perform their brain activities and they will have very serious health problems such as Alzheimer in 40s years old [6]. Therefore, it is tried to determine whether the baby has any genetic disorder before birth. These trials put forward a probability or sometimes a definite result with various data mining techniques and tests. Tests that give definite results are used in the latest stage since they may create danger for mother and fetus. The mentioned methods are referred to as "*invasive methods*" in the medical literature. However, non-invasive methods performed do not constitute any health threat to the fetus and the mother during pregnancy. Moreover, in terms of cost, invasive tests are more costly than non-invasive tests. Although invasive tests require surgical intervention and more medical equipment, non-invasive tests just need blood serum sample. Therefore, non-invasive methods are the most common techniques for using to predict Down syndrome. Also, the most common non-invasive method is a triple screening test in Turkey. The triple screening test was based on some data analysis. It measures three markers that are called alpha fetoprotein (AFP), human-chorionic-gonadotropin (hCG), and unconjugated-estriol (uE3) in a maternal serum sample. The main goal of this thesis is improving the accuracy of non-invasive triple screening tests to estimate chromosomal abnormalities in a cheaper and less dangerous way.

1.4 Problem Statement

Today, states and persons seriously contemplate for healthcare costs. Besides healthcare costs, human life and quality of life are becoming more important day after day. As a result, many studies are carried out in the medical field. One of these studies is the triple screening test which has estimated the risk of a baby birth with Down syndrome. It has been developed for reducing needs of invasive methods that are caused deaths and miscarriages. This test also gives parents a choice to the quality of their life by themselves, because test estimates whether the baby has chromosomal abnormalities before the birth. Triple screening tests were originally developed by scientist N. J. Wald [7]. In this test, it is estimated if a baby has or has not chromosomal abnormalities (Down syndrome) by using the variables like

maternal age, AFP, hCG, uE3, ethnicity, gestational age and smoking status, etc. When analyzing maternal information, historical records of the previous pregnants are used. Multiples of the median (MoM) term is used to provide a specific standard when using past records. MoM is used to standardize blood serum sample variables (AFP, hCG, and uE3) that are affected by ethnicity, geographic location, smoking etc. As a result, the triple test is an economical and non-invasive test. Therefore, it has been a common and widely accepted test throughout the world.

The working process of the triple screening test consists of three steps. First of all, a set of sample is prepared. This sample set obtained from women who have given birth and obtained maternal ages, maternal serum samples, weights, etc. At the second stage, the maternal serum sample, age, the weight of current pregnancy of a woman is examined. At the last stage, the probability of having Down syndrome in the fetus is calculated as a percentage by correlating statistically of the sample set with data of newly pregnancy. However, the detection rate of the triple screening test is reported as 60% [7]. This detection rate may not be at the desired level when the subject is mother and fetus. The test has a false positive rate (FPR) of 5%. This means that the fetus in the pregnancy that is healthy can be misdiagnosed as 5%. In this regard, the aim of this thesis is to improve the test by increasing the detection rate. Thereby the invasive test preference will reduce. As a consequence, the abortion rate is expected to decrease.

1.5 Contributions

The health status of the baby seriously affects both the life quality of the family and the life quality of the baby. The birth of infants with Down syndrome is also one of these factors. Today, amniocentesis application which has a definite result and triple screening test which has average estimation in the diagnosis of this chromosomal abnormality are widely used. Amniocentesis is an invasive test, which is preferred if there is a high probability of a problematic fetus indicated by the triple screening test. This is endangering the health of mother and baby. Improving the performance of the triple screening test for minimizing the necessity of invasive tests also is aimed in the thesis. If better detection rate is achieved by keeping the cost of the test constant, the genetic disorder that is affecting daily life can be predicted earlier with more precise estimations and the preventions can be increased. The confidence of

pregnant women may increase when the output of the test will be a higher likelihood of success and they will not be mentally tired of thinking about the Down syndrome until the end of pregnancy.

1.6 Organization of the Thesis

This thesis is organized as follows. The current algorithm behind a triple screening test is explained in chapter 2. This is followed by the review of studies related to triple screening test, Down syndrome, invasive and non-invasive test and data mining in medicine. Postnatal and prenatal studies to detect Down syndrome have also been examined.

In chapter 3, triple screening test performance is investigated in details. This section describes the subprocesses of analysis: preprocessing, feature extraction, and classification.

In chapter 4 concludes the thesis. Analyzes and results to improve the performance of the test are shown in this chapter.

In chapter 5, conclusion, discussion and future work are given. Performances are compared with other studies in the literature. It includes contributions besides potential future works.

CHAPTER 2

RELATED WORKS

Genetic anomalies occur approximately 10% - 20% of live births [8]. General chromosomal abnormality, which is trisomy 21, usually seems in live born infants [9]. Structural and nonstructural variables are included by sonographic findings in fetuses with Down syndrome. Various techniques which are including maternal age [10], biochemical markers [11], prenatal ultrasound and amniocentesis [12] have been used to describe females at risk of carrying a fetus with aneuploidy. On the other hand, 0.5% - 1% of the fetal mortality occurs in the invasive techniques [13]. Failure to treat Down syndrome has led to the development of prenatal screening tests. These screening tests aim to detect the anomalous fetus as early as possible in the pregnancy. A triple screening test is one of these tests. Estimating Down syndrome with prenatal tests was found after long-term studies. Before talking about prenatal tests, Down syndrome will be defined. Also, it is called trisomy 21. In the following section, the studies about methods of determining Down syndrome in the gestation period are discussed. These methods are divided into two categories. They are invasive methods and non-invasive methods. The development of non-invasive screening tests used during the pregnancy is closely related to the thesis. Studies on these will also be examined in this chapter. Then the historical development of the triple screening tests, which are widely used today, will be examined. Finally, the relationship between data mining and triple screening tests will be presented in this chapter.

2.1 Down Syndrome

There is a nucleus in each cell in a human body where the chromosomal substance is collected in the genes. The genes have codes which are liable for the whole of our heritage features. Also, they are collected throughout some structures that are named chromosomes. Classically, each cell has 23 pairs of chromosomes in its core and half

of them are inherited from mother and father [14]. Figure 2.1 represents the chromosome and gene schematically. If a person has the complete or fractional additional copy of chromosome 21, that person has Down syndrome.

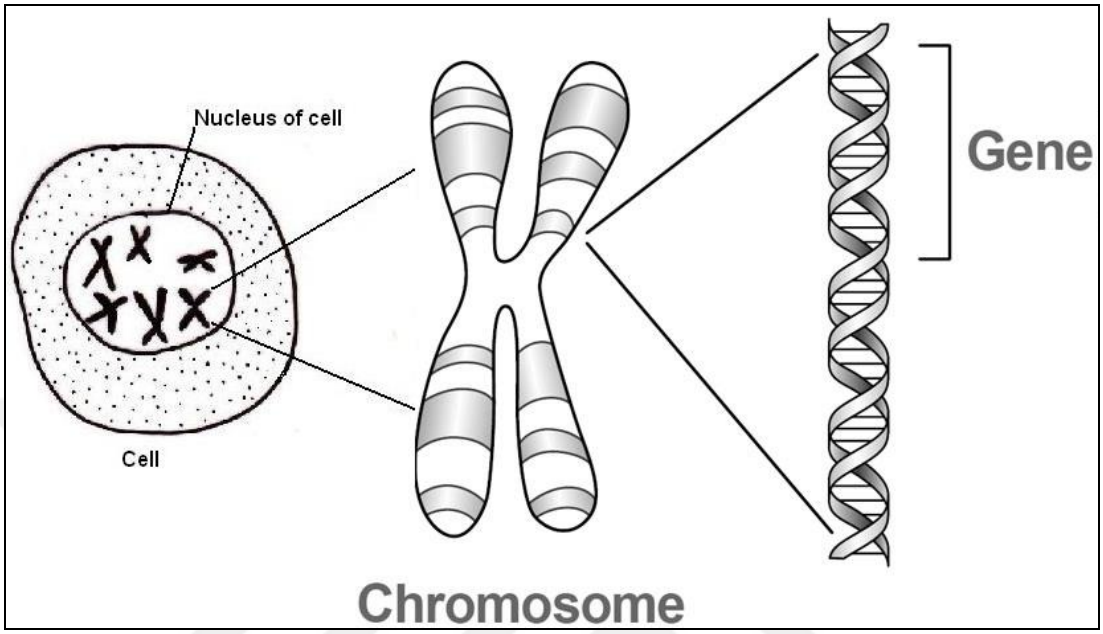


Figure 2.1 A typical cell, nucleus, chromosome and gene

This additional chromosomal substance switches the process of growth and reasons characteristics related to Down syndrome. The upward slant to the eyes, small stature, a single deep wrinkle across the center of the palm, and a low muscle tone are occurred in Down syndrome individuals [15]. Although people with Down syndrome are physically similar to each other, each of them has a unique personality.

2.2 History of Down Syndrome

Down syndrome was defined by John Langdon Down in 1866 [16]. Trisomy 21 or Down syndrome [17] is created by a defect of the 21st chromosome until the egg or sperm is developed [18]. Consequently, a sperm or egg cell is created with an additional copy of chromosome 21 and the cell has 24 chromosomes. In recent history, researchers have found an opportunity to study about the features of a person with Down syndrome due to improvements in medicine and science. Scientist Jérôme Lejeune described Down syndrome as a chromosomal disorder in 1959 [19]. Jérôme detected 47 chromosomes in the person's cells with Down syndrome instead of a typical 46 chromosomes exhibit. Later, he decided that the

additional fractional or full copy of the chromosome 21 outputs in features to with Down syndrome. After that, 329 genes on chromosome 21 was effectively defined and classified by some group of researchers until 2000. As a result, these successful studies and the results helped to be large improvements in Down syndrome research [20].

2.3 Trisomy 21 (Nondisjunction)

Nondisjunction is generally brought some problems in cell divisions [21]. Trisomy 21 occurs with three copies of chromosome 21 in an embryo that is in a habitation of the usual 2. At perception, a couple of chromosomes 21 in both a sperm and an egg crashes for separating. An extra chromosome is duplicated in every cell of a human body as the embryo improves. General cell division is shown in Figure 2.2. Also, Trisomy21 cell division is shown in Figure 2.3.

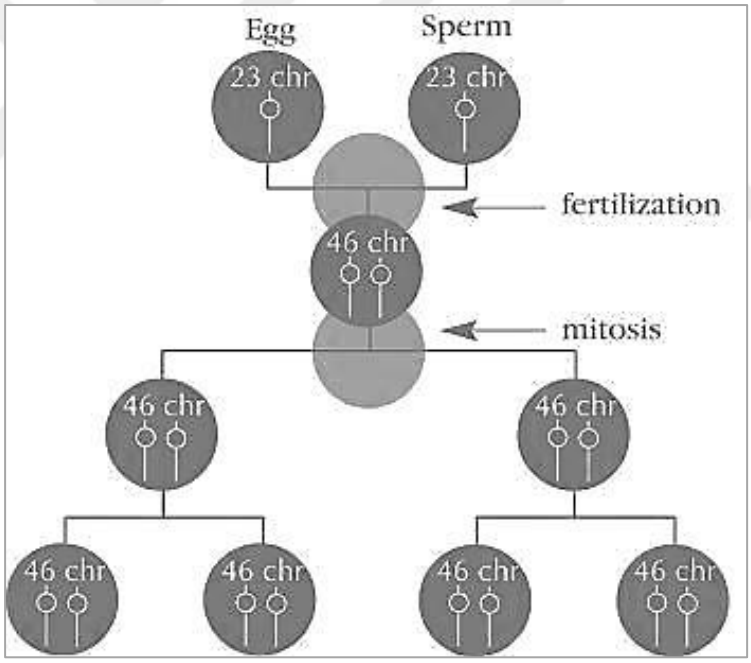


Figure 2.2 Typical Cell Division [14]

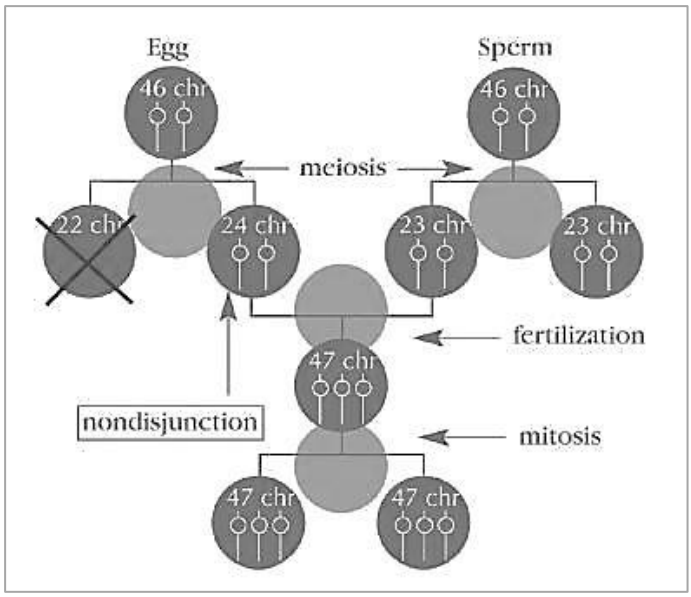


Figure 2.3 Nondisjunction Cell Division [14]

2.4 Causes of Down Syndrome

Whatever the type of Down's syndrome is, everybody with Down syndrome has the additional, serious part of the chromosome 21 present in several of his or her cells. The extra chromosomal substance changes the route of improvement and reasons personalities related to Down syndrome. The additional full or fractional chromosome is yet unidentified. Maternal age is first reason, which has increased the likelihood of nondisjunction or mosaicism of the baby with Down's syndrome [22]. Nevertheless, 80% of infants with Down syndrome who has a mother younger than 35 years are born due to higher birth rates in younger women. There isn't a whole and complete scientific study, which signifies that environmental factors or the parents' activities affect Down syndrome before or during pregnancy. The added part or full copy chromosome 21 that reasons Down syndrome may create from both father and mother. Nearly 5% of the instances have been followed up to father [14].

2.5 Relationship between Maternal Age and Down Syndrome

Down syndrome may happen in persons of all nationalities and at different economic stages. However, mature females are more likely to have a child with Down syndrome. If the female is 35 years old and then she has about a 1:350 probability of considering a baby with Down syndrome, and the probability rises regularly to the

1:100 by forty year age. When age comes to 45, the frequency turns into nearly 1:30. Maternal age doesn't appear to be connected to a substitution risk [23]. Figure 2.4 and Table 2.1 shows the relationship between maternal age and Down syndrome.

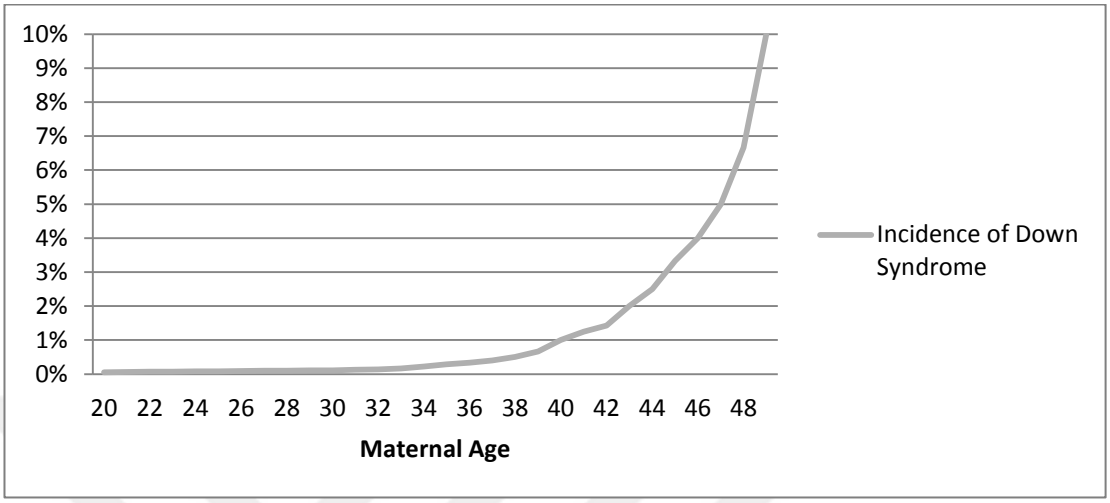


Figure 2.4 Line chart of relation maternal age and Down syndrome

Several partners delay giving birth to a baby later in their life, a frequency of the Down syndrome concepts is estimated to rise. Consequently, the genetic consultation is happening progressively significant for partners. Nevertheless, most physicians do not fully inform partners about an incidence of the Down syndrome, the progress of identification, and cure of infants for Down's baby.

Table 2.1 Relations between maternal age and Down syndrome [14]

Maternal Age	Incidence of Down syndrome	Maternal Age	Incidence of Down syndrome	Maternal Age	Incidence of Down syndrome
20	1:2000	30	1:900	40	1:100
21	1:1700	31	1:800	41	1:80
22	1:1500	32	1:720	42	1:70
23	1:1400	33	1:600	43	1:50
24	1:1300	34	1:450	44	1:40
25	1:1200	35	1:350	45	1:30
26	1:1100	36	1:300	46	1:25
27	1:1050	37	1:250	47	1:20
28	1:1000	38	1:200	48	1:15
29	1:950	39	1:150	49	1:10

2.6 Diagnosis of Down Syndrome

Down syndrome may be detected in the two methods that are prenatally or postnatally (after birth). The postnatally, Trisomy 21 is generally defined by the existence of particular bodily characteristics: an upward slope to eyes, a solo deep crease across the palm of the hand, the mildly pressed facial profile and the weak muscle. Since the characteristics can be given in the infants without Down syndrome, the chromosomal examination named the karyotype is completed to verify identification. Clinicians take a blood sample to analyze the cells of the baby to get a karyotype. They take pictures of the chromosomes and categorize chromosomes by dimension, quantity, and figure. Figure 2.5 shows karyotype of the woman with Down syndrome. Clinicians can diagnose Trisomy 21 after the analyzing the karyotype. Another genetic test is called FISH (fluorescence in situ hybridization) can implement same procedures and check a diagnosis in a smaller sum of time [14].

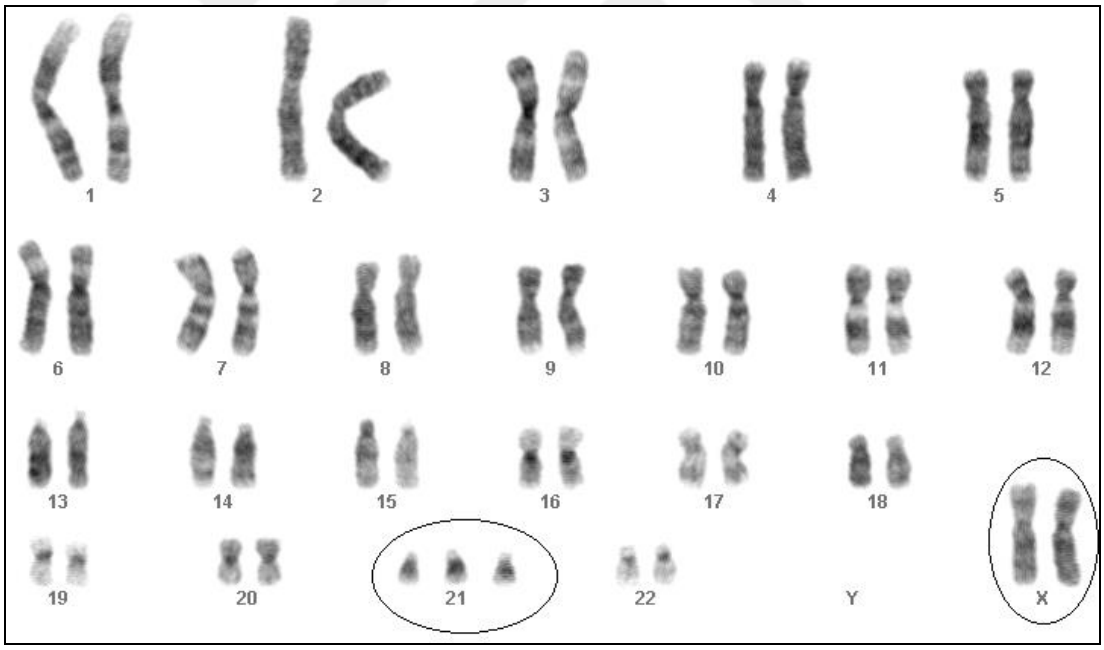


Figure 2.5 Karyotype of a female with trisomy 21 [14]

Before birth, there are 2 special tests which may be made for finding Down syndrome along the prenatal period [24]. They are a diagnostic test and the screening test. All female is under the risk of getting a baby with Down syndrome. In Turkey, the screening test is recommended for every female to look at the risk of baby's Down syndrome during early pregnancy but there is a critical aspect for

understanding which the screening test doesn't provide an exact answer to whether the infant has Down syndrome. The diagnostic test will generally be recommended if the screening test indicates that the infant has a high risk of Down syndrome. A diagnostic prenatally tests are used to see a growing infant really does have Down syndrome. The main accessible prenatal diagnostic tests are two: chorionic villus sampling (CVS), and amniocentesis. Besides, if the pregnant make a diagnostic test, there may be risks of complications. It contains miscarriages, injuries, and infections. This is the reason why the diagnostic test is not recommended for all pregnancies [24].

2.6.1 Diagnostic tests for Down Syndrome

There are 2 important tests which are amniocentesis and chorionic villus sampling to diagnose trisomy 21 prenatally.

2.6.1.1 Amniocentesis

Amniocentesis is a method that draws amniotic liquid that is coming from an uterine cavity with a needle, by a transabdominal method and below nonstop ultrasound support, so as to get a specimen of fetal exfoliated cells, transudates, urine or secretions [25]. A graph of amniocentesis method is shown Figure 2.6. The fluid sample is taken from the uterus by doctors and then they do several microbial, biochemicals, chromosomal, and molecular researches becoming acted on the amniotic liquid specimen [26–27]. A sample of the amniotic fluid in pregnant woman's uterus, which is an immersive growing baby, is taken by a good injector during amniocentesis. The amniotic fluid sample contains specific cells which are coming from an infant's derma, as well as unwanted materials such as urine, etc. Cells in the amniotic liquid include the infant's chromosomal substance. The test is mostly recommended after the 15^{th} week of pregnancy. This is because it has been shown to be safe during this period of pregnancy. Amniocentesis has a risk of problems. After amniocentesis test, all risk of a woman who has a miscarriage is approximately 1% [14]. So, approximately 1 in every 100 females with an amniocentesis may have an abortion. There is also another risk that is less than 1:1000 which test will end in serious infection. However, these risks cannot be ignored when the subject comes to the human health. But the reasons of miscarriages after the rest are still unknown certainly. Research has shown that

when a miscarriage occurs, there is a general reason from the amniotic sac is damaged or infected. It is also difficult to tell when a pregnant woman is most probable to miscarry after she has had an amniocentesis. Also, many of the miscarriages occur in two weeks after amniocentesis [14].

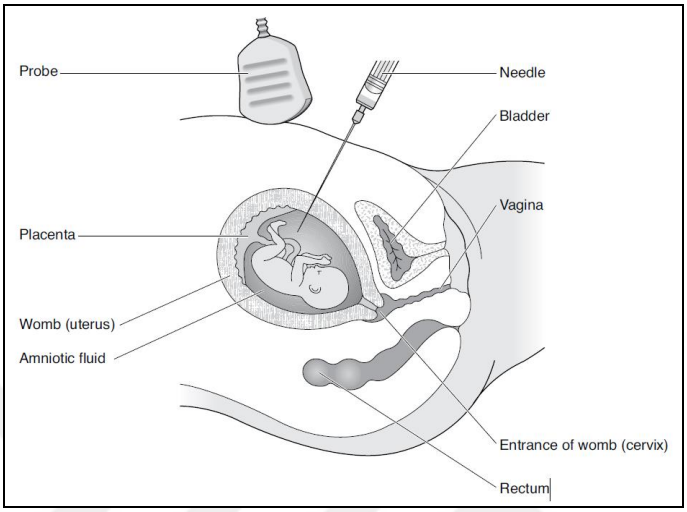


Figure 2.6 Amniocentesis

2.6.1.1.1 Miscarriage on Amniocentesis

Every each pregnancy has a risk of miscarriage, whether or not have an amniocentesis. Pregnant women with amniocentesis were thought to have an abortion risk of a 1 in 100 at the end of 15 weeks of gestation [28]. If the test is done before 15 weeks of pregnancy, the risk of the miscarriage increases. However, the reason of this miscarriage status is not clearly known. Reasons of the miscarriages may be a bleeding, infection, or harm to the amniotic tissue due to a process. The process associated miscarriages are unusual more than two weeks after amniocentesis [28].

2.6.1.1.2 Infection on Amniocentesis

After an amniocentesis, infection rarely occurs. Less than 1:1000 pregnant individuals that have the amniocentesis will occur the critical infection in woman amniotic liquid. Something may cause infection [29]. For instance, it is used by injury to pregnant woman's intestine with an injector through a process so microbes, which are typically included in the intestine, escape. By microbes, which exist on the derma of her abdomen, through a trail of the injector, circulating in the

abdomen/uterus. The microbes which are seen on ultrasound examine and in the ultrasound lotion can go throughout the trail of the injector into her abdomen or womb. Indicates of the infection may contain pyrexia, tenderness of her womb and spasms of her abdomen [29]. Nevertheless, if the right techniques are used to reduce infection, infection is unlikely.

2.6.1.1.3 Injury to the Developing Baby on Amniocentesis

When performing the amniocentesis, the baby can be damaged with a needle. Nevertheless, nonstop ultrasound control through amniocentesis is decreased a probability of the complication [30].

2.6.1.2 Chorionic Villus Sampling (CVS)

The name of this test comes from a very small tissue sample that is taken from the piece of afterbirth named chorionic villi. During CVS, it is received from the placenta. Chorionic villi cells include the similar chromosomal substance like the developing baby's cells. For this reason, tests can be performed in placental tissue in a laboratory to observe a genetic structure of infant. In other words, chromosome and genetic situations involving Down syndrome may be detected [31]. There are two methods of CVS as shown in Figure 2.7 and Figure 2.8. They are CVS throughout the abdomen and CVS throughout the cervix. 2.6.1.2 Chorionic villus sampling is frequently applied with passing well injector throughout the derma of the pregnant woman abdomen and inside of her womb to get the tissue of the afterbirth and it is called transabdominal CVS. The test is generally performed between an 11^{th} and 14^{th} weeks of the prenatal period. After CVS test, the risk of miscarriage is approximately 1% - 2%. Also, roughly one or two pregnant women per a hundred females with CVS may lose their baby [32].

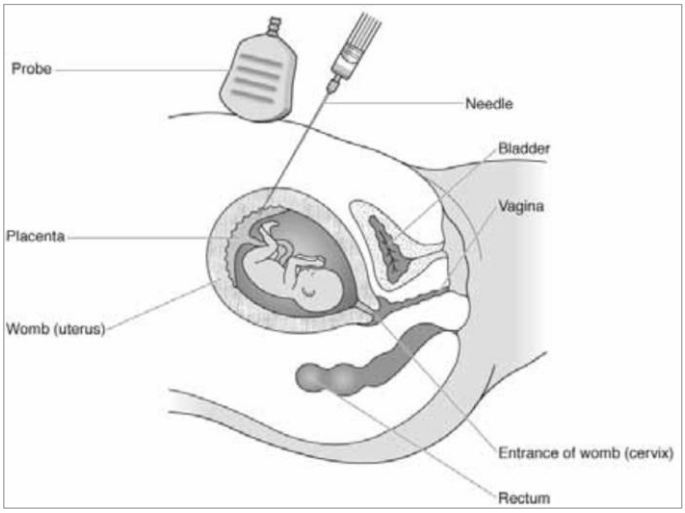


Figure 2.7 Chorionic Villus Sampling -through the abdomen

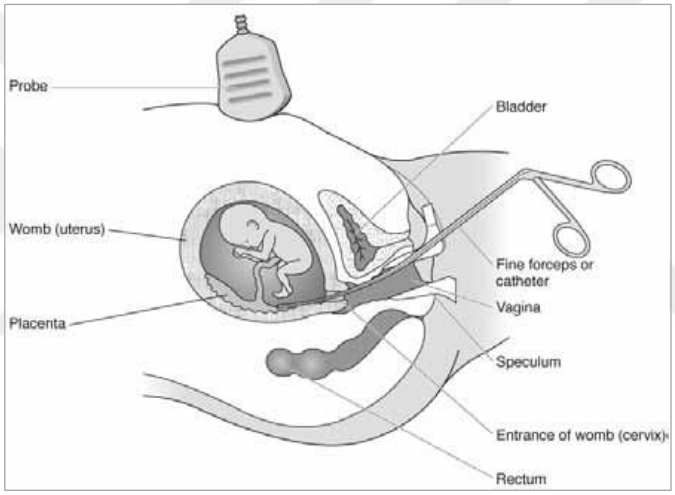


Figure 2.8 Chorionic Villus Sampling - through the cervix (entrance to the womb)

2.6.1.2.1 Miscarriage on CVS

Regardless of whether your mother is CVS, every pregnancy has a risk of miscarriage. However, there is an additional risk for women who perform the CVS test. This additional risk of miscarriage is approximately 3 in 200. The extra risk of miscarriage after the CVS test is more than that after an amniocentesis. The reason for this more risk of a miscarriage of CVS when it is compared with amniocentesis might be as CVS is performed previous periods in pregnancy. Another factor is that Chorionic villus sampling is generally performed due to the suspicious troubles of the growing infant. If the troubles exist, they can increase the risk of baby's life. Many of miscarriages in women who perform the CVS test occur within two weeks.

When passed the three weeks from the performing test, then miscarriage is less likely. Certainly, miscarriages depending on the CVS are not known [33].

2.6.1.2.2 Infection on CVS

After CVS, an infection can be seen. Its percentage is about 0.1% (it means 1 pregnant woman in 1000 pregnant women) who perform CVS can face with a critical infection. There may be many reasons for the infection. For instance, germs that are generally included inside the bowel can mess by a needle when the pregnant woman's bowel is injured with a needle in transabdominal CVS. The microbes, which are currently on the derma of her abdomen, can travel throughout the trail of the injector inside her abdomen. Also, microbes may be located on an ultrasound examine and in the ultrasound lotion, so they can mess with a needle and they may travel throughout the trace of the injector inside the pregnant woman womb or uterus. On the other side, if germs mess with the biopsy needle, the infection may pass through the neck of her womb after transcervical CVS. Symptoms of infection can contain pyrexia, tenderness of her abdomen and spasms of her uterus [34].

2.6.1.2.3 Limb Abnormalities in the Developing Baby on CVS

As a result of the research, five babies were reported to have limb abnormalities such as missing fingers and toes after CVS in the 1990s, researchers' interests increased. However, CVS was performed before the 10^{th} week of pregnancy in all cases. Later studies showed that such problems were a source of concern did not have a higher risk than the general population. Nevertheless, CVS is not recommended prior to completing 10^{th} week of pregnancy. This is because CVS is more difficult to implement at this stage of the pregnancy [33].

2.6.1.2.4 Some Researches about Miscarriage on Diagnostic Test

The landmark publication of Tabor and his friends in 1986 [35], clarifying the outputs of a randomized check trial of abortion next amniocentesis, lead to a suggestion via RCOG (Royal College of Obstetrics and Gynecology) Guidelines which pregnant females should be warned that low risk is 1% after amniocentesis. There haven't been studied researches, which are associated entering the chorionic villus sample (CVS) and women who underwent no process; therefore the risk of CVS has been estimated from researches associating amniocentesis with CVS, led

to the RCOG direction to declare a process associated risk of 1.5% [36–38]**.** In North America, the risk of pregnancy loss based on professional opinion associated with prolonged amniocentesis is 0.5% [39]**.** Nevertheless, many publications from the scientist studies have named into issue these risk schemes and it is usually recognized, which developments in technology and improve in operative ability, are probable significantly to alter the risks [40]. The authors draw attention to the calculation process of associated risk for miscarriages by collation of prenatal period outputs in females who have CVS. Therefore, approximating process associated risk next prenatal specimen in studies is essential to accommodate for pregnancy and maternal features before competing them [39]. The last analysis of proof on "*process associated miscarriages"* prove, which the suggestion is clearly conflicting, however highlighted, which the generality of researches that analyze pregnancy loss proportions in females with the invasive process and they do not, tell no important change among a clusters; analysis accomplishes, which risk of miscarriage is dissimilar for woman, and according to maternal demographical features and constituent parts of its unlikely that the first trimester screening test and invasive example will contribute significantly to the individual patient [41]**.**

2.6.2 Screening Test

More than one study was conducted to improve the non-invasive tests. Firstly, Penrose, scientist, found an association between maternal age and Down syndrome [22]. The results of his study showed that paternal age was not a significant factor, while maternal age was to be regarded as very critical. In 1984, Merkeatz and partners reported a relationship between low maternal serum AFP and Down syndrome pregnancies [42]. Many teams were tried to estimate Down syndrome before birth in 1987. These were the studies done by looking maternal age and the AFP value. Cuckle and N.J. Wald [43], scientists who compiled these studies, found that the likelihood of having the Down syndrome pregnancy was 28% with FPR of 2% to 8%. N. J. Wald's study showed that the detection rate with AFP, hCG, and uE3 was 61% and the false positive rate was 5% in 1988 [7]. After that, N.J. Wald developed new screening test with 4 markers which were AFP, hCG, uE3, and inhibin-A. Performance of this test was 70% DR (detection rate) and 5% FPR [44]. However, this study increased the cost of the test. Lately, many scientists focused on baby DNA which is in maternal blood and recent studies show that DNA of a

baby can find in maternal blood serum and these studies called as cell-free DNA for Down syndrome. Its detection rate is 99% and it is a quite successful test [45]. However, this test is extremely expensive and many pregnant women don't prefer the cell-free DNA tests due to the costs. Therefore, the triple screening test is still popular nowadays.

A screening test, which is a non-invasive test, is a procedure of to measure an inhabitant by a particular indicator or indicators described screening threshold value for recognizing persons in inhabitants at more risk for the specific disorder. A screening test is valid to the inhabitants; identification is used at the personal patient level [46]. An abnormality scan should only be done if the abnormality is considered important sufficient to require interference. The indicators that are utilized in the screening test should be of appropriate sensitivity to determine the maximum of the persons affected by the minus misidentification of unaffected persons. At the same time, there should be a perfect diagnostic test to decide whether the screen-positive individual is indeed abnormal and whether there is an interference with all the people classified as affected. Screen testing, including screening and intervention, should be inexpensive. For this reason, the screening test should be acceptable to everyone [47].

2.6.2.1 Reasons to Perform Prenatal Screening Test

Some patients may decide to terminate their pregnancy if they are exposed to a fatal anomaly. Abnormality recognition can permit specific prenatal care and replace perinatal care. It may be reasonable to prevent cesarean delivery because of a fetal distress in a child with a fatal abnormality if the patient doesn't select to end a gestation. To learn abnormalities earlier is much better for the parents. The families have time to be psychologically and economically prepared. They may train themselves about a family abnormality [48].

There are 2 main techniques of screening tests for Down syndrome; biochemical serum screening, and the ultrasound scan [32]:

Nuchal translucency ultrasound scan (NT scan): It is a specific ultrasound scan which is completed between 11^{th} weeks 2 days and 14^{th} weeks 1 day of the prenatal period. It calculates the liquid group below the derma at a rear of the infant's collum

as shown in Figure 2.9. All infants contain a group of fluids at this place; however infants with Down syndrome are inclined to have extra liquid in this part. Subsequently, other details such as fluid measurement, maternal age, baby size and maternal weight, ethnicity, and smoking status are inserted the computer software to calculate the risk of Down syndrome. It can be occasionally problematic to measure Nuchal Translucency correctly. This may be due to the infant's condition or because the pregnant woman is overweight [32].

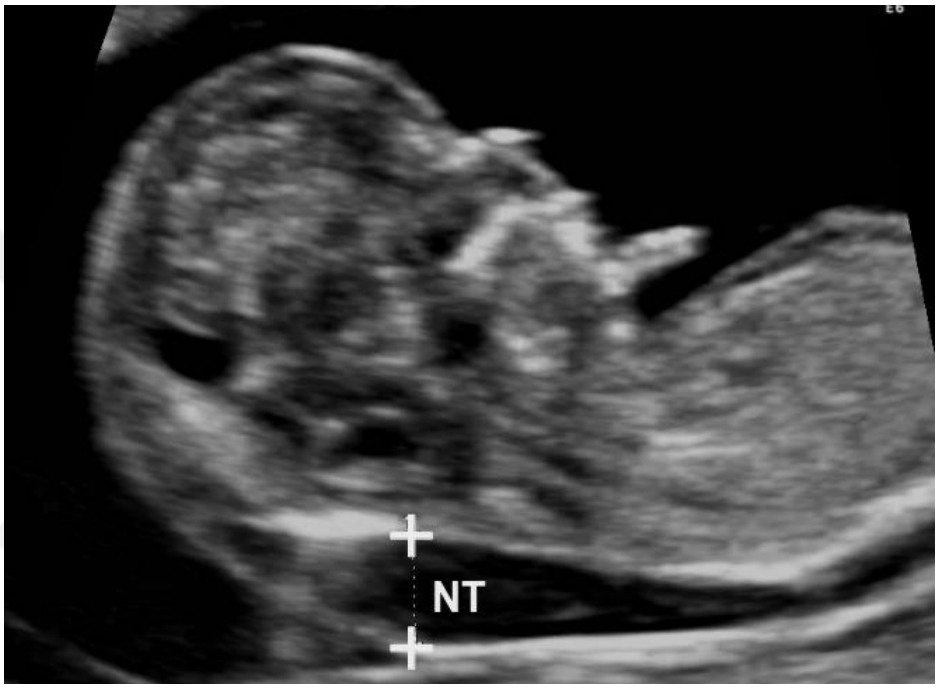


Figure 2.9 Nuchal translucency ultrasound scanning of fetus (NT)

Blood tests: Blood tests are made to evaluate the stages of proteins and hormones in the bloodstream. The proteins and hormones are generated via a placenta or the growing infant. Samples contain unconjugated estriol (uE3), human chorionic gonadotropin and alpha-feta-protein. The quantities of the materials may be influenced if the infant is with Down syndrome. The computer software is worked to produce a risk of the infant who has Down syndrome, according to the blood test outputs, maternal age, and phase which a pregnant woman is in her pregnancy, her weight, ethnicity, and smoking condition [32]**.**

Down syndrome can be estimated in two ways. These are the first trimester and second trimester screening tests [13].

2.6.2.2 First Trimester Screening Tests

In the early 1990s, Nicolaides tried to predict the Down syndrome by measuring nuchal truanslucency at the ultrasound, so that the estimate of Down syndrome was widespread considering the nuchal truanslucency [49–52]. Another study that contains 8514 pregnancies reported a 79% DR at the 5% FPR [53]. It is compounded maternal age, NT ultrasonography, and measurement of maternal serum hCG and PAPP-A in the first-trimester screening [25, 26, 54, 55]. Blood collection for the biochemical examination and ultrasound the evaluation for NT is generally done between 11^{th} and 13^{th} weeks of pregnancy. The raised NT, reduce in prenatal period that is related PAPP-A stages, and a rise in hCG may be a signal of Down syndrome, and help doctors in recognizing pregnancies for Down syndrome risk. NT measurement has the Down syndrome DR approximately 70% with a 5% FPR; however detection rates rise to 79%–90%, with the 5% FPR when mixed with the PAPP-A and hCG [54–57]. Many studies were done to calculate the optimum period to do the first trimester screening, with the aim of giving the highest DR when yet keeping a low FPR [58–61]. The researches advise that previous PAPP-A and hCG evaluations are obtained at 9^{th} and 10^{th} weeks of pregnancy, with the NT scan obtained at 12^{th} weeks of pregnancy, may raise DR to 90%, with a 3% FPR [58, 60, 61]. If the outputs find out a raised fetal abnormalities risk, a mother may be recommended hereditary guidance with a choice to select either first trimester CVS or second trimester amniocentesis [62]. A combined screening test is one of the first screening tests.

2.6.2.2.1 Combined Screening Test

Combining of nuchal translucency scan output, PAPP-A value and hCG value indicate a risk of trisomy 21. The blood tests are taken on the same day that the screening tests are done [58].

2.6.2.3 Second Trimester Prenatal Screening Tests

Second trimester prenatal screening might contain numerous blood tests, named several indicators. The indicators supply info about a mother's risk of an infant with definite hereditary situations or birth abnormalities. In general, a screening test is used to predict the risk by getting a sample of the maternal blood between the 15^{th}-20^{th} weeks of prenatal period. These multiple markers are AFP screening, hCG, uE3

and Inhibin A [63]. In addition, anomalous test outputs of alpha fetoprotein and other indicators mean that it might require more examination. The ultrasound is generally used to verify the times of the prenatal period and for observing at the fetal vertebral column and other body elements for disorders. The amniocentesis can be required for precise identification. When looking to screening test with another perspective, multiple markers screening test is not invasive. This means that it is not completely correct and it gives only probability to determine who may be recommended extra tests for her pregnancy in the population. The FPR results may represent a problem about the infant on the contrary baby is an actually healthy or the normal result is represented with the false negative result when the infant really has a health difficulty [13].

2.6.2.3.1 Triple Screening Test

The second trimester scan is usually based on the triple screening test [64]. Three maternal serum indicators, which are alpha-feto-protein, human-chorionic-gonadotropin and unconjugated-estriol, are calculated and used to adjust the mother's risk according to her age for producing a mother specific Down syndrome risk in this test [65]. So, compensate for the difference of the indicators with of gestational, evaluated concentration is separated by the median indicator stages in the important week of gestational yielding MoM (Multiples of Median). Additionally, MoM is modified to atone for aspects with another Down syndrome those different indicator stages [66]. Generally, marketable software packages trust a similar process for the risk evaluation. An accuracy of the outputs depends upon the specific markers: methodical performing of the immunoassays used [67]; correct dating of prenatal period [68], properly select of medians, which is used to evaluate MoM, also and acceptable elements in MoM setting and on the choice of convenient inhabitants limits [69].

2.6.2.3.2 Quadruple Screening Test

This test is based on the results of a blood test between 14^{th} weeks and 20^{th} weeks of gestation [70]. The test doesn't contain the ultrasound. As previously mentioned, the test presents the risk in the prenatal period of Down's syndrome infant. For instance, the test can present that there is 1:1000 risks of Down's syndrome infant. It indicates that one pregnant woman will have an infant with Down syndrome for

every thousand pregnancies and rest of them will have a baby born without Down syndrome. As a consequence, that is a largely small risk. National Screening Committee, advised a threshold limit to separate between screening test outputs in the United Kingdom with an upper risk that infant is born with Down syndrome and persons has a minor risk. The threshold level is 1:150 and it represents that infant has Down syndrome if screening test outputs present a risk of between 1:2 and 1:150, that is categorized like a maximum risk output. It is classified as a minimum risk output if the output shows the risk 1:151 or higher, then. The second number gets higher, the risk becomes lower [71].

However, all of the tests don't say for definite whether the infant has Down syndrome, so tests just ensure likelihood. Most ladies who have a screening test for trisomy 21 may have a lower risk output. She may assure again by it if this is the case. It doesn't indicate her infant certainly doesn't have Down syndrome. In some cases, a baby does have Down syndrome. If the mother has been given a higher risk output, it doesn't indicate her infant absolutely has Down syndrome. Extra tests are required to approve the diagnosis and provide a certain response if she is given a higher risk output. In cases where infant doesn't have Down syndrome, it is named false positive output [32]. Alternatively, diagnostic tests may supply the absolute diagnosis with 100% precision [14]. Also, the new progressive prenatal screening test can establish chromosomal substance from infant which is rotating in the maternal blood serum sample. It isn't invasive such as the diagnostic tests, however the tests support a high precision degree. Nevertheless, it is much expensive. All screen tests still won't absolutely diagnose Down syndrome. Because of that, the thesis aims to improve accuracy and reduce the deviations with minimum cost or economical price.

2.7 Estimation Down Syndrome with Triple Screening Test

The triple screening test is currently used based on statistical analysis of gestational age, maternal age and ethnicity, together with hCG, AFP, and uE3 values measured during the 15-21 weeks of pregnancy. Values obtained in the triple screening test may be different depending on factors such as race, geographic location, socio-economic level, and prevalence of anomaly [72]. Additionally, the triple screening

tests can never give definite positive or negative results. They can only report a probability. However, the closeness of this probability to the exact findings can be expressed as the reliability of the test. Invasive tests such as amniocentesis may detect Down syndrome nearly 100%, but it should be the last option while it may cause the loss of mother's life or end of pregnancy [73].

Biochemical markers which are AFP, hCG, and uE3 in the maternal serum sample and other factors that are age, gestational age, weight, multiple pregnancies, family health story, ethnicity, smoking, and diabetes are used in the triple screening test.

2.7.1 Biochemical Markers

Alpha-fetoprotein screening (AFP), Human chorionic gonadotropin (hCG) and Unconjugated estriol (uE3) are biochemical indicators or markers.

2.7.1.1 Alpha-Fetoprotein (AFP)

AFP was firstly identified as a fetal particular globulin in 1956 [74]. It is produced in a yolk sac, which is a membranous sac attached to an embryo, liver of the infant, and gastrointestinal tract. Fetal plasma quantities are highest between 10^{th} to 13^{th} weeks of pregnancy and decrease continuously till period [42] when maternal levels highest in 3^{rd} trimester [75]. Level of the AFP value during the unaffected pregnancy is given in Figure 2.10. Multiples of the Median, MoM, reports Laboratory measurements of AFP levels.

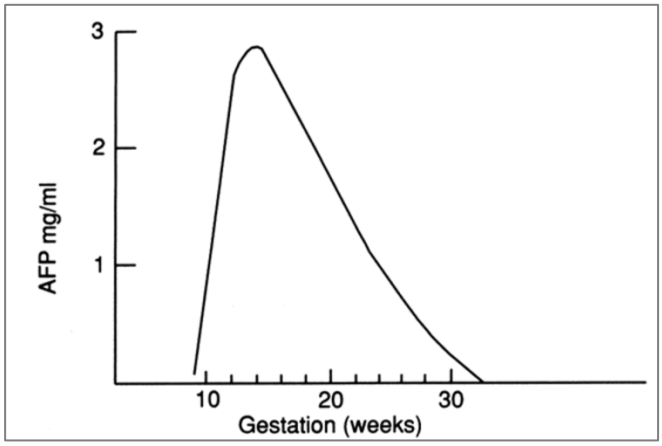


Figure 2.10 AFP value in fetal serum [76]

Doctors require identifying how their reference laboratories tell AFP outputs. Several causes and conditions related to raised and dejected AFP quantities [75]. The general cause is an inaccurate guessed gestational age for an unnatural AFP level [77]. On the other side, AFP increases continuously in unaffected pregnancies as presented in Figure 2.11. AFP is better than ultrasonography to detect Neural tube defects and it is the only indicator on triple screening test suitable for Neural tube defect recognition. It may discover 90% of anencephalies pregnancies and 80% of spina bifida situations [75].

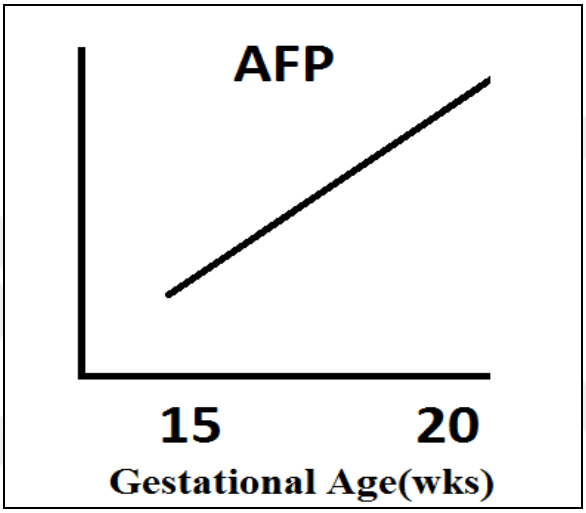


Figure 2.11 AFP levels during the unaffected pregnancies [78]

Unnatural amounts of AFP can indicate:

- Open neural tube defects like spina bifida
- Trisomy 21
- Chromosomal anomalies
- Failures in an abdominal wall of the infants
- Multiple pregnancies
- A get it wrong due date, as the levels differ through the prenatal period

2.7.1.2 Human Chorionic Gonadotropin (hCG)

hCG, which is a complicated glycoprotein, is created completely via syncytiotrophoblast later sewing the uterus tissue. That rises quickly in the first 8 weeks of pregnancy [79]. hCG reduces gradually till twenty weeks, when it plateaus

[80]. Figure 2.12 shows that the median level of hCG during the screening period in unaffected pregnancies. Maternal weight and parity have an impact hCG level [75]. The risen hCG amount is visible for the most delicate indicator [81, 82]. The low human chorionic gonadotropin level is related to Trisomy 18 [80]. The hCG levels are normal in Neural tube defects. Amniocentesis for females who are older than 35 years and who are a younger than 35 years of an age calibrated AFP level signifying a risk of Down syndrome equal to that of a 35 year old, cases of trisomy 21 can be detected with 25% – 50% [75]. The insertion of human chorionic gonadotropin to the alpha-fetoprotein screen rises discovery of Trisomy 21 nearly 40%- 50% over AFP alone [82].

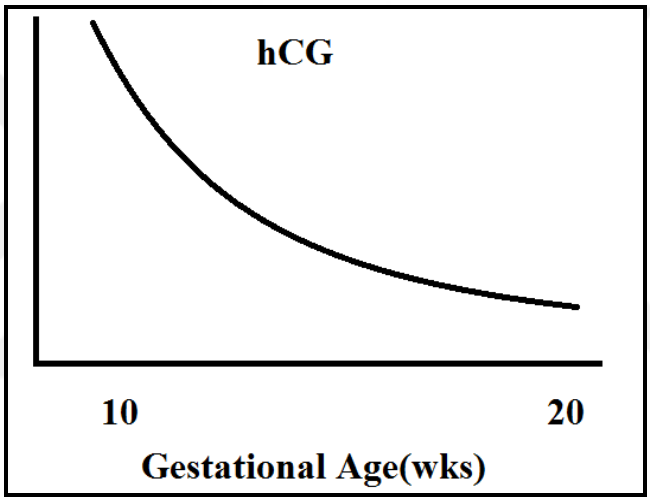


Figure 2.12 Median levels of hCG during the screening period in unaffected pregnancies [83]

2.7.1.3 Unconjugated Estriol (uE3)

Placenta creates uE3 from precursors prepared by the fetal adrenal glands and the liver [80]. The uE3 rises continuously along the gestation to an upper level than ovaries produce [84]. The uE3 levels are reduced in trisomy 21 and trisomy 18 [75]. In unaffected pregnancies, it increases during the screening period as shown in Figure 2.13. The supplement of uE3, hCG, and AFP raises the recognition of Down syndrome in females who are younger than 35 years old [82, 85].

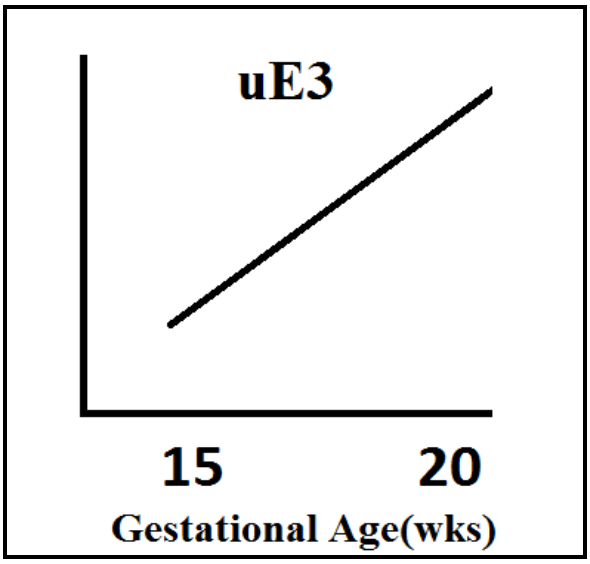


Figure 2.13 Median levels of uE3 during the screening period in unaffected pregnancies [86]

2.7.2 Other Factors Potentially Affecting Screening Test Performance

Various variables were detected that affect the results while performing the triple tests. These include age, the method of gestational age determination, maternal weight, multiple pregnancies, family history, ethnicity, smoking, fetal sex, and diabetes [47].

2.7.2.1 The Method of Gestational Age Determination

The generally used method is the start date of last menstrual period and various ultrasonographic measurements. In the first trimester of pregnancy, CRL (Crown Rump Length) is important in defining the gestational age. In the second trimester, Biparietal Diameter (BPD) or long bone measurements that are femur or humeral length become important. CRL is very important in 7 days. As the baby grows, the measurement accuracy also reduces. Because of that, frequent ultrasonographic examination at first trimester increases sensitivity [47]. The false positive rate is lower about 2% for all marker combinations when pregnancy period is predicted with scanning. For instance, scan timing might decrease the false positive rate of maternal blood from 4.2% to 2.7% for a DR of 85% [13].

2.7.2.2 Maternal Weight

Studies showed that there was an association between maternal weight and Down syndrome. One of these studies was done by Reynold and Penney in 1991 [87]. Blood samples were taken from the pregnancies and biochemical markers were observed in the study. After analysis, there was a correlation between maternal weight and AFP, hCG. However, there was no relationship between uE3 and maternal weight. The relationship between maternal weight and markers is presented in Table 2.2 [87].

Table 2.2 Median serum AFP, hCG, and uE3 MoMs related to maternal weight [87]

Maternal Weight(kg)	AFP(MoM)	hCG(MoM)	uE3(MoM)
< = 50	1.37	1.36	0.98
51 – 60	1.30	1.04	0.95
61 – 70	1.08	1.06	0.95
71 – 80	0.94	1.00	0.89
81 – 90	0.85	0.95	0.93
> 90	0.83	0.75	0.94

As the AFP value increases, the risk factor decreases. In addition, the increments in hCG value cause the risk factor to increase. As shown in Table 2.2, the value of AFP decreases as maternal weight increases. Low AFP causes a high risk factor. As a result, the possibility of Down syndrome baby increases according to AFP marker when maternal weight increases.

The same probability is valid for hCG. Its value decreases as maternal weight increases. When hCG value decreases, the risk factor decreases. As a result, the possibility of having a baby with Down syndrome decreases according to AFP marker when maternal weight increases. In the study, two markers were combined and the effect of weight correction on the numbers of women designated as "*increased risk*" by calculation of Down syndrome risk factors was shown in Table 2.3 [87].

Table 2.3 Risk effects between weight and markers [87]

	No. of women in risk group (% of 1408)	**No. of women reclassified no longer at increased risk**	**No. of women reclassified as now at increased risk**
No correction	78(5.5%)		
AFP only	91(6.5%)	4(89 ± 7 kg)	17(56 ± 6 kg)
hCG only	74(5.3%)	9(56 ± 6 kg)	5(91 ± 13 kg)
AFP+hCG	82(5.8%)	0	4(55 ± 9 kg)

Use each of AFP and hCG correction decreases the women number who are reclassified as a modification of an analyze partially retrieve for the influence of the other. The average weight of women is low that is described by the bigger tilt of the association between weights and AFP.

2.7.2.3 Multiple Pregnancies

Since the marker values in multiple pregnancies are obtained by the contribution of both babies, they have generally doubled to 2.0 MoM instead of 1.0 MoM. uE3 is not like that. When compared to single pregnancies, the change in MoM in twin pregnancies is around 1.7 MoM. This screening test is not considered compatible with twin pregnancies [88].

2.7.2.4 Family History

If family members have Down syndrome or other chromosomal abnormality stories, then the mother who is in the family has a higher risk [89].

2.7.2.5 Ethnicity

Dissimilarities have in the stages of screening test indicators between females of various national ancestries afterward accounting for parental mass [90]. Maternal serum AFP is 15% higher, total hCG is 18% higher, Inhibin A is 8% lower, and PAPP-A is 35% higher in Black women than in Caucasian women [91]. AFP is 6% lower, uE3 is 7% higher, total hCG is 6% higher and PAPP-A is 17% higher in South Asian women. Higher levels of the first trimester PAPP-A and hCG are seen in Asian women, and higher uE3 is seen in Aboriginal Canadian women [92]. Adjusting for ethnic origin slightly increases the detection rate (DR) for a given

false positives (FP), but, more importantly, it tends to equalize the FP among women of different ethnic groups [92]. In a statistical way, considerable differences in Neural tube measurement have been found between ethnic groups [91, 93, 94]. Nevertheless, it appears these differences might be very small to assure correction [11].

2.7.2.6 Smoking

The risk ratios for the association of cigarette smoking around the time of beginning with Down syndrome was 58% in the case defect control comparison and 56% in the case normal control comparison [95]. First and second trimester hCG levels are 25% lower, while Inhibin A is 50% - 60% higher in smoker mothers [95].

2.7.2.7 Insulin-Dependent Diabetes Mellitus

Some second trimester blood indicators incline to be lesser in pregnancies with insulin dependent diabetes mellitus. AFP and hCG are lower in diabetic females after weight checking. Other indicators don’t change in diabetic women [92, 96, 97]. In order not to differ, the observed MoM for a diabetic female is divided into related median MoM in diabetic females without Down's syndrome gestations. Due to deficiency of information in diabetic females that have a Down syndrome gestation, the pseudo risk may be computed for diabetic females [92]. So that the amounts of hCG, PAPP-A, and nuchal translucency aren't differed so much with or without insulin dependent diabetes [98].

2.8 Mathematical Expressions and Calculations Used in Screening Tests

Screening tests can never be expressed as positive or negative. However, the likelihood of that may be related to the reliability of the test. Some mathematical expressions are used to estimate the Down syndrome risk.

These expressions are:

- Detection rate (DR)
- F-Measure
- Specifity
- False positive rate (FPR)
- Receiver Operating Characteristic (ROC) Area
- Multiples of Median (MoM)

- Normal Distribution
- Likelihood Ratio (LR)

2.8.1 Detection Rate (DR)

The accepted description of DR is a proportion of Down syndrome fetuses which are accurately recognized. Otherwise, it is a ratio of true positives to all positives [99]. Also, the detection rate is called sensitivity and recall [100]. Equation 2.1 defines detection rate.

Detection Rate = True Positives / All Positives (2.1)

2.8.2 F-Measure

The F-measure is identified by measuring of a test's precision in a statistical analysis of binary classification. A precision and a recall are analyzed of the test to evaluate the value. Its value can be clarified as a weighted mean of the precision and recall. The best score of F-measure is 1 and worst value is 0 [101]. It is as defined in Equation 2.2.

F-Measure = (2 x Precision x Recall) / Precision + Recall (2.2)

2.8.3 Precision

Precision can be named true positive accuracy, being a compute of an accuracy of estimated positives contrary to the rate of detection of real positives [100]. Precision is as defined in Equation 2.3.

Precision = True Positives / (True Positives + False Positives) (2.3)

2.8.4 Specifity

It is true negatives in all negative (unaffected) results [100]. It is as defined in Equation 2.4.

Specifity = True Negatives / (False Positives + True Negatives) (2.4)

2.8.5 False Positive Rate (FPR)

Actually, the statistical examination is an error examination. A statistical test does not assure confidential results; it just calculates the possibility of error of a given the

result [102]. FPR means that a subject without the chromosomal aneuploidy is misclassified like having the chromosomal aneuploidy on the essential of the screening test. The subject gives the uncertain impression that the baby has the disease and therefore endures the redundant psychological results as well as having to undertake possibly invasive diagnostic or treatment procedures [103]. It is a percentage of all negative results in all positive results as defined in Equation 2.5.

False Positive Rate=False Positives / (False Positives + True Negatives) (2.5)

2.8.6 Receiver Operating Characteristic (ROC) Area

Receiver operating characteristic arc [104] is a diagrammatic depiction of a diagnostic capability of the binary classifier technique like the discriminative cut off value changes. The ROC curve is plotted by plotting the true positive ratio at various threshold settings with a false positive rate. The ROC arc can be created by drawing the accumulative distribution function of the true positive rates on the "y axis" against the accumulative distribution function of the false positive rates on the "x axis". ROC analyses support instruments to choose the best possible patterns and to reject non-optimal models regardless of the cost circumstance or class distribution. The ROC analysis is clearly and definitely associated with cost and benefit analysis of diagnostic decision making. A point that shows 100% accuracy and 100% specificity in the upper left corner or coordinate (0, 1) of the ROC area are given by the best possible estimation method. (0, 1) is also called an excellent classification. In other words, area 1 represents an excellent test; an area of 0.5 represents a worthless test.

2.8.7 Multiples of Median (MoM)

It means that an amount of how far a single test output diverges from the median. The outputs of the single tests are highly flexible and it is generally declared outputs of the screening test [105, 106]**.** The risk factor is assessed based on normal blood tests after blood test results are calculated. A mean of normal results is named as population median. As a result, the "*mean*" value is standardized 1.0 MoM. AFP and uE3 levels are lower in pregnancies of Trisomy 21, so the levels are below the average and therefore lower than 1.0 MoM. Likewise, hCG would be more than 1.0

MoM in a trisomy 21 pregnancy. The laboratory result as a total risk factor is computed by a specific computer program in the screening tests. Then, these statistically significant normal screening tests are calculated for each gestational week (8^{th} -11^{th} weeks and 15^{th} – 22^{nd} weeks). Two large organizations are preparing data for these processes. The SURUSS (Serum Urine Research and Ultrasound Screening Study) and FASTER (First and Second Trimester Evaluation of Risk) are research institutes that collect data on this issue. The findings are subjected to a different calculation. The values for each gestational week are sorted from small to large or from large to small. The value in the middle is considered "*average*" (median). Then, the result of each test is divided by a median. The last column, which is MoM in Table 2.4, was obtained by dividing the values in a first column by a middle value in a sequence.

Table 2.4 Calculation of MoM

Marker Value	**Sorted Marker Value** (Ng/ml)			**MoM**
10	10	(10/40)	=	0.25
40	20	(20/40)	=	0.50
20	**40**	(40/40)	=	1.00
80	60	(60/40)	=	1.50
60	80	(80/40)	=	2.00

2.8.7.1 Importance of the MoM

1- It prevents systematic changes between laboratories from making it difficult evaluation. (Unit, mode of operation, method, etc.)

2- It stabilizes marker levels that have fluctuations variables with pregnancy. It makes evaluation easier.

3- MoM values are unitless and can be evaluated on the same chart, regression curve, and criteria. Thus, it is easier to reach the patient's MoM value from a flat skew by working with fewer women [107].

The previous table has shown the calculation of the MoM. However, if it is desired to compare the results of two different laboratories as shown in Figure 2.14, the marker values that are under different conditions must be standardized and converted to MoM. As shown in the Figure 2.15, the Laboratory B looks at the

marker value at 11th weeks while the Laboratory A looks at the marker value at 13th weeks.

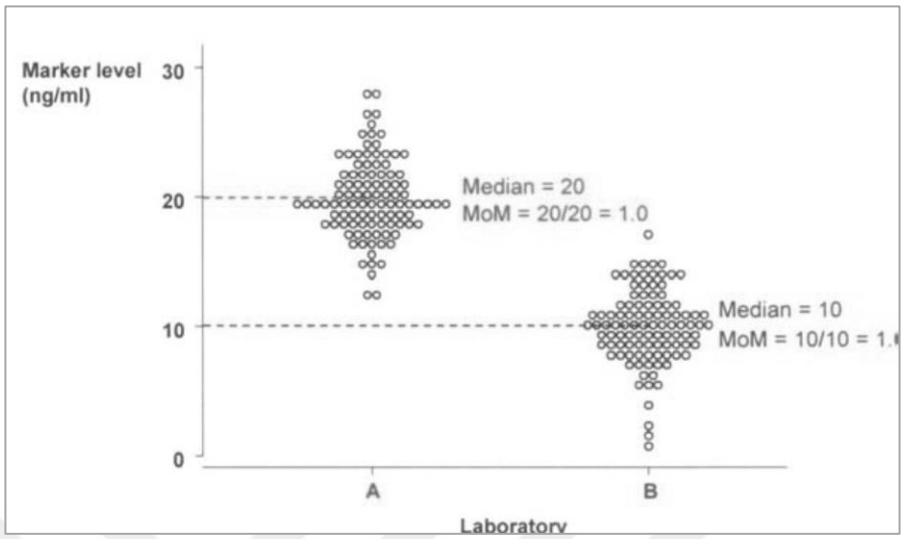


Figure 2.14 Calculation of median and MoM graphically.

It is necessary to eliminate the difference between the laboratories and to standardize their values. Standardized and converted values, which are under the different conditions, are shown as MoM in Figure 2.14. So that the measured marker values at different conditions in different laboratories can be compared by means of the MOM.

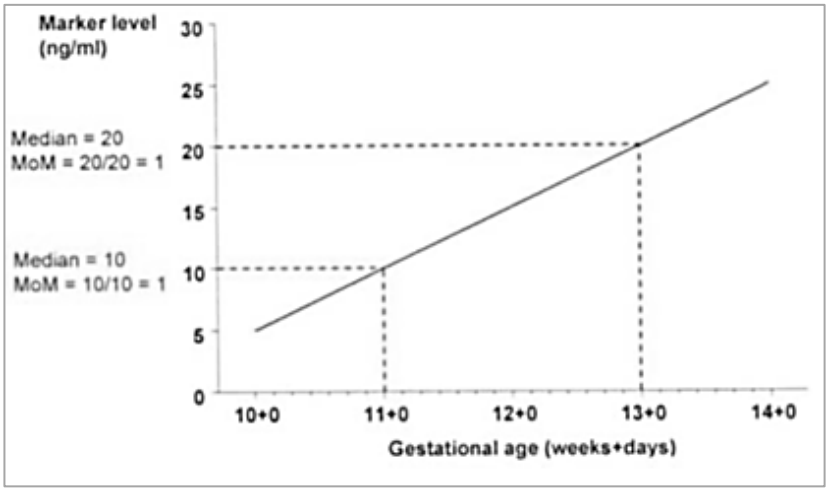


Figure 2.15 Two marker values on the same chart with MoM

2.8.8 Normal Distribution

The Gaussian distribution is widely used for probability. In statistics, Normal distribution is critical and it is frequently used in the natural and social sciences to present actual valued random variables whose distributions aren't known [108, 109]**.** The normal distribution graph is shown in Figure 2.16.

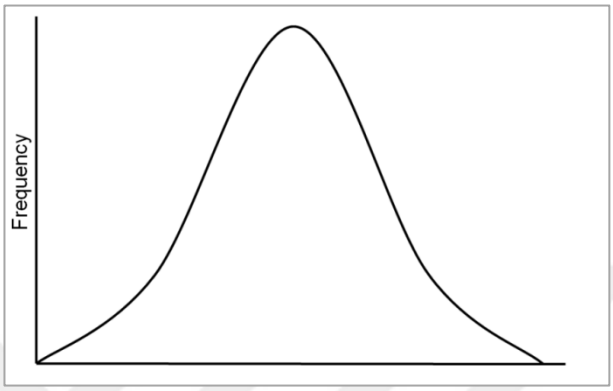


Figure 2.16 Normal distribution graph

The normal distribution is important due to the central limit theorem. Physical amounts that are estimated to be the sum of many independent procedures frequently have distributions that are closely normal. Moreover, numerous outputs and techniques may be created from systematically in the specific model when the related variables are normally distributed [110].

Though various other distributions are bell-shaped, the normal distribution is occasionally named the bell curve [111]. The first object to estimate Down syndrome is to draw the Gaussian curve for the markers and reach from there to the average. These values are the MoM values for each patient on the x axis and the number of cases on the y-axis for the test. Example of the subject is shown in the Figure 2.17.

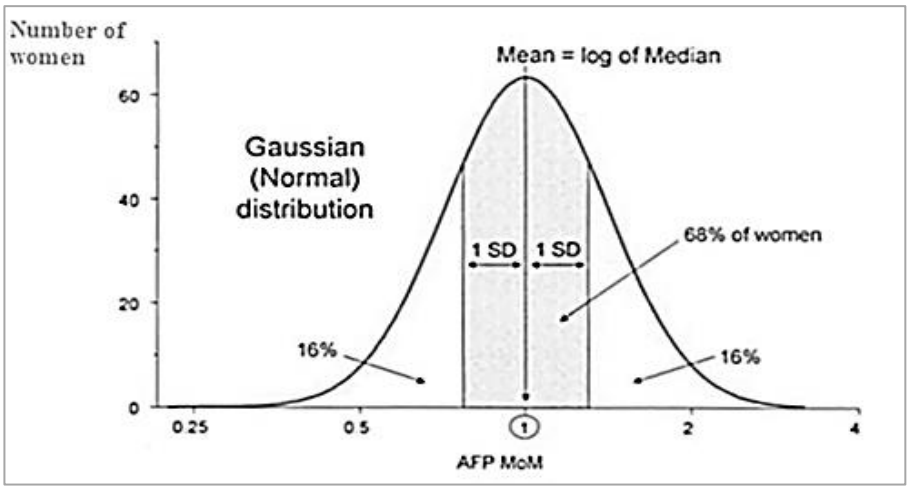


Figure 2.17 Distribution of second trimester AFP MoM values in unaffected pregnancies

2.8.9 Likelihood Ratio (LR)

The first definition of probability ratios for decision rules was made on the information theory in 1954 [112]. Probability ratios were represented between 1975 and 1980 in medicine [113–115]. It is the mathematical expression of the possibility which the given test is truly positive and it is defined in Equation 2.6.

Likelihood Ratio = Sensitivity / (1- Specifity) (2.6)

The affected and unaffected pregnancies of positive results are drawn as a Gaussian curve on the same coordinate plane to calculate the LR. It is seen that two curves coincide in one place after drawing. If a straight line is drawn upward from the mean (MoM) value which is previously determined, with the help of the Gaussian curve, the ratio of the unaffected of the curve to the affected curve pregnancies gives the LR of that test. The purpose of searching this value is to show whether the test is a good parameter for a screening test or not. The following curves in Figure 2.18 show the DR (detection rate) value of that parameter. The middle overlap area of both curves (darkened area) is a false positive area. This value will be given as "*Odds of being affected given a positive result (OAPR)"* of any test.

Example:

The value, which is set for the marker, is 2.5 MoM

A vertical line drawn up from here will give the LR of the patient. For this example, it is LR = 75/3 = 25.

The prevalence of the disease in a certain period of time in that society is 2/1000. So 2 out of 1000 people get Down syndrome.

Then;

OAPR = 25x2/1000 = 50/1000 = 1/20

So, on this value of MoM;

The risk of Down syndrome will be 1 negative patient in 20 positive patients with the same parameter value at the same age and condition. Figure 2.18 shows all calculations and its graph.

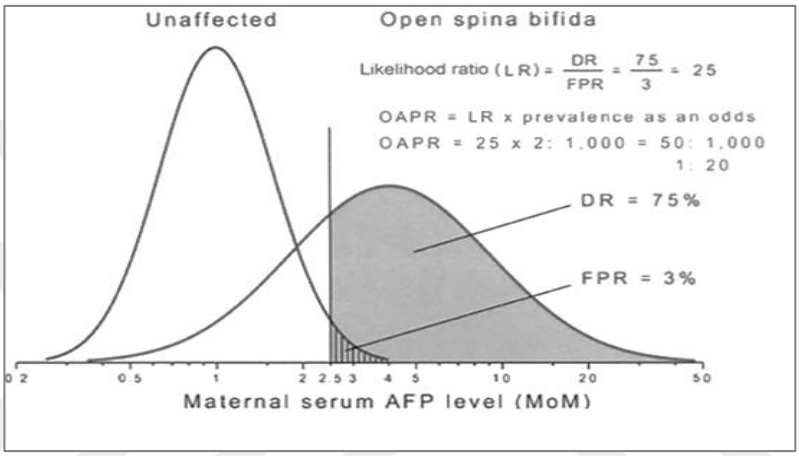


Figure 2.18 Calculation of Likelihood ratio graphically

Another example:

The value, which is set for the marker, is 3.0 MoM

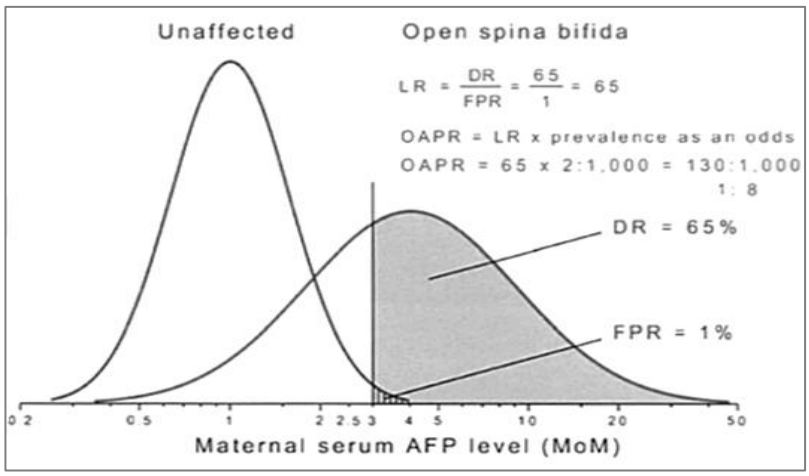


Figure 2.19 Likelihood ratio (LR), Detective rate (DR), false positive rate (FPR)

In Figure 2.19, the area of the affected and unaffected curves in the line drawn up from 3.0 MoM is 6.5 and 1. They are 65% for DR and 1% for FPR.

Also assume that the prevalence is 2: 1000

OAPR = 65x2/1000 = 130/1000 = 1/8.

That is 1 out of every 8 positive results are negative (false positive rate).

2.8.10 Calculation of Specific Risk Analysis for Triple Screening Test

Consider a patient with a serum AFP value of 2.5 MoM:

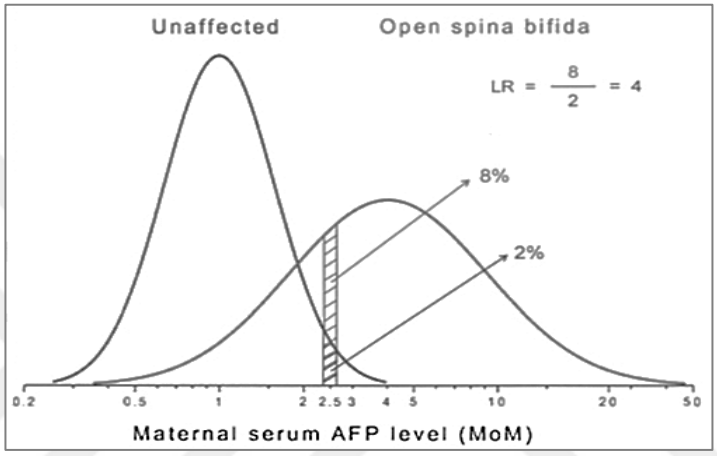


Figure 2.20 Likelihood for calculation specific risk analysis

In Figure 2.20, the unaffected area is 8 and affected area is 2 in the vertical line which is drawn up.

LR = 8/2 = 4

OAPR = 4x2/1000 = 8/1000 or 0.8%

The result is expressed as 8:1000. The test result for this patient is negative as shown in Figure 2.21.

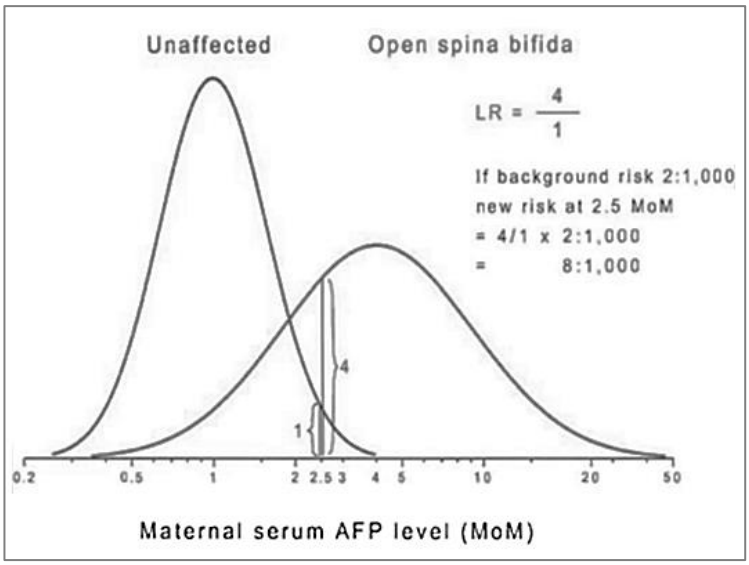


Figure 2.21 The result of specific risk analysis

A patient with a MoM value of 1.0:

The vertical line drawn to the unaffected section is mostly in the area of the affected pregnancies.

LR = 1/10

In Figure 2.22 is given that the prevalence is 2:1000, and the result is 1x2/10000 = 2/10000.

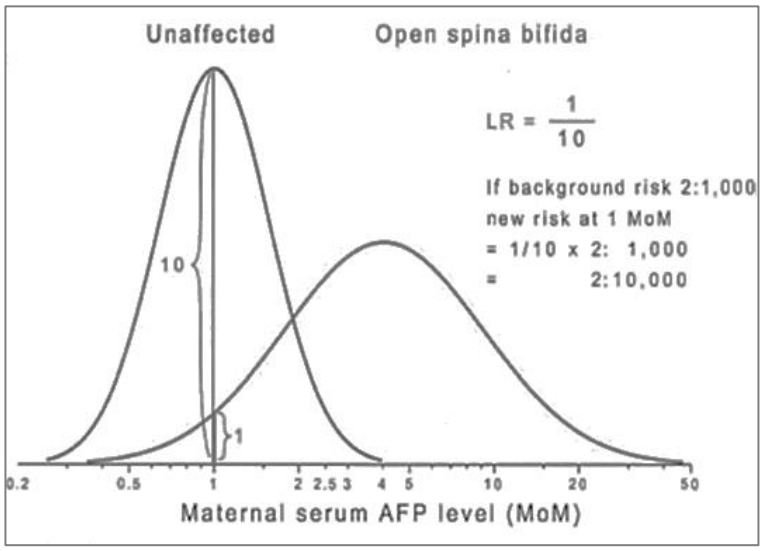


Figure 2.22 Another result of specific risk analysis

2.8.11 Calculation of Risk Analysis with Combining All Markers

A complex formula called Bayes Theorem [114] is used to find the correlation that determines the relationship of each value. Also, N.J. Wald and colleagues reported that the calculation using Likelihood Ratio, DR, FPR, and the prevalence is more accurate in 2004 [97]. All of these parameters are mathematically combined with the previously prepared log linear curves and the reflection of the healthy populations to the current markers described above. Then the computerized combination is used to analyze the risk of the mother. A correlation coefficient for each marker is calculated and a mathematical indicator of the relationship between them is formulated. Age, ethnicity, smoking, illnesses in the family, individual and family stories, last menstrual period, age, weight and pre-calculated median values, table values are included in this set. When the information from the patient is entered into the system, they are mathematically combined with the tables in the memory and a risk analysis of the patient is given.

CHAPTER 3

METHODOLOGY

In this part of the study, the types of data used in the study, the characteristics of the data used, the elimination of data, a balancing of an unbalanced dataset, the variables in the balanced data, the software used in the study, the relationship between the variables, a brief summary of the algorithms which are used in the data classification, and the analysis of the classification will be explained.

3.1 Data Acquisition for Study

In order to develop a classification performance of a triple screening test via using data mining, this study was carried out to examine the pregnancies that have records in Sahinbey Research and Practice Hospital of Gaziantep University. The permissions for this study were obtained from the ethics committee of Gaziantep University, the head of the Industrial Engineering Department of Gaziantep University and Administration of Sahinbey Research and Practice Hospital of Gaziantep University. Ethics committee decision papers can be found in APPENDIX A.

The thesis population consisted of women who attended Sahinbey Training and Research Hospital, Gaziantep, for their antenatal care between 2010 and 2016. Patient records or data were received from the department of statistics, obstetrics & gynaecology clinic, biochemistry laboratory and molecular genetics laboratory of the hospital. Amniocentesis records from patient data were added to the data set by accessing the amniocentesis report of each patient from the molecular genetics laboratory. Maternal serum samples that had AFP, hCG, uE3 levels, and maternal age, weight, etc., were taken from the triple screening test results reports in the biochemistry laboratory. Samples of amniocentesis and triple test reports can be

found in APPENDIX B. 6340 patients who have given birth, 324 patients who aborted, and 2815 patients who have amniocentesis data were reached. Approximately 8,000 patient data were examined in total. Also, the hospital uses an algorithm to determine Down syndrome (Trisomy 21). Figure 3.1 represents a specific algorithm for Down syndrome.

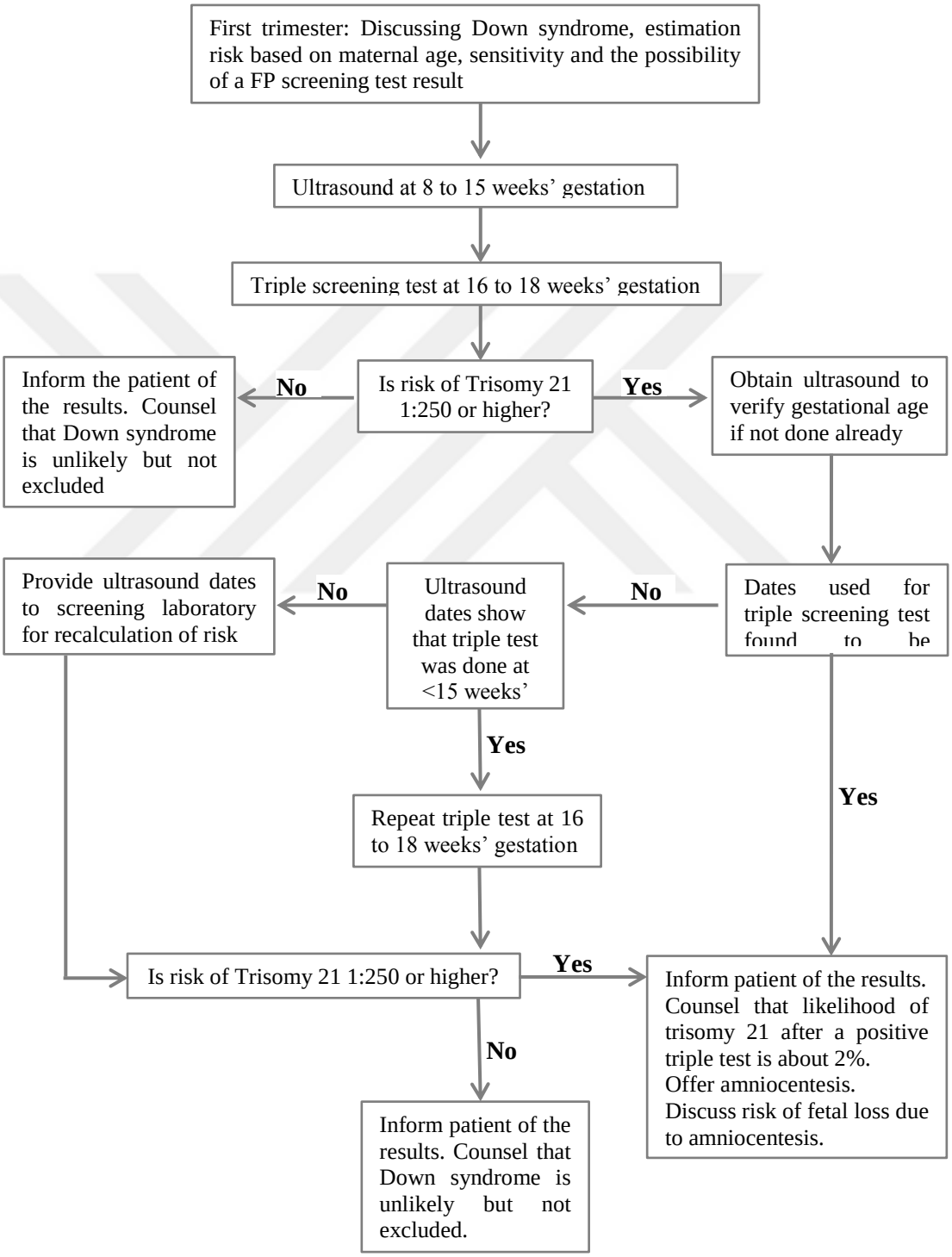


Figure 3.1 Algorithm for determining Down syndrome

3.2 Elimination of Data

The pregnant women must have all examinations and surveys in the same hospital until the birth so that the patient records could be part of the study. Thus a patient who gives birth healthy or with an anomaly, and who had the amniocentesis and triple screening result entered into the data set with matching patient numbers, file numbers and patient names. The remaining patient data was deleted from the data set. As a result, there were 81 pregnant women who have the whole record indicating the baby's genetic disorder status. 76 of them had no genetic disorder and 5 of them had trisomy 21. The number of patients extremely decreased to 81. The reason of decreasing is some women who have not gone the same hospital for all examinations and surveys during the pregnancy and births were not in the same hospital.

3.3 Variables Used in the Study

Variables used in the study are gestational age, maternal age, smoking status, ethnicity, pregnancies by IVF method, amniocentesis result or health status of baby AFP, hCG and uE3 levels. 81singleton pregnancies had triple screening tests and complete records of whether or not the baby had Down syndrome. The markers AFP, hCG, and uE3 were observed from maternal serum samples. Gestational age based on ultrasonography bi-parietal diameter (BPD) measurements. An example figure of the BPD is as shown in Figure 3.2.

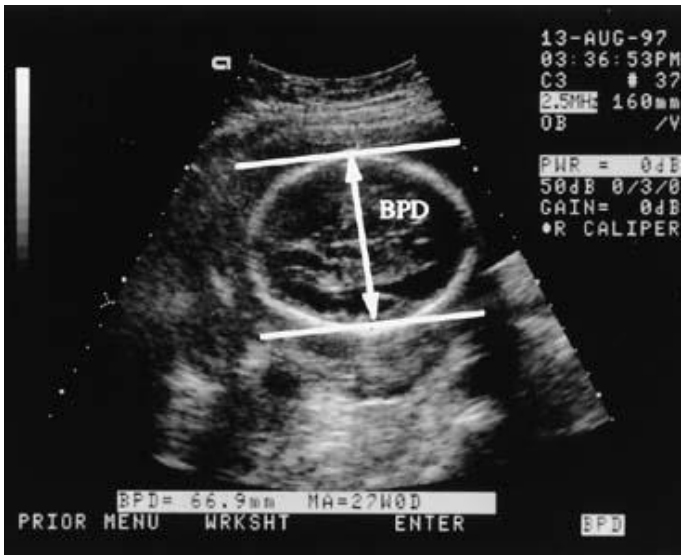


Figure 3.2 BPD example

All biochemical tests were performed on a Beckman Coulter Unicel DXI 800 medical device as shown in Figure 3.3. Besides these data, maternal age, gestational age, maternal weight, ethnicity, and smoking condition were also examined in the study.

Figure 3.3 Beckman Coulter Unicel DXI 800 Synchron Clinical Systems

The data set is composed of 14 different variables of 81 patients in the initial unregulated case. The initial size of the dataset is 81x14 and statistical information about the initial dataset of patients is given in APPENDIX C. The description of the variables that make up the data set is as follows:

Patient ID: It is the identification number for each patient from 1 to 81.

Age: It includes the age of the patient. The unit is years.

Weight: Includes the weight of the patient. The unit is kg.

GestationalAge: It indicates the gestational week and day when measured the maternal serum values of the patient.

Smoking: It includes the status of the maternal smoking status.

IVF (In Vitro Fertilization): Express whether or not the mother candidate has performed the pregnancy with the tube baby method.

Ethnicity: It indicates the race of the mother.

AFP: Specifies the value of the AFP marker in the maternal serum sample. The unit is in the form of IU / ml.

AFPMoM: Standardized AFP level.

uE3: Specifies the value of the uE3 marker in the maternal serum sample. The unit is in the form of IU / ml.

uE3MoM: Standardized uE3 level.

hCG: Specifies the value of the hCG marker in the maternal serum sample. The unit is in the form of IU / ml.

hCGMoM: Standardized hCG level.

Result: It indicates the genetic disorder status of the baby. It is defined Negative or trisomy 21.

3.4 Used Software for Data Mining

Weka and SPSS software products were used in this thesis.

3.4.1 Weka

Weka includes a collection of imagining instruments and algorithms for the data examination and estimator forming, together with graphical user interfaces (GUI) for simple entry to the functions [116]. The non-Java version of Weka is a Tcl / Tk front end for modeling algorithms implemented in other programming languages, as well as data preprocessing utilities in C and a Makefile based system for running machine learning experiments. The original version of Weka was first prepared to analyze the data in agricultural areas [117, 118], but nowadays the new version of Weka is used in many areas such as education and researches. WEKA has many advantages and these are:

- Free accessibility under the GNU General Public License.

- Compactness, because it is entirely applied in Java programming language and therefore runs on approximately several latest computing platform.
- The extensive accumulation of the data preprocessing and modeling methods.
- It provides ease of use due to user friendly GUI.

Various standard data mining missions, extra specially, data preprocessing, clustering, classification, regression, visualization, and feature choice are supported by the Weka. All of Weka's methods are established on a hypothesis which the data is accessible like one suitable file or relationship, where all data element is defined by a rigid number of features. Weka ensures access to Structured Query Language (SQL) databases through Java Database Connectivity and it might manage an output reverted by the database inquiry. It isn't skillful of multi relational data mining; however there is various computer programming to transform a group of related database boards into the single board which is proper for evaluating with using Weka [119]. Additional significant area, which isn't involved by the algorithms currently in Weka distribution, is array modeling. There are 3 base modes in the main menu of Weka as given in Figure 3.4.

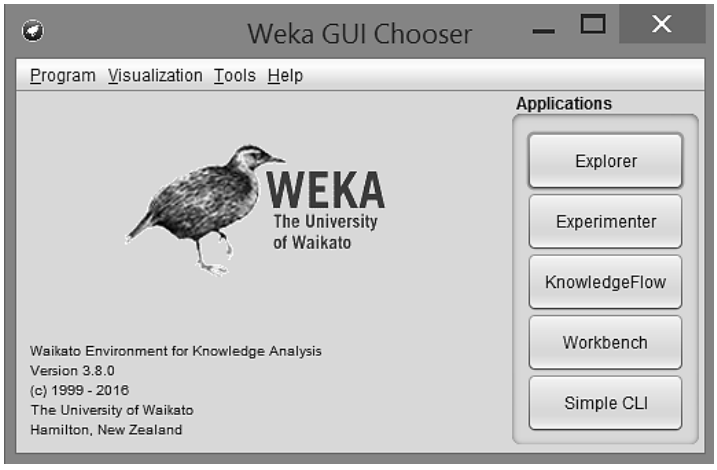


Figure 3.4 Main menu of Weka

Explorer is the main user interface of the Weka and it is as presented in Figure 3.5, but basically, the similar functionality may be reached from the component based interface and command line.

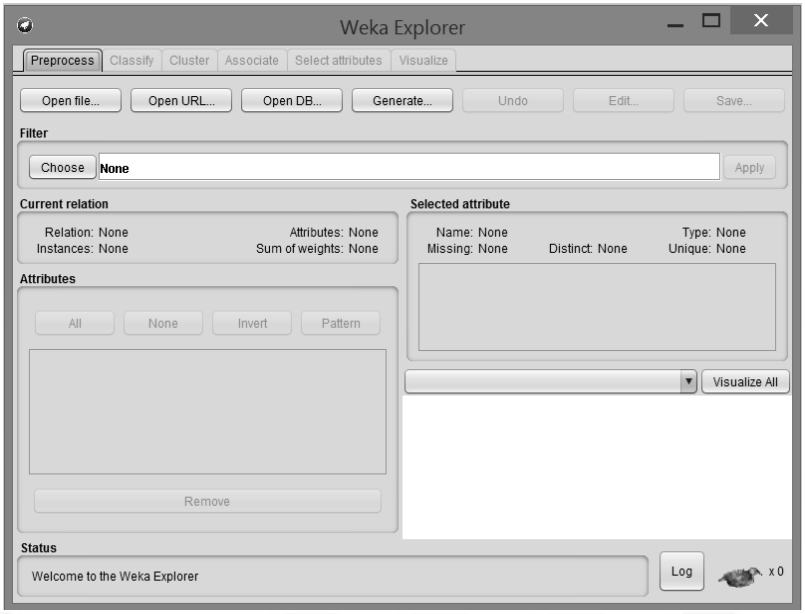


Figure 3.5 Explorer mode of Weka

There is also the Experimenter as given in Figure 3.6 that allows an organized collation of the predicted performance of Weka machine learning algorithms in a set of data.

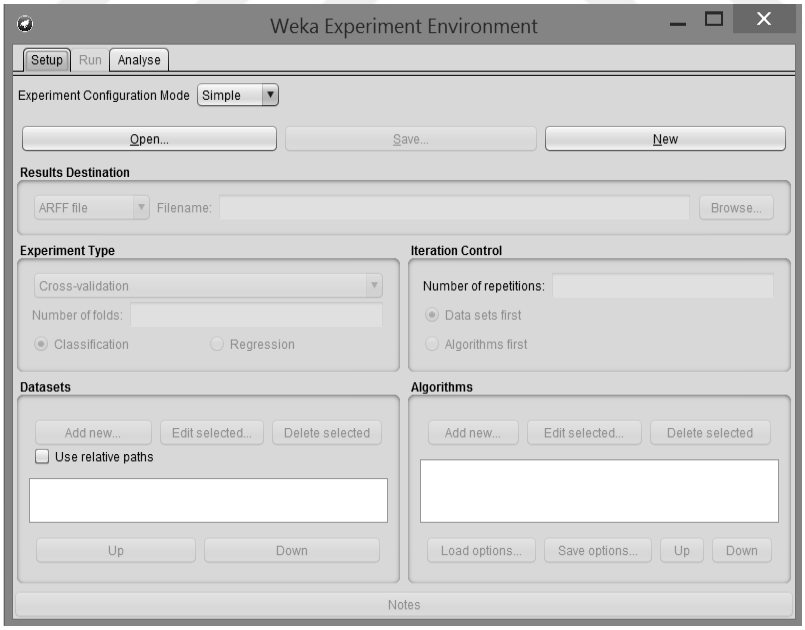


Figure 3.6 Experimenter mode of Weka

Explorer has a few panels. They provide access to the main components of the workbench [120]:

- ❖ Preprocessing panel contains a possibility to import data from a database, a comma-separated values (CSV) file, etc. These filters can be used to convert the data and delete the samples and build based on certain criteria.
- ❖ Classification panel allows classification and regression algorithms to be applied to the resulting data set.
- ❖ The association panel has access to the rule learner, which tries to recognize every important relation between the attributes in the data.
- ❖ Cluster module provides reach to clustering methods in the software.
- ❖ The Selection module enables algorithms for determining the cleverest estimator options in the data set.
- ❖ The Visualize module represents a scatter plot matrix, where individual scatter draws may be chosen, enlarged or examined with extra several choice operatives.

Weka was used in the thesis to oversample, evaluate relations between markers, classify data and analyze results.

3.4.2 SPSS Software

SPSS is a software package that is logically used for batch and non-blended statistical analysis. The SPSS graphical user interface shows in Figure 3.7. The software name was originally Statistical Package for the Social Sciences (SPSS) [121]. Software that is widely used for statistical analysis in social sciences is SPSS. Marketplace analysts, medicine analysts, questionnaire firms, administration, education analysts, advertising societies, data miners, and the others use SPSS. The original SPSS manual that was found by Nie, Bent, and Hull in 1970 has been identified for permitting common scientists to analyze their specific statistical analysis [122].

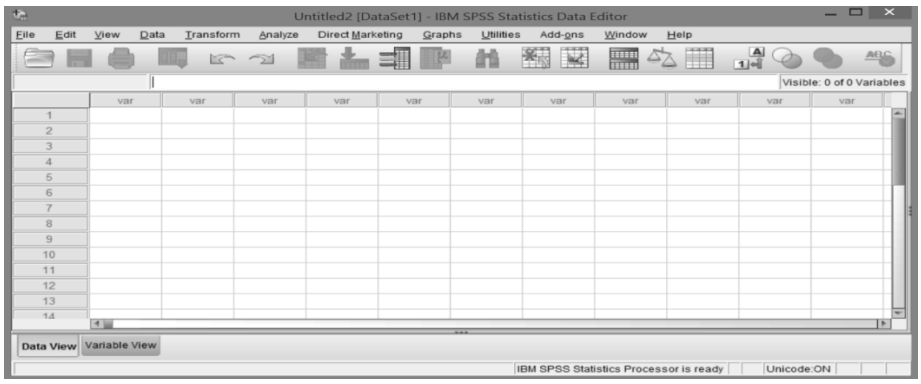


Figure 3.7 User interface of SPSS

Furthermore, data management and data documentation are basic characteristics of SPSS to analyze statistically.

Statistics that are contained in the main SPSS:

- Descriptive statistics: Cross tabulation, Frequencies, Descriptive, Explore, Descriptive Ratio Statistics
- Bivariate statistics: Means, t-test, ANOVA, Correlation (bivariate, partial, distances), Nonparametric tests
- Estimation for numerical results: Linear regression
- Estimation to identify clusters: Factor analysis, cluster analysis (two-step, K-means, hierarchical), Discriminant

Several characteristics of SPSS Statistics can be accessed by pop-up menus and programmable in copyrighted 4GL command syntax language. It provides a beneficial repeatable result, making simpler cyclical missions, and using complicated data processing and analysis. Also, specific multipart applications may just be designed in syntax and aren't available throughout the menu structure. A pop-up menu list additionally generates command syntax: it may be shown in the output; however the original sets must be transformed for doing the syntax to the user. The settings might additionally be attached into a syntax file using the "paste" button present in every menu. Programs may be compete interactively or unclaimed by using a provided Production Job Facility.

SPSS Statistics brings restrictions on interior folder construction, data kinds, data processing and matching file restrictions and simplifies programming significantly. SPSS data sets have 2 dimensional chart construction where the rows characteristically signify situations and the columns signify measurements. Barely, 2 data styles are described. They are numeric and string. Each data operations take place in order throughout the data set. Files may be paired one-to-one and one-to-many, however they cannot be many-to-many. Also, SPSS cannot store data cells may just include numbers and text, and formulas in cells [123]. Cells can be manually edited, describe the file architecture, and permit data input devoid of using command syntax. It can be enough for minor datasets. But major datasets like statistical examinations are further frequently used by online surveys. The datasets are computed into SPSS [124].

SPSS software was used in the thesis because Weka software was lacking in examining the relationship between variables. The relationship between variables was examined in SPSS and variables that did not satisfy $p < 0.01$ were deleted from the data set.

3.5 Data Preprocessing

Data pre-processing is the primary most important step for data mining or data analysis. Data preprocessing outputs entered directly into the data mining model and last outputs are detailed. The useful data resource doesn't just raise a precision of mining furthermore dramatically increases the productivity of the algorithm. Before the data mining algorithm is used, generally the data is processed [125]. Figure 3.8 presents preprocessing stages graphically.

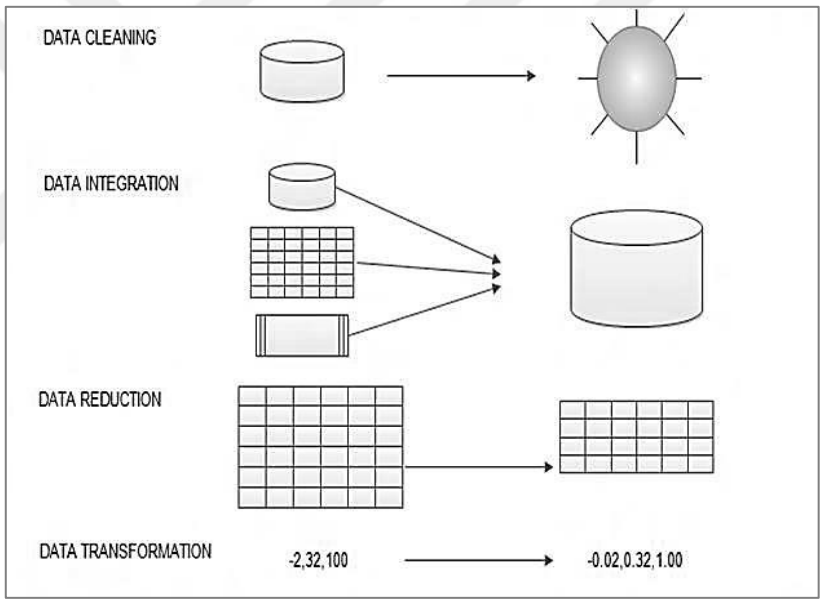
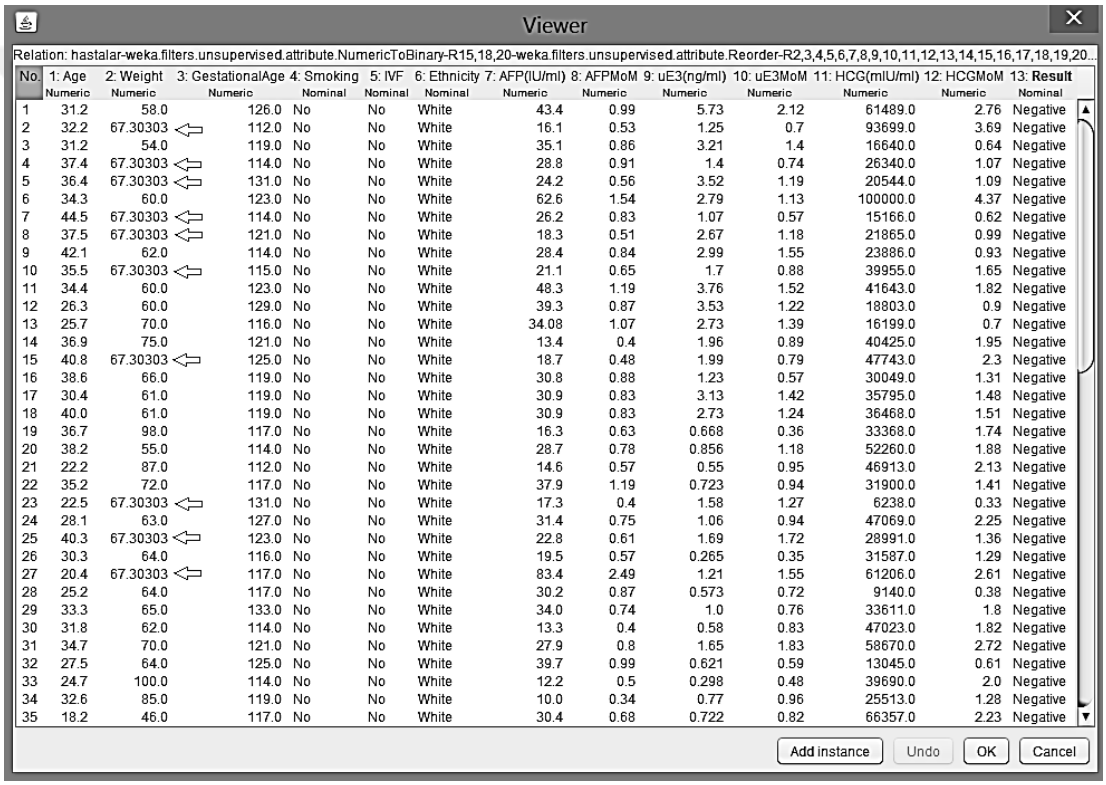


Figure 3.8 Data preprocessing stages

There are various tools and methods for selecting a sample cluster from the population. One of them is the transformation which changes raw data to create a particular record. Another one is denoising that eliminates noise from the data sets. Another one is normalization that arranges dataset for additional effective access and lastly, feature subtraction that gives specific data which is important in a particular context.

3.5.1 Completion of Missing Values

Weights of patients were found to lack in collected data. In total, 15 of 81 patients' weights were found to be missing. "*ReplaceMissingValues*" method is used to complete this missing data. It allows replacing every missing value of nominal and numeric attributes in a data set with modes and means in a training dataset. Thus, the missing 15 values were completed through the Weka by looking at the average of 66 patients. Average of them is 67.30303 as given in Figure 3.9. The statistical information about full data set of patients with completed missing values is given APPENDIX C.



No.	1: Age	2: Weight	3: GestationalAge	4: Smoking	5: IVF	6: Ethnicity	7: AFP(IU/ml)	8: AFPMoM	9: uE3(ng/ml)	10: uE3MoM	11: HCG(mIU/ml)	12: HCGMoM	13: **Result**
	Numeric	Numeric	Numeric	Nominal	Nominal	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
1	31.2	58.0	126.0	No	No	White	43.4	0.99	5.73	2.12	61489.0	2.76	Negative
2	32.2	67.30303	112.0	No	No	White	16.1	0.53	1.25	0.7	93699.0	3.69	Negative
3	31.2	54.0	119.0	No	No	White	35.1	0.86	3.21	1.4	16640.0	0.64	Negative
4	37.4	67.30303	114.0	No	No	White	28.8	0.91	1.4	0.74	26340.0	1.07	Negative
5	36.4	67.30303	131.0	No	No	White	24.2	0.56	3.52	1.19	20544.0	1.09	Negative
6	34.3	60.0	123.0	No	No	White	62.6	1.54	2.79	1.13	100000.0	4.37	Negative
7	44.5	67.30303	114.0	No	No	White	26.2	0.83	1.07	0.57	15166.0	0.62	Negative
8	37.5	67.30303	121.0	No	No	White	18.3	0.51	2.67	1.18	21865.0	0.99	Negative
9	42.1	62.0	114.0	No	No	White	28.4	0.84	2.99	1.55	23886.0	0.93	Negative
10	35.5	67.30303	115.0	No	No	White	21.1	0.65	1.7	0.88	39955.0	1.65	Negative
11	34.4	60.0	123.0	No	No	White	48.3	1.19	3.76	1.52	41643.0	1.82	Negative
12	26.3	60.0	129.0	No	No	White	39.3	0.87	3.53	1.22	18803.0	0.9	Negative
13	25.7	70.0	116.0	No	No	White	34.08	1.07	2.73	1.39	16199.0	0.7	Negative
14	36.9	75.0	121.0	No	No	White	13.4	0.4	1.96	0.89	40425.0	1.95	Negative
15	40.8	67.30303	125.0	No	No	White	18.7	0.48	1.99	0.79	47743.0	2.3	Negative
16	38.6	66.0	119.0	No	No	White	30.8	0.88	1.23	0.57	30049.0	1.31	Negative
17	30.4	61.0	119.0	No	No	White	30.9	0.83	3.13	1.42	35795.0	1.48	Negative
18	40.0	61.0	119.0	No	No	White	30.9	0.83	2.73	1.24	36468.0	1.51	Negative
19	36.7	98.0	117.0	No	No	White	16.3	0.63	0.668	0.36	33368.0	1.74	Negative
20	38.2	55.0	114.0	No	No	White	28.7	0.78	0.856	1.18	52260.0	1.88	Negative
21	22.2	87.0	112.0	No	No	White	14.6	0.57	0.55	0.95	46913.0	2.13	Negative
22	35.2	72.0	117.0	No	No	White	37.9	1.19	0.723	0.94	31900.0	1.41	Negative
23	22.5	67.30303	131.0	No	No	White	17.3	0.4	1.58	1.27	6238.0	0.33	Negative
24	28.1	63.0	127.0	No	No	White	31.4	0.75	1.06	0.94	47069.0	2.25	Negative
25	40.3	67.30303	123.0	No	No	White	22.8	0.61	1.69	1.72	28991.0	1.36	Negative
26	30.3	64.0	116.0	No	No	White	19.5	0.57	0.265	0.35	31587.0	1.29	Negative
27	20.4	67.30303	117.0	No	No	White	83.4	2.49	1.21	1.55	61206.0	2.61	Negative
28	25.2	64.0	117.0	No	No	White	30.2	0.87	0.573	0.72	9140.0	0.38	Negative
29	33.3	65.0	133.0	No	No	White	34.0	0.74	1.0	0.76	33611.0	1.8	Negative
30	31.8	62.0	114.0	No	No	White	13.3	0.4	0.58	0.83	47023.0	1.82	Negative
31	34.7	70.0	121.0	No	No	White	27.9	0.8	1.65	1.83	58670.0	2.72	Negative
32	27.5	64.0	125.0	No	No	White	39.7	0.99	0.621	0.59	13045.0	0.61	Negative
33	24.7	100.0	114.0	No	No	White	12.2	0.5	0.298	0.48	39690.0	2.0	Negative
34	32.6	85.0	119.0	No	No	White	10.0	0.34	0.77	0.96	25513.0	1.28	Negative
35	18.2	46.0	117.0	No	No	White	30.4	0.68	0.722	0.82	66357.0	2.23	Negative

Figure 3.9 Replace missing values on Weka

3.5.2 Balancing Work for Imbalanced Data

Imbalanced data often indicate a problem with classification troubles where classes are not evenly represented. It was observed that the pregnancies with the definite records of 76 of the 81 singleton pregnancies were detected unaffected and 5 of them associated with Down syndrome. This makes an imbalanced situation. Figure 3.10 shows this imbalance data attributes as a column chart in Weka.

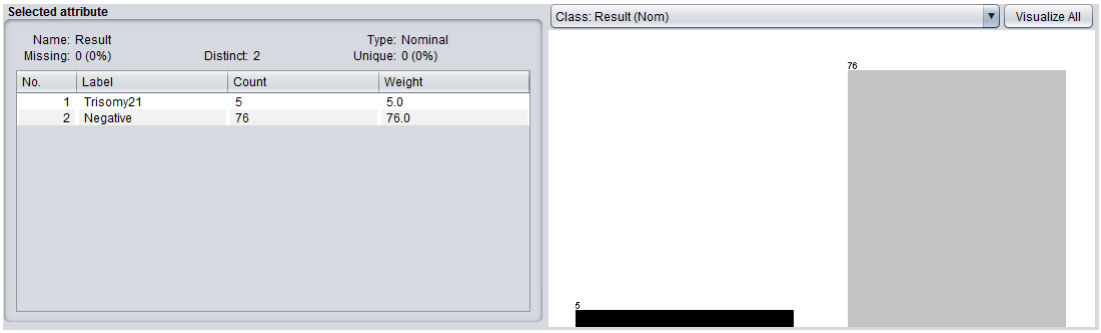


Figure 3.10 Imbalanced data set

3.5.2.1 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE (Synthetic Minority Oversampling Technique) [126] algorithm was found by Nitesh V. Chawla, Kevin W. Bowyer, and their teammates. Their approach about SMOTE was encouraged by a method which demonstrated accomplished in handwritten character identification [127]. In the examples, the minority class was over-sampled by taking each minority class sample and presenting synthetic samples along line segments participating all of the closest neighbors of the minority class. Neighbors are randomly selected from nearest neighbors as a requirement of excessive sampling. They used five close neighbors in the study. For example, just 2 neighbors from the 5 nearest neighbors are selected and 1 sample is created in the way of every one if the quantity of oversampling required is 200%. SMOTE algorithm is as shown in Figure 3.11 and synthetic examples are created as follows: Property taken into consideration sample is taken with the difference between its closest neighbors. This difference is multiplied by a random number between 0 and 1 and added to the feature vector of interest. This causes a random point to be selected along the line between the two special properties.

Algorithm *SMOTE*(T, N, k)
Input: Number of minority class samples T; Amount of SMOTE $N\%$; Number of nearest neighbors k
Output: $(N/100) * T$ synthetic minority class samples

1. (* *If N is less than 100%, randomize the minority class samples as only a random percent of them will be SMOTEd.* *)
2. **if** $N < 100$
3. **then** Randomize the T minority class samples
4. $T = (N/100) * T$
5. $N = 100$
6. **endif**
7. $N = (int)(N/100)$ (* *The amount of SMOTE is assumed to be in integral multiples of 100.* *)
8. k = Number of nearest neighbors
9. $numattrs$ = Number of attributes
10. $Sample[\,][\,]$: array for original minority class samples
11. $newindex$: keeps a count of number of synthetic samples generated, initialized to 0
12. $Synthetic[\,][\,]$: array for synthetic samples
(* *Compute k nearest neighbors for each minority class sample only.* *)
13. **for** $i \leftarrow 1$ **to** T
14. Compute k nearest neighbors for i, and save the indices in the $nnarray$
15. Populate(N, i, $nnarray$)
16. **endfor**

Populate(N, i, $nnarray$) (* *Function to generate the synthetic samples.* *)
17. **while** $N \neq 0$
18. Choose a random number between 1 and k, call it nn. This step chooses one of the k nearest neighbors of i.
19. **for** $attr \leftarrow 1$ **to** $numattrs$
20. Compute: $dif = Sample[nnarray[nn]][attr] - Sample[i][attr]$
21. Compute: gap = random number between 0 and 1
22. $Synthetic[newindex][attr] = Sample[i][attr] + gap * dif$
23. **endfor**
24. $newindex{+}{+}$
25. $N = N - 1$
26. **endwhile**
27. **return** (* *End of Populate.* *)
End of Pseudo-Code.

Figure 3.11 Algorithm of SMOTE [126].

According to their study, the method successfully makes the determination area of minority class to happen more common. Also, it combines oversampling the minority (abnormal) class and under sampling the majority (normal) class to do improved classifier execution [128]. Therefore, SMOTE was used to increase the data set over Weka which is data mining software. The minority (trisomy 21) class was over-sampled at 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, 1100%, 1200%, 1300%, 1400%, 1500%, and 2000% of its original size. Accuracy was calculated by the area under the ROC curve. SMOTE percentages and ROC are values are as shown in Table 3.1.

Table 3.1 Percentages of SMOTE and ROC are values

Percentages of SMOTE	ROC Area Value	Percentages of SMOTE	ROC Area Value
0	0.207	*900**	*0.897*
100	0.424	1000	0.802
200	0.634	1100	0.838
300	0.765	1200	0.870
400	0.842	1300	0.848
500	0.783	1400	0.886
600	0.799	1500	0.843
700	0.825	2000	0.882
800	0.842		
** The best percentage is given by ROC Area value*			

An area of 1.0 represents a great accuracy, for example, the ROC curve transfers towards the left and top limits of the ROC chart [128]. The best ROC Area value was provided with 76 negatives and 50 positive patients. Totally there were 126 patient data in the thesis. Column chart of balanced data set is shown in Figure 3.12. Also, statistical information about balanced data set of patients is as presented in APPENDIX C.

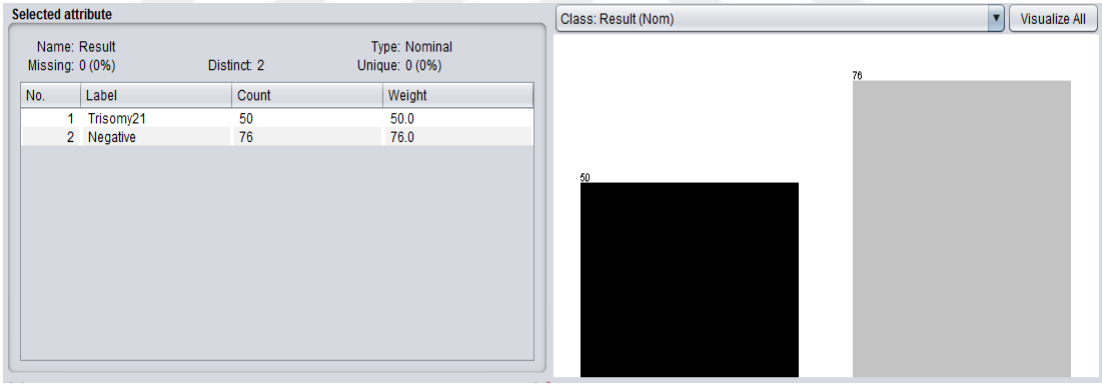


Figure 3.12 Balanced data set

3.5.3 Relations between Variables in the Balanced Data Set

Correlation coefficients were analyzed to observe a relationship between independent variables of the 126 acquired pregnancies. A correlation coefficient is a number between –1 and 1 which decides whether two combined sets of data are associated linearly. It becomes a positive linear correlation and more "confident" relation as an approach to 1. On the other side, when the approach to -1, it has a negative linear correlation. Also, if it closes to zero then there is no evidence about

relations [129]. Figure 3.13 shows correlations examples with some sample variables.

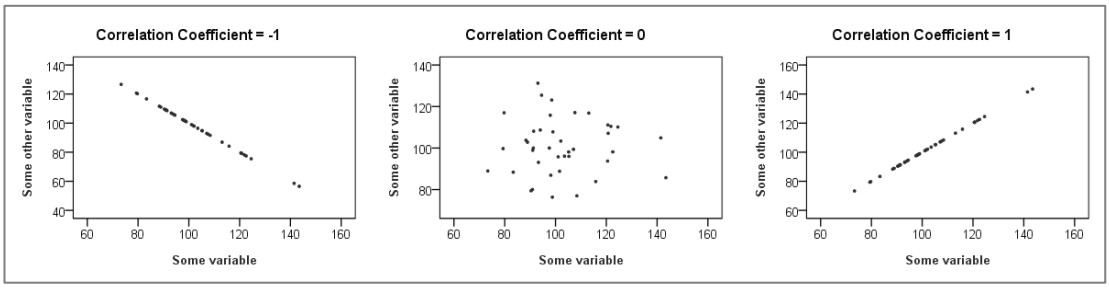


Figure 3.13 Correlation coefficient relations

The confidence in an association is properly decided not only by the correlation coefficient but also by the number of couples in the dataset. The coefficient should be very close to 1 or -1 in order to be accepted as "*statistically significant*" if there are very few couples, but if there are a huge number of couples, a coefficient close to 0 may yet be measured "*highly significant*". A normal technique used by mathematicians to calculate the "*significance"* of the experimental analysis is the "*p value*". p value represents probability by taking a number between 0 and 1. Statisticians say that a p-value of 0.01 is "*highly significant"* or say that "*the data is significant at the 0.01 level*" [129].

The small significance level "*0.01*" was chosen and any possible significant difference in variables was tried to ensure. In addition, Logistic Regression was used to look at the relationship between variables because the dependent variable was a categorical variable [130]. As a result of analyzes, when looking into the relationship between the variables, only AFP, uE3, and hCG markers ensured significance level and other markers such as maternal age, weight, etc. were eliminated from the data set.

3.6 Algorithms Used in Classifying Data

Many classification algorithms have been tried in this thesis. These algorithms were fundamental and well-known algorithms of data mining. The summary description of the classifiers is as follows;

3.6.1 ZeroR

It is the easiest classification technique. It trusts the goal and it does not take into account all estimators. It easily forecasts the majority group. Though there isn't liability control in ZeroR, it is practical to determine a reference point performance like a standard with other classification techniques. The logic of ZeroR is pretty simple. It looks at the ratio between the results in the train data set, and the result in the most adjacent is used as the predictor in the next data. In other words, the accuracy of the ZeroR algorithm is calculated dividing the true positives to all positives. Figure 3.14 shows the model evaluation of ZeroR. The confusion matrix shows that ZeroR only predicts the majority class correctly.

Confusion Matrix		Play Golf			
		Yes	No		
ZeroR	Yes	9	5	*Positive Predictive Value*	0.64
	No	0	0	*Negative Predictive Value*	0.00
		Sensitivity	*Specificity*	**Accuracy** = 0.64	
		1.00	0.00		

Figure 3.14 Model Evaluation of ZeroR

3.6.2 The k-Nearest Neighbors (k-NN)

The K-nearest neighbors observe all existing events and classify new events according to similarities. k-NN was found by Cover and Hart in statistical approximation and model identification even now in the beginning of 1967's [131]. It is used for classification and regression. In both situations, the input contains k-nearest training samples in a characteristic field. The output changes due to whether k-NN is used for classification or regression [132]. "k" represents odd numbers since there may be equality in even numbers for the nearest neighbors. The nearest neighbor is examined as many as k when a new member needs to be classified. The distance between the new member and its neighbors are taken into consideration. There are three different distance calculation functions. These are the Euclidean, Manhattan and Minkowski distances. Their formulas are as given in Equation 3.1, 3.2 and 3.3. After the nearest neighbors are identified, a new member is assigned to the nearest neighbors' class.

Euclidean Distance Function = $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ (3.1)

Manhattan Distance Function = $\sum_{i=1}^{k}|x_i - y_i|$ (3.2)

Minkowski Distance Function = $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$ (3.3)

3.6.3 OneR

It is an easy and precise classification algorithm and it composes of one rule for any estimators in a dataset and it chooses the rule with the smallest total error [133]. To generate a rule for an estimator, a frequency table for each estimator in contrast to the target is built. OneR algorithm is presented as Figure 3.15.

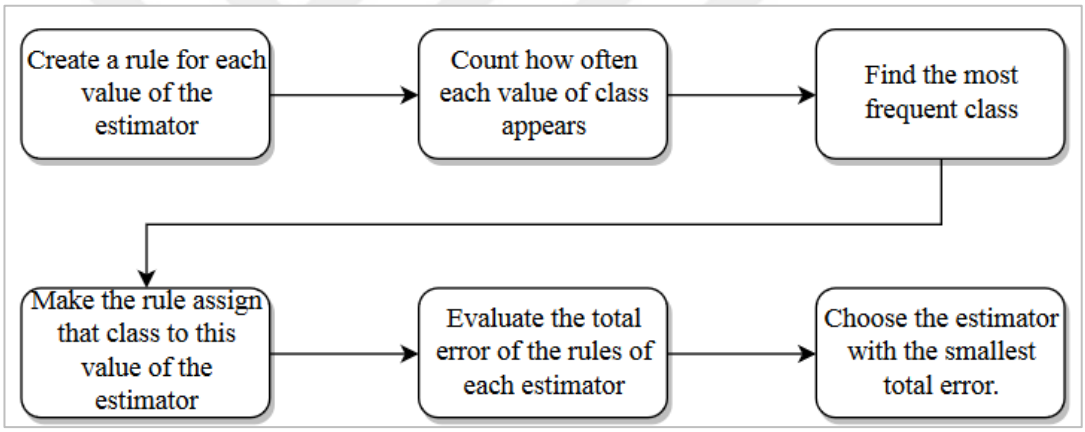


Figure 3.15 OneR Algorithm

3.6.4 Bayesian Network

Judea Pearl was one of the pioneers of Bayesian networks [134]. It is one of the methods used to express data modeling and state transition in computer science. In the literature, the characteristics of networks, which are also known as Bayesian network or belief network, are statistical networks, and the transition edges between nodes are chosen as stated by statistical choices. Bayesian networks are guided acyclic graphs (DAGs), and each node represents the separate variable. Additionally, the gathering of the variables may be indicated by Bayesian networks. A broader form of Bayesian networks is uncertain decision trees.

Nodes in Bayesian networks define variables, and connections define the relationship between nodes. Dependencies are quantified by the probabilities of the conditions given to the parents on the network. The network supports the calculation of the probabilities of the subset of variables which are given evidence for any subset. Pearl explained a general Bayesian network in his study [135]. Figure 3.16 shows the causal associations between the seasons of a year, whether it's raining, whether the fountain is on, whether the sidewalk is wet, and whether the sidewalk is slippery and these states are expressed as X_1, X_2, X_3, X_4, and X_5. For example, the nonappearance of the direct relationship between X_1 and X_5 indicates that the season is not a direct effect on slipperiness and that the effect is due to the wetness of the pavement.

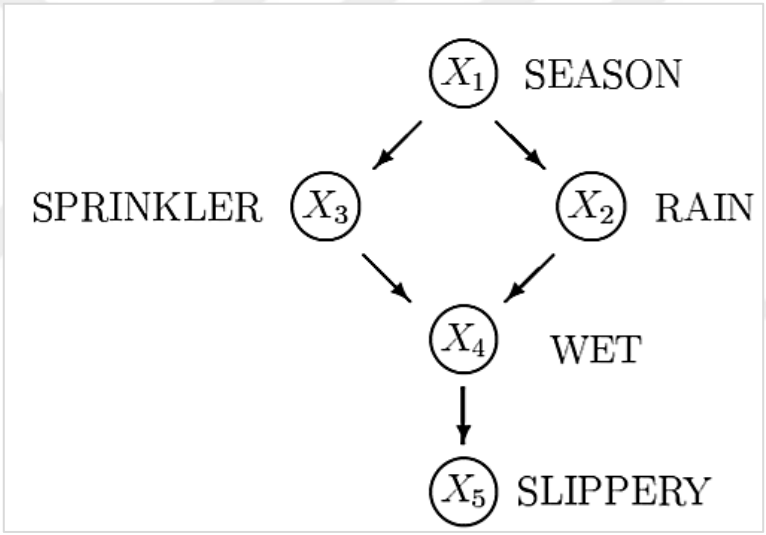


Figure 3.16 Bayesian network representing casual influences among five variables[135].

Arrows on the chart represent true causal connections, and they aren’t a flow of knowledge through the causing. The causing procedures may work on Bayesian networks by spreading the knowledge in any route. For instance, if fountain system is on, the sidewalk is probably wet (estimated). Also, if somebody slides on the sidewalk, it gives evidence which the sidewalk is wet. Conversely, if it is seen which sidewalk is wet, it is possible that rain system is open or it is raining; however if it is observed that the fountain system is open, it decreases a chance of rainfall. It is particularly problematic to demonstrate naturally on rule based organizations and neural networks.

3.6.5 Naïve Bayes

This classifier gets its name from famous mathematician Thomas Bayes who lived in the 17th century. Naïve Bayesian classifier is a simplified version of the Bayesian theory with suggesting independence [136]. The Bayes theorem is as shown in Equation 3.4. The Naïve Bayes classification target to determine a class or a category of the dataset which is shown to a structure by a series of computations described in compliance with likelihood values.

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.4}$$

- P(A|B) : The probability of event A occurring when event B occurs.
- P(B|A) : The probability of event B occurring when event A occurs.
- P(A) and P(B) : The posterior probabilities of events A and B.

Learning data set at a certain rate is given to the system in Naïve Bayes classification. The Naïve Bayes theorem [137] is as shown in Equation 3.5 and 3.6. There must be a class of data presented for learning. The probability tests on the learning data and the new test data that are shown to a structure are performed in compliance with a previously obtained likelihood values and it is tried to decide which group of training dataset is presented. If there is a big number of the learned dataset, it may be so accurate for detecting the true category of the test data.

Naïve Bayes Theorem:

$$P(c|X_1, X_2, \dots, X_n) = \frac{P(c)P(X_1,X_2,\dots,X_n|c)}{P(X_1,X_2,\dots,X_n)} \tag{3.5}$$

$$P(c|X) = P(X_1|c) \times P(X_2|c) \times \dots\dots \times P(X_n|c) \times P(c) \tag{3.6}$$

- P(c|X) is the posterior probability of class given predictor (attribute).
- P(c) is the prior likelihood of class.
- P(X|c) is the probability that is the likelihood of estimator given class.

- P(X) is the prior probability of estimator.

3.6.6 C4.5

Ross Quinlan proposed this algorithm again in 1993 to handle the restrictions of the Iterative Dichotomiser 3 (ID3) algorithm which is discussed earlier [138]. A restriction of ID3 is extremely sensitive to features with large numbers of values. The sensitivity of ID3 to features with many values is indicated by citizenship identification numbers. Since the numbers of citizenship identification numbers are specific to each individual, testing a value will continuously give a small conditional entropy value. Nevertheless, it isn't a functional test. C4.5 uses "Information gain" to overcome that problem. The calculation creates nothing new. Nonetheless, it permits measuring a gain ratio. The gain ratio is defined as in Equation 3.7.

$$Gain\ Ratio(p,t) = \frac{Gain(p,t)}{SplitInfo(p,t)} \tag{3.7}$$

where *SplitInfo* is;

$$SplitInfo(p,test) = \sum_{j=1}^{n} p'\left(\frac{j}{p}\right) \times \log\left(p'\left(\frac{j}{p}\right)\right) \tag{3.8}$$

"*P' x (j/p)*" is the ratio of elements shows at a location p that is getting the value of the j^{th} test. Consider the dissimilar entropy; an above description is independent of the distribution of samples within dissimilar classes. As ID3, the dataset is arranged on each node of a tree to decide the finest discrimination characteristic. Gain ratio impurity technique is used to assess the discrimination qualities. As in ID3, decision trees were constructed using training data or datasets in C4.5 [139]. At each node of the tree, C4.5 chooses a data attribute that best divides its sample set into subgroups which are developed in a class. The criterion is the normalized information gain resulting from selecting an attribute to divide the data. The attribute with the highest normalized knowledge gain is selected to make a decision.

3.6.7 Fisher Linear Discriminant Analysis (FLDA)

It is a classification technique that is improved by R. A. Fisher in 1936 [140]. Although the method is simple, it produces good results in complex problems. FLDA is based on the search for a linear combination that greatest divides the

variables between the 2 classes. The function of Fisher score [141] is as shown in Equation 3.9.

$$S(\beta) = \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \tag{3.9}$$

When considering the score function, it is to estimate the linear coefficients that yield the problem solvable score to the maximum with the Equations 3.10 and 3.11.

Model coefficients: $\beta = C^{-1}(\mu_1 - \mu_2)$ (3.10)

Pooled covariance matrix: $C = \frac{1}{n_1+n_2}(n_1 C_1 + n_1 C_2)$ (3.11)

where;

β: Linear model coefficients

C_1, C_2: Covariance matrices

μ_1, μ_2: Mean vectors

The best way to determine discrimination is to calculate the Mahalanobis distance, which is as shown in Equation 3.12, between 2 groups. A distance bigger than 3 between the groups means that the two averages are more different than the 3 standard deviations. This means that the overlap is very little.

$$\Delta^2 = \beta^T(\mu_1 - \mu_2) \tag{3.12}$$

Δ: Mahalanobis distance between two groups

Finally, if the condition in Equation 3.13 is satisfied, the new incoming point is classified as "*c1*".

$$\beta^T \left(x - \left(\frac{\mu_1+\mu_2}{2} \right) \right) > \log \frac{p(c_1)}{p(c_2)} \tag{3.13}$$

β^T : Coefficients vector

x : Data vector

$\left(\frac{\mu_1+\mu_2}{2}\right)$: Mean vector

$\log\frac{p(c_1)}{p(c_2)}$: Class probability

3.6.8 Logistic Regression

Logistic regression estimates a possible output which may just have two values. Forecasting is based on using one or more forecasts. Linear regression [142] isn't suitable to estimate the value of a binary variable due to 2 causes. First one is that a linear regression will estimate values outdoor the suitable limit. Next, because the dichotomous trials may barely be carried out with two probable values for every trial, the residuals won't be normally distributed along the estimated edge [143]. Alternatively, the logistic regression creates a logistic arc that is restricted to values between zero and one. The logistic regression is same to the linear regression, except that the arc is structured with using the normal algorithm of the odds of the goal variable instead of likelihood. Furthermore, the estimators don't need to be normally distributed or evenly distributed in every group. Figure 3.17 shows differences between linear and logistic models. The constant (b_0) changes the arc left and right and the slope (b_1) describes the steepness of the arc curve in the logistic regression.

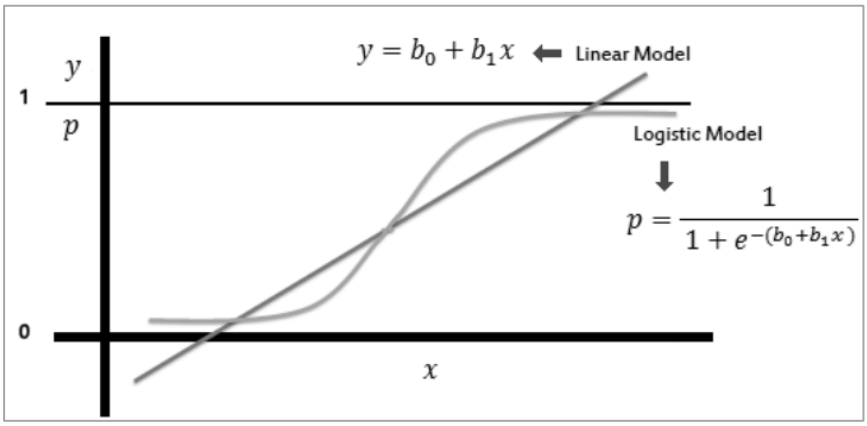


Figure 3.17 Logistic and Linear Models [144].

The logistic regression equation, which is as shown in Equation 3.14, can be written as a probability ratio.

$$p = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots+b_px_p)}} \tag{3.14}$$

There are a few analogies between linear regression and logistic regression. The logistic regression also uses the maximum likelihood estimation to get the equation coefficients which link the estimators to the target, for example, the technique used to predict the most suitable line coefficients in the linear regression is the regression of ordinary least squares [145]. MLE is as represented in Equation 3.15. After the first function is predicted, the procedure is recurred until Log Likelihood (LL) doesn't modify notably.

$$\beta^1 = \beta^0 + [X^T W X]^{-1}.X^T(\gamma - \mu) \tag{3.15}$$

β: Vector of the logistic regression coefficients.

W: Square matrix of order N with elements $n_i\pi_i(1-\pi_i)$ on the diagonal and zeros everywhere else.

μ: A vector of length N with elements $\mu_i = n_i\pi_i$

3.6.9 Multilayer Perceptron (MLP)

The perceptron was first used in the visual perception model (retina) [146]. Although Single Layer Perceptron (SLP) is extremely limited, it is one of the oldest neural networks. Perceptron produces a single output from input into the nerve cell. A single layer perceptron (SLP) is a feed-forward network based on a threshold transfer function. SLP, which is as shown in Figure 3.18, is the easiest kind of artificial neural networks and may just classify linear discrete problems with a binary target (1, 0). Examples of these problems are "AND, OR, NOT" states.

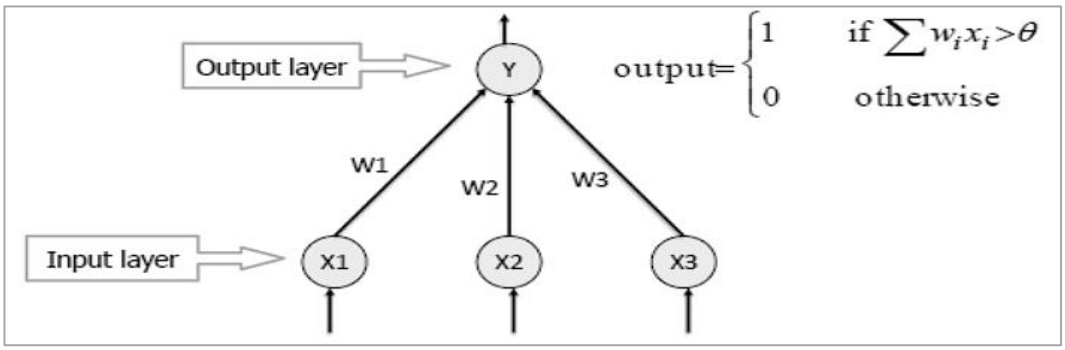


Figure 3.18 Single Layer Perceptron (SLP) [147].

SLP does not already have the knowledge, so the initial weights are randomly assigned. Single layer perceptron collects every the weighted entry and if the total is overhead the cut off level, the SLP is said to be active (output=1) as represented in Figure 3.19.

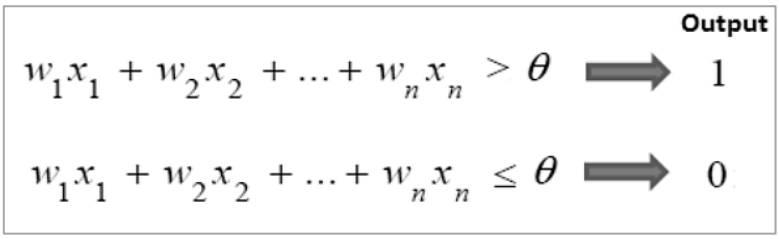


Figure 3.19 SLP Algorithm

Also, the input values are given to the perceptron. The performance is measured acceptable and no revolutions are completed in weight if the estimated result is similar to the expected result. Nevertheless, the weights have to be modified to decrease the error if the output doesn't complement the expected output. Perceptron weight adjustment is as shown in Equation 3.16.

$$\Delta w = \eta \times d \times x \tag{3.16}$$

d: Predicted output

η: Learning rate, usually between 0 and 1

x: Input data

The most well-known instance of an inadequacy of the sensor for solving difficulties with nonlinear no separation cases is the XOR (exclusive or) problem. But, MLP, which uses the back propagation algorithm, may satisfactorily classify the XOR data. Multilayer perceptron has the similar construction as a single layer perceptron with one or more hidden layers that are as shown in Figure 3.20.

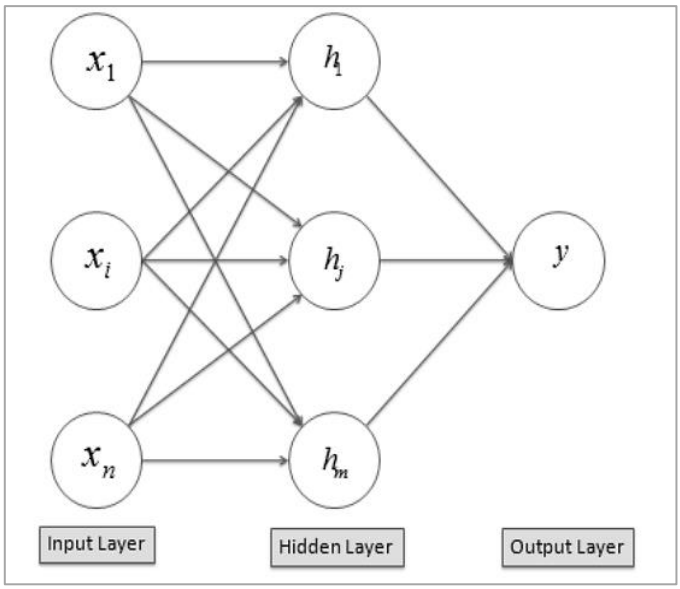


Figure 3.20 Multilayer Perceptron [147].

The back propagation algorithm contains 2 steps. The forward phase (Figure 3.21) in which the activations are forwarded from the input to the output stage and backward to change the weight and bias values of the error between the real and the desired nominal value observed in the output stage.

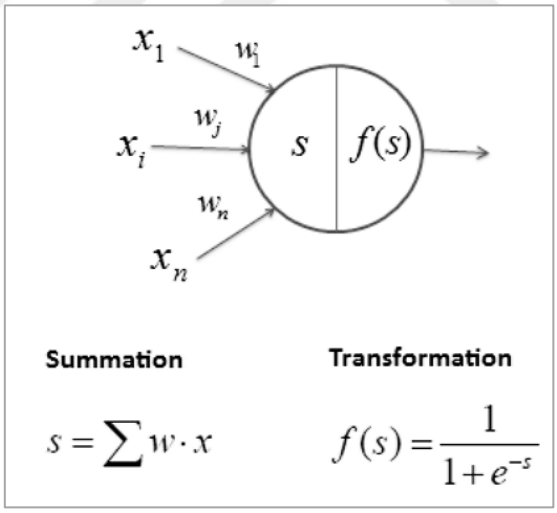


Figure 3.21 Forward phase in MLP

Spread the inputs by enhancing every the weighted input and then compute the outputs using the sigmoid cut off level in forwarding phase. On the other side, in backward phase, spreads the errors backward by sharing out them to all elements in accordance with the quantity of this error the element is liable for [148]. Error in any output neuron is as defined in Equation 3.17 and error in any hidden neuron is as

identified in Equation 3.18. Also, the equation of change in the weights is same in SLP.

$$d_o = y \times (1 - y) \times (t - y) \tag{3.17}$$

$$d_i = y_i \times (1 - y_i) \times (w_i - d_o) \tag{3.18}$$

3.6.10 Sequential Minimal Optimization (SMO)

SMO is an optimization algorithm used to train a Support Vector Machine (SVM) on a data set. The SVM makes classification by discovering the hyperplane which increases the margin between the two classes [149]. The vectors (cases), which are shown in Figure 3.22, describe the hyperplane are the support vectors. It is possible to separate two groups by drawing a border between the two groups in a plane for classification. The place where this border can be drawn is that the two groups should be the farthest from their members. SVM determines how this vector or line is drawn.

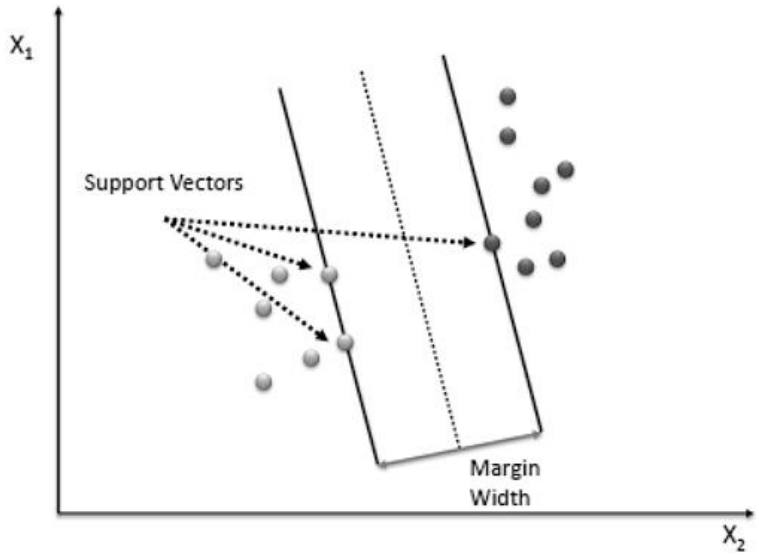


Figure 3.22 Support Vectors

SVM algorithm has three steps; Define an optimal hyperplane (maximize margin), spread the overhead description for the nonlinearly separable problem, and map the data to the high dimensional area where it is simpler for classifying with linear

decision surfaces. To describe the ideal hyperplane, it is necessary to maximize the width (w) of the margin. Figure 3.23 shows the width of the margin (w) and its equation.

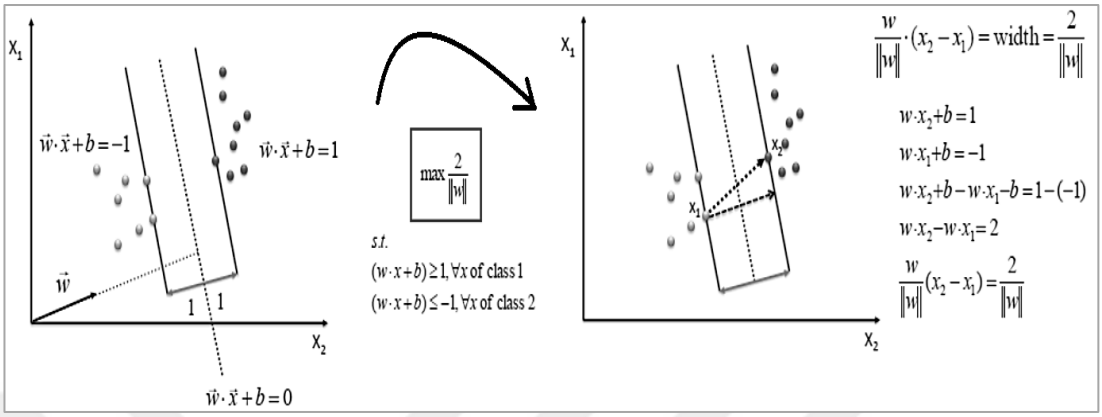


Figure 3.23 Width of the margin (w)

Also, "*w*" and "*b*" is found by solving the objective function, as shown in Equation 3.19, with using Quadratic Programming (QP). A solution of the QP problems is hard and it takes a long time. The SMO may rapidly answer the SVM QP problems without any additional matrix storage and without using numerical QP optimization stages at all [150].

$$min\frac{1}{2}\|w\|^2 \tag{3.19}$$

s.t. $y_i(w\cdot x_i+b)\geq 1, \forall x_i$

Two sets of data can be extracted from each other in three ways. They are a straight line, a flat plane, and hyperplane but, there are states in which a nonlinear area may distribute datasets more effectively. SMO processes use a kernel function, non-linear, to match the data to a dissimilar area where a hyperplane (linear) cannot be used to differentiate. It means that a nonlinear function is learned by a linear learning machine in the area of a high dimensional feature when the capability of the system is measured by a parameter which isn't dependent on the dimensionality of the field. This is named the *"kernel trick"* [151] that is as shown in Figure 3.24; it transforms the data into a higher dimensional feature space to allow the kernel for performing the linear division of the data.

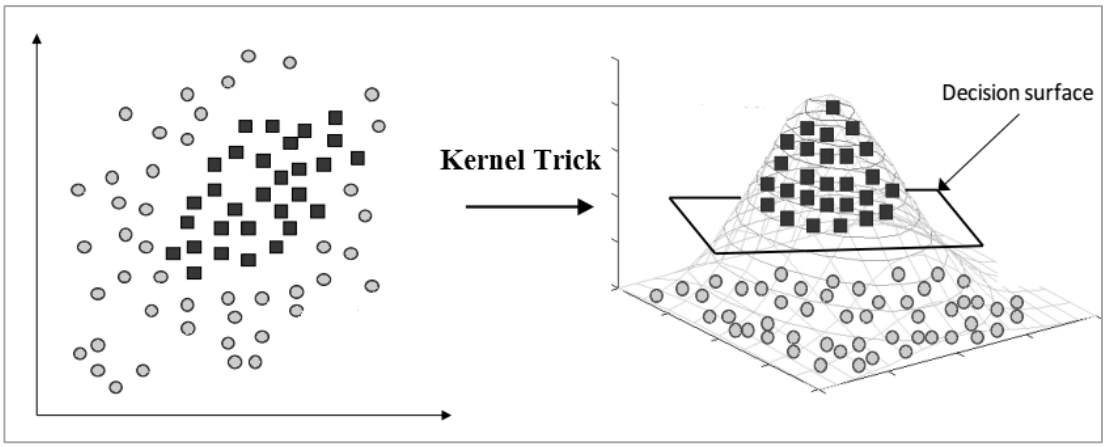


Figure 3.24 Kernel Trick

SMO able to use various kernel functions for determining the support vectors. These are the polynomial kernel, normalized polynomial kernel, Pearson VII universal function kernel (PUK), and radial basis function kernel.

Polynomial Kernel

The polynomial kernel is the continuous nucleus. The polynomial kernel is finely adapted for difficulties where all training data are normalized. Equation 3.20 shows polynomial kernel function. Adjustable parameters are the slope α, the constant term "*c"* and the polynomial degree "d". Where x and y are vectors in the input field, that is the vectors of the features computed from training or test samples, and $c \geq 0$ is a free parameter that loses the effect of high order terms in the polynomial. When c = 0, the kernel is named homogeneous.

$$K(x, y) = (\alpha x^T y + c)^d \tag{3.20}$$

Normalized Polynomial Kernel

It is as defined in Equation 3.21 by normalizing the norm in high dimensional feature space. The kernel value is between 0 and 1 by normalizing an output of standard polynomial kernel [152].

$$K(x, y) = \frac{(1+x^T y)^d}{\sqrt{(1+x^T x)^d (1+y^T y)^d}} \tag{3.21}$$

Pearson VII Universal Function Kernel (PUK)

The Pearson VII function kernel [153] for two vectors is as given in Equation 3.22

$$K(x_i, x_j) = \frac{1}{\left[1+\left(\left(2\sqrt{|x_i-x_j|^2\sqrt{2^{(1/\omega)}-1}}\right)/\sigma\right)^2\right]^{\omega}} \quad (3.22)$$

where;

xi and x_j are two vector arguments,

σ and ω control the half-width and the tailing factor of the peak.

Radial Basis Function Kernel

The Radial basis function (RBF) kernel is a kernel which is in the form of a radial basis function. The RBF kernel is as defined in Equation 3.23 [154].

$$K(x, x') = exp[-\gamma\|x - x'\|^2] \quad (3.23)$$

where;

γ is a parameter that sets the "spread" of the kernel,

σ is a free parameter,

$\gamma = 1/2\sigma^2$,

$\|x - x'\|^2$: may be known like the squared Euclidean distance between the two feature vectors.

CHAPTER 4

RESULTS

The results of the SMOTE over-sampling minority class according to ROC Area are shown in Table 4.1.

Table 4.1 Percentages of SMOTE and patient numbers with ROC area values

Percentages of SMOTE	ROC Area Value	# of Unaffected	#of Down
0	0.207	76	5
100	0.424	76	10
200	0.634	76	15
300	0.765	76	20
400	0.842	76	25
500	0.783	76	30
600	0.799	76	35
700	0.825	76	40
800	0.842	76	45
*900**	*0.897*	*76*	*50*
1000	0.802	76	55
1100	0.838	76	60
1200	0.870	76	65
1300	0.848	76	70
1400	0.886	76	75
1500	0.843	76	80
2000	0.882	76	105

* The best percentage is given by ROC Area value

The ROC area values were examined in the modified data set and minority class (negative result pregnancies) was oversampled by certain coefficients. The best ROC Area value was obtained with 76 negatives and 50 positive pregnancies with a value of ROC Area value 0.897 at 900% coefficient. Figure 4.1 shows the best ROC Area graph which is at 900% rate.

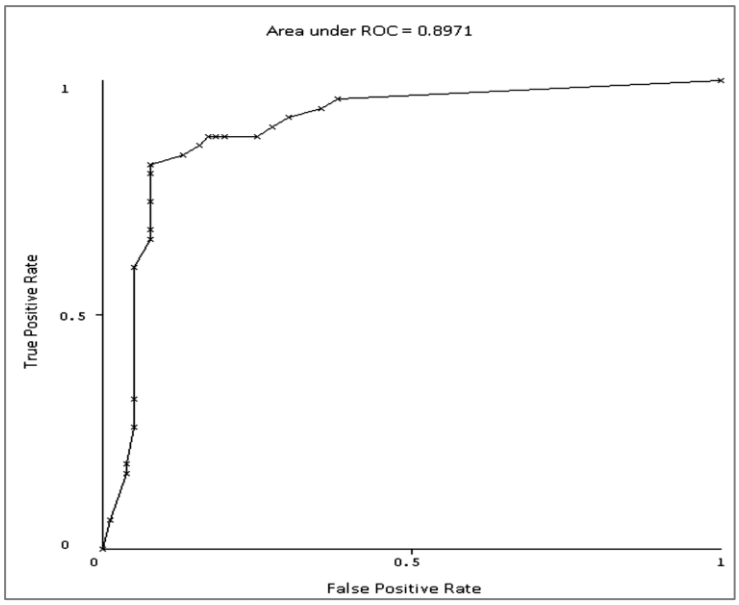


Figure 4.1 The best ROC Area graph with 900% coefficient of SMOTE

The relationship between variables was examined as shown in Table 4.2. Logistic regression analysis was used for relationships on SPSS statistical software because there was a categorical (nominal) dependent variable and Logistic regression method has the capability to calculate the relations between nominal variables and numeric variables. Variables that provide $p<0.01$ significance level subjected to the thesis and other variables eliminated from the study. As a result, it was found acceptable to use AFP, hCG and uE3 variables for analysis. The statistical information about eliminated data set for the thesis is given in APPENDIX C.

Table 4.2 The results of Logistic Regression analysis

		Score	df	Significance
Variables	Maternal Age	3,515	1	,061
	Weight	2,814	1	,093
	Gestational Age	5,626	1	,018
	AFP*	7,586	1	,006
	uE3*	16,709	1	,000
	hCG*	19,003	1	,000
*Variables that provide $p<0.01$ inequality subject to the thesis.				

The modified data set had been run on experimenter mode of Weka software via 10 fold cross validation technique with classification algorithms which are zeroR, k-NN, OneR, Bayesian Network, Naive Bayes, J48(C4.5), FLDA, Logistic Regression, Multilayer Perceptron and SMO respectively. All classification algorithms were run only a time one by one.

Weka software has two options to verify the test. These are percentage split and k-fold cross-validation techniques. Percentage split is usually used in datasets with a big number of samples. Because the percentage split divides the data set into the two parts according to the given percentages and one part becomes the training set while the remaining part becomes the test set. For example, if a selected percentage is 66%, then the remaining part, 33%, will be used as the test set, and 66% of the data set will be used for training and test set result will be the final result of the classifier test set. If the number of instances in the data set is very large, dividing the certain parts (training and test parts) of the data set does not affect results of the test. However, if the sample size is small, using most of the samples for training part and testing part at the same time are more reliable. In such a case, using k-fold cross-validation technique will give more reliable results [155]. Because this technique splits the dataset into k subsets that consist randomly selected samples and *k-1* subsets are used for training and 1subset are used for testing. Then the chosen subset for the test is changed by another and the process is repeated. This process is performed *k* times in total and the average of the test results is calculated. So, more valid and reliable results are obtained in the small data set. Table 4.3 shows the k-fold cross-validation technique, training, and test subsets in *k* steps. When determining the value of *k*, some values such as mean absolute error and F-measure are taken into account. The *k* value, which is used commonly in the literature, is 10 but it may vary according to the number of samples in the dataset. As the number of *k* increases, the number of samples that is divided into subsets will decrease and the number of training subsets will increase, so it may affect the success rate and accuracy in the wrong direction due to the decrease in the number of samples to be tested. Also, if the number of *k* decreases, the number of samples, which is divided into subsets increases and the number of training subsets decreases. Because of this situation, the performance of the algorithm that is generated by the restricted training set may be affected wrongly.

Table 4.3 k-fold cross-validation technique (k=10)

k=10	1. Subset	2. Subset	3. Subset	4. Subset	5. Subset	6. Subset	7. Subset	8. Subset	9. Subset	10. Subset
1. Step	Test	Training	Training	Training	Training	Training	Training	Training	Training	Training
2. Step	Training	Test	Training	Training	Training	Training	Training	Training	Training	Training
3. Step	Training	Training	Test	Training	Training	Training	Training	Training	Training	Training
4. Step	Training	Training	Training	Test	Training	Training	Training	Training	Training	Training
5. Step	Training	Training	Training	Training	Test	Training	Training	Training	Training	Training
6. Step	Training	Training	Training	Training	Training	Test	Training	Training	Training	Training
7. Step	Training	Training	Training	Training	Training	Training	Test	Training	Training	Training
8. Step	Training	Training	Training	Training	Training	Training	Training	Test	Training	Training
9. Step	Training	Training	Training	Training	Training	Training	Training	Training	Test	Training
10. Step	Training	Training	Training	Training	Training	Training	Training	Training	Training	Test

Therefore, the *k* value should be determined correctly by taking into account various experiments and the number of samples in the dataset. When criteria, which are mean absolute error and F-measure, are taken into consideration, *k=10* is found to be correct. In the data set of the study, the *k* parameter was tried with the Bayes network classifier at values 9, 10, 11 and 15 and the smallest mean absolute error with the largest F-measure value was found at the *k = 10,* which is as shown in Table 4.4.

Table 4.4 Determination of the *k* value

		k Value			
		k=9	k=10	k=11	k=15
Bayesian Network	Mean Absolute Error	0.1915	0.1654	0.1779	0.1799
	F-Measure	0.835	0.874	0.859	0.859

Details of the results are shown in figures for each classification algorithm. First of all, ZeroR was applied to the data set. It is the most basic classification algorithm and it simply estimates majority class. Because of that, it is practical for a deciding a baseline performance as a benchmark for other classifiers. The results of the classification are given in Figure 4.2.

The results show that the correctly classifies percentage of other classifiers should be at least 60%. A lower success rate means that the using classifier has failed. Another

parameter passed in the output is a total number of instances 126. This means that 126 patients (instances) were used in this test. Also, ZeroR predicts negative class. Because ZeroR chooses a base class which has more instances then another class. On the other hand, the coefficient of Kappa statistic measures the fit between two observers in the evaluation of categorical items [156]. Kappa can take a value between -1 and +1. A value of -1 indicates that the incompatibility between the two observers is perfect, on the other side, when the value is +1 indicates that the Kappa value is a perfect fit between the two observers [157]. If a kappa value of 0 is found, then the harmony between these two observers shows that there is a purely random association. Kappa value is 0 in the result of the zeroR classifier and it means that association between observers was purely chance. In addition, the output shows the error rates that mean the distinctness between the real and estimated value. In a confusion matrix, a classifier shows which class the instances placed and which class is the actually for the instances.

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     hastalar-weka.filters.unsupervised.attribute.NumericToBinary-R15,18,20-weka.filters.unsupe
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: Negative

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          76               60.3175 %
Kappa statistic                          0
Mean absolute error                      0.4791
Root mean squared error                  0.4893
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,000    0,000    0,000      0,000    0,000      0,000    0,484     0,389     Trisomy21
                 1,000    1,000    0,603      1,000    0,752      0,000    0,484     0,596     Negative
Weighted Avg.    0,603    0,603    0,364      0,603    0,454      0,000    0,484     0,514

=== Confusion Matrix ===

  a  b   <-- classified as
  0 50 |  a = Trisomy21
  0 76 |  b = Negative
```

Figure 4.2 Detailed results of the zeroR classification algorithm for the data set

Next classification algorithm was the k-NN algorithm. Firstly *k* parameter should be determined. Weka helps for determining to *k* parameter with cross validation techniques. When the cross validation was performed for *k*, Weka found best *k* value was 3. Also, *k* parameter could be found with trial and error method or trial and observation method. The *k* parameter was observed at various values according to some criteria as shown in Figure 4.3. The best *k* value was found when *k* = 3. Although some criterion values were the same at k = 1, some criterion values were lower than k = 3. Therefore the value of the parameter *k* was chosen 3.

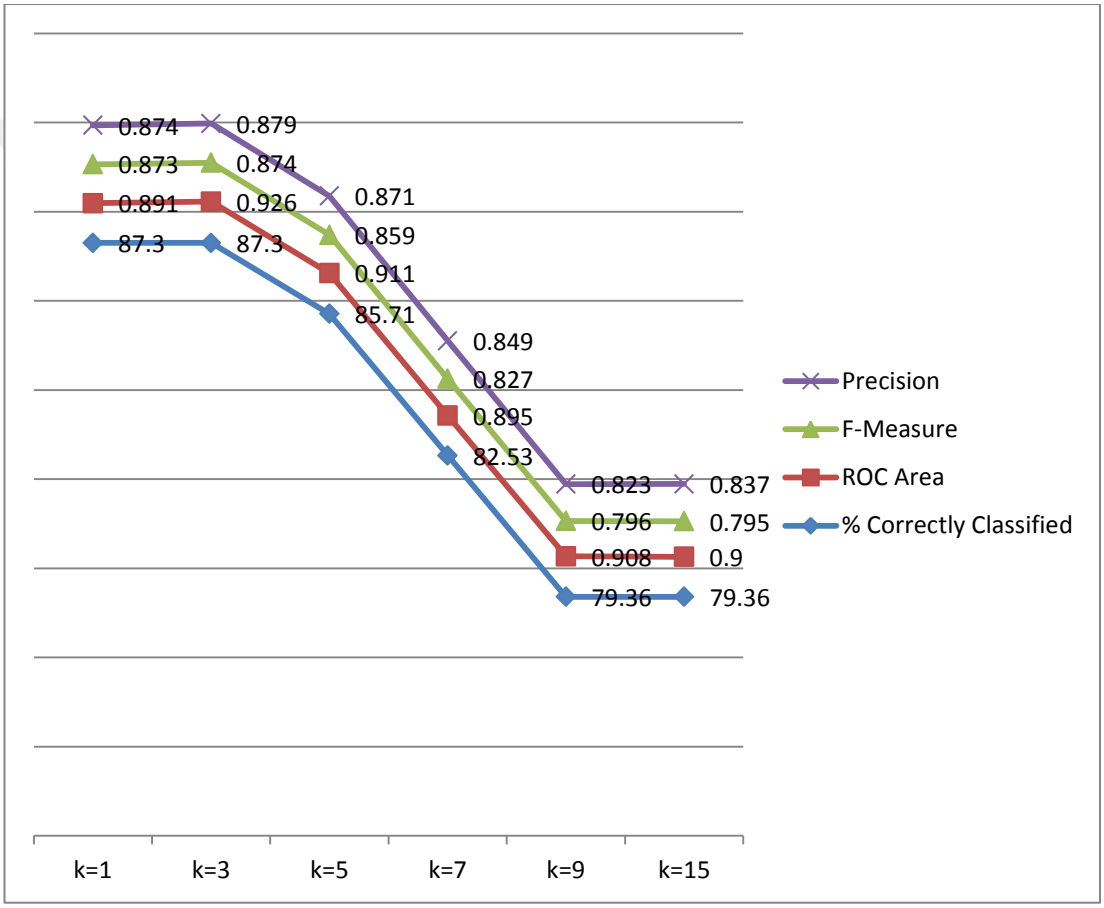


Figure 4.3 Determining *k* parameter for the k-NN classifier

Also, k-NN classifier uses training and classification algorithms which are as shown in Figure 4.4. This algorithm supposes all samples communicate to points in the n-dimensional space $\mathcal{R}^n$ [158].

Training Algorithm:

- For each training instance $\langle x, f(x) \rangle$ which is in training set, add the instance to the training examples list.

Classification Algorithm:

- Given a query instance x_q to be classified,
  - Let $x_1 \ldots x_k$ denote the k instances from training example list that are nearest to x_q
  - Return

$$\hat{f}(x_q) \leftarrow \underbrace{argmax}_{v \in V} \sum_{i=1}^{k} \delta(v, f(x_i))$$

where $\delta(a, b) = 1$ if $a = b$ and where $\delta(a, b) = 0$ otherwise.

Figure 4.4 Algorithms of the k-NN classifier [158].

The output of the k-NN classifier is shown in Figure 4.5. The Kappa statistic value was calculated 0.74 for this classifier. It means that the harmony between the observers was close to perfection. Other parameters, which are notable, are TP and FP ratios. 90% detection rate (TPR) and 14.5% FP rate was found by the k-NN classifier. The aim of the study is to increase the detection rate as much as possible and decrease the FP rate at the same time. When looking at the confusion matrix, 45 of the 50 Trisomy21 samples were estimated correctly and 5 of them were mispredicted. Also, 65 of the 76 negative samples were correctly estimated and 11 of them were mispredicted and 11 mispredictions are called false positives at the same time.

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka
Relation:     hastalar-weka.filters.unsupervised.attribute.NumericToBinary-R15,18,20-weka.filters.unsupe
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         110               87.3016 %
Kappa statistic                          0.7401
Mean absolute error                      0.1607
Root mean squared error                  0.3248
Relative absolute error                 33.5439 %
Root relative squared error             66.3877 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,900    0,145    0,804      0,900    0,849      0,744    0,926     0,860     Trisomy21
                 0,855    0,100    0,929      0,855    0,890      0,744    0,926     0,940     Negative
Weighted Avg.    0,873    0,118    0,879      0,873    0,874      0,744    0,926     0,908

=== Confusion Matrix ===

  a  b   <-- classified as
 45  5 |  a = Trisomy21
 11 65 |  b = Negative
```

Figure 4.5 Detailed results of the k-NN classification algorithm for the data set

The third classification algorithm was OneR algorithm. It is a simple and precise classification algorithm and it composes of one rule for any estimators in the data and it chooses the rule with the smallest total error. Results of the OneR algorithm are as shown in Figure 4.6. Firstly, OneR found the most frequent class which was Negative class with 76 instances. Secondly, it made the rules for each predictor (AFP, hCG, uE3) and calculated the total errors of the rules. The error rate for AFP was 0.2063, for HCG was 0.4524 and for uE3 was 0.3016. So, AFP had the minimum error rate with 0.2063. Then the classifiers selected the AFP predictor and estimated the results. The predictor rule for the AFP is as following;

- If AFP is smaller than 23.33 or bigger/equal than 29.92 then the prediction is Negative.
- If AFP is between 23.33 and 29.92 then the prediction is Trisomy21.

Kappa statistics of OneR shows that compatibility between observers is neither perfect and nor bad. Percentage of correctly classified is %79 and it's better than zeroR classifier. The OneR correctly predicted 34 of the 50 Trisomy21 instances and 66 of the 76 Negative instances. It has not bad result although it is a simple classifier. Also, the F-Measure parameter is important in the results. In this result, F-Measure is 0.791 and when the value of F-Measure increases to 1, it means good, since F-measure is associated with the precision and recall.

```
=== Run information ===

Scheme:       weka.classifiers.rules.OneR -B 6
Relation:     hastalar-weka.filters.unsupervised.attribute.NumericToBinary-R15,18,20-weka.filters.unsupe
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

AFP(IU/ml):
        < 23.331342     -> Negative
        < 29.9203915    -> Trisomy21
        >= 29.9203915   -> Negative
(101/126 instances correct)


Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         100               79.3651 %
Kappa statistic                          0.5599
Mean absolute error                      0.2063
Root mean squared error                  0.4543
Relative absolute error                 43.066  %
Root relative squared error             92.8383 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,680    0,132    0,773      0,680    0,723      0,563    0,774     0,652     Trisomy21
                 0,868    0,320    0,805      0,868    0,835      0,563    0,774     0,778     Negative
Weighted Avg.    0,794    0,245    0,792      0,794    0,791      0,563    0,774     0,728

=== Confusion Matrix ===

  a  b   <-- classified as
 34 16 |  a = Trisomy21
 10 66 |  b = Negative
```

Figure 4.6 Detailed results of the OneR classification algorithm for the data set

Next one was Bayesian Network classification algorithm. It is a probabilistic graphical model that shows a series of random variables and an acyclic graph guided by their conditional dependencies. Figure 4.7 shows detailed results of the Bayesian Network classification algorithm for the data set. There are 3 nodes (AFP, hCG, and uE3) in output. All nodes are associated with Result class. Also, the results show 5 types of local score metrics. LogScore returns the log of the quality of a network. Score-based algorithms constitute a Bayesian network that maximizes the score function, which indicates the correctness of the causality structure for a multivariate data set [159]. The lines, which are in output, list the logarithmic score of the network structure for the network for various methods of scoring. The score type specifies the measure used to evaluate the quality of a network structure. It may be one of Bayes, Bayesian Dirichlet equivalence (BDeu), Minimum Description Length (MDL), Akaike Information Criterion (AIC), and Entropy. In addition, a number of correctly classified instances are 110. The classifier incorrectly predicted 16 instances and all of them are negative class. They classified as Trisomy21, however, they should have been Negative. Also, the classifier predicted correctly all of the Trisomy21 instances. But, it should be run many times such as 100 run for stable and accurate results. On the other hand, FP rate is 21% and it is too much for the aim of the study. Kappa statistic value is nearly same with k-NN classifier kappa statistic and compatibility or harmony between observers are near the perfection.

```
=== Run information ===

Scheme:       weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -
Relation:     hastalar-weka.filters.unsupervised.attribute.NumericToBinary-R15,18,20-weka.filters.unsupe
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Bayes Network Classifier
not using ADTree
#attributes=4 #classindex=3
Network structure (nodes followed by parents)
AFP(IU/ml)(3): Result
uE3(ng/ml)(3): Result
hCG(mlU/ml)(2): Result
Result(2):
LogScore Bayes: -338.07213077574994
LogScore BDeu: -350.7498644005968
LogScore MDL: -356.7322007576922
LogScore ENTROPY: -330.1326502694591
LogScore AIC: -341.1326502694591


Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         110               87.3016 %
Kappa statistic                          0.7485
Mean absolute error                      0.1654
Root mean squared error                  0.2835
Relative absolute error                 34.5134 %
Root relative squared error             57.9458 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                1,000    0,211    0,758      1,000    0,862      0,773    0,928     0,817     Trisomy21
                0,789    0,000    1,000      0,789    0,882      0,773    0,928     0,960     Negative
Weighted Avg.   0,873    0,084    0,904      0,873    0,874      0,773    0,928     0,903

=== Confusion Matrix ===

  a  b   <-- classified as
 50  0 |  a = Trisomy21
 16 60 |  b = Negative
```

Figure 4.7 Detailed results of the Bayesian Network classification algorithm for the data set

Another algorithm was Naive Bayes classifiers that are a family of simple probability classifiers. It is based on the application of Bayesian theory with strong (naive) independence estimates between features. The results are shown in Figure 4.8. The probability of the Trisomy21 class is 0.4 and probability of the Negative

class is 0.6 in the output. Mean, standard deviation and precision of each attribute were calculated by Naïve Bayes classifier. The mean and standard deviation are classical statistical terms. Weight sum means a number of instances for each class. Another term is precision and it is the percentage of positive predictions which are correctly predicted [True Positives / (True Positives + False Positives)]. Also, the precision is the minimum standard deviation permitted for the characteristic in the problem. It is taken by a heuristic executed in Naïve Bayes that calculates a mean of the difference between adjacent values of a characteristic. The classifier estimated 102 of 126 instances correctly and its percentage is 80.92%. Kappa statistic is the average value. Detection rate (TP rate) is high value, but FPR is also high and that is the unwanted situation. A value of 1.0 represents a perfect accuracy as the ROC curve moves towards the left and top boundaries of the ROC graph. Because of that, Naïve Bayes has good accuracy with 0.933 ROC are value. Also as seen in the confusion matrix, the classifier predicted most of the Trisomy21 instances correctly. However, the same thing cannot be said for the Negative class instances, because the third one of the instances in this class was predicted wrongly.

```
=== Classifier model (full training set) ===

Naive Bayes Classifier

                    Class
Attribute       Trisomy21   Negative
                    (0.4)      (0.6)
===================================
AFP(IU/ml)
  mean              24.9684    30.0541
  std. dev.          4.0979    15.2997
  weight sum             50         76
  precision           0.622      0.622

uE3(ng/ml)
  mean               0.6124     1.1517
  std. dev.          0.1342     1.0421
  weight sum             50         76
  precision          0.0456     0.0456

hCG(mlU/ml)
  mean           46779.3216 34391.7802
  std. dev.      15437.1761 21250.3504
  weight sum             50         76
  precision         771.936    771.936



Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         102               80.9524 %
Kappa statistic                          0.6276
Mean absolute error                      0.2033
Root mean squared error                  0.3523
Relative absolute error                 42.424  %
Root relative squared error             72.0042 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,960    0,289    0,686      0,960    0,800      0,660    0,933     0,885     Trisomy21
                 0,711    0,040    0,964      0,711    0,818      0,660    0,933     0,963     Negative
Weighted Avg.    0,810    0,139    0,854      0,810    0,811      0,660    0,933     0,932

=== Confusion Matrix ===

  a  b   <-- classified as
 48  2 |  a = Trisomy21
 22 54 |  b = Negative
```

Figure 4.8 Detailed results of the Naïve Bayes classification test

C4.5 is an extension of Quinlan's previous ID3 algorithm. Decision trees created by C4.5 may be used for classification, so C4.5 is frequently called a statistical classifier. Details of the classifying results with Weka are found in Figure 4.9. A decision tree was created as seen in Figure 4.9. According to the decision tree, if the uE3 value is greater than 0.8, the result is Negative. Otherwise, the value of AFP is

checked. If the AFP value is greater than 30.1 then the result is Negative. If the AFP value is between 30.1 and 24.2, the result is Trisomy21. But if the AFP value is less than or equal to 24.2, the hCG value is checked. If hCG value is greater than 40425, the result is negative, otherwise the result is Trisomy21. Also, a decision tree is as shown in Figure 4.10.

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     hastalar-weka.filters.unsupervised.attribute.NumericToBinary-R15,18,20-weka.filters.unsupe
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

uE3(ng/ml) <= 0.8
|   AFP(IU/ml) <= 30.1
|   |   AFP(IU/ml) <= 24.2
|   |   |   hCG(mlU/ml) <= 40425: Negative (22.0/2.0)
|   |   |   hCG(mlU/ml) > 40425: Trisomy21 (17.0/6.0)
|   |   AFP(IU/ml) > 24.2: Trisomy21 (41.0/4.0)
|   AFP(IU/ml) > 30.1: Negative (13.0)
uE3(ng/ml) > 0.8: Negative (33.0)

Number of Leaves  :     5

Size of the tree :      9


Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         105               83.3333 %
Kappa statistic                          0.6506
Mean absolute error                      0.1946
Root mean squared error                  0.3657
Relative absolute error                 40.6137 %
Root relative squared error             74.7313 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,780    0,132    0,796      0,780    0,788      0,651    0,863     0,717     Trisomy21
                0,868    0,220    0,857      0,868    0,863      0,651    0,863     0,894     Negative
Weighted Avg.   0,833    0,185    0,833      0,833    0,833      0,651    0,862     0,823

=== Confusion Matrix ===

  a  b   <-- classified as
 39 11 |  a = Trisomy21
 10 66 |  b = Negative
```

Figure 4.9 Detailed results of the C4.5 classification algorithm for the data set

There are 5 leaves and 9 trees in the Figure 4.10. Estimates were made according to the decision tree and the number of correctly classified samples was 105. F-Measure and precision are 0.833 and ROC area value is 0.862. However, the detection rate is 0,780 and the detection rate is lower for this classifier when compared to the classifiers which were tested before.

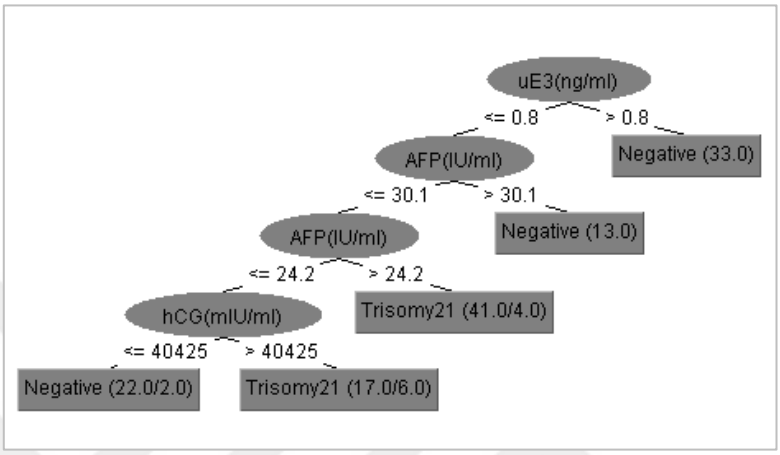


Figure 4.10 Decision tree of C4.5

Next classifier was FLDA. This technique searches for directives on the data with the greatest variance and subsequent project the data onto it. The results are given in Figure 4.11. The discriminant equation should be remembered in order to understand the weights in the output. Linear discriminant analysis yields equilibrium as a linear combination of independent variables which will the greatest distinguishes between clusters in the dependent variable. The linear combination is identified as the discriminant function [160]. The weights allocated to every independent variable are corrected for the associations between all variables. The weights are mentioned to as discriminant coefficients. The discriminant equation is as defined in Equation 4.1.

$$F = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (4.1)$$

where, F is a latent variable made by the linear combination of the dependent variable, X_1, X_2,…, X_p are the p independent variables, ε is the error term and $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are the discriminant coefficients.

Weights of AFP, uE3, and hCG are $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ in the equation and classifier use of them to find specific discriminant function. The results show that the FLDA performs worse than the classifiers that are previously used. Kappa statistic value shows higher dependence on chance and false positive rate is excessive value. In general, FLDA classifier has shown poor performance four the study.

```
=== Run information ===

Scheme:       weka.classifiers.functions.FLDA -R 1.0E-6
Relation:     hastalar-weka.filters.unsupervised.attribute.NumericToBinary-R15,18,20-weka.filters.unsupe
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Fisher's Linear Discriminant Analysis

Threshold: -0.048966373579728106

Weights:

AFP(IU/ml):     -0.04409745408155738
uE3(ng/ml):     -0.9990272328637937
hCG(mlU/ml):     5.039877175946904E-5


Time taken to build model: 0.38 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          86               68.254  %
Kappa statistic                          0.3672
Mean absolute error                      0.3788
Root mean squared error                  0.4458
Relative absolute error                 79.0552 %
Root relative squared error             91.1069 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,740    0,355    0,578      0,740    0,649      0,377    0,773     0,617     Trisomy21
                 0,645    0,260    0,790      0,645    0,710      0,377    0,773     0,862     Negative
Weighted Avg.    0,683    0,298    0,706      0,683    0,686      0,377    0,773     0,765

=== Confusion Matrix ===

  a  b   <-- classified as
 37 13 |  a = Trisomy21
 27 49 |  b = Negative
```

Figure 4.11 Detailed results of the FLDA classification algorithm for the data set

Another classifier was the Logistic Regression. The detailed results are given in Figure 4.12. As mentioned before in Equation 3.14, coefficients for Trisomy21 class refer to b_i. Odds ratios are exponential of the coefficients. For example, $exp(0.0319)$ is equal to 0.9686 and the log of the odds ratio is the value of the coefficient attached to the variable AFP in the logistic regression. So, logistic regression equation can be written as in Equation 4.2.

$$\frac{p}{1-p} = exp\big(f(x)\big) = exp(-0.0319 \times AFP)exp(-1.4476 \times uE3)exp(0 \times hCG) \tag{4.2}$$

If AFP increases for 1 unit, then $exp\big(f(x)\big)$ will be 0.9686 $exp\big(f(x)\big)$. This means that the predicted odds of Trisomy21increases for 0.9686 times when AFP increases for 1 unit. So, when looking the odds ratios in the output, firstly hCG, secondly AFP and lastly uE3 are most favorable to the Trisomy21 output. On the other hand, 87 samples from 126 are correctly classified when look at the number of correctly classified samples. 28 of them are instances of Trisomy21class and the rest of them are instances of Negative class. TP rate is 56% and FP rate is 22%, so these values are lower than the other studies which were presented before. Also, Matthews's correlation coefficient (MCC) is 0.343. The MCC is firstly used in machine learning by biochemist Brian W. Matthews in 1975 [161]. It is often regarded as a balanced criterion that can be used even if the classes are very different in size. In essence, MCC is a correlation coefficient between observed and predicted binary classifications. It returns a value between −1 and +1. The +1 coefficient represents the perfect estimate, 0 is not better than the random estimate, and -1 represents the total disagreement between the estimate and the observation. So MCC of the logistic regression classifier is close to 0 and it means that predictions are attached to the chance.

```
=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...
                  Class
Variable       Trisomy21
=======================
AFP(IU/ml)       -0.0319
uE3(ng/ml)       -1.4476
hCG(mlU/ml)            0
Intercept         0.1467


Odds Ratios...
                  Class
Variable       Trisomy21
=======================
AFP(IU/ml)        0.9686
uE3(ng/ml)        0.2351
hCG(mlU/ml)            1


Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          87               69.0476 %
Kappa statistic                          0.3422
Mean absolute error                      0.3804
Root mean squared error                  0.4409
Relative absolute error                 79.3974 %
Root relative squared error             90.1153 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,560    0,224    0,622      0,560    0,589      0,343    0,759     0,592     Trisomy21
                0,776    0,440    0,728      0,776    0,752      0,343    0,759     0,859     Negative
Weighted Avg.   0,690    0,354    0,686      0,690    0,687      0,343    0,759     0,753

=== Confusion Matrix ===

  a  b   <-- classified as
 28 22 |  a = Trisomy21
 17 59 |  b = Negative
```

Figure 4.12 Detailed results of the Logistic regression algorithm for the data set

MLP (Multilayer Perceptron) is a feed forward artificial neural network model which maps sets of input data onto a set of suitable outputs. MLP involves multiple layers of knots in an oriented graph, with each layer fully attached to the next layer. It was used in the thesis and it had best results between other classifiers. Detail of the results is presented in Figure 4.13. The MLP used 3 input variables that were AFP, hCG, and uE3 to predict classes (Trisomy21 and Negative). The training stage was to adjust the internal weights to get as close as possible to the known classes values.

```
=== Classifier model (full training set) ===

Sigmoid Node 0
    Inputs    Weights
    Threshold    3.7828455624026294
    Node 2    -8.761344323823842
    Node 3    -11.124350387877087
Sigmoid Node 1
    Inputs    Weights
    Threshold    -3.782898504706192
    Node 2    8.761491959006735
    Node 3    11.124519747913117
Sigmoid Node 2
    Inputs    Weights
    Threshold    -17.226796387664635
    Attrib AFP(IU/ml)    -16.294398012932433
    Attrib uE3(ng/ml)    -3.310396915955904
    Attrib hCG(mlU/ml)    -7.59671841765014
Sigmoid Node 3
    Inputs    Weights
    Threshold    16.37299312465918
    Attrib AFP(IU/ml)    15.535909207037049
    Attrib uE3(ng/ml)    12.241964781268676
    Attrib hCG(mlU/ml)    0.27209562455665126
Class Trisomy21
    Input
    Node 0
Class Negative
    Input
    Node 1


Time taken to build model: 0.47 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         112               88.8889 %
Kappa statistic                          0.7726
Mean absolute error                      0.1534
Root mean squared error                  0.2991
Relative absolute error                 32.0168 %
Root relative squared error             61.1295 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,920    0,132    0,821      0,920    0,868      0,776    0,910     0,764     Trisomy21
                 0,868    0,080    0,943      0,868    0,904      0,776    0,910     0,949     Negative
Weighted Avg.    0,889    0,100    0,895      0,889    0,890      0,776    0,910     0,876
```

Figure 4.13 Detailed results of the Multilayer Perceptron classification algorithm for the data set

In the output, there are 4 sigmoid nodes. Node 0 and node 1 are output nodes and node 2 and node 3 hidden nodes which are attribute nodes. The number of nodes or hidden layers was defined as (number of attributes+ number of class) / 2. This equation was called wildcard values on Weka. There are some wildcard values on Weka and they are "*a"* = (attributes + classes) / 2, "*i"* = attributes, "*o"* = classes, "*t"* = attributes + classes. *"a"* was selected in the study and number of nodes was

calculated as 2. Figure 4.14 represents a neural network of MLP classifier for the study. All nodes have weight values of attributes and all the values given are interconnection weights between hidden and output nodes.

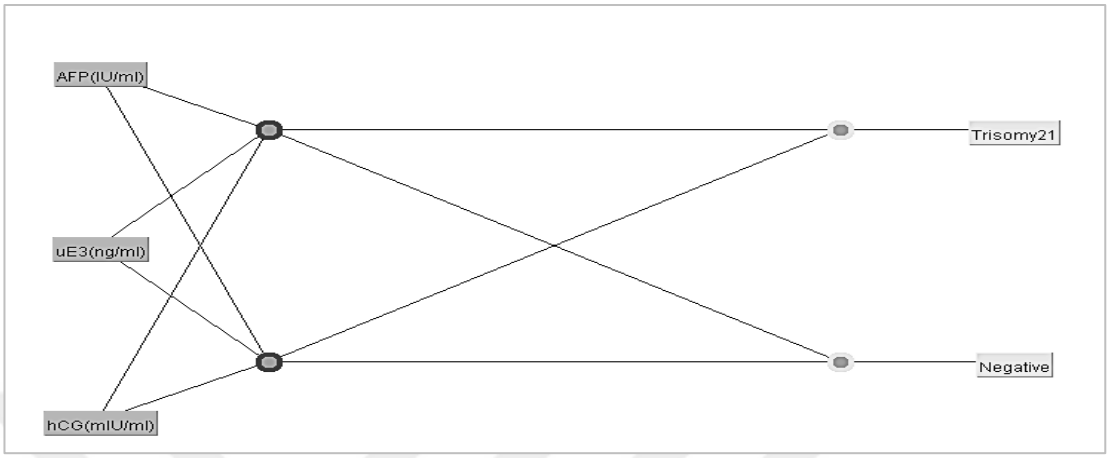


Figure 4.14 Neural network of the MLP classifier test

As a result, classified correctly instances are 112 and it has high percentage rate with 88.89%. Alternatively, kappa statistic value is near to perfect harmony between observers. The high value of precision-recall curve area (PRC) and ROC areas values indicate that accuracy of the classifier is sufficiently high. Also, error rates are low and they show that differences between predicted and observed values are small. MLP is the classifier with the best detection rate (92%) among all classifiers, which are tested so far.

As the last classifier algorithm used in the study was Sequential Minimal Optimization (SMO) that was designed by John Platt in 1998 at Microsoft Research. Figure 4.15 shows the results of the SMO classifier. Normalized polynomial kernel function was used to normalization in the kernel function. The normalized polynomial function was intended to normalize the mathematical expression of the polynomial kernel instead of normalizing the data set [162]. It can be said that the normalized polynomial kernel is a generalized version of the polynomial kernel. As you can see on the output, 95 support vectors are found along with 7686 kernel evaluations. Some of these vectors are seen on the output. A new instance was predicted with these support vectors. It is seen that the number of true positives is high with a number of false positives. However, the aim of the study is high true

positive rate with the low positive rate. Also, that kappa statistic values shows that predictions were most dependent on chance. On the other side, the F-Measure value influenced the accuracy of classifier negatively. In general, it cannot be said that the performance of the classifier is very good because the parameter values in the output have average values.

```
Instances:    126
Attributes:   4
              AFP(IU/ml)
              uE3(ng/ml)
              hCG(mlU/ml)
              Result
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

SMO

Kernel used:
  Normalized Poly Kernel: K(x,y) = <x,y>^2.0/(<x,x>^2.0*<y,y>^2.0)^(1/2)

Classifier for classes: Trisomy21, Negative

BinarySMO

      1       * <0.098093 0.080949 0.207313 > * X]
 +        1       * <0.224796 0.103223 0.372518 > * X]
 +        1       * <0.06267 0.061934 0.44983 > * X]
 -        1       * <0.197029 0.041965 0.316876 > * X]
 -        1       * <0.153978 0.051727 0.590655 > * X]
 +        1       * <0.226158 0.061391 0.20183 > * X]
 -        1       * <0.209159 0.102985 0.416287 > * X]
 +        1       * <0.181199 0.04473 0.715852 > * X]
 +        1       * <0.716621 0.467584 1 > * X]
 -        1       * <0.090384 0.082118 0.589071 > * X]
 +        1       * <0.550409 0.06954 0.591759 > * X]
 -        1       * <0.11156 0.087152 0.459393 > * X]
 -        0.3686 * <0.25605 0.033039 0.594311 > * X]
 +        1       * <0.717984 0.06628 0.521681 > * X]
 +        0.7913 * <1 0.181456 0.597956 > * X]
 -        1       * <0.251351 0.065525 0.691207 > * X]
 -        1       * <0.221663 0.096082 0.575402 > * X]
 +        1       * <0.764305 0.036581 0.906915 > * X]
 -        1       * <0.27031 0.077359 0.611092 > * X]
 +        1       * <0.27248 0.070989 0.476651 > * X]
 +        1.4746

Number of support vectors: 95

Number of kernel evaluations: 7686 (94.093% cached)



Time taken to build model: 0.11 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          87               69.0476 %
Kappa statistic                          0.3968
Mean absolute error                      0.3095
Root mean squared error                  0.5563
Relative absolute error                 64.599  %
Root relative squared error            113.7032 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,820    0,395    0,577      0,820    0,678      0,420    0,713     0,545     Trisomy21
                 0,605    0,180    0,836      0,605    0,702      0,420    0,713     0,744     Negative
Weighted Avg.    0,690    0,265    0,734      0,690    0,693      0,420    0,713     0,665

=== Confusion Matrix ===

  a  b   <-- classified as
 41  9 |  a = Trisomy21
 30 46 |  b = Negative
```

Figure 4.15 Detailed results of the SMO classification algorithm for the data set

After running all algorithms one by one, the results had to be run multiple times to be stable and accurate results. Also, randomness is an important criterion for completing the study. Many runs of the test provide randomness of the instances and randomness

of other parameters which are used in classification. Thus, bias and deviation to any direction in the results can be avoided with randomness. Therefore, the selected three independent variables and a dependent variable examined 1000 runs with classification algorithms which are zeroR, k-NN, OneR, BayesNET, Naive Bayes, J48, FLDA, Logistic Regression, MLP and SMO on experimenter mod of WEKA software. In addition, the confidence level was set to 0.01 for all criteria.

Firstly, correct classifying percentages were observed. Figure 4.16 represents correct classification percentages of each classifier. The ZeroR was selected as the base classifier and the results of the other classifiers were compared with ZeroR. The expression used by Weka is "(*v/ /*)". "*v*" signifies that the result is significantly better or more than base classifier. "*" indicates that the result is significantly worse or less than base classifier and the space in the middle indicates that the result is neither good nor bad from the base classifier. For example, the result of the 7th classifier was neither good nor bad; it approximately had the same result with the base classifier. 6 classifiers gave better results than base classifier when it continued to compare. Also, MLP had the best result with 89.93% for a correct classification percentage. The worst classifier was SMO with 68.17% according to correct classification percentages.

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.01 -result-matrix "weka.experiment.ResultMatrixPlainText
Analysing:  Percent_correct
Datasets:   1
Resultsets: 10
Confidence: 0.01 (two tailed)
Sorted by:  -
Date:       16.10.2017 16:40


Dataset                   (1) rules.Ze | (2) lazy. (3) rules (4) bayes (5) bayes (6) trees (7) funct (8) funct (9) funct (10) func
-----------------------------------------------------------------------------------------------------------------------------------
'hastalar-weka.filters.un(1000   60.26 |   86.26 v   77.98 v   86.10 v   81.18 v   82.35 v   67.91     67.55     89.93 v   68.17
-----------------------------------------------------------------------------------------------------------------------------------
                             (v/ /*) |   (1/0/0)   (1/0/0)   (1/0/0)   (1/0/0)   (1/0/0)   (0/1/0)   (0/1/0)   (1/0/0)   (0/1/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) lazy.IBk '-K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\"\"' -30801
(3) rules.OneR '-B 6' -3459427003147861443
(4) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E bayes.net.estimate.SimpleEstimator -- -A 0.5' 74603744325877
(5) bayes.NaiveBayes '' 5995231201785697655
(6) trees.J48 '-C 0.25 -M 2' -217733168393644444
(7) functions.FLDA '-R 1.0E-6' -9212385698193681291
(8) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
(9) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 5' -5990607817048210779
(10) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.NormalizedPolyKernel -E 2.0 -C 250007\"
```

Figure 4.16 Percentages of correctly classified for classifiers

Detection rate or recall was defined before in Equation 2.1. Figure 4.17 presents detection rates of classifiers and base classifier (ZeroR) had 0 values. It means that the base classifier did not predict any instances of Trisomy 21 class correctly. In earlier cases, it was explained why ZeroR classified all instances of Trisomy21 wrong. So, all classifiers had the better results than the ZeroR. But the best one was Bayesian Network algorithm with 97% percentage. The one with the closest result to Bayesian Network was Naïve Bayes classifier. Also, Logistic regression classifier had the worst detection rate.

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.01 -result-matrix "weka.experiment.ResultMatri
Analysing:  IR_recall
Datasets:   1
Resultsets: 10
Confidence: 0.01 (two tailed)
Sorted by:  -
Date:       16.10.2017 15:07


Dataset                  (1) rules.Z | (2) lazy (3) rule (4) baye (5) baye (6) tree (7) func (8) func (9) func (10) fun
-----------------------------------------------------------------------------------------------------------------------------
'hastalar-weka.filters.un(1000   0.00 |   0.92 v   0.66 v   0.97 v   0.96 v   0.81 v   0.74 v   0.55 v   0.94 v   0.81 v
-----------------------------------------------------------------------------------------------------------------------------
                            (v/ /*) |  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) lazy.IBk '-K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\
(3) rules.OneR '-B 6' -3459427003147861443
(4) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E bayes.net.estimate.SimpleEstimator -- -A 0.5' 746
(5) bayes.NaiveBayes '' 5995231201785697655
(6) trees.J48 '-C 0.25 -M 2' -217733168393644444
(7) functions.FLDA '-R 1.0E-6' -9212385698193681291
(8) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
(9) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 5' -5990607817048210779
(10) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.NormalizedPolyKernel -E 2.0 -
```

Figure 4.17 Detection rates of classifiers

FPR means that a subject without the chromosomal aneuploidy is misclassified like having the chromosomal aneuploidy on the essential of the screening test. The subject gives the uncertain impression that the baby has the disease and therefore endures the redundant psychological results as well as having to undertake possibly invasive diagnostic or treatment procedures [103]. It is a percentage of all negative results in all positive results. The false positive rates of classifiers are presented in Figure 4.18. The smallest values are more important for the study because the objective of the study is to minimize the FPR. Therefore, MLP is the best classifier according to false positive rates. Also, the base classifier has 0.0 FP rate. Its reason

that ZeroR classified all instances as Negative and thus there was no actual Negative instance (false positive) which was classified as Trisomy21. Lastly, the SMO has the worst false positive rate with 41%.

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.01 -result-matrix "weka.experiment.ResultMatri
Analysing:  False_positive_rate
Datasets:   1
Resultsets: 10
Confidence: 0.01 (two tailed)
Sorted by:  -
Date:       16.10.2017 15:08


Dataset                  (1) rules.Z | (2) lazy (3) rule (4) baye (5) baye (6) tree (7) func (8) func (9) func (10) fun
-----------------------------------------------------------------------------------------------------------------------------
'hastalar-weka.filters.un(1000   0.00 |   0.17 v   0.14 v   0.21 v   0.28 v   0.17 v   0.36 v   0.24 v   0.13 v   0.41 v
-----------------------------------------------------------------------------------------------------------------------------
                           (v/ /*) |  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) lazy.IBk '-K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\
(3) rules.OneR '-B 6' -3459427003147861443
(4) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E bayes.net.estimate.SimpleEstimator -- -A 0.5' 746
(5) bayes.NaiveBayes '' 5995231201785697655
(6) trees.J48 '-C 0.25 -M 2' -217733168393644444
(7) functions.FLDA '-R 1.0E-6' -9212385698193681291
(8) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
(9) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 5' -5990607817048210779
(10) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.NormalizedPolyKernel -E 2.0 -
```

Figure 4.18 False positive rates of classifiers

Another result is the area under ROC. The best possible estimation method gives a point that represents 100% accuracy (no false negatives) and 100% specificity (no false positives) in the upper left corner or coordinate (0, 1) of the ROC area. (0, 1) is also called an excellent classification. In other words, area 1 represents an excellent test; an area of 0.5 represents a worthless test. The best result should have a maximum value. Therefore, Bayesian Network and Naïve Bayes classifiers have the best results with 94% as shown in Figure 4.19. So, they are close to the excellent test. After them, MLP and k-NN classifiers have the good ROC area values with 93%. The worst classifier is the ZeroR (base classifier) according to ROC area.

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.01 -result-matrix "weka.experiment.ResultMatri
Analysing:  Area_under_ROC
Datasets:   1
Resultsets: 10
Confidence: 0.01 (two tailed)
Sorted by:  -
Date:       16.10.2017 16:44


Dataset                   (1) rules.Z | (2) lazy (3) rule (4) baye (5) baye (6) tree (7) func (8) func (9) func (10) fun
----------------------------------------------------------------------------------------------------------------------------
'hastalar-weka.filters.un(1000   0.50 |   0.93 v   0.76 v   0.94 v   0.94 v   0.86 v   0.77 v   0.76 v   0.93 v   0.70 v
----------------------------------------------------------------------------------------------------------------------------
                            (v/ /*) |  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) lazy.IBk '-K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\
(3) rules.OneR '-B 6' -3459427003147861443
(4) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E bayes.net.estimate.SimpleEstimator -- -A 0.5' 74
(5) bayes.NaiveBayes '' 5995231201785697655
(6) trees.J48 '-C 0.25 -M 2' -217733168393644444
(7) functions.FLDA '-R 1.0E-6' -9212385698193681291
(8) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
(9) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 5' -5990607817048210779
(10) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.NormalizedPolyKernel -E 2.0 -
```

Figure 4.19 Area under ROC of classifiers

Last analysis was F-measure. In a statistical analysis of binary classification, the F-measure is a measure of a test's accuracy. It analyzes the precision and the recall of the test to calculate the score. Its score can be clarified as a weighted mean of the precision and recall, and the best value of F-measure is 1 and the worst value is 0 [163]. As a result, Multilayer Perceptron had best F-measure value (88%) for the study as given in Figure 4.20. After the MLP classifier, k-NN and Bayesian Network classifiers had the best F-measure values with 85%. Logistic regression had the worst F-measure value when the base classifier excluded from comparing.

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.01 -result-matrix "weka.experiment.ResultMatri
Analysing:  F_measure
Datasets:   1
Resultsets: 10
Confidence: 0.01 (two tailed)
Sorted by:  -
Date:       16.10.2017 16:44


Dataset                   (1) rules.Z | (2) lazy (3) rule (4) baye (5) baye (6) tree (7) func (8) func (9) func (10) fun
----------------------------------------------------------------------------------------------------------------------------
'hastalar-weka.filters.un(1000   0.00 |   0.85 v   0.70 v   0.85 v   0.81 v   0.78 v   0.65 v   0.56 v   0.88 v   0.67 v
----------------------------------------------------------------------------------------------------------------------------
                            (v/ /*) |  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)  (1/0/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) lazy.IBk '-K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\
(3) rules.OneR '-B 6' -3459427003147861443
(4) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E bayes.net.estimate.SimpleEstimator -- -A 0.5' 74
(5) bayes.NaiveBayes '' 5995231201785697655
(6) trees.J48 '-C 0.25 -M 2' -217733168393644444
(7) functions.FLDA '-R 1.0E-6' -9212385698193681291
(8) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
(9) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 5' -5990607817048210779
(10) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.NormalizedPolyKernel -E 2.0 -
```

Figure 4.20 F-measures of classifiers

Table 4.5 shows the averages which were calculated for each algorithm after 100 runs of the data set. 1000 runs were totally made for the study. Percentage of Correctly Classified, Detection Rate, False Positive Rate, Area under ROC and F-Measure were the most important criteria for the study. While using these criteria in the study, the criteria were weighted according to their importance.

Table 4.5 The results of the classification algorithms tests

	% Correctly Classified	% Detection Rate (DR)	% False Positive Rate (FPR)	% Area Under ROC	% F Measure	Total Weighted Scores
Weights	*35*	*50*	*-40*	*35*	*30*	
ZeroR (Base)	0.6	0	0	0.5	0	
Weighted Score	*0*	*0*	*0*	*0*	*0*	*0*
k-NN	0.86	0.92	0.17	0.9	0.85	
Weighted Score	*9.1*	*46*	*-6.8*	*14*	*25.5*	*87.8*
OneR	0.78	0.66	0.14	0.66	0.7	
Weighted Score	*6.3*	*33*	*-5.6*	*5.6*	*21*	*60.3*
Bayesian Network	0.86	0.97	0.21	0.93	0.85	
Weighted Score	*9.1*	*48.5*	*-8.4*	*15.05*	*25.5*	*89.75*
Naïve Bayes	0.81	0.96	0.28	0.92	0.81	
Weighted Score	*7.35*	*48*	*-11.2*	*14.7*	*24.3*	*83.15*
C4.5	0.82	0.81	0.17	0.85	0.78	
Weighted Score	*7.7*	*40.5*	*-6.8*	*12.25*	*23.4*	*77.05*
FLDA	0.68	0.74	0.36	0.78	0.65	
Weighted Score	*2.8*	*37*	*-14.4*	*9.8*	*19.5*	*54.7*
Logistic Regression	0.68	0.55	0.24	0.77	0.56	
Weighted Score	*2.8*	*27.5*	*-9.6*	*9.45*	*16.8*	*46.95*
MLP	0.9	0.94	0.13	0.92	0.88	
Weighted Score	*10.5*	*47*	*-5.2*	*14.7*	*26.4*	*93.4**
SMO	0.68	0.81	0.41	0.7	0.67	
Weighted Score	*2.8*	*40.5*	*-16.4*	*7*	*20.1*	*54*
** The best test according to total weighted scores*						

The weight given for the percentage correctly classified criteria was 35, and the weight for the detection rate was 50. The detection rate was the most important criteria for the study. Therefore the weight of the detection rate should have had the largest value. The ROC area was weighted with 35 and the F-measure was weighted with 30. Lastly, the weight of the FPR was determined as -40. It had the minus sign because FPR never gives any benefit to the improvement of the triple screening test. In other words, if the weight of the FPR had a plus value, then every increment in the FPR would appear as a benefit in the scoring model and an error would occur in the calculations. Also, an equation that considered the base classifier was also added to the scoring model. The weighted score for any classifier was described as in Equation 4.3.

$$W(x_i) = (x_i - y_i) \times w_i \tag{4.3}$$

where
x is the classifier,
y is the base classifier,
i is the criteria,
x_i is the value of the specified criterion of the particular classifier,
y_i is the value of specified criterion of the base classifier,
w_i is the weight of the specified criterion.

After calculating the weighted scores for all the classifiers and criteria, these scores were summed and the total weighted score was calculated. Equation 4.4 represents total weighted score formula.

$$\sum W(x_i) = \sum (x_i - y_i) \times w_i \tag{4.4}$$

Thus, the best test and improvement was observed with the weighted scoring model. As a result, the best test was provided Multilayer Perceptron algorithm with 94.24% detection rate and 13% false positive rate.

CHAPTER 5

CONCLUSION & DISCUSSION

In this thesis, it is aimed to predict Down syndrome in a more precise manner. Improving the performance of the triple screening test will reduce the need for invasive testing so that there will be no risk of the pregnant woman and fetus due to invasive tests. The detection rate of the triple screening test, which was used in the thesis, was increased from 61% to 94%. However, the false positive rate increased from 5% to 13%. In addition, while N.J. Wald's study had 80% detective rate versus 16% false positive rate, the thesis had 94% detection rate and false positive rate 13%. Detection rates and false positive rates of N. J. Wald's study are shown in Table 5.1. When detection rate increased, the false positive rate increased proportionally. Their relations and equation are as shown in Figure 5.1. As a result, if detection rate is 94%, the false positive rate will be approximately 25.2% according to their relationship. So, the false positive rate in the thesis is smaller than Wald's study.

Table 5.1 Detection rates and false positive rates for N. J. Wald's study

Detection rate (%) of Down syndrome	**False positive rate (%)**
80%	16
75%	12
70%	8.6
65%	6.4
60%	4.7
55%	3.4
50%	2.5
45%	1.7
40%	1.2
35%	0.8
30%	0.5
25%	0.3
20%	0.2

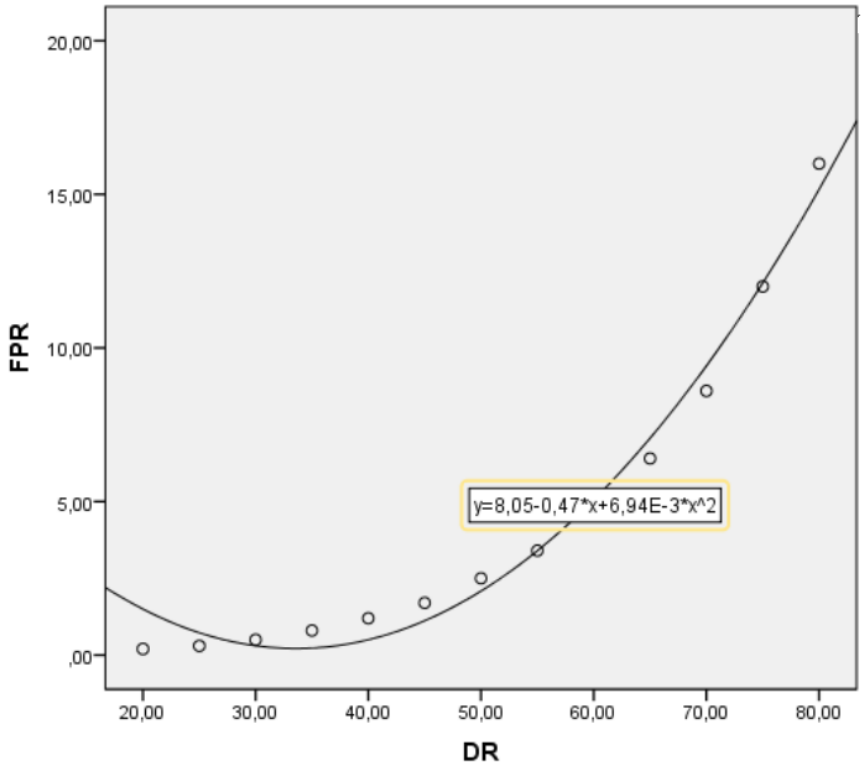


Figure 5.1 Relationship and equation between DR and FPR for Wald's study

This demonstrates that the thesis improved the performance of the triple screening test in a significant amount. Future works can be improving the performance of the triple screening test by reducing the false positive rate from 13% to less or detection rate can be increased when the false positive rate is constant. Hereby, invasive tests like amniocentesis won't be needed and life of baby and mother won't risk anymore.

REFERENCES

[1] R. Agrawal, T. Imieliński, and A. Swami. (1993). Mining Association Rules Between Sets of Items in Large Databases. *ACM SIGMOD International Conference on Management of Data*. New York, NY, USA. 207–216.

[2] Paramjit Kaur and Kanwalpreet Singh Attwal (2014). Data Mining:Review. *Int. J. Comput. Sci. Inf. Technol.* **5(5)**, 6225–6228.

[3] J. Han, M. Kamber, and J. Pei. (2006). *Data Mining: Concepts and Techniques, Second Edition*, 2 edition. Amsterdam ; Boston : San Francisco. CA: Morgan Kaufmann.

[4] J. Han, M. Kamber, and J. Pei. (2011). *Data Mining: Concepts and Techniques, Third Edition*, 3 edition. Amsterdam ; Boston : San Francisco. Morgan Kaufmann.

[5] M. Shin *et al.*(2009). Prevalence of Down syndrome among children and adolescents in 10 regions of the United States. *Pediatrics*. **124(6)**, 1565–1571.

[6] A. V. Kenshole, D. Gallichan, S. Pahl, and J. Clibbens. (2017). Lifestyle factors and Alzheimer's disease in people with Down syndrome. *J. Appl. Res. Intellect. Disabil. JARID*.

[7] N. J. Wald *et al.* (1988). Maternal serum screening for Down's syndrome in early pregnancy. *BMJ*. **297(6653)**, 883–887.

[8] D. A. Driscoll and S. Gross. (2009). Clinical practice. Prenatal screening for aneuploidy. *N. Engl. J. Med.***360(24)**, 2556–2562.

[9] C. Papp, Z. Szigeti, E. Tóth-Pál, J. Hajdú, J. G. Joó, and Z. Papp. (2008). Ultrasonographic Findings of Fetal Aneuploidies in the Second Trimester – Our Experiences. *Fetal Diagn. Ther.* **23(2)**, 105–113.

[10] R. G. Resta. (2005). Changing demographics of advanced maternal age (AMA) and the impact on the predicted incidence of Down syndrome in the United States: Implications for prenatal screening and genetic counseling. *Am. J. Med. Genet. A.***133(1)**, 31–36.

[11] K. Spencer, C. Y. Ong, A. W. Liao, and K. H. Nicolaides. (2000). The influence of ethnic origin on first trimester biochemical markers of chromosomal abnormalities. *Prenat. Diagn.* **20(6)**, 491–494.

[12] T. D. Shipp and B. R. Benacerraf. (2002). Second trimester ultrasound screening for chromosomal abnormalities. *Prenat. Diagn.* **22(4)**, 296–307.

[13] N. J. Wald *et al.* (2003). First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *Health Technol. Assess. Winch. Engl.* **7(11)**, 1–77.

[14] What Is Down Syndrome? - National Down Syndrome Society. Intl.: http://www.ndss.org/Down-Syndrome/What-Is-Down-Syndrome/., 28.08.2017.

[15] R. Bhattacharyya, D. Sanyal, K. Roy, and S. Bhattacharyya. (2010). Correlation between physical anomaly and behavioral abnormalities in Down syndrome. *J. Pediatr. Neurosci.* **5(2)**, 105–110.

[16] J. Van Robays. (2016). John Langdon Down (1828 – 1896). *Facts Views Vis. ObGyn*. 8**(2)**, 131–136.

[17] E. Fletcher-Janzen and C. R. Reynolds. (2007). *Encyclopedia of Special Education* .Volume 1, 3rd Edition. New York: Wiley.

[18] E. R. MD and H. R. PhD. (2013). *Essentials of Rubin's Pathology*. Sixth, None edition. Philadelphia: LWW.

[19] J. Lejeune, M. Gautier, and R. Turpin. (1959). *Étude des chromosomes somatiques de neuf enfants mongoliens.*

[20] E. Prost and N. Roberts. (2013). Downs: The History of a Disability. *J. Can. Acad. Child Adoles, Psychiatry*. **22(2)**, 180–181.

[21] S. E. Antonarakis *et al.* (1992). The meiotic stage of nondisjunction in trisomy 21: determination by using DNA polymorphisms. *Am. J. Hum. Genet.* **50(3)**, 544–550.

[22] L. S. Penrose. (2009). The relative effects of paternal and maternal age in mongolism. 1933. *J. Genet.* **88(1)**, 9–14.

[23] J. K. Morris, D. E. Mutton, and E. Alberman. (2002). Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *J. Med. Screen.* **9(1)**, 2–6.

[24] E. S. Mansfield. (1993). Diagnosis of Down syndrome and other aneuploidies using quantitative polymerase chain reaction and small tandem repeat polymorphisms. *Hum. Mol. Genet.* **2(1)**, 43–50.

[25] ACOG Committee on Practice Bulletins, ACOG Practice Bulletin No. 77. (2007). Screening for fetal chromosomal abnormalities. *Obstet. Gynecol.* **109(1)**, 217–227.

[26] ACOG Committee on Practice Bulletins, ACOG Practice Bulletin No. 88 (2007). (2007). Invasive prenatal testing for aneuploidy. *Obstet. Gynecol.* **110(6)**, 1459–1467.

[27] M. Cruz-Lemini *et al.* (2014). How to perform an amniocentesis," *Ultrasound Obstet. Gynecol.* **44(6)**, 727–731.

[28] C. Ogilvie and R. Akolekar. (2014). Pregnancy Loss Following Amniocentesis or CVS Sampling-Time for a Reassessment of Risk. *J. Clin. Med.* **3(3)**, 741–746.

[29] F. Tara, M. Lotfalizadeh, and S. Moeindarbari. (2016). The effect of diagnostic amniocentesis and its complications on early spontaneous abortion. *Electron. Physician*. **8(8)**, 2787–2792.

[30] Amniocentesis, Pregnancy and amniocentesis decision aid. Intl.: https://patient.info/health/amniocentesis-leaflet., 28.08.2017.

[31] Antenatal care for uncomplicated pregnancies | Guidance and guidelines | NICE. Intl.: https://www.nice.org.uk/guidance/cg62., 28.08.2017.

[32] Prenatal Screening and Diagnosis of Down's Syndrome | Health. Intl.: https://patient.info/health/prenatal-screening-and-diagnosis-of-downs-syndrome., 28.08.2017.

[33] Chorionic Villus Sampling, Medical investigation in pregnancy. Intl.: https://patient.info/health/chorionic-villus-sampling-leaflet., 28.08.2017.

[34] Amniocentesis and Chorionic Villus Sampling (Green-top Guideline No. 8), *Royal College of Obstetricians & Gynaecologists*. Intl.: https://www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg8/., 28.08.2017.

[35] A. Tabor, D. Jerne, and J. E. Bock. (1986). Incidence of rhesus immunisation after genetic amniocentesis. *Br. Med. J. Clin. Res. Ed*. **293(6546)**, 533–536.

[36] K. Nicolaides, M. de L. Brizot, F. Patel, and R. Snijders. (1994). Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10-13 weeks' gestation. *Lancet Lond. Engl.* **344(8920)**, 435–439.

[37] K. Sundberg *et al.* (1997). Randomised study of risk of fetal loss related to early amniocentesis versus chorionic villus sampling. *Lancet Lond. Engl.* **350(9079)**, 697–703.

[38] A. Tabor and Z. Alfirevic. (2010). Update on procedure-related risks for prenatal diagnosis techniques. *Fetal Diagn. Ther.* **27(1)**, 1–7.

[39] R. D. Wilson, S. Langlois, J.-A. Johnson, and Society of Obstetricians and Gynaecologists of Canada. (2007). Mid-trimester amniocentesis fetal loss rate. *J. Obstet. Gynaecol. Can. JOGC J. Obstet. Gynecol. Can. JOGC.* **29(7)**, 586–595.

[40] R. Akolekar, S. Bower, N. Flack, C. M. Bilardo, and K. H. Nicolaides. (2011). Prediction of miscarriage and stillbirth at 11-13 weeks and the contribution of chorionic villus sampling. *Prenat. Diagn.* **31(1)**, 38–45.

[41] C. M. Ogilvie and R. Akolekar. (2013). Procedure-related pregnancy loss following invasive prenatal sampling: time for a new approach to risk assessment and counseling. *Expert Rev. Obstet. Gynecol.* **8(2)**, 135–142.

[42] I. R. Merkatz, H. M. Nitowsky, J. N. Macri, and W. E. Johnson. (1984). An association between low maternal serum α-fetoprotein and fetal chromosomal abnormalities. *Am. J. Obstet. Gynecol.* **148(7)**, 886–894.

[43] H. S. Cuckle, N. J. Wald, and S. G. Thompson. (1987). Estimating a woman's risk of having a pregnancy associated with Down's syndrome using her age and serum alpha-fetoprotein level. *Br. J. Obstet. Gynaecol.* **94(5)**, 387–402.

[44] N. J. Wald, J. W. Densem, L. George, S. Muttukrishna, and P. G. Knight. (1996). Prenatal screening for Down's syndrome using inhibin-A as a serum marker. *Prenat. Diagn.* **16(2)**, 143–153.

[45] W.-L. Ke, W.-H. Zhao, and X.-Y. Wang. (2015). Detection of fetal cell-free DNA in maternal plasma for Down syndrome, Edward syndrome and Patau syndrome of high risk fetus. *Int. J. Clin. Exp. Med.* **8(6)**, 9525–9530.

[46] N. Wald and I. Leck, Eds. (2000). *Antenatal and Neonatal Screening*, 2 edition. Oxford, New York: OUP Oxford.

[47] D. Chitayat, S. Langlois, R. D. Wilson, Genetics Committee of the Society of Obstetricians and Gynaecologists of Canada, and Prenatal Diagnosis

Committee of the Canadian College of Medical Geneticists. (2011). Prenatal screening for fetal aneuploidy in singleton pregnancies. *J. Obstet. Gynaecol. Can. JOGC J. Obstet. Gynecol. Can. JOGC*. **33(7)**, 736–750.

[48] D. Neagos, R. Cretu, R. C. Sfetea, and L. C. Bohiltea. (2011). The Importance of Screening and Prenatal Diagnosis in the Identification of the Numerical Chromosomal Abnormalities. *Mædica*. **6(3)**, 179–184.

[49] K. H. Nicolaides, G. Azar, D. Byrne, C. Mansur, and K. Marks. (1992). Fetal nuchal translucency: ultrasound screening for chromosomal defects in first trimester of pregnancy. *BMJ*. **304(6831)**, 867–869.

[50] R. J. Snijders, P. Noble, N. Sebire, A. Souka, and K. H. Nicolaides. (1998). UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10-14 weeks of gestation. Fetal Medicine Foundation First Trimester Screening Group. *Lancet Lond. Engl.* **352(9125)**, 343–346.

[51] N. J. Wald and A. K. Hackshaw. (1997). Combining ultrasound and biochemistry in first-trimester screening for Down's syndrome. *Prenat. Diagn.* **17(9)**, 821–829.

[52] F. D. Malone, M. E. D'Alton, and Society for Maternal-Fetal Medicine. First-trimester sonographic screening for Down syndrome. *Obstet. Gynecol.* **102(5)**, 1066–1079.

[53] R. Wapner *et al.* (2003). First-Trimester Screening for Trisomies 21 and 18. *N. Engl. J. Med.* **349(15)**, 1405–1413.

[54] J. L. Simpson. (2012). Invasive procedures for prenatal diagnosis: any future left?. *Best Pract. Res. Clin. Obstet. Gynaecol.* **26(5)**, 625–638.

[55] D. A. Driscoll, S. J. Gross, and Professional Practice Guidelines Committee. (2009). Screening for fetal aneuploidy and neural tube defects. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **11(11)**, 818–821.

[56] F. D. Malone *et al.* (2005). First-trimester or second-trimester screening, or both, for Down's syndrome. *N. Engl. J. Med.* **353(19)**, 2001–2011.

[57] K. Spencer, C. E. Spencer, M. Power, C. Dawson, and K. H. Nicolaides. (2003). Screening for chromosomal abnormalities in the first trimester using ultrasound and maternal serum biochemistry in a one-stop clinic: a review of three years prospective experience. *BJOG Int. J. Obstet. Gynaecol.* **110(3)**, 281–286.

[58] D. Wright *et al.* (2010). First-trimester combined screening for trisomy 21 at 7-14 weeks' gestation. *Ultrasound Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstet. Gynecol.* **36(4)**, 404–411.

[59] N. Tørring. (2009). Performance of first-trimester screening between gestational weeks 7 and 13. *Clin. Chem.* **55(8)**, 1564–1567.

[60] I. Kirkegaard, O. B. Petersen, N. Uldbjerg, and N. Tørring. (2008). Improved performance of first-trimester combined screening for trisomy 21 with the double test taken before a gestational age of 10 weeks. *Prenat. Diagn.* **28(9)**, 839–844.

[61] A. Borrell *et al.* (2004). First-trimester screening for trisomy 21 combining biochemistry and ultrasound at individually optimal gestational ages. An interventional study. *Prenat. Diagn.* **24(7)**, 541–545.

[62] M. Smith and J. Visootsak. (2013). Noninvasive screening tools for Down syndrome: a review. *Int. J. Womens Health.* **5**, 125–131.

[63] P. O'Leary *et al.* (2016). Screening for Down syndrome in the second trimester of pregnancy. *Aust. N. Z. J. Obstet. Gynaecol.* **56(1)**, 19–21.

[64] H. S. Abou-Youssef, M. M. Kamal, and D. A. Mehaney. (2014). Triple Test Screening for Down Syndrome: An Egyptian-Tailored Study. *PLOS ONE*. **9(10)**, 110-370.

[65] G. E. Palomaki and J. E. Haddow. (1987). Maternal serum alpha-fetoprotein, age, and Down syndrome risk. *Am. J. Obstet. Gynecol.* **156(2)**, 460–463.

[66] T. M. Reynolds and M. D. Penney. (1990). The mathematical basis of multivariate risk screening: with special reference to screening for Down's syndrome associated pregnancy. *Ann. Clin. Biochem.* **27(5)**, 452–458.

[67] N. J. Wald, A. K. Hackshaw, and L. M. George. (2000). Assay precision of serum alpha fetoprotein in antenatal screening for neural tube defects and Down's syndrome. *J. Med. Screen.* **7(2)**, 74–77.

[68] J. Bishop, F. D. Dunstan, B. J. Nix, and T. M. Reynolds. (1995). The effects of gestation dating on the calculation of patient specific risks in Down's syndrome screening. *Ann. Clin. Biochem.* **32(5)**, 464–477.

[69] T. Reynolds, A. Ellis, and R. Jones. (2004). Down's syndrome risk estimates demonstrate considerable heterogeneity despite homogeneity of input. *Ann. Clin. Biochem.* **41(6)**, 464–468.

[70] K. Bricelj, M. Vuković, I. Verdenik, J. Osredkar, and K. Geršak. (2014). Analysis of Second Trimester Maternal Quadruple Screening Test for Trisomy 21 and Trisomy 18. *Slov. Med. J.* **83(9).**

[71] S. Yazdani, R. Rouholahnejad, N. Asnafi, M. Sharbatdaran, M. Zakershob, and Z. Bouzari. (2015). Correlation of pregnancy outcome with quadruple screening test at second trimester. *Med. J. Islam. Repub. Iran*. **29**, 281.

[72] H. C. Watt, N. J. Wald, D. Smith, A. Kennard, and J. Densem. (1996). Effect of allowing for ethnic group in prenatal screening for Down's syndrome. *Prenat. Diagn.* **16(8)**, 691–698.

[73] H. Güner. (2008). *Yüksek riskli gebelikler: Yönetim seçenekleri*. İstanbul: Güneş Tıp Kitabevleri.

[74] C. G. Bergstrand and B. Czar. (1956). Demonstration of a new protein fraction in serum from the human fetus. *Scand. J. Clin. Lab. Invest.* **8(2)**, 174.

[75] F. G. Cunningham and J. W. Williams, Eds. (1997). *Williams obstetrics*. 20th ed. Stamford, Conn: Appleton & Lange.

[76] Z. A. Habib. (1977). Maternal serum alpha-feto-protein: Its value in antenatal diagnosis of genetic disease and in obstetrical-gynaecological care. *Acta Obstet. Gynecol. Scand.* **56(61)**, 1–91.

[77] H. S. Cuckle, N. J. Wald, and R. H. Lindenbaum. (1984). Maternal serum alpha-fetoprotein measurement: a screening test for Down syndrome. *Lancet Lond. Engl.* **1(8383)**, 926–929.

[78] A. Zhang *et al.* (2016). A Chemiluminescent Protein Microarray Method for Determining the Seroglycoid Fucosylation Index. *Sci. Rep.* **6**, 31132.

[79] J. N. Macri *et al.* (1990). Maternal serum Down syndrome screening: free beta-protein is a more effective marker than human chorionic gonadotropin. *Am. J. Obstet. Gynecol.* **163(4)**, 1248–1253.

[80] D. N. Saller and J. A. Canick. (1996). Maternal serum screening for fetal Down syndrome: clinical aspects. *Clin. Obstet. Gynecol.* **39(4)**, 783–792.

[81] L. H. Kellner *et al.* (1995). Triple marker (alpha-fetoprotein, unconjugated estriol, human chorionic gonadotropin) versus alpha-fetoprotein plus free-beta subunit in second-trimester maternal serum screening for fetal Down syndrome: a prospective comparison study. *Am. J. Obstet. Gynecol.* **173(4)**, 1306–1309.

[82] M. L. MacDonald, R. M. Wagner, and R. N. Slotnick. (1991). Sensitivity and specificity of screening for Down syndrome with alpha-fetoprotein, hCG, unconjugated estriol, and maternal age. *Obstet. Gynecol.* **77(1)**, 63–68.

[83] V. Subbiah *et al.* (2014). Next generation sequencing analysis of platinum refractory advanced germ cell tumor sensitive to Sunitinib (Sutent®) a VEGFR2/PDGFRβ/c-kit/ FLT3/RET/CSF1R inhibitor in a phase II trial. *J. Hematol. Oncol.J Hematol Oncol*. **7(1)**.

[84] R. E. B. MD, R. M. K. MD, and H. B. J. MD. (2000). *Nelson Textbook of Pediatrics.* 16 edition. Philadelphia: Saunders.

[85] J. E. Haddow, G. E. Palomaki, G. J. Knight, G. C. Cunningham, L. S. Lustig, and P. A. (1994). Boyd. Reducing the need for amniocentesis in women 35 years of age or older with serum markers for screening. *N. Engl. J. Med.* **330(16)**, 1114–1118.

[86] F. Ren *et al.* (2016). Second trimester maternal serum triple screening marker levels in normal twin and singleton pregnancies. *Biomed. Rep.* **4(4)**, 475–478.

[87] T. M. Reynolds, M. D. Penney, H. Hughes, and R. John. (1991). The effect of weight correction on risk calculations for Down's syndrome screening. *Ann. Clin. Biochem.* **28(3)**, 245–249.

[88] A. Matias, N. Montenegro, and I. Blickstein. (2005). Down syndrome screening in multiple pregnancies. *Obstet. Gynecol. Clin. North Am.* **32(1)**, 81–96.

[89] S. S. Malini and N. B. Ramachandra. (2007). Possible risk factors for Down syndrome and sex chromosomal aneuploidy in Mysore, South India. *Indian J. Hum. Genet.* **13(3)**, 102–108.

[90] D. A. Krantz, T. W. Hallahan, V. J. Macri, and J. N. Macri. (2005). Maternal weight and ethnic adjustment within a first-trimester Down syndrome and trisomy 18 screening program. *Prenat. Diagn.* **25(8)**, 635–640.

[91] K. Spencer, V. Heath, A. El-Sheikhah, C. Y. T. Ong, and K. H. Nicolaides. Ethnicity and the need for correction of biochemical and ultrasound markers of chromosomal anomalies in the first trimester: a study of Oriental, Asian and Afro-Caribbean populations. *Prenat. Diagn.* **25(5)**, 365–369.

[92] N. J. Wald, A. Kennard, A. Hackshaw, and A. McGuire. (1997). Antenatal screening for Down's syndrome. *J. Med. Screen.* **4(4)**, 181–246.

[93] M. Chen *et al.* (2002). The effect of ethnic origin on nuchal translucency at 10-14 weeks of gestation. *Prenat. Diagn.* **22(7)**, 576–578.

[94] B. Thilaganathan, M. Khare, B. Williams, and N. C. Wathen. (1998) Influence of ethnic origin on nuchal translucency screening for Down's syndrome. *Ultrasound Obstet. Gynecol.* **12(2)**, 112–114.

[95] E. B. Hook and P. K. Cross. (1985). Cigarette smoking and Down syndrome. *Am. J. Hum. Genet.* **37(6)**, 1216–1224.

[96] G. E. Palomaki, G. J. Knight, and J. E. Haddow. (1994). Human chorionic gonadotropin and unconjugated oestriol measurements in insulin-dependent diabetic pregnant women being screened for fetal Down syndrome. *Prenat. Diagn.* **14(1)**, 65–68.

[97] W. Huttly, A. Rudnicka, and N. J. Wald. (2004). Second-trimester prenatal screening markers for Down syndrome in women with insulin-dependent diabetes mellitus. *Prenat. Diagn.* **24(10)**, 804–807.

[98] K. Spencer, S. Cicero, A. Atzei, C. Otigbah, and K. H. Nicolaides. (2005). The influence of maternal insulin-dependent diabetes on fetal nuchal translucency thickness and first-trimester maternal serum biochemical markers of aneuploidy. *Prenat. Diagn.* **25(10)**, 927–929.

[99] F. D. J. Dunstan and A. B. J. Nix. (1998). Screening for Down's Syndrome: The Effect of Test Date on the Detection Rate. *Ann. Clin. Biochem.*. **35(1)**, 57–61.

[100] D. Powers. (2011). Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2(1)**, 37–63.

[101] C. J. V. Rijsbergen. (1979). *Information Retrieval*. 2nd ed. Newton, MA, USA: Butterworth-Heinemann.

[102] D. Normando, M. A. de O. Almeida, and C. C. A. Quintão. (2011). Análise do emprego do cálculo amostral e do erro do método em pesquisas científicas publicadas na literatura ortodôntica nacional e internacional. *Dent. Press J. Orthod.* **16(6)**, 33–35.

[103] L. D. Maxim, R. Niebo, and M. J. Utell. (2014). Screening tests: a review with examples. *Inhal. Toxicol.* **26(13)**, 811–828.

[104] J. A. Hanley and B. J. McNeil. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* **143(1)**, 29–36.

[105] J. C. Bishop, F. D. Dunstan, B. J. Nix, T. M. Reynolds, and A. Swift. (1993). All MoMs are not equal: some statistical properties associated with reporting results in the form of multiples of the median. *Am. J. Hum. Genet.* **52(2)**, 425–430.

[106] N. J. Wald, H. Cuckle, J. H. Brock, R. Peto, P. E. Polani, and F. P. Woodford. (1977). Maternal serum-alpha-fetoprotein measurement in antenatal screening for anencephaly and spina bifida in early pregnancy. Report of U.K. collaborative study on alpha-fetoprotein in relation to neural-tube defects. *Lancet Lond. Engl.* **1(8026)**, 1323–1332.

[107] G. Vranken, T. Reynolds, and J. Van Nueten. (2006). Medians for second-trimester maternal serum markers: geographical differences and variation caused by median multiples-of-median equations. *J. Clin. Pathol.* **59(6)**, 639–644.

[108] N. J. Salkind. (2007). *Encyclopedia of measurement and statistics*. SAGE Publications.

[109] A. Lyon. (2014). Why are Normal Distributions Normal? *Br. J. Philos. Sci.* **65(3)**, 621–649.

[110] J. Krithikadatta. (2014). Normal Distribution. *J. Conserv. Dent. JCD*. **17(1)**, 96–97.

[111] M. Abramowitz and I. A. Stegun, Eds. (1995). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, 0009. Revised edition. New York, NY: Dover Publications.

[112] J. A. Swets. (1973). The Relative Operating Characteristic in Psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*. **182(4116)**, 990–1000.

[113] J. R. Thornbury, D. G. Fryback, and W. Edwards. (1975). Likelihood ratios as a measure of the diagnostic usefulness of excretory urogram information. *Radiology*. **114(3)**, 561–565.

[114] H. J. van der Helm and E. A. Hische. (1979). Application of Bayes's theorem to results of quantitative clinical chemical determinations. *Clin. Chem.* **25(6)**, 985–988.

[115] S. G. Pauker and J. P. Kassirer. Therapeutic decision making: a cost-benefit analysis. *N. Engl. J. Med.* **293(5)**, 229–234.

[116] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. (2016). *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. 4th edition. Amsterdam: Morgan Kaufmann.

[117] G. Holmes, A. Donkin, and I. H. Witten. (1994). *WEKA: A Machine Learning Workbench*. Department of Computer Science, University of Waikato.

[118] S. J. Cunningham. .Applying a Machine Learning Workbench: Experience with Agricultural Databases.

[119] P. Reutemann, B. Pfahringer, and E. Frank. (2004). A Toolbox for Learning from Relational Data with Propositional and Multi-instance Learners. *Advances in Artificial Intelligence.* 1017–1023.

[120] Weka 3 - Data Mining with Open Source Machine Learning Software in Java. Intl.: http://www.cs.waikato.ac.nz/ml/weka/index.html., 29.08.2017.

[121] D. Quintero *et al.* (2013). *Workload Optimized Systems: Tuning POWER7 for Analytics*. IBM Redbooks.

[122] D. Clawson, Ed. (1998). *Required Reading: Sociology's Most Influential Books*. Amherst: University of Massachusetts Press.

[123] C. Müller, B. Wellman, and A. Marin. (1999). How to Use SPSS to Study Ego-Centered Networks. *Bull. Sociol. Methodol. Methodologie Sociol.* **64(1)**, 83–100.

[124] IBM SPSS - IBM Analytics. (2016). Intl.: http://www.ibm.com/analytics/us/en/technology/spss/., 29.08.2017.

[125] L. Xiang-wei and Q. Yian-fang. (2012). A Data Preprocessing Algorithm for Classification Model Based On Rough Sets. *Phys. Procedia*. **25**, 2025–2029.

[126] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Int Res*. **16(1)**, 321–357.

[127] T. M. Ha and H. Bunke. (1997). "Off-Line, Handwritten Numeral Recognition by Perturbation Method. *IEEE Trans Pattern Anal Mach Intell*. **19(5)**, 535–539.

[128] D. J. Goodenough, K. Rossmann, and L. B. Lusted. 1974. Radiographic applications of receiver operating characteristic (ROC) curves. *Radiology*. **110(1)**, 89–95.

[129] N. Fenton and M. Neil. (2012). *Risk Assessment and Decision Analysis with Bayesian Networks*. 1 edition. Boca Raton: CRC Press.

[130] D. A. Freedman. (2009). *Statistical Models: Theory and Practice*. 2 edition. Cambridge ; New York: Cambridge University Press.

[131] T. Cover and P. Hart. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory*. **13(1)**, 21–27.

[132] N. S. Altman. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **46(3)**, 175–185.

[133] R. C. Holte. (1993). Very Simple Classification Rules Perform Well on Most Commonly Used Data sets. *Mach. Learn.* **11(1)**, 63–90.

[134] J. Pearl. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. *Proceedings of the 7th Conference of the Cognitive Science Society*. University of California, Irvine. 329–334.

[135] J. Pearl. (1998). The Handbook of Brain Theory and Neural Networks. *M. A. Arbib, Ed. Cambridge*. MA, USA: MIT Press.149–153.

[136] S. J. R. Norvig. (2002). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

[137] Naïve Bayesian. Intl.: http://www.saedsayad.com/naive_bayesian.html., 10.10.2017.

[138] J. R. Quinlan. (1993). *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

[139] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali. (2014). A comparative study of decision tree ID3 and C4.5. *Int. J. Adv. Comput. Sci. Appl.* **4(2)**.

[140] R. A. Fisher. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugen.* **7(2)**, 179–188.

[141] LDA. Intl.: http://www.saedsayad.com/lda.htm., 12.10.2017.

[142] A. Schneider, G. Hommel, and M. Blettner. (2010). Linear Regression Analysis. *Dtsch. Ärztebl. Int.* **107(44)**, 776–782.

[143] D. W. Hosmer and S. Lemeshow. (2000). Introduction to the Logistic Regression Model. In *Applied Logistic Regression*. John Wiley & Sons, Inc. 1–30.

[144] Logistic Regression. Intl.:http://www.saedsayad.com/logistic_regression.htm. 13.10.2017.

[145] D. W. Hosmer and S. Lemeshow. (2000). Multiple Logistic Regression. In *Applied Logistic Regression*. John Wiley & Sons, Inc. 31–46.

[146] F. Rosenblatt. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65(6)**, 386–408.

[147] Perceptron. Intl.: http://www.saedsayad.com/artificial_neural_network_bkp.htm., 13.10.2017.

[148] C. V. D. Malsburg. (1986). Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. In *Brain Theory*. Springer, Berlin, Heidelberg. 245–248.

[149] Support Vector Machine. Intl.: http://www.saedsayad.com/support_vector_machine.htm., 14.10.2017.

[150] J. Platt. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.

[151] B. Schölkopf, A. Smola, and K.-R. Müller. (1998). Nonlinear Component Analysis As a Kernel Eigenvalue Problem. *Neural Comput*. **10(5)**, 1299–1319.

[152] A. Gasteratos, M. Vincze, and J. K. Tsotsos. (2008). *Computer Vision Systems: 6th International Conference on Computer Vision Systems*. ICVS 2008 Santorini, Greece, Proceedings.

[153] B. Üstün, W. J. Melssen, and L. M. C. Buydens. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. *Chemom. Intell. Lab. Syst.* **81(1)**, 29–40.

[154] M. A. Aizerman, E. A. Braverman, and L. Rozonoer. (1964). Theoretical foundations of the potential function method in pattern recognition learning. In *Automation and Remote Control.* 821–837.

[155] R. Kohavi. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* San Francisco, CA, USA. **2**, 1137–1143.

[156] J. Cohen. (1960). A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **20(1)**, 37–46.

[157] J. L. Fleiss and others. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76(5)**, 378–382.

[158] T. M. Mitchell, *Machine Learning*. (1997). 1st ed. New York, NY, USA: McGraw-Hill, Inc.

[159] R. R. Bouckaert. (2008). Bayesian network classifiers in Weka for version 3-5-7. *Artif. Intell. Tools*. **11**, 369–387.

[160] R. Ho. (2013). *Handbook of Univariate and Multivariate Data Analysis with IBM SPSS*. 2 edition. Boca Raton: Chapman and Hall/CRC.

[161] B. W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta BBA - Protein Struct.* **405(2)**, 442–451.

[162] A. B. A. Graf and S. Borer. (2001). Normalization in support vector machines. *Proc. DAGM 2001 Pattern Recognition*. 277–282.

[163] D. M. Powers. (2011). Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation.

APPENDICIES

APPENDIX A – Ethics Committee Decision Papers

GAZİANTEP ÜNİVERSİTESİ KLİNİK ARAŞTIRMALAR ETİK KURULU KARAR FORMU

ARAŞTIRMANIN AÇIK ADI	Veri Madenciliği Kullanarak Üçlü Testin Sınıflandırma Performansının İyileştirilmesi
VARSA ARAŞTIRMANIN PROTOKOL KODU	310

ETİK KURUL BİLGİLERİ		
	ETİK KURULUN ADI	Gaziantep Üniversitesi Klinik Araştırmalar Etik Kurulu
	AÇIK ADRESİ:	Gaziantep Üniversitesi Sağlık Bilimler Fakültesi 2. Kat Şehitkamil/Gaziantep
	TELEFON	0342 360 07 53 /-77704
	FAKS	0342 360 39 27
	E-POSTA	gaunetikkurul@gmail.com

BAŞVURU BİLGİLERİ					
	KOORDİNATÖR/SORUMLU ARAŞTIRMACI UNVANI/ADI/SOYADI	Yrd .Doç.Dr. Alptekin Durmuşoğlu			
	KOORDİNATÖR/SORUMLU ARAŞTIRMACININ UZMANLIK ALANI	Endüstri Mühendisliği Bölümü			
	KOORDİNATÖR/SORUMLU ARAŞTIRMACININ BULUNDUĞU MERKEZ	Gaziantep Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği Bölümü			
	VARSA İDARİ SORUMLU UNVANI/ADI/SOYADI				
	DESTEKLEYİCİ				
	PROJE YÜRÜTÜCÜSÜ UNVANI/ADI/SOYADI (TÜBİTAK vb. gibi kaynaklardan destek alanlar için)				
	DESTEKLEYİCİNİN YASAL TEMSİLCİSİ				
	ARAŞTIRMANIN FAZI VE TÜRÜ	FAZ 1 ☐			
		FAZ 2 ☐			
		FAZ 3 ☐			
		FAZ 4 ☐			
		Gözlemsel ilaç çalışması ☐			
		Tıbbi cihaz klinik araştırması ☐			
		İn vitro tıbbi tanı cihazları ile yapılan performans değerlendirme çalışmaları ☐			
		İlaç dışı klinik araştırma ☐			
		Diğer ise belirtiniz :			
	ARAŞTIRMAYA KATILAN MERKEZLER	TEK MERKEZ ☐	ÇOK MERKEZLİ ☐	ULUSAL ☐	ULUSLARARASI ☐

DEĞERLENDİRİLEN BELGELER	Belge Adı	Tarihi	Versiyon Numarası	Dili
	ARAŞTIRMA PROTOKOLÜ			Türkçe ☐ İngilizce ☐ Diğer ☐
	BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU			Türkçe ☐ İngilizce ☐ Diğer ☐
	OLGU RAPOR FORMU			Türkçe ☐ İngilizce ☐ Diğer ☐
	ARAŞTIRMA BROŞÜRÜ			Türkçe ☐ İngilizce ☐ Diğer ☐
DİĞER İLİLENDİRİLEN	Belge Adı		Açıklama	

Etik Kurul Başkanının
Unvanı/Adı/Soyadı: Prof. Dr. Belgin ALAŞEHİRLİ
İmza:

Not: Etik kurul başkanı, imzasının yer almadığı her sayfaya imza atmalıdır.

GAZİANTEP ÜNİVERSİTESİ KLİNİK ARAŞTIRMALAR ETİK KURULU KARAR FORMU

ARAŞTIRMANIN AÇIK ADI	Veri Madenciliği Kullanarak Üçlü Testin Sınıflandırma Performansının İyileştirilmesi
VARSA ARAŞTIRMANIN PROTOKOL KODU	310

	SİGORTA	☐	
	ARAŞTIRMA BÜTÇESİ	☐	
	BİYOLOJİK MATERYEL TRANSFER FORMU	☐	
	İLAN	☐	
	YILLIK BİLDİRİM	☐	
	SONUÇ RAPORU	☐	
	GÜVENLİLİK BİLDİRİMLERİ	☐	
	DİĞER:	☐	
KARAR BİLGİLERİ	**Karar No:2016 /310**	**Tarih: 28.11 .2016**	
	Yukarıda bilgileri verilen başvuru dosyası ile ilgili belgeler araştırmanın/çalışmanın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş ve uygun bulunmuş olup araştırmanın/çalışmanın başvuru dosyasında belirtilen merkezlerde gerçekleştirilmesinde etik ve bilimsel sakınca bulunmadığına toplantıya katılan etik kurul üye tam sayısının salt çoğunluğu ile karar verilmiştir. İlaç ve Biyolojik Ürünlerin Klinik Araştırmaları Hakkında Yönetmelik kapsamında yer alan araştırmalar/çalışmalar için Türkiye İlaç ve Tıbbi Cihaz Kurumu'ndan izin alınması gerekmektedir.		

KLİNİK ARAŞTIRMALAR ETİK KURULU	
ETİK KURULUN ÇALIŞMA ESASI	İlaç ve Biyolojik Ürünlerin Klinik Araştırmaları Hakkında Yönetmelik, İyi Klinik Uygulamaları Kılavuzu
BAŞKANIN UNVANI / ADI / SOYADI:	Prof. Dr.Belgin ALAŞEHİRLİ

Unvanı/Adı/Soyadı	Uzmanlık Alanı	Kurumu	Cinsiyet		Araştırma ile ilişki		Katılım *		İmza
Prof. Dr.Belgin ALAŞEHİRLİ	FARMAKOLOJİ	Gaziantep Üniversitesi Tıp Fakültesi	E☐	K x ☐	E☐	H x☐	E☐	H☐	
Prof.Dr. Mehmet KESKİN	PEDİATRİ	Gaziantep Üniversitesi Tıp Fakültesi	E x ☐	K☐	E☐	H x ☐	E☐	H☐	
Prof. Dr Feridun IŞIK	GÖĞÜS CERRAHİ	Gaziantep Üniversitesi Tıp Fakültesi	E x ☐	K☐	E☐	H x ☐	E☐	H☐	
Prof. .Dr. İlker SEÇKİNER	ÜROLOJİ	Gaziantep Üniversitesi Tıp Fakültesi	E x ☐	K☐	E☐	H x ☐	E☐	H☐	
Prof. Dr. Ramazan BAL	FİZYOLOJİ	Gaziantep Üniversitesi Tıp Fakültesi	E x ☐	K☐	E☐	H x ☐	E☐	H☐	
Prof. .Dr.Yasemin ZER	MİKROBİYOLOJİ	Gaziantep Üniversitesi Tıp Fakültesi	E☐	K x ☐	E☐	H x☐	E☐	H☐	
Doç. Dr. Zeynel Abidin ÖZTÜRK	İÇ HASTALIKLARI	Gaziantep Üniversitesi Tıp Fakültesi	E x ☐	K☐	E☐	H x ☐	E☐	H☐	
Doç.Dr.Seval KUL	BİYOİSTATİSTİK	Gaziantep Üniversitesi Tıp Fakültesi	E☐	K x ☐	E☐	H x ☐	E☐	H☐	
Yrd. Doç. Dr Betül TAŞ	AĞIZ DİŞ ve ÇENE CERRAHİSİ	Gaziantep Üniversitesi Diş Hekimliği Fakültesi	E☐	K x ☐	E☐	H x ☐	E☐	H☐	
Uzm.Dr. Cahide Elif ORHAN	FARMAKOLOJİ	Gaziantep İl sağlık Müdürlüğü	E☐	K x ☐	E☐	H x☐	E☐	H☐	
Eyüp ÇELİK	AVUKAT	Gaziantep Barosu	E x ☐	K☐	E☐	H x ☐	E☐	H☐	
İrem ELBEYLİ	MİMAR	Gaziantep Büyükşehir Belediyesi	E☐	K x ☐	E☐	H x☐	E☐	H☐	

*:Toplantıda Bulunma

Etik Kurul Başkanının
Unvanı/Adı/Soyadı: Prof. Dr. Belgin ALAŞEHİRLİ
İmza:

Elden teslim aldım
Memet Mehmet

Not: Etik kurul başkanı, imzasının yer almadığı her sayfaya imza atmalıdır.

APPENDIX B - Samples Of Amniocentesis And Triple Test Reports

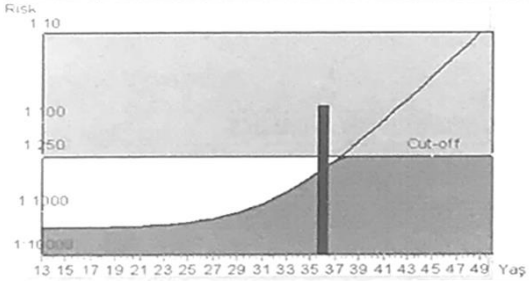
Gaziantep Üniversitesi Tıp Fakültesi
Biyokimya Laboraturi

DOWN SENDROMU TARAMA SONUÇLARI

İsim		Örnek No		Diyabet	Hayır
		Hasta No		Fetus sayısı	1
Serum alma tarihi	11.03.2015	Dogum Tarihi	1975	Sigara	Hayır
Rapor tarih	12.03.2015	Dogumdaki yaş	36.0	IVF	Hayır
Önceki trisomy 21 hamilelikleri	bilinmiyor	Agırlık(kg	58	Etnik köken	Beyaz

DÜZELTİLMİŞ MoM'LAR ve RİSKLER

AFP	30.1	IU/ml	0.88	Düzeltilmiş	MoM	Örnek alınma tarihindeki gebelik yaş	16 + 0
uE3	0.386	ng/ml	0.60	Düzeltilmiş	MoM	Tarama metodu	BPD Hadlock
HCG	71080	mIU/ml	2.56	Düzeltilmiş	MoM	Doktor	



Tr.21 riski
doğumdaki
1:66

Yaş riski
doğumdaki
1: 347

DOWN SENDROMU

Hesaplanmış Trisomy 21 riski cut off değerinin üzerindedir ve bu değer artan riski temsil etmektedir.
Trisomy 21 sonucuna göre .aynı degerleri gösteren 66 kadından sadece 1'inde trisomy 21 li gebeliğe rastlanmıştır. 65 kadında ise normal gebelik görülmüştür.
Yüksek HCG seviyesi şüphelidir.
Hesaplanan tüm değerler istatistiksel risk degerleri olup tarama amacı ile kullanılmaktadır Bu bir teşhis programı değildir
Bu sonuçların istatistiksel sonuçlar olduğu ve teşhis değeri olmadığı unutulmamalıdır.

NÖRAL TÜP DEFEKT	TRİSOMY 18
Nöral tüp defekt için düzeltilmiş AFP MoM değeri (0,88) risksiz bölgede bulunmaktadır.	**Hesaplanan Trisomy 18 riski < 1:10000 dir. Trisomy 18 için istatistiksel bir risk tespit edilmemiştir.**

Cut off değerinin altında | Cut off değerinin altında, Yaş Riskinin üzerind | Cut off değerinin üzerinde | Prisca 5.0.2.37

T.C.
GAZİANTEP ÜNİVERSİTESİ
ŞAHİNBEY ARAŞTIRMA VE UYGULAMA HASTANESİ

TIBBİ BİYOLOJİ ANABİLİM DALI
MOLEKÜLER GENETİK TANI, HEMATOLOJİ VE DOKU TİPLENDİRME LABORATUVARI

Adı Soyadı	:	Rapor Tarihi	: 03.04.2015 12:39
T.C Kimlik No	:	Dosya no	:
Baba Adı	:	Başvuru No	:
Kurumu	: SOSYAL GÜVENLİK KURUMU BŞK.GAZİ	Doğum Yeri - Tarih	: ADIYAMAN - 1979
İstem Tarihi	: 13.03.2015	İstem Kabul Tarihi	: 03.04.2015
Hizmet Adi	: AMNIYON SIVISINDAN KROMOZOM ANALIZI		: 03.04.2015 09:01
Yaş	: 35 Cinsiyet : KADIN		
İstem Bölüm	: KADIN HASTALIKLARI PLK	İstem Doktor	:

Tanı :	Kodu	Adı
	Z33	GEBELİK DURUMU

AMNİOSENTEZ MATERYALİ SİTOGENETİK ANALİZ RAPORU

İncelenen Materyal	: Amniyon Sıvısı
Uygulanan Bantlama Yöntemi	: Konvansiyonel Sitogenetik,G bandlama
Analiz Edilen Bant Sayısı	: 450-550
İncelenen Metafaz Sayısı	: 20

SONUÇ : 47,X*,+21(Trizomi 21)

Açıklama : Amniyosentez materyalinden hazırlanan kültürlerden elde edilen metafazlarda 47,X*,+21 saptanmıştır.Saptanan Trizomi 21 kromozom kuruluşu Down Sendromu'na yol açmaktadır.Genetik danışmanlık verilmesi önerilir.

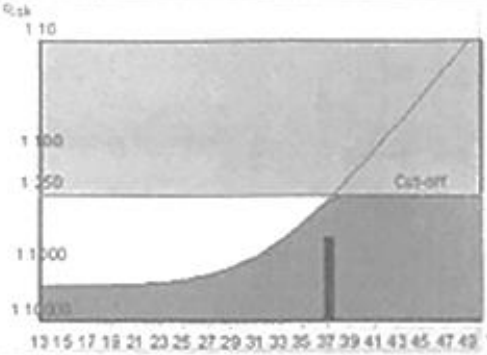
Gaziantep Üniversitesi Tıp Fakültesi
Biyokimya Laboratuarı

DOWN SENDROMU TARAMA SONUÇLARI

İsim		Örnek No		Diyabet	Hayır
		Hasta No		Fetus sayısı	1
Serum alma tarihi	26.09.2012	Doğum Tarihi	1976	Sigara	Hayır
Rapor tarihi	26.09.2012	Doğumdaki yaş	37.2	IVF	Hayır
Önceki trisomy 21 hamilelikleri	bilinmiyor	Ağırlık(kg)	74	Etnik köken	Beyaz

DÜZELTİLMİŞ MoM'LAR ve RİSKLER

AFP	26.2	IU/ml	0.81	Düzeltilmiş	MoM	Örnek alınma tarihindeki gebelik yaşı	17 +0
uE3	0.74	ng/ml	0.69	Düzeltilmiş	MoM	Tarama metodu	BPD
hCG	22731	mIU/ml	1.06	Düzeltilmiş	MoM	Doktor	



Tr.21 riski
doğumdaki
1:575

Yaş riski
doğumdaki
1: 261

DOWN SENDROMU

Hesaplanmış Trisomy 21 riski cut off değerinin altındadır ve bu değer risksiz bölgede bulunmaktadır.
Trisomy 21 sonucuna göre ,aynı değerleri gösteren 575 kadından sadece 1'inde trisomy 21 li gebeliğe rastlanmıştır. 574 kadında ise normal gebelik görülmüştür.
Hesaplanan tüm değerler istatistiksel risk değerleri olup tarama amacı ile kullanılmaktadır Bu bir teşhis programı değildir
Bu sonuçların istatistiksel sonuçlar olduğu ve teşhis değeri olmadığı unutulmamalıdır.

NÖRAL TÜP DEFEKT	TRİSOMY 18
Nöral tüp defekt için AFP (MoM: 0.81) değeri risksiz bölgede bulunmaktadır.	**Hesaplanan Trisomy 18 riski < 1:10000 dir. Trisomy 18 için istatistiksel bir risk tespit edilmemiştir.**

Cut off değerinin altında Cut off değerinin altında, Yaş Riskinin üzerinde Cut off değerinin üzerinde Prisca 5.0.1.0

T.C.
GAZİANTEP ÜNİVERSİTESİ
ŞAHİNBEY ARAŞTIRMA VE UYGULAMA HASTANESİ

TIBBİ BİYOLOJİ ANABİLİM DALI
MOLEKÜLER GENETİK TANI, HEMATOLOJİ VE DOKU TİPLENDİRME LABORATUVARI

Adı Soyadı	:	Rapor Tarihi	: 19.10.2012 09:45
T.C Kimlik No	:	Dosya no	:
Baba Adı	:	Başvuru No	:
Kurumu	: SOSYAL GÜVENLİK KURUMU BŞK GAZI	Doğum Yeri - Tarih	: OĞUZELİ - 1976
İstem Tarihi	: 27.09.2012	İstem Kabul Tarihi	: 27.09.2012
Hizmet Adi	: AMNİOTİK MAYİİ KÜLTÜRÜ		: 27.09.2012 16:49
Yaş	: 36 Cinsiyet : KADIN		
İstem Bölüm	: KADIN HASTALIKLARI PLK	İstem Doktor	:

Tanı :

Kodu	Adı
Z33	GEBELİK DURUMU

AMNİOSENTEZ MATERYALİ SİTOGENETİKANALİZ RAPORU

İncelenen Materyal : Amniosentez materyali

Uygulanan bantlama yöntemi : GTG-Bantlama

İncelenen metafaz sayısı : 20

SONUÇ : 47,X*,+21 (Trizomi 21)

NOT : Amnion sıvısından hazırlanmış kültürlerden elde edilen metafaz plaklarında 47,X*,+21 kromozom kuruluşu saptanmıştır.

APPENDIX C

Section A

The Initial Data Set of Patients

Attributes	Age	Weight	GestationalAge	Smoking	IVF	Ethnicity	AFP	AFPMoM	uE3	uE3MoM	hCG	hCGMoM	Result
Type of Attribute	Numeric	Numeric	Numeric	Nominal	Nominal	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
Unit of Attribute	Years	Kg	Days	Yes/No	Yes/No	Black/White	IU/ml	-	ng/ml	-	mIU/ml	-	Trisomy21/Negative
Number of Values	81	66	81	81	81	81	81	81	81	81	81	81	81
Number of Categorical Values	-	-	-	1 Yes 80 No	0 Yes 81 No	0 Black 81 Black	-	-	-	-	-	-	5 Trisomy21 76 Negative
Number of Missing Values	0	15	0	0	0	0	0	0	0	0	0	0	0
Minumum Value	18.2	45	112	0	0	0	10	0.34	0.208	0.19	3508	0.18	0
Maximum Value	45.2	100	137	0	0	0	83.4	2.49	5.73	2.17	100000	4.37	0
Mode	23.9	60	116	-	-	-	30.9	0.83	2.73	0.57	-	1.07	-
Median	34.5	65	118	-	-	-	26.6	0.74	0.723	0.75	30683	1.34	-
Mean	33.89	67.30	119.83	-	-	-	29.70	0.82	1.12	0.85	35217.32	1.53	-
Standard Deviation	6.24	11.90	5.82	-	-	-	15.02	0.40	1.02	0.40	21546.09	0.90	-
Variance	38.42	139.48	33.48	-	-	-	222.91	0.16	1.04	0.16	458502561	0.80	-

Section B

Data Set of Patients with Completed Missing Values

Attributes	Age	Weight	GestationalAge	Smoking	IVF	Ethnicity	AFP	AFPMoM	uE3	uE3MoM	hCG	hCGMoM	Result
Type of Attribute	Numeric	Numeric	Numeric	Nominal	Nominal	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
Unit of Attribute	Years	Kg	Days	Yes/No	Yes/No	Black/White	IU/ml	-	ng/ml	-	mIU/ml	-	Trisomy21/Negative
Number of Values	81	81	81	81	81	81	81	81	81	81	81	81	81
Number of Categorical Values	-	-	-	1 Yes 80 No	0 Yes 81 No	0 Black 81 Black	-	-	-	-	-	-	5 Trisomy21 76 Negative
Number of Missing Values	0	0	0	0	0	0	0	0	0	0	0	0	0
Minumum Value	18.2	45	112	0	0	0	10	0.34	0.208	0.19	3508	0.18	0
Maximum Value	45.2	100	137	0	0	0	83.4	2.49	5.73	2.17	100000	4.37	0
Mode	23.9	67.30303	116	-	-	-	30.9	0.83	2.73	0.57	-	1.07	-
Median	34.5	67.30303	118	-	-	-	26.6	0.74	0.723	0.75	30683	1.34	-
Mean	33.89	67.30	119.83	-	-	-	29.70	0.82	1.12	0.85	35217.32	1.53	-
Standard Deviation	6.24	10.73	5.82	-	-	-	15.02	0.40	1.02	0.40	21546.09	0.90	-
Variance	38.42	113.65	33.48	-	-	-	222.91	0.16	1.04	0.16	458502562	0.80	-

Section C

Balanced Data Set of Patients

Attributes	Age	Weight	GestationalAge	Smoking	IVF	Ethnicity	AFP	AFPMoM	uE3	uE3MoM	hCG	hCGMoM	Result
Type of Attribute	Numeric	Numeric	Numeric	Nominal	Nominal	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
Unit of Attribute	Years	Kg	Days	Yes/No	Yes/No	Black/White	IU/ml	-	ng/ml	-	mIU/ml	-	Trisomy21/Negative
Number of Values	126	126	126	126	126	126	126	126	126	126	126	126	126
Number of Categorical Values	-	-	-	1 Yes 80 No	0 Yes 81 No	0 Black 81 Black	-	-	-	-	-	-	50 Trisomy21 76 Negative
Number of Missing Values	0	0	0	0	0	0	0	0	0	0	0	0	0
Minumum Value	18.2	45	112	0	0	0	10	0.34	0.208	0.19	3508	0.18	0
Maximum Value	45.2	100	137	0	0	0	83.4	2.49	5.73	2.17	100000	4.37	0
Mode	23.9	67.30303	116	-	-	-	30.9	0.83	2.73	0.57	-	1.07	-
Median	34.35	67.30303	118.1997785	-	-	-	26.4094645	0.730331	0.664727	0.710629	36164	1.5981665	-
Mean	34.23	68.38	119.43	-	-	-	28.04	0.79	0.94	0.79	39303.63	1.69	-
Standard Deviation	5.17	9.30	5.09	-	-	-	12.45	0.33	0.86	0.34	20175.55	0.82	-
Variance	26.52	85.82	25.67	-	-	-	153.72	0.11	0.73	0.11	403822158	0.67	-

Section D

Data Set of Patients after the Elimination

Attributes	AFP	uE3	hCG	Result
Type of Attribute	Numeric	Numeric	Numeric	Nominal
Unit of Attribute	IU/ml	ng/ml	mlU/ml	Trisomy21/Negative
Number of Values	126	126	126	126
Number of Categorical Values	-	-	-	50 Trisomy21 76 Negative
Number of Missing Values	0	0	0	0
Minumum Value	10	0.208	3508	0
Maximum Value	83.4	5.73	100000	0
Mode	30.9	2.73	-	-
Median	26.41	0.66	36164	-
Mean	28.04	0.94	39303.63	-
Standard Deviation	12.45	0.86	20175.55	-
Variance	153.72	0.73	403822157.62	-