COMPARATIVE ANALYSIS OF CLASSIFICATION MODELS FOR PROSTATE CANCER

By

KHALED SAAD S ALQAHTANI

A Dissertation Submitted to

Rutgers, The State University of New Jersey

Rutgers Biomedical and Health Sciences

School of Health Related Professions

In partial fulfillment of the Requirements

For the Degree of

Doctoral of Philosophy

Department of Health Informatics

December, 2014

RUTGERS
BIOMEDICAL AND HEALTH SCIENCES

Final Dissertation Approval Form

Comparative Analysis of Classification Models for Prostate Cancer

BY

Khaled Saad Alqahtani

Dissertation Committee:

Shankar Srinivasan PhD, Committee Chair

Syed Haque PhD, Committee Member

Dinesh Mital PhD, Committee Member

Approved by the Dissertation Committee:

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

©2014

Khaled Saad S. Alqahtani

ALL RIGHTS RESERVED

ABSTRACT

COMPARATIVE ANALYSIS OF CLASSIFICATION MODELS FOR PROSTATE CANCER

By

KHALED SAAD S ALQAHTANI

Among different types of cancers which occur in men, prostate cancer is the most commonly occurring one. However, prostate cancer epidemiology is not completely identified. Neither the causation nor pathogenesis of prostate cancer can be totally understood by today's information. As a result, prostate cancer screening tests cannot always detect the disease. Prostate-specific antigen (PSA) blood test is the most widely used test to screen men for prostate cancer. However, PSA blood test adversaries accuse this test as rendering misguided results that lead to over-diagnosis and overtreatment. According to The National Cancer Institute, men who go through a prostate biopsy procedure because of an elevated PSA test result, only about 25 percent of them actually have prostate cancer. The other 75% of men might face the side effects of prostate biopsy which includes serious infections, pain, and bleeding.

This study utilized the Nationwide Inpatient Sample (NIS), and the Surveillance, Epidemiology, and End Results (SEER) Program data to identify some key risk factors for those patients who are more likely to be diagnosed with prostate cancer. In addition, an Artificial Neural Network has been implemented a long with the most used classification methods that include Logistic Regression, k-Nearest Neighbors, Naïve Bayes classifier, Decision Tree classifier, and Support Vector Machine, in order to

recognize prostate cancer in an early stage. All these classification methods' results were analyzed using confusion matrix and Receiver Operating Characteristic (ROC) analyses.

This study found that age, ethnicity, family history of cancer, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension are positively associated with prostate cancer. Although, obesity, alcohol abuse, and smoking were significantly associated with the prostate cancer, this association found to be negative. The result of classification methods' tests showed that the Artificial Neural Network had success rates of 87.53% on NIS data and 99.31% on SEER data compared to Logistic Regression (81.71%, 84.95%), k-Nearest Neighbors (73.46%, 91.62%), Naïve Bayes classifier (70.86%, 86.56%), Decision Tree classifier (78.02%, 90.34%), and Support Vector Machine (72.33%, 88.52%).

In conclusion, this study tried to minimize the PSA false result by identifying more key risk factors and providing a prediction tool based on Artificial Neural Network to predict and to support the clinical decision in prostate cancer screening.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the invaluable support and guidance of my advisor Dr. Shankar Srinivasan. I am very thankful to him for his endless support.

I would like to thank the members of my committee Dr. Syed Haque, Dr. Dinesh Mital and Dr. Masayuki Shibata for their kindness, supportive, and useful advice during my dissertation process.

I would also like to thank all the faculty and staff of the Department of Health Informatics for their kindness and support, especially Dr. Syed Haque for his outstanding leadership. Thank you, Dr. Haque, for the support and kindness.

Finally, I would like to thank my family for their endless love and support throughout my life.

To

Whom I care most

My beloved father and my sweet mother

I love you

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

I. INTRODUCTION

Cancer is a major cause of morbidity and mortality worldwide. In 2012, around 14 million new cases were registered and 8 million cancer-related deaths across the world [1]. Among men, the five most common sites of cancer diagnosed in 2012 were the lung (16.7% of the total), prostate (15.0%), colorectum (10.0%), stomach (8.5%), and liver (7.5%) [1]. Prostate Cancer had the second highest incidence (31.1 per 100 000) worldwide [1]. However, in the United States, prostate cancer is the most common cancer in men. With an aging male population, the 2014 new cases estimated for prostate cancer is 233,000 and 29,480 deaths [2]. As shown in Figure 1, prostate cancer is number one in the top ten cancer sites for men in the United States for all races.

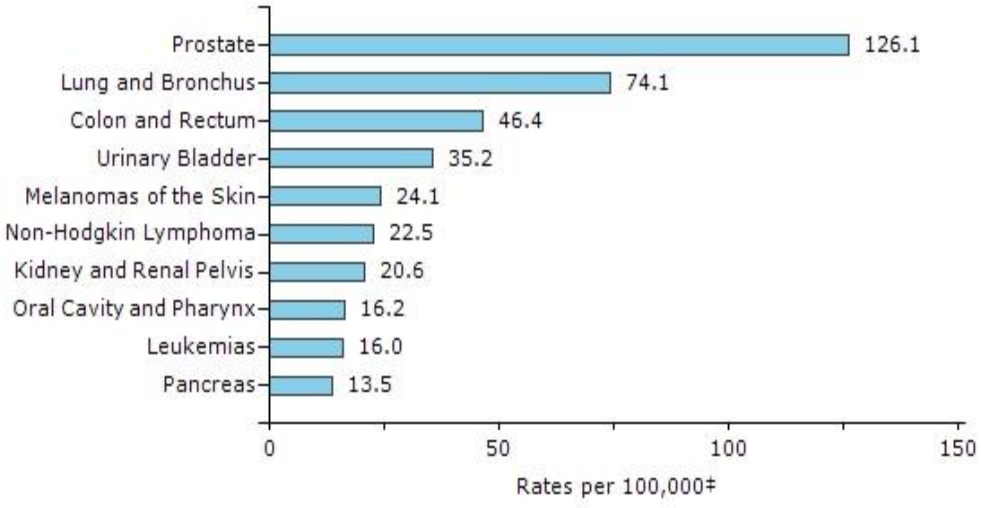

Figure 1: Top 10 Cancer Sites: 2010, Male, United States - All Races [3]

Early detection of prostate cancer by using the prostate-specific antigen (PSA) blood test has increased the number of new incidents. However, prostate cancer death rates are showing a significant decline of 40 percent since 1993. Many experts believe that using prostate-specific antigen (PSA) blood test is the main reason for the declining death rates as shown in Figure 2.

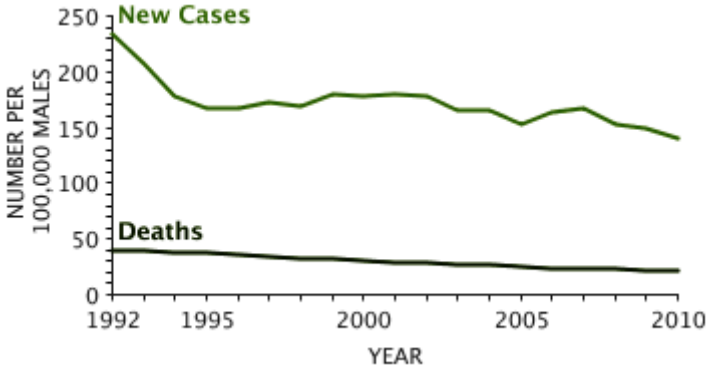

Figure 2: Number of New Cases and Deaths per 100,000, in the United States [4]

According to [4] and [5] a male born today faces a risk of 15.3 percent of being diagnosed with prostate cancer at some point during his lifetime. In addition, he has a 3 percent chance of dying of prostate cancer. Despite of these numbers, a statistical numbers from a large national health database indicate that a significant number of men who have been diagnosed with prostate cancer are more likely to live as long as men in the same age group who have not been diagnosed with prostate cancer.

Prostate cancer is more common in older men than younger men. Almost 60 percent of new cases occur at age of 65 years old and older [4]. Men of African American descent have the highest numbers of incidents and death rates [4]. Also, men with a positive family history of prostate cancer are more likely to be diagnosed with prostate cancer.

1.1 Background and Statement of the Problem

Among different types of cancers which occur in men, prostate cancer is the most commonly occurring one. However, prostate cancer epidemiology is not completely identified. Neither the causation nor pathogenesis of prostate cancer can be totally understood by today's information. As a result, prostate cancer screening tests cannot always detect the disease. Prostate-specific antigen (PSA) blood test is the most widely used test to screen men for prostate cancer. Since it became available 20 years ago, the effect of the PSA blood test on prostate cancer screening has been more noticeable through the rise of the total of new prostate cancer incidences.

The cells of the prostate gland produce a serum protein called Prostate-Specific Antigen (PSA) which is released into the blood. The PSA test measures the serum PSA level in a patient's blood. The normal level of PSA is 4.00 nanograms per milliliter and lower, but it might be different from one patient to another. If the PSA level is above 4.00 ng/mL, physicians might request a prostate biopsy in order to determine the existent of cancer.

However, PSA blood test adversaries accuse this test as rendering misguided results that lead to over-diagnosis and overtreatment. According to [6], men who go through a prostate biopsy procedure because of an elevated PSA test result, only about 25 percent of them actually have prostate cancer. The other 75% of men might face the side effects of prostate biopsy which includes serious infections, pain, and bleeding.

1.2 Objectives and Significance of the Research

The main objectives of this study were, first, to identify the key risk factors of prostate cancer; and then, develop an artificial neural network model for early prediction of this

disease; and to examine whether an artificial neural network model can effectively predict the likelihood of prostate cancer. Five years (2007, 2008, 2009, 2010, and 2011) of the Nationwide Inpatient Sample (NIS) data were used to investigate the association of these factors with prostate cancer. In addition, the Surveillance, Epidemiology, and End Results (SEER) Program data was also used to test an artificial neural network model based on age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size variables.

This study tried to identify some key risk factors for those patients who are more likely to be diagnosed with prostate cancer. These risk factors were classified into four categories. First category, general factors include age, race/ethnicity. Genetic factors are the second category, which include family history of cancer. Third category, lifestyle factors include obesity, physical activity, smoking, alcohol abuse, dietary fat, and vitamin D deficiency. Last category is chronic disease and surgical procedure which include inflammation of prostate, vasectomy, and hypertension.

The aforementioned variables have been found in previous studies to affect the likelihood of having prostate cancer. From the years 2006 – 2010, the median age at diagnosis was 66, indicating that 50% of the men who were diagnosed during these years were 66 or younger when they developed prostate cancer [7]. This study is necessary to further build the research and provide implications and useful information about prostate cancer screening.

The research questions are:

- Is Artificial Neural Network a significant technique in predicting the risk of prostate cancer?
- Does Artificial Neural Network predict the risk of prostate cancer better than logistic regression the popular used technic for prediction or the other classification methods?
- Is Family history of any other type of cancer a significant factor in predicting the risk of prostate cancer?
- Is Obesity a significant factor in predicting the risk of prostate cancer?
- Is Physical activity a significant factor in predicting the risk of prostate cancer?
- Is Alcohol abuse a significant factor in predicting the risk of prostate cancer?
- Is Smoking a significant factor in predicting the risk of prostate cancer?
- Is Fatty acid Deficiency a significant factor in predicting the risk of prostate cancer?
- Is Cholesterol a significant factor in predicting the risk of prostate cancer?
- Is Vitamin D Deficiency a significant factor in predicting the risk of prostate cancer?
- Is Inflammation of Prostate a significant factor in predicting the risk of prostate cancer?
- Is Vasectomy a significant factor in predicting the risk of prostate cancer?
- Is Hypertension a significant factor in predicting the risk of prostate cancer?

1.3 Research Hypotheses

The hypotheses are:

1. Artificial Neural Network is a significant technique in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

2. Artificial Neural Network predicts the risk of prostate cancer better than logistic regression.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

3. Family history of any other type of cancer is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

4. Obesity is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

5. Physical Activity is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

6. Alcohol Abuse is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

7. Smoking is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

8. Fatty acid Deficiency is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

9. Cholesterol is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

10. Vitamin D Deficiency is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

11. Inflammation of Prostate is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

12. Vasectomy is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

13. Hypertension is a significant factor in predicting the risk of prostate cancer.

Null Hypothesis: $H_0 = H_1$

Alternative Hypothesis: $H_0 \neq H_1$

1.4 Study Limitation

The Nationwide Inpatient Sample (NIS) data set does not include PSA test result. Therefore, the Surveillance, Epidemiology, and End Results (SEER) Program data which include PSA test result was used beside the Nationwide Inpatient Sample (NIS) data. In addition, this study did not investigate some factors details such as alcohol consumption or smoking habit. This study relied on the diagnosis codes that registered on NIS data and SEER data only.

CHAPTER II

II. LITERATURE REVIEW

2.1 Prostate Cancer Overview

The prostate is part of the male's reproductive and urological systems. Its location is in front of the rectum, under the bladder, and behind the public bone. The prostate is a solid organ that surrounds the urethra which is connected to the bladder. The urethra carries urine and semen out of the body. Seminal vesicles are located behind the prostate and their function is to provide fluid for semen. The normal healthy prostate in younger men is about the size of a walnut (4×3×2cm3) and typically weighs 20-30 grams. Size and weight will vary with age [8].

The prostate has different types of cells. Gland cells which make the prostate fluid are the most common cells that develop cancer. In some rare situations, other cells like sarcomas, small cell carcinomas, and transitional cell carcinomas can develop a cancer.

Most prostate cancers grow slowly, but in some cases they can grow and spread quickly. According to the American Cancer Society, many older men (and even some younger men) who died of other diseases also had prostate cancer that never affected them during their lives. In many cases, neither they nor their doctors even knew they had it [8].

Prostate cancer cells look different under the microscope. Based on their abnormal shape, they can be classified into two types. The first type is called "Low-grade Prostatic intraepithelial neoplasia" which the patterns of prostate cells appear almost normal. The second type is "High-grade Prostatic intraepithelial neoplasia" which the patterns of cells

look more abnormal. In as early as their 20s, some men have shown prostatic intraepithelial neoplasia in their prostate [8]. By the age of 50, around 50 percent of men have prostatic intraepithelial neoplasia [8]. Prostate cancer can invade (metastasize) to different organs of the body such as rectum, lymph nodes, bones and bladder. If the cancer spreads to other organs, additional symptoms appear such as bone pain[9].

The diagnosis process for prostate cancer takes several steps before cancer has been determined. First, the patient's medical history and symptoms such as frequent urination, blood in the urine or painful urination, is reviewed. The second step is screening. There are two options for the screening: digital rectal exam (DRE) which is a physical exam that can be done by inserting a gloved finger into the patient's rectum to palpate hard areas in the prostate. The second screening option is a prostate-specific antigen (PSA) blood test. This test is designed to measure the level of PSA in the patient's blood. When the DRE and/or PSA test result show that the patient may have prostate cancer, a biopsy from the prostate is necessary to confirm the cancer.

The clinical stages of prostate cancer are classified into four stages. The clinical stage tells how much the cancer may have grown within the prostate and whether it has spread to other tissues or organs. In the first two stages, the cancer is undetectable during a DRE exam. The third stage is the most often found when a prostate biopsy is done because of a PSA test result that showed a high PSA blood level. This is the most commonly diagnosed stage of prostate cancer. The fourth stage means that prostate cancer can be felt during a DRE, but is still only in the prostate.

2.2 Prostate Cancer Screening

The screening of prostate cancer is very vital tool towards the prevention of secondary prostate cancer. During annual screening, questions are asked about urination, prostate-specific antigen (PSA) blood tests are evaluated and digitally examinations of the rectum (DRE) are carried out. The American Cancer Society and the American Urological Association presently have made recommendations for males who seek information and desire that their condition be evaluated in order to increase life expectancy.

Recommendations made by American Cancer Society suggest that screening should be initiated at the age of 50 years for common men, 45 for those who have increased risk factors. And 40 for those who have risks associated because of hereditary causes. Guidelines presented by American Cancer Society in the year 2009 showed prominent variations from previous suggestions rectum examination was made an option, those who had PSA of more than 2.5 nanograms per millimeters should go through screening on a yearly basis and those who have PSA less than the above-mentioned value should go through screening every two years. Also evaluations of risks on an individual basis must be a component of the referred decisions for patients with PSA in the range 2.5 to 4.0 nanogram per millimeter [10].

There are some present who also oppose prostate cancer screening. They say that it generally carries out the detection of irrelevant cancers and doesn't produce an influence upon entire survival and can produce adverse reactions upon a patient's quality of life. In place of regular prostate cancer screening, adversaries suggest making this an option.

Supporters of prostate cancer screening share the opinion that major number of patients who are detected with an elevated PSA are clinically significant and if this is not given proper treatment then life can be at risk [11]. The information which is being presented suggests that regular screening brings an improvement in survival [11].

Two larger control studies were carried out for determining the information which supports prostate cancer screening. The studies provided various results. 7 or 10 years Prostate, Lungs, Colorectal, and Ovarian (PLCO) screening trial study failed to show reduction in deaths related with prostate cancer in males who went through prostate cancer screening [12]. On the other hand, European Randomized Study for Prostate Cancer (ERSPC) found that deaths in screening arm is less than control arm by 20% after mean of 8.8 years [13]. For preventing a single death because of prostate cancer, 1410 males will require screening and in addition to this 48 males will require treatment. Also reduction in death rate associated with prostate cancer was noted for males who were in trials for twelve years. In addition to this incidences of tumors related with T3 & T4 were twenty two percent low and incidences of lesions of M1 were forty one percent less in ERSPC trials. Basic difference among two researches may be responsible for conflicting conclusions.

162000 males were studied in the ERSPC trials from 7 countries of Europe, the PLCO trial focused upon 76,693 males from the same country. Eighty five percent males who had biopsy indications in the trial of ERSPC showed acceptance towards biopsy of prostate. Contrary to this, 30% males in the screening arm of the research of PLCO with irregular level of PSA went with the biopsy of prostate. Also in the trials of PLCO 52% of males in control arm passed from Screening of PSA. This can provide an explanation

why incidences of death from the prostate cancer were not significant among the control group and screening group.

Information from the United States is congruent with the results of ERSPC trials. The Surveillance, Epidemiology, and End Results (SEER) data show that the new cases of prostate cancer have reduced more than three times ever since prostate cancer screening was introduced in the early 1990s. In addition, the American Cancer Society data indicate that mortality from the prostate cancer has reduced to fifty percent during the past twenty years.

Table 1: PSA Thresholds for age and race referance

Age Group	Black	White	Asian
40–49	0–2	0–2.5	0–2
50–59	0–4	0–3.5	0–3
60–69	0–4.5	0–4.5	0–4
70–79	0–5.5	0–5.5	0–5

Usual levels of PSA can vary with the age and race as shown in Table 1 [14]. Volume of prostate and reinforcements of the testing of PSA like the PSA free from percentage and velocity PSA with age and race all together can play a vital role in identifying males with higher risks of developing prostate cancer [11].

Ratio among overall PSA and the unbound PSA is measured by percentage free PSA. This kind of testing is vital for the values of PSA in 4.0 and10 nanogram per millimeter lower percentage indicates higher risk of the prostate cancer [15]. The determination of

PSAV is carried out by evaluating three values of PSAV with the gap of six months [11]. The calculation of PSAV in continuous mean of alteration in 3 visits is done in accordance with the formula which follows [16]:

$$\frac{1}{2} \times \left(\left[\frac{(PSA_2 - PSA_1)}{elapsed\ time\ in\ years} \right] + \left[\frac{(PSA_3 - PSA_2)}{elapsed\ time\ in\ years} \right] \right) \quad (1)$$

The [17] screening based on PSA with four years gap shows the outcome in the death rate associated with decrease of 32 percent prostate cancer related deaths in males who were between 55 to 69 years of age at the start of testing. The [18] study compared the influence of screening for a male selecting to undergo screening on a repeated basis with the PSA testing in greater than the influence produced by larger random trials. That gives estimates of the level of population.

The [19] study suggested that prostate cancer is crucial for the health related issues and it has been deduced through various researches that screening is vital in bringing reduction in the risks that are related with malignancy related pathologies and deaths associated with the prostate cancer. It is important to take steps for addressing the issues related with wrong evaluation and wrong treatment of pathologies prior to screening plans based on population can be endorsed.

2.3 prostate cancer risk factors

The fundamental cause of prostate cancer is unknown. However, cancer in general takes a long period of time to develop and produce a cancerous cell. This development classified as two-step process. First, there are initial factors that prompt the cell to be changed to a cancerous cell. Secondly, there are motivated and supported factors that

allow the cancerous cells to continue to grow and progress. These factors are under study and investigation and some of them are still unknown how they associate with cancer.

In prostate cancer, some factors such as age, race, family history, and lifestyle have been shown as risk factors to certain degree. These risk factors are classified to four main categories: General factors which include age and race, Genetic factors which are family history, and lifestyle factors that include obesity, smoking, alcohol abuse, physical activity, dietary fat, and vitamin D deficiency. Last category is chronic disease and surgical procedure which include inflammation of prostate, vasectomy, and hypertension. Each one of these categories will be discussed with more details as follow.

2.3.1 General factors

2.3.1.1 Age

Prostate cancer is a cancer that affects the older male population. According to [4] the average age at diagnosis is 66 years while 65 to 74 years are the most common diagnosed age as shown in Figure 3. Nevertheless, prostate cancer can occur to men of any age group. The Prostate Cancer Prevention Trial (PCPT) demonstrated that with an increase in age by 1 year the risk associated with it increases by 3% in the presence of Gleason score 8 to 10 prostate cancer subsequent to adjusting for identified risk factors of high-grade prostate cancer [20].

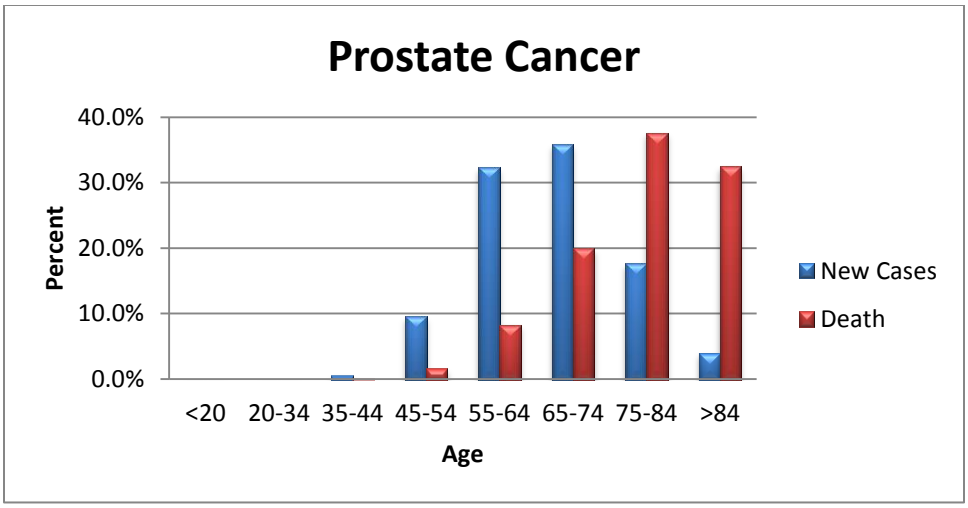

Figure 3: Prostate Cancer Percent of New Cases and Deaths by Age Group [4]

Occurrence and death of prostate cancer are both strongly related to age. The influence of age on this cancer is further made difficult by the race/ethnicity of the patient, which is another chief risk factor for prostate cancer. [21] and [22] declare that about 80% of men will develop cancerous cells in their prostate gland when they reach 80 years of age. It has been found via autopsy and research reports that around 70% to 90% of men above the age of 80 years have prostates with cancer undetected [23]; some type of cancers spread quickly, while most of the cancers show slow growth and spread.

Prostate cancer is more common in older male population and it becomes essential to undergo screening for its early detection and management as the age increases [24].

Surveillance, Epidemiology, and End Results (SEER) dataset has recently carried out an analysis of this cancer. This analysis specifies that following the use of the PSA blood test in the year of 1986 and onward, the high rate of growth of prostate cancer in men of 20 to 49 years of age was seen [25]. This study [25] recommended that this growth would be

the result of an extensive use of prostate-specific antigen (PSA) test for the purpose of screening.

Prostate cancer usually does not manifest itself via symptoms and signs in its early stages, and thus, people do not know of its existence in their body during this dormant period. The men of age 40 to 44 are of particularly at risk, [26] discovered an average marked increase in risk by 3.5% every year. Nonetheless, [27] identifies that certain subgroups of male in age group of 40 to 54 years may be at benefit because early screening can be done with resultant earlier detection of the cancer in them.

For instance, men at high risk for prostate cancer, like those people who have a strong family history or those who belong to African-American race, may get benefit from early detection, because of their increased chances of disease [28]. These particular men should be notified of both the recognized harms and the possible benefits of early screening, and mutual decision-making should proceed with the belief that there are no proportional data to show that men at increased risk than borderline for prostate cancer will benefit more from screening when compared to those at average risk [27].

With the increase in age, an increase in the level of Prostate-Specific Antigen (PSA) in the blood also increases; a high level of PSA may point towards the development of prostate cancer [29]. This test becomes specific with advancing age. A correlational research study performed by [30] discovered that with advancing age, men showed less eagerness to undergo PSA screenings. Additionally, men that have an apparently normal-sized prostate gland and their PSA levels are below 4 micrograms per liter, encompass a 15% risk of developing this cancer; men with a PSA level between 4 and 10 micrograms

per liter show a 25% risk of prostate cancer, and lastly, men with PSA greater than 10 is suggestive of a 67% risk of developing a prostate cancer [30].

Older men with numerous co-morbidities should be enlightened of their increased chance of death from other causes prior to getting a survival advantage from surgical procedures or radiotherapy for low- and intermediate-risk illness. In the USA, about 58% of men with diagnosed prostate cancer are above 75 years of age [7].

It is also identified that men of an increasing age will exhibit an increase in the size of the gland because of benign prostatic hyperplasia (BPH). This is very common, in about 70% of men with aged over 70 years [31]. The danger of prostate cancer-specific mortality raises with an increase of age are shown in men who have Gleason score 6 and 7, but not 8 to 10, prostate cancer [31].

Pertaining to the benefits and disadvantages of yearly screening in men of age group 40 to 55 years with a less risk for developing prostate cancer and the scarcity of lethal prostate cancers occurring in this age group, [27] does not advise routine screening. This does not suggest that there is totally no advantage of the PSA screening at this age group, instead that there are substantial issues related with this screening that the gains are not good enough to overshadow the problems [27].

[32] says that studies performed on prostate cancer in middle-aged male population are small in number, because more emphasis is given on findings involving men over age 60 years.

Currently, there is no definite information highlighting the etiology of prostate cancer present, though, research demonstrates that a positive family history of prostate cancer increases the chance of development of this cancer at an early age [32].

At present, more advanced research is needed to comprehensively become aware of the preventive measures, the etiology, the commencement, and the management of this cancer in people of age 40 and above. Men will age devoid of any prostate related issues if they become acquainted with information regarding the prostate health and its maintenance at an early age.

2.3.1.2 Race/Ethnicity

In the United States, prostate cancer claims men's lives more prolifically than any other cancer [6,33]. Although there are many protocols for prevention, detection, research, and treatments for prostate cancer, many elements remain undiscovered [23]. Regarding cancer and ethnicity in general, the death rate from cancer for African Americans is 34% higher than that of Caucasian Americans [23]. Death rates tabulated for the years between 1975 and 2005 revealed variation according to ethnicity [7]. For example, during this period, African American men were observed as those most likely to die from prostate cancer, specifically, above all other ethnic groups. Caucasian American men were recorded with the second highest mortality rate, followed by Native American, Hispanic and then Asian/Pacific Islander ethnicities [6,7].

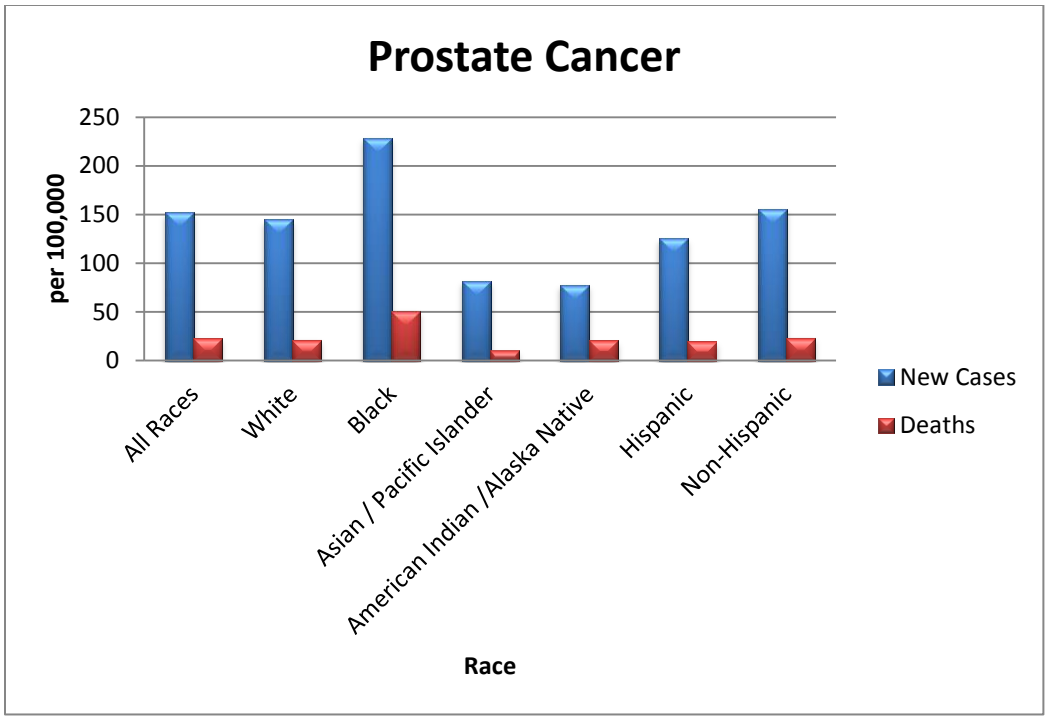

Figure 4: Prostate Cancer Number of New Cases and Deaths by Race/Ethnicity [4]

Race plays an important role in determining susceptibility to many diseases including cancer. There is enough evidence available to generate wide acceptance that ethnicity is a factor that is strongly associated with the likelihood of contracting prostate cancer [22,34,35] with thousands of men – particularly those of African American and Afro-Caribbean (most notably Jamaican) origin, falling victim to this disease every year [36].

Internationally, men who descend from sub-Saharan Africa are at a substantially higher risk of developing prostate cancer in their lifetime [1]. In the United Kingdom, the numbers of Black men who were diagnosed with prostate cancer are double all men from other races in the United Kingdom [34]. On the other hand, men of Asian descendants have the lowest probability of suffering from the disease [4]. Asian descendants who remain residents of Asia are much less likely to acquire prostate cancer than Asian males who

live in the United States [37], or than those men who are born in the United Kingdom. Furthermore Asians who migrate to the United States are less likely to be afflicted than Caucasian Americans [38].

According to the Center for Disease Control and Prevention (CDC), since 1975, African Americans have had and continue to have the highest incidence of prostate cancer - with Caucasian American men notable for the second highest incidence, followed by Latinos, Asian/Pacific Islander and Native American men [33].

Many studies including [39-41] find that African American men are less likely than Caucasian American men to be screened for prostate cancer. [42] reported that African American men are less likely to maintain regular screening practices, perhaps leading to a delay in diagnosis. Of those patients who did in fact present for screening. African Americans were found to be younger and indicated higher PSA values compared with other race patients [43].

Studies by [44,45] found that men are more likely to contract advanced prostate cancer due to lower socioeconomic status, which is associated with disparities in areas such as lack of access to healthcare, understanding of obesity issues, lack of insurance, low income, and lower levels of literacy and general education. However, data suggest that, even when controlling for these and other socioeconomic considerations, African Americans' ancestry remained an independent predictor of disease recurrence, poor prognosis (particularly in cases of obesity [46,47]) or mortality [48]. Moreover, African Americans are more likely to be diagnosed with a higher grade (degree of aggression) prostate cancer and/or higher stage (level of advancement) than Caucasian Americans [49].

However, African Americans men have around 1.8 times the risk of developing cancer as likely as the Caucasian Americans men, but African Americans are about 2.5 times more likely to die from cancer than Caucasian Americans [50]. Considering the disparities in the treatment of cancer and prostate health in the United States alone, where social, political, and economic processes are observed to be of measurable influence, perhaps it stimulates thought that, at least theoretically, the incidence and advancement of prostate cancer might be slowed, prevented from occurring or even be eliminated once these socioeconomic factors are identified and addressed [49,51,52].

2.3.2 Genetic factors

2.3.2.1 Family History

After age the biggest risk cause for prostate cancer is the family history, with an incidence of about 10% to 20% among men who develop prostate cancer because of a positive family history of this disease [53]. The most major clinical aspect of prostate cancer perhaps in men with a positive family history is its relatively early development, and these men characteristically get this disease being diagnosed 6 - 7 years before as compared to men with no such family history of this cancer [53 54].

The family history of prostate cancer put the men at high risk of developing it as well as to high mortality. Men, whose father had a diagnosed prostate cancer, have double the risk as compared to men with no family history of this disease. Moreover, if anyone has a brother with diagnosed prostate cancer, then the risk for him increases threefold. The presence of diagnosed family history in both the brother and the father boosts the risk approximately ninefold [55]. Even a positive family history in male population is linked

with fatal prostate cancer. Death danger from this tumor is about double more for men with a brother or a father who passed away of this cancer than men with this cancer devoid of any positive family history [56].

The evaluation of family history of prostate cancer reported by the affected individual has become slightly harder in the time of PSA. This is because of the associated exposure to this test within members of the same family, and the high utilization of PSA testing by men having a family history of prostate cancer [57]. Results from two studies put forward that the incidence of familial aggregation of prostate cancer in great part is attributable to genetic factors [58].

Certainly, prostate cancer is expected to have one of the extreme heritability, while above 40% of unpredictability of prostate cancer is due to genetic factors. Genome-wide association studies have tried to find a common single-nucleotide poly-morphisms (SNPs) related to the incidence of prostate cancer [59]. Up till now, scientists have discovered 41 risk loci and also verified them across numerous studies, this is a huge number of loci than any other cancer has [60]. It is anticipated that these forty one loci interpret almost one-quarter of the variability of the frequency of prostate cancer associated with genetic factors [60]. It is worth mentioning that the inheritance of prostate cancer emerges to be the consequences of small positive connections of genetic variants of low-penetrant type instead of big connections with high-penetrance alleles [60].

Most of the recognized germline dangerous loci seem to be more weakly linked with fatal or benign prostate cancer [61]. Nevertheless, family studies show a familial constituent of prostate cancer-associated survival [62]. So far, only one study based on genome-based

association of deadly prostate carcinoma has been published [63]. Though no SNPs attained genome-associated implication, three SNPs were related to this malignant cancer at 10^{-5}, and out of these three one was consequently confirmed in an autonomous cohort [60]. One of the problems in prostate cancer survival and studying germline variants is the recognition of cohorts with adequate numbers of cancer-specific incidents [60]. Consequently, more research is required to approximate the inheritance of prostate cancer survival, in addition to bigger genome-wide association studies of malignant disease.

A positive history of this cancer in the family of the male individual affected by prostate cancer basically shows a multifaceted blend of environmental and genetic factors. About 5% to 10% of cases of prostate cancer has been approximated that are the result of dominant inherited susceptibility to this cancer [64], and autosomal recessive and X-linked sorts of inheritance have in addition been implied [65]. The effect of genetics on the growth of prostate cancer was demonstrated in a study comprising of 44,788 duos of twins scheduled in registries such as Danish, Finnish, and Swedish identical registries [58]. The researchers observed the statistically significant outcomes of genetic factor responsible for causing cancer of the prostate gland, with innate genes causing forty two percent to the entire possibility of onset of prostate cancer and unknown environmental factors forming the residual 58% of the danger. Latterly, genome-associated research has been employed to recognize inherited threat for prostate cancer, with more than 24 prostate cancer risk is related to solitary-nucleotide polymorphisms (SNPs) that have been determined and further predicted being found out from studies conducting from time to time [66,67]. Albeit all of the SNPs found out to date are only somewhat linked with the risk of prostate cancer, collectively they show a stronger connection [67]. Besides, a SNP that is

considerably connected to the risk of development of most malignant nature of prostate cancer and not the benign prostate cancer has in recent times been reported [68].

The risk of this cancer for life has been expected to be about 12 percent for a male having a father with prostate cancer diagnosed at an age above 60 years while it is 35% to 45% for a male with 3 or more diagnosed male family members. The absolute risk of prostate cancer is only 8% in males with no such family history of the disease [69].

There is huge evidence present to demonstrate that men with a positive family history of prostate cancer are at high risk of developing it than those lacking a family history [70]. Furthermore, a man's risk can be influenced additionally in accordance with the level of association with the affected family member, his age at the moment of cancer detection and the entire number of family members who got affected with this cancer [70]. Provided that family history is a plain aspect to evaluate regular clinical practice via essential queries from the patient, it should be considered as an imperative factor to think about alongside PSA for assessment of the prostate cancer risk [70]. Additionally, family history can be evaluated from a comparatively less than or around 40 years of age (in case a father is diagnosed with this cancer), possibly providing the chance for medical involvement beforehand for male individuals at elevated danger of prostate cancer by better examination and, at some point, vigorous strategies for reduction of risk may become available [70].

Prostate cancer has three different epidemiological varieties, namely familial, hereditary and sporadic [71]. Different studies have established a fact that there are no such differences with respect to pathology among these epidemiological varieties of prostate cancer [72-74].

At present, no such clinical testing is present for genes implicated in hereditary prostate cancer. [71] conducted the sentinel study at Johns Hopkins University and gave the first definition for inherited prostate cancer, called as the Hopkins Criteria, and it is still the only way even after twenty years to imply hereditary prostate cancer.

In 1992, [71] described hereditary prostate cancer by using family history. A person was believed to have hereditary prostate cancer, if his family ancestry disclosed a history of prostate cancer in the following cases:

1. Family having 3 or more first-degree relatives i.e. comprising father, son, and brother.

2. Three consecutive generations of either the paternal or maternal ancestries; or

3. No less than two relatives diagnosed at or before the 55 years of age.

Males with familial prostate cancer also showed a positive family history, but it was not so encouraging but inadequate to come true to the hereditary criteria of prostate cancer. Those patients who didn't have any family history of prostate cancer were thought to suffer from sporadically induced prostate cancer [71]. In one of the study [75] made the different definitions of hereditary prostate cancer simply by saying that sporadic prostate cancer is a type of cancer developing haphazardly in the population while the familial type of prostate cancer is random grouping of prostate cancer in families and, hereditary prostate cancer is a powerful grouping with formation of prostate cancer prematurely in families.

[71] stated that about 43 percent of prostate cancer onset at an early age (i.e. an age less than 55 years) was the result of an autosomal dominant inheritance of an uncommon

allele. Above all, a number of researchers discovered that hereditary prostate cancer of an early age onset comprised of only a small amount (such as only 9% by the age of 85 years) of all prostate cancer incidences. This research results showed that only 2 percent of prostate cancer takes place in men of EuAm with an age below 55 years. These researchers concluded that the influence of hereditary prostate cancer in these people is greatest in the younger ages, i.e. below age 55 years [76].

Men belonging to the African lineage may possess a powerful genetic tendency to develop cancer of the prostate as compared to the men belonging from other lineages [77].

Though, positive screening rates have been seen in men of the different populations having a positive family history of prostate cancer, but studies conducted recently shows that not all men's groups show same screening rates. These studies show that men of the African-American ancestry have considerably lower screening rates even with a strong family history of prostate cancer than men of the Caucasian origin with a strong family history [78]. This study shows that only 45 percent of men of the African-American origin, having a positive family history of prostate cancer with more than four affected relatives, ever underwent a PSA testing while only 35 percent of men ever underwent a DRE [78].

In another study carried out in the United States consisting of 56 men having an affected FDR were not like to undergo this DRE or PSA testing as compared to 100 men without FDRs who were affected in the past by prostate cancer [79].

A number of studies have evaluated the psychological, sociodemographic, and medical features linked to screening of prostate cancer in males with a positive family history of prostate cancer. It has been noted that men having a family history of this cancer who are

elder, [79-82] married [83] and have much incomes [81-84] are more expected to receive a PSA screening. Also, having a high number of diagnosed relatives [82,84,85] and having talked about screening with a doctor [80] have been found as predictors of a high screening rate. In contrary, men appear least interested to undergo screening frequently even if they have an elevated cancer associated risk. A Swedish study particularly demonstrated that men with an established family history of the prostate cancer (with 3 or more relatives with this disease) were least interested of having a screening test in case they had an increased tendency to avoid cancer related risk [85].

Latest data propose not only a family history of prostate cancer does affect its risk, but possibly family history of breast cancer too. A number of studies put forward that a family history of breast cancer boosts the risk of prostate cancer by70 percent [86] and death by 16 percent [87]. On the other hand, others studies have found no such connection of risk between family histories of breast cancer and prostate cancer [88-90].

A family history of prostate cancer in either brother(s) or father raised cancer risk by 48 percent and 11 percent, respectively [91].Given that prostate cancer history in a sibling (brother) was more mightily related to high cancer risk compared to history of prostate cancer in a father only, we assumed that this might consequence from the likely early age at diagnosis of the sibling (average age of 64 years) against average father's age of 71 years [91].

About 46% men with positive family history of prostate cancer were found to have low grade disease while 51% of such men were found to have a high-grade disease, than men with no family history [91]. The existence of family history of both prostate and breast

cancer raised the risk of development of low- and high-grade cancer by 119 percent and 247 percent compared with men with no such family histories of either cancer [91]. When a comparison was done in men with a family history of prostate cancer only, it was found that those with a positive family history of breast and prostate cancer were 69% and 133% more expected to develop low- and high-grade cancer, respectively [91].

2.3.3 Lifestyle factors

2.3.3.1 Obesity

Stronger evidences in relation to the significant part played by lifestyle and surroundings in the causes of prostate cancer are there [92]. In the West, Obesity is very common and the prostate cancer is one of the most significant health issues.

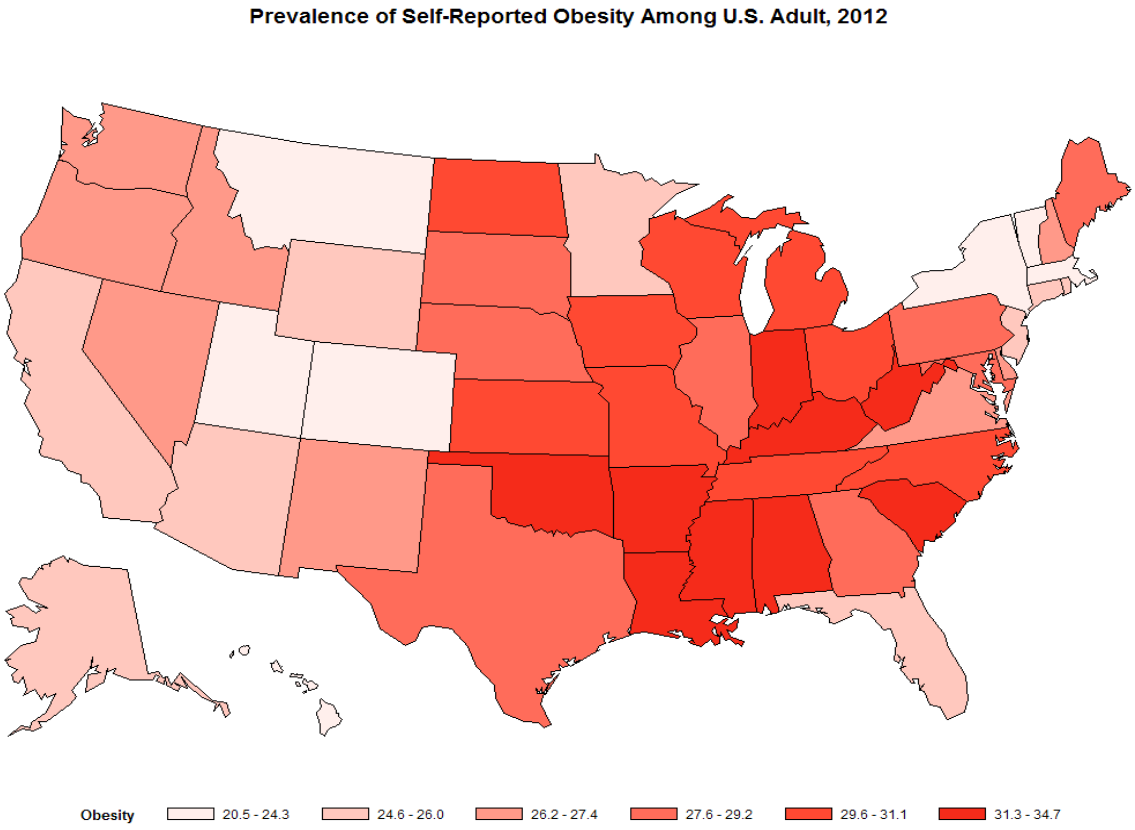

Figure 5: Self-Reported Obesity Among U.S. Adults in 2012 [93]

In Unites States, 14 to 20 percent of deaths associated with cancer are related with obesity [94]. Adding to this it has also been found that obesity also increases the risks associated with the several type of cancer and aggressive types of prostate cancer [23].

More than 66% of population of America is suffering from obesity [95].Small variations were seen in the percentage of obese adults aging between twenty to seventy four years from the year 1960 -1962 to 1976 - 1980. However much different to all this, the rates associated with Obesity witnessed a great increase in the years 1976 to 1980 &1999 to 2002. It was more than doubled from 15.1% to 31%. The increment was noticed across races, ethnic group and genders[23].

Differences based upon races and ethnicities are normally constant among states as shown in Figure 5. In the years 2006 to 2008, the rates of Obesity among states were in the range of 23-45 percent in the African American population. In Hispanic population, the rates are twenty one to thirty seven percent and nine to thirty percent in the white population [96]. The increase in rates of adults whom are categorized as extremely obese has prominently contributed to the raise of the rates of obesity in the last twenty five years. The rates associated with severe obesity in adults aging between 24-74 faced an increase from the percentage of 1.4 in between the years 1976 to 1980 to the percentage of 6.3 in between the year 2009 to 2010 [23].

The situation deserves attention because over the past few decades the raise of obesity rates has become a major public health concern [97]. Adults are facing the obesity threat, but at the same time children are also being affected by this disease as well. As result, this can further bring increase in the adult obesity rates in United States with the passage of time [98]. It is very important to develop a clear and deep understanding about the effects which are produced by obesity upon developing, preventions of diseases related with Obesity. This understanding can play an important part in results related with health concern.

Body Mass Index (BMI) is the popularly used ratio by many academies and health care practitioners for obesity measurement [99]. The Obesity of the body of individuals can be checked with the help of BMI in which they are classified as normal, overweight or obese. Body Mass Index (BMI) can be calculated by dividing the mass in kilograms on the height in meters squared [100]. According to the World Health Organization (WHO), Body Mass Index (BMI) result can be interpreted as follow [99]:

- Body mass index less than 18.5 kilograms per meter square is considered Underweight.
- Body mass index in the limit between 18.5 and 24.99 kilograms per meter square is considered normal.
- Body mass index in the limit between 25.0 and 29.99 kilograms per meter square is considered Overweight.
- Body mass index greater than 30 kilograms per meter square is considered Obese.

Researches have shown links between chronic diseases and Obesity like inflammations, dyslipidemia. Diabetes, blood pressure, cardiac anomalies, osteoarthritis, cancers and stroke [101]. On the other hand, losing weight toward normal weight reduce the risks and aggressive form of co-illnesses that are related to Obesity.

Weight loss can also act as a way for preventing cancer development [50,102,103]. The chances of developing breast cancer in women who have had their menopause are increased by obesity also risks associated with the colon cancer and esophagus are increased [104-106]. Moreover, it needs to be mentioned that Obesity is associated with aggressive form of prostate and breast cancer [107].

In prostate cancer cases, the gaining of function of the receptor of androgen is the main step in the development of cancer which is dependent upon hormone to the cancer which is not dependent upon hormone like the prostate cancer [108]. Males who are obese are at lower risks of developing the prostate cancer in comparison to those obese individuals who are higher risk of dying because of this pathology in comparison to those who are lean [109]. Therefore, obese men have higher risk of developing prostate cancer.

The majority of male population can survive the cancer of prostate, but those men who are associated with hormone-refractory metastatic prostate cancer have a chance less than six percent to survive another five years. Never the less, Obese men are at higher risk of facing death than those men who have normal body mass index [109-111].

Obesity might be link to prostate cancer mortality. Obesity may cause false-negative PSA test result which lead to treatment delay [23]. Men who are obese more often face failures related with biochemistry; they develop highly metastatic tumors and lead to prostate cancer which is not dependent upon androgen [111]. As mention before, Obesity links with advanced stage of prostate cancer. Also, surgery can be very complicated and difficult radiation treatment for prostate cancer patients who are obese [112]. It has been found that obese prostate cancer patients sometime have a true PSA test result, but they still develop aggressive form of prostate cancer which indicates there are biological mechanisms played part in the development of prostate cancer[112].

Obesity is related with changed levels of different hormones that include testosterone, insulin, IGF-1, estrogen, leptin etc [105]. They all have been associated with prostate cancer. Because Obesity is a result of dietary style, those who are obese possess positive balance

of energy and show consumption of great quantity of fats from diet and all of them are related with cancer [113]. Also obesity is related with mediator of inflammation and it can be a developing factor of prostate cancer [113].

Researches which examine the relation between adult's body mass index and development risks of prostate cancer have provided mixture of outcomes. Many larger cohort studies showed increase body mass index related with a great risk of the prostate cancer [114-117], but some of these studies described these relations as not sufficiently strong. On the other hand, some different researches showed no relationship between body mass index and risks of prostate cancer [118,119]. A recent group study from the United States showed inverse relation among obesity and the diagnosis of prostate cancer, but males who were less than sixty years or those who have a positive family history of prostate cancer [120].

In regarding to race, great numbers of African-American males were obese in present regiment. Obesity can be regarded as a risk that can be modified, which can be related with more violet tumors [46]. Obese African-American males have higher risks pathology related aspects than obese non-African-American males [46]. False PSA test result is more often occur with African-American men; and obesity contributes to increases this false result of PSA test [46]. Obesity may be responsible for the racial discrimination that is present in prostate cancer.

Epidemiological relation between Obesity and violent prostate cancer is specifically related because of universal nature of pathology and the numbers of males that are victims [121]. Apart from race and age there are some developed risks for the prostate

cancer. The recognition of Obesity as another risk for violent prostate cancer is of great importance for the health of public because of its nature, which can be modified. The fat tissues are different in their capacities to undergo expansion and reduction during the life of a person and reduction can be brought by bringing changes in lifestyle [121]. Epidemiological evidences which link obesity with violent prostate cancer shows the significance of considering size of body into consideration while carrying out the screening, monitoring and treatment of patients of prostate cancer also guiding patients about healthy choices [121].

2.3.3.2 Physical Activity

Physical activity plays a role in various ways to decrease the risk of numerous varieties of cancer, along with cancers of the endometrium, breast, prostate and colon [122]. Additionally, physical activity if performed regularly assists to keep a healthy body weight by harmonizing caloric intake with energy utilize. The physiological benefits of a physically active way of life outdo decreasing the risk of cancer, and incorporate decreased risk of death [23].

Though, the best possible intensity, frequency, and duration of physical activity required to lower cancer risk are unidentified, researchers recommend that increased levels of physical activity (such as performing an activity of or more than 300 minutes weekly or of activity of high energy comprising 150 minutes weekly) may offer even better declining in cancer risk [94]. A number of other research have revealed that being energetic at these high degrees of physical activity aids in preventing weight gain and Obesity [122,123]. By serving as a source of maintaining weight, this level of physical activity can have an influence on decreasing the risk of onset of obesity-associated cancers [94].

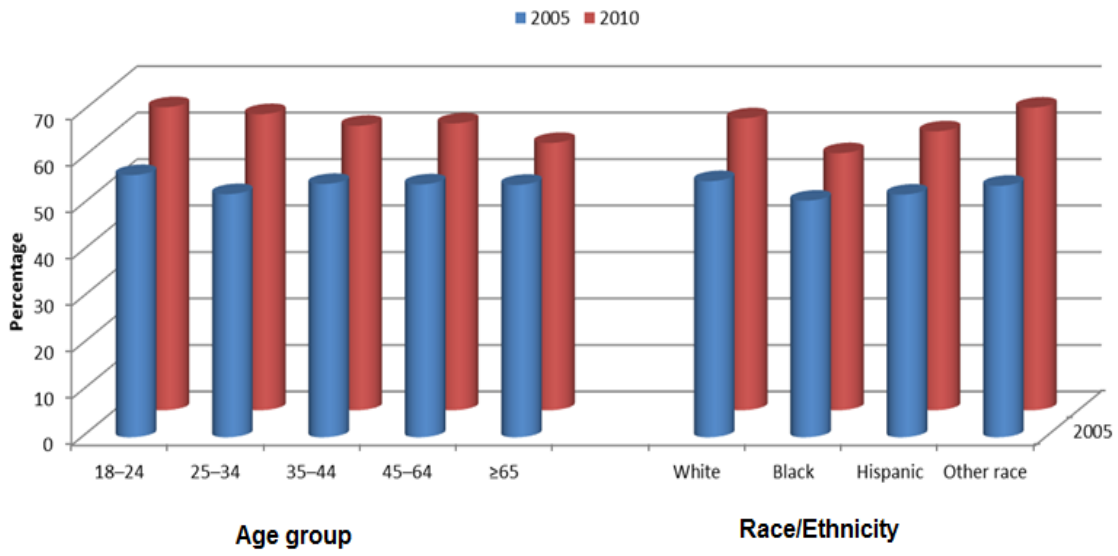

Figure 6: Percentage of adults who reported recent walking, by age group and race, 2005 and 2010 [124]

In the year of 2011, approximately 25.4% of adults informed that they have no free time for physical activity. The fraction of adults notifying no free time for physical activity varied from 16.5% and 36% in Colorado and Mississippi, respectively [23]. In the same year, about 20.6% of adults informed that they spare extra time to engage themselves in no less than 150 minutes of moderate activity or 75 minutes of strong activity every week [23].

Physical activity is proven to be helpful at any age and it helps claim a positive health condition. So as to observe the consequence of physical activity, the difference is made between levels of physical activity, like light, light-moderate, moderate, strong and intense exercises. A study demonstrates that moderate- strong physical activity, for example sports, is helpful in reducing the risk for the development of all types of malignant cancers [125]. A study discovered a strong opposite connection between prostate cancer and physical activity [32]. Moreover, [32] stated that the results of physical activity

were conflicting. [125] found that the regular performing of strong physical activity reduces the risk for prostate cancer among males of the middle-age group.

Physical activity has been considered a main tool by researchers for making the overall life quality better. Among older people (above age 60 years old) who were the prostate cancer survivors for a long time, [126] discovers that steady moderate-to-strong exercise performed weekly makes the physical function, quality of life and health perception better. One study illustrates that energetic physical activity boosts the levels of androstenedione hormone in the body. It is, therefore, very important that middle-aged as well as older adults keep themselves busy in adapting physical activity routines. One study discovered associations between education level and physical activity. With increasing age and low education levels, cancer patients don't bother about the importance of physical activity usefulness for them whereas a people of higher education level were found to be engaged in a regular weekly exercise [126]. In general, physical activity has been believed to be defensive against hormone-related tumor development because the physical activity decreases the power of endogenous hormones travelling all through the body [127]. Though, the results concerning physical activity and the danger for prostate cancer are indecisive.

However, the accurate biologic mechanisms that connect physical activity and risk of prostate cancer remain uncertain, it is conceivable that physical activity may perhaps decrease the risk of this cancer via this pathway, i.e. It may impact the levels of some endogenous hormones associated with risk of prostate cancer, for instance androgen levels [128], levels of insulin [129], as well as insulin-like growth factors [130],while daily activity may decrease testosterone levels [131], levels of serum insulin [132], and of IGF-I [133].

On the whole, there was 10% i.e. a statistically significant decreased risk of prostate cancer when evaluated against the maximum versus the minimum degree of activity [134]. Physical activity if done in the age group of 20 to 45 years and the 45 to 65 years considerably decreased the risk of prostate cancer, while physical activity earlier than 20 years of age and after 65 years of age did not decrease the risk of prostate cancer [134].

The [134] results of physical activity on the risk of prostate cancer varied among dissimilar populations. For Europeans, Occupational physical activity (OPA) and Total physical activity (TPA) markedly diminished the risk of this cancer, while the Recreational physical activity (RPA) was not associated with the reduction in risk. A feeble but noteworthy outcome of the RPA was discovered when limited to cohort studies. The collective RRs from the USA populations were also considerable for occupational physical activity and total physical activity but not for recreational physical activity. On the contrary, there was no clear link between physical activity and possibility of prostate cancer in Canadian people found. On top, we discovered a little proof for any effect of physical activity on the risk of this malignancy among populations of the Asia-Pacific [134].

[134] noticed a 19% reduction in risk of prostate cancer for occupational physical activity and only a five percent slightly statistically significant reduction of risk for RPA. When many researches were limited to those of superior quality, the RPA as well as OPA consequences were satisfied with a drop of risk by 14% and 3%, respectively [134].

In regard to the race, [135] found that there was a major association noticed between more than 9 metabolic equivalent (MET) hours weekly and a reduced risk of this cancer on biopsy in white men than men who were inactive (less than 3 MET hours weekly). White

men who completed 3 to 8.9 MET hours every week did not receive the same advantage [135].in addition, [135] observed that high levels of exercise have an opposite connection with the danger of development of prostate carcinoma particularly in men of white origin but not in men of black origin. It is noteworthy to consider that there was no significant variation found pertaining to the level of exercise between the two groups. During the last decades exercise has been found to be discreetly related to a high risk of prostate cancer in black men [135]. There was no association found between any degree of exercise and the danger of positive biopsy results in black men [135].

2.3.3.3 Alcohol Abuse

Alcohol has been in use and abuse for long in a great diversity of cultures all over the world, and its consumption has been rising fast in several countries [1]. According to The National Institute on Alcohol Abuse and Alcoholism (NIAAA), in 2012, the percentage of adults who aged 18 or older drank alcohol at some point in their lifetime is 87.6%. Adults who drank in the past year are 71%; and 56.3% for they who drank in the past month [136]. Substance use is everywhere in the general public, with 60- 80% of people consuming alcohol, and about 10% of people use some type of substance to the extent of addiction or abuse [137]. Given that the link between alcohol and a variety of types of cancer is comparatively modest at reduced levels of consumption, there have been various observational studies, showing most of the participants using low to moderate quantity of alcohol, and thus, an important positive link, a nonexistence of a considerable association, or a noteworthy negative link between alcohol consumption and the danger of death and disease were found from certain type of cancers [1].

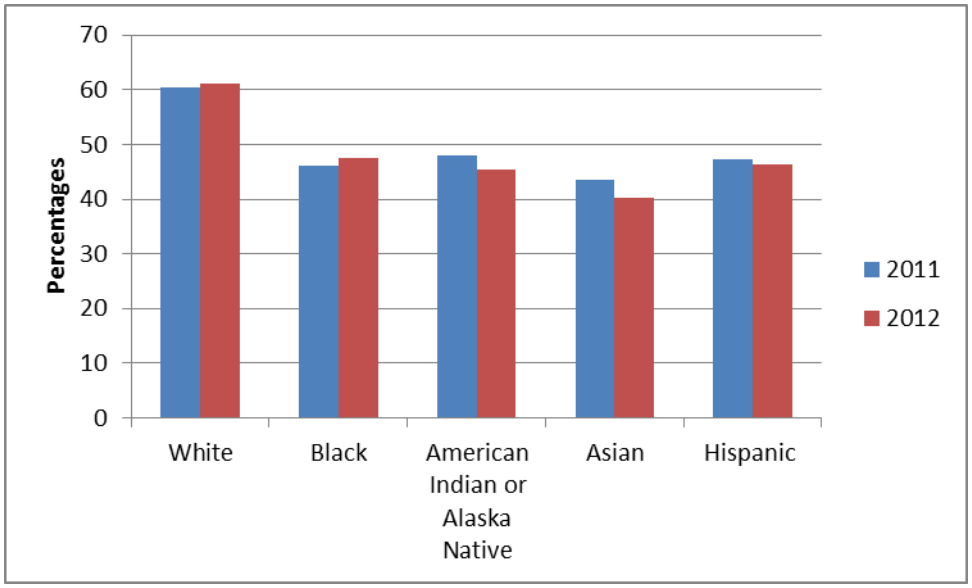

Figure 7: Alcohol Use, Persons Aged 18 or Older, by Demographic Characteristics: Percentages, 2011 and 2012

The precise nature of the connection between alcohol drinking and prostate cancer risk has been vague regardless of the big number of research studies undertaken as reviewed in [138-143]. Despite the fact that there seems to be small evidence for an association as a whole, a little risk increase with high levels of alcohol drinking has been found in [144] but not always has been found in [138]. In view of the fact that these early reviews and meta-analyses [140-143], 6 extra cohort [145-150], and 6 case–control [151-156] research studies have been issued on the connection between risk of prostate cancer and alcohol drinking of which 1 case–control [156] and 2 cohorts [146,148] studies revealed a significant larger risk associated with heavy alcohol drinking. [157] found a link between lifetime alcohol drinking and high risk of prostate cancer.

A latest meta-analysis that tried to separate methodological concerns with a scrupulous attention on amount of alcohol use concluded that alcohol drinking of two or more drinks daily was linked to a high risk of prostate cancer derived from confirmation from case–

control population- based studies [143]. One concern that has not been sufficiently identified is how this association may differ by the aggressiveness of the tumor. Until now, 9 studies [145,149,153,154,158-162] have regarded at least several features of the severity of the cancer and the consequences have been uncertain.

[157] studied the association of different type of alcohol with prostate cancer. They found that there was a noteworthy higher possibility of occurrence of prostate cancer when measured using statistical data with regard to excessive drinking of beer, during evaluating the dose of alcohol drinking for whole life. 1.65 as odds ratio obtained for those who had drunk 7 beers or more weekly in comparison to nondrinkers for benign cases versus controls, with (p= 0.003). For the aggressive cases versus controls, odds ratio of 1.99 was obtained, with (p=0.002) for males who used to drink 7 alcoholic drinks or more weekly than who does not drinker. The insignificant associations were found for liquor or wine lifetime drinking.

Those males who drank about twenty one alcohol drinks weekly or more had notably elevated risk of developing prostate cancer, irrespective of any type of alcohol[157]. For increasing intake of alcohol, there was a statistically significant high risk of prostate cancer with rising quartiles of alcohol drinking [157].

Heavy consumption of alcohol (i.e. greater than 50 g of alcohol every day) and customary heavy drinking (i.e. greater than 4 drinks on a daily basis or more than 5 days weekly) were related to an increased risk of prostate cancer of high-grade nature (RR, 2.01 [95% CI, 1.33 to 3.05] and 2.17 [95% CI, 1.42 to 3.30], respectively); a lesser amount of drinking was not linked with cancer risk [159].

(Middleton Fillmore et al., 2009) meta- analysis consisting of 16 cohort studies and 31 case–controls reported a significant pooled RR for any alcohol consumption drinking against no alcohol consumption of 1.08 (95% CI, 1.01 to 1.16), and proposed that prostate cancer occurrence is certainly linearly linked with alcohol drinking, and with an RR of 1.16 (95% CI, 1.06 to 1.26) for an increase in everyday alcoholic drink unit. However, the authors declared that this association needs more exploration, particularly with regard to heavy drinking.

With reference to the link between prostate cancer and alcohol by stage, alcohol drinking raise the risk of terminal or advanced prostate cancer, [159] though a current potential U.S. cohort study stated a less important connection between alcohol and malignant cancers and an augmented risk of benign cancer [162]. Other research that demonstrated no linkage between prostate cancer and alcohol may have been distressed by the discovery of an error in screening, which would disguise this link if people who drink heavily were less interested to receive PSA screening [163].

[164] confirmed a positive connection between alcohol consumption and high-grade prostate cancer. On the other hand, [138] comprehensive meta-analysis offered no confirmation about a material linkage between prostate cancer and alcohol drinking, even at very high drinking doses.

In China,[142] meta-analysis found that no major association found between drinking alcohol and certain cancers such as that of colon and rectum, lung, ampulla of Vater, prostate, extra hepatic bile duct or pancreas. The consequence of alcohol consumption on

the risk of several cancers may differ with traditions, the form of alcohol used, its dose, or the standard of living of the participants.

2.3.3.4 Smoking

The consumption of tobacco is regarded as the most significant risk of cancer, which can be avoided. Around hundred million individuals died in 20th century all around the globe because of the diseases, which are related with tobacco, such as chronic lung disease, cancer, stroke and cardiovascular disease. A quarter of those who smoke will meet death in a premature manner during the age of 35 to 69[23]. The prevalence of smoking in the grownups of the United States faced a reduction between 2005 and 2011 it has a percentage of 20.9% to 19% and prominent reductions were there for both males and females (23.9% to 21.6%) and (18.1% to 16.5%) respectively [23].

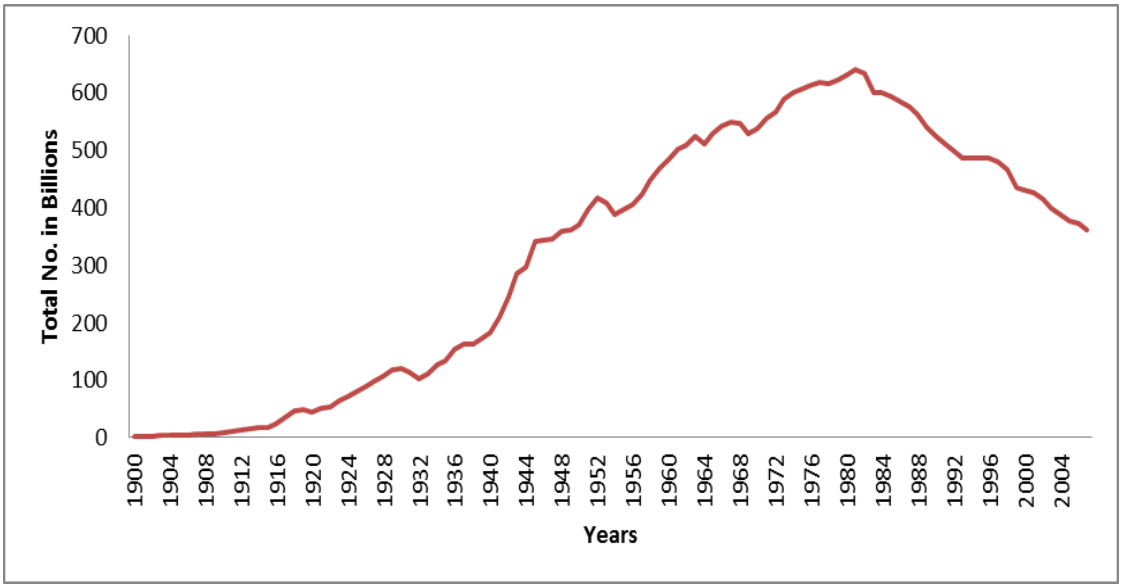

Figure 8: Cigarette Consumption, United States, 1900-2007 [165]

A great part of smokers develop tobacco addiction prior to the legal age for purchasing cigarettes. The younger generation shows more sensitivity towards nicotine [166]. The use

of tobacco improves risks of malignancy of mouth, lungs, larynx, esophagus, pharynx, liver, bladder, ovary, pancreas and cervix [167,168] . Smoking of cigarette is the supreme cause of death and illness that can be prevented in the United States [169].

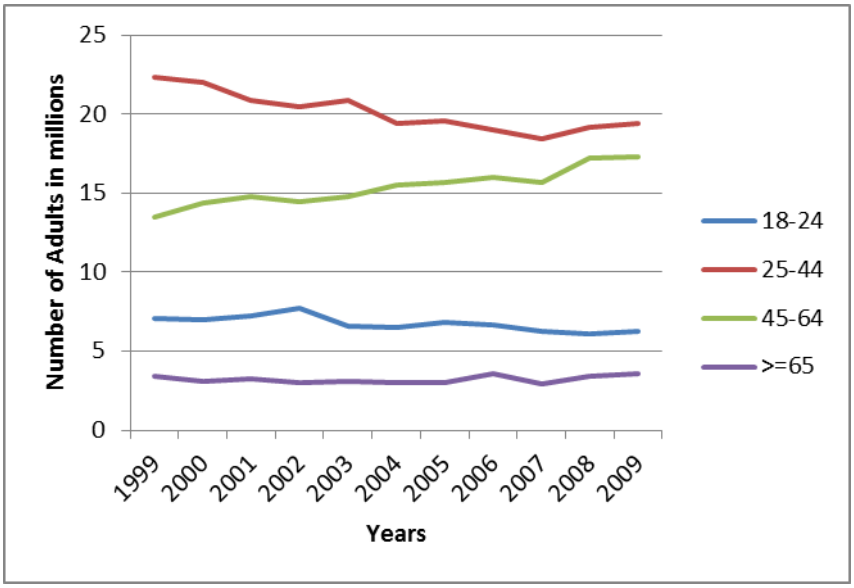

Figure 9: Number of Adults in millions, Who Were Current Smokers by Age, from 1999 to 2009 [170]

No doubt smoking is related with improved incidences of many cancers which include lung, bladder, and kidney cancers; and its relation with the prostate cancer stays not clear. In addition to this the influence of smoking on the treatment of prostate cancer is a matter under consideration. A new study showed that smoking of cigarette is related with poorer diagnosis and high prostate cancer related deaths irrespective of the approach of treatment [171]. In patients who undergo radical prostatectomy, smoking is related with improved risks for metastasis. Smoking is related with a high risk of CRPC, BCR and death rate [172].

Heavy smokers possess an improved risk of 24% to 30 % death from prostate cancer in comparison to those who don't smoke [171]. Moreover, previous smokers possess small

raise in the risks of prostate cancer; for them the risks of tumors were nine percent greater than those who don't smoke [171]. A great part of the potential cohort studies indicated that existing smoking is related with medium level increase of thirty percent in lethal prostate cancer risks in comparison to those who don't smoke [173]. Smoking is an important risk factor as far as the development of the prostate cancer is concerned and must be taken as related exposures in the research of prostate cancer and preventing death from this cancer [173].

However, in [174] recent study conclude some opposite finding. In their study they found that current smokers have a decreased risk of prostate cancer in comparison with those who don't smoke, with statistically significant for localised and low-grade disease, but this is not for higher grade disease.[174] found that males who used to smoke at enrollment possessed ten percent low risks for the prostate cancer in comparison to those who don't smoke. On the other hand, those who smoke more than 25 cigarettes per day and males who had smoked for a longer time period more than 40 years have a greater risk of prostate cancer deaths [174]. It appears that smoking is related with decreased risks of less violent prostate cancer. On the other hand, heavy smokers are related with high risks of the death due to prostate cancer [174].

In another recent study from Japan, the results of [164] depict smoking brought a reduction in the risks of cancers in subjects, but this negative relation may have impaired in subjects in which cancer has been detected. The outcome gets support from a recent analysis of 24 groups which depicted improved risks of prostate cancer [171].

Smoking is not a very significant risk as far as the males who are smoking in decade prior to the diagnosing show a poor prognosis and great death rate from this pathology [173]. However the cause for this relation is not apparent the evidences show that smoking in a direct manner contributes towards a violent prostate cancer phenotype instead of improving risks in an indirect manner via a delay in diagnosis. Therefore, smoking must be taken as a related exposure in the prostate cancer research and death prevention from this pathology [173].

2.3.3.5 Dietary Fat

Aspects related with diet are considered as very important environmental factors which influence the development and progress of the prostate cancer. In accordance with the opinion provided by experts 30 % to 35 % of all the malignancies are related with aspects of diet [175]. Higher intake of calories, lower fibers and higher intake of saturated fats are taken as risks for the prostate cancer. On the other hand, vegetables, vitamin D, vitamin E, tomatoes, selenium can act as protective agents [176,177].

The link between fats from diet and prostate cancer cases has been researched in detail after early ecological evidences which link per capita consumption of fat to the prostate cancer death rate [178]. The intake of fatty acids has also passed through investigations in relation to the prostate cancer occurrence and report suggests that a lower risk of the development of prostate cancer was reported with higher diet and blood concentrations of the marine omega-3, polyunsaturated fatty acids and positive links among alpha-linolenic acid and advanced [179] fatal [180,181] cancer of prostate, but the results are not converging every time [182,183].

The consumption of fat is related with prostate cancer [184,185], however, the part played by fats of diet and other specified types is not clear [184,186]. International ecological research showed that the total intake of animal fat relates with the prostate cancer death rate [187]. Researchers have indicated that saturated fats [188,189] monounsaturated fats [190,191], alpha-linolenic fatty acids [192,193] related with the prostate cancer, but no prominent relation has been figured out by others [194,195]. Researchers have shown that intake of the specified saturated fatty acid [196] was strongly associated with prostate cancer risks.

A large nationwide, case-control study, [197]found that pointed parts of diet show association with prostate cancer risks and diet lower in trans-fats can bring reduction in risks associated with prostate cancer.

The migration researches showed that prostate cancer incidence is high in immigrants who come from Japan to the United States than in Japan and this indicates towards the point that lifestyles or deity factors can play part in the development of prostate cancer [198,199]. Western diets can be described as higher fat intake, diet fat and many fatty acids. Fatty acids like n-6 polyunsaturated fatty acids (PUFA) are highlighted for playing a part in the prostate cancer cause and development [200]. Fatty acids and their related metabolite show involvement in many paths with significant effects on the development of prostate cancer [200].

Consumption of fat remained the main consideration of diet related studies and risk of the prostate cancer. Fat is that part of diet which is most energy crowded. It produces 9 calories per gram upon full oxidation. Looking at its chemical makeup it consists of a backbone of glycerine with which attachment of fatty acids are placed. Fatty acids are

categorized in three ways, saturated fatty acid (SFA), monounsaturated fatty acid (MUFA) and polyunsaturated fatty acid (PUFA). There is difference present between their biological qualities because of saturation degrees and the number of carbons in a chain [186]. It has been indicated that forty percent of entire intake of energy in a conventional diet of West comes from fat [201]. Association between diet related intake of fats and the prostate cancer has been indicated via a relationship of per capita consumption of fat and the larger international level variation in the rates of prostate cancer [202,203].

[204] found that there is a relation between early onset of prostate cancer and higher consumption of fats. A significant risk was linked to higher consumption of all fatty acid, MUFA, PUFA and SFA. A prominent dose-respond association was figured out for entire fat and its types (p= 0.001), indicating that there is incremental in the risks associated with the prostate cancer with high consumption of fat [204]. Causative fractions indicate that if these relationships were informal, then almost forty percent of the early onset cases of prostate cancer can have aetiological existence associated with the consumption of fat [204].

[200] cohort study found that the association of fat and fatty acids showed difference by the severity of the prostate cancer. Consumption of fat and fatty acid was not associated with risks of prostate cancer, which has not reached advanced levels [200]. On the other hand, consumption of saturated fat (EPA and ALA) is associated with high risk of progressed prostate cancer [200].

It was found by [205]that higher total dietary fat and fatty acid with short chain consumption were not positively associated with survival after the confirmation of prostate cancer which is localized in group study of Swedish males.

For the cholesterol association with prostate cancer, [206] shows that the present of cholesterol metabolism is essential in prostate cancer development. In addition, [207] in their study suggest that hypercholesterolemia has an essential role in prostate cancer development. Cholesterol works as a mediator of cell growth, inflammation, steroidogenesis, and membrane dynamics; therefore, there is a link between cholesterol and prostate cancer development. [208] study and [209] population-based cohort study both provide further evidence that cholesterol levels in blood are associated with increased risk of prostate cancer. Moreover, they suggest that parts of aggressive and non-aggressive prostate cancer are related to cholesterol levels. On the other hand, [210] find that a low cholesterol levels might be an effective strategy to prevent and delay prostate cancer development.

2.3.3.6 Vitamin D Deficiency

The importance of Vitamin D as one of the secosteroid hormones exhibiting pleiotropic actions can never be ignored. It is already very famous as far as the calcium regulatory action is concerned, these days it is getting great attention because of the fact that Vitamin D possesses the potential of regulating immunological changes along with this it can also perform against inflammations and is equipped with properties which enable it to cause regression of fibrosis. In addition it has its part in cell division as well.

Prior to interacting with the receptor an activation procedure consisting of three steps is required in the case of Vitamin D. A major amount of this vitamin is manufactured in skin because of sunlight. Initially UVB converts 7-dehydrocholesterol to the pre-vitamin D3 that is then changed into Vitamin D3 via a procedure that is dependent upon heat. Here the important point to highlight is that disproportionate exposure to sun is not responsible for causing the intemperance of Vitamin D as additional Vitamin D3 is demolished by sun. From diet a smaller percentage of this vitamin is obtained and it can be obtained from fishes, egg, ultraviolet irradiated mushroom, supplement etc. Vitamin D2 and D3 that are derived from diet their absorption takes place from procedures that are related with bile acids. On the other hand, Vitamin D is merged in the lumen of intestines and from there enterocytes absorb it turning them into chylomicrons which enter the circulatory system through lymphatic system. The synthesized vitamins from both the above mentioned sources can be placed in fat tissues or they can witness 25-hydroxylation in the liver.

Ecological researches conducted have shown that mortality due to prostate cancer is inversely proportional to sun levels [211,212]. It can also depict that Vitamin D is inversely proportional to the prostate cancer [213-215] as the standing of this vitamin is related with the amount of sunlight which skin is exposed to and also the potential of skin related with its manufacturing as a result it carries out the regulation of the growth of cells and their division [213,216]. The studies, which have been conducted provide evidences that risks associated with prostate cancer are reduced in those males who are more exposed to the light of sun [215,217,218] also the pigmentary properties which can stop the synthesis of Vitamin D, like darker skin[35,215] or easy tanning [219] related with prostate cancer.

Ecological indications containing contraindications that higher level of ultra-violet exposure are related with increase risks associated with the death rate related with prostate cancer [220].

There are evidences present which indicated that Vitamin D indeed plays a part in the development of the prostate cancer. Genetic variations in the receptor are related with the Gleason score 139 also the variation causing agents in the pathway of Vitamin D are related with risks leading to reappearance of the death rate associated with prostate cancer [221]. Higher expressions of the receptor proteins in the prostate cancer are related with low risks of deadly cancers in males this was exhibited in Health Professionals Follow-up study and Physicians. Males in the highest versus low quarter of the expressions of Vitamin D receptors possessed risks of 0.37 with the PSA adjustments at diagnosing [222]. Also a research related with prostate cancer death rate carried out at HPFS and the PHS figured that the prostate cancer sufferers that have low concentrations of the prediagnostic 25 D possessed a prominent threat related with the prostate cancer deaths and 1.59 RR for high versus low quarter [223]. The levels of vitamin D that are pre-diagnostic are prominently related with stages as well as grades of the research. So it is found that the exposure to the discussed vitamin apparently is not related with low rates of the incident of the cancer of prostate, but different evidences show that its pathway can play its part in the development of the cancer of prostate. It was found in a research that in migrants from Asia increase risks of the incidents of prostate cancer is present because of the adaptation of western ways of living a diet that is not good in containing larger quantity of Vitamin D enriched fish oil [224].

[225] study provides weak supporting to the concept that good amount of skin exposed to sun can provide protection against prostate cancer development. However the [226] study showed that 25(OH) D amounts are related with more violent cancers; and there are no evidences of relation between the risks of cancer of prostate. The little exposure to sunlight for obtaining vitamin D can bring reduction in violent prostate cancer. Epidemiology based relations in prostate cancer and Vitamin D can play an important part in growth and division of prostate cancer [224]. But the researches which link serum levels of vitamin D and the prostate cancer death rate have no strong correlation [227].

[228] found that the deficiency of Vitamin D is related with risks of prostate cancer as basically as exposure to ultraviolet B radiations or in the form of diet related factors or in the form of endogenous units.

2.3.4 Chronic Disease and Surgical Procedure

2.3.4.1 Inflammation of prostate

Chronic inflammation diseases have been linked to cancer development such as bladder, liver, large-intestine cancers, and stomach [229,230]. Inflammation plays an important role in angiogenesis, cancer cell invasion, cancer initiation and growth, and metastatic dissemination [231]. Therefore, inflammation has been considered as a risk factor of prostate cancer [231]. In fact, inflammatory infiltrates are commonly found in prostate biopsies [232]. [233] find that prostate inflammation, mostly chronic type, is associated with an increased the likelihoods of prostate cancer. Moreover, prostate inflammation increases high-grade prostate cancer which leads to develop high aggressive cancer. On the other hand, [234] found that chronic inflammation is associated with a lower risk of prostate

cancer. Also, inflammation that found in prostate biopsies might lower the risk of subsequent prostate cancer detection.

2.3.4.2 Vasectomy

According to [235] there are studies did not find a clear association between vasectomy and prostate cancer development. However, there are several studies indicate that there are a positive association between vasectomy and increased risk of prostate cancer. [236] found that there is a small association between vasectomy and the risk of prostate cancer. They describe the biological mechanism that supports the relationship between vasectomy and prostate cancer as unlikely. Also, they sagest that this association must be tested again if there are more prostate cancer factors found in order to clarified the association with vasectomy. [237] did not find a clear association between vasectomy and prostate cancer risk.

2.3.4.3 Hypertension

Early studies do not have a sufficient amount of data to confirm the association between high blood pressure and prostate cancer. Some of these studies found that there are no a significant association between prostate cancer and high blood pressure [238,239]. However, [240] a case control study in African Americans community found that the prostate cancer risk was increased by 2.4 fold among patients with hypertension. In 2011, an animal experiment done by [241] on rat laboratory found some evidence that link hypertension with increase the risk of prostate cancer. [242] recent study on a large sample of patients showed that there is an association between hypertension and obesity and prostate cancer risk. Moreover, hypertension was associated with poor diagnosis of prostate cancer patients;

and also, patient who is overweight and has hypertension had a statically significant shorter survival time ($p = 0.037$) [243].

2.4 Artificial Neural Network

2.4.1 Artificial Neural Networks Overview

Artificial Neural Networks (ANN) that work in a parallel as well-structured information processing units, act just like a human brain and nervous system by imitating their great computational ability. In comparison to serial processors, the fundamental computational elements of a human brain are quite slow. However, they're still able to perform some tasks that a traditional computer may consume an enormous amount of time to do. And in most cases, computers may not even perform at all. These neural networks learn from experiences, generalize from past examples, outline important aspects from the input of unnecessary data, and deal with ambiguous situations, all by simulating the brain. ANN includes many neurons and synaptic strengths called weights. They emulate the nervous system in a way the weight signals move through the network.

Lately the research community has started to advocate the use of ANNs as a means of solving biomedical and healthcare problems [244]. Artificial Neural Networks are similar to biological neurons due to the functional similarity. However, they are much simpler and therefore any similarity between the two is still at the fundamental level [244].

Apart from just emulating the human intelligence, neural networks have great ability in learning the relationship between the input and output mapping of a dataset without having any earlier knowledge or assumptions of the statistical distribution of the data.

This remarkable capability makes the neural network the right tool for the task of classification and also the regression in practical circumstances. These two tasks are pretty essential and are marked as a fundamental part of biomedical applications. Also, in contrast to many conventional methods based on linear techniques, neural networks are quite applicable for correct modeling of intricate data patterns mainly because they are non-linear [244].

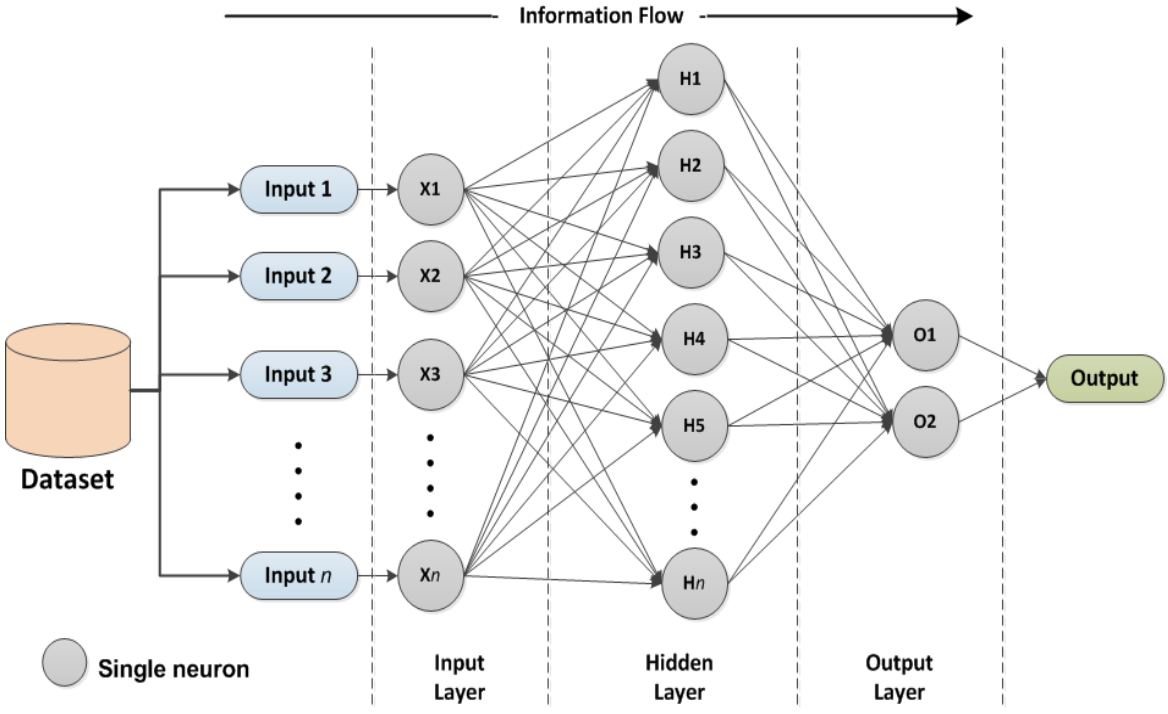

Figure 10: Artificial Neural Network Structure

These computer systems known as ANN have been constructed to automatically carry out tasks such as deducing new information by learning and formulating new information and discovering. All these abilities are traits of a human brain. As shown in Figure 10, ANN is made up of several layers; input layer, one or more hidden layers, and an output layer. On each layer, there are particular components connected to one another. These are called

nodes or neurons. Each of the neuron's connected to weights and the following communication network. In this network, signals travel via neurons over weight. Collective signals are received by each neuron depending on their weights and bring about an output signal that may also be generated by other neurons. Effectively, ANN has been used to predict the future occurrence of diseases such as reoccurrence of cancer, cardiology ailment and to help physicians with prognostic and decision support [245].

Neural networks can emulate feed forward systems and can also administer a large quantity of input data. They can, in fact, function as or take the place of the missing data given their parallel design. They have the ability to pick on the built-in rules of a given system, look after long term memory, and recognize patterns in both changing and noisy environments.

Classification is one of the many uses of an ANN. For this task, each pattern of input is forced systematically to output the pattern indicators that are part of the training data; this training set contains the input covariate x with the corresponding class labels. Also called multilayer perceptrons (MLP), feed forward networks are made to mold a set of input signals (X) into a set of output signals (G). The feedback networks commence with the beginning activity state of a feedback system. Once the state transition has been done, the asymptotic final state is established as a result of the computation. Associative memories are one of the many uses of the feedback network. For example, when presented with a pattern near the prototype X, the output should produce pattern X. One other use is that of autoassociative memory or contents addressable memory with the help of which the requested output is completed to become an X. In every circumstance, the network picks on and is trained by the repetitive pattern with known outputs, also called pattern

indicators. Supervised neural networks find an association f: X →G for pairs of giving sets of inputs and outputs [246].

2.4.2 Artificial Neural Networks for Prediction of Prostate Cancer

Among different types of cancers which occur in men prostate cancer is the most commonly occurring one. Both prostate specific antigen (PSA) and digital rectal examination (DRE) are the most common tools for early detection and screening for prostate cancer. No doubt that PSA has the reputation of being an important marker for the prostate cancer [247]. However, PSA has important limitation related with its use in the diagnosis of cancer, which is the prominent overlapping of patients with prostate cancer and those who suffer from other prostate pathologies. When PSA test indicate that the serum level in blood is higher than normal, the patient has prostate cancer. However, the serum PSA level can be affected by benign prostate hyperplasia (BPH), prostatitis or prostate manipulations, especially when the range of serum PSA is between 4.0 and 10.0 nanogram per millimeter[248].

The utilization of prostate specific antigen (PSA) has resulted in unnecessary diagnosis and treatment of the prostate cancer because of its accuracy limitation and poor specificity for early detection [249]. A great part of the problem is related with the limitations of the prostate specific antigen test. Different types of efforts have been put in for improving the situation, but the results have not been very convincing.

During the past ten year artificial intelligence (AI) has gained decent recognition in different medical practices. Artificial neural network (ANN) is one of the important technic used in artificial intelligence. Neural network emerges as a famous help for

assisting when it is about making prognosis related choices for patients. These types of models can give deep understanding into choices based on decisions and aids these choices by the development of prognostics which allow getting information from previous clinical information [245].

The data which has been published states that the models of ANN have appeared as important tools when it's about decreasing the work pressure on the clinicians by carrying out the detection and giving choice related support.

In comparison to other concepts related with machine learning, many appreciating points are present in neural networks, which deserve the attention of potential users. Great architecture combined with other tools makes ANNs a great tool for processing. It can be used for any quantity of output as well as input and they have a good support in various languages of programming. Via alteration of weight before training, imposition of restrictions related with custom changes the present knowledge can included in the design and construction. In addition to this neural network is normally not very expensive for utilization after training and this makes it decent for real application. Latest outcomes show that these carry out the generation of results that can be compared to any state-of-art classifiers [250,251].

Utilization of ANNs in the prostate cancer is great option as:

1. More than one factor to make prediction which leave impact on the results.
2. The demands of offering individualized consulting that utilize several different tests result together.

3. Free limitation in compere with logistic regression analysis which has serious limitations.
4. The demand for an up-to-date tool which can be applied without difficulty to all who need it.

In regard to prostate cancer, ANNs improves the precision associated with prediction of the primary biopsy of prostate in comparison to parameters related with PSA for patients that are examined for early detection or screening of prostate cancer [252].

In [253] study, they use neural network model to provide a predictive outcome for prostate cancer patients whether they have cancer or not by using free prostate-specific antigen, total prostate-specific antigen and age information. Their study outcomes of ANN, rate of success 94.11 percent and 94.44 percent were accomplished for the diagnosis of cancer as well as validity correspondingly. This setting doesn't carries out the diagnosis of malignancy in a conclusive manner it aids clinicians in making the decision that if biopsy is required by giving data related with patient having prostate cancer or not.

In 2012, the [254] independent cohort and [255] show that the ANN is a good aid in regular activities for improving the detection rate of prostate cancer and brings reduction in not required biopsies. Moreover, [256]study shows that the models of ANN are aids in assessing the risk of cancer of prostate and for making the decision that if biopsy is required or not. [257] utilizes ANN and generally used analysis techniques logistic regression for developing predictor models for the cancer of prostate survival rate. Their outcomes show that the ANNs, is more precise than the logistic regressions with a precision of 91.07 percent and 89.61 percent correspondingly. Similar study by [249] carried out the comparison of neural

network to the logistic regressions. It was shown that ANN is more precise. [258] used the neural network study of the history of employment as one of the risk factors for prostate cancer.

In [259] study that based on a large population aims to distinguish between patients with prostate cancer and those who do not show evidence of cancer by using benign prostate-specific antigen (bPSA) and benign prostatic hyperplasia-associated (BPHA). They exam the result by using a percent free PSA (%fPSA)-based artificial neural network (ANN) model. Their result shows that BPHA/tPSA-based artificial neural network was most accurate compering to all other parameters.

On the other hand, [260] study shows that PSA velocity (PSAV) has restricted value for detecting prostate cancer with seventy one percent improving PSA value; and twenty nine percent of all prostate cancers don't have usual PSAV. Although they indicate that artificial neural network velocity which based on the combination of PSAV and a %free PSA-based ANN cannot improve the rate of prostate cancer detection, it might reduce unnecessary prostate biopsy by 11.00% to 17.00% .

A review by [261] of some PubMed publications that utilized artificial neural network in clinical trials to draw a future trends in some clinical areas such as diagnosis, prognosis and recovery guidance for cancer patients, in addition to the need for detailed application of hard methods. It has suggestions for design of study for addressing some other empirical model for medical detection based on generic non-linear function approximations, which includes artificial neural networks.

CHAPTER III

III. METHODS

3.1 Overview

This study utilizes series of data processing, statistical procedures, and data mining techniques to achieve its goals. Data modeling, extraction, cleaning, and recoding were used to prepare the row data for the study. Logistic regression, analysis of variance (ANOVA), and descriptive analysis were used in this study as statistical procedures. Artificial neural network is the main technique used in this study. All these methods were used to process and analyze a large data set driven from multiple years.

3.2 Data Sources

3.2.1 The Nationwide Inpatient Sample (NIS)

The Nationwide Inpatient Sample (NIS) is one of the main sources of data for this study. The NIS is considered to be the largest database in the United States for all-payer inpatient health care. This study will utilize NIS data for the following years 2007, 2008, 2009, 2010, and 2011, with a total of 12,004,120 records as shown in Table 2. These records were selected for each man with age of 35 years and older. Approximately 5.35% of the data population has prostate cancer.

Table 2: The distribution of NIS data by years

Year	Prostate Cancer		Total	
2007	115,628	4.96%	2,333,312	19.44%
2008	125,571	5.13%	2,446,883	20.38%
2009	127,685	5.41%	2,358,344	19.65%

2010	127,656	5.37%	2,378,964	19.82%
2011	146,113	5.88%	2,486,617	20.71%
Total	**642,653**	**5.35%**	**12,004,120**	**100.00%**

3.2.2 The National Health Interview Survey (NHIS)

The National Health Interview Survey (NHIS) is part of the major data collection programs of the National Center for Health Statistics (NCHS). NHIS data is targeting the health information of the United States civilian who is 18 years and old. The major goal of NHIS is to monitor the health of the United States population. This study utilizes NHIS data for the years of 2009 to 2012 to exam the association of physical activity with prostate cancer. Men with age of 35 years and older were selected to be analyzed for this study. In NHIS data, there are three main questions about the type of physical activity (vigorous, light or moderate, and strengthening). Also, three to one sub questions for each one of the main questions which asked about the length of doing physical activity and time period for length.

3.2.3 Surveillance, Epidemiology, and End Results (SEER) Program

In the United States, the National Cancer Institute (NCI) fund and run the Surveillance, Epidemiology, and End Results (SEER) Program which consider a very important source of cancer patients' data and information. SEER program cover around 28 percent of the United States population. The SEER program collects data from several different geographic areas of the United States, which include Alaska Native Tumor Registry, Arizona Indians, Cherokee Nation, Connecticut, Detroit, Atlanta, Greater Georgia, Rural Georgia, San Francisco-Oakland, San Jose-Monterey, Greater California, Hawaii, Iowa,

Kentucky, Los Angeles, Louisiana, New Jersey, New Mexico, Seattle-Puget Sound, and Utah.

In addition, it is the only source of cancer patients' data collects stage of cancer at the time of diagnosis and patient survival data.

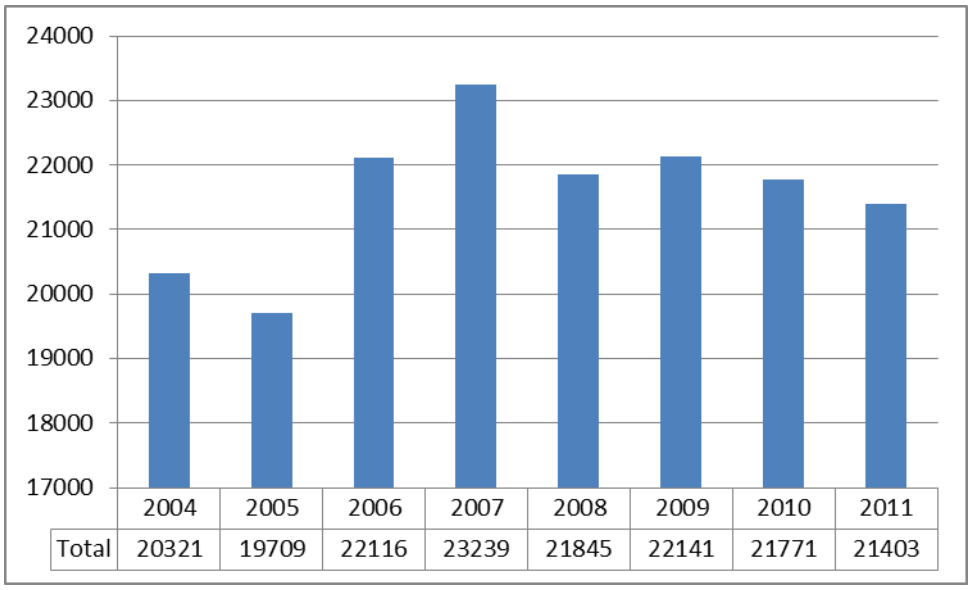

Figure 11 : The distribution of SEER data by years

This study utilizes SEER data for the years of 2004 to 2011 as shown in Figure 11, with a total of 172,545 records. All patients whose age between 35 years and 100 years were selected to analyze their data which include age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data.

3.3 Statistical Analysis and Tools

In this study, all preliminary analyses were performed using SAS version 9.3(SAS Institute, Cary, NC, USA). In addition, MATLAB version 8.3(The MathWorks, Inc., Natick, MA USA) with Artificial Neural Network toolbox was used to simulate the classification models and exam their result. Microsoft Excel 2010 and Microsoft Vision

2010(Microsoft Corp, Redmond, WA USA) were also used to create charts, diagrams, and figures. All calculated *p* values were two-sided and *p* value less than 0.05 were considered statistically significant.

3.3.1 Analysis of Variance (ANOVA)

Analysis of variance was conducted to determine which group is significantly different than other groups. Age and race variables were examined by analysis of variance.

3.3.2 k-Fold Cross-Validation

The k-Fold Cross-Validation technique was used in this study to ensure the selection of the classifier models with the highest accuracy and the lowest prediction errors.

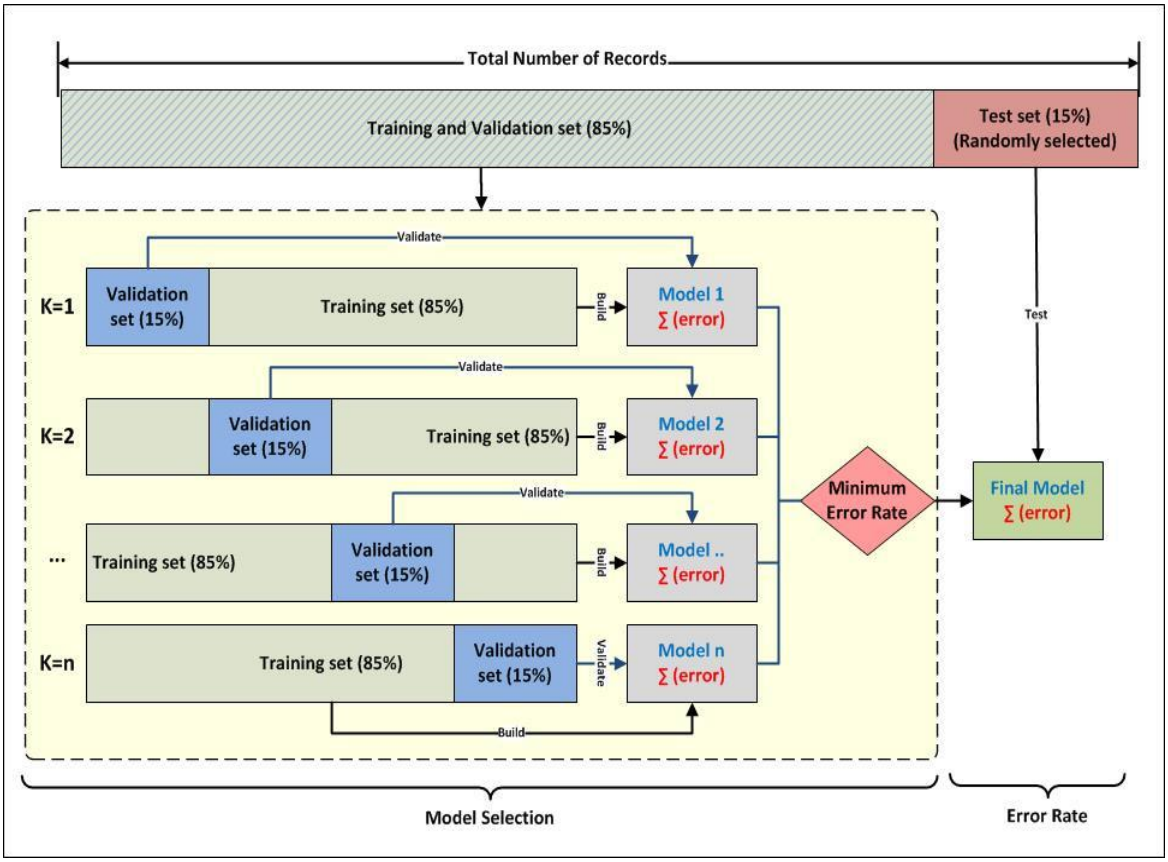

Figure 12 : k-fold Cross-Validation technique

As shown in Figure 12, the main dataset was divided into a training and validation dataset (85%) and testing dataset (15%). The testing dataset was created randomly. The training and validation dataset was partitioned into k (a positive integer) equal size sub-dataset. Subsequently k iterations of training and validation were performed. In each iteration cycle, a different sub-dataset was held-out for validation while the remaining sub-datasets (k -1) were used for training. At the end of each training cycle, the classifier models were run against the validation dataset to make predictions and then calculate the errors. Each model was stored in models array with its own error rate. Upon completion, all observations were used for both training and validation; and the model with lower error rate was selected as the final model.

Algorithm 1 k-Fold Cross-validation.

```
Require: T, a training data set
Require: k, the number of folds
Require: a, a learning algorithm
F = Split Into Folds(T, k)
modelsArray = [a, error rate]
for i = 1 to k do
        for j = 1 to k do
                if j ≠ i then
                        Train(a, F[j])
                end if
        end for
        r ← calculate errors by Validate(a, F[i])
        modelsArray[i] ← [a, r]
end for
c ← Selected Classifier(modelsArray [a, Minimum (r)])
```

3.3.3 Logistic Regression

This study used logistic regression analysis to exam the association probability of each risk factor and the incidence of prostate cancer. In the logistic regression model, prostate

cancer is the dependent variable. The prostate cancer risk factors which include age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension are the independent variables. The Logistic Regression model helped to identify the critical variables that can be used to predict prostate cancer incidence. Moreover, odds ratios were used for each independent variable to measure the expected number of times prostate cancer risk factor will occur relative to the number of times it will not occur.

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k}}, \text{where } X_1, \ldots, X_k \text{ are the } k \text{ independent variables} \quad (2)$$

In addition, logistic regression classification models were created for both NIS dataset and SEER dataset by using MATLAB 8.3. For NIS dataset, prostate cancer is the dependent variable in the logistic regression model. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension. The logistic regression model for SEER dataset use prostate cancer as the dependent variable. On the other hand, age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data are independent variables. In both models, the distribution is set to binomial. The link function is set to “logit”.

$$X\beta = \log\left(\frac{\mu}{1-\mu}\right) \tag{3}$$

$$where\ \ \mu = \frac{1}{1+\exp(-X\beta)} \tag{4}$$

3.3.4 k-Nearest Neighbors

k-Nearest Neighbors (KNN) algorithm is a classification method that classifies unlabeled data to the nearest most similar labeled data. In other words, a set of observation data (X) is classified based on its label (Y). A new data that is unlabeled will be assigned a label similar to the nearest most similar labeled data. For distance measures, there are several distance functions; and all of them are used with continuous variables only.

$$Pr(Y = j\ |X = x_0) = \frac{1}{K}\sum_{i\epsilon N_0} I\ (y_i = j) \tag{5}$$

$$Euclidean\ distance\ function = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{6}$$

KNN classifier is simple but very effective with high accuracy. Also, KNN is insensitive to outliers; and does not make assumptions about the underlying data distribution. On the other hand, KNN does not generate a model and requires a large amount of memory.

In this study, KNN classifier models were created for both NIS dataset and SEER dataset by using MATLAB 8.3. For NIS dataset, prostate cancer is the dependent variable in the K-Nearest neighbors' model. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol

abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension. The K-Nearest neighbors' classifier model for SEER dataset uses prostate cancer as the dependent variable. On the other hand, age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data are independent variables. In both models, the distance measure function is Euclidean.

3.3.5 Naive Bayes Classifier

Naïve Bayes classifier is utilizing Bayesian methods to classify data. During the training phase, Naïve Bayes classifier calculates the probability of every class based on independent variables. After that, these probabilities will be used to classify unlabeled data to the most likely class.

$$p(y, x_1, \dots, x_k) = q(y) \prod_{i=1}^{k} q_i(x_i|y) \qquad (7)$$

Naïve Bayes classifier is simple, it is fast and very effective, and works well with small datasets as well as large datasets. However, Naïve Bayes classifier has the assumption that each variable is important and independent, and it is not always true.

In this study, Naïve Bayes classifier models were created for both NIS dataset and SEER dataset by using MATLAB 8.3. For NIS dataset, prostate cancer is the dependent variable in the Naïve Bayes classifier model. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension. The Naïve Bayes classifier model for SEER dataset use

prostate cancer as the dependent variable. On the other hand, age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data are independent variables.

3.3.6 Decision Tree

Decision Tree is a hierarchical data structure which is started with one root and ends up with leaves by applying the divide-and-conquer technique. As a result, a decision tree is built by decision nodes and terminal leaves. Each decision node has two branches as an output path. Inside each decision node, a test function that determines which branch path should be taken by the input variable to reach a leaf node as the final destination.

$$I[Y;X] = \sum_{i=1} \Pr(X = x_i)\, I[Y;X = x_i] \qquad (8)$$

Decision Tree is simple and more understandable when is converted to a set of IF-THEN statements. However, Decision Tree model sometimes can be over-fit or under-fit.

In this study, Decision Tree classifier models were created for both NIS dataset and SEER dataset by using MATLAB 8.3. For NIS dataset, prostate cancer is the dependent variable in the Decision Tree classifier model. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension. The Decision Tree classifier model for SEER dataset use prostate cancer as the dependent variable. On the other hand, age, race,

Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data are independent variables.

3.3.7 Support Vector Machines

Support Vector Machine (SVM) classifies data by creating a flat boundary between data. This boundary called a hyper-plane, which separate data into two fairly homogeneous partitions on both sides. The best hyper-plane is one that gives the largest minimum distance to data, which is called a margin. All data points that are located on the margin boundary and are the nearest data points to the hyper-plane are called the support vectors.

The Support Vector Machine (SVM) is a state-of-the-art classification method that is referred to as black box processes. Although, SVM can be slow during the training phase, it has high accuracy and superiority over other classification methods.

In this study, Support Vector Machine classifier models were created for both NIS dataset and SEER dataset by using MATLAB 8.3. For NIS dataset, prostate cancer is the dependent variable in the Support Vector Machine classifier model. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension. The Support Vector Machine classifier model for SEER dataset use prostate cancer as the dependent variable. On the other hand, age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data are independent variables.

3.3.8 Artificial Neural Network

Artificial Neural Network (ANN) is a very sophisticated prediction method which can deal with non-linear functions effectively. This study uses multilayer perceptron (MLP), which is the most used artificial neural network architecture. Multilayer perceptron has a strong ability to classify and predict complex problems.

The Multilayer perceptron structure is designed to group the neurons into layers as shown in Figure 10. The first layer called input layer; and it represents the input variables which can be *n* number of neurons. The last layer is called the output layer and it contains the result variables which can be also *n* number of neurons. The middle layer is called the hidden layer and it can be more than one layer. A typical artificial neural network is represented mathematically by the following equation.

$$y(k) = F\left(\sum_{i=1}^{n} w_i(k) \cdot x_i(k) + b\right) \tag{9}$$

Where:

- *y(k) is output value in discrete time k*
- *F is a transfer function*
- *n is total of input variables*
- *w_i is weight value in discrete time k where i = from0 to n*
- *x_i is input value in discrete time k where i= from 0 to n*
- *b is bias*

Activation functions:

Equation 1: Sigmoid

$$f(x) = \frac{1}{1 + e^{-\beta x}} \tag{10}$$

Equation 2: Gaussian

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{11}$$

Equation 3: Threshold

$$f(x) = \begin{cases} 1 \; if \; x_i \geq threshold \\ 0 \; if \; x_i < threshold \end{cases} \tag{12}$$

In this study, Artificial Neural Network classifier models were created for both the NIS dataset and SEER dataset by using MATLAB 8.3. Both models are feed-forward networks which data moves only forward from the input neurons to the output neurons through the hidden layers.

For the NIS dataset, prostate cancer is the dependent variable in the Artificial Neural Network classifier model. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension. The Artificial Neural Network classifier model for the SEER dataset uses prostate cancer as the dependent variable. On the other hand, age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size data are independent variables.

Both data was divided randomly into three sub-datasets: Training Data (70%), Validating Data (15%), and Testing Data (15%). The hidden layer size was determined by

Minimum-Hidden-Layer-Size algorithm. The Neural Network model for NIS dataset had best performance with hidden layer size of 44 neurons. And for the Neural Network model for SEER dataset, the best performance was with hidden layer size of 52 neurons.

Algorithm 2 Minimum-Hidden-Layer-Size.

```
Require: D, a data set
Require: m, maximum number of hidden layer neurons, where m ≥ s
mimHiddenLayerSize ← 0
s ← total number of D variables
v ← 0
for i = s to m do
        net ← create Neural Network with hidden layer size (i)
        Train (net, D)
        Validate(net, D)
        p ← calculate performance (Test(net, D))
        if p > v then
                mimHiddenLayerSize  = i
                v = p
        end if
end for
return mimHiddenLayerSize
```

The final Artificial Neural Network classifier model combines the Neural Network model for NIS dataset and the Neural Network model for SEER dataset with their weights and biases values.

$$y(k) = F\left(\left(\sum_{i=1}^{n} w_i(k) \cdot x_i(k) + b\right).\left(\sum_{j=1}^{m} w_j(k) \cdot x_j(k) + b\right) + \bar{b}\right) \qquad (13)$$

3.4 Research Design

In this study, datasets were obtained from the Nationwide Inpatient Sample (NIS), the National Health Interview Survey (NHIS), and the Surveillance, Epidemiology, and End Results (SEER) Program. The raw datasets were extracted and stored in a relational database. This study used the open source MySQL database to store the data. ICD 9 codes for every related factor were extracted and stored as shown in Table 3.

Table 3 : ICD-9-CM Diagnosis Codes

	Disease / Procedure	Clinical Classifications Software	ICD-9-CM diagnosis
1	Cancer of Prostate	29	185, 2334, V1046
2	Family history of malignant neoplasm of prostate	-	V1642
3	Family history of malignant neoplasm	-	V160, V161, V162, V163, V164, V1640, V1641, V1643, V1649, V165, V1651, V1652, V1659, V166, V167, V168, V169
4	Obesity	-	2780, 27800, 27801, 27802, V8521, V8522, V8523, V8524, V8525, V8530, V8531, V8532, V8533, V8534, V8535, V8536, V8537, V8538, V8539, V854
5	Alcohol-related disorders	660	2910, 2911, 2912, 2913, 2914, 2915, 2918, 29181, 29182, 29189, 2919, 30300, 30301, 30302, 30303, 30390, 30391, 30392, 30393, 30500, 30501, 30502, 30503, 76071, 9800, 3575, 4255, 53530, 53531, 5710, 5711, 5712, 5713
6	Tobacco use disorder	-	3051

7	Disorders of fatty acid oxidation	-	27785
8	Pure hypercholesterolemia	-	2720
9	Vitamin D Deficiency	-	2680, 2681, 2682, 2689
10	Inflammatory diseases of prostate	-	6010, 6011, 6012, 6013, 6014, 6018, 6019
11	Vasectomy	-	V2652
12	Essential hypertension	98	4011, 4019

The data was cleaned and recoded as binary data (0 and 1). The prostate cancer column was created; and assigned "1" value for each patient who has been diagnosed with prostate cancer; and "0" value for those patients who do not have prostate cancer. The risk factors variables were also treated the same as prostate cancer variable. Age variable was tested as the quantitative variable in all datasets. However, in order to exam the relationship between age groups and Prostate Cancer in NIS dataset, age was divided into subgroups and recoded as a binary data as shown in Figure 14. These age subgroups were 35-44 group, 45-54 group, 55-64 group, 65-74 group, 75-84 group, and 85 and older group. Also, categorical variables that have more than two option values such as race were converted to dummy indicator variables. Each indicator variable has the values 0 and 1. Any categorical variable with *n* categories has been split into *n* indicator variables. For example, race variable was divided into 7 variables: White, Black, Hispanic, Asian or Pacific Islander, Native American, Other Race, and Missing.

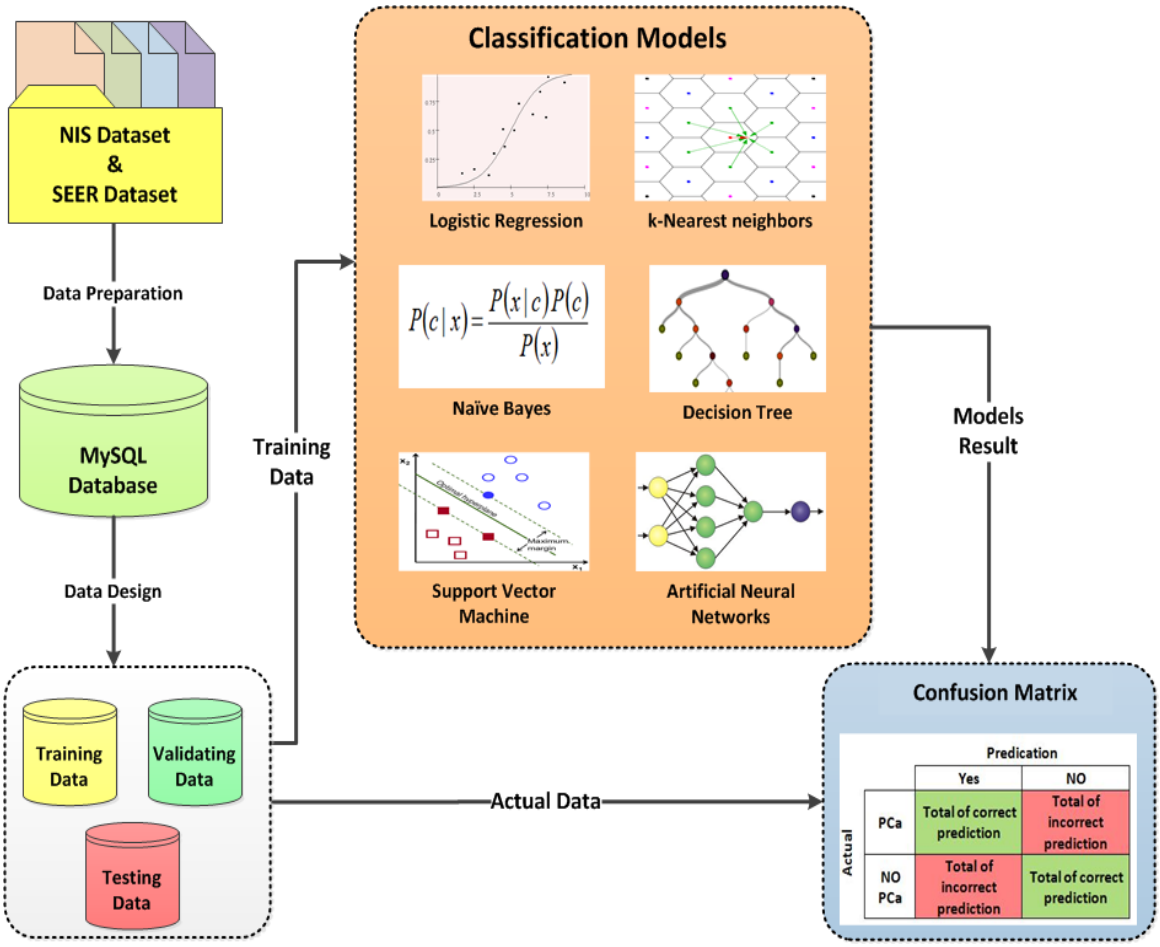

Figure 13: A graphical representation of the study process.

Artificial Neural Network (ANN) classification method was used to predict the occurrence of prostate cancer. ANN classifier models were built for NIS data and for SEER data, as well. In addition, five classification methods were built for each data to evaluate our ANN models. These classification methods were Logistic Regression, k-Nearest Neighbors, Naïve Bayes classifier, Decision Tree classifier, and Support Vector Machine. All these classifiers were simulated by MATLAB version 8.3.

Data was divided randomly into three sub-datasets: Training Data (70%), Validating Data (15%), and Testing Data (15%) as shown in Figure 13. The outcome of each model was stored and evaluated against other classification models.

VIEWTABLE: C.Data

	pca	PCa Family History	Cancer Family History	Obese	Alcohol	smoking	Fatty Acid	Vitamin D	Inflammation	Vasectomy	Age35_44	Age45_54	Age55_64
1	0	0	0	0	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	1	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	1	0	0	0	0	1	0	0
6	0	0	0	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	0	0	1	0	0
10	0	0	0	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	1	0	0	0	0	1	0	0
12	0	0	0	0	0	1	0	0	0	0	1	0	0
13	0	0	0	0	0	1	0	0	0	0	1	0	0
14	0	0	0	0	0	0	0	0	0	0	1	0	0
15	0	0	0	1	0	0	0	0	0	0	1	0	0
16	0	0	0	0	0	1	0	0	0	0	1	0	0

Figure 14: Binary Data

The effectiveness of each model was tested using two methods: Receiver Operating Characteristic (ROC) Analysis and Confusion Matrix. ROC curve was the main evaluation method used. The area under the curve (AUC) indicating the accuracy of the model as shown in Figure 15.

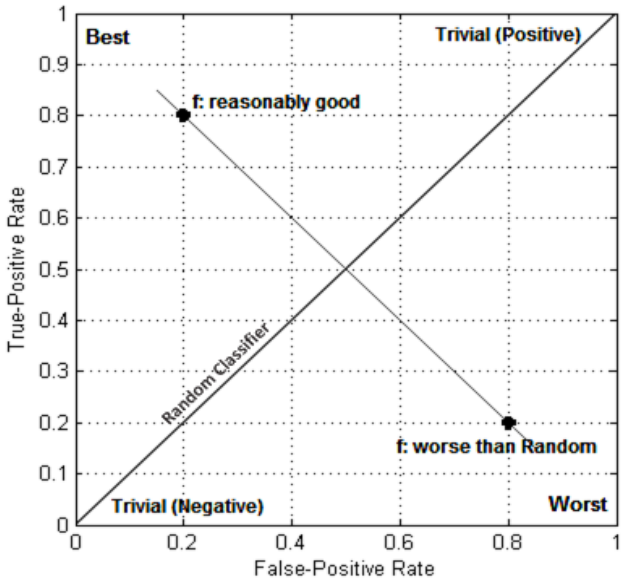

Figure 15 : The ROC space.

The AUC can be calculated as a normalized Wilcoxon-Mann-Whitney statistic[262] as follow.

$$AUC = \frac{1}{P.N}\sum_{i=1}^{P}\sum_{j=1}^{N} C(x_i y_j) \tag{14}$$

$$where\ C(x_i y_j) = \begin{cases} 1\ if\ x_i > y_j\ , \\ 0\ otherwise, \end{cases}$$

$$P\ is\ the\ total\ number\ of\ positive\ label\ '1'$$

$$N\ is\ the\ total\ number\ of\ negative\ label\ '0'$$

Confusion Matrix was the second evaluation method. It evaluated the models performed by calculating false positives classification and false negatives classification; and then, calculated the models' accuracy and the miss-classification rate.

Table 4 : Confusion Matrix

	Predicted Class (0)	Predicted Class (1)
Actual Class (0)	True Negative (TN)	False Positives (FP)
Actual Class (1)	False Negatives (FN)	True Positive (TP)

$$Model\ Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{15}$$

$$Misclassification\ Rate = \frac{FP + FN}{TP + FP + TN + FN} \tag{16}$$

$$Sensitivity\ (True\ Positive\ Rate) = \frac{TP}{TP + FN} \tag{17}$$

$$Specificity\ (True\ Negative\ Rate) = \frac{TN}{FP + TN} \tag{18}$$

The main goal was to determine which model has the highest performances and accurate classification for Prostate Cancer diagnose.

Both Neural Network models' outputs with their weights and biases values were combined to calculate the final result.

For the ethical consideration, all patients' identifications were removed from the data. Also, all participants in this study were anonymous. The HCUP Data Use Agreement (DUA) along with training course on data security were completed and signed before having access to the data. In addition, the Surveillance, Epidemiology, and End Results (SEER) Use Agreement was signed and submitted before having access to SEER data.

CHAPTER IV

IV. RESULTS

4.1 Introduction

A promising combination of some techniques of state of the art classifiers and sample data of 12,004,120 patients' record from the Nationwide Inpatient Sample (NIS) and 172,545 patients' records from the Surveillance, Epidemiology, and End Results (SEER) Program. This study helps in identifying some of the risk factors of prostate cancer and to use them as possible prediction factors for early detection of prostate cancer. Age, race, family history of prostate cancer, family history of any other cancer, obesity, physical activity, alcohol abuse, smoking, fatty acid deficiency, hypercholesterolemia, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension factors are expected to have a vary association with prostate cancer as well as early signs of detection prostate cancer. This study expects to find age, race, and positive family history of cancer are statistically significant risk factors and strong early warning of prostate cancer. Although, this study expects to find different association between prostate cancer and the other factors which include obesity, physical activity, alcohol abuse, smoking, fatty acid deficiency, hypercholesterolemia, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension from strong to weak association, they will be very helpful in early detection of prostate cancer.

4.2 NIS Data Analysis

4.2.1 Descriptive Statistics

The main data are derived from a larger dataset of The Nationwide Inpatient Sample (NIS) through the years 2007, 2008, 2009, 2010 and 2011. The sample (12,004,120) is of men age 35 and older, and the average age of the patients is 64 years old with a standard deviation of 14.7. The demographic characteristics of the data showed in Table 2.

Table 5: The demographic characteristics of the NIS data

	Prostate Cancer		*No Prostate Cancer*		*Total*	
Number of patient	642,653	5.35%	11,361,467	94.65%	12,004,120	100.00%
Mean [SD] age	74.59 [10.64]		63.58 [14.66]		64.17 [14.69]	
	n	*%*	*n*	*%*	*n*	*%*
Age group (years)						
35-44	1,503	0.23%	1,256,845	11.06%	1,258,348	10.48%
45-54	24,178	3.76%	2,229,455	19.62%	2,253,633	18.77%
55-64	93,504	14.55%	2,468,735	21.73%	2,562,239	21.34%
65-74	176,363	27.44%	2,375,514	20.91%	2,551,877	21.26%
75-84	225,644	35.11%	2,072,487	18.24%	2,298,131	19.14%
85 and older	121,461	18.90%	958,431	8.44%	1,079,892	9.00%
Race/Ethnicity						
White	425,067	66.14%	6,899,518	60.73%	7,324,585	61.02%
Black	73,731	11.47%	1,262,011	11.11%	1,335,742	11.13%
Hispanic	29,487	4.59%	797,556	7.02%	827,043	6.89%

Asian or Pacific Islander	7,644	1.19%	193,006	1.70%	200,650	1.67%
Native American	2,126	0.33%	62,830	0.55%	64,956	0.54%
Other	12,203	1.90%	275,990	2.43%	288,193	2.40%
Missing	92,395	14.38%	1,870,556	16.46%	1,962,951	16.35%
Risk Factors						
PCa Family History	5,504	0.86%	10,598	0.09%	16,102	0.13%
Other Cancer Family History	7,104	1.11%	93,025	0.82%	100,129	0.83%
Obesity	32,907	5.12%	962,369	8.47%	995,276	8.29%
Alcohol	7,649	1.19%	400,058	3.52%	407,707	3.40%
Smoking	46,716	7.27%	1,864,834	16.41%	1,911,550	15.92%
Fatty Acid deficiency	0	0.00%	29	0.00%	29	0.00%
Hypercholesterolemia	62,119	9.67%	831,685	7.32%	893,804	7.45%
Vitamin D deficiency	2,702	0.42%	37,323	0.33%	40,025	0.33%
Inflammation of prostate	3,054	0.48%	34,591	0.30%	37,645	0.31%
Vasectomy	519	0.08%	2,880	0.03%	3,399	0.03%
Hypertension	417,999	65.04%	6,669,577	58.70%	7,087,576	59.04%

In the sample data, 642,653 patients were diagnosed with prostate cancer and represent approximately 5.35% of the entire data. The age distribution of prostate cancer patients is shown in Figure 16. Prostate cancer patients whose age 65 years and older are the majority with 81.45% of all prostate cancer data.

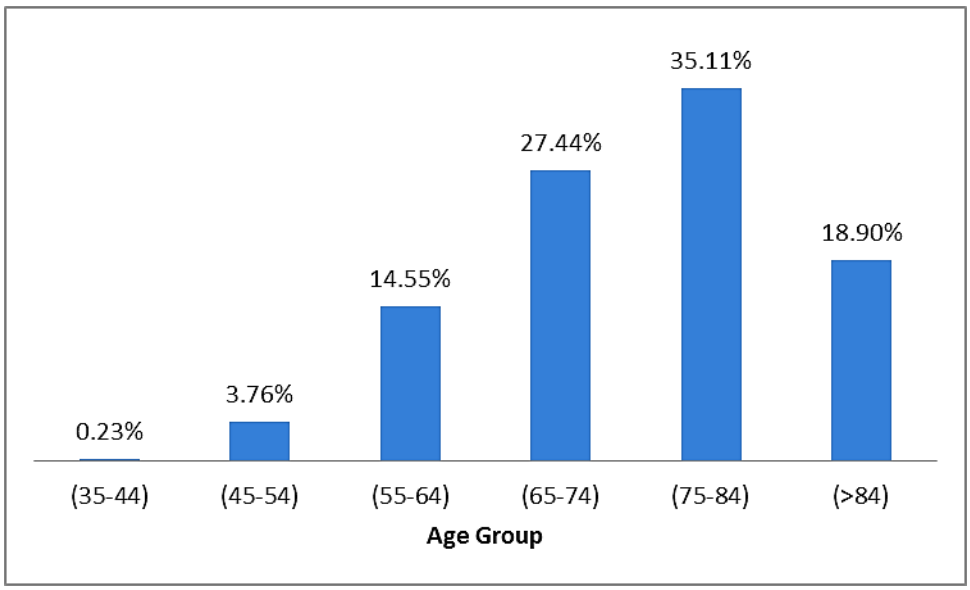

Figure 16: The age distribution of prostate cancer patients

On the other hand, the distribution of race is shown in Figure 17. Unfortunately, there are around 14% of race data is recorded as missing from the source.

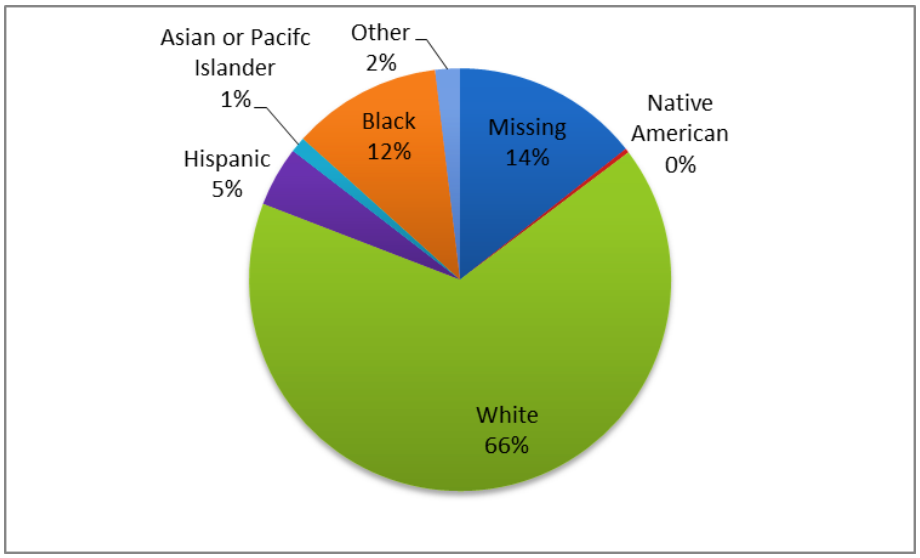

Figure 17: The race distribution of prostate cancer patients

Patients who are obese represent 8.29% of the whole data; 5.12% of them were diagnosed with prostate cancer. Alcoholic patients represent 3.40% of the whole data; 1.19% of

them were diagnosed with prostate cancer. Patients who are smoking represent 15.92% of the whole data; and 7.27% of them were diagnosed with prostate cancer. Also, patients who were diagnosed with fatty acid deficiency, vitamin D deficiency, inflammation of prostate, and vasectomy each one of them represent less than 0.35% of the whole data. Patients who were diagnosed with hypercholesterolemia represent 7.45% of the whole data. Hypertension patients represent 59.04% of the whole data.

4.2.2 Analysis of Risk Factors for Prostate Cancer

The relationship probability of the independent variables which are age, race, family history of prostate cancer, family history of any other type of cancer, obesity, physical activity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension were examined against prostate cancer diagnosis as dependent variable by using logistic regression analysis. The model of logistic regression assisted to identify which variables are critical in prediction of prostate cancer. The Receiver Operating Characteristic (ROC) curve has been used.

4.2.2.1 Dependent Variable: Prostate Cancer

With a total of 642,653 patients were diagnosed with prostate cancer in the simple data, prostate cancer variable was signed the value “1” for the patient who were diagnosed with prostate cancer; and “0” value for patient who were not diagnosed with prostate cancer. All patients are men with age of 35 years and older.

Table 6: Odds Ratio Estimates result

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Age	1.04	1.039	1.041
Age 35-44	0.069	0.064	0.074
Age 45-54	0.421	0.406	0.437
Age 55-64	0.988	0.962	1.015
Age 65-74	1.288	1.265	1.312
Age 75-84	1.256	1.242	1.27
White	1.139	1.13	1.147
Black	1.649	1.632	1.666
Hispanic	0.902	0.89	0.915
Asian	0.72	0.702	0.737
Native American	0.787	0.753	0.823
Other Race	0.99	0.97	1.009
Family history of prostate cancer	7.752	7.407	8.065
Family history of cancer	1.708	1.666	1.751
Obese	0.807	0.798	0.817
Alcohol Abuse	0.655	0.64	0.67
Smoking	0.746	0.739	0.754
Hypercholesterolemia	1.243	1.233	1.254
Vitamin D deficiency	1.082	1.04	1.126
Inflammation of prostate	1.376	1.325	1.429
Vasectomy	2.906	2.641	3.197
Hypertension	1.184	1.178	1.19

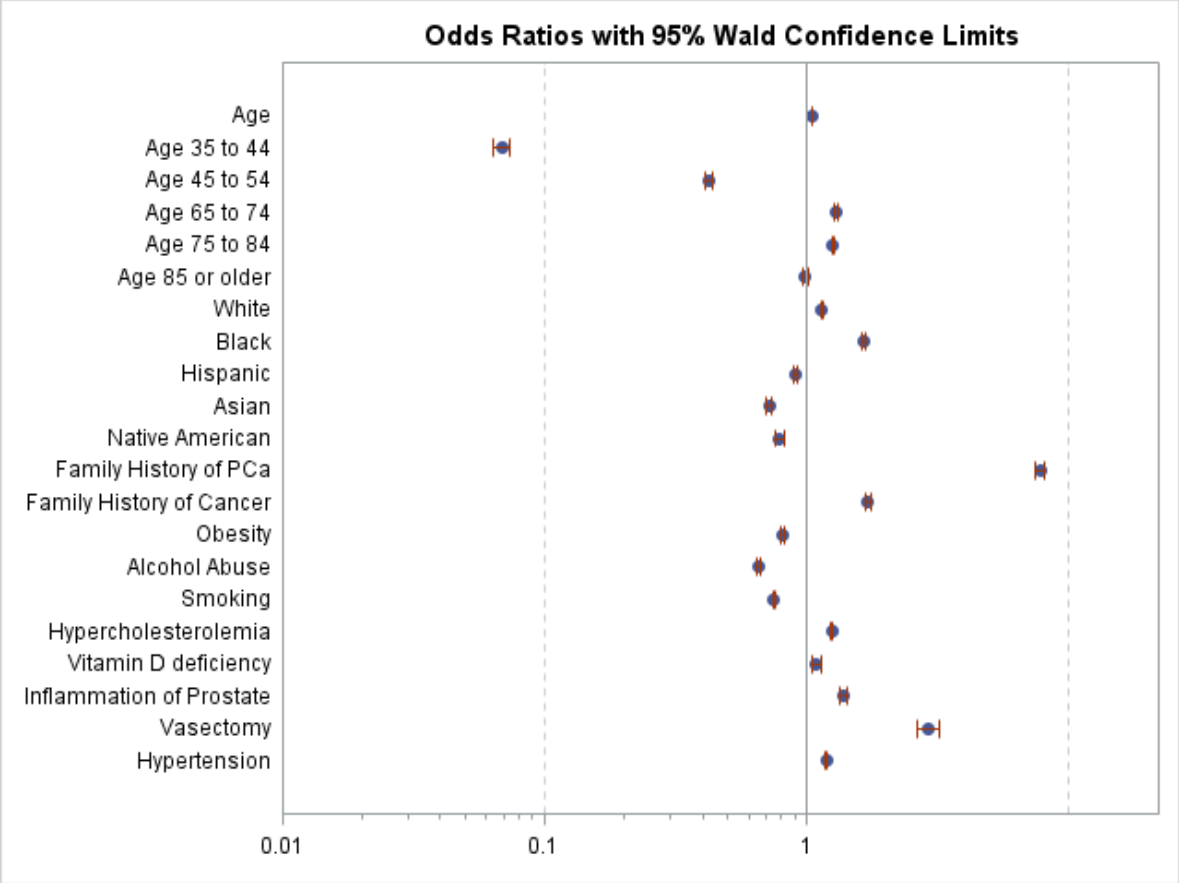

Figure 18: Odds Ratios Result

4.2.2.2 Independent Variable: Age

The variable age is shown to be predictive of prostate cancer ($p < 0.0001$) with an Odd Ratio = 1.041. The odds of developing prostate cancer are 1.041 times greater for each year increase in age.

The age group 35 to 44 years is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.070. This indicates that the odds of men with age between 35 and 44 years have decreased the risk of developing prostate cancer by 0.070 compering to older groups.

The age group 45 to 54 years is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.421. This indicates that the odds of men with age between 45 and 54 years have decreased the risk of developing prostate cancer by 0.421compering to older groups.

The age group 55 to 64 years is not significantly associated with the prediction of prostate cancer ($p = 0.2505$).

The age group 65 to 74 years is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.287. This indicates that the odds of men with age between 65 and 74 years have increased risk of developing prostate cancer by 1.287 greater than younger groups.

The age group 75 to 84 years is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.258. This indicates that the odds of men with age between 75 and 84 years have increased risk of developing prostate cancer by 1.258 greater than younger groups.

4.2.2.3 Independent Variable: Race

Race is statically significant with ($p <0.0001$). The white race is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.134. This indicates that the odds of white men have increased risk of developing prostate cancer by 1.134 greater than other race.

The black race is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.650. This indicates that the odds of black men have increased risk of developing prostate cancer by 1.650 greater than other race.

The Hispanic race is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.902. This indicates that the odds of Hispanic men have decreased the risk of developing prostate cancer by 0.902 compering to other race.

The Asian race is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.720. This indicates that the odds of Asian men have decreased the risk of developing prostate cancer by 0.720 compering to other race.

The Native American race is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.789. This indicates that the odds of Native men have decreased the risk of developing prostate cancer by 0.789 compering to other race. Patients who are classified as other race are not significantly associated with the prostate cancer ($p = 0.3178$).

4.2.2.4 Independent Variable: Positive family history of prostate cancer

The variable family history of prostate cancer is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 7.752. This indicates that the odds of men with positive family history of prostate cancer have increased the risk of developing prostate cancer by 7.752 greater than men with no family history of prostate cancer.

4.2.2.5 Independent Variable: Positive family history of any other type of cancer

The variable family history of any other type of cancer is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.717. This indicates that the odds of men with positive family history of any other type of cancer have increased the risk of developing prostate cancer by 1.717 greater than men with no family history of other cancer.

4.2.2.6 Independent Variable: Obesity

The variable obesity is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.808. This indicates that obesity is not good in predicting prostate cancer.

4.2.2.7 Independent Variable: Physical activity

For all three type of physical activity, vigorous ($p = 0.3255$), light-moderate ($p = 0.5866$), strengthening activity ($p = 0.9724$), none of them show a significant association with prostate cancer.

4.2.2.8 Independent Variable: Alcohol Abuse

The variable alcohol abuse is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.657. This indicates that alcohol is not good in predicting prostate cancer.

4.2.2.9 Independent Variable: Smoking

The variable smoking is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 0.747. This indicates that smoking is not good in predicting prostate cancer.

4.2.2.10 Independent Variable: Fat intake

Fatty acid deficiency is not statically significant to be associated with the prostate cancer ($p = 0.5404$). However, the variable hypercholesterolemia is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.243. This indicates that the odds of hypercholesterolemia patients have increased the risk of developing prostate cancer by 1.243 greater than men who do not diagnosis with hypercholesterolemia.

4.2.2.11 Independent Variable: Vitamin D deficiency

The variable vitamin D deficiency is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.079. This indicates that the odds of men who have vitamin D deficiency have increased the risk of developing prostate cancer by 1.079 greater than men who do not have vitamin D deficiency.

4.2.2.12 Independent Variable: Inflammation of prostate

The variable inflammation of prostate is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.351. This indicates that the odds of men who have prostate inflammation have increased the risk of developing prostate cancer by 1.351 greater than men who do not have prostate inflammation.

4.2.2.13 Independent Variable: Vasectomy

The variable vasectomy is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 2.906. This indicates that the odds of men who have vasectomy have increased the risk of developing prostate cancer by 2.906greater than men who do not have vasectomy.

4.2.2.14 Independent Variable: Hypertension

The variable hypertension is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 1.184. This indicates that the odds of men who have hypertension have increased the risk of developing prostate cancer by 1.184 greater than men who do not have hypertension.

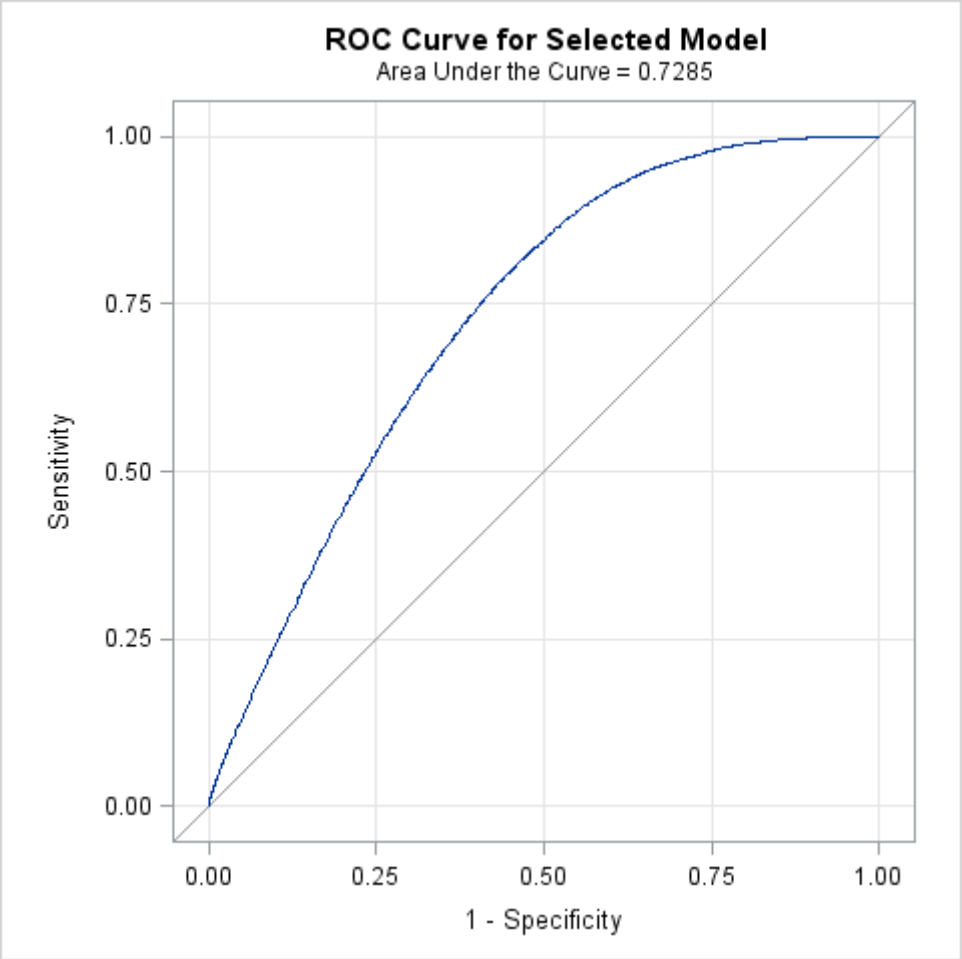

Figure 19: ROC Curves

4.2.3 Logistic Regression Classification Model

Logistic Regression Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The bias error is the differences between mean classification and true mean, while variance error is the differences between prediction and mean prediction on average. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the logistic regression model case, the testing error is less than the training error as shown in Figure 20.

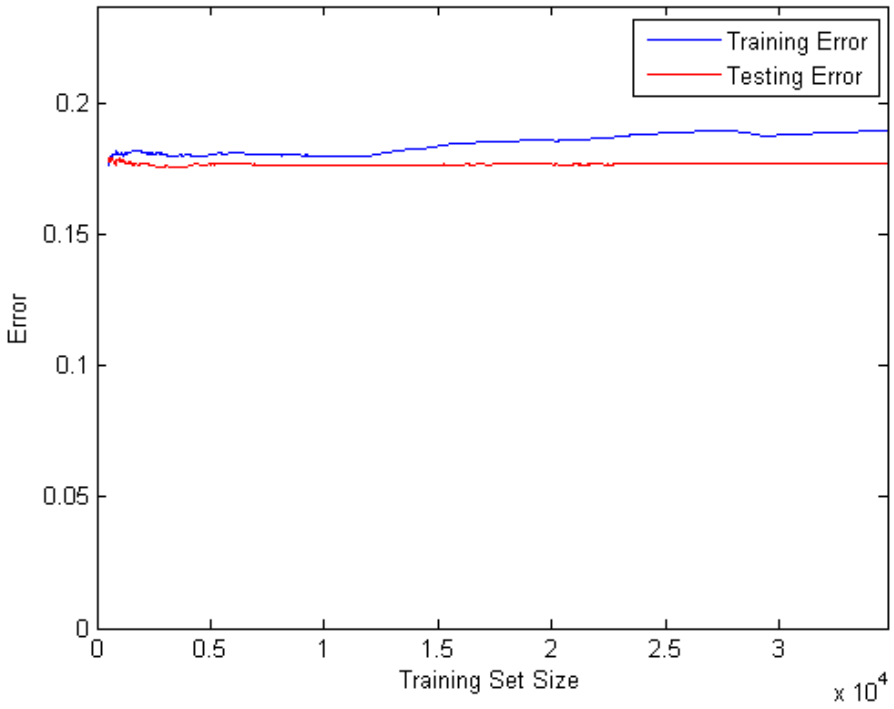

Figure 20: Logistic Regression Bias-Variance Learning Curve

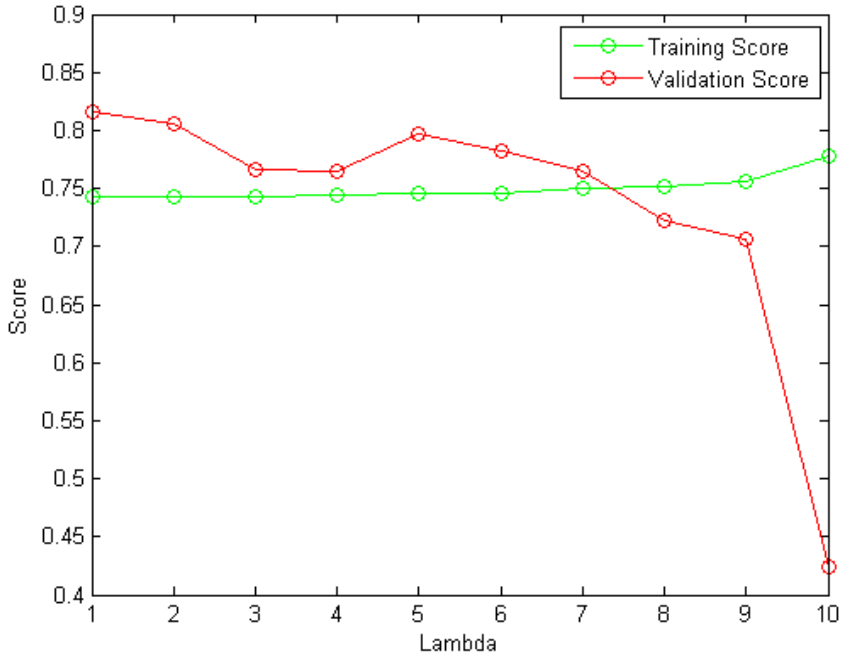

Figure 21 : Logistic Regression Validation Curve

The both curves started almost from same point. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above 0.15.

On the other hand, the Logistic Regression Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 21. In this case, the performance of Logistic Regression classifier model is acceptable.

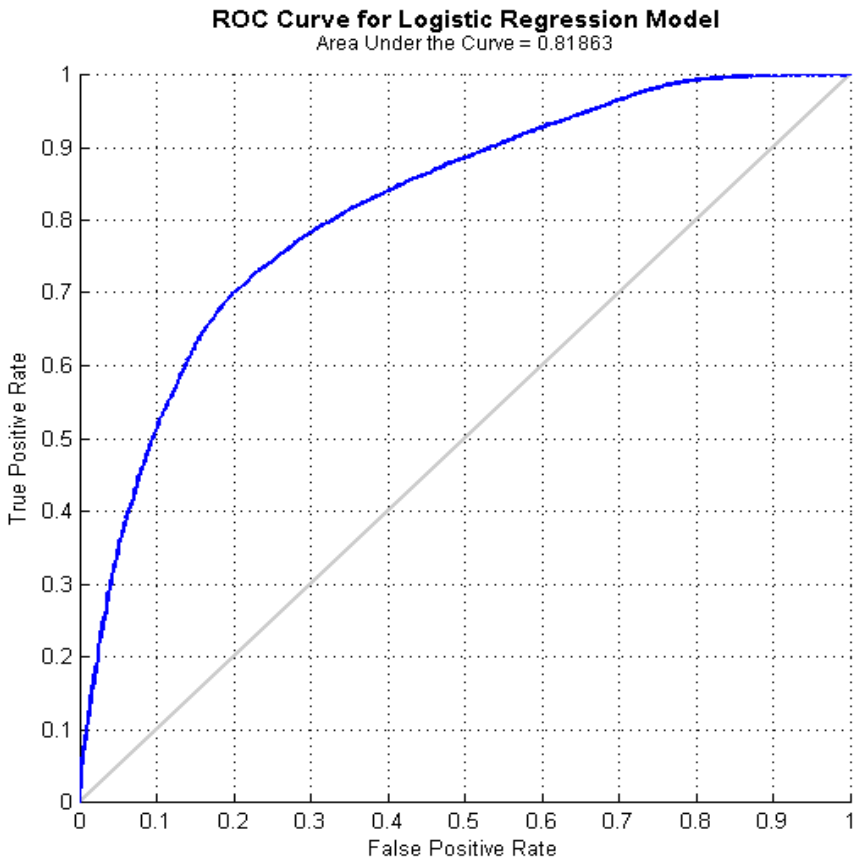

Figure 22 : Logistic Regression Classifier Model ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.82 as shown in Figure 22. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 75.2%, while the misclassification rate is around 24.8%. In addition, the sensitivity (True Positive Rate) is around 71.6%; and the specificity (True Negative Rate) is around 78.1%.

4.2.4 k-Nearest Neighbors (KNN)

k-Nearest Neighbors Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the k-Nearest Neighbors model case, the testing error is higher than the training error as shown in Figure 23. At the beginning, testing error showed high bias by decreasing. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level.

The k-Nearest Neighbors model was suffering from high variance. As a result, some features were removed to minimize the variance errors.

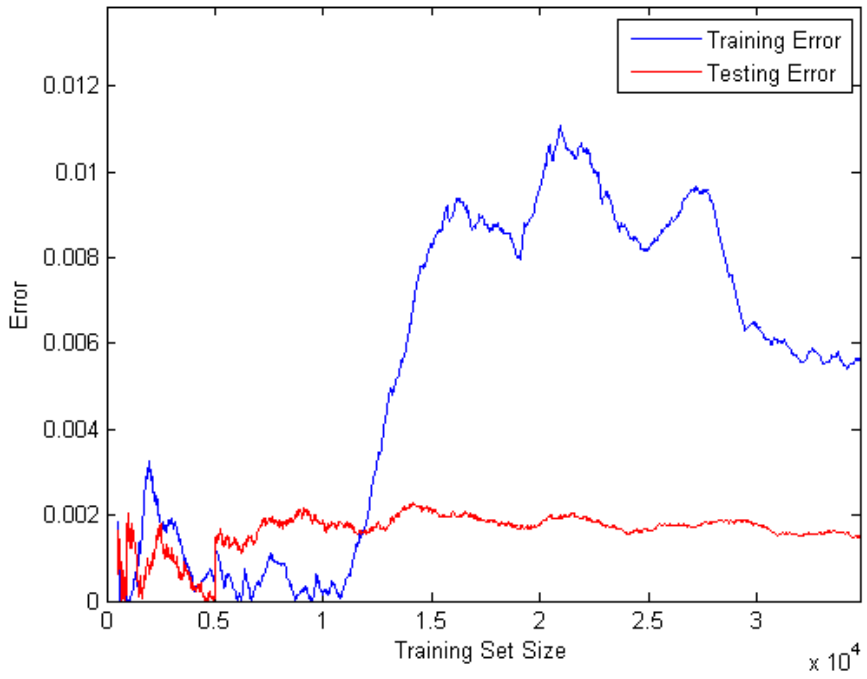

Figure 23 : k-Nearest Neighbors Bias-Variance Learning Curve

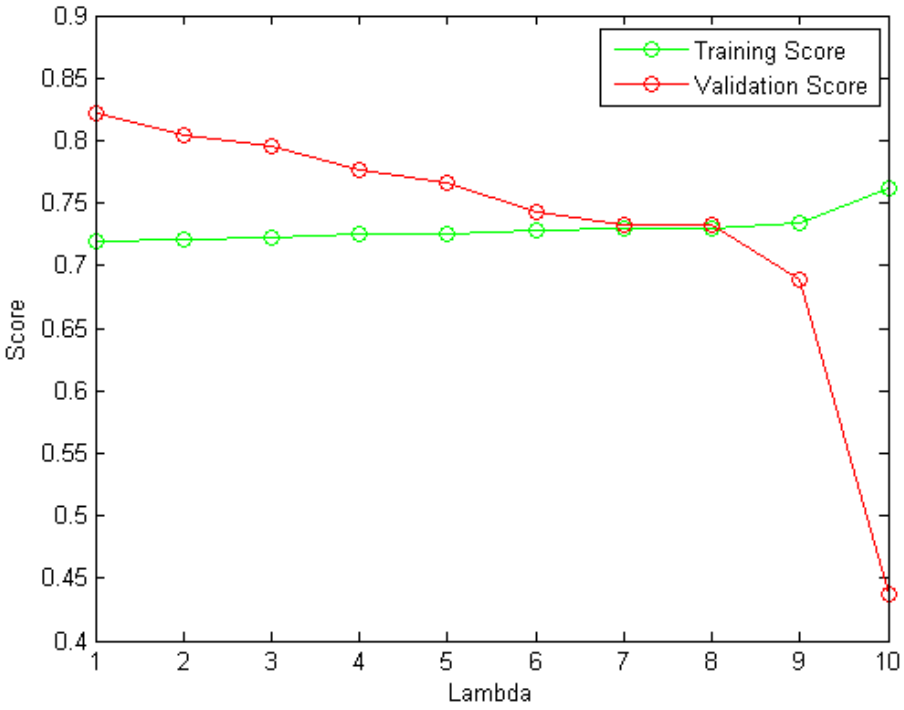

Figure 24 : k-Nearest Neighbors Validation Curve

On the other hand, the k-Nearest Neighbors Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 24. In this case, the performance of k-Nearest Neighbors classifier model is acceptable.

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.73 as shown in Figure 25. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 72.8%, while the misclassification rate is around 27.2%. In addition, the sensitivity (True Positive Rate) is around 75.6%; and the specificity (True Negative Rate) is around 70.4%.

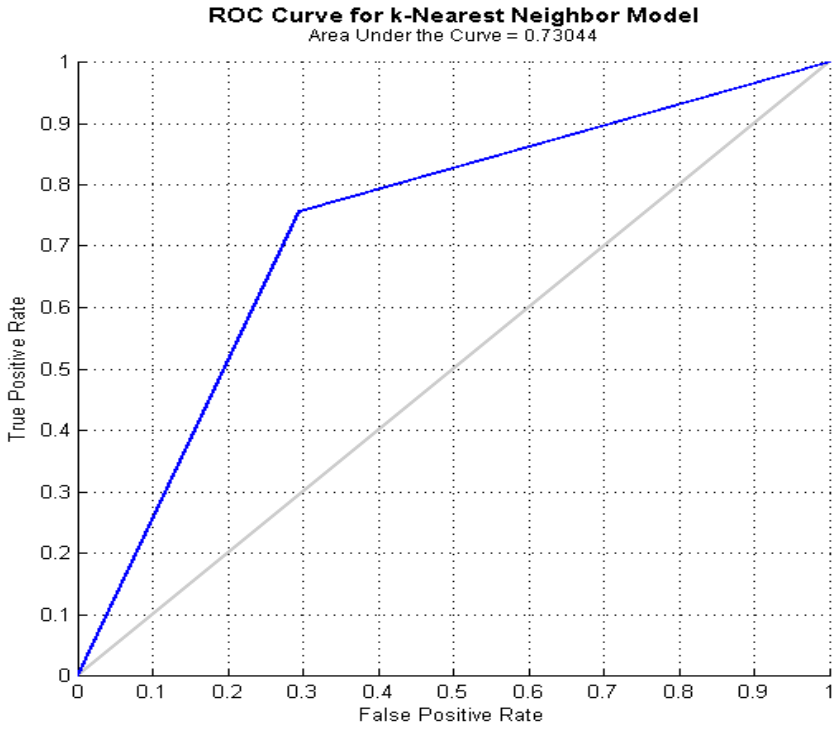

Figure 25 : k-Nearest Neighbors ROC Curve

4.2.5 Naive Bayes Classifier

Naive Bayes Classifier Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the Naive Bayes Classifier model case, the testing error is less than the training error as shown in Figure 26. At the beginning, testing error showed high bias by decreasing. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level.

The Naive Bayes Classifier model was suffering from high variance. As a result, some features were removed to minimize the variance errors.

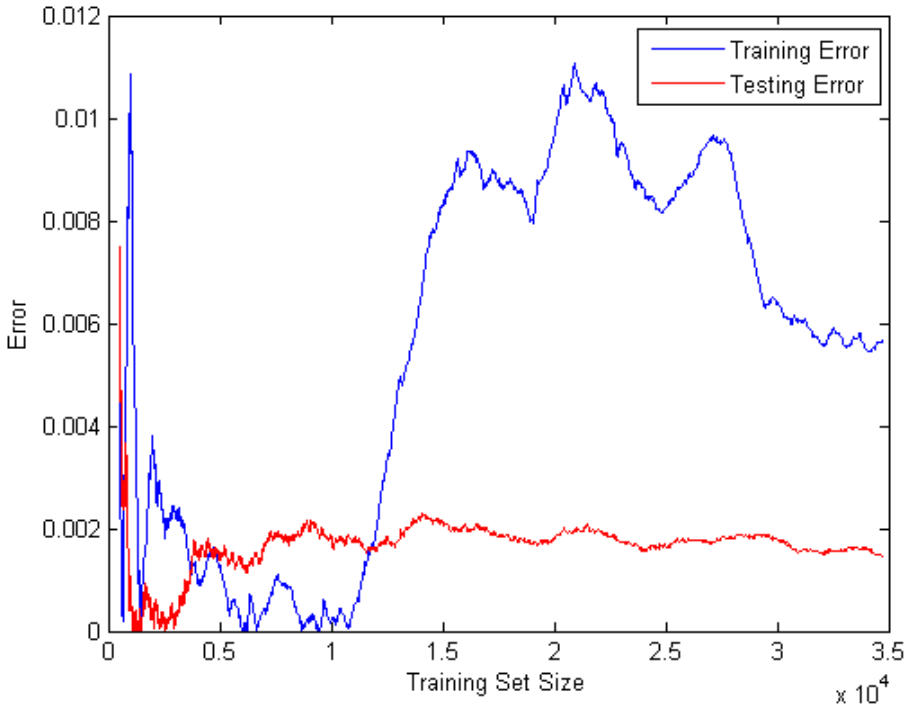

Figure 26 : Naive Bayes Classifier Bias-Variance Learning Curve

On the other hand, the Naive Bayes Classifier Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 27. In this case, the performance of Naive Bayes Classifier model is acceptable.

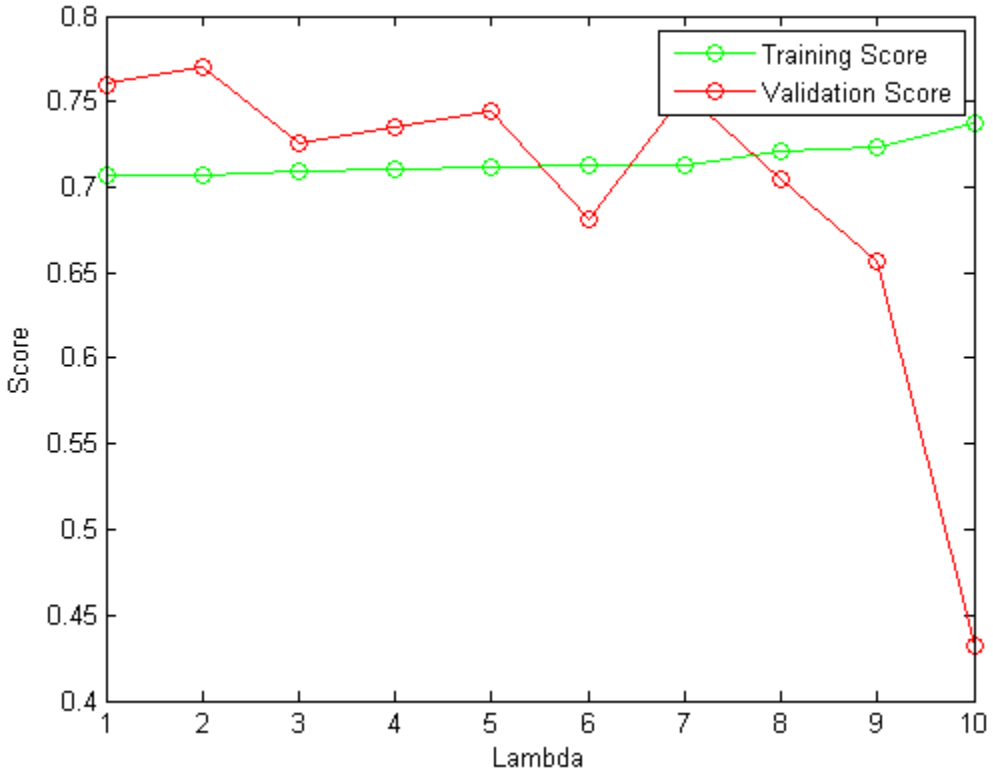

Figure 27 : Naive Bayes Classifier Validation Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.71 as shown in Figure 28. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 71.2%, while the misclassification rate is around 28.8%. In addition, the sensitivity (True Positive Rate) is around 66.4%; and the specificity (True Negative Rate) is around 75.1%.

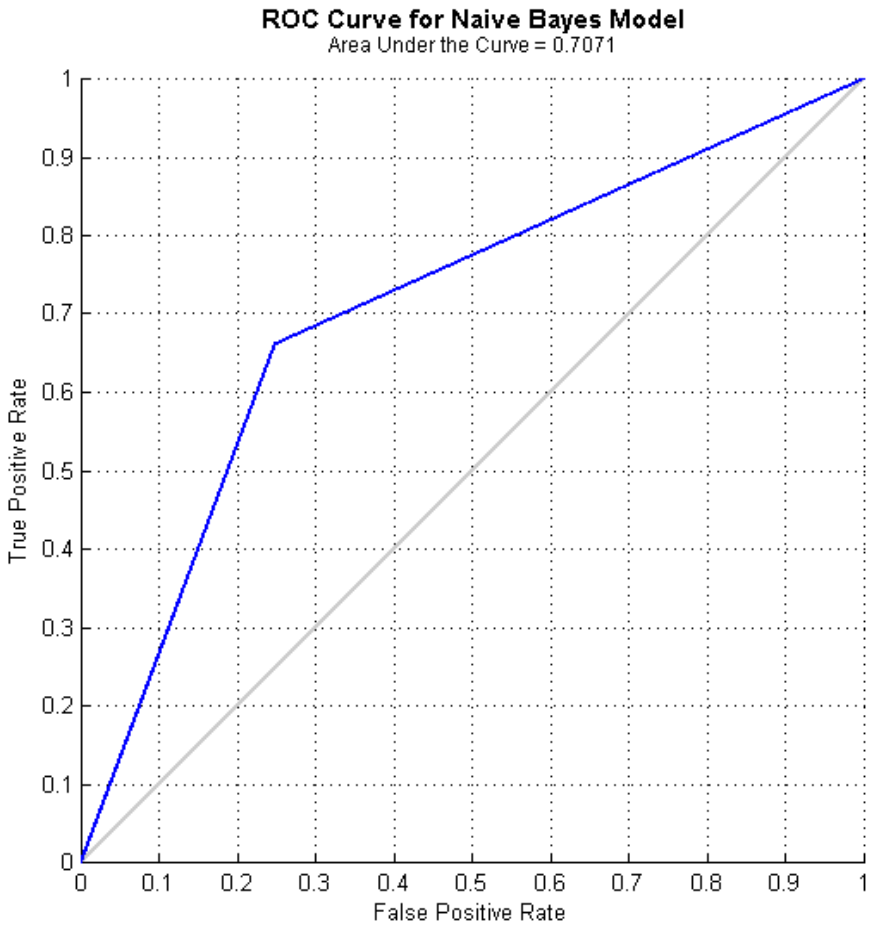

Figure 28 : Naive Bayes Classifier ROC Curve

4.2.6 Decision Tree Classifier

Decision Tree Classifier Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the Decision

Tree Classifier model case, the testing error is less than the training error as shown in Figure 29. At the beginning, testing error showed high bias by decreasing. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level.

The Decision Tree Classifier model was suffering from high variance. As a result, some features were removed to minimize the variance errors.

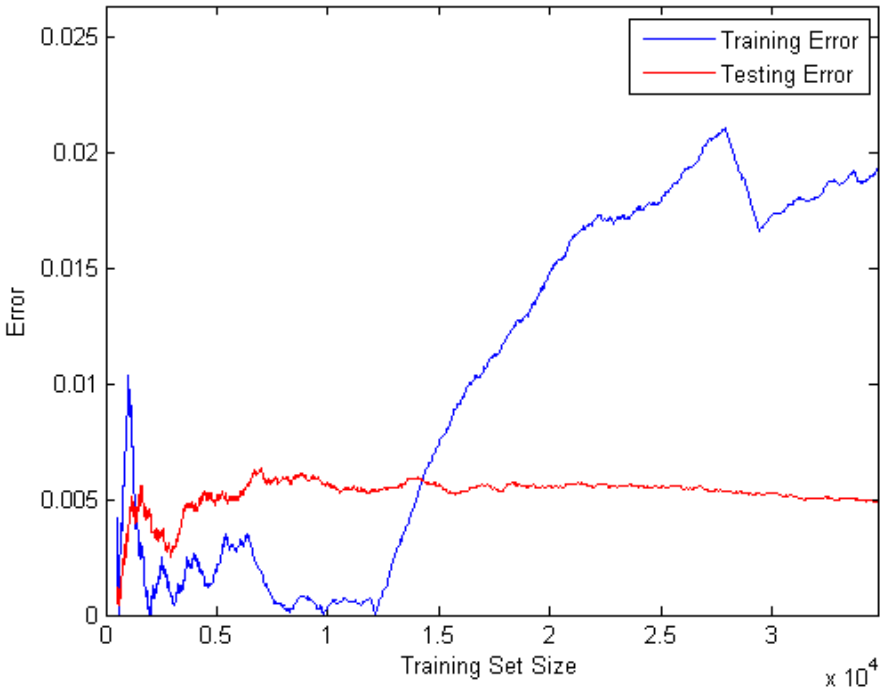

Figure 29 : Decision Tree Classifier Bias-Variance Learning Curve

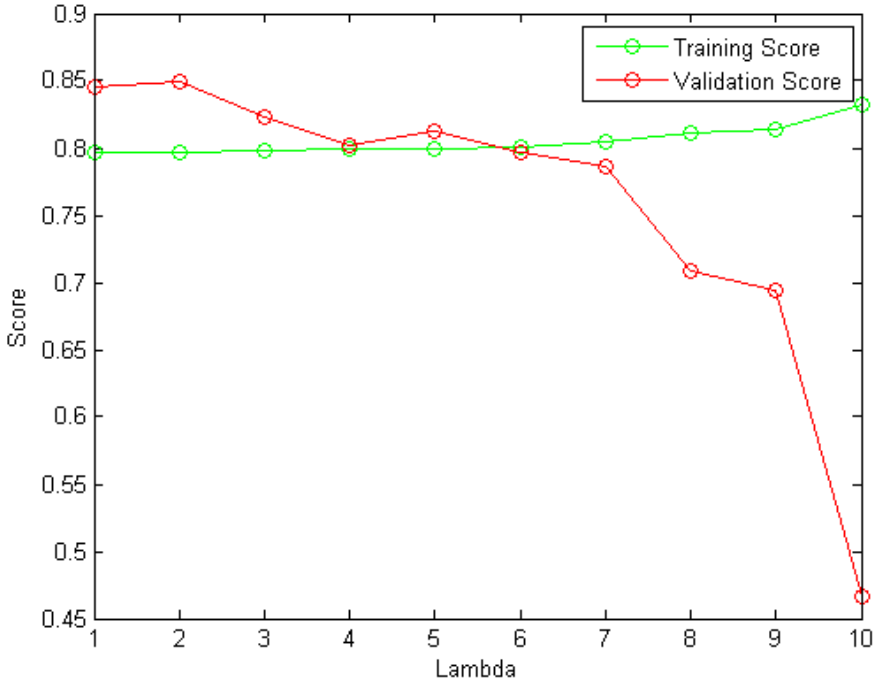

Figure 30 : Decision Tree Classifier Validation Curve

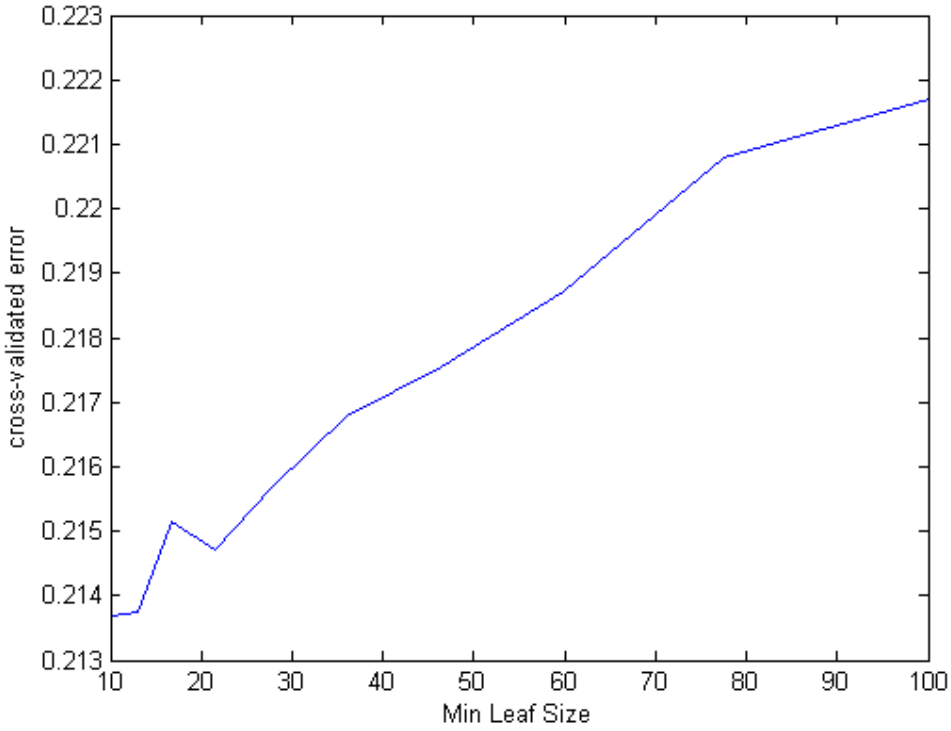

Figure 31 : Decision Tree Classifier Validation Curve Leaf Size

On the other hand, the Decision Tree Classifier Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 30. In addition, the best leaf size is under 23 as shown in Figure 31. In this case, the performance of Decision Tree Classifier model is acceptable.

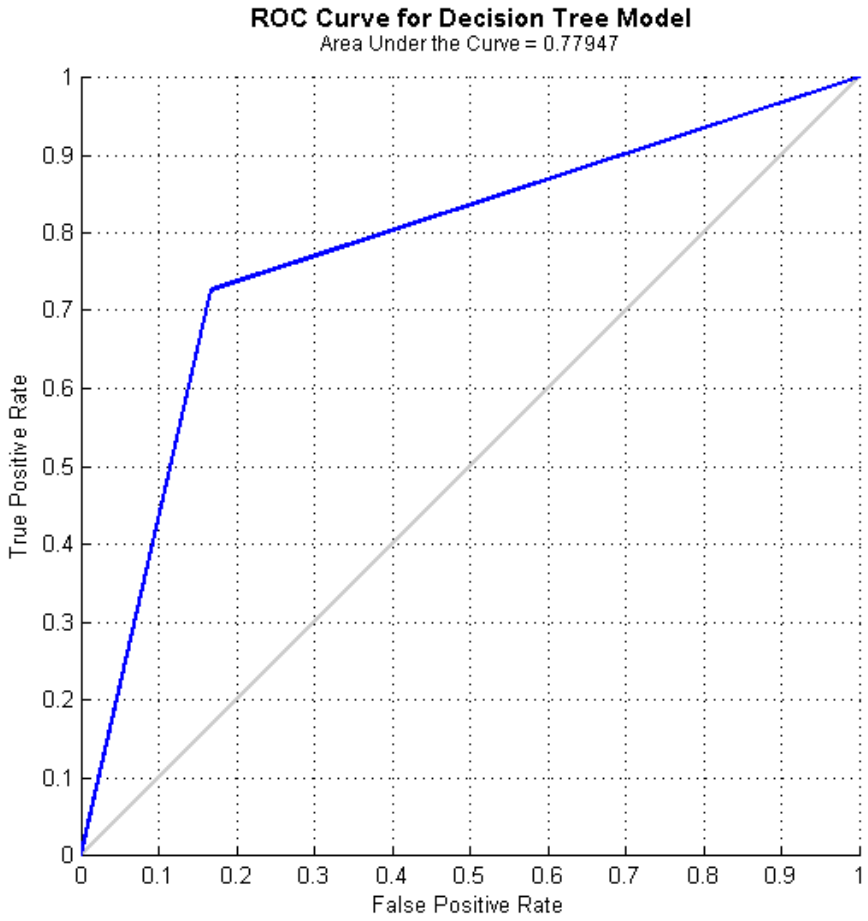

Figure 32 : Decision Tree Classifier ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.78 as shown in Figure 32. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 78.5%, while the misclassification rate is around 21.5%. In addition, the sensitivity (True Positive Rate) is around 72.7%; and the specificity (True Negative Rate) is around 83.2%.

4.2.7 Support Vector Machines

Support Vector Machines Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the Support Vector Machines model case, the testing error is higher than the training error as shown in Figure 33.

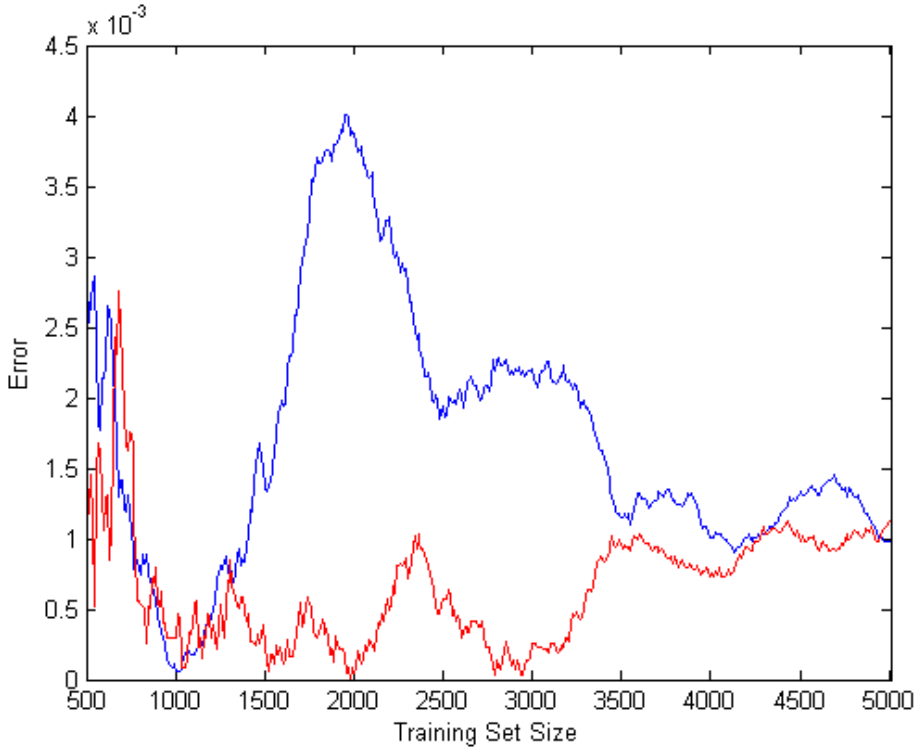

Figure 33 : Support Vector Machines Bias-Variance Learning Curve

At the beginning, testing error showed high bias by decreasing. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level.

The Support Vector Machines model was suffering from high variance.

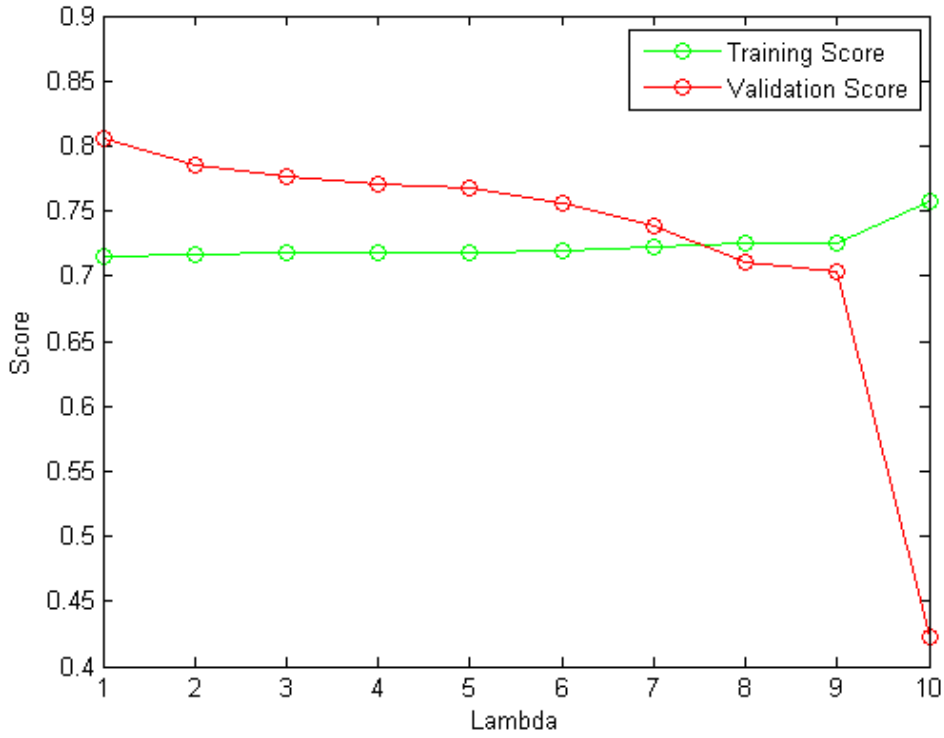

Figure 34 : Support Vector Machines ROC Curve

On the other hand, the Support Vector Machines Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 34. In this case, the performance of Support Vector Machines model is acceptable.

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.72 as shown in Figure 35. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 72.2%, while the misclassification rate is around 27.8%. In addition, the sensitivity (True Positive Rate) is around 73.2%; and the specificity (True Negative Rate) is around 71.4%.

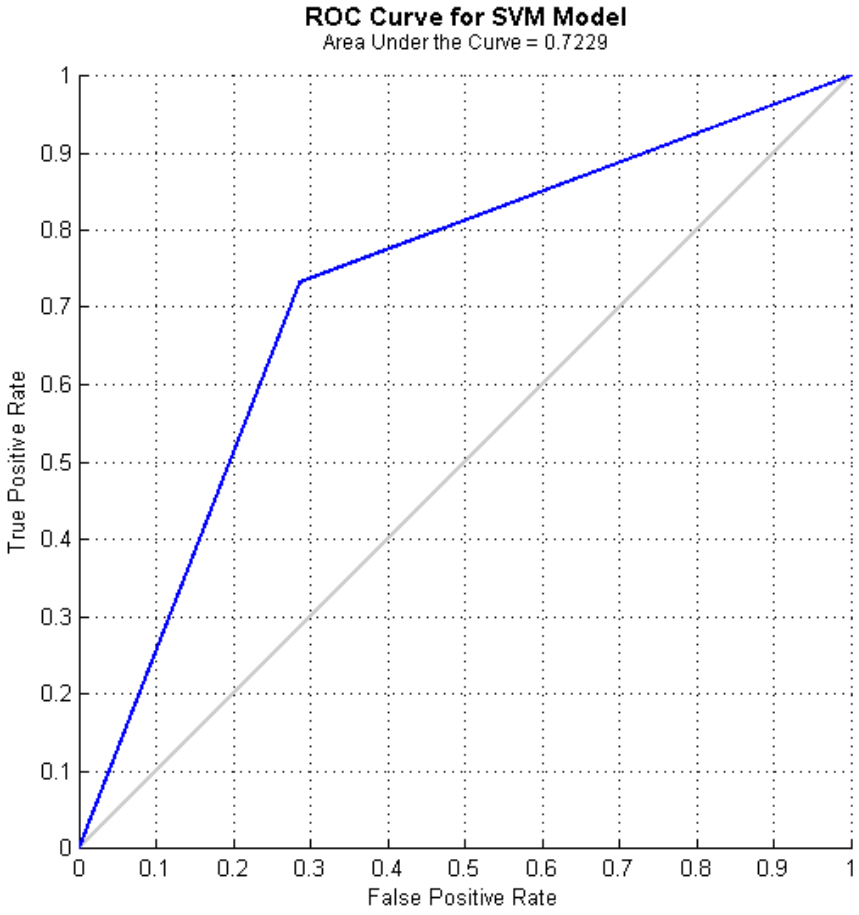

Figure 35 : Support Vector Machines ROC Curve

4.2.8 Artificial Neural Network

First, data was divided randomly to three sets (training, validation, and testing). Then, the network used scaled conjugate gradient back-propagation function to update weight and bias values during training phase. The Neural Network training, validation, and test performance is shown in Figure 36. The network stopped training when the validation error increased for six iterations, which occurred at iteration 190. The best validation performance occurred at iteration 184, which means there is no significant over-fitting had occurred by iteration 184 as shown in Figure 37.

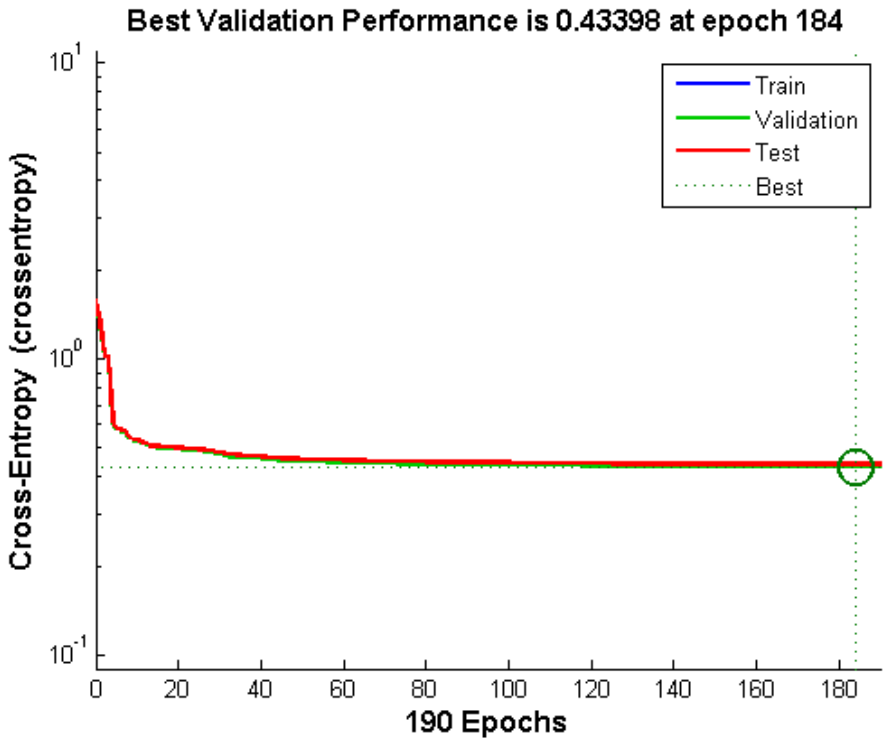

Figure 36 : Artificial Neural Network training, validation, and test performance

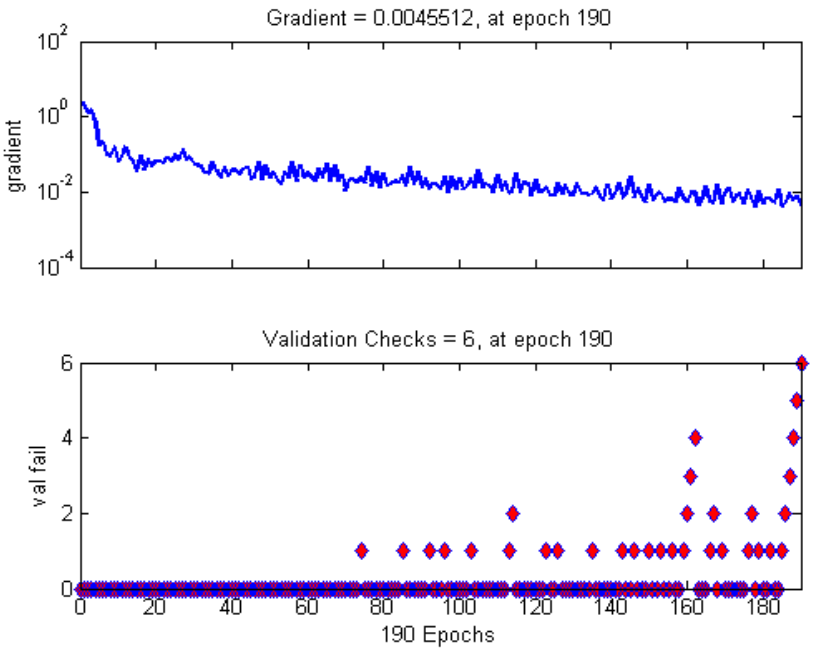

Figure 37 : Artificial Neural Network training state values

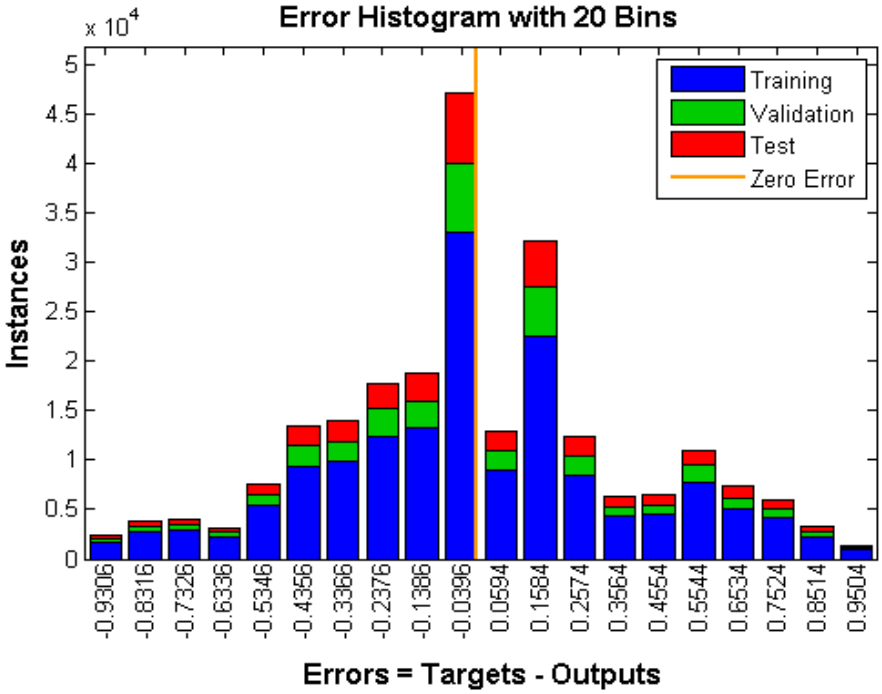

Figure 38 : Artificial Neural Network Error Histogram

The error histogram in Figure 38 show more verification of network performance. Most errors are fallen close to zero error. However, there are training points located far from zero error representing outliers' data.

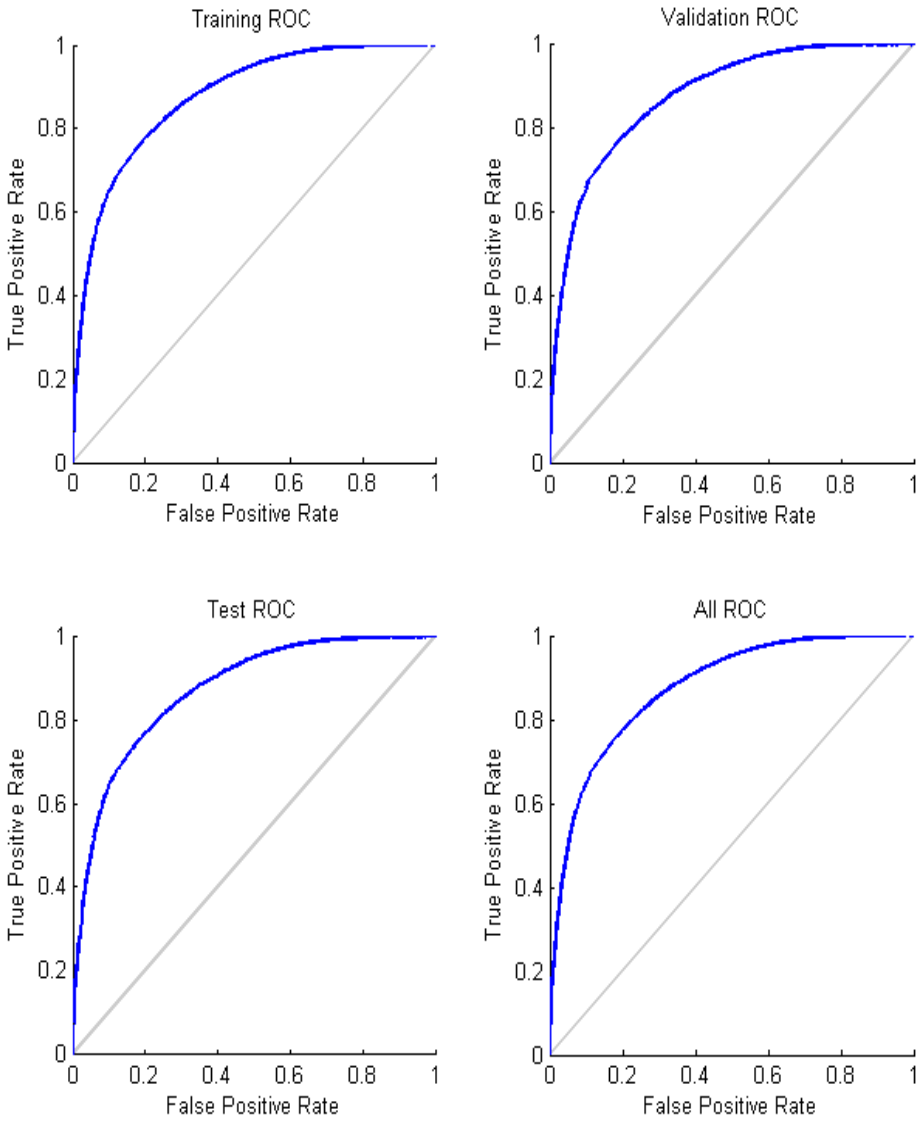

Figure 39 : Artificial Neural Network ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) for all data is around 0.88 as shown in Figure 39. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 79.1%, while the misclassification rate is around 20.9%. In addition, the sensitivity (True Positive Rate) is around 71.6%; and the specificity (True Negative Rate) is around 85.2%.

For training data, the model accuracy is approximately 79.1%, while the misclassification rate is around 20.9%. In addition, the sensitivity (True Positive Rate) is around 71.5%; and the specificity (True Negative Rate) is around 85.3%.

For validation data, the model accuracy is approximately 79.3%, while the misclassification rate is around 20.7%. In addition, the sensitivity (True Positive Rate) is around 71.9%; and the specificity (True Negative Rate) is around 85.4%.

For test data, the model accuracy is approximately 78.9%, while the misclassification rate is around 21.1%. In addition, the sensitivity (True Positive Rate) is around 71.7%; and the specificity (True Negative Rate) is around 84.7%.

4.3 SEER Data Analysis

4.3.1 Descriptive Statistics

The Surveillance, Epidemiology, and End Results (SEER) Program is the leading source of information for cancer patients. SEER data from 2004 to 2011 with 172,545 samples are utilized for this study. Patients' age is between 35 years and 100 years with average age of 66 years old; and the standard deviation is 10.3. Patients' age distribution is shown in Figure 40.

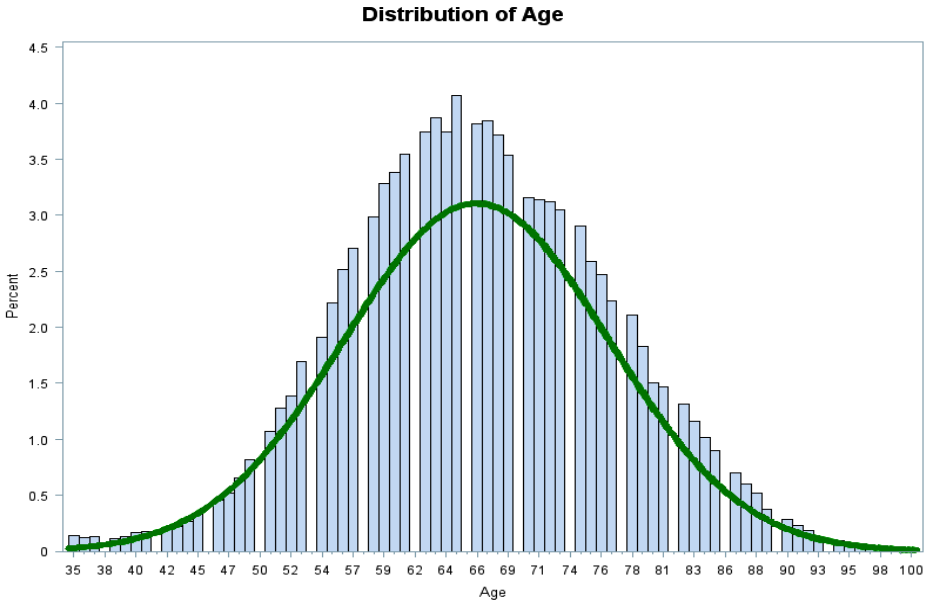

Figure 40 : The age distribution of prostate cancer patients in SEER data

The majority of patients are white race represent approximately 79% of all data. Black race represent roughly around 13%. Patients' race distribution is shown in Figure 41.

Data contains Prostatic Specific Antigen (PSA) lab result and its interpretation as positive or negative. Data shows that 76.85% of patients PSA results are positive; and around

6.68% are negative, while 1.89% is on border neither positive nor negative. However, there is around 14.58% of patients have no PSA result interpretation.

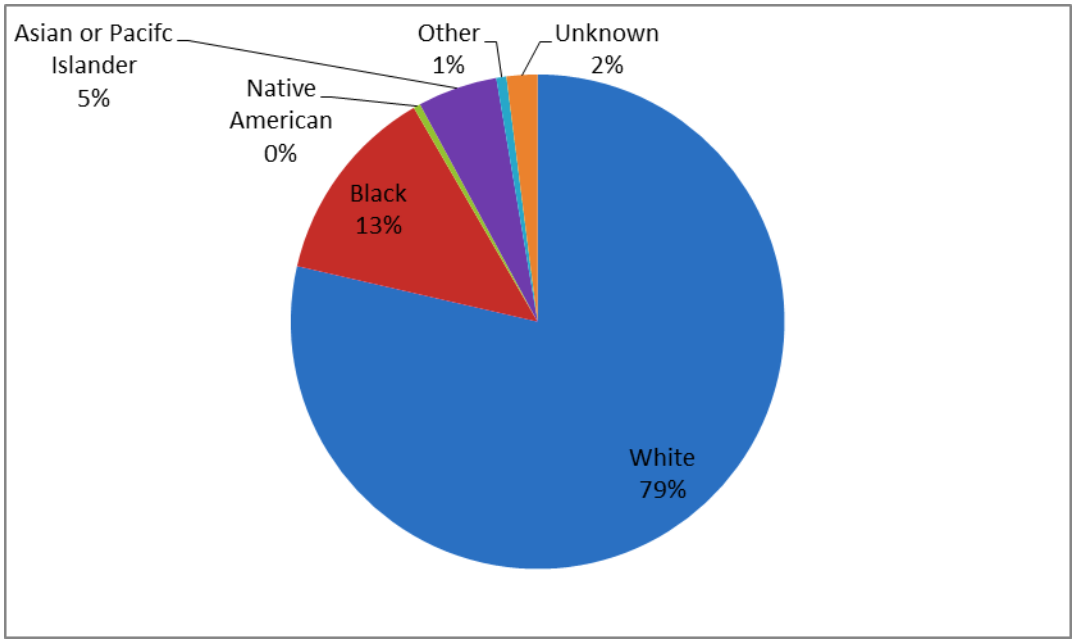

Figure 41: SEER race distribution

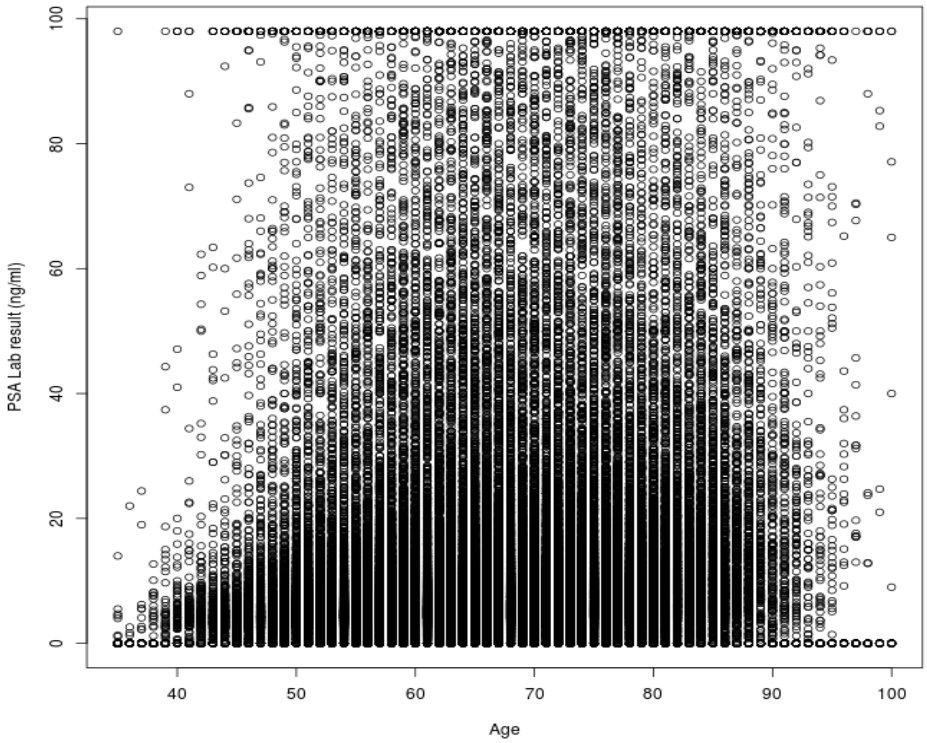

Figure 42 : PSA lab value distribution over age

Prostatic Specific Antigen (PSA) lab value distribution over patients' age is shown in Figure 42. Unfortunately, there are around 14.77% of patients PSA lab value have not been documented in their records. Also, around 2.27% of patients did not have the PSA teat.

4.3.2 Logistic Regression Classification Model

Logistic Regression Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The bias error is the differences between mean classification and true mean, while variance error is the differences between prediction and mean prediction on average. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the logistic regression model case, the testing error is less than the training error as shown in Figure 43.

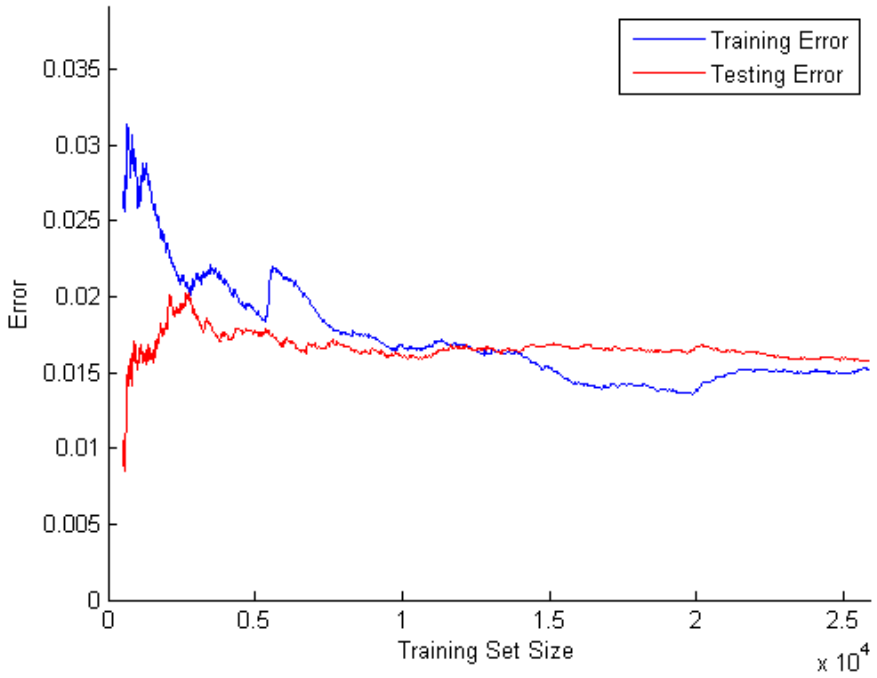

Figure 43 : Logistic Regression Bias-Variance Learning Curve

At the beginning, testing error showed high bias. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above 0.015.

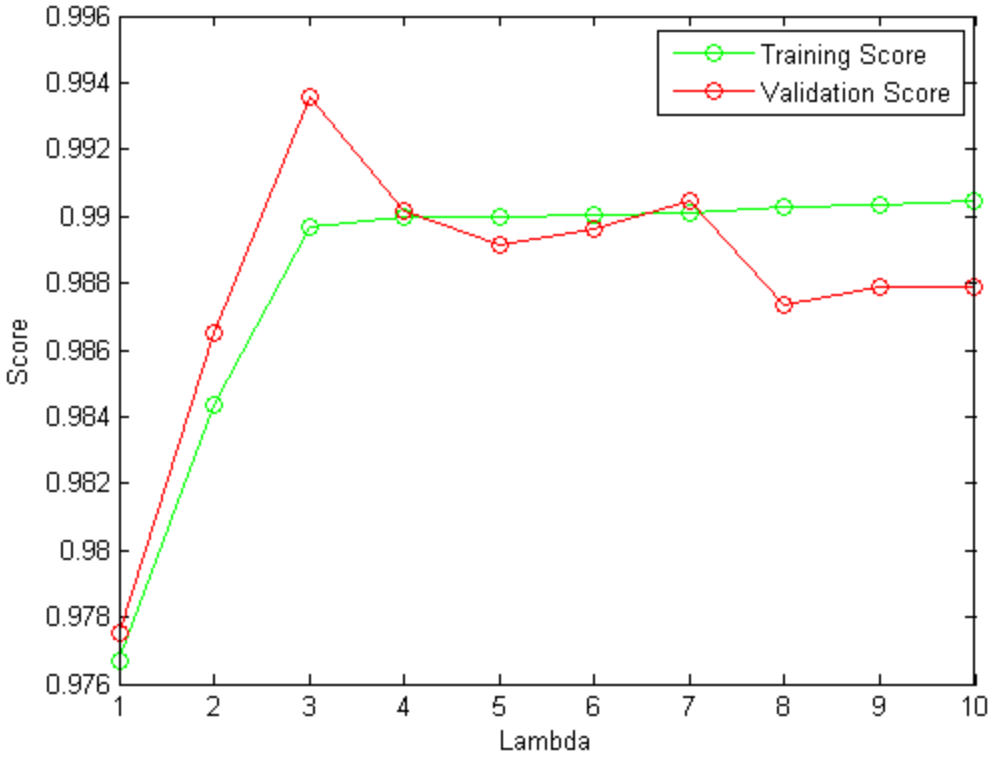

Figure 44 : Logistic Regression Validation Curve

On the other hand, the Logistic Regression Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 44. In this case, the performance of Logistic Regression classifier model is acceptable.

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.85 as shown in Figure 45. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 98.5%, while the misclassification rate is around 1.5%. In addition, the sensitivity (True Positive Rate) is around 99.4%; and the specificity (True Negative Rate) is around 67.3%.

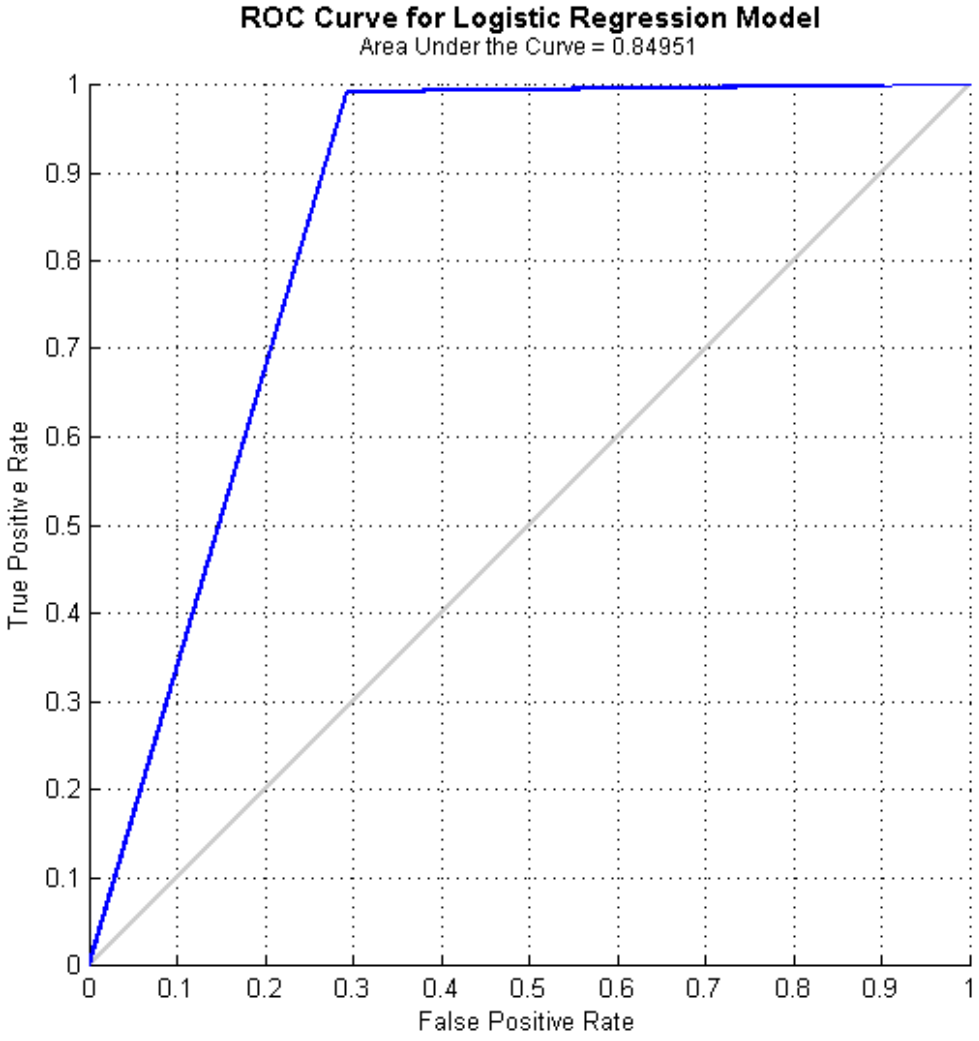

Figure 45 : Logistic Regression ROC Curve

4.3.3 k-Nearest neighbors (KNN)

k-Nearest Neighbors Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the k-Nearest Neighbors model case, the testing error is higher than the training error as shown in Figure 46.

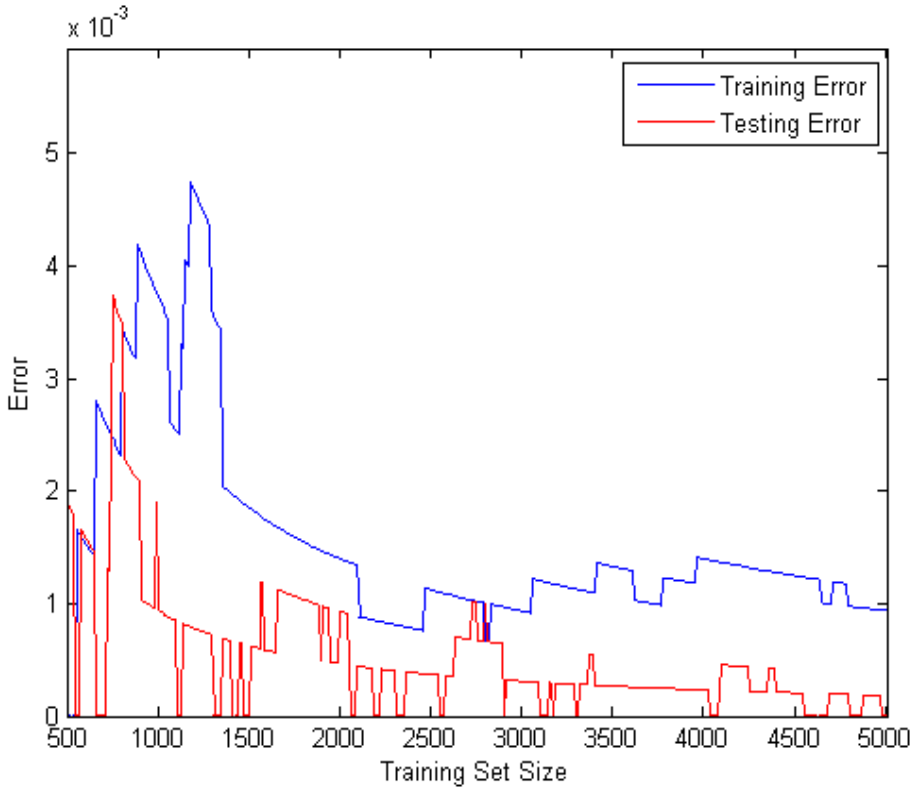

Figure 46 : k-Nearest Neighbors Bias-Variance Learning Curve

At the beginning, testing error showed high bias. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance.

It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level.

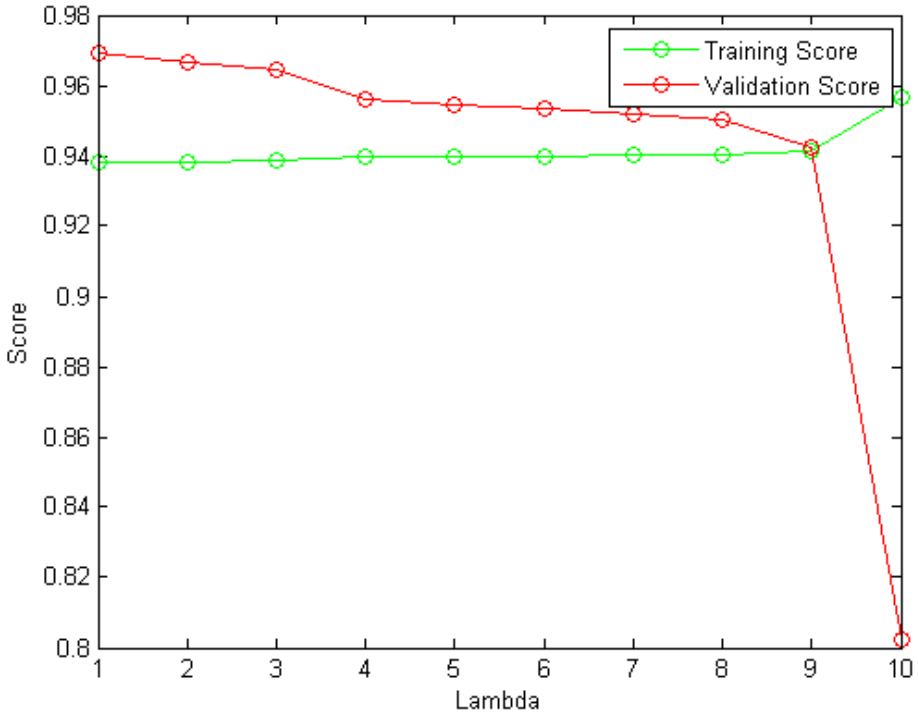

Figure 47 : k-Nearest Neighbors Validation Curve

On the other hand, the k-Nearest Neighbors Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 47. In this case, the performance of k-Nearest Neighbors classifier model is acceptable.

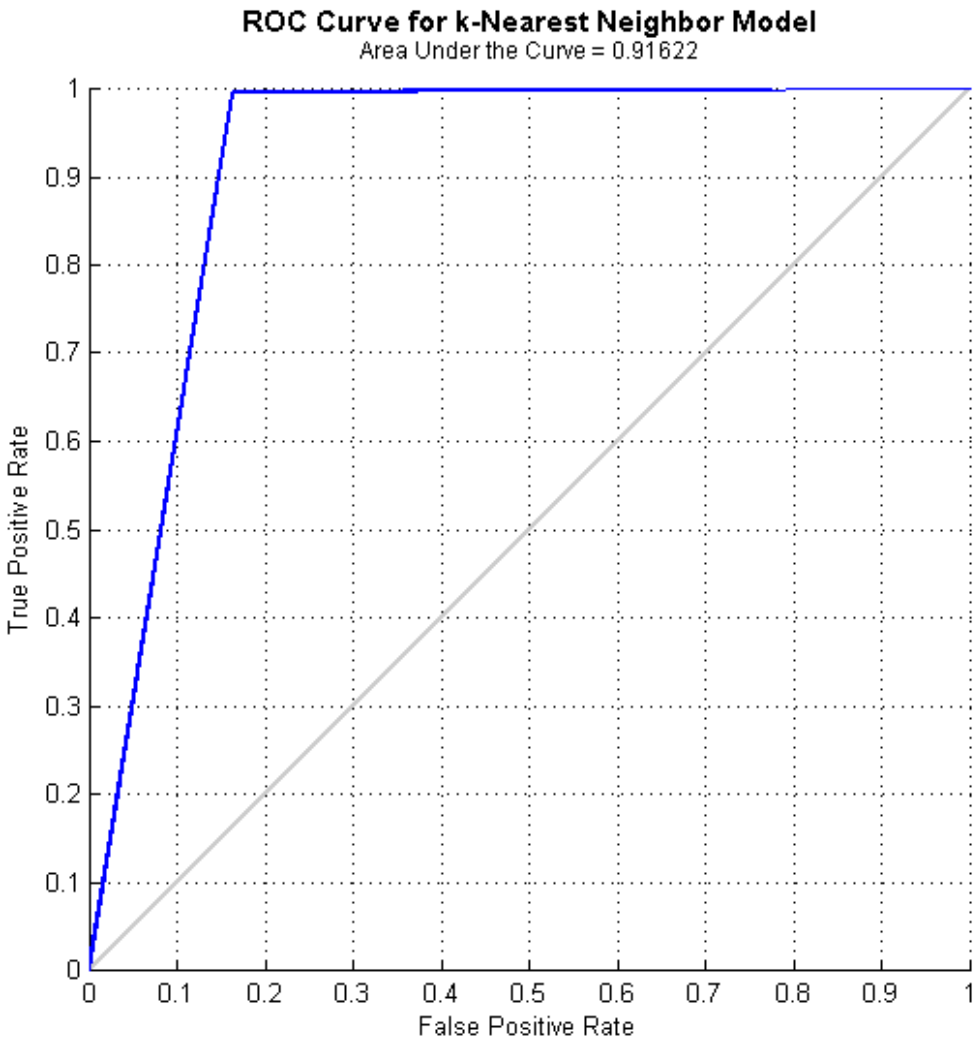

Figure 48 : k-Nearest Neighbors ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.92 as shown in Figure 48. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 98.6%, while the misclassification rate is around 1.4%. In addition, the sensitivity (True Positive Rate) is around 99.2%; and the specificity (True Negative Rate) is around 77.4%.

4.3.4 Naive Bayes Classifier

Naive Bayes Classifier Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the Naive Bayes Classifier model case, the testing error is higher than the training error as shown in Figure 49. At the beginning, testing error showed high bias. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above 0.06.

Some features that have zero variance were removed to correct the classifier performance.

On the other hand, the Naive Bayes Classifier Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 50. In this case, the performance of Naive Bayes Classifier model is acceptable.

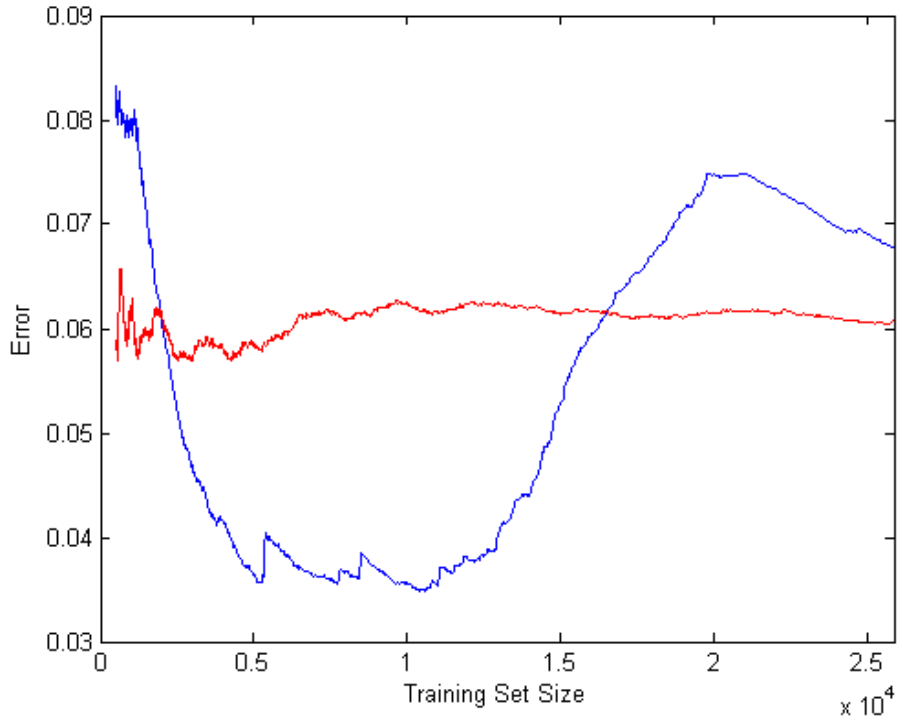

Figure 49 : Naive Bayes Classifier Bias-Variance Learning Curve

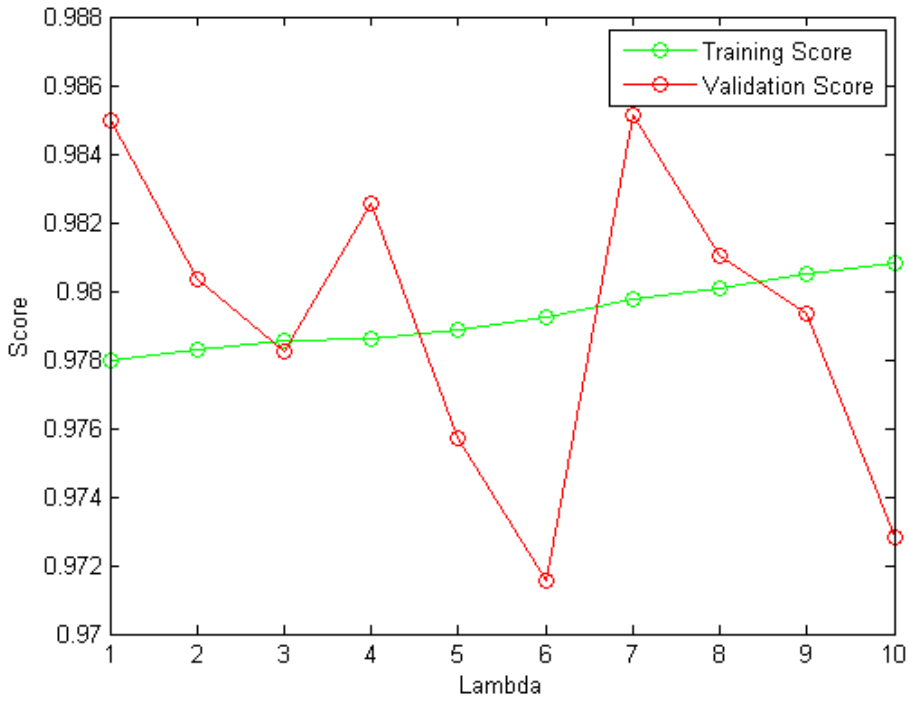

Figure 50 : Naive Bayes Classifier Validation Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.85 as shown in Figure 51. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 98.0%, while the misclassification rate is around 2.0%. In addition, the sensitivity (True Positive Rate) is around 98.8%; and the specificity (True Negative Rate) is around 71.7%.

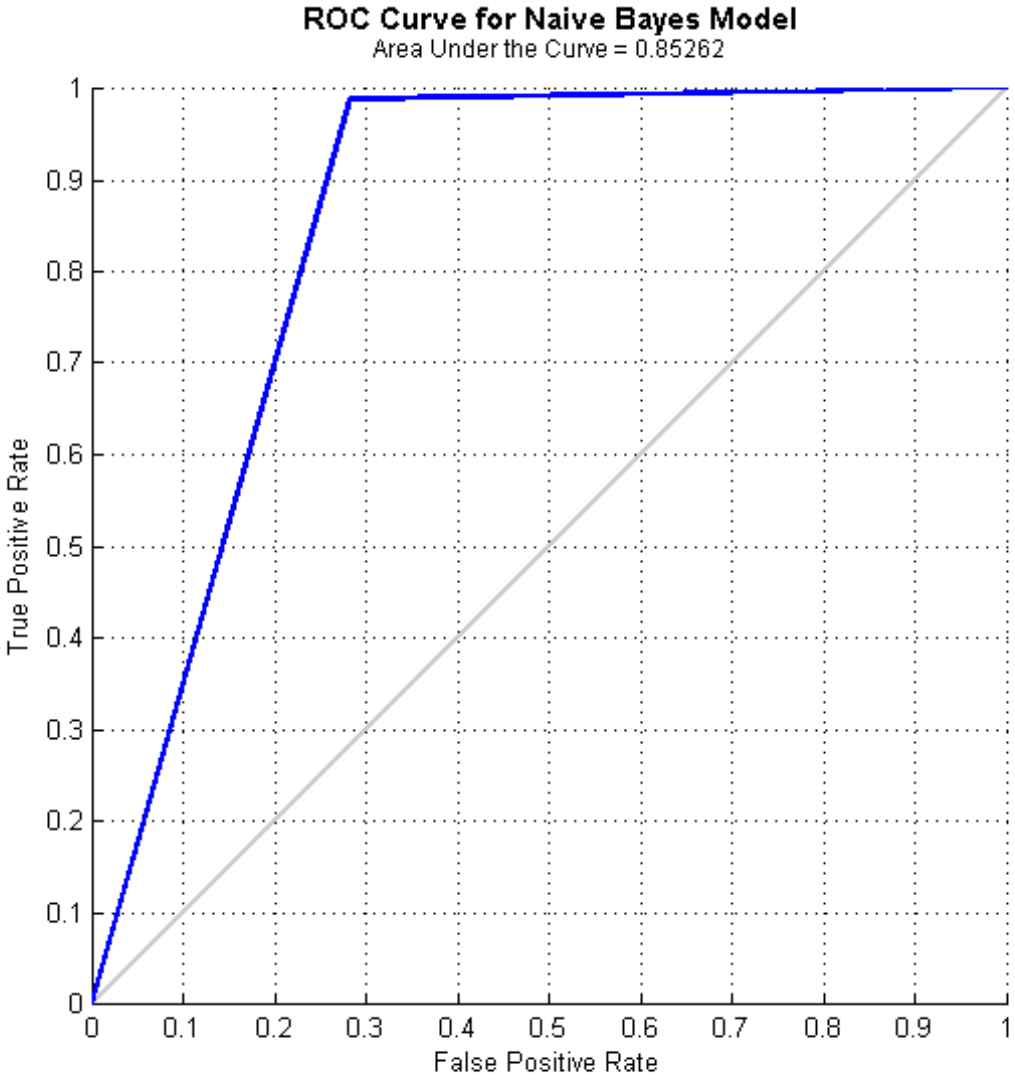

Figure 51 : Naive Bayes Classifier ROC Curve

4.3.5 Decision Trees

Decision Tree Classifier Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the Decision Tree Classifier model case, the testing error is less than the training error as shown in Figure 52. At the beginning, testing error showed high bias. However, adding more training data increased the gap between training errors and testing errors, which indicate high variance. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level. The Decision Tree Classifier model was suffering from high variance.

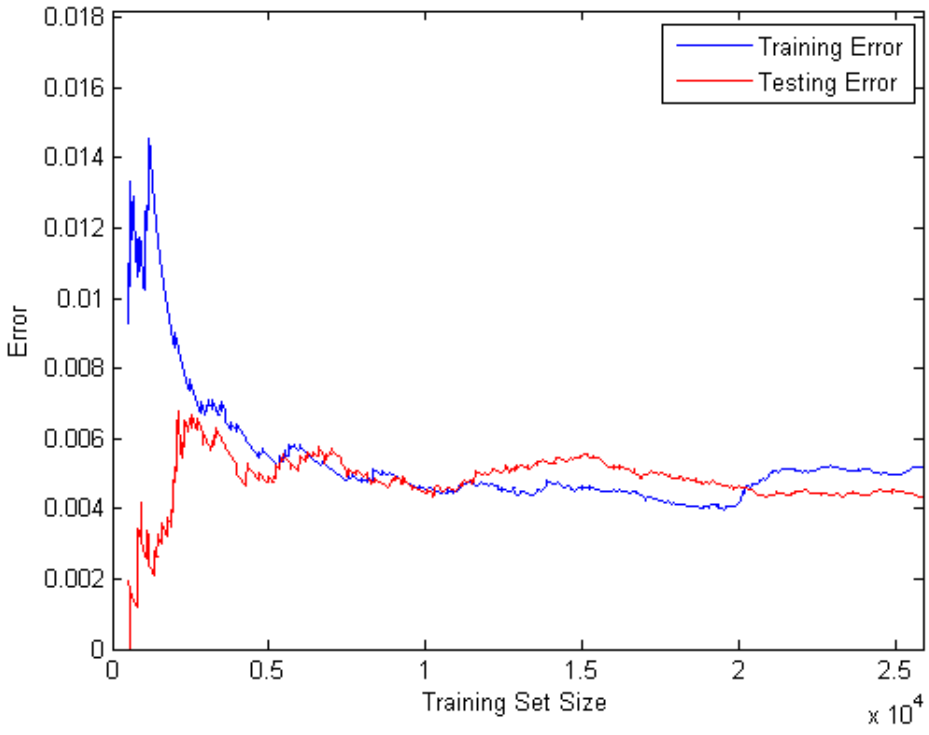

Figure 52 : Decision Tree Classifier Bias-Variance Learning Curve

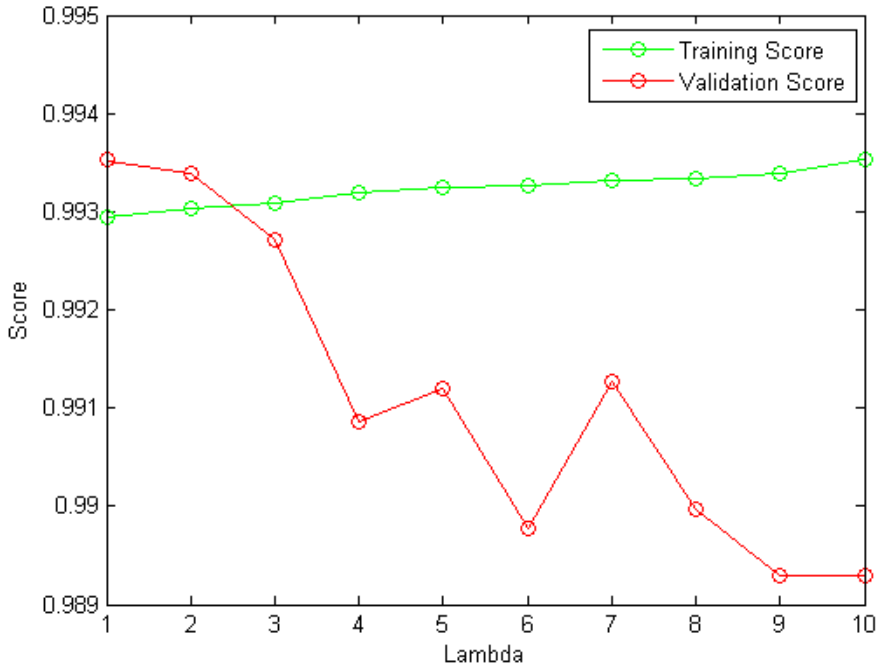

Figure 53 : Decision Tree Classifier Validation Curve

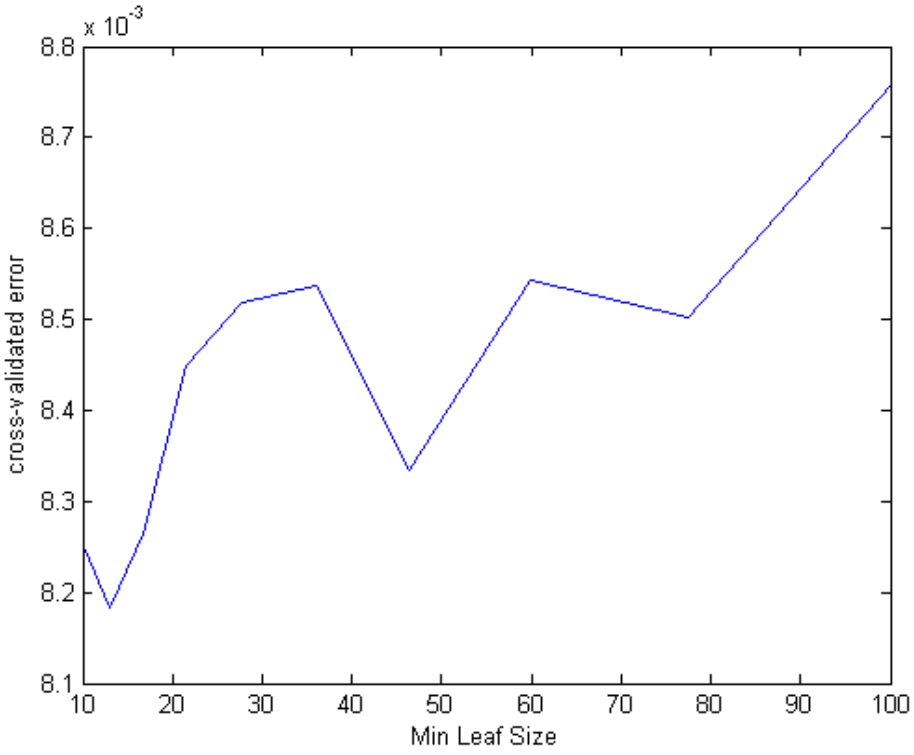

Figure 54: Decision Tree Classifier Validation Curve Leaf Size

On the other hand, the Decision Tree Classifier Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 53. In addition, the best leaf size is under 80 as shown in Figure 54. In this case, the performance of Decision Tree Classifier model is acceptable.

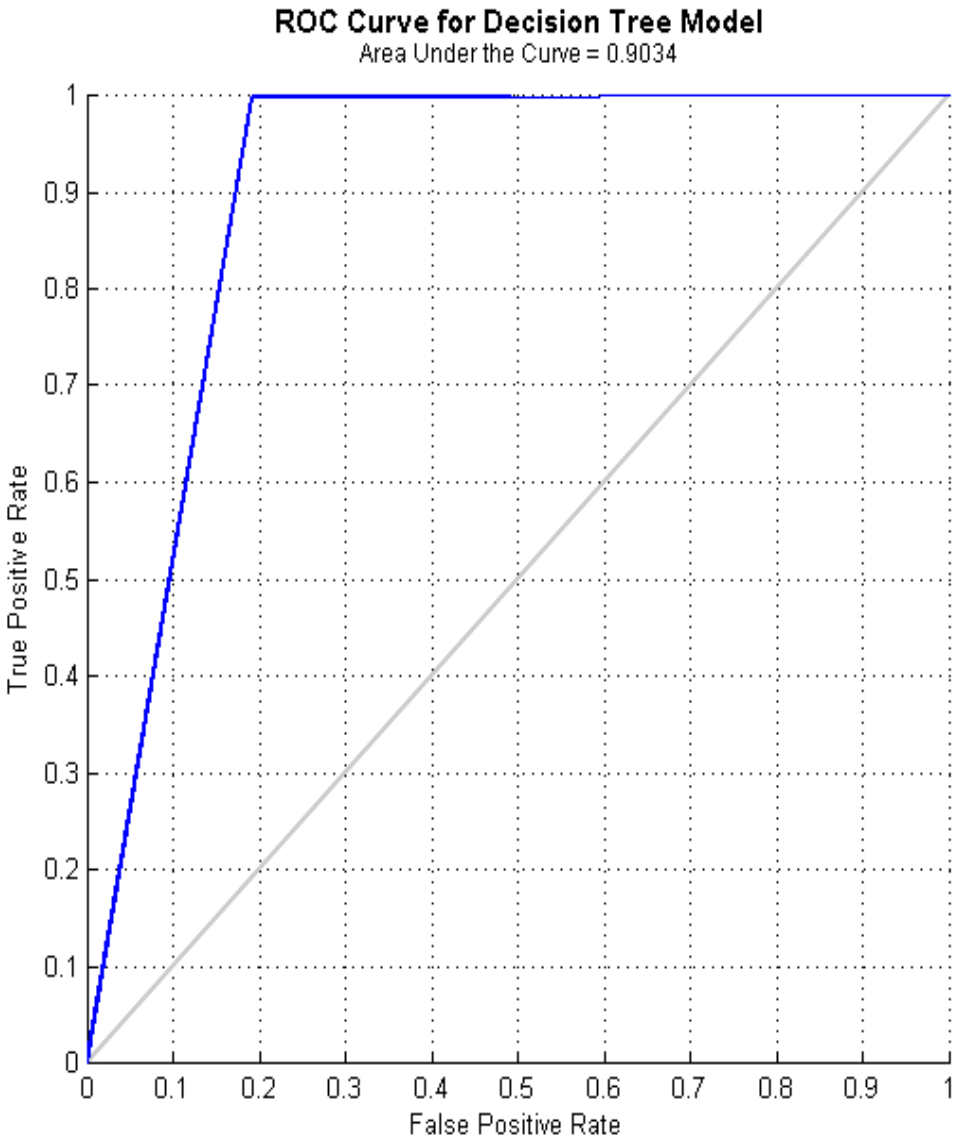

Figure 55 : Decision Tree Classifier ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.90 as shown in Figure 55. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 99.2%, while the misclassification rate is around 0.8%. In addition, the sensitivity (True Positive Rate) is around 99.8%; and the specificity (True Negative Rate) is around 78.6%.

4.3.6 Support Vector Machines

Support Vector Machines Bias-Variance Learning Curve was plotted to ensure whether the classifier suffers more from a variance error or a bias error. The goal of Bias-Variance Learning Curve is to show the advantage of adding more training data. In the Support Vector Machines model case, the testing error is higher than the training error as shown in Figure 56. At the beginning, testing error showed high bias. It was clear that adding more training data will not improve the model performance, as the testing errors seemed to stay above zero level.

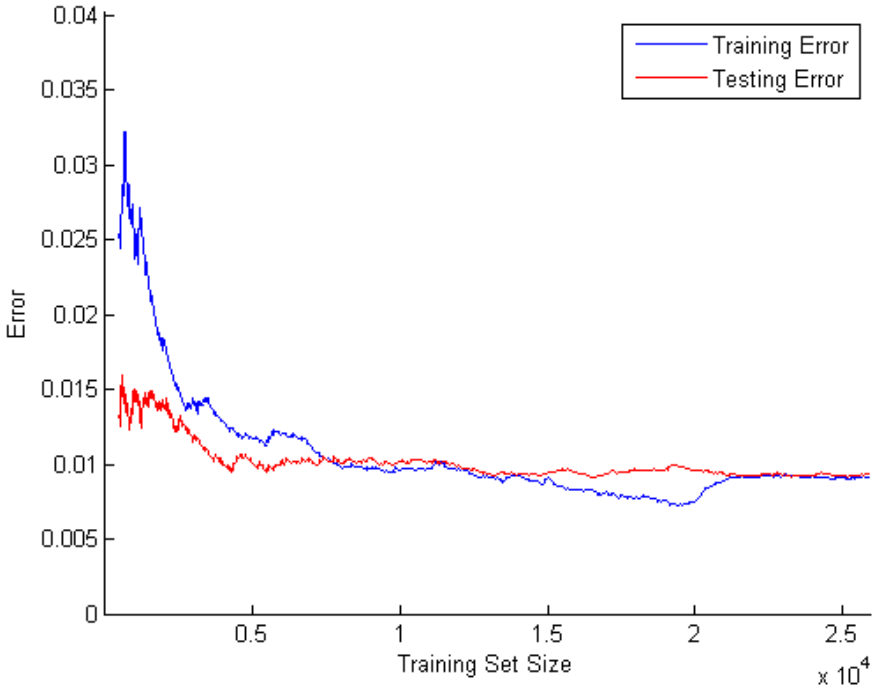

Figure 56 : Support Vector Machines Bias-Variance Learning Curve

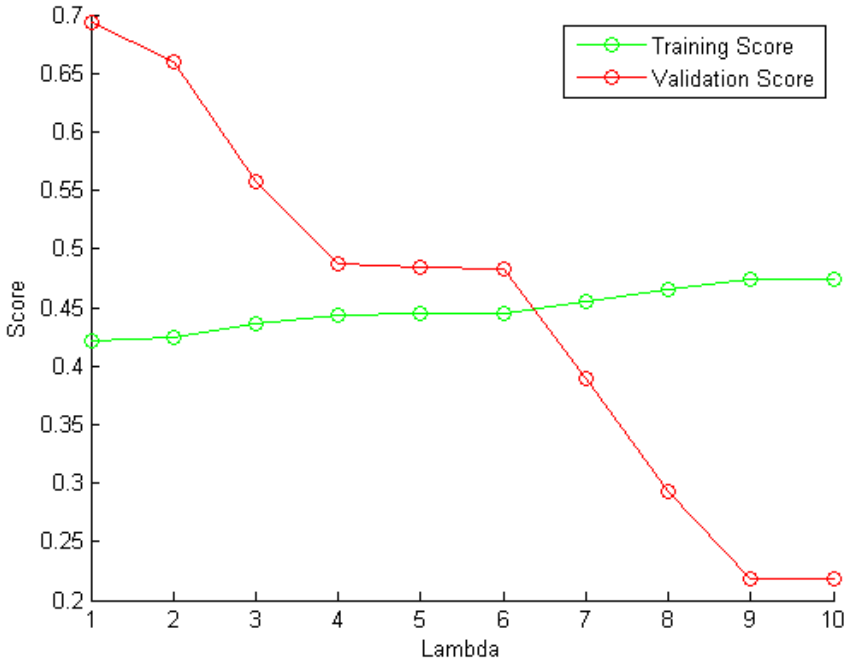

Figure 57 : Support Vector Machines Validation Curve

On the other hand, the Support Vector Machines Validation Curve shows that in the end of the training process the training score is high and the validation score is low as shown in Figure 57. In this case, the performance of Support Vector Machines model is acceptable.

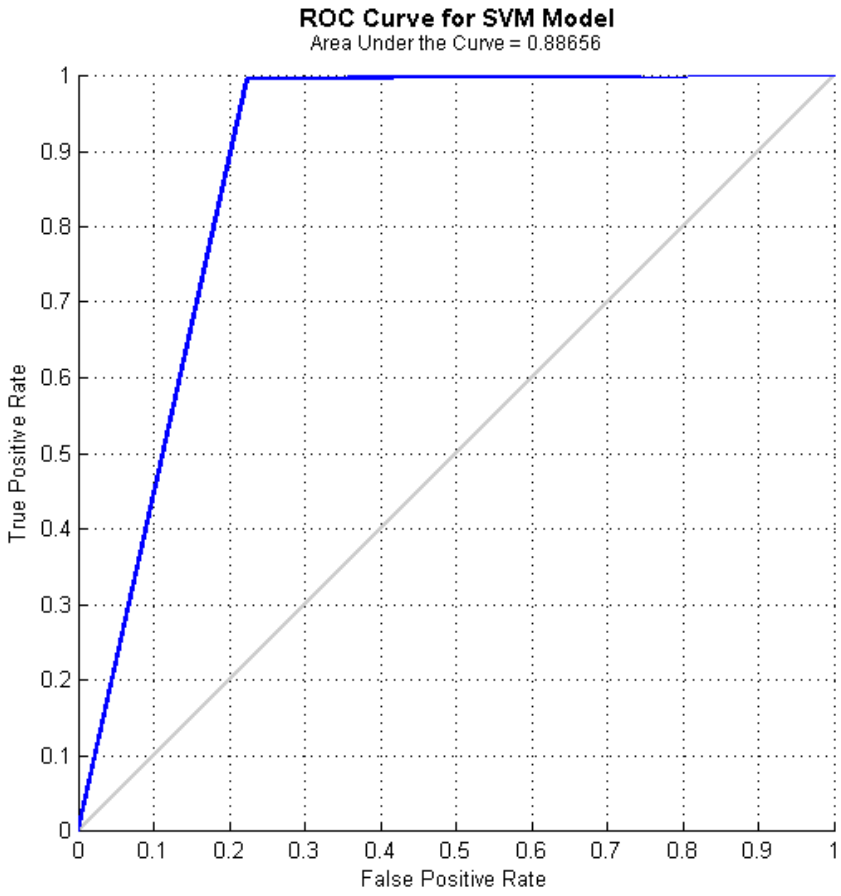

Figure 58 : Support Vector Machines ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) is around 0.89 as shown in Figure 58. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 99.0%, while the misclassification rate is around 1.0%. In addition, the sensitivity (True Positive Rate) is around 99.7%; and the specificity (True Negative Rate) is around 77.6%.

4.3.7 Artificial Neural Network

First, data was divided randomly to three sets (training, validation, and testing). Then, the network used scaled conjugate gradient back-propagation function to update weight and bias values during training phase. The Neural Network training, validation, and test performance is shown in Figure 59. The network stopped training when the validation error increased for six iterations, which occurred at iteration 142. The best validation performance occurred at iteration 136, which means there is no significant over-fitting had occurred by iteration 136 as shown in Figure 60.

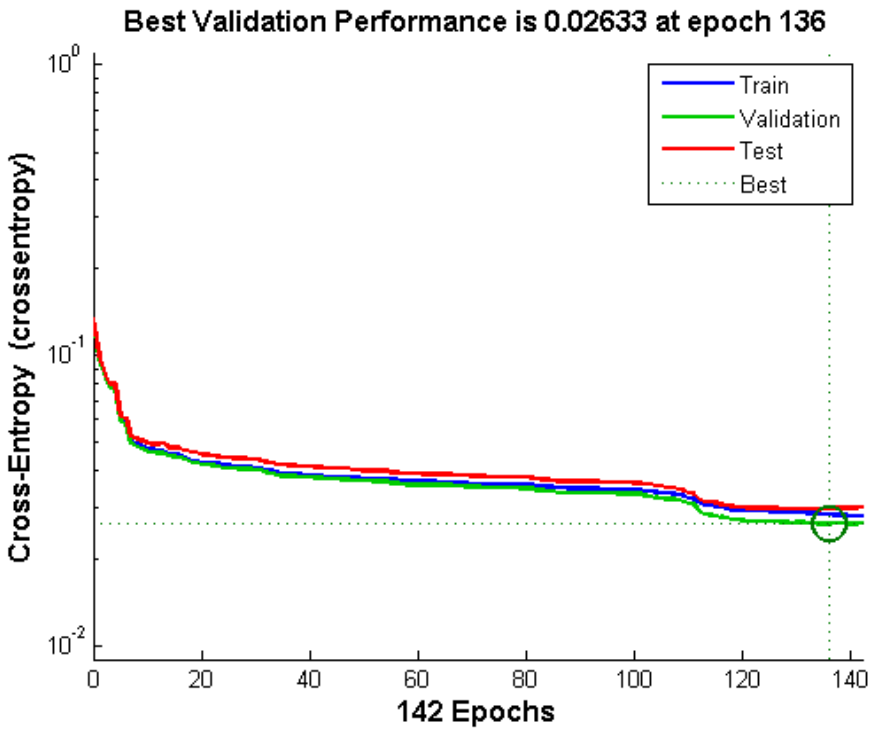

Figure 59 : Neural Network training, validation, and test performance

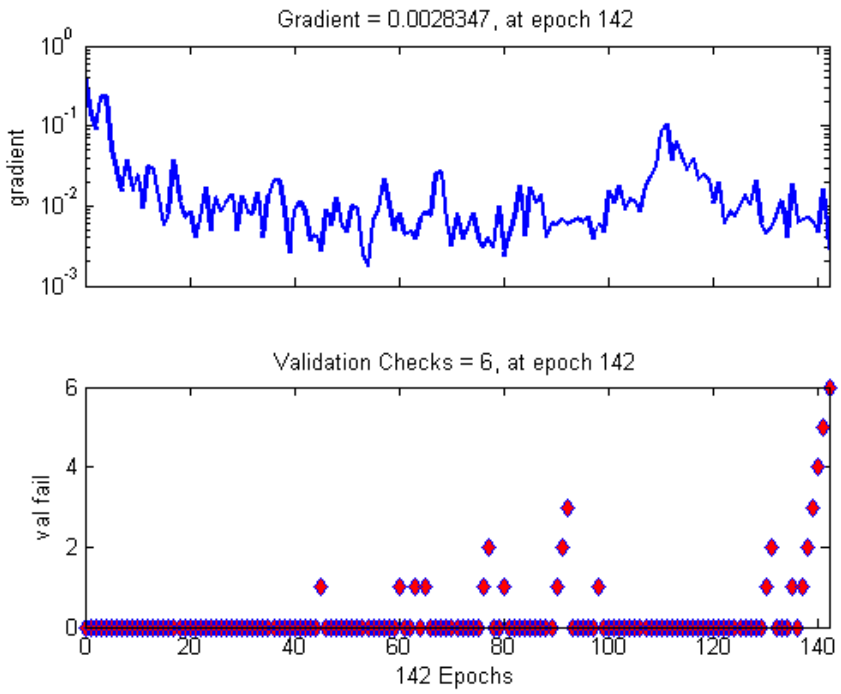

Figure 60 : Artificial Neural Network training state values

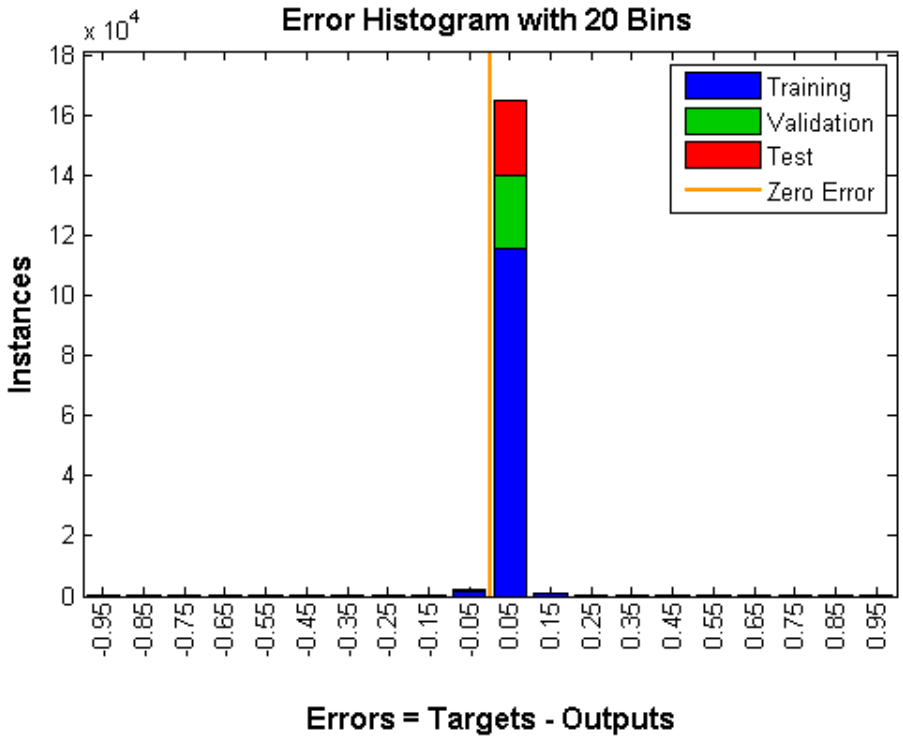

Figure 61 : Artificial Neural Network Error Histogram

The error histogram in Figure 61 show more verification of network performance. Most errors are fallen close to zero error.

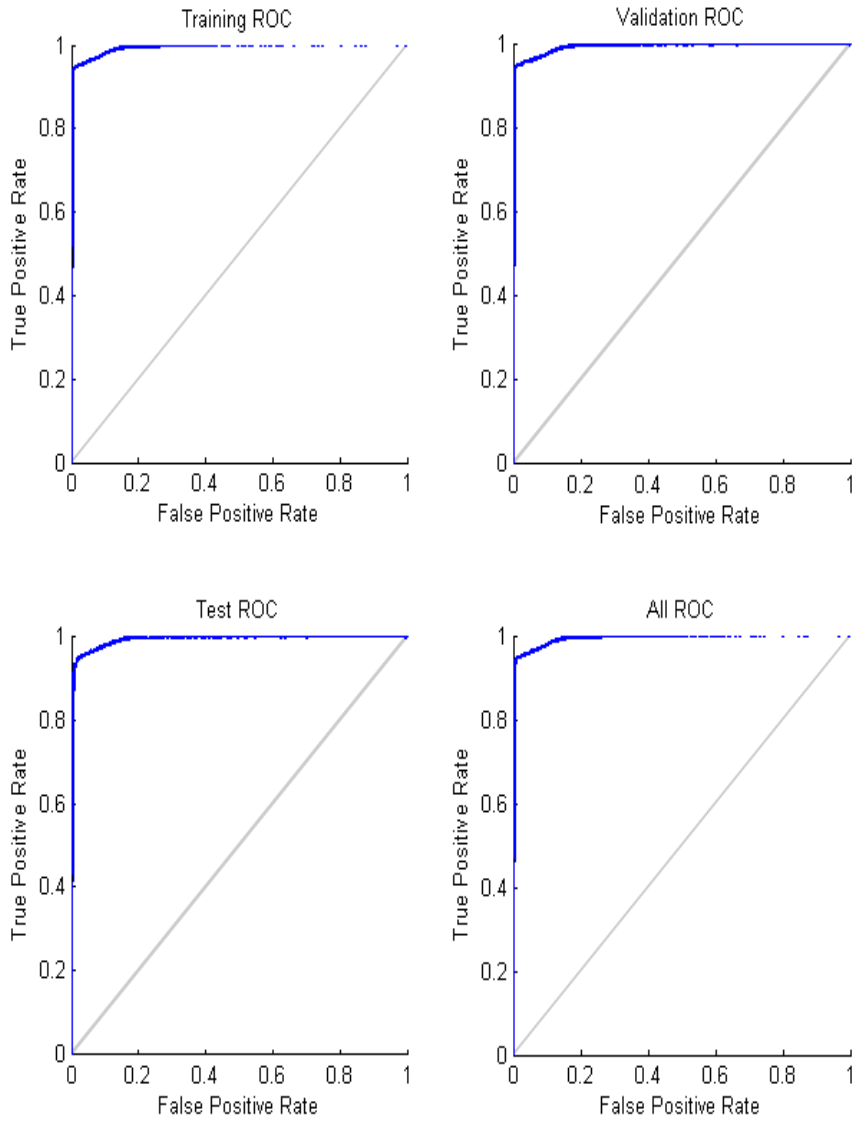

Figure 62 : Artificial Neural Network ROC Curve

Receiver Operating Characteristic (ROC) Analysis shows that the area under the curve (AUC) for all data is above 0.99 as shown in Figure 62. AUC value is acceptable since it is higher than random classifier (AUC=0.5).

In the Confusion Matrix analysis, the model accuracy is approximately 99.1%, while the misclassification rate is around 0.9%. In addition, the sensitivity (True Positive Rate) is around 99.7%; and the specificity (True Negative Rate) is around 76.4%.

For training data, the model accuracy is approximately 99.1%, while the misclassification rate is around 0.9%. In addition, the sensitivity (True Positive Rate) is around 99.7%; and the specificity (True Negative Rate) is around 76.1%.

For validation data, the model accuracy is approximately 99.1%, while the misclassification rate is around 0.9%. In addition, the sensitivity (True Positive Rate) is around 99.7%; and the specificity (True Negative Rate) is around 77.7%.

For test data, the model accuracy is approximately 99.1%, while the misclassification rate is around 0.9%. In addition, the sensitivity (True Positive Rate) is around 99.8%; and the specificity (True Negative Rate) is around 76.7%.

CHAPTER V

V. DISCUSSION

5.1 Analysis of Risk Factors for Prostate Cancer

Age is considered to be a critical factor for predicting prostate cancer. Prostate cancer is more common in the older male population and it becomes essential to undergo screening for its early detection and management as the age increases [24]. This study results show that age is significantly associated with prostate cancer. The odds of developing prostate cancer are 1.041 times greater for each year increase in age. The authors of [32] says that studies performed on prostate cancer in the middle-aged male population are small in number, because more emphasis is placed on findings involving men over age 60 years of age. In this study, men who are 65 year old and older are the majority of prostate cancer patients; therefore, having an age of 65 and older is a highly significant predictor for prostate cancer. This study's results show that an age of 65 years or older is positively associated with prostate cancer. Men between the ages of 65 and 74 have an increased risk of developing prostate cancer by as much as 1.287 times greater than younger aged men. On the other hand, an age under 65 years is negatively associated with prostate cancer, which [27] does not advise routine screening at an early age. The odds of men between the ages of 45 and 54 years show a decreased the risk of developing prostate cancer by 0.421compared with older aged men.

Regarding cancer and ethnicity in general, the death rate from cancer for African Americans is 34% higher than that of Caucasian Americans [23]. Caucasian American men

were recorded with the second highest mortality rate, followed by Native American, Hispanic, and then Asian/Pacific Islander ethnicities [6,7]. In regard to new incidence and according to the Center for Disease Control and Prevention (CDC), since 1975, African Americans have had and continue to have the highest incidence of prostate cancer - with Caucasian American men being notable for the second highest incidence, followed by Latinos, Asian/Pacific Islander and Native American men [33]. This study's results show that race is significantly associated with prostate cancer. Men with African-American ancestry are at high risk of prostate cancer. An African-American man has an increased risk of developing prostate cancer by as much as 1.650 times greater than other ethnicities. On the other hand, Asian men have a low risk of developing prostate cancer by 0.720 compared men of other ethnic backgrounds. For Caucasian American men, the odds show an increased risk of developing prostate cancer by 1.134 greater than other ethnicities, which rank Caucasians as second in prostate cancer incidence African American men.

One of the biggest risk factors for prostate cancer is having a family history of cancer, with an incidence of about 10% to 20% among men who develop prostate cancer because of a positive family history of this disease[53]. The most major clinical aspect of prostate cancer perhaps in men with a positive family history is its relatively early development, and these men characteristically present with this disease and are diagnosed 6 - 7 years earlier than men with no such family history of this cancer [53 54]. This study's findings show that family history of prostate cancer is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 7.752. This indicates that the odds of men with positive

family history of prostate cancer have increased the risk of developing prostate cancer by 7.752 greater than men with no family history of prostate cancer.

In addition, men with positive family history of cancer other than prostate cancer also have increased the risk of developing prostate cancer by 1.717 greater than men with no family history of cancer.

In the United States, between 14 and 20 percent of deaths associated with cancer are complicated by obesity [94]. Adding to this phenomenon, it has also been found that obesity also increases the risks associated with the several types of cancer and aggressive types of prostate cancer [23]. On the other hand, some different research studies showed no relationship between body mass index and risks of prostate cancer [118,119]. A recent group study from the United States showed an inverse relationship between obesity and a diagnosis of prostate cancer, except in males who were less than sixty years old or those who had a positive family history of prostate cancer [120]. This study's results show that obesity is significantly associated with prostate cancer. However, the statistical analysis indicates that an obese man has a low risk of developing prostate cancer by 0.808 compering to non-obese.

Physical activity is proven to be beneficial at any age and it helps individuals to attain a positive health condition. However, the results concerning physical activity and the danger for prostate cancer are indecisive. This study's results show that physical activity is not significantly associated with prostate cancer for all three types of physical activity, vigorous (p = 0.3255), light-moderate (p =0.5866), strengthening activity (p =0.9724).

The link between alcohol and a variety of types of cancer is comparatively modest at reduced levels of consumption, there have been various observational studies, showing most of the participants using low to moderate quantity of alcohol, and thus, an important positive link, a nonexistence of a considerable association, or a noteworthy negative link between alcohol consumption and the danger of death and disease were found from certain type of cancers [1]. The precise nature of the connection between drinking alcohol and prostate cancer risk has been vague regardless of the big number of research studies undertaken as reviewed in the literature [138-143]. Despite the fact that there seems to be little evidence for an association as a whole, a minor risk increase with high levels of alcohol drinking has been found in [144] but this is not always found to be true as has been found in [138]. This study's results found that alcohol abuse is significantly associated with prostate cancer. However, the statistical odds indicate that alcohol abuse has a low incidence of risk at promoting the development of prostate cancer by 0.657.

The relationship between smoking and developing prostate cancer remains unclear. The result of this research found that smoking is significantly associated with prostate cancer. However, the statistical odds indicate that a male smoker has a low risk of developing prostate cancer by 0.747 as compared to non-smoker. This result contradicts some similar finding in [164] a study, which found smoking brought a reduction in the risks of cancers.

The consumption of fat is related with prostate cancer [184,185], however, the part played by fats of diet and other specified types is not clear [184,186]. This study found that fatty acid deficiency is not significantly associated with prostate cancer (p = 0.3178). However, hypercholesterolemia is significantly associated with the prostate cancer. The statistical odds indicate that the likelihood of a hypercholesterolemia patient has increased risk of

developing prostate cancer by 1.243 greater than men who have not been diagnosed with hypercholesterolemia.

There are evidences presented which indicated that Vitamin D indeed plays a part in the development of the prostate cancer. Genetic variations in the receptor are related with the Gleason score of 139 as well as the variation causing agents in the pathway of Vitamin D are related with risks leading to reappearance of the death rate associated with prostate cancer [221]. This study found that vitamin D deficiency is significantly associated with the prostate cancer. Men who are diagnosed with vitamin D deficiency have increased the risk of developing prostate cancer by 1.079 greater than men who have not diagnosed with vitamin D deficiency.

Inflammation of prostate has been considered as a risk factor of prostate cancer [231]. In fact, inflammatory infiltrates are commonly found in prostate biopsies [232]. Researchers have found that prostate inflammation, mostly the chronic variety, is associated with an increased likelihood of prostate cancer [233]. This study found that inflammation of the prostate is significantly associated with the prostate cancer. Men who have inflammation of prostate have an increased risk of developing prostate cancer by 1.351.

According to [235] there are studies did not find a clear association between vasectomy and prostate cancer development. However, this study found that vasectomy is significantly associated with the prostate cancer ($p < 0.0001$) with odds of 2.906. This indicates that the odds of men who has vasectomy have increased the risk of developing prostate cancer by 2.906 greater than men who does not have vasectomy.

Early studies do not have a sufficient amount of data to confirm the association between high blood pressure and prostate cancer. Some of these studies found that there are no a significant association between prostate cancer and high blood pressure [238,239]. However, this study found that hypertension is significantly associated with the prostate cancer. Men who have hypertension have increased risk of developing prostate cancer by 1.184.

5.2 Prostate Cancer Classification Models for NIS Data

Since the most of classification methods require a large amount of training data, the NIS dataset is large enough to satisfy the classifiers demand. All models used prostate cancer as a dependent variable with value of '0' for negative prostate cancer diagnosis and the value of '1' for positive prostate cancer diagnosis. The independent variables are age, race, family history of prostate cancer, family history of any other cancer, obesity, alcohol abuse, smoking, fat intake, vitamin D deficiency, inflammation of prostate, vasectomy, and hypertension.

5.2.1 Confusion matrix analysis

From the confusion matrix, the top model accuracy rate is Neural Network classifier with 79.1%. Decision Tree classifier model is the second position with 78.5%. Third classifier model is logistic regression with accuracy rate of 75.2%. k-Nearest Neighbors is in the fourth with accuracy rate of 72.8%. The Support Vector Machines and the Naive Bayes Classifier have accuracy rate of 72.2% and 71.2%, respectively.

The sensitivity rate (True Positive Rate) for the Neural Network classifier is 71.6%, which ranks it along with logistic regression in the fourth position. The k-Nearest Neighbors has the highest sensitivity rate with 75.6%, followed by the Support Vector

Machines (73.2%) and the Decision Tree classifier (72.7%). Finally, the Naive Bayes Classifier has sensitivity rate of 66.4%.

The specificity (True Negative Rate) for the Neural Network classifier is 85.2%, which ranks it at the top. The Decision Tree classifier model is in the second position with a specificity rate of 83.2%, followed by the Logistic Regression (78.1%), the Naive Bayes Classifier (75.1%), the Support Vector Machines (71.4%) and the k-Nearest Neighbors (70.4%).

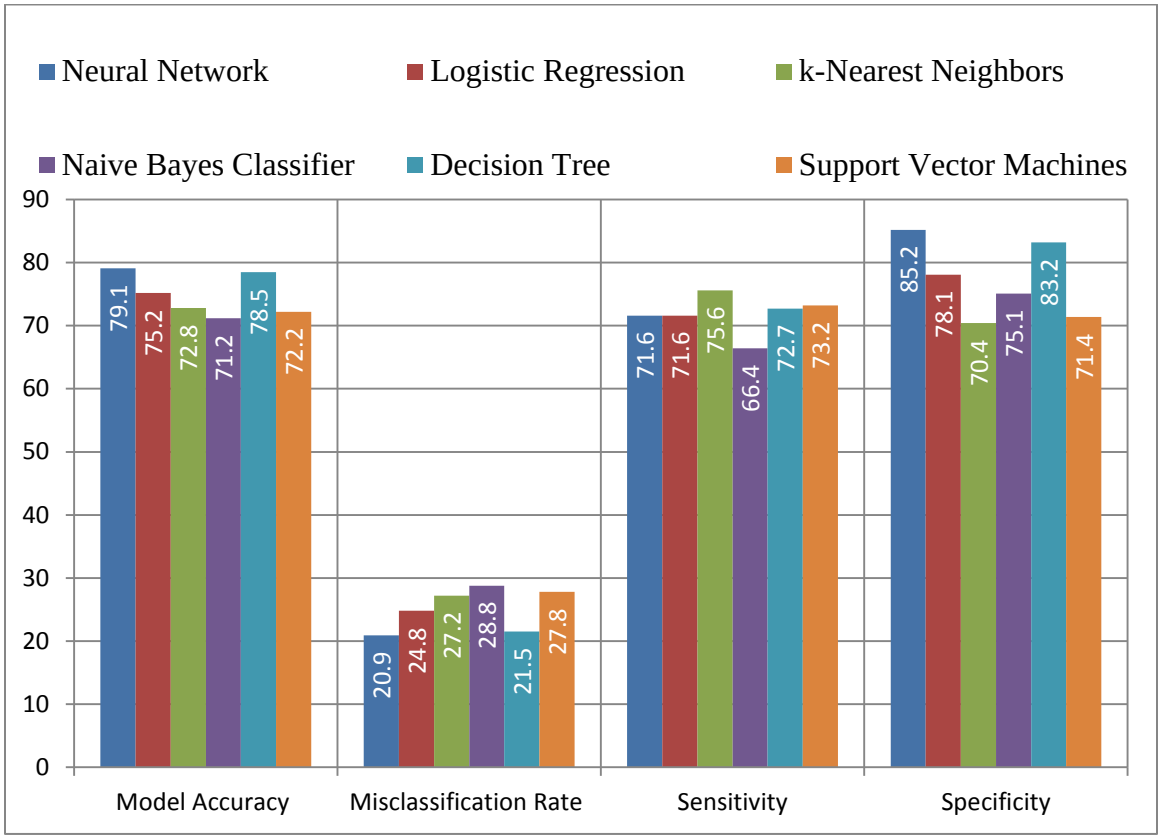

Figure 63 : Confusion Matrix Results for All Models (NIS Data)

5.2.2 Receiver Operating Characteristic (ROC) Analysis

Receiver Operating Characteristic (ROC) Analysis is the main measurement for this study. A ROC analysis is used to determine the accuracy of all models. A ROC curve is used to define the relationship between the sensitivity and specificity of the classifier model. The Area Under the Curve (AUC) represents the model accuracy. The larger value of AUC is the better model accuracy.

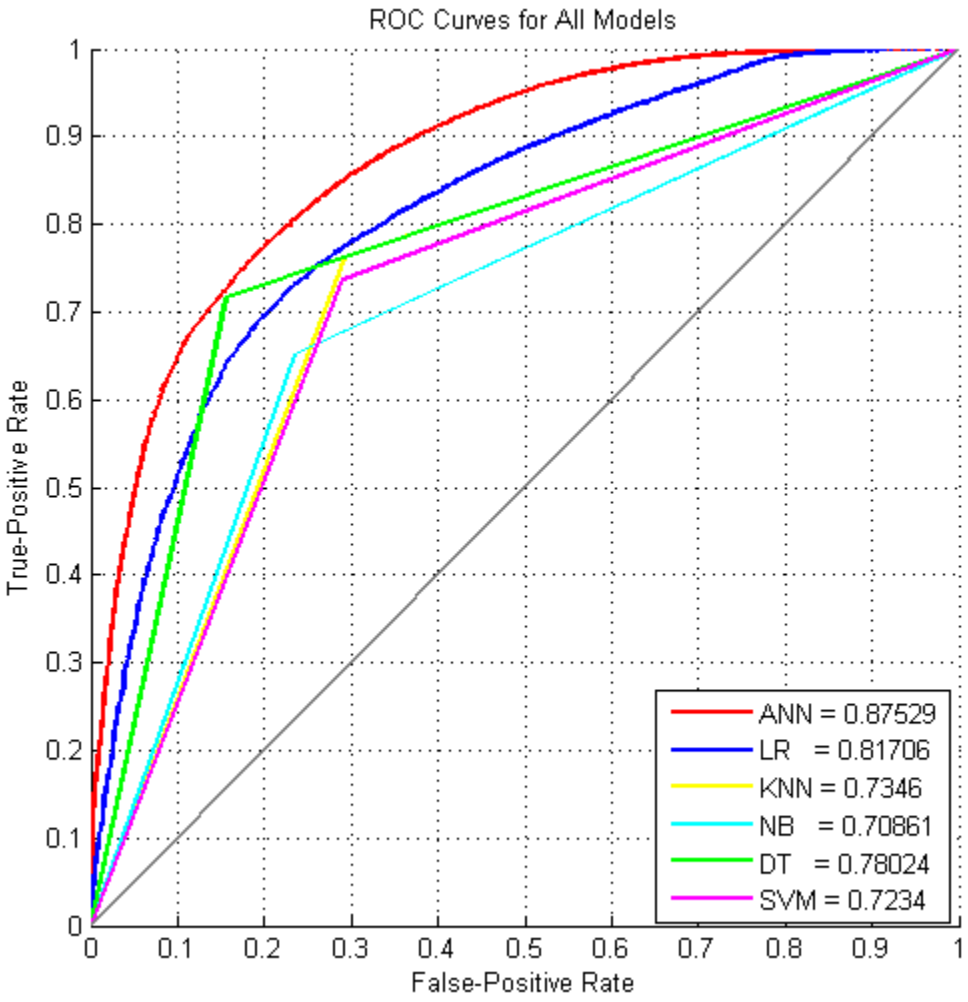

Figure 64 : ROC Curves for All Models (NIS Data)

As shown in Figure 64, the Neural Network classifier is the highest model accuracy with the AUC of 0.875. The Logistic Regression is the second with an AUC rate of 0.817, followed by the Decision Tree classifier (AUC= 0.78), the k-Nearest Neighbors (AUC= 0.735), the Support Vector Machines (AUC= 0.723) and the Naive Bayes Classifier (AUC= 0.709).

5.3 Prostate Cancer Classification Models for SEER Data

All models used prostate cancer as a dependent variable with value of '0' for negative prostate cancer diagnosis and the value of '1' for positive prostate cancer diagnosis. The independent variables are age, race, Prostatic Specific Antigen (PSA) lab value, PSA interpretation, Gleason's score, and tumor size.

5.3.1 Confusion matrix analysis

From the confusion matrix, the top model accuracy rate is the Decision Tree classifier model with 99.2%. The Neural Network classifier is in the second position with 99.1%. Third classifier model is the Support Vector Machines with accuracy rate of 99.0%. The k-Nearest Neighbors is in the fourth with accuracy rate of 98.6%. The Logistic Regression and the Naive Bayes Classifier have accuracy rate of 98.5% and 98.0%, respectively.

The sensitivity rate (True Positive Rate) for the Neural Network classifier is 99.7%, which ranks it along with the Support Vector Machines at second position. The Decision Tree classifier has the highest sensitivity rate with 99.8%. The Logistic Regression has

sensitivity rate of 99.4%; followed by the k-Nearest Neighbors (99.2%). Finally, the Naive Bayes Classifier has sensitivity rate of 98.8%.

The specificity (True Negative Rate) for the Neural Network classifier is 76.4%, which ranks it in the fourth position. The Decision Tree classifier model has the highest specificity rate of 78.6%, followed by the Support Vector Machines (77.6%), and the k-Nearest Neighbors (77.4%). Finally, the Naive Bayes Classifier is in the fifth position with sensitivity rate of 71.7%, followed by the Logistic Regression (67.3%).

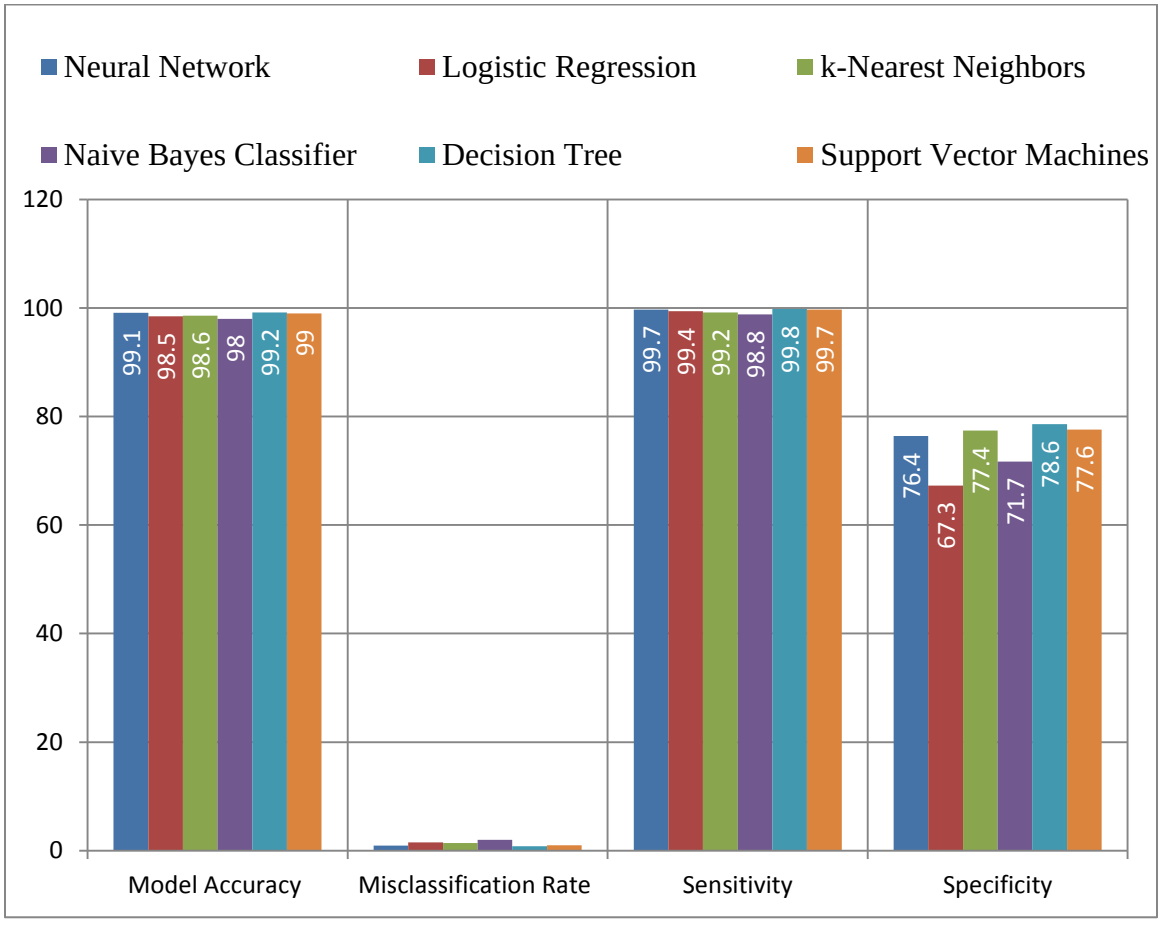

Figure 65 : Confusion Matrix Results for All Models (SEER Data)

5.3.2 Receiver Operating Characteristic (ROC) Analysis

Receiver Operating Characteristic (ROC) Analysis is the main measurement for this study. ROC analysis is used to determine the accuracy of all models. ROC curve is used to define the relationship between the sensitivity and specificity of the classifier model. The Area Under the Curve (AUC) represents the model accuracy. The larger value of AUC is the better model accuracy.

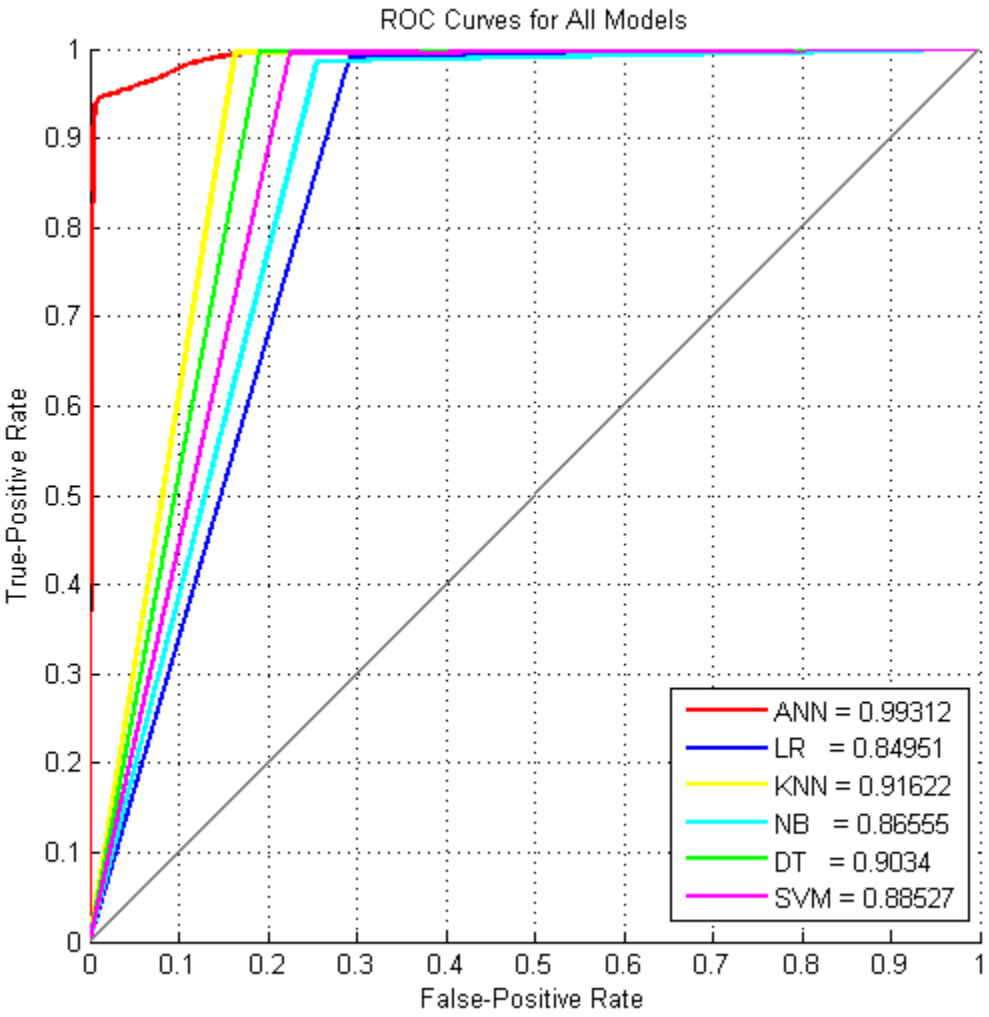

Figure 66 : ROC Curves for All Models (SEER Data)

As shown in Figure 66, the Neural Network classifier is the highest model accuracy with AUC of 0.993. the k-Nearest Neighbors is in the second with AUC rate of 0.916, followed by the Decision Tree classifier (AUC= 0.903), the Support Vector Machines (AUC= 0.885), the Naive Bayes Classifier (AUC= 0.866), and the Logistic Regression (AUC= 0.85).

Web Application

Both Neural Network models for NIS data and SEER data were combined together in web application to be used and tested.

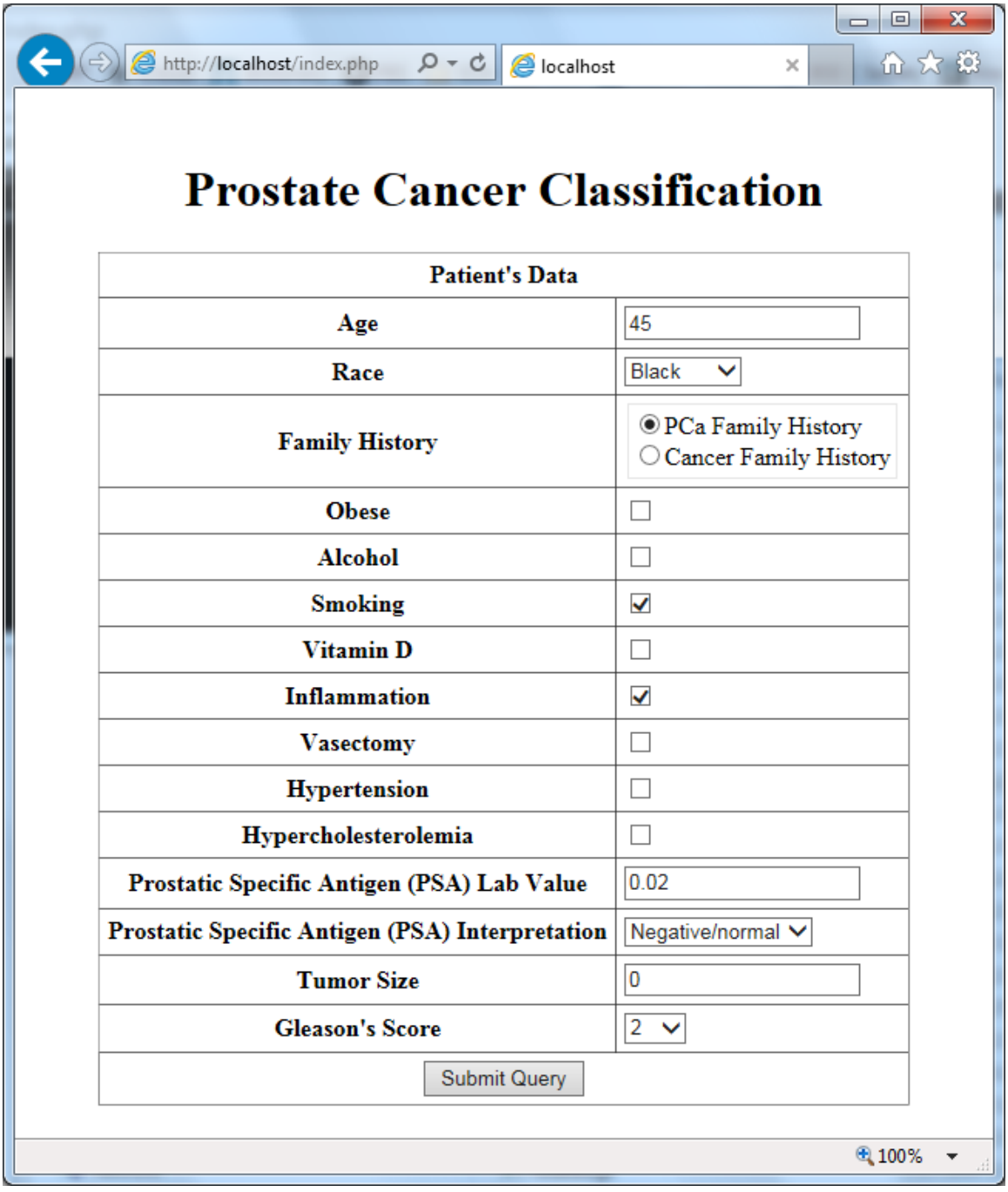

Figure 67 : Neural Network Models Web Interface

The result was presented as probability chart.

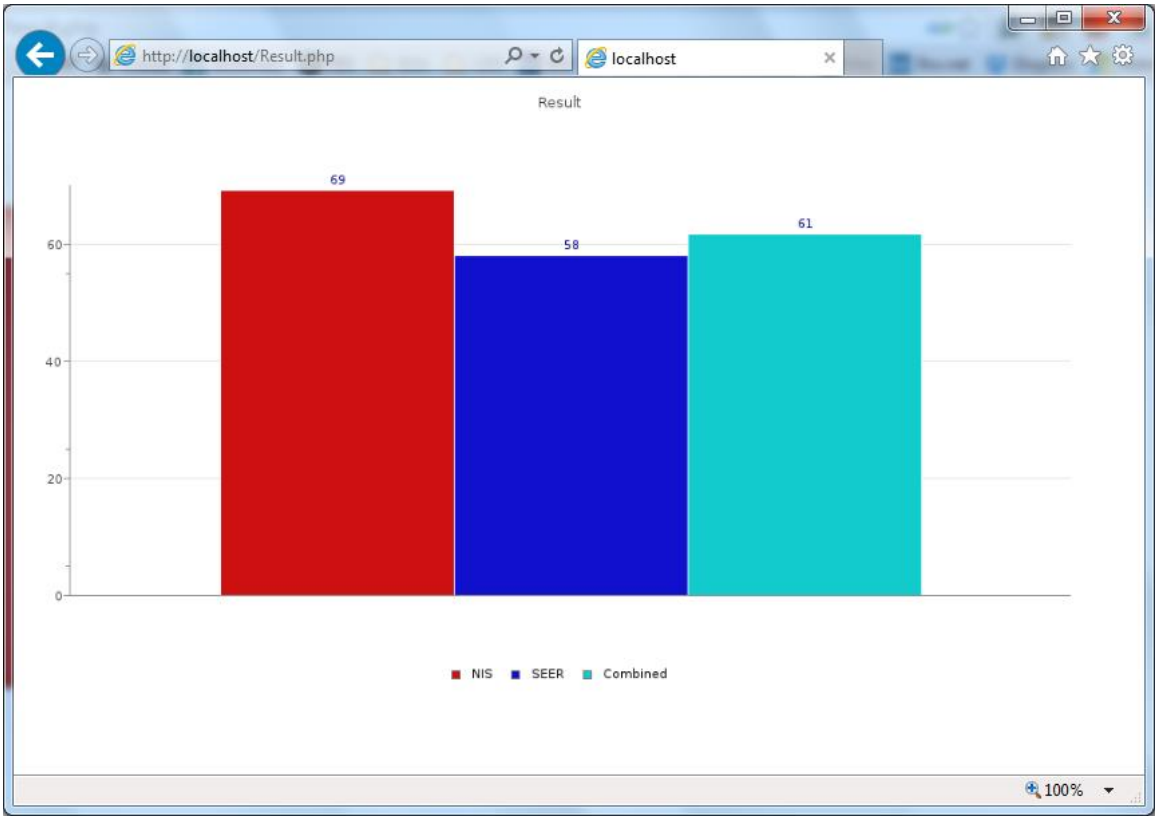

Figure 68 : Neural Network Models Web Output

CHAPTER VI

VI. SUMMARY AND CONCLUSIONS

6.1 Summary and Conclusions

Prostate cancer is a cancer that typically affects men in the older population. Age is considered to be a critical factor for predicting prostate cancer. The primary conclusion being age is significantly associated with overall prostate cancer incidence. Moreover, men who are 65 years old and older are the majority of prostate cancer patients; therefore, an age of 65 and older is a highly significant predictor for prostate cancer.

Ethnicity is also significantly associated with prostate cancer. Men with African American ancestry are at high risk of prostate cancer. From the primarily result, an African American man has increased risk of developing prostate cancer by 1.650 greater than other races. On the other hand, Asian men have a low risk of developing prostate cancer by 0.720 compared to other race.

A family history of cancer is the most critical variable in predicting prostate cancer. Men with positive family history of prostate cancer have increased risk of developing prostate cancer by 7.752 greater than men with no family history of prostate cancer. In addition, men with positive family history of cancer other than prostate cancer have also increased the risk of developing prostate cancer by 1.717 greater than men with no family history of other cancer.

Obesity, alcohol abuse, and smoking are significantly associated with prostate cancer, but they are not good enough to predict prostate cancer. Physical activity and fatty acid

deficiency are not significantly associated with prostate cancer. However, hypercholesterolemia is significantly associated with the prostate cancer; and it can be a risk factor to predict prostate cancer. Moreover, vitamin D deficiency is significantly associated with the prostate cancer. Men who are diagnosed with vitamin D deficiency have increased risk of developing prostate cancer by 1.079 times greater than men who have not diagnosed with vitamin D deficiency.

Vasectomy is significantly associated with the prostate cancer. Men who had vasectomy procedure have increased risk of developing prostate cancer by 2.906 greater than men who do not have vasectomy. For the chronic diseases, inflammation of prostate and hypertension are significantly associated with the prostate cancer. Men who have inflammation of prostate or hypertension have an increased risk of developing prostate cancer by 1.351 and 1.184, respectively.

The Artificial Neural Network (ANN) is very sophisticated classification method which can deal with more advance problems effectively. This study utilizes ANN to classify the diagnosis of prostate cancer; and when the result is compared with other classification methods, including, logistic regression, k-Nearest Neighbors, Naïve Bayes classifier, Decision Tree classifier, and Support Vector Machine. The Nationwide Inpatient Sample (NIS) data and the Surveillance, Epidemiology, and End Results (SEER) data were used to train and test all the classification models. ANN performed very well in both data with accuracy rate of 0.875 on NIS data and 0.993 on SEER data. These promising results can lead to more comprehensive classification models for prostate cancer by combining both Neural Networks results.

In conclusion, among different types of cancers, which occur in men, prostate cancer is the most commonly occurring one. Although, Prostate Specific Antigen (PSA) is most used screening test for prostate cancer, PSA test result sometime can be false which leads to over-diagnosis. This study tried to minimize the false result by providing a prediction tool based on Artificial Neural Network to predict and to support the clinical decision in prostate cancer screening.

6.2 Future Research

The Nationwide Inpatient Sample (NIS) data and the Surveillance, Epidemiology, and End Results (SEER) data are fully rich data with a large number of variables. Exploring both data and exam more variables will definitely increase the accuracy of Artificial Neural Network models which lead to more accurate prostate cancer screening.

Medical imaging is playing a major role in all cancer screening. A combination of medical imaging and artificial intelligence methods such Artificial Neural Network can be used in clinical practice to increase the successful diagnosis and eliminate the medical errors. Many studies have made important contributions to this field, but the demand for more studies and innovation are still high.

References

1. Cancer IAR. *World Cancer Report 2014.* World Health Organization; 2014.
2. Institute NC. Prostate Cancer. 2014; http://www.cancer.gov/cancertopics/types/prostate. Accessed April 2nd, 2014.
3. Group USCSW. United States Cancer Statistics: 1999–2010 Incidence and Mortality Web-based Report. 2013; www.cdc.gov/uscs. Accessed April 2nd, 2014.
4. Institute NC. SEER Cancer Statistics Factsheets: Prostate Cancer. 2014; http://seer.cancer.gov/statfacts/html/prost.html. Accessed April 2nd, 2014.
5. Carter HB. - Prostate Cancer Y1 - 2013 Y2 - 01. - *Prostate Disorders.* 2013:- 25.
6. Institute NC. *Prostate Cancer.* Vol 2014. National Cancer Institute at National Institutes of Health: U.S. Department of Health and Human Services; 2014.
7. Institute NC. *SEER Cancer Statistics Factsheets: Prostate Cancer.* Vol 2014. Bethesda, MD: National Cancer Institute; 2014.
8. Society AC. Cancer Facts & Figures 2014. 2014; http://www.cancer.org/cancer/prostatecancer/index. Accessed April 2nd, 2014.
9. Society AC. *Cancer Facts & Figures 2014.* Vol 2014. Atlanta, GA: American Cancer Society; 2014.
10. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J. Clin.* 2010;60(2):70-98.
11. Greene KL, Albertsen PC, Babaian RJ, et al. Prostate specific antigen best practice statement: 2009 update. *J. Urol.* 2013;189(1 SUPPL):S2-S11.
12. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *N. Engl. J. Med.* 2009;360(13):1310-1319.
13. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N. Engl. J. Med.* 2009;360(13):1320-1328.
14. Mark W. McClure M. **Chapter 77 – Prostate Cancer**. In: Inc E, ed. *Integrative Medicine*. 3rd ed. Elsevier Inc: Elsevier Inc.; 2012:10.
15. Catalona WJ, Partin AW, Walsh PC, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: A prospective multicenter clinical trial. *J. Am. Med. Assoc.* 1998;279(19):1542-1547.
16. Fang J, Metter EJ, Landis P, Carter HB. PSA velocity for assessing prostate cancer risk in men with PSA levels between 2.0 and 4.0 ng/ml. *Urology.* 2002;59(6):889-893.
17. Roobol MJ, Kranse R, Bangma CH, et al. Screening for Prostate Cancer: Results of the Rotterdam Section of the European Randomized Study of Screening for Prostate Cancer. *Eur. Urol.* 2013;64(4):530-539.
18. Bokhorst LP, Bangma CH, van Leenders GJLH, et al. Prostate-specific Antigen–Based Prostate Cancer Screening: Reduction of Prostate Cancer Mortality After Correction for Nonattendance and Contamination in the Rotterdam Section of the European Randomized Study of Screening for Prostate Cancer. *Eur. Urol.* 2014;65(2):329-336.
19. Schröder FH. Landmarks in prostate cancer screening. *BJU Int.* 2012;110:3-7.
20. Thompson IM, Ankerst DP, Chi C, et al. Assessing Prostate Cancer Risk: Results from the Prostate Cancer Prevention Trial. *JNCI: Journal of the National Cancer Institute.* 2006;98(8):529.

21. Institute NC. *Cancer Trends Progress Report – 2011/2012 Update.* Bethesda, MD: National Cancer Institute; August 2012.
22. Mackie A. *Screening for Prostate Cancer. Review against programme appraisal criteria for the UK National Screening Committee (UK NSC).* UK National Screening Committee; June 2010.
23. Society AC. *Cancer Facts & Figures 2013.* Vol 2013. Atlanta, GA: American Cancer Society; 2013.
24. Banthia R, Malcarne VL, Varni JW, Ko CM, Sadler GR, Greenbergs HL. The Effects of Dyadic Strength and Coping Styles on Psychological Distress in Couples Faced with Prostate Cancer. *J. Behav. Med.* 2003;26(1):31.
25. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *JNCI: Journal of the National Cancer Institute.* 2009;101(19):1325-1329.
26. Li J, German R, King J, et al. Recent trends in prostate cancer testing and incidence among men under age of 50. *Cancer Epidemiol.* 2012;36(2):122.
27. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. *The Journal of urology.* 2013;190(2):419.
28. Brawley OW. Prostate cancer epidemiology in the United States. *World J. Urol.* 2012;30(2):195-200.
29. Bellenir K. Prostate and urological disorders sourcebook. 2006.
30. Chiu BCH, Anderson JR, Corbin D. Predictors of prostate cancer screening among health fair participants. *Public Health (Elsevier).* 2005;119(8):686-693.
31. Russo AL, Chen M-H, Aizer AA, Hattangadi JA, D'Amico AV. Advancing age within established Gleason score categories and the risk of prostate cancer-specific mortality (PCSM). *BJU Int.* 2012;110(7):973.
32. Gallus S, Foschi R, Talamini R, et al. Risk Factors for Prostate Cancer in Men Aged Less Than 60 Years: A Case–Control Study from Italy. *Urology.* 2007;70(ISS6):1121-1126.
33. CDC CfDCaP. Prostate Cancer Statistics. 2010; www.cdc.gov/cancer/prostate/. Accessed April 2nd, 2014.
34. UK PC. Stand By Your Man. 2013; http://prostatecanceruk.org/. Accessed April 2nd, 2014.
35. Ben-Shlomo Y, Evans S, Ibrahim F, et al. The risk of prostate cancer amongst black men in the United Kingdom: the PROCESS cohort study. *Eur. Urol.* 2008;53(1):99-105.
36. Anderson B, Marshall-Lucette S, Webb P. African and Afro-Caribbean men's experiences of prostate cancer. *Br. J. Nurs.* 2013;22(22):1296-1307.
37. Villeneuve PJ, Johnson KC, Kreiger N, Mao Y. Risk Factors for Prostate Cancer: Results from the Canadian National Enhanced Cancer Surveillance System. *Cancer Causes Control.* 1999(5):355.
38. Platz EA, Giovannucci E. *Prostate Cancer.* Oxford University Press; 2006.
39. Etzioni R, Berry KM, Legler JM, Shaw P. Prostate-specific antigen testing in black and white men: an analysis of medicare claims from 1991–1998. *Urology.* 2002;59(2):251.
40. Breen N, Wagener DK, Brown ML, Davis WW, Ballard-Barbash R. Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *JNCI: Journal of the National Cancer Institute.* 2001;93(22):1704-1713.
41. Hoffman RM, Gilliland FD, Eley JW, et al. Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study. *JNCI: Journal of the National Cancer Institute.* 2001;93(5):388-395.

42. Gonzalez JR, Consedine NS, McKiernan JM, Spencer BA. Barriers to the initiation and maintenance of prostate specific antigen screening in Black American and Afro-Caribbean men. *The Journal Of Urology.* 2008;180(6):2403-2408.
43. Koscuiszka M, Hatcher D, Christos PJ, et al. Impact of race on survival in patients with clinically nonmetastatic prostate cancer who deferred primary treatment. *Cancer (0008543X).* 2012;118(12):3145.
44. Liu L, Cozen W, Bernstein L, Ross RK, Deapen D. Changing relationship between socioeconomic status and prostate cancer incidence. *J. Natl. Cancer Inst.* 2001;93(9):705-709.
45. Roetzheim RG, Pal N, Tennant C, et al. Effects of health insurance and race on early detection of cancer. *JNCI: Journal of the National Cancer Institute.* 1999;91(16):1409-1415.
46. Caire AA, Sun L, Polascik TJ, Albala DM, Moul JW. Obese African-Americans with prostate cancer (T1c and a prostate-specific antigen, PSA, level of <10 ng/mL) have higher-risk pathological features and a greater risk of PSA recurrence than non-African-Americans. *BJU Int.* 2010;106(8):1157.
47. Tewari AK, Gold HT, Demers RY, et al. Effect of socioeconomic factors on long-term mortality in men with clinically localized prostate cancer. *Urology.* 2009;73(3):624-630.
48. Du XL, Fang S, Coker AL, et al. Racial disparity and socioeconomic status in association with survival in older men with local/regional stage prostate carcinoma: findings from a large community-based cohort. *Cancer.* 2006;106(6):1276-1285.
49. Penner LA, Eggly S, Griggs JJ, Underwood W, Orom H, Albrecht TL. Life-Threatening Disparities: The Treatment of Black and White Cancer Patients. *J.Soc.Iss.* 2012;68(2):328.
50. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J. Clin.* 2014;64(1):9-29.
51. Stokes WA, Hendrix LH, Royce TJ, et al. Racial differences in time from prostate cancer diagnosis to treatment initiation. *Cancer (0008543X).* 2013;119(13):2486.
52. Gilligan T. Social disparities and prostate cancer: mapping the gaps in our knowledge. *Cancer Causes & Control: CCC.* 2005;16(1):45-53.
53. Roehl KA, Antenor JAV, Catalona WJ, Loeb S, Corbin N. Characteristics of Patients With Familial Versus Sporadic Prostate Cancer. *J. Urol.* 2006;176(6):2438-2442.
54. Wakefield CE, Meiser B, Gaff CL, et al. Issues faced by unaffected men with a family history of prostate cancer: a multidisciplinary overview. *The Journal Of Urology.* 2008;180(1):38.
55. Hemminki K, Czene K. Attributable risks of familial cancer from the family-cancer database. *Cancer Epidemiology Biomarkers and Prevention.* 2002;11(12):1638-1644.
56. Brandt A, Sundquist J, Hemminki K. Risk for incident and fatal prostate cancer in men with a family history of any incident and fatal cancer. *Ann. Oncol.* 2012;23(1):251-U298.
57. Ross LE, Berkowitz Z, Ekwueme DU. Use of the prostate-specific antigen test among U.S. men: Findings from the 2005 National Health Interview Survey. *Cancer Epidemiology Biomarkers and Prevention.* 2008;17(3):636-644.
58. Lichtenstein P, Iliadou A, Holm NV, et al. Environmental and heritable factors in the causation of cancer: Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* 2000;343(2):78-85.
59. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* 2008;40(3):310-315.
60. Wilson KM, Giovannucci EL, Mucci LA. Lifestyle and dietary factors in the prevention of lethal prostate cancer. *Asian Journal Of Andrology.* 2012;14(3):365.

61. Pomerantz MM, Freedman ML. Genetics of Prostate Cancer Risk. *Mt. Sinai J. Med.* 2010;77(6):643-654.
62. Hemminki K. Familial risk and familial survival in prostate cancer. *World J. Urol.* 2012;30(2):143-148.
63. Penney KL, Mucci LA, Kraft PL, et al. Genome-wide association study of prostate cancer mortality. *Cancer Epidemiology Biomarkers and Prevention.* 2010;19(11):2869-2876.
64. Bruner DW, Dorgan J, Engstrom P, Moore D, Parlanti A. Relative risk of prostate cancer for men with affected relatives: Systematic review and meta-analysis. *Int. J. Cancer.* 2003;107(5):797-803.
65. Schleutker J, Matikainen M, Smith J, et al. A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: Frequent HPCX linkage in families with late-onset disease. *Clin. Cancer Res.* 2000;6(12):4810-4815.
66. Christensen GB, Farnham JM, Cannon-Albright LA, et al. Genome-wide linkage analysis of 1,233 prostate cancer pedigrees from the International Consortium for prostate cancer genetics using novel sumLINK and sumLOD analyses. *Prostate.* 2010;70(7):735-744.
67. Sun J, Kader AK, Hsu FC, et al. Inherited genetic markers discovered to date are able to identify a significant number of men at considerably elevated risk for prostate cancer. *Prostate.* 2011;71(4):421-430.
68. Xu J, Zheng SL, Sun J, et al. Inherited genetic variant predisposes to aggressive but not indolent prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* 2010;107(5):2136-2140.
69. Bratt O. Hereditary prostate cancer: clinical aspects. *The Journal Of Urology.* 2002;168(3):906-913.
70. Madersbacher S, Alcaraz A, Emberton M, et al. The influence of family history on prostate cancer risk: implications for clinical management. *BJU Int.* 2011;107(5):716.
71. Carter BS, Beaty TH, Steinberg GD, Walsh PC, Childs B. Mendelian inheritance of familial prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* 1992;89(8):3367-3371.
72. Bastacky SI, Wojno KJ, Walsh PC, Carmichael MJ, Epstein JI. Original Articles: Prostate Cancer: Pathological Features of Hereditary Prostate Cancer. *The Journal of Urology.* 1995;153(3 SUPPL.):987-992.
73. Kang JS, Maygarden SJ, Mohler JL, Pruthi RS. Comparison of clinical and pathological features in African-American and Caucasian patients with localized prostate cancer. *BJU Int.* 2004;93(9):1207-1210.
74. Keetch DW, Humphrey PA, Smith DS, Stahl D, Catalona WJ. Clinical and pathological features of hereditary prostate cancer. *J. Urol.* 1996;155(6):1841-1843.
75. Sacco E, Prayer-Galetti T, Pinto F, et al. Familial and hereditary prostate cancer by definition in an Italian surgical series: Clinical features and outcome. *Eur. Urol.* 2005;47(6):761-768.
76. Carter BS, Bova GS, Beaty TH, et al. Hereditary prostate cancer: Epidemiologic and clinical features. *J. Urol.* 1993;150(3):797-802.
77. Salami MA, Etukakpan B, Olapade-Olaopa EO. Update on prostate cancer in black men. *Journal of Men's Health and Gender.* 2007;4(4):456-463.
78. Weinrich SP. Prostate cancer screening in high-risk men: African American Hereditary Prostate Cancer Study Network. *Cancer.* 2006;106(4):796-803.
79. Miller SM, Diefenbach MA, Kruus LK, Watkins-Bruner D, Hanks GE, Engstrom PF. Psychological and Screening Profiles of First-Degree Relatives of Prostate Cancer Patients. *J. Behav. Med.* 2001;24(3):247-258.

80. Cormier L, Mangin P, Guillemin F, et al. Impact of prostate cancer screening on health-related quality of life in at-risk families. *Urology.* 2002;59(6):901-906.
81. Vadaparampil ST, Jacobsen PB, Kash K, Watson IS, Saloup R, Pow-Sang J. Factors predicting prostate specific antigen testing among first-degree relatives of prostate cancer patients. *Cancer Epidemiology, Biomarkers & Prevention: A Publication Of The American Association For Cancer Research, Cosponsored By The American Society Of Preventive Oncology.* 2004;13(5):753-758.
82. Meiser B, Cowan R, Costello A, Giles GG, Lindeman GJ, Gaff CL. Prostate cancer screening in men with a family history of prostate cancer: the role of partners in influencing men's screening uptake. *Urology.* 2007;70(4):738-742.
83. Spencer BA, Babey SH, Etzioni DA, et al. A population-based survey of prostate-specific antigen testing among California men at higher risk for prostate carcinoma. *Cancer.* 2006;106(4):765-774.
84. Sweetman J, Watson M, Norman A, et al. Feasibility of familial PSA screening: psychosocial issues and screening adherence. *Br. J. Cancer.* 2006;94(4):507-512.
85. Bratt O, Lundgren R, Damber JE, et al. Risk perception, screening practice and interest in genetic testing among unaffected men in families with hereditary prostate cancer. *Eur. J. Cancer.* 2000;36(2):235-241.
86. Cerhan JR, Parker AS, Putnam SD, et al. Family history and prostate cancer risk in a population-based cohort of Iowa men. *Cancer Epidemiology Biomarkers and Prevention.* 1999;8(1):53-60.
87. Rodríguez C, Calle EE, Tatham LM, et al. Family history of breast cancer as a predictor for fatal prostate cancer. *Epidemiology.* 1998;9(5):525-529.
88. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J. Natl. Cancer Inst.* 1994;86(21):1600-1608.
89. Isaacs SD, Kiemeney LA, Baffoe-Bonnie A, Beaty TH, Walsh PC. Risk of cancer in relatives of prostate cancer probands. *J. Natl. Cancer Inst.* 1995;87(13):991-996.
90. Hayes RB, Liff JM, Pottern LM, et al. Prostate cancer risk in U.S. blacks and whites with a family history of cancer. *Int. J. Cancer.* 1995;60(3):361-364.
91. Thomas JA, Gerber L, Moreira DM, et al. Prostate cancer risk in men with prostate and breast cancer family history: results from the REDUCE study (R1). *J. Intern. Med.* 2012;272(1):85.
92. Holmberg L. Obesity, nutrition, and prostate cancer: insights and issues. *Eur. Urol.* 2013;63(5):821.
93. System BRFS. Obesity and Overweight for Professionals: Data and Statistics: Adult Obesity - DNPAO - CDC. 2014; http://www.cdc.gov/obesity/data/adult.html. Accessed April 23, 2014.
94. Kushi LH, Doyle C, McCullough M, et al. American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J. Clin.* 2012;62(1):30-67.
95. Ogden CL CM, McDowell MA, Flegal KM. *Obesity Among Adults in the United States--No Statistically Significant Change Since 2003-2004.* Hyattsville, MD: National Center for Health Statistics;2007.
96. Prevention CfDCa. Differences in prevalence of obesity among black, white, and Hispanic adults -- United States, 2006-2008. *MMWR Morb. Mortal. Wkly. Rep.* 2009;58(27):740-744.

97. Services UDoHaH. THE SURGEON GENERAL'S CALL TO ACTION TO PREVENT AND DECREASE OVERWEIGHT AND OBESITY. *Medical Benefits.* 2002;19(2):12.
98. Rodriguez C, Patel AV, Calle EE, Jacobs EJ, Chao A, Thun MJ. Body mass index, height, and prostate cancer mortality in two large cohorts of adult men in the United States. *Cancer Epidemiology Biomarkers and Prevention.* 2001;10(4):345-353.
99. Organization WH. *Obesity: preventing and managing the global epidemic.* Geneva: World Health Organization;2000. 894.
100. NHLBI Obesity Education Initiative Expert Panel on the Identification E, and Treatment of Obesity in Adults (US). *Clinical Guidelines: Identification, Evaluation, and Treatment of Overweight and Obesity.* Bethesda (MD): Legal Books Depot;1999. 10859985.
101. Calle EE, Kaaks R. OVERWEIGHT, OBESITY AND CANCER: EPIDEMIOLOGICAL EVIDENCE AND PROPOSED MECHANISMS. *Nature Reviews Cancer.* 2004;4(8):579-591.
102. Hursting SD, Smith SM, Lashinger LM, Harvey AE, Perkins SN. Calories and carcinogenesis: Lessons learned from 30 years of calorie restriction research. *Carcinogenesis.* 2009;31(1):83-89.
103. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J. Clin.* 2009;59(4):225-249.
104. Brown KA, Simpson ER. Obesity and Breast Cancer: Mechanisms and therapeutic implications. *Frontiers in Bioscience - Elite.* 2012;4 E(7):2515-2524.
105. Fenton JI, Hord NG, Lavigne JA, Perkins SN, Hursting SD. Leptin, insulin-like growth factor-1, and insulin-like growth factor-2 are mitogens in ApcMin/+ but not Apc+/+ colonic epithelial cell lines. *Cancer Epidemiology, Biomarkers & Prevention: A Publication Of The American Association For Cancer Research, Cosponsored By The American Society Of Preventive Oncology.* 2005;14(7):1646-1652.
106. Doyle SL, Donohoe CL, Finn SP, et al. IGF-1 and its receptor in esophageal cancer: association with adenocarcinoma and visceral obesity. *The American Journal Of Gastroenterology.* 2012;107(2):196-204.
107. Parekh N, Chandran U, Bandera EV. Obesity in Cancer Survival. *ANNUAL REVIEW OF NUTRITION, VOL 32.* 2012;32:311-342.
108. Edwards J, Krishna NS, Grigor KM, Bartlett JMS. Androgen receptor gene amplification and protein expression in hormone refractory prostate cancer. *Br. J. Cancer.* 2003;89(3):552-556.
109. Gong Z, Agalliu I, Lin DW, Stanford JL, Kristal AR. Obesity is associated with increased risks of prostate cancer metastasis and death after initial cancer diagnosis in middle-aged men. *Cancer.* 2007;109(6):1192-1202.
110. Efstathiou JA, Shipley WU, Bae K, et al. Obesity and mortality in men with locally advanced prostate cancer: Analysis of RTOG 85-31. *Cancer.* 2007;110(12):2691-2699.
111. Freedland SJ, Bañez LL, Sun LL, Fitzsimons NJ, Moul JW. Obese men have higher-grade and larger tumors: an analysis of the duke prostate center database. *Prostate Cancer Prostatic Dis.* 2009;12(3):259-263.
112. Freedland SJ, Aronson WJ, Kane CJ, et al. Impact of obesity on biochemical control after radical prostatectomy for clinically localized prostate cancer: A report by the shared equal access regional cancer hospital database study group. *J. Clin. Oncol.* 2004;22(3):446-453.
113. Buschemeyer WC, Freedland SJ, Allott EH, Masko EM, Freedland SJ. Obesity and Prostate Cancer: Epidemiology and Clinical Implications▪. *Eur. Urol.* 2013;52; 63(2; 5):331; 800.

114. Veierød MB, Laake P, Thelle DS. Dietary fat intake and risk of prostate cancer: A prospective study of 25,708 Norwegian men. *Int. J. Cancer.* 1997;73(5):634-638.
115. Engeland A, Tretli S, Bjørge T. Height, body mass index, and prostate cancer: a follow-up of 950000 Norwegian men. *Br. J. Cancer.* 2003;89(7):1237-1242.
116. Andersson SO, Wolk A, Bergström R, et al. Body size and prostate cancer: a 20-year follow-up study among 135006 Swedish construction workers. *JNCI: Journal of the National Cancer Institute.* 1997;89(5):385-389.
117. Putnam SD, Cerhan JR, Parker AS, et al. Lifestyle and anthropometric risk factors for prostate cancer in a cohort of Iowa men. *Ann. Epidemiol.* 2000;10(6):361-369.
118. Schuurman AG, Goldbohm RA, Dorant E, van den Brandt PA. Anthropometry in Relation to Prostate Cancer Risk in the Netherlands Cohort Study. *Am. J. Epidemiol.* 2009;151(6):541-549.
119. Lund Nilsen TI, Vatten LJ. Anthropometry and prostate cancer risk: A prospective study of 22,248 Norwegian men. *Cancer Causes Control.* 1999;10(4):269-275.
120. Giovannucci E, Rimm EB, Stampfer MJ, et al. Body mass index and risk of prostate cancer in U.S. health professionals. *J. Natl. Cancer Inst.* 2003;95(16):1240-1244.
121. Allott EH, Masko EM, Freedland SJ. Obesity and Prostate Cancer: Weighing the Evidence. *Eur. Urol.* 2013;63(5):800.
122. Miles L. Physical activity and the prevention of cancer: a review of recent findings. *Nutrition Bulletin.* 2007;32(3):250-282.
123. Services UDoHaH. Dietary Guidelines for Americans, 2010 | Dietary Guidelines for Americans | Health.gov (ODPHP). 2010; http://www.health.gov/dietaryguidelines/2010.asp. Accessed April 25, 2014.
124. Services USDoHaH. Vital Signs: Walking Among Adults — United States, 2005 and 2010. 2012; http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6131a4.htm. Accessed April 25, 2014.
125. Wannamethee SG, Shaper AG, Walker M. Physical activity and risk of cancer in middle-aged men. *Br. J. Cancer.* 2001;85(9):1311.
126. Mosher CE, Sloane R, Morey MC, et al. Associations between lifestyle factors and quality of life among older long-term breast, prostate, and colorectal cancer survivors. *Cancer.* 2009;115(17):4001-4009.
127. Thune I, Lund E. Physical activity and the risk of prostate and testicular cancer: A cohort study of 53,000 Norwegian men. *Cancer Causes Control.* 1994;5(6):549-556.
128. McTiernan A. Mechanisms linking physical activity with cancer. *Nature Reviews Cancer.* 2008;8(3):205-211.
129. Hsing AW, Chua S, Jr., Gao YT, et al. Prostate cancer risk and serum levels of insulin and leptin: a population-based study. *J. Natl. Cancer Inst.* 2001;93(10):783-789.
130. Chan JM, Stampfer MJ, Giovannucci E, et al. Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. *Science (New York, N.Y.).* 1998;279(5350):563-566.
131. Hackney AC, Sinning WE, Bruot BC. Reproductive hormonal profiles of endurance-trained and untrained males. *Med. Sci. Sports Exerc.* 1988;20(1):60-65.
132. Assah FK, Brage S, Ekelund U, Wareham NJ. The association of intensity and overall level of physical activity energy expenditure with a marker of insulin resistance. *Diabetologia.* 2008;51(8):1399-1407.
133. Ngo TH, Barnard RJ, Tymchuk CN, Cohen P, Aronson WJ. Effect of diet and exercise on serum insulin, IGF-I, and IGFBP-1 levels and growth of LNCaP cells in vitro (United States). *Cancer Causes Control.* 2002;13(10):929-935.

134. Liu Y, Hu F, Li D, et al. Does Physical Activity Reduce the Risk of Prostate Cancer? A Systematic Review and Meta-analysis. *Eur. Urol.* 2011;60(5):1029.
135. Singh AA, Jones LW, Antonelli JA, et al. Association between exercise and primary incidence of prostate cancer. *Cancer.* 2013;119(7):1338.
136. (NIAAA) NIoAAaA. Alcohol Facts and Statistics | National Institute on Alcohol Abuse and Alcoholism (NIAAA). 2014; http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-facts-and-statistics. Accessed April 25, 2014.
137. Sadeu JC, Hughes CL, Agarwal S, Foster WG. Alcohol, drugs, caffeine, tobacco, and environmental contaminant exposure: Reproductive health consequences and clinical implications. *Crit. Rev. Toxicol.* 2010;40(7):633-652.
138. Rota M, Scotti L, Turati F, et al. Alcohol consumption and prostate cancer risk: a meta-analysis of the dose-risk relation. *Eur. J. Cancer Prev.* 2012;21(4):350-359.
139. Dagnelie PC, Schuurman AG, Goldbohm RA, Van Den Brandt PA. Diet, anthropometric measures and prostate cancer risk: a review of prospective cohort and intervention studies. *BJU Int.* 2004;93(8):1139-1150.
140. Leitzmann MF, Rohrmann S. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. *Clin. Epidemiol.* 2012;4:1-11.
141. Rizos C, Papassava M, Golias C, Charalabopoulos K. Alcohol consumption and prostate cancer: a mini review. *Exp. Oncol.* 2010;32(2):66-70.
142. Li Y, Yang H, Cao J, et al. Association between Alcohol Consumption and Cancers in the Chinese Population -- A Systematic Review and Meta-Analysis. *Eur. Urol.* 2013;6; 24; 63(4; 3; 5):1; 451; 810.
143. Middleton Fillmore K, Chikritzhs T, Stockwell T, Bostrom A, Pascal R. Alcohol use and prostate cancer: a meta-analysis. *Molecular Nutrition & Food Research.* 2009;53(2):240-255.
144. Dennis LK. Meta-analysis for combining relative risks of alcohol consumption and prostate cancer. *Prostate.* 2000;42(1):56-66.
145. Rohrmann S, Linseisen J, Kaaks R, et al. Alcohol consumption and the risk for prostate cancer in the european prospective investigation into cancer and nutrition. *Cancer Epidemiology Biomarkers and Prevention.* 2008;17(5):1282-1287.
146. Sesso HD, Paffenbarger Jr RS, Lee IM. Alcohol consumption and risk of prostate cancer: The Harvard alumni health study. *Int. J. Epidemiol.* 2001;30(4):749-755.
147. Platz EA, Leitzmann MF, Rimm EB, Willett WC, Giovannucci E. Alcohol Intake, Drinking Patterns, and Risk of Prostate Cancer in a Large Prospective Cohort Study. *Am. J. Epidemiol.* 2004;159(5):444-453.
148. Velicer CM, Kristal A, White E. Alcohol use and the risk of prostate cancer: results from the VITAL cohort study. *Nutr. Cancer.* 2006;56(1):50-56.
149. Baglietto L, Severi G, English DR, Giles GG, Hopper JL. Alcohol consumption and prostate cancer risk: Results from the Melbourne collaborative cohort study. *Int. J. Cancer.* 2006;119(6):1501-1504.
150. Weinstein SJ, Stolzenberg-Solomon R, Pietinen P, Taylor PR, Virtamo J, Albanes D. Dietary factors of one-carbon metabolism and prostate cancer risk. *The American Journal Of Clinical Nutrition.* 2006;84(4):929-935.
151. Jie Y, Li-Xin Q, Hong-Fei W, et al. Genetic polymorphisms in the cytochrome P450 1A1 and 2E1 genes, smoking, drinking and prostate cancer susceptibility: A case-control study in a Han nationality population in Southern China. *Int. J. Urol.* 2006;13(6):773-780.
152. Cox B, Sneyd MJ, Paul C, Skegg DCG. Risk factors for prostate cancer: A national case-control study. *Int. J. Cancer.* 2006;119(7):1690-1694.

153. Schoonen WM, Salinas CA, Stanford JL, Kiemeney LALM. Alcohol consumption and risk of prostate cancer in middle-aged men. *Int. J. Cancer.* 2005;113(1):133-140.
154. Chang ET, Hedelin M, Adami HO, Bälter KA, Grönberg H. Alcohol drinking and risk of localized versus advanced and sporadic versus familial prostate cancer in Sweden. *Cancer Causes Control.* 2005;16(3):275-284.
155. Cáceres L DD, Iturrieta G J, Varela F N, Quiñones S L, Acevedo C C, Huidobro A C. Relationship among metabolizing genes, smoking and alcohol used as modifier factors on prostate cancer risk: Exploring some gene-gene and gene-environment interactions. *Eur. J. Epidemiol.* 2005;20(1):79-88.
156. Sharpe CR, Siemiatycki J. Case-control study of alcohol consumption and prostate cancer risk in Montréal, Canada. *Cancer Causes Control.* 2001;12(7):589-598.
157. McGregor SE, Courneya KS, Kopciuk KA, Tosevski C, Friedenreich CM. Case-control study of lifetime alcohol intake and prostate cancer risk. *Cancer Causes Control.* 2013;24(3):451.
158. Sutcliffe S, Platz EA, Giovannucci E, et al. A prospective cohort study of red wine consumption and risk of prostate cancer. *Int. J. Cancer.* 2007;120(7):1529-1535.
159. Gong Z, Kristal AR, Schenk JM, Tangen CM, Goodman PJ, Thompson IM. Alcohol consumption, finasteride, and prostate cancer risk: results from the Prostate Cancer Prevention Trial. *Cancer.* 2009;115(16):3661.
160. Shannon J, Phoutrides E, Palma A, et al. Folate intake and prostate cancer risk: a case-control study. *Nutr. Cancer.* 2009;61(5):617-628.
161. Chao C, Haque R, Poon KYT, Quinn VP, Van Den Eeden SK, Caan BJ. Red wine consumption and risk of prostate cancer: The California men's health study. *Int. J. Cancer.* 2010;126(1):171-179.
162. Watters JL, Park Y, Hollenbeck A, Schatzkin A, Albanes D. Alcoholic Beverages and Prostate Cancer in a Prospective US Cohort Study. *Am. J. Epidemiol.* 2010;172(7):773.
163. Kristal AR, Stanford JL. Cruciferous vegetables and prostate cancer risk: confounding by PSA screening. *Cancer Epidemiology, Biomarkers & Prevention: A Publication Of The American Association For Cancer Research, Cosponsored By The American Society Of Preventive Oncology.* 2004;13(7):1265-1265.
164. Sawada N, Inoue M, Iwasaki M, et al. Alcohol and smoking and subsequent risk of prostate cancer in Japanese men: The Japan Public Health Center-based prospective study. *Int. J. Cancer.* 2014;134(4):971-978.
165. Association AL. *Trends in Tobacco Use.* Chicago, IL: American Lung Association;2011.
166. Services UDoHaH. Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General. 2012.
167. Carmona RH. The Health Consequences of Smoking A Report of the Surgeon General. *Medical Benefits.* 2004;21(13):3-3.
168. Secretan B, Straif K, Baan R, et al. A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *The Lancet Oncology.* 2009;10(11):1033-1034.
169. Danaei G, Ding EL, Mozaffarian D, et al. The preventable causes of death in the United States: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med.* 2009;6(4).
170. SAMHSA CfBHSaQ. Results from the 2012 National Survey on Drug Use and Health: Detailed Tables. *National Survey on Drug Use and Health.* 2013(130319).
171. Huncharek M, Haddock KS, Reid R, Kupelnick B. Smoking as a risk factor for prostate cancer: a meta-analysis of 24 prospective cohort studies. *Am. J. Public Health.* 2010;100(4):693.

172. Moreira DM, Aronson WJ, Terris MK, et al. Cigarette smoking is associated with an increased risk of biochemical disease recurrence, metastasis, castration-resistant prostate cancer, and mortality after radical prostatectomy: results from the SEARCH database. *Cancer.* 2014;120(2):197.
173. Zu K, Giovannucci E. Smoking and Aggressive Prostate Cancer: A Review of the Epidemiologic Evidence. *Cancer Causes Control.* 2009(10):1799.
174. Rohrmann S, Linseisen J, Allen N, et al. Smoking and the risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition. *Br. J. Cancer.* 2013;108(3):708.
175. Nelson EC, Rodriguez RL, Dawson K, Galvez AF, Evans CP. The interaction of genetic polymorphisms with lifestyle factors: implications for the dietary prevention of prostate cancer. *Nutr. Cancer.* 2008;60(3):301.
176. Divisi D, Di Tommaso S, Salvemini S, Garramone M, Crisci R. Diet and cancer. *Acta Biomedica de l'Ateneo Parmense.* 2006;77(2):118-123+127.
177. Gonzalez CA. Nutrition and cancer: the current epidemiological evidence. *The British Journal Of Nutrition.* 2006;96 Suppl 1:S42-S45.
178. Rose DP, Boyar AP, Wynder EL. International comparisons of mortality rates for cancer of the breast, ovary, prostate, and colon, and per capita food consumption. *Cancer.* 1986;58(11):2363-2371.
179. Giovannucci E, Rimm EB, Colditz GA, et al. A Prospective Study of Dietary Fat and Risk of Prostate Cancer (vol 85, pg 1571, 1993). *JNCI-JOURNAL OF THE NATIONAL CANCER INSTITUTE.* 2013;105(13):994-994.
180. Giovannucci E, Stampfer MJ, Willett WC, Liu Y, Platz EA. Risk factors for prostate cancer incidence and progression in the health professionals follow-up study. *Int. J. Cancer.* 2007;121(7):1571-1578.
181. Leitzmann MF, Stampfer MJ, Michaud DS, et al. Dietary intake of n-3 and n-6 fatty acids and the risk of prostate cancer. *Am. J. Clin. Nutr.* 2004;80(1):204-216.
182. Koralek DO, Kirsh V, Schatzkin A, et al. A prospective study of dietary alpha-linolenic acid and the risk of prostate cancer (United States). *Cancer Causes Control.* 2006;17(6):783-791.
183. Brasky TM, Till C, White E, et al. Serum Phospholipid Fatty Acids and Prostate Cancer Risk: Results From the Prostate Cancer Prevention Trial. *Am. J. Epidemiol.* 2011;173(12):1429-1439.
184. Fleshner N, Zlotta AR. Prostate cancer prevention: past, present, and future. *Cancer.* 2007;110(9):1889-1899.
185. Fleshner N, Bagnell PS, Klotz L, Venkateswaran V. Dietary fat and prostate cancer. *The Journal Of Urology.* 2004;171(2 Pt 2):S19-S24.
186. Park SY, Murphy SP, Wilkens LR, Kolonel LN, Henderson BE. Fat and meat intake and prostate cancer risk: The Multiethnic Cohort Study. *Int. J. Cancer.* 2007;121(6):1339-1345.
187. Colli JL, Colli A. International comparisons of prostate cancer mortality rates with dietary practices and sunlight levels. *Urologic Oncology: Seminars and Original Investigations.* 2006;24(3):184-194.
188. Rohan TE, Howe GR, Burch JD, Jain M. Dietary factors and risk of prostate cancer: A case-control study in Ontario, Canada. *Cancer Causes Control.* 1995;6(2):145-154.
189. Lee MM, Wang RT, Gu FL, Wang T, Spitz M, Hsing AW. Case-control study of diet and prostate cancer in China. *Cancer Causes Control.* 1998;9(6):545-552.

190. Bidoli E, Talamini R, Bosetti C, et al. Macronutrients, fatty acids, cholesterol and prostate cancer risk. *Ann. Oncol.* 2005;16(1):152-157.
191. Kristal AR, Cohen JH, Qu P, Stanford JL. Associations of energy, fat, calcium, and vitamin D with prostate cancer risk. *Cancer Epidemiology Biomarkers and Prevention.* 2002;11(8):719-725.
192. Godley PA, Sandier RS, Campbell MK, Mohler JL, Gallagher P, Martinson FEA. Biomarkers of essential fatty acid consumption and risk of prostatic carcinoma. *Cancer Epidemiology Biomarkers and Prevention.* 1996;5(11):889-895.
193. De Stéfani E, Deneo-Pellegrini H, Boffetta P, Ronco A, Mendilaharsu M. Alpha-linolenic acid and risk of prostate cancer: a case-control study in Uruguay. *Cancer Epidemiology, Biomarkers & Prevention: A Publication Of The American Association For Cancer Research, Cosponsored By The American Society Of Preventive Oncology.* 2000;9(3):335-338.
194. Key T, Silcocks PB, Davey GK, Appleby PN, Bishop DT. A case-control study of diet and prostate cancer. *Br. J. Cancer.* 1997;76(5):678-687.
195. McCann SE, Ambrosone CB, Moysich KB, et al. Intakes of selected nutrients, foods, and phytochemicals and prostate cancer risk in western New York. *Nutr. Cancer.* 2005;53(1):33-41.
196. Kurahashi N, Inoue M, Iwasaki M, Sasazuki S, Tsugane S. Dairy product, saturated fatty acid, and calcium intake and prostate cancer in a prospective cohort of Japanese men. *Cancer Epidemiology Biomarkers and Prevention.* 2008;17(4):930-937.
197. Hu J, La Vecchia C, Gibbons L, Negri E, Mery L. Nutrients and risk of prostate cancer. *Nutr. Cancer.* 2010;62(6):710-718.
198. Shimizu H, Ross RK, Bernstein L, Yatani R, Henderson BE, Mack TM. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. *Br. J. Cancer.* 1991;63(6):963-966.
199. Yu H, Harris RE, Gao YT, Gao R, Wynder EL. Comparative epidemiology of cancers of the colon, rectum, prostate and breast in Shanghai, China versus the United States. *Int. J. Epidemiol.* 1991;20(1):76-81.
200. Pelser C, Mondul AM, Hollenbeck AR, Park Y. Dietary fat, fatty acids, and risk of prostate cancer in the NIH-AARP diet and health study. *Cancer Epidemiol. Biomarkers Prev.* 2013;22(4):697-707.
201. Pandian SS, Eremin OE, McClinton S, Wahle KWJ, Heys SD. Fatty acids and prostate cancer: current status and future challenges. *J. R. Coll. Surg. Edinb.* 1999;44(6):352-361.
202. Gann PH. Risk factors for prostate cancer. *Reviews In Urology.* 2002;4 Suppl 5:S3-S10.
203. Park SY, Kolonel LN. *Nutrition and Prostate Cancer.* Elsevier Inc.; 2013.
204. Lophatananon A, Archer J, Easton D, et al. Dietary fat and early-onset prostate cancer risk. *Br. J. Nutr.* 2010;103(9):1375-1380.
205. Epstein MM, Kasperzyk JL, Mucci LA, et al. Dietary fatty acid intake and prostate cancer survival in Orebro County, Sweden. *Am. J. Epidemiol.* 2012;176(3):240-252.
206. Krycer JR, Brown AJ. Does Changing Androgen Receptor Status during Prostate Cancer Development Impact upon Cholesterol Homeostasis? *PLoS One.* 2013;8(1):1-10.
207. Pelton K, Freeman MR, Solomon KR. Cholesterol and prostate cancer. *Curr. Opin. Pharmacol.* 2012;12(6):751-759.
208. Kok DEG, van Roermund JGH, Aben KKH, et al. Blood lipid levels and prostate cancer risk; a cohort study. *Prostate Cancer Prostatic Dis.* 2011;14(4):340-345.

209. Shafique K, McLoone P, Qureshi K, Leung H, Hart C, Morrison DS. Cholesterol and the risk of grade-specific prostate cancer incidence: evidence from two large prospective cohort studies with up to 37 years' follow up. *BMC Cancer.* 2012;12(1):1-8.
210. Murtola TJ, Syvälä H, Pennanen P, et al. The Importance of LDL and Cholesterol Metabolism for Prostate Epithelial Cell Growth. *PLoS One.* 2012;7(6):1-8.
211. Hanchette CL, Schwartz GG. Geographic patterns of prostate cancer mortality: Evidence for a protective effect of ultraviolet radiation. *Cancer.* 1992;70(12):2861-2869.
212. Schwartz GG, Hanchette CL. UV, latitude, and spatial trends in prostate cancer mortality: All sunlight is not the same (United States). *Cancer Causes Control.* 2006;17(8):1091-1101.
213. Giovannucci E. The epidemiology of vitamin D and cancer incidence and mortality: a review (United States). *Cancer Causes & Control: CCC.* 2005;16(2):83-95.
214. Schwartz GG. Vitamin D, Sunlight, and the Epidemiology of Prostate Cancer. *Anticancer Agents Med. Chem.* 2013;13(1):45-57.
215. Garland CF, Garland FC, Gorham ED, et al. The Role of Vitamin D in Cancer Prevention. *Am. J. Public Health.* 2006;96(2):252-261.
216. Lou YR, Qiao S, Talonpoika R, Syvälä H, Tuohimaa P. The role of Vitamin D3 metabolism in prostate cancer. *The Journal Of Steroid Biochemistry And Molecular Biology.* 2004;92(4):317-325.
217. van der Rhee HJ, de Vries E, Coebergh JWW. Does sunlight prevent cancer? A systematic review. *Eur. J. Cancer.* 2006;42(14):2222-2232.
218. Kricker A, Armstrong B. Does sunlight have a beneficial influence on certain cancers? *Prog. Biophys. Mol. Biol.* 2006;92(1):132-139.
219. Strange RC, Luscombe CJ, French ME, et al. Outcome in prostate cancer: Associations with skin type and polymorphism in pigmentation-related genes. *Carcinogenesis.* 2001;22(9):1343-1347.
220. Grant WB. Geographic variation of prostate cancer mortality rates in the United States: Implications for prostate cancer risk related to vitamin D. *International Journal Of Cancer. Journal International Du Cancer.* 2004;111(3):470-471.
221. Holt SK, Koopmeiners JS, Feng Z, et al. Vitamin D pathway gene variants and prostate cancer prognosis. *Prostate.* 2010;70(13):1448-1460.
222. Hendrickson WK, Flavin R, Kasperzyk JL, et al. Vitamin d receptor protein expression in tumor tissue and prostate cancer progression. *J. Clin. Oncol.* 2011;29(17):2378-2385.
223. Fang F, Kasperzyk JL, Irene S, et al. Prediagnostic Plasma Vitamin D Metabolites and Mortality among Patients with Prostate Cancer. *PLoS One.* 2011;6(4):1-6.
224. Schwartz GG, Hulka BS. Is vitamin D deficiency a risk factor for prostate cancer? (hypothesis). *Anticancer Res.* 1990;10(5 A):1307-1312.
225. Gilbert R, Metcalfe C, Oliver SE, et al. Life course sun exposure and risk of prostate cancer: population-based nested case-control study and meta-analysis. *Int. J. Cancer.* 2009;125(6):1414-1423.
226. Gilbert R, Metcalfe C, Fraser WD, et al. Associations of circulating 25-hydroxyvitamin D with prostate cancer diagnosis, stage and grade. *Int. J. Cancer.* 2012;131(5):1187-1196.
227. Barnett CM, Beer TM, Nielson CM, et al. Serum 25-OH vitamin D levels and risk of developing prostate cancer in older men. *Cancer Causes Control.* 2010;21(8):1297-1303.
228. Gupta D, Lammersfeld CA, Trukova K, Lis CG. Vitamin D and prostate cancer risk: a review of the epidemiological literature. *Prostate Cancer Prostatic Dis.* 2009;12(3):215-226.

229. Ames BN, Gold LS, Willett WC. The Causes and Prevention of Cancer. *Proc. Natl. Acad. Sci. U. S. A.* 1995(12):5258.
230. Campisi J. Cancer and ageing: rival demons? *Nature Reviews Cancer.* 2003;3(5):339.
231. De Marzo AM, Platz EA, Sutcliffe S, et al. Inflammation in prostate carcinogenesis. *Nature Reviews Cancer.* 2007;7(4):256-269.
232. Gerstenbluth RE, Seftel AD, MacLennan GT, et al. Distribution of chronic prostatitis in radical prostatectomy specimens with up-regulation of bcl-2 in areas of inflammation. *J. Urol.* 2002;167(5):2267-2270.
233. Gurel B, Lucia MS, Thompson IM, Jr., et al. Chronic inflammation in benign prostate tissue is associated with high-grade prostate cancer in the placebo arm of the prostate cancer prevention trial. *Cancer Epidemiology, Biomarkers & Prevention: A Publication Of The American Association For Cancer Research, Cosponsored By The American Society Of Preventive Oncology.* 2014;23(5):847-856.
234. Moreira DM, Nickel JC, Gerber L, et al. Baseline prostate inflammation is associated with a reduced risk of prostate cancer in men undergoing repeat prostate biopsy: Results from the REDUCE study. *Cancer (0008543X).* 2014;120(2):190-196.
235. Carter HB. Prostate Disorders 2013. *Prostate Disorders.* 2013:25.
236. Dennis LK, Dawson DV, Resnick MI. Vasectomy and the risk of prostate cancer: a meta-analysis examining vasectomy, status, age at vasectomy, and time since vasectomy. *Prostate Cancer Prostatic Dis.* 2002;5(3):193.
237. Holt SK, Salinas CA, Stanford JL. Adult Urology: Vasectomy and the Risk of Prostate Cancer. *The Journal of Urology.* 2008;180:2565-2568.
238. Friedman GD. Blood pressure and heart rate: No evidence for a positive association with prostate cancer. *Ann. Epidemiol.* 1997;7(7):486-489.
239. Fitzpatrick AL, Daling JR, Furberg CD, Kronmal RA, Weissfeld JL. Hypertension, heart rate, use of antihypertensives, and incident prostate cancer. *Ann. Epidemiol.* 2001;11(8):534-542.
240. Jennifer LBD, Rodney LD, Aruna VS, James EM, Kathleen AC. Features of the metabolic syndrome and prostate cancer in African-American men. *Cancer (0008543X).* 2007;109(5):875-881.
241. Takeshita K, Takahashi S, Tang M, Seeni A, Asamoto M, Shirai T. Hypertension is positively associated with prostate cancer development in the TRAP transgenic rat model. *Pathol. Int.* 2011;61(4):202-209.
242. Asmar R, Beebe-Dimmer JL, Korgavkar K, Keele GR, Cooney KA. Hypertension, obesity and prostate cancer biochemical recurrence after radical prostatectomy. *Prostate Cancer Prostatic Dis.* 2013;16(1):62-66.
243. Xu H, Zhang LM, Liu J, Ding GX, Ding Q, Jiang HW. The association between overall survival of prostate cancer patients and hypertension, hyperglycemia, and overweight in Southern China: A prospective cohort study. *J. Cancer Res. Clin. Oncol.* 2013;139(6):943-951.
244. Kamruzzaman J, Begg R, Sarker R. *Overview of Artificial Neural Networks and Their Applications in Healthcare.* IGI Global; 2006.
245. Ghavami P, Kapur K. Prognostics & artificial neural network applications in patient healthcare. *2011 IEEE Conference on Prognostics & Health Management (PHM).* 2011:1.
246. Micheli-Tzanakou E. Artificial Neural Networks: An Overview. *Network: Computation in Neural Systems.* 2011;22(1-4):208-230.

247. Polascik TJ, Oesterling JE, Partin AW. Prostate specific antigen: a decade of discovery--what we have learned and where we are going. *The Journal Of Urology.* 1999;162(2):293-306.
248. Hudson MA, Bahnson RR, Catalona WJ. Clinical use of prostate specific antigen in patients with prostate cancer. *J. Urol.* 1989;142(4):1011-1017.
249. Bob D, Amir K, Lori D, et al. Diagnostic Strategies for Prostate Cancer. *European Urology Supplements.* 2011;10:e26-e37.
250. Kim SY, Moon SK, Jung DC, et al. Pre-operative prediction of advanced prostatic cancer using clinical decision support systems: accuracy comparison between support vector machine and artificial neural network. *Korean journal of radiology : official journal of the Korean Radiological Society.* 2011;12(5):588-594.
251. Chen ST, Kuo SJ, Tseng HS, et al. Comparative analysis of logistic regression, support vector machine and artificial neural network for the differential diagnosis of benign and malignant solid breast tumors by the use of three-dimensional power Doppler imaging. *Korean Journal of Radiology.* 2009;10(5):464-471.
252. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M. Novel artificial neural network for early detection of prostate cancer. *J. Clin. Oncol.* 2002;20(4):921-929.
253. Saritas I, Ozkan IA, Sert IU. Prognosis of prostate cancer by artificial neural networks. *Expert Systems with Applications.* 2010;37(9):6646-6650.
254. Ecke TH, Bartel P, Hallmann S, et al. Outcome prediction for prostate cancer detection rate with artificial neural network (ANN) in daily routine. *Urol. Oncol.* 2012;30(2):139-144.
255. Ecke TH, Hallmann S, Koch S, et al. External validation of an artificial neural network and two nomograms for prostate cancer detection. *ISRN urology.* 2012;2012:643181.
256. Stephan C, Cammann H, Meyer HA, et al. An artificial neural network for five different assay systems of prostate-specific antigen in prostate cancer diagnostics. *BJU Int.* 2008;102(7):799-805.
257. Delen D. Analysis of cancer data: a data mining approach. *Expert Systems.* 2009;26(1):100-112.
258. Dombi GW, Rosbolt JP, Severson RK. Neural network analysis of employment history as a risk factor for prostate cancer. *Comput. Biol. Med.* 2010;40(9):751-757.
259. Stephan C, Cammann H, Deger S, et al. Benign Prostatic Hyperplasia-associated Free Prostate-specific Antigen Improves Detection of Prostate Cancer in an Artificial Neural Network. *Urology.* 2009;74(4):873-877.
260. Stephan C, Büker N, Cammann H, Meyer H-A, Lein M, Jung K. Artificial neural network (ANN) velocity better identifies benign prostatic hyperplasia but not prostate cancer compared with PSA velocity. *BMC Urol.* 2008;8:10-10.
261. Lisboa PJ, Taktak AF. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural networks : the official journal of the International Neural Network Society.* 2006;19(4):408-415.
262. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.