A Thesis/Project/Dissertation Report
On

PROGNOSIS OF ACUTE AND CHRONIC DISEASE

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE & ENGINEERING

GALGOTIAS UNIVERSITY
(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under the supervision of:
Mr. Lalit Kumar
(Assistant Professor of Galgotias University)

Summitted By:

Uday Gupta & Ved Prakash Verma

18SCSE1010316 & 18SCSE1010346

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

INDIA

DECEMBER,2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"PROGNOSIS OF ACUTE AND CHRONIC DISEASE"** in partial fulfillment of the requirements for the award of the Bachelors of Technology submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Name **"Mr. Lalit Kumar"** Designation, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering, Galgotias University, Greater Noida.

The matter presented in the thesis/project/dissertation has not been submitted by us for the award of any other degree of this or any other places.

Uday Gupta18SCSE10101316)

Ved Prakash Verma(18SCSE1010346)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor Name

Designation

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **"UDAY GUPTA (18SCSE1010316) & VED PRAKASH VERMA (18SCSE1010346)"** has been held on __________________ and their work is recommended for the award of B. Tech (CSE).

Signature of Examiner(s)

Signature of Supervisor(s)

Signature of Project Coordinator

Signature of Dean

Date: December, 2021

Place: Greater Noida

ACKNOWLEDGEMENT

Apart from the efforts of our, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We would like to show my greatest appreciation to “**Mr. Lalit Kumar”** and also our dean “**Dr. Munish Sabharwal”**. We can’t say thank you enough for his tremendous support and help. We feel motivated and encouraged every time, We attend his meeting. Without his encouragement and guidance this project would not have materialized.

The guidance and support received from all the members who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Abstract

After pandemic has gone everybody is trying to be health conscious to their health. There are some possibilities that we have a disease in our body but we are unaware of the disease. It could be any type of disease at any level of its phenomena. So, we have a dataset based on patient's health history and using this data we are going to create an algorithm using Machine Learning by which we will train our model. And this model will predict some sufferable disease for patients which will be going to affect the patients in future.

Whenever we are visiting a doctor, then a doctor can give you the best advice only if he knows accurate and full information of your past health history. Here our model has a unique feature where we can store a patient health history online and this health history can be sharable among all of the hospitals where the patient will be visiting for his/her treatment in future.

We are using Azure cloud computing to store the data, machine learning to train the model and seaborn, SciPy, sklearn to predict the diseases.

With the following data the model will predict the disease which will come in future and aware to all peoples for future. I have used the Standard file from Kaggle who predict at most 40+ diseases at a time.

With the help of this project, all individuals will aware of their health histories and we can also share our data with the government if they follow our terms and conditions. The government can use our data in terms of maintaining and taking measure steps for the betterment of the health of each citizen in a particular region or state.

List of Tables

List of Figures

Table Of Contents

Chapter-1

Introduction

Many older adults will die from complications of chronic diseases such as congestive heart failure (CHF), dementia, and chronic obstructive pulmonary disease (COPD). In 2001, heart disease was the leading cause of death in the United States, with chronic lower respiratory tract illnesses, dementia, and renal disease ranked in the top ten, according to the Centers for Disease Control and Prevention statistics.1 Rather than dying suddenly, patients with these diseases often experience a gradual decline in health punctuated by exacerbations of their disease.2 These patients may live for years with multiple serious illnesses.

On one hand, this illness trajectory can help clinicians to initiate end-of-life discussions because a slow period of decline with multiple intercurrent events offers numerous prompts and opportunities. However, this extended illness course also makes prognosis very difficult to predict. Moreover, this uncertainty in estimating prognosis can be particularly difficult to explain to patients and families.

Chronic diseases/non-communicable diseases are currently the major cause of death among adults in almost all countries and the toll is projected to increase by a further 17% in the next 10 years. Globally, approximately one in three of all adults suffer from multiple chronic conditions. Six in ten adults in the US have a chronic disease and four in ten adults have two or more. It has been calculated that, of the 58 million deaths in 2005, approximately 35 million will be as a result of chronic diseases. Health damaging behaviors - particularly tobacco use, lack of physical activity, and poor eating habits, excessive alcohol use - are major contributors to the leading chronic diseases. The leading chronic diseases in developed countries include (in alphabetical order) arthritis, cardiovascular disease e.g., heart attacks and stroke, cancer e.g. breast and colon cancer, diabetes, epilepsy and seizures, obesity, and oral health problems. Each of these conditions plagues older adults. This rise in Chronic diseases (CDs) is a very serious situation, both for public health and for the societies and economies affected. Until recently, the impact and profile of chronic diseases have generally been insufficiently appreciated.

1.2 Formulation of Problem:

In India, as in many developing countries, public health advocacy to date has been mainly devoted to infectious diseases. However, we now have major public health issues due to chronic diseases that need to be addressed with equal energy and focus.

This World Health Organization report, preventing chronic diseases: a vital investment, is of relevance to me, as Indian Minister for Health, as my country tackles the increasing number of issues relating to chronic disease. The scale of the problem we face is clear with the projected number of deaths attributable to chronic diseases rising from 3.78 million in 1990 (40.4% of all deaths) to an expected 7.63 million in 2020 (66.7% of all deaths).

A number of my fellow citizens are featured within this report, as Faces of Chronic Disease. You will read about K. Sridhar Reddy, who, like a huge proportion of Indians, consumed tobacco and battled both serious cancer and associated financial debts. His story is all too familiar in a country which is the world's second largest producer, as well as consumer, of tobacco, where we consequently experience huge rates of cancer, including the largest numbers of oral cancer in the world. This costs the country dearly, for the individuals affected, but also in terms of treatment costs for tobacco-related diseases, estimated at US$ 7.2 billion just for the year 2002–2003.

A chronic condition "is a physical or mental health condition that lasts more than one year and causes functional restrictions or requires ongoing monitoring or treatment" [1,2]. Chronic diseases are among the most prevalent and costly health conditions in the United States. Nearly half (approximately 45%, or 133 million) of all Americans suffer from at least one chronic disease [3,4,5], and the number is growing. Chronic diseases—including, cancer, diabetes, hypertension, stroke, heart disease, respiratory diseases, arthritis, obesity, and oral diseases—can lead to hospitalization, long-term disability, reduced quality of life, and death [6,7]. In fact, persistent conditions are the nation's leading cause of death and disability [6].

Globally, chronic diseases have affected the health and quality of life of many citizens [8,9]. In addition, chronic diseases have been a major driver of health care costs while also impacting workforce patterns, including, of course, absenteeism. According to the Centers for Disease Control, in the U.S. alone, chronic diseases account for nearly 75 percent of aggregate healthcare spending, or an estimated $5300 per person annually. In terms of public insurance, treatment of chronic diseases comprises an even larger proportion of spending: 96 cents per dollar for Medicare and 83 cents per dollar for Medicaid [4,10,11,12]. Thus, the understanding, management, and prevention of chronic diseases are important objectives if, as a society, we are to provide better quality healthcare to citizens and improve their overall quality of life.

More than two thirds of all deaths are caused by one or more of these five chronic diseases: heart disease, cancer, stroke, chronic obstructive pulmonary disease, and diabetes. Additional statistics are quite stark [5,13]: chronic diseases are responsible for seven out of 10 deaths in the U.S., killing more than 1.7 million Americans each year; and more than 75% of the $2 trillion spent on public and private healthcare in 2005 went toward chronic diseases [5]. What makes treating chronic conditions (and efforts to manage population health) particularly challenging is that chronic conditions often do not exist in isolation. In fact, today one in four U.S. adults have two or more chronic conditions [5], while more than half of older adults have three or more chronic conditions. And the likelihood of these types of comorbidities occurring goes up as we age [5]. Given America's current demographics, wherein 10,000 Americans will turn 65 each day from now through the end of 2029 [5], it is reasonable to expect that the overall number of patients with comorbidities will increase greatly.

Trends show an overall increase in chronic diseases. Currently, the top ten health problems in America (not all of them chronic) are heart disease, cancer, stroke, respiratory disease, injuries, diabetes, Alzheimer's disease, influenza and pneumonia, kidney disease, and septicemia [14,15,16,17,18]. The nation's aging population, coupled with existing risk factors (tobacco use, poor nutrition, lack of physical activity) and medical advances that extend longevity (if not also improve overall health), have led to the conclusion that these problems are only going to magnify if not effectively addressed now [19].

A recent Milken Institute analysis determined that treatment of the seven most common chronic diseases coupled with productivity losses will cost the U.S. economy more than $1 trillion dollars annually. Furthermore, compared with other developed nations, the U.S. has ranked poorly on cost and outcomes. This is predominantly because of our inability to effectively manage chronic disease. And yet the same Milken analysis estimates that modest reductions in unhealthy behaviors could prevent or delay 40 million cases of chronic illness per year [11]. If we learn how to effectively manage chronic conditions, thus avoiding hospitalizations and serious complications, the healthcare system can improve quality of life for patients and greatly reduce the ballooning cost burden we all share [10].

The success of population health and chronic disease management efforts hinges on a few key elements: identifying those at risk, having access to the right data about this population, creating actionable insights about patients, and coaching them toward healthier choices. Methods such as data-driven visual analytics help experts analyze large amounts of data and gain insights for making informed decisions regarding chronic diseases [10,20]. According to the U.S.-based Institute of Medicine and the National Research, the vision for 21st century healthcare includes increased attention to cognitive support in decision making [21]. This encompasses computer-based tools and techniques that aid comprehension and cognition. Visualization techniques offer cognitive support by offering mental models of the information through a visual interface [22]. They combine statistical methods and models with advanced interactive visualization methods to help mask the underlying complexity of large health data sets and make evidence-based decisions [23]. Chronic diseases are characterized by high prevalence among populations, rising complication rates, and increased incidence of people with multiple chronic conditions, to name a few. In this scenario, visualization can represent association between preventive measures and disease control, summary health dimensions across diverse patient populations and, timeline of disease prevalence across regions/populations, to offer actionable insights for effective population management and national development [24]. Additionally, visual techniques offer the ability to analyze data at multiple levels and dimensions starting from population to subpopulation to the individual [25]. This paper addresses the challenge of understanding large amounts of data related to chronic diseases by applying visual analytics techniques and producing descriptive analytics. Our overall goal is to gain insight into the data and make policy recommendations.

Given that large segments of the U.S. population suffer from one or more chronic disease conditions, a data-driven approach to the analysis of the data has the potential to reveal patterns of association, correlation, and causality. We therefore studied the variables extracted from a highly reliable source, the Centers for Disease Control. Data for variables pertaining to several categories, namely chronic condition ("condition" is used interchangeably with "disease"), behavioral health, mental health, preventive health, demographics, overarching conditions, and location for several years (typically 2012 to 2014). We analyzed relationships within each category and across categories to obtain multi-dimensional views and insight into the data. The analytics provide insights and implications that suggest ways for the healthcare system to better manage population health.

Main Problem Arises, there are some possibilities that we have a disease in our body but we are unaware of the disease. It could be any type of disease at any level of its phenomena. Whenever we are visiting a doctor, then a doctor can give you the best advice only if he knows accurate and full information of your past health history. Here our model has a unique feature where we can store a patient health history online and this health history can be sharable among all of the hospitals where the patient will be visiting for his/her treatment in future.

1.2.1 Tools and Technology Used:

1.2.1.1 Hardware Requirements: Processor RAM Disk Space Pentium II, Pentium III, Pentium IV or higher 64 Mb or Higher 130Mb, Minimum 8 Gb RAM.

1.2.1.2 Software Requirements: Operating System DatabaseWin-98, Win-XP, Linux or any other higher version MS Access.

1.2.1.3 Others Tools:

- Google Collab
- Anaconda
- Azure
- python.

Chapter-2

Literature Survey

Chronic diseases include heart disease, stroke, cancer, chronic respiratory diseases and diabetes. Visual impairment and blindness, hearing impairment and deafness, oral diseases and genetic disorders are other chronic conditions that account for a substantial portion of the global burden of disease.

From a projected total of 58 million deaths from all causes in 2005,1 it is estimated that chronic diseases will account for 35 million, which is double the number of deaths from all infectious diseases (including HIV/AIDS, tuberculosis and malaria), maternal and perinatal conditions, and nutritional deficiencies combined

We use visualization and descriptive analytics to explore chronic conditions, preventive healthcare, mental health, and overarching conditions, with the objective of deciphering relationships and patterns that emerge from the visualization. We would like to point out that since our sample includes adults aged 18 and over our results are applicable for adults in that age group.Figure 1 models the average prevalence of diagnosed diabetes among adults aged ≥18 years in the period 2012 to 2014. Puerto Rico leads the pack, followed by Mississippi.

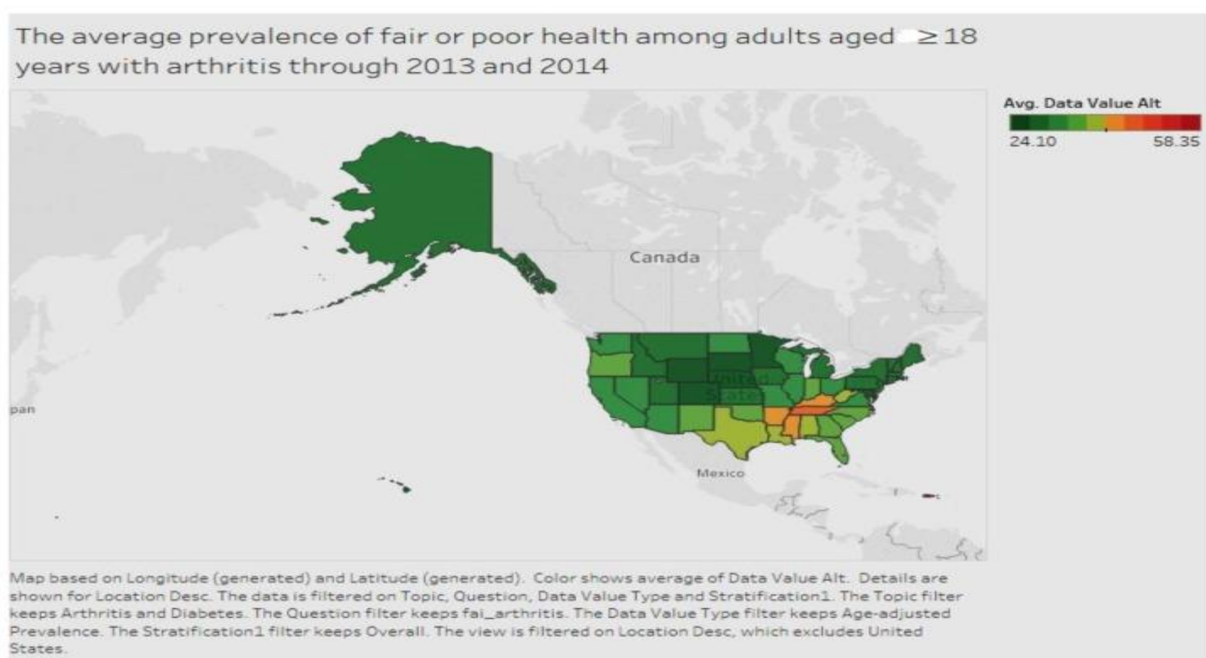


Figure 1 Diabetes by state

As Figure 2 below shows, Puerto Rico has the highest number of citizens among adults aged ≥18 years, in fair or poor health with arthritis for the period 2013 to 2014. Puerto Rico is followed by Tennessee and Mississippi.

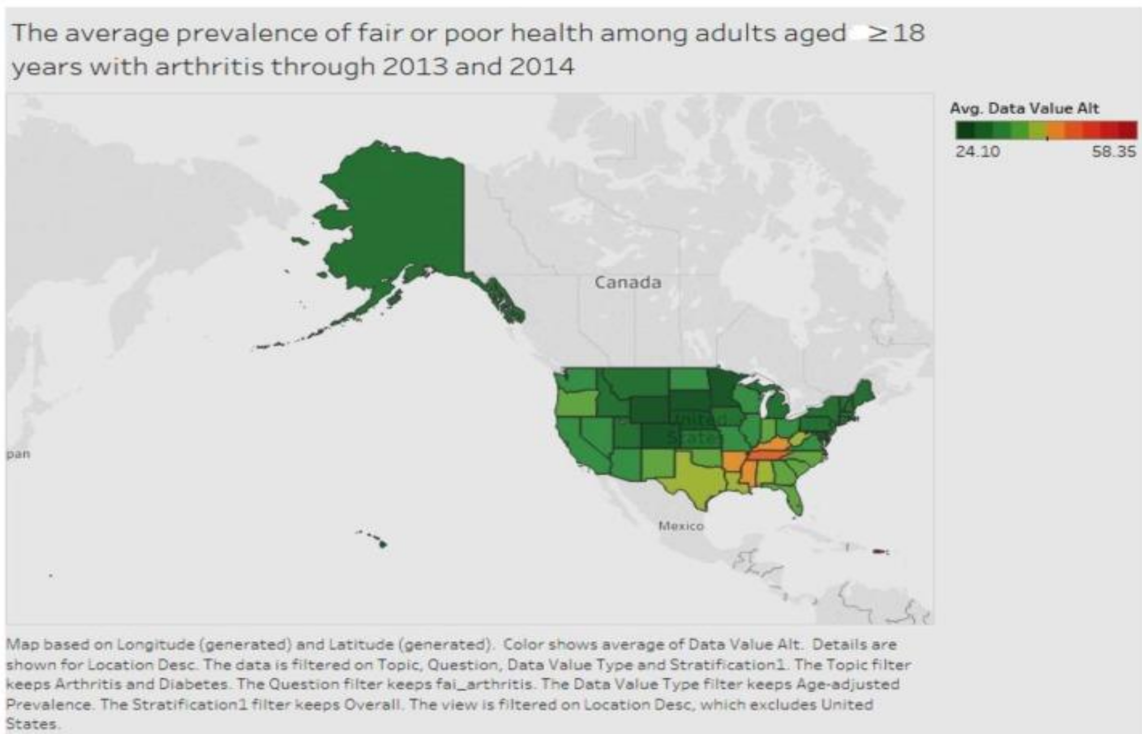


Figure 2 Arthritis by state.

In exploring chronic conditions by location in the U.S., we see that some conditions, such as diabetes, arthritis, and obstructive pulmonary diseases, are more prevalent in eastern states, while others, such as asthma, occur more often in northeastern states. For diabetes, listed as a cause of death for the years 2010 to 2014, the states of Oklahoma and West Virginia had the relatively high average threshold of over 100 (age adjusted rate per 100,000). In the case of asthma, West Virginia has the highest prevalence of the condition (among adults), while Maryland, Massachusetts, and New York had the highest number of hospitalizations. With regard to chronic obstructive pulmonary disease, Kentucky and West Virginia had the most hospitalizations compared to other states. The majority of states are indeed below 45 cases per 100,000. With respect to arthritis among adults, a majority of states average below 25%, with the exception of West Virginia, which averaged 34.15%. In summary, West Virginia ranks high in prevalence for most chronic conditions, such as diabetes, asthma, chronic pulmonary disease, and arthritis when compared to all other states for the period 2000 to 2014.

We looked at the distribution of chronic conditions by gender and race to identify relevant trends and patterns (Figure 3 and Figure 4). Chronic conditions differ by gender. Women tend to have significantly higher cases per 100,000 of hospitalizations for asthma. Whereas men tend to have a higher mortality rate from chronic obstructive pulmonary disease, diabetes, chronic kidney, and other conditions, as shown in Figure 3.

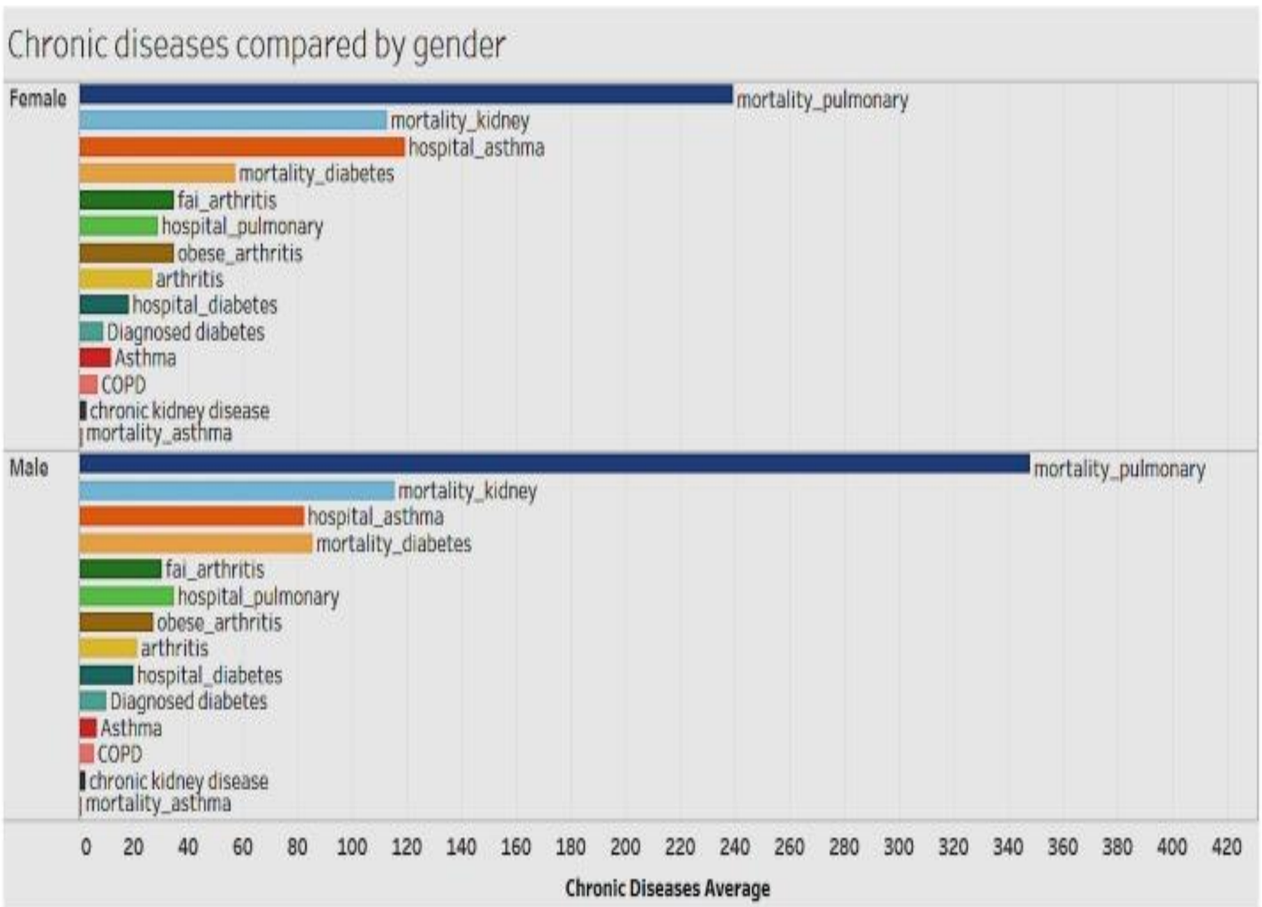


Figure 3 Chronic Condition by gender

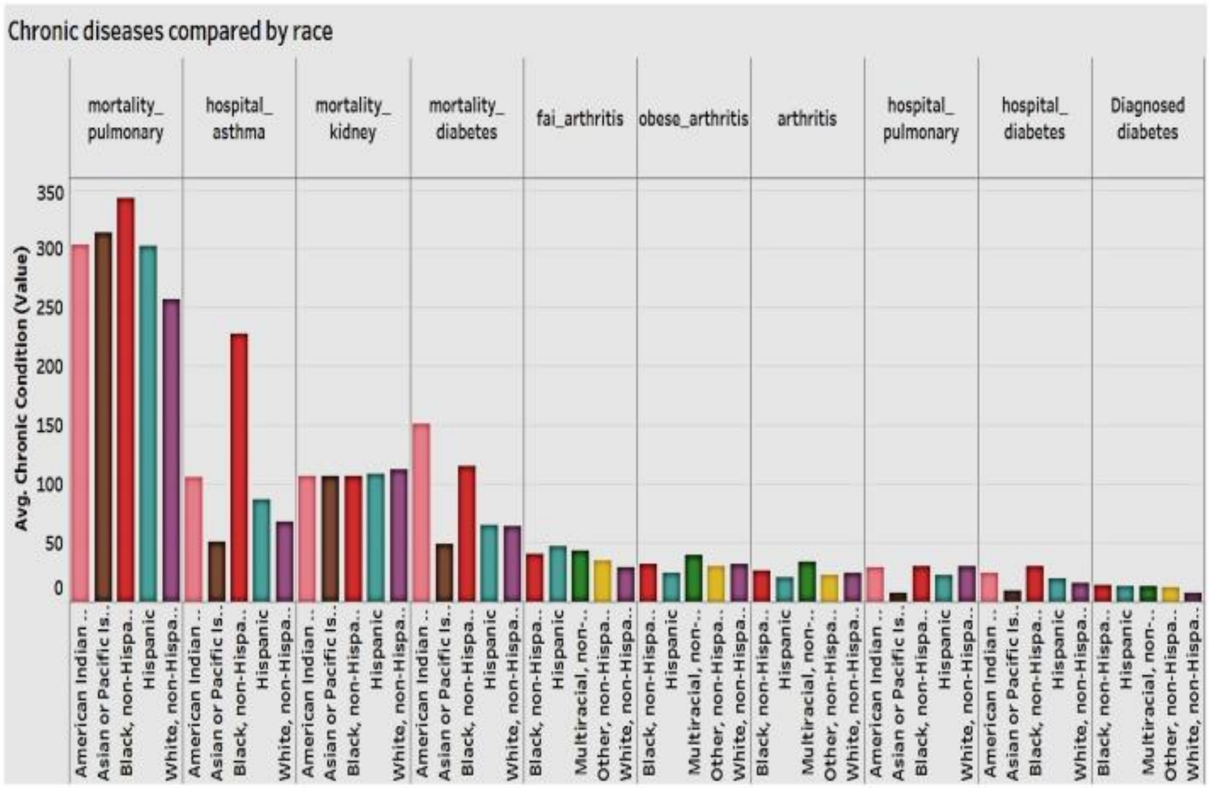


Figure 4 Chronic Condition by race

We also examined all chronic conditions by race (Figure 4) and found that non-Hispanic Blacks have higher mortality rate for pulmonary disease and asthma and a higher hospitalization from diabetes. They are followed by Pacific Islander and American Indians. All categories of arthritis are fairly evenly distributed among Black, non-Hispanic, Multiracial, Whites, and other.

Females have a higher hospitalization rate for asthma (per 100,000), while in terms of mortality rate for chronic obstructive pulmonary disease, diabetes, and chronic kidney disease, males have the higher hospitalization rate. Again, American Indian or Alaskan Natives have higher mortality rate for chronic obstructive pulmonary disease, diabetes, and kidney disease. They're followed by Blacks and non-Hispanics.

Mental Health by Gender and Race

Mental health is an important aspect of national healthcare impacting chronic diseases. We analyzed mental health by gender (Figure 5) and by race (Figure 6). When we examine how many days an individual feels “mentally unhealthy" for the years 2012 to 2014, women are more likely to have more unhealthy days than men, as shown in Figure 5.

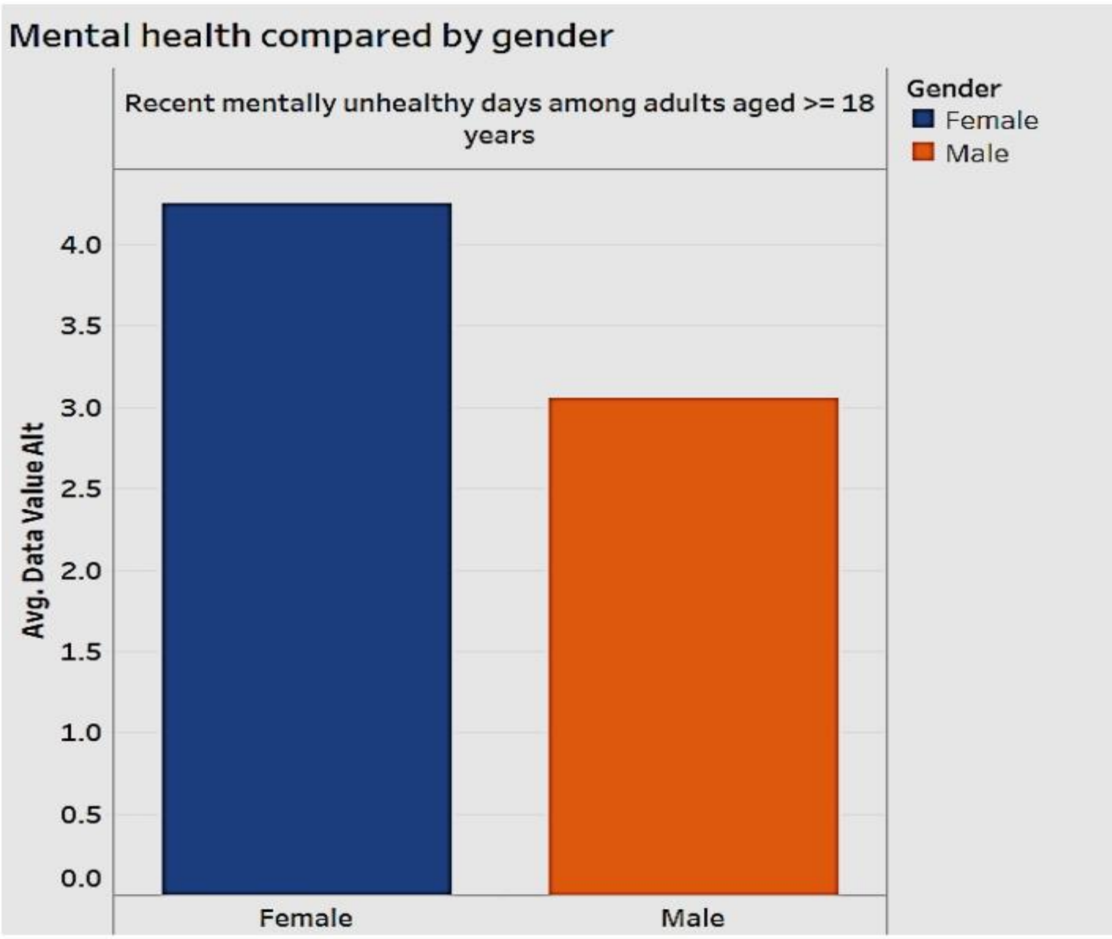


Figure 5 Mental Health by gender

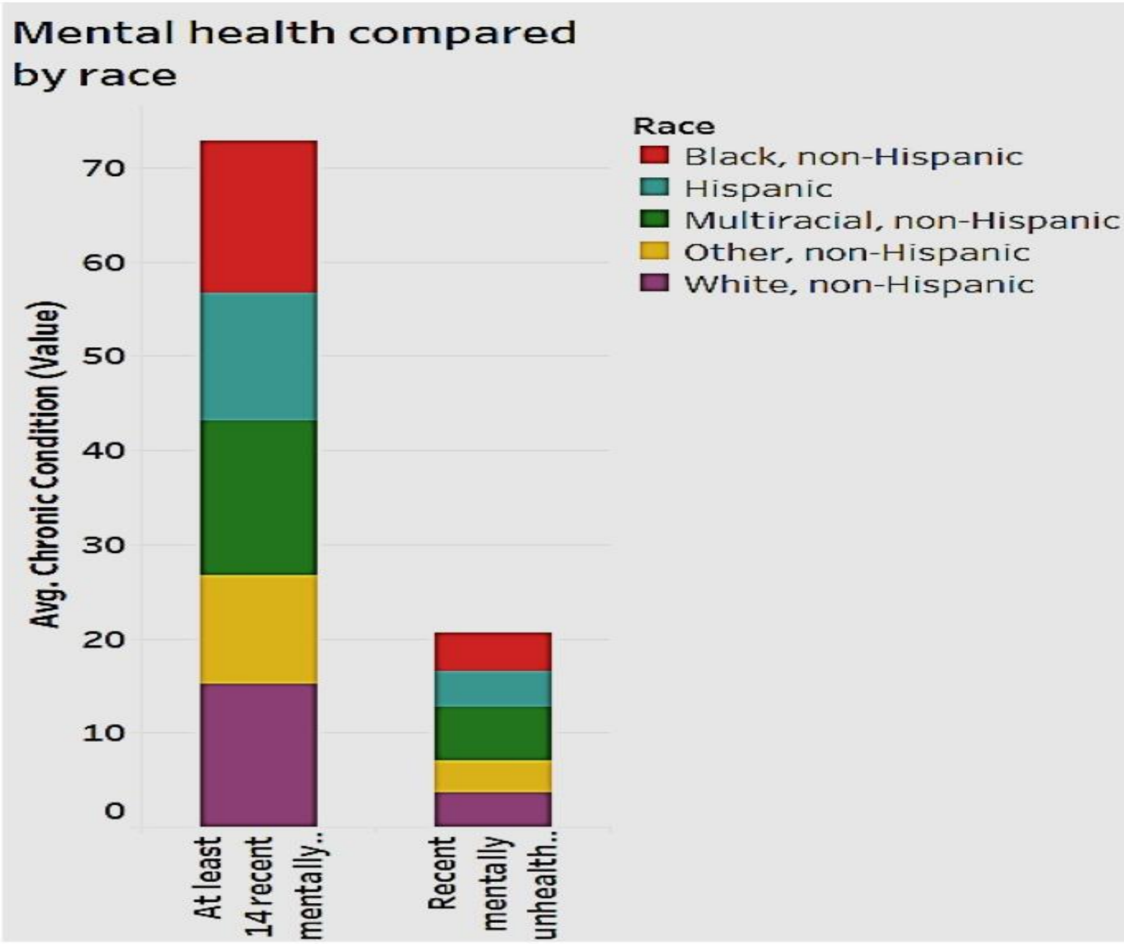


Figure 6 Mental Health by Race

Simultaneously, multi-racial, non-Hispanic women in the age group 18 to 44 have a higher crude prevalence rate of at least 14 recent "mentally unhealthy" days. This group is followed by black non-Hispanics.

We then studied behavioral habits in the data set to gain insight into noticeable patterns, if in fact any exist.

Behavioral Habits by Gender and Race

Figure 7 charts behavioral habits by gender.

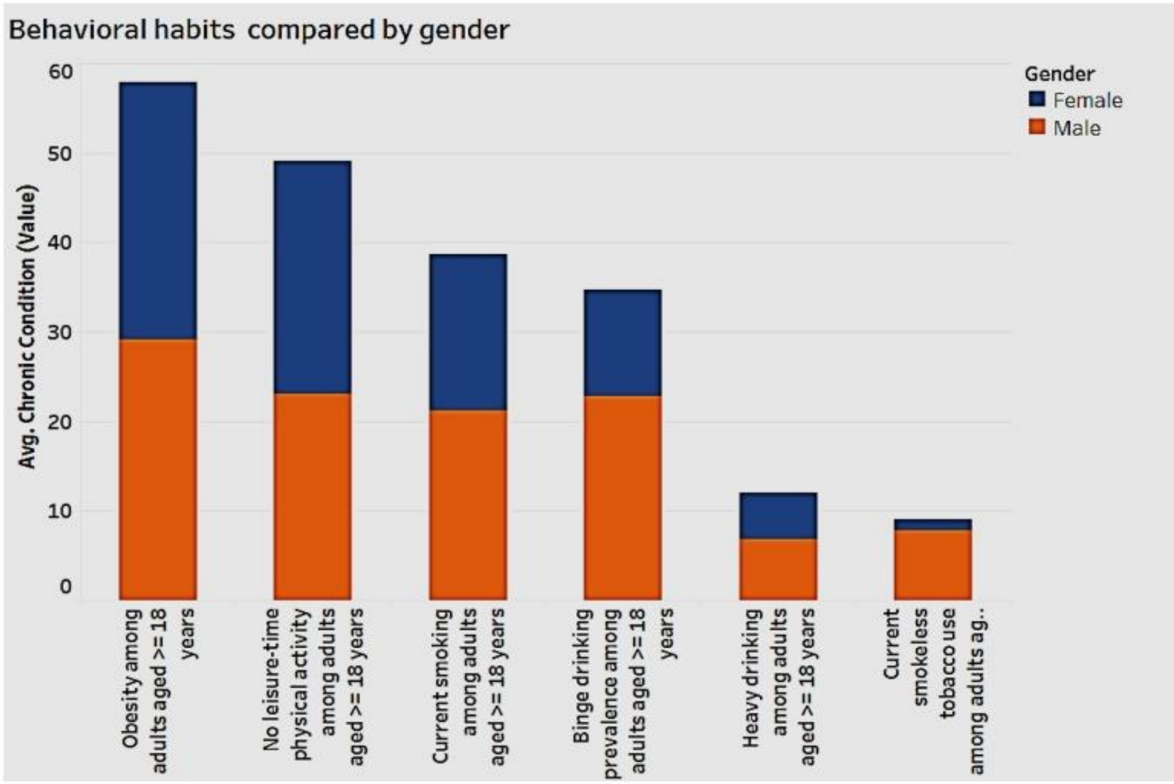


Figure 7 Behavioral Habits by Gender

As seen in the chart above, men display higher numbers in the alcohol categories of "binge drinking" and "heavy drinking", as well as in "current smokeless" tobacco use among adults. In terms of engaging in "current smoking", "obesity", and "no leisure-time" physical activity, both men and women experience similar complications, that highlights the need for positive behavior modification.

Figure 8 illustrates the analysis of behavioral habits by race.

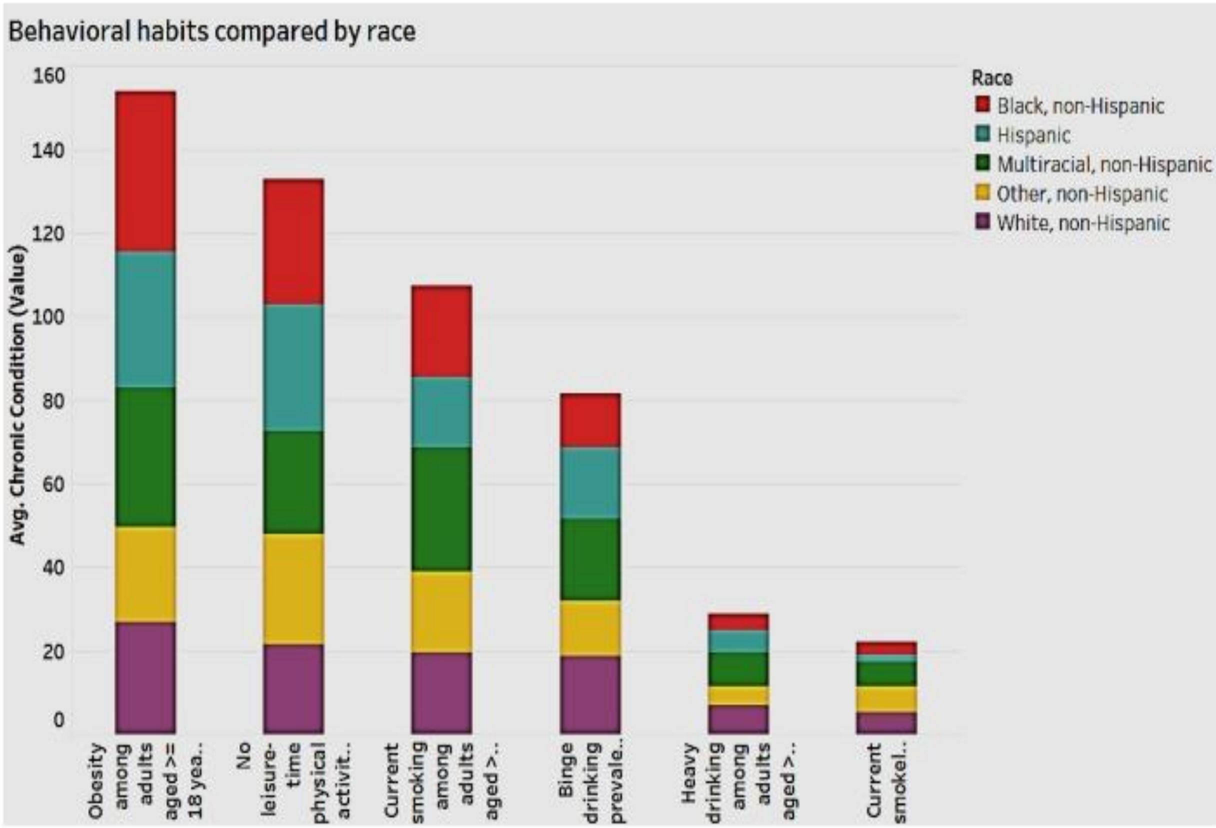


Figure 8 Behavioral Habits by race

Figure 8 reveals that for the behavioral habits of "obesity" and "no leisure-time" physical activity among adults aged 18 and over, the black non-Hispanic and Hispanic races have the highest frequency, while white non-Hispanics have the lowest. By and large, in most behavioral habits, the other non-Hispanics have the lowest frequency.

Preventive Health and Chronic Conditions

We analyzed the data to detect associations between demographics and preventive health. As Figure 9 indicates, both men and women appear to engage in preventive health, though women have the edge. With regard to race, Blacks and Hispanics engage less in preventive health overall, as shown in Figure 10.

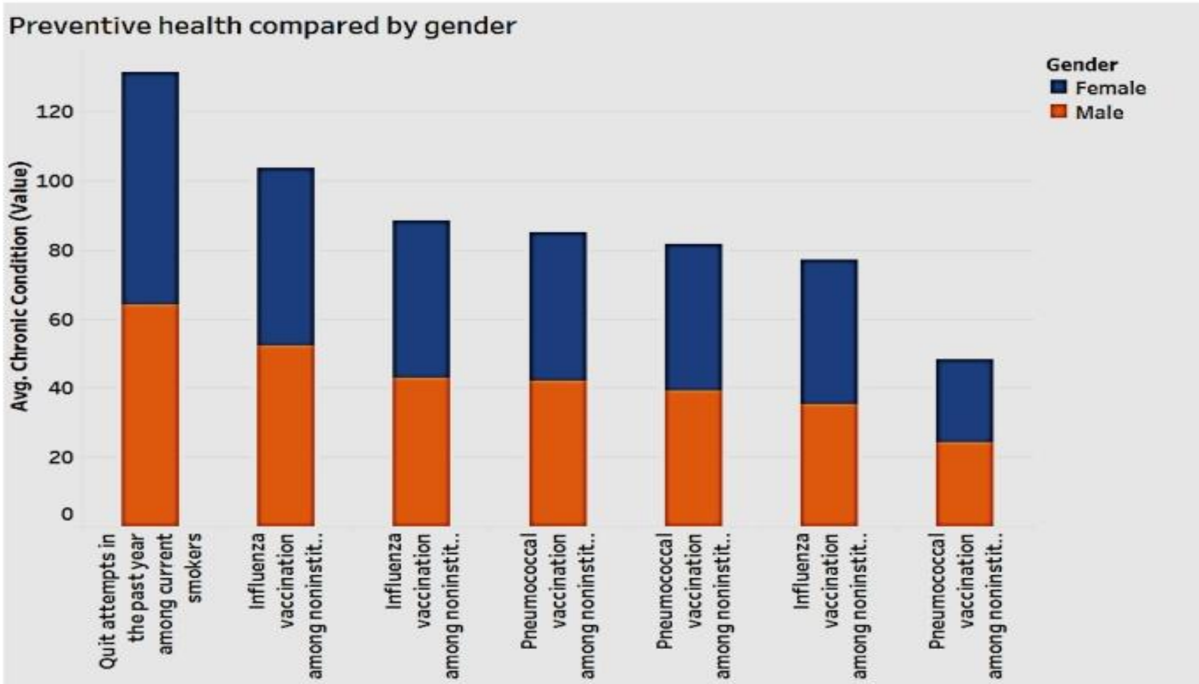


Figure 9 Preventive health by gender.

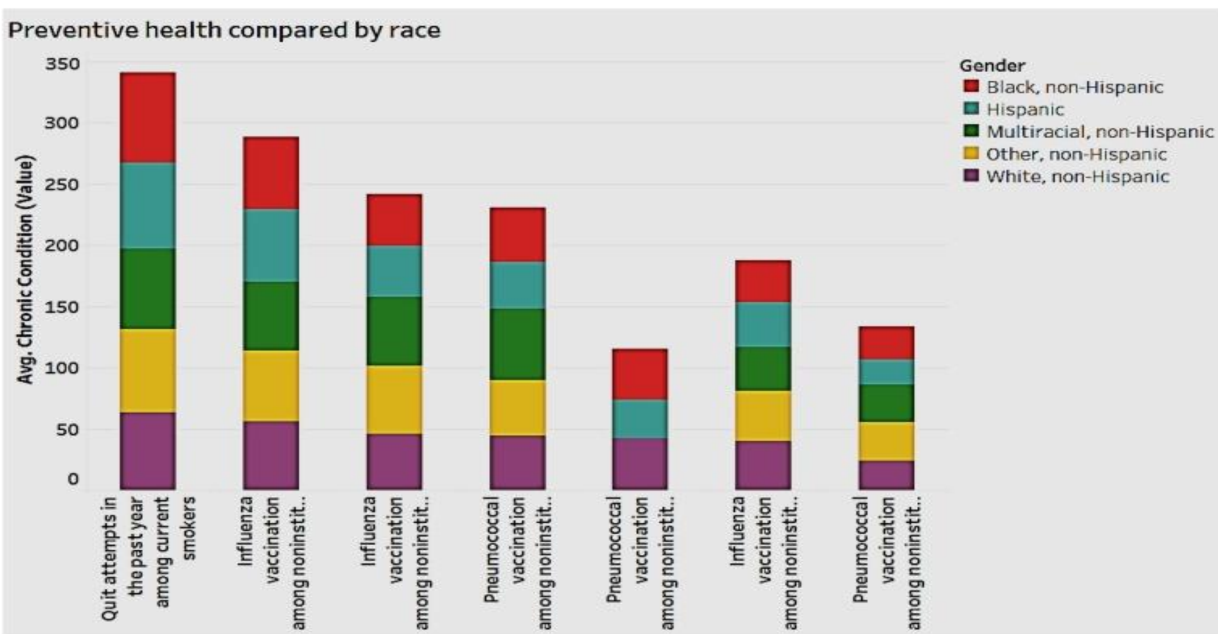


Figure 10 Preventive health by race.

While all chronic conditions are debilitating on the economy, for the sake of scope, we selectively analyze the influence of a few conditions such as diabetes and asthma. By 2034, the population with diabetes is expected to increase by 100% and the cost expected to increase by 53% [33]. Figure 11 depicts the association between diabetes and pneumococcal vaccination for diabetes.

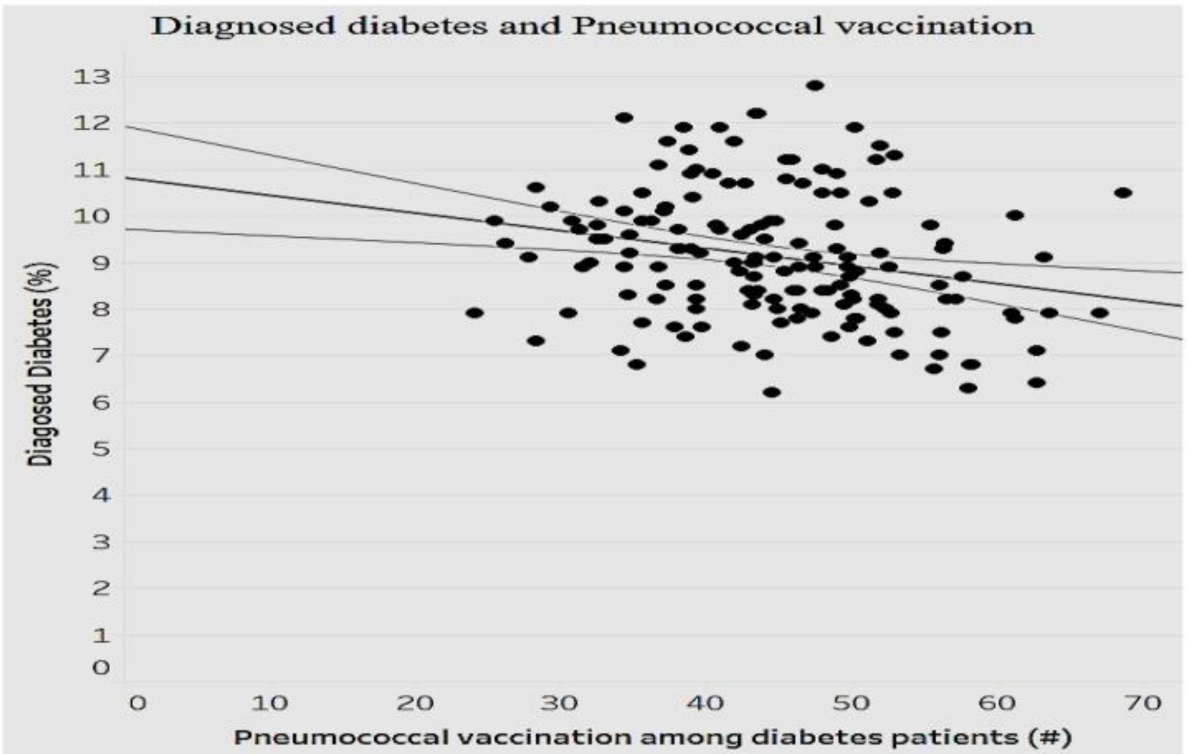


Figure 11 Diagnosed diabetes ratio by pneumococcal vaccination ratio. (#: number).

As indicated in Figure 11, there is a significant negative relationship between the average pneumococcal vaccination among diabetes patients and the average diagnosed diabetes ratio among the population ($p < 0.0001$). As the average pneumococcal vaccination among diabetes patients increases, the average diagnosed diabetes ratio decreases (fewer cases of diabetes). Given the importance of asthma as another prevalent chronic condition, we decided to analyze the relationship between the mortality ratio and influenza vaccinations for asthma to determine the efficiency of preventive measures (Figure 12).

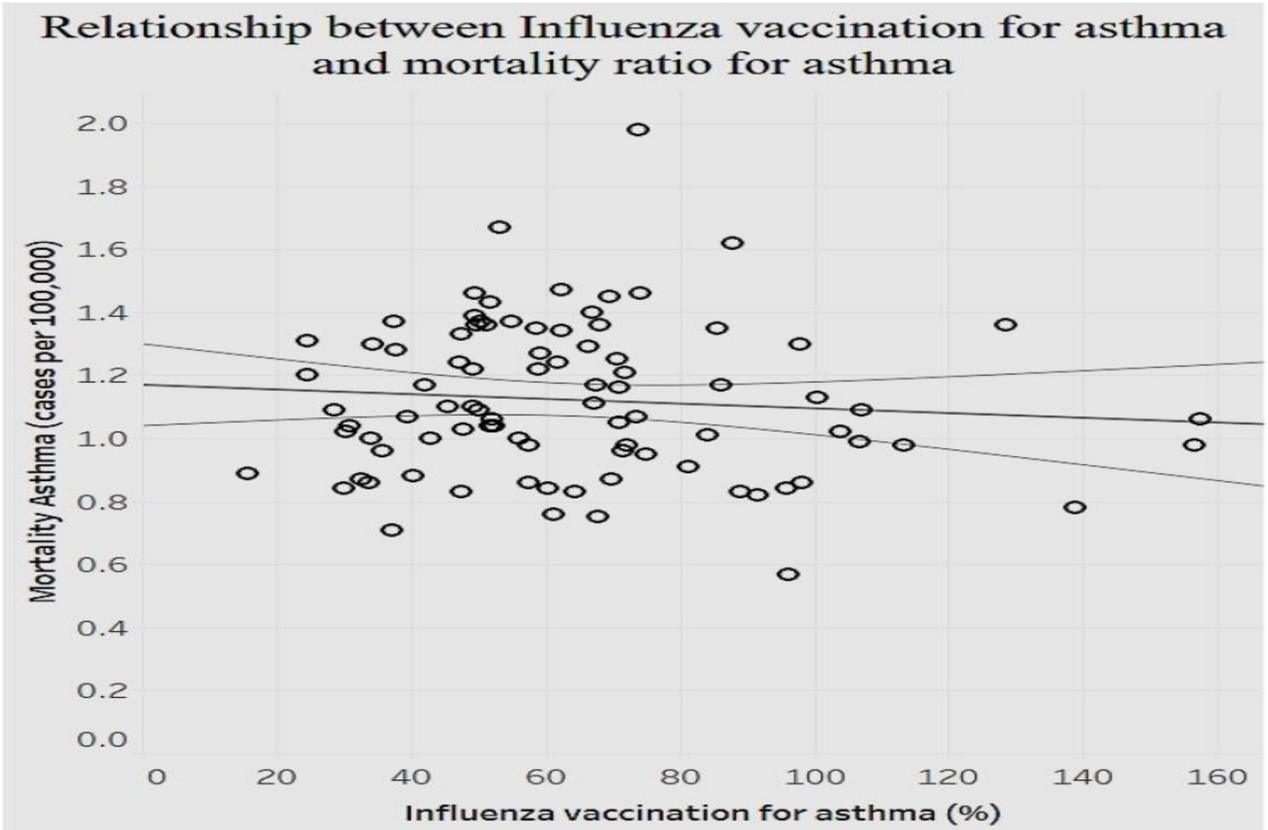


Figure 12 Mortality ratio for asthma and influenza vaccination for asthma.

Figure 12 shows a significant negative association ($p < 0.0001$): as the rate of influenza vaccination for asthma increases, the mortality ratio of asthma declines. Analysis of the above preventive health variables shows that resources and efforts dedicated to preventive healthcare offer promise. The importance of managing chronic diseases is also highlighted when we examine the association between behavioral habits and overarching conditions.

Behavioral Health and Overarching Conditions

Overarching conditions represent situations or factors that directly or indirectly influence the area of study. In our research we look at the influence of these conditions on chronic diseases, behavioral health, and preventive health. The overarching conditions include lack of health insurance (%), self-rated health status (good, fair, poor), and prevalence of sufficient sleep (%) for which data was available.

We explored the association of self-assessed health statuses among adults with the behavioral habits of binge drinking and heavy drinking (Figure 13).

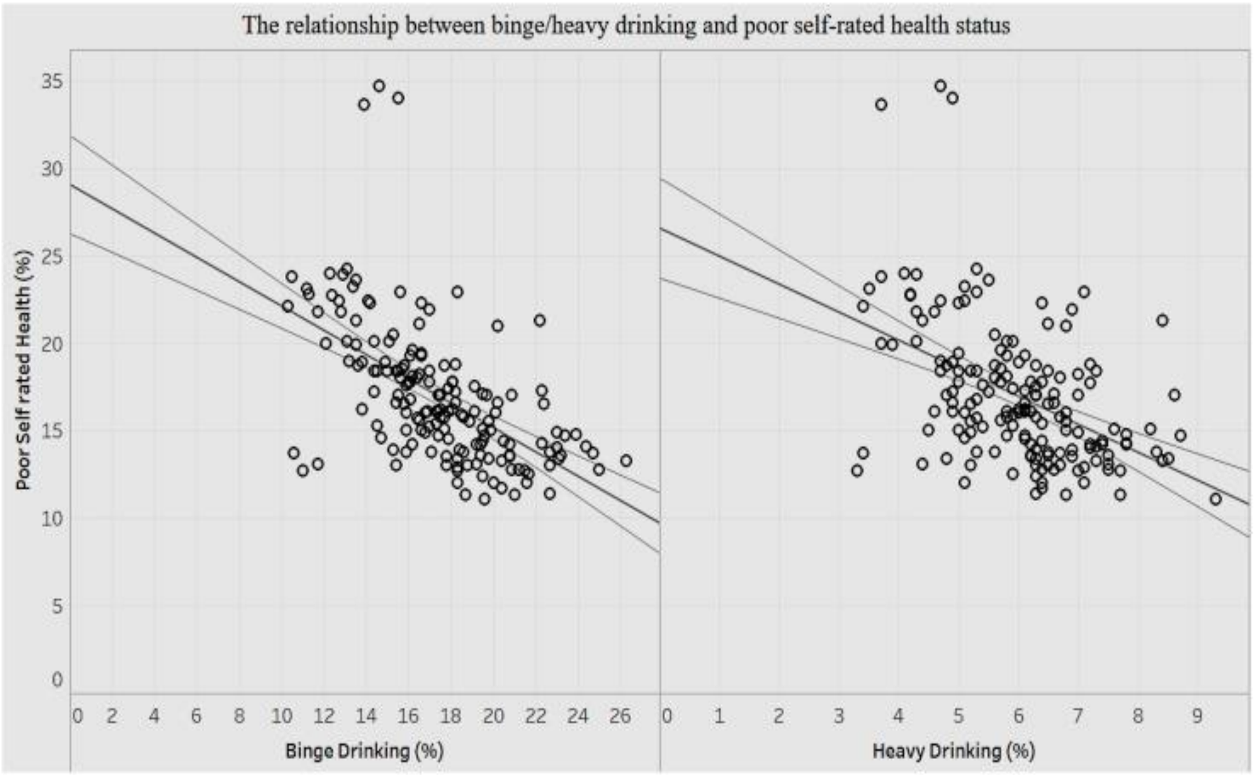


Figure 13 Binge/heavy drink and poor self-rated health status.

There is a significant negative correlation ($p < 0.0001$) between binge drinking and self-assessment of health. That is to say that the lower the health self-assessment, the higher is the percentage of binge drinking. A decrease of less than 1% (0.69%) in self-assessed health is associated with a 1% increase in binge drinking. Likewise, there is a significant negative association ($p < 0.0001$) between self-assessment of health and percentage of heavy drinking. A decrease of 1.6% in self-assessed health is associated with a 1% increase in heavy drinking. We can surmise that reduced self-assessment of health has a stronger influence on heavy drinking than binge drinking among adults.

Next, we looked at the association between current smoking prevalence and presence of sufficient sleep among adults (Figure 14).

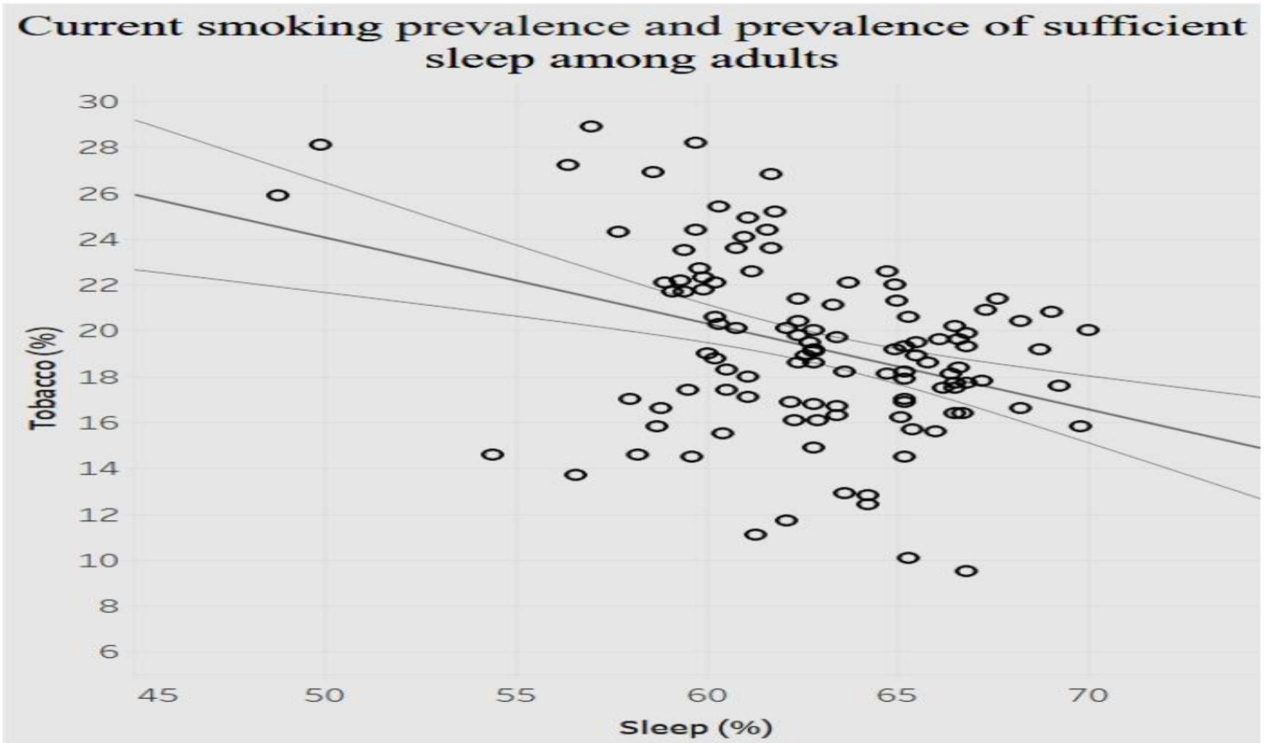


Figure 14 Current smoking prevalence by presence of sufficient sleep among adults.

Figure 14 above shows a significant negative association ($p < 0.0001$) between prevalence of current smoking and prevalence of sufficient sleep. When current smoking prevalence decreases by less than 1% (0.38%), the prevalence of sufficient sleep increases by 1%.

The relationship between poor self-rated health status and obesity is positive (Figure 18). The higher the prevalence of fair or poor self-rated health, the higher is the prevalence of obesity. When poor self-rated health increases by 1%, the prevalence of obesity increases by 0.468779%.

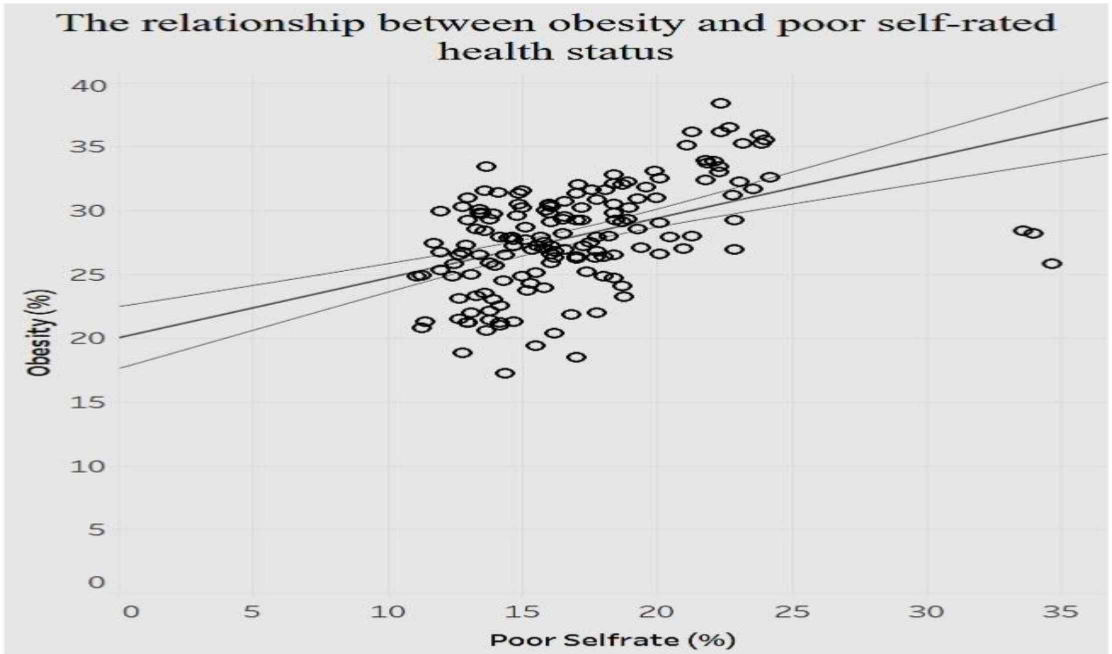


Figure 15 Obesity by poor self-rated health status.

Similarly, poor self-rated health has a positive association with current smoking, as indicated in Figure 16. As the prevalence of poor self-rated health increases by 1%, the prevalence of current smoking increases by 0.30425%.

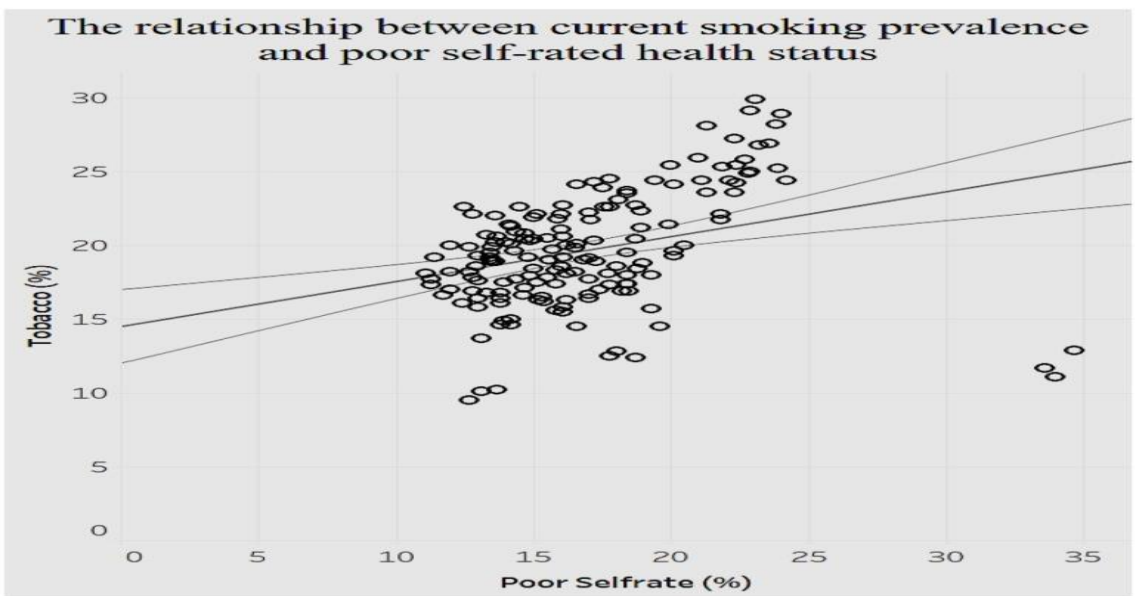


Figure 16 Smoking by self-rated health status.

Chronic Conditions and Overarching Conditions

In the analysis of various chronic conditions, there are significant clusters of conditions among men and women, such as the prevalence of asthma, with the women tending to have a higher prevalence of asthma than men. Regarding such chronic conditions as diabetes, there is a significant positive relationship ($p < 0.001$) between lack of health insurance and prevalence of diagnosed diabetes (Figure 17).

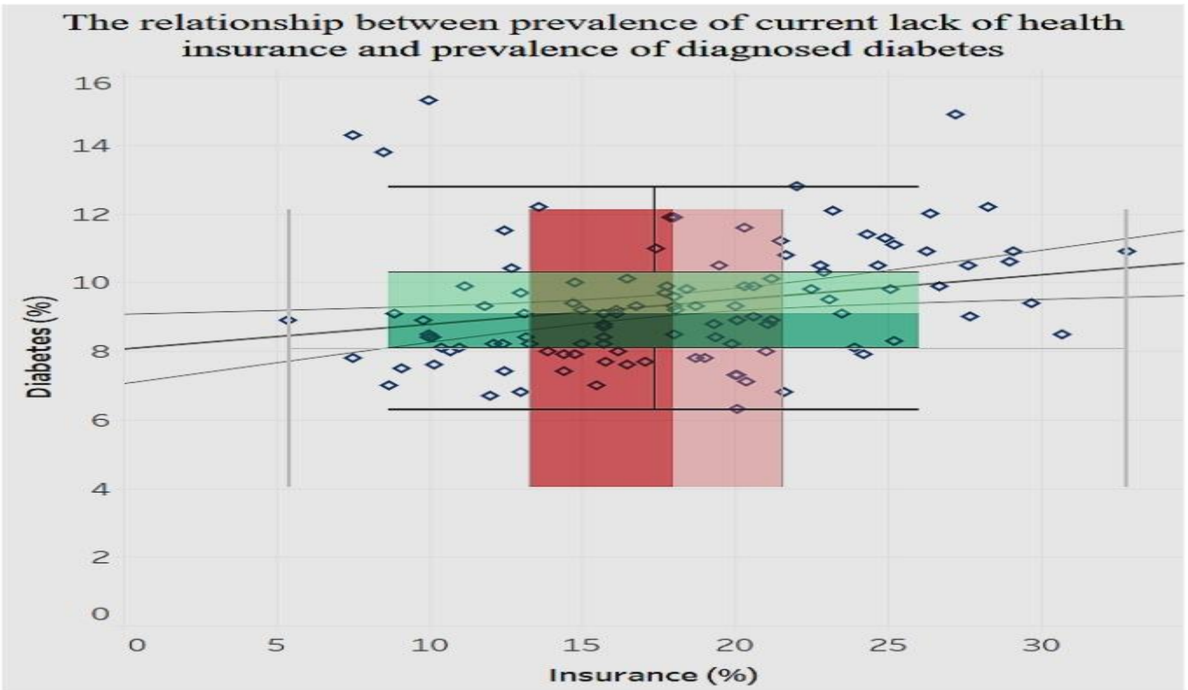


Figure 17 Current lack of health insurance by diagnosed diabetes.

We notice in Figure 17 that the distribution of lack of health insurance is sparse compared to that of diagnosed diabetes among adults aged 18 and older. Likewise, for chronic kidney disease (Figure 18) there is a significant positive relationship ($p < 0.0001$) with lack of health insurance**.**

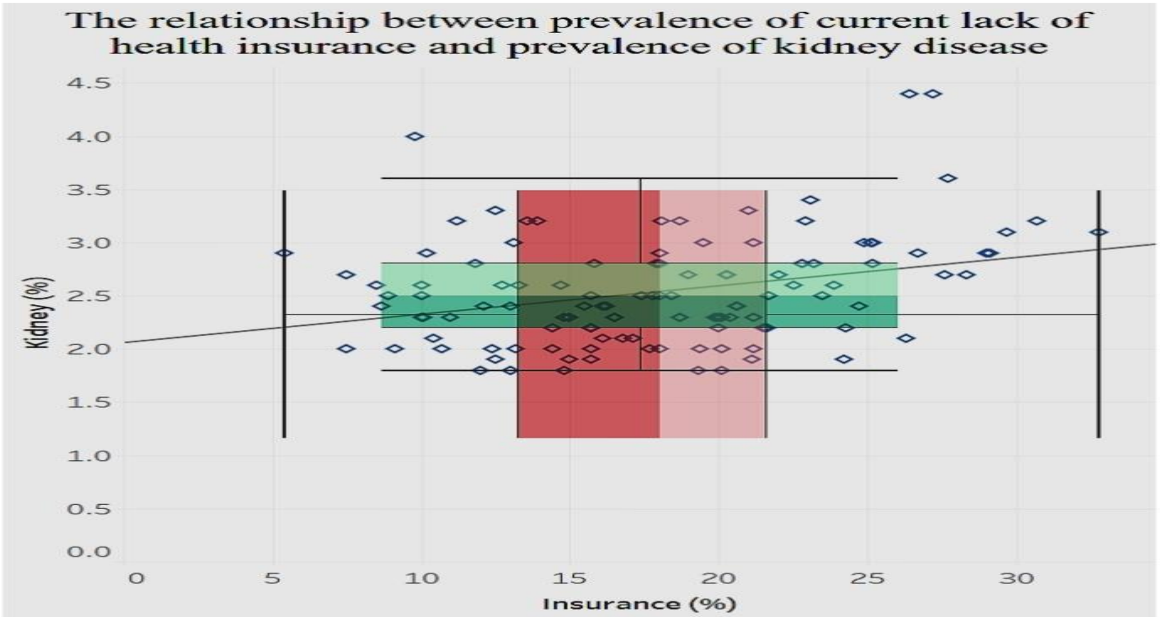


Figure 18 Lack of insurance by chronic kidney disease.

The relationship between lack of insurance and hospitalization for chronic pulmonary disease is positive and significant ($p < 0.0001$), as shown in Figure 19. An increase in the lack of insurance is associated with an increase in hospitalization for chronic pulmonary disease.

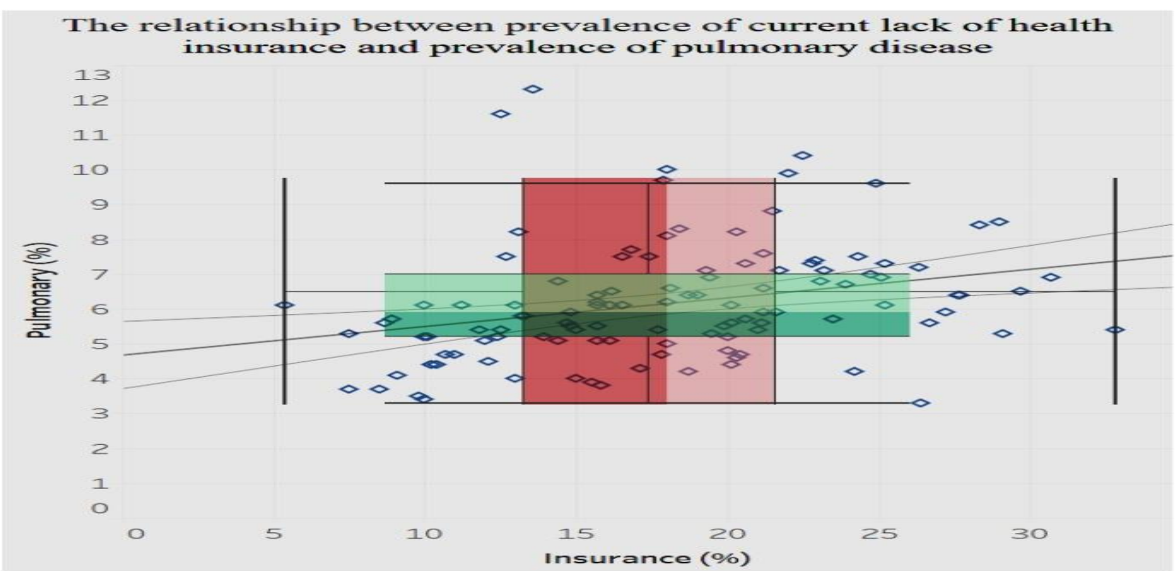


Figure 19 Lack of insurance by pulmonary disease.

Association between Chronic Conditions

We analyzed for any associations between different chronic conditions. It is important to incorporate gender as a factor in the association and prevalence of chronic diseases, so as to develop customized plans for diagnoses and treatments. A linear trend model was developed for the relationship between asthma and diabetes (Figure 20).

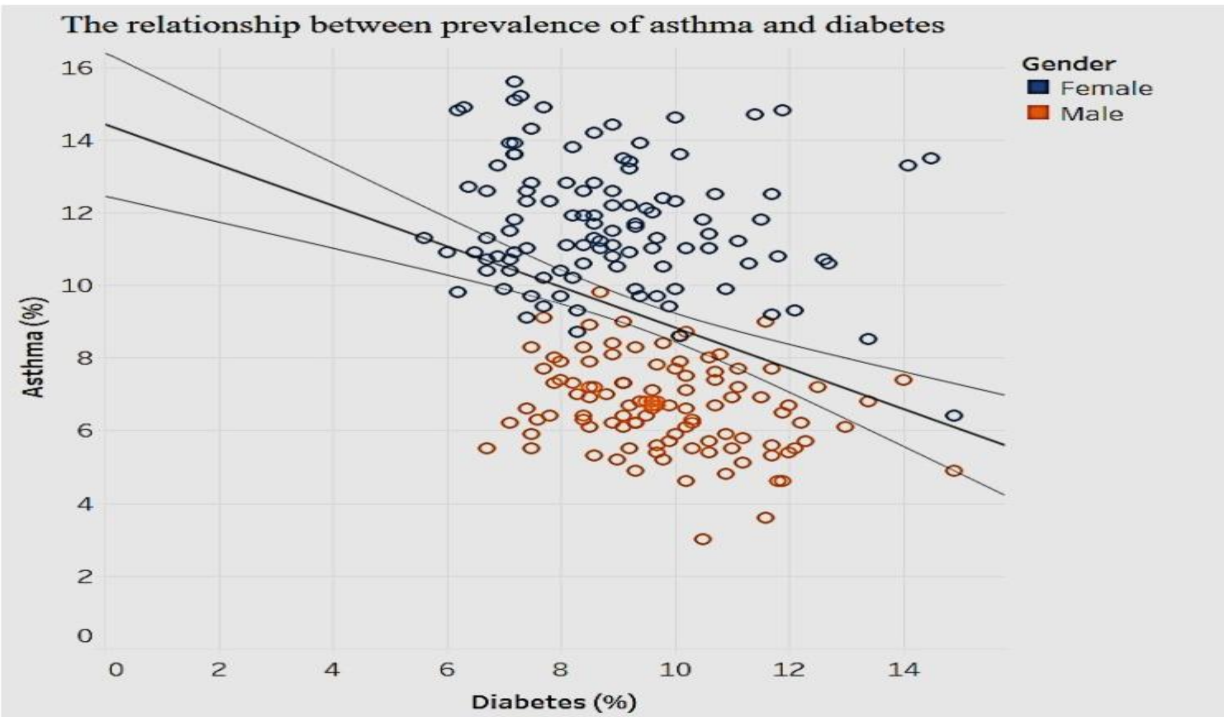


Figure 20 Asthma by diabetes.

The model in Figure 20 shows a significant negative relationship ($p < 0.01$) between asthma and diabetes. We can see gender clusters for the prevalence of asthma. Women tend to have higher prevalence of asthma compared to men. Overall, prevalence of asthma is negatively related to the prevalence of diabetes. On average, a high prevalence of asthma is associated with a low prevalence of diabetes. In terms of gender differences our results are consistent with other studies that have shown that women are more prone to develop asthma. Contributing factors include puberty, menstruation, pregnancy, menopause, and oral contraceptives [26,27]. There is potential for more research in this area.

The association between diabetes and kidney disease is shown in Figure 21.

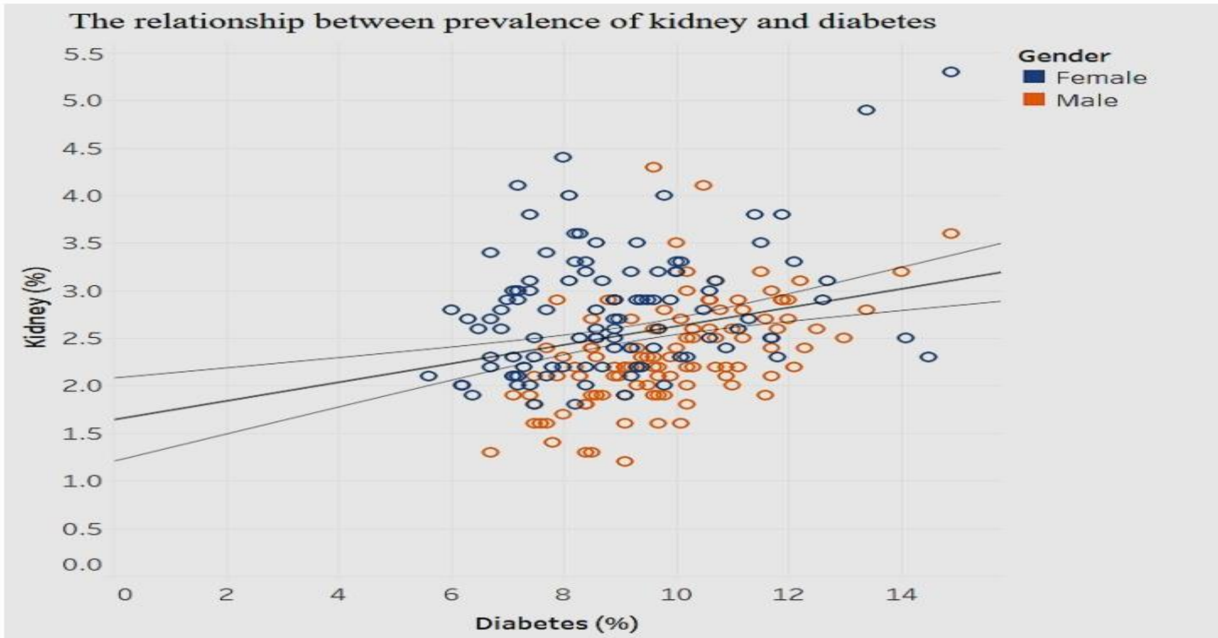


Figure 21 Diabetes by kidney disease.

Figure 21 shows a moderate, positive association ($p < 0.01$) between prevalence of kidney disease and diabetes. As the prevalence of diagnosed diabetes increases by 1%, the prevalence of chronic kidney disease increases by 0.09%. There are no obvious differences in gender here.

The association between diabetes and chronic pulmonary disease is shown in Figure 22, and that between arthritis and asthma is shown in Figure 23.

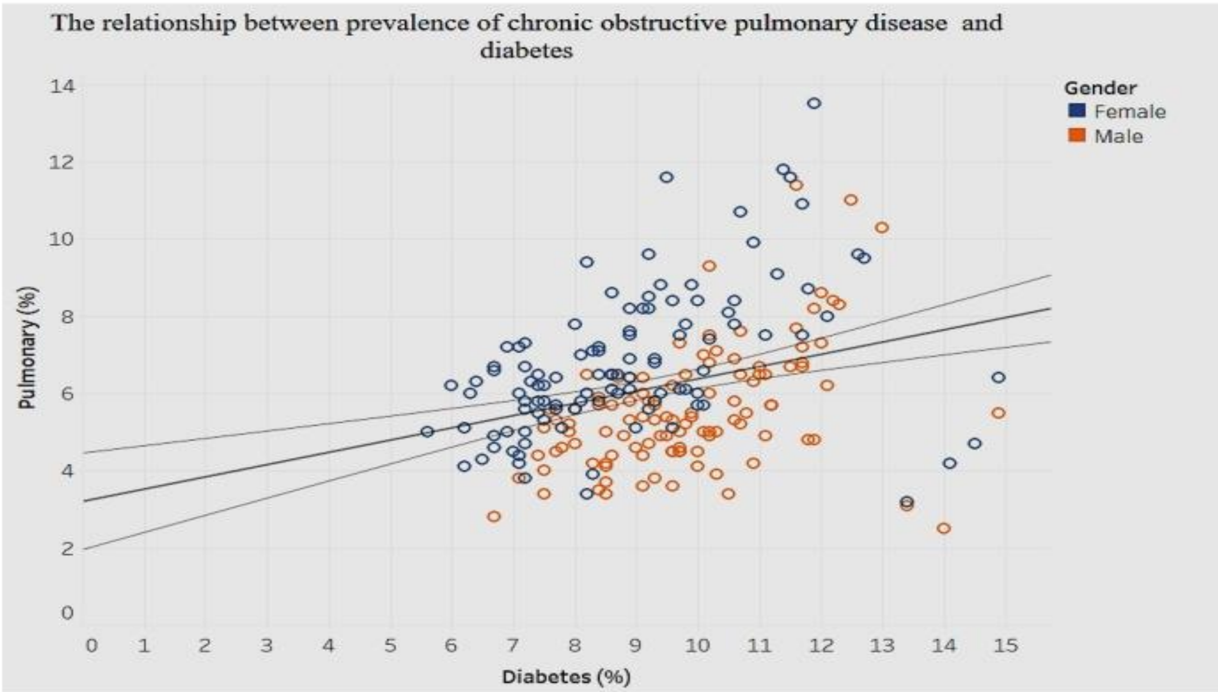


Figure 22 Diabetes by obstructive pulmonary disease.

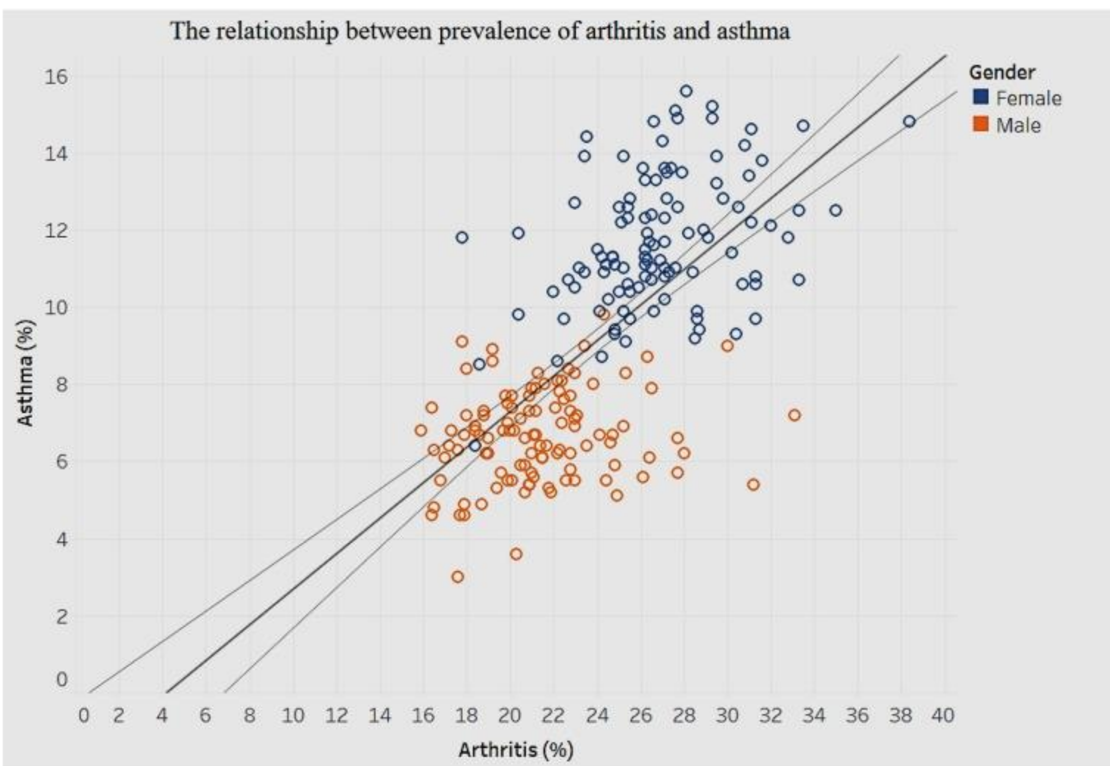


Figure 23 Arthritis by asthma.

Our aims to implement a robust machine learning model that can efficiently predict the disease of a human, based on the symptoms that he/she possess. Let us look into how we can approach this machine learning problem:

Approach:

Gathering the Data: Data preparation is the primary step for any machine learning problem. We will be using a dataset from Kaggle for this problem. This dataset consists of two CSV files one for training and one for testing. There is a total of 133 columns in the dataset out of which 132 columns represent the symptoms and the last column is the prognosis.

Cleaning the Data: Cleaning is the most important step in a machine learning project. The quality of our data determines the quality of our machine learning model. So it is always necessary to clean the data before feeding it to the model for training. In our dataset all the columns are numerical, the target column i.e. prognosis is a string type and is encoded to numerical form using a label encoder.

Model Building: After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We will be using this cleaned data to train the Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier. We will be using a confusion matrix to determine the quality of the models.

Inference: After training the three models we will be predicting the disease for the input symptoms by combining the predictions of all three models. This makes our overall prediction more robust and accurate.

A Disease is damage that affects the human system or one of its parts. As, for example, people are affected by bacteria, viruses, etc.…

As in today, we all are fighting with a threatening virus known to be Coronavirus. It is a contagious disease caused by the SARS-CoV-2 Virus. People infected with this virus are generally suffering from respiratory problems. People with ill medical history like cardiovascular disease, diabetes, chronic respiratory disease, or cancer developed many types of viruses. This virus is spreading in a huge amount irrespective of age group and a large amount of death can also be observed in this pandemic.

The disease could be of two types: Acute and Chronic. Acute diseases are those which occur suddenly and last for a short period like flu, fever, typhoid, etc. While Chronic diseases are those which occur for a prolonged time and last for a longer period, even a lifetime like Cancer, HIV, etc.

Chapter-3

Project Design

Data Flow Diagram:

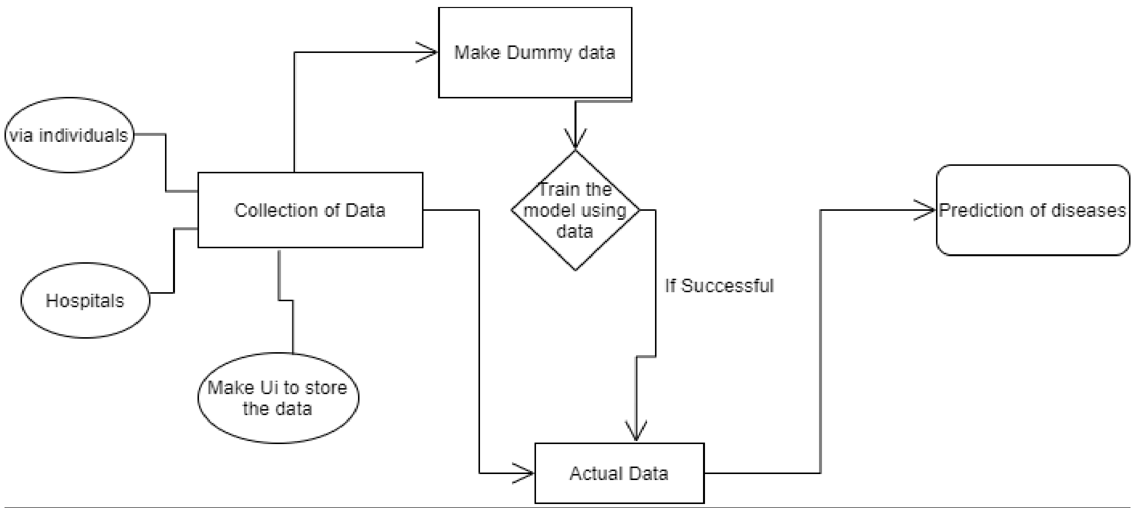
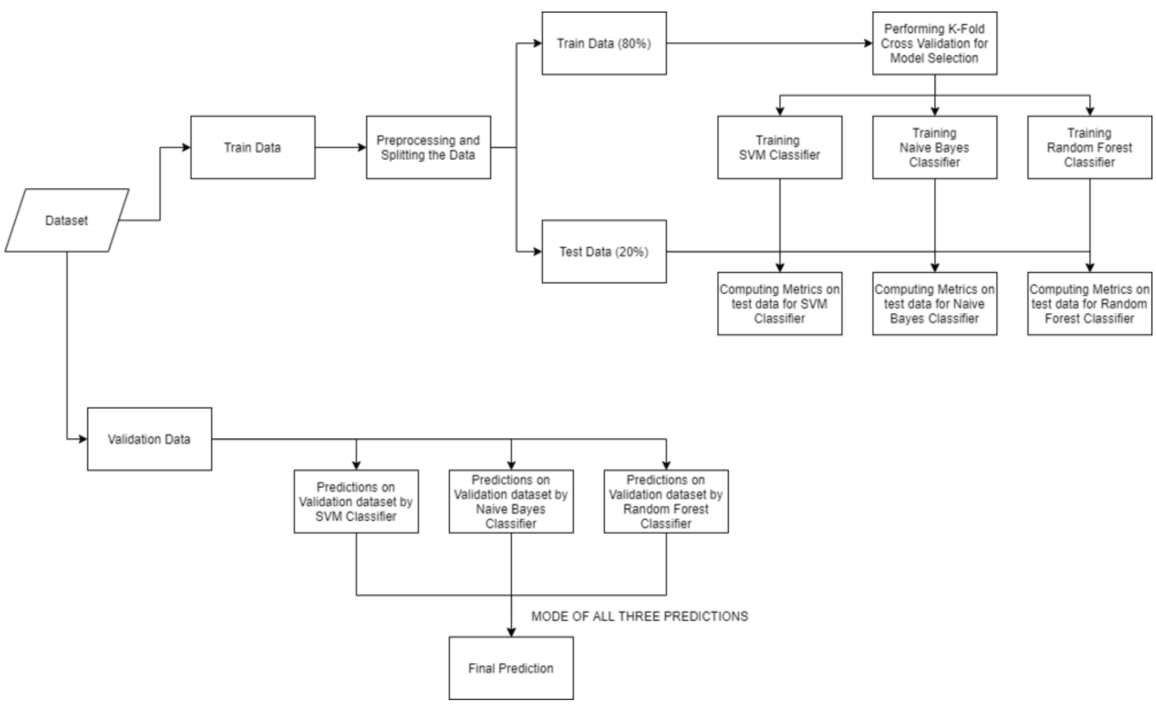


Fig [24]

UMLDiagram:



Fig[25]

Chapter-4

Modules Description

We have experimented on the real dataset taken through the Kaggle team for disease prediction and our survey is showing 100% accuracy experimented on different classifiers for both training and testing data.

4.1. First Strategy:

First, we will be talking about the workflow for our implementation to get a rough idea that how we are going to implement the project Fig [1].

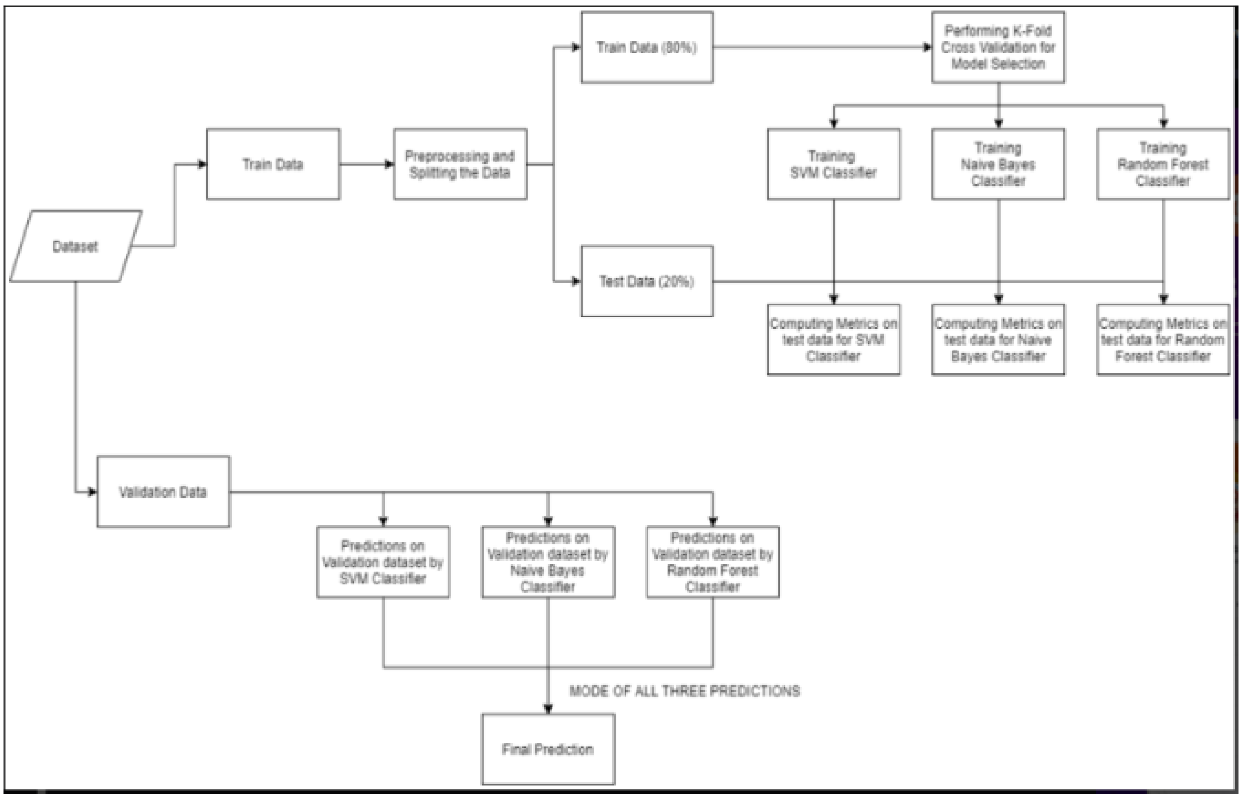


Fig [26]. Workflow

We have used the training and testing data for the prediction. Our main motive is to reading of the dataset, i.e., is to check whether our data is consisting null values. If it is so then we have to drop that null column and whole our dataset should be consisting of 0's and 1's values. Then we have balanced our dataset using a bar plot Fig [27].

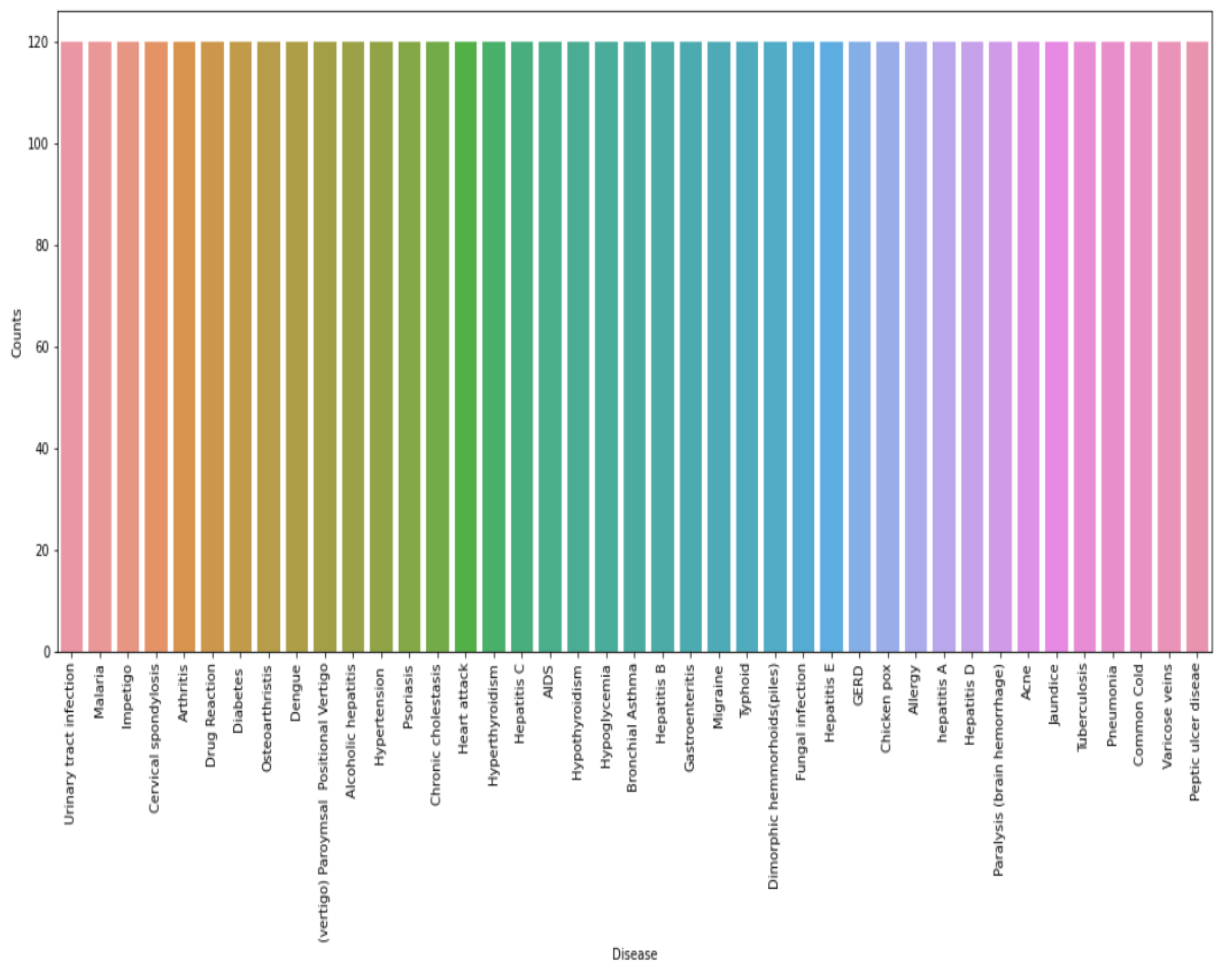


Fig [27]. Dataset Bar Plot

Then the dataset will be converted into the training and testing criteria using 80:20. In which 80% will be of training data and 20% is of evaluating the performance of the model i.e., testing data.

After splitting the data, we have gone to modeling part. We have used many classifiers like: K-fold Cross-Validation, Support vector Classifier, Gaussian Naïve Bayes Classifier, and Random Forest Classifier for the cross-validation method.

After experimenting with our data with K-fold Cross-Validation we have got accurate mean scores to be very high. Then again after experimenting we have got 100% accuracy in plot-confusion-matrix for all three classifiers. And then we have fitted the model on whole data and also validated that on the dataset that we have tested.

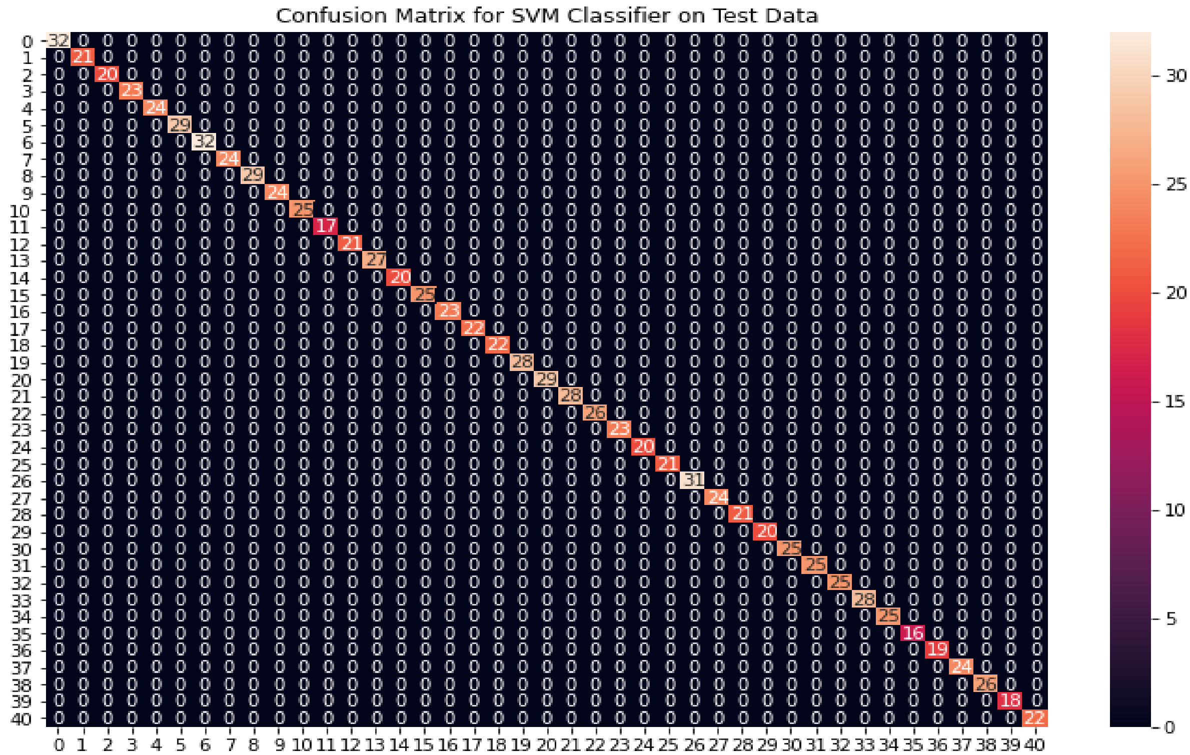


Fig [28]. Train data via SVM Classifier and get approx. 100.0 Accuracy

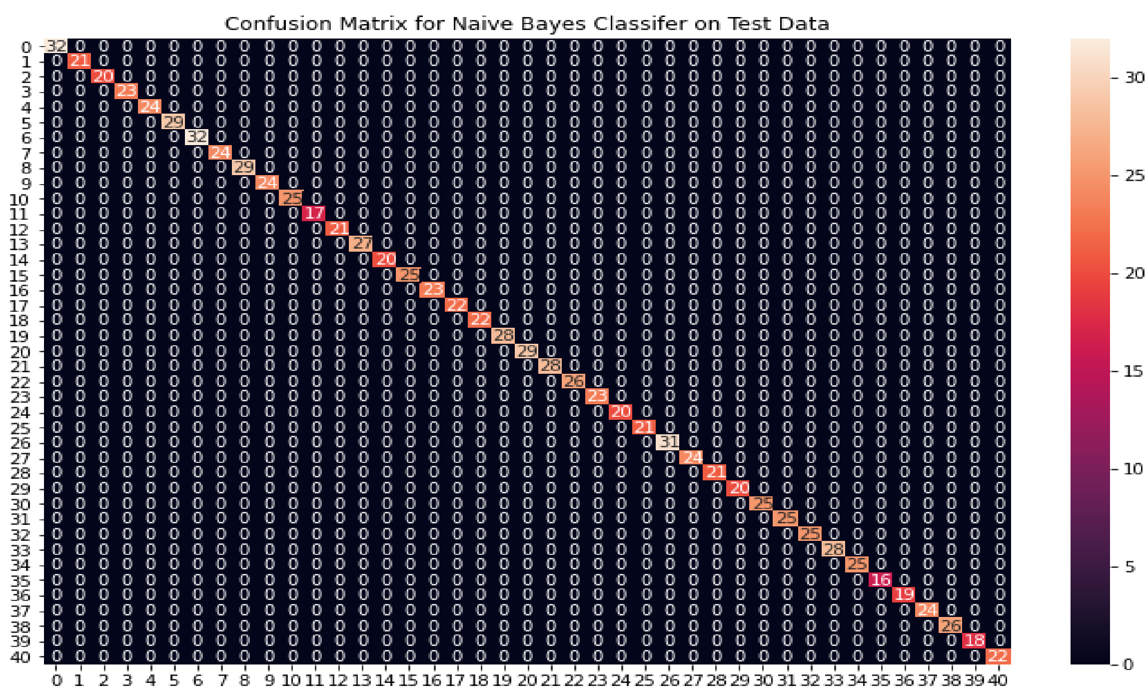


Fig [29]. Train data via Naïve Byes Classifier and get approx. 100.0 Accuracy

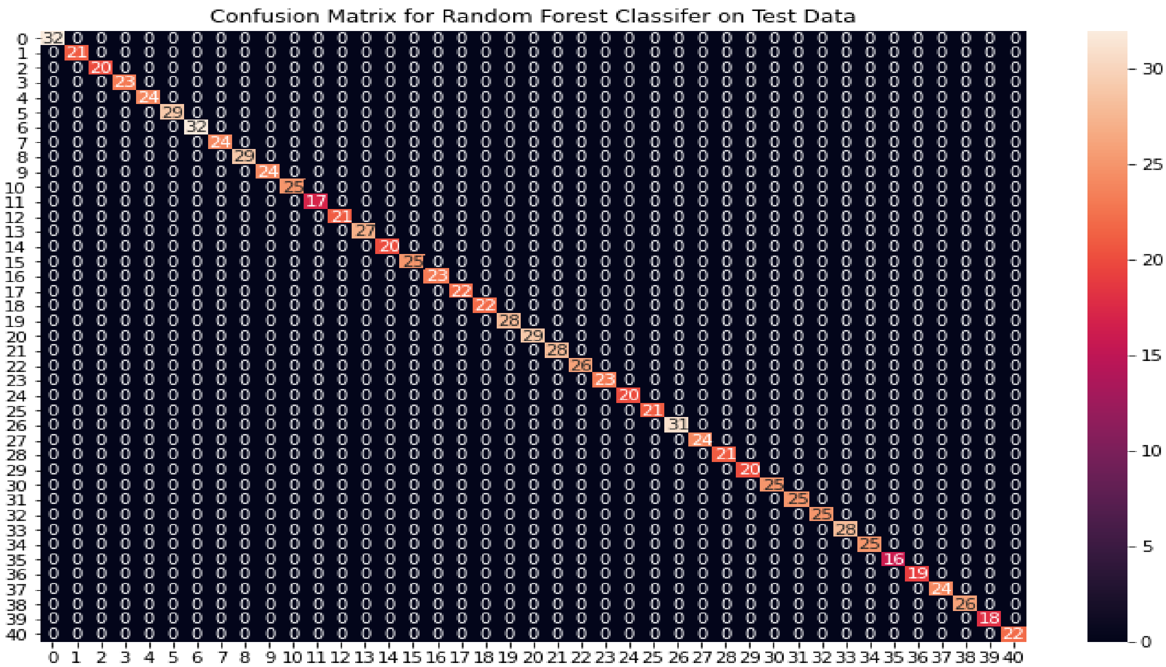


Fig [30]. Train data via Random Forest and get approx. 100.0 Accuracy

We had also combined all the classifier then again test then Accuracy will come:100.0 like Fig [5].

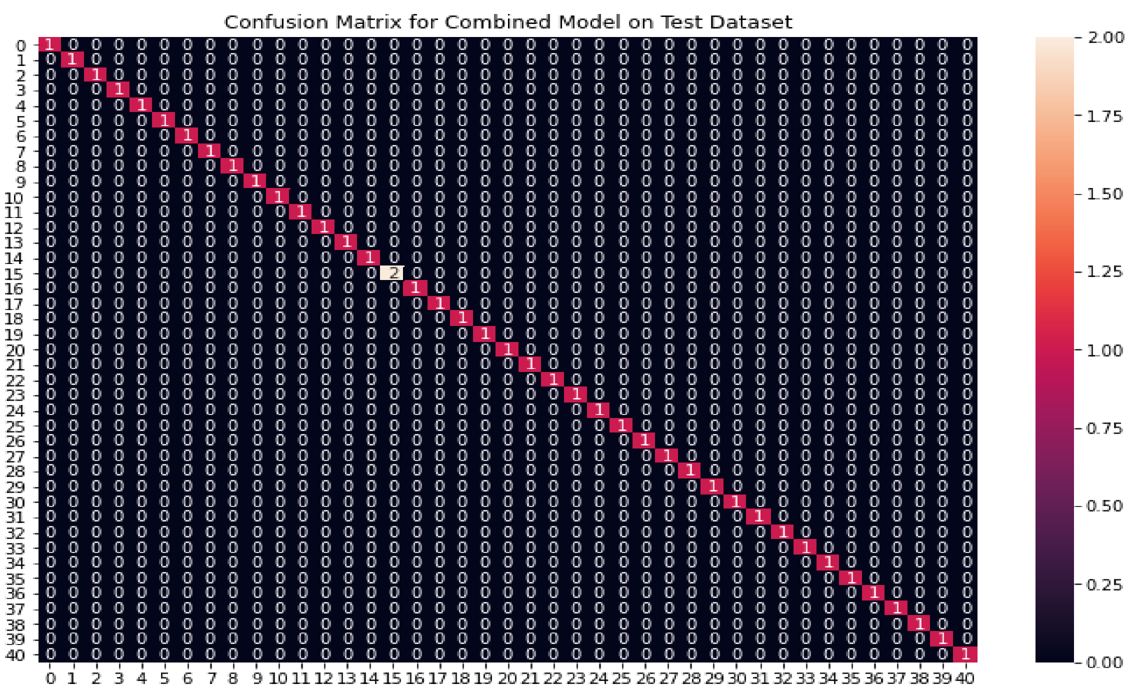


Fig [31]. Train data via combined classifier and get approx. 100.0 Accuracy

Chapter-5

Results

This study has analyzed chronic conditions in conjunction with several demographic variables, including gender and race. There are widespread variations in the prevalence of diverse chronic diseases, the number of hospitalizations for specific diseases, and the diagnosis and mortality rates for different states. For some chronic diseases—such as diabetes, arthritis, and obstructive pulmonary —the prevalence in the east is higher than in other regions, while, there is higher prevalence for other conditions, such as asthma, in the northeast. The south and midwest also show their own prevalence of chronic diseases. Likewise, there are variations for hospitalization and mortality rates. In addition, there are gender differences related to chronic conditions. For example, women tend to have higher cases per 100,000 for asthma-related hospitalizations. Men, on the other hand, appear to have higher mortality rates for chronic obstructive pulmonary disease, diabetes, chronic kidney, and others. Also, when we examined chronic conditions by race, we noticed that American Indian or Alaska Natives had higher mortality rates for chronic obstructive pulmonary disease, diabetes, chronic kidney, and so on, followed by Black and non-Hispanic groups.

In addition, the study analyzed demographics of mental health, behavior habits, and preventive health. The associations between behavioral health and chronic conditions and between preventive health and chronic conditions were also analyzed. There is a positive relationship between average female coronary heart disease mortality ratio and average female tobacco use ratio. There is a negative relationship between the average pneumococcal vaccination among diabetes patients and the average diagnosed diabetes ratio among the population. Referring to the relationship between behavioral health and overarching conditions, the study found a negative correlation between age-adjusted prevalence percentage of fair or poor self-rated health status among adults aged ≥18 years and binge drinking adults. The current smoking prevalence and sufficiency of sleep among adults is negatively related. The current lack of health insurance is negatively related to both prevalence of current smoking and that of current smokeless tobacco use. The relationship between obesity and poor self-rated health status is positively related. Similarly, current smoking prevalence has a strong, positive correlation with fair or poor self-rated health status. There are different negative or positive correlations between overarching conditions and chronic conditions. For instance, there is a significant positive relationship between the prevalence of a lack of health insurance and that of diagnosed diabetes. But the relationship between prevalence of a lack of health insurance and prevalence of asthma is negatively related.

Finally, we conducted analyses of the differences among chronic conditions. There are obvious clusters between men and women for asthma, although women tend to have a higher prevalence of asthma than men. Overall, prevalence of asthma is negatively related to the prevalence of diabetes. There is a moderate, positive correlation between prevalence of kidney and diabetes, which is akin to the positive correlation between the prevalence of chronic obstructive pulmonary disease and diabetes, arthritis and asthma, arthritis and chronic obstructive pulmonary disease, and asthma and chronic obstructive pulmonary.

As we have taken the data of numerous patients and predicted their associated disease based on their symptoms by using different machine learning algorithms like the K-Fold Cross-Validation method, Random Forest Classifier, Support Vector Classifier, and Gaussian Naive Bayes Classifier. At last, it was concluded that patients based on their symptoms are suffering from “FUNGAL INFECTION”.

We can see that our combined model has classified all the data points accurately. We have come to the final part of this whole implementation; we will be creating a function that takes symptoms separated by commas as input and outputs the predicted disease using the combined model based on the input symptoms.

The function is created which has taken symptoms as input and has generated predictions for a disease. And according to dataset output came to be as “FUNGAL INFECTION”.

```
{'rf_model_prediction': 'Fungal infection',
'naive_bayes_prediction': 'Fungal infection',
'svm_model_prediction': 'Fungal infection',
'final_prediction': 'Fungal infection'}
```

Fig [32]. Final Result

Note: The symptoms that are given as input to the function should be exactly the same among the 132 symptoms in the dataset.

Chapter-6

Conclusion

We have taken the data from Kaggle and have applied the machine learning algorithm on that dataset which consists of symptoms to predict disease.

We have used different algorithms such as the K-Fold Cross-Validation method, Random Forest Classifier, Support Vector Classifier, and Naïve Bayes Classifier to get the accuracy in our prediction. We have gained 100% accuracy on all four models.

With the following data the model will predict the disease which will come in future and aware to all peoples for future. I have used the Standard file from Kaggle who predict at most 40+ diseases at a time.

With the help of this project, all individuals will aware of their health histories and we can also share our data with the government if they follow our terms and conditions. The government can use our data in terms of maintaining and taking measure steps for the betterment of the health of each citizen in a particular region or state.

The study makes multiple essential contributions to chronic disease analysis at the patient/physician and the state levels. At the patient level, analysis of chronic conditions and related behavioral factors allows patients to be proactive in managing their conditions as well as modifying behavioral health. In this day and age, patients are eager to assimilate health information from various sources [28,29]. Being informed allows patients to self-monitor and seek appropriate and timely medical care [30,31], contributing to an ultimate care model that is increasingly personalized.

Similar to patients, physicians too have varying information needs in healthcare that need to be satisfied [32]. To physicians, information on chronic conditions and more importantly, associations between multiple conditions and between categories of healthcare, enable developing personalized treatment plans based on patient-specific profiles that integrate various symptoms with environmental and other health data [33]. Additionally, the array of information increases their ability to guide patients in towards lifestyle medicine (making lifestyle changes in healthy diet, exercise etc.) in the management of chronic diseases [34]. The road from sickness to wellness requires integrated efforts from physicians and patients—physicians can coach and guide the patients but the ultimate cross-over to wellness lies in the patients' hands.

Whereas most studies on chronic diseases focus on specific chronic diseases and are somewhat limited, this study offers comprehensive analysis over multiple categories of chronic diseases at the state-level. By utilizing visual analytics and descriptive analytics, our study offers methods for gaining insight into the relationships between behavior habits, preventative health and demographics, and chronic conditions. Moreover, this study contributes in terms of the methodology of analytics used in the research. It demonstrates the efficacy of data-driven analytics, which can help make informed decisions on chronic diseases.

Going forward, more theoretical and empirical research is needed. Additional studies can address the relationship between chronic disease conditions and other indicators, such as economic, financial, and social. While chronic disease management has become the focus in modern medicine as our population ages and medical costs continue to rise, research should focus on preventive and mitigating policies. The benefits of prevention and its potential to reduce costs and improve outcomes have received the attention of insurance companies, health care plans, and the U.S. Congress. Healthcare systems are now incentivized to reduce readmissions and physicians are encouraged to meet evidence-based quality measures to provide the best outcomes for patients with chronic disease states.

Acute and Chronic Disease Reports is a peer-reviewed journal that aims to publish research dealing with chronic and acute diseases such as: Severe Acute Respiratory Syndrome, Acute Bronchitis, Acute Lymphocytic Leukemia, Cardiac failure, Crohn's disease, and chronic renal disease.

The primary focus of the Journal lies in exploring the genetic and molecular mechanisms underlying these conditions and achieving a better understanding of the body's response to these conditions.

The Acute and Chronic Disease Reports Journal has assembled together renowned scientists in the Editorial Board. All the manuscripts are subject to vigorous peer-review process to ensure quality and originality. In addition to Research Articles, the Journal also publishes high quality Commentaries, Reviews, and Perspectives aimed at encapsulating the latest knowledge that synthesizes new theories and treatment strategies for the better management of chronic and acute disorders.

The team at the Acute and Chronic Disease Reports Journal takes immense pride in providing a streamlined and unbiased publishing process. Acute and Chronic Disease Reports Journal provides an encouraging platform for the scientists to share their invaluable contributions towards this field.

Acute and Chronic Disease Reports & Therapy is an academic journal which aims to publish most complete and reliable source of information on the discoveries and current developments in the mode of Research articles, Review articles, Case reports, short communications, etc. in all areas of the field and making them freely available through online without any restrictions or any other subscriptions to researchers worldwide.

Chapter-7

References

1. Basu J., Avila R., Ricciardi R. Hospital readmission rates in U.S. States: Are readmissions higher where more patients with multiple chronic conditions cluster? *Health Serv. Res.* 2016;**51**:1135–1151. [PMC free article] [PubMed] [Google Scholar]

2. Buttorff C., Ruder T., Bauman M. Multiple Chronic Conditions in the United States. [(accessed on 1 January 2018)]; Available online: https://www.rand.org/pubs/tools/TL221.html.

3. American Association of Retired Persons Chronic Conditions among Older Americans. [(accessed on 1 January 2018)]; Available online: https://assets.aarp.org/rgcenter/health/beyond_50_hcr_conditions.pdf.

4. Fried L. America's Health and Health Care Depend on Preventing Chronic Disease. [(accessed on 31 December 2017)]; Available online: https://www.huffingtonpost.com/entry/americas-health-and-healthcare-depends-on-preventing_us_58c0649de4b070e55af9eade.

5. Tinker A. How to Improve Patient Outcomes for Chronic Diseases and Comorbidities. [(accessed on 30 December 2017)]; Available online: http://www.healthcatalyst.com/wp-content/uploads/2014/04/How-to-Improve-Patient-Outcomes.pdf.

6. Comlossy M. *Chronic Disease Prevention and Management.* National Conference of State Legislatures; Denver, CO, USA: 2013. [Google Scholar]

7. Centers for Disease Control The Power of Prevention: Chronic Disease the Public Health Challenge of the 21st Century. [(accessed on 31 December 2017)]; Available online: www.cdc.gov/chronicdisease/pdf/2009-Power-of-Prevention.pdf.

8. O'Grady M.J., Capretta J.C. Health-Care Cost Projections for Diabetes and Other Chronic Diseases. [(accessed on 30 December 2017)]; Available online: http://www.civicenterprises.net/MediaLibrary/Docs/HR%20-%20health%20care%20cost%20projections%20for%20diabetes%20and%20other%20chronic%20diseases.pdf.

9. World Health Organization Ten Facts about Chronic Disease. [(accessed on 31 December 2017)]; Available online: http://www.who.int/features/factfiles/chp/10_en.html.

10. Becker's Hospital Review The Key to Population Health: Know Your Chronic Disease Patients and Coach Them. [(accessed on 31 December 2017)]; Available online: https://www.beckershospitalreview.com/healthcare-information-technology/the-key-to-population-health-know-your-chronic-disease-patients-and-coach-them.html.

11. National Association of Chronic Disease Directors Why Public Health is Necessary to Improve Healthcare. [(accessed on 31 December 2017)]; Available online: http://www.chronicdisease.org/?page=whyweneedph2imphc.

12. Trotter P., Lobelo F., Heather A.J. Chronic Disease Is Healthcare's Rising Risk. [(accessed on 31 December 2017)]; Available online: https://www.healthitoutcomes.com/doc/chronic-disease-is-healthcare-s-rising-risk-0001.

13. The Growing Crisis of Chronic Disease in the United States. [(accessed on 31 December 2017)]; Available online: https://www.fightchronicdisease.org/sites/default/files/docs/GrowingCrisisofChronicDiseaseintheUSfactsheet_81009.pdf.

14. Anderson G., Horvath J. The growing burden of chronic disease in America. *Public Health Rep.* 2004;**119**:263–270. doi: 10.1016/j.phr.2004.04.005. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

15. Beaton T. Top 10 Most Expensive Chronic Diseases for Healthcare Payers 2017. [(accessed on 1 January 2018)]; Available online: https://healthpayerintelligence.com/news/top-10-most-expensive-chronic-diseases-for-healthcare-payers.

16. Guidestone The Cost of Chronic Disease and Obesity 2011. [(accessed on 30 December 2017)]; Available online: https://www.guidestone.org/-/media/Insurance/WorksiteWellness/CostOfChronicDiseaseAndObesity.pdf?la=en.

17. Healthcare Information and Management Systems Society Keeping Communities Healthy: Wellness and Chronic Disease Management Initiatives. [(accessed on 1 January 2018)]; Available online: http://www.himssanalytics.org/news/top-pop-health-initiatives-essentials-brief-update.

18. U.S. Department of Labor, Bureau of Labor Statistics Consumer Expenditure Survey. [(accessed on 1 January 2018)]; Available online: http://www.bls.gov/cex/#overview.

19. U.S. Department of Health and Human Services . *Multiple Chronic Conditions—A Strategic Framework: Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions.* U.S. Department of Health and Human Services; Washington, DC, USA: 2010. [Google Scholar]

20. Raghupathi W., Raghupathi V. An overview of health analytics. *J. Health Med. Inform.* 2013;**4**:1–11. doi: 10.4172/2157-7420.1000132. [CrossRef] [Google Scholar]

21. Committee on Engaging the Computer Science Research Community in Health Care Informatics. Nat'l Research Council . In: *Computational Technology for Effective Healthcare: Immediate Steps and Strategic Directions.* Stead W.W., Lin H.S., editors. National Academies Press; Washington, DC, USA: 2009. [PubMed] [Google Scholar]

22. Khan M., Khan S.S. Data and information visualization methods, and interactive mechanisms: A survey. *Int. J. Comp. Appl.* 2011;**34**:1–14. [Google Scholar]

23. Caban J.J., Gotz D. Visual analytics in healthcare—Opportunities and research challenges. *J. Am. Med. Assoc.* 2015;**22**:260–262. doi: 10.1093/jamia/ocv006. [PubMed] [CrossRef] [Google Scholar]

24. Gotz D., Borland D. Data-driven healthcare: Challenges and opportunities for interactive visualization. *IEEE Comp. Graphic Appl.* 2016;**36**:90–96. doi: 10.1109/MCG.2016.59. [PubMed] [CrossRef] [Google Scholar]

25. Harle C., Neill D., Padman R. Development and evaluation of an information visualization system for chronic disease risk assessment. *IEEE Intell. Syst.* 2012;**27**:81–85. doi: 10.1109/MIS.2012.112. [CrossRef] [Google Scholar]

26. Keselman A., Heller N. Estrogen signaling modulates allergic inflammation and contributes to sex differences in asthma. *Front. Immunol.* 2015;**6**:568. doi: 10.3389/fimmu.2015.00568. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

27. Zein J.G., Erzurum S.C. Asthma is different in Women. *Curr. Allergy. Asthma Rep.* 2015;**15**:28. doi: 10.1007/s11882-015-0528-y. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

28. Amante D.J., Hogan T.P., Pagoto S.L., English T.M., Lapane K.L. Access to care and use of the Internet to search for health information: Results from the U.S. National Health Interview Survey. *J. Med. Internet Res.* 2015;**17**:106. doi: 10.2196/jmir.4126. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

29. Eysenbach G., Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Br. Med. J.* 2002;**324**:573–577. [PMC free article] [PubMed] [Google Scholar]

30. Bodenheimer T., Wagner E.H., Grumbach K. Improving primary care for patients with chronic illness. *J. Am. Med. Assoc.* 2002;**288**:1775–1779. doi: 10.1001/jama.288.14.1775. [PubMed] [CrossRef] [Google Scholar]

31. Himes B.E., Weitzman E.R. Innovations in health information technologies for chronic pulmonary diseases16. *Respir. Res.* 2016;**17**:1–7. doi: 10.1186/s12931-016-0354-3. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

32. Kourouthanassis P.E., Mikalef P., Ioannidou M., Pateli A. Exploring the online satisfaction gap of Medical doctors: An expectation-confirmation investigation of information needs. *Springer.* 2014;**820**:217–228. [PubMed] [Google Scholar]

33. Mikalef P., Kourouthanassis P.E., Pateli A.G. Online information search behaviour of physicians. *Health Inf. Lib. J.* 2017;**34**:58–73. doi: 10.1111/hir.12170. [PubMed] [CrossRef] [Google Scholar]

34. Hayes C., West C., Egger G. *Chapter 22—Rethinking Chronic Pain in a Lifestyle Medicine Context.* Academic Press; Cambridge, MA, USA: 2017. pp. 339–353. [Google Scholar]