

**A COMPARATIVE STUDY ON PREDICTION OF HEPATITIS B VIRUS (HBV) USING
MACHINE LEARNING ALGORITHM**

BY

**MUHAMMAD MUSABBIR
ID: ID: 172-15-10027**

AND

**MD. MEHEDI HASAN
ID: 172-15-9849**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE
Daffodil International University

Co-Supervised By

Dewan Mamun Raza

Lecturer

Department of CSE
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2021**

APPROVAL

This Project titled “A Comparative Study on Prediction of Hepatitis B virus (HBV) using Machine Learning Algorithm”, submitted by Muhammad Musabbir, ID No: 172-15-10027. Md. Mehedi Hasan, ID No: 172-15-9849 to Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 2020.

BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Rubaiya Hafiz
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Shamim H Ripon
Professor
Department of Computer Science and Engineering
East West University

External Examiner

DECLARATION

We hereby declare that; this project has been done by us under the supervision of **Zerin Nasrin Tumpa, Lecturer Department of CSE** Daffodil International University. We Also declare that neither this project nor any part of this project has been submitted anywhere else for award of any degree or diploma.

Supervised by:



Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Muhammad Musabbir
ID:172-15-10027
Department of CSE
Daffodil International University



Md. Mehedi Hasan
ID: 172-15-9849
Department CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness Zerine Nasrin Tumpa. Lecturer, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Syed Akther Hossain, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

Abstract

The Hepatitis B virus (HBV) can cause acute and chronic liver infections. It is transmitted through infected blood products, unprotected sex, infected items such as needles, razor blades, dental or medical equipment, unscreened blood transfusions, or from mother to child at birth. Despite a considerable body of published research on hepatitis B in Bangladesh, researchers continue to lament the lack of reliable information about hepatitis B virus (HBV) infection epidemiology. The present review aims to provide a comprehensive survey of the literature with particular focus on a number of epidemiological questions, as well as a commentary on the trends of hepatitis B research as it has taken place in Bangladesh. It is difficult to provide conclusive estimates about HBV prevalence in the general population of Bangladesh.

So, the aim of this piece of research work is to analyze the symptoms of Hepatitis B and Hepatitis C fever and early prediction of the symptoms that can be seen in years ahead. For predicting the symptoms two different Machine Learning algorithms have been used. Support vector machine (SVM) and random forest classifier algorithm have been used. Finally, the accuracy of these two has been evaluated and the confusion matrix has been shown and then we have talked about the algorithm which is better for our dataset.

TABLE OF CONTENTS

CONTENTS PAGES

Board of examiners	i
Declaration	ii
Acknowledgement	iii
Abstract	iv

CHAPTER

CHAPTER 1: INTRODUCTION **1-5**

1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the study	3
1.4 Research Questions	4
1.5 Expected Output	4
1.6 Report Layout	5

CHAPTER 2: BACGROUND **6-9**

2.1 Introduction	6
2.2 Related Works	7
2.3 Cost effectiveness assessment	7
2.4 Cost effectiveness estimates	8
2.5 Research Summary	8
2.6 Scope of the Problem	9
2.7 Challenges	9

CHAPTER 3: Research Methodology **10-15**

3.1 Introduction	10
3.2 Research Subject and Instrumentation	10
3.3.1 Data Collection Procedure	10
3.4 Statistical Analysis	12
3.5 Implementation Requirements	14

CHAPTER 4: Experimental Results and Discussion	16-25
4.1 Introduction	16
4.2 Experimental Results	16
4.3 Descriptive Analysis	24
Chapter 5: Summary, Conclusion, Recommendations and Implication for Future Research	26-27
5.1 Summary of the Study	26
5.2 Conclusions	26
5.3 Recommendations	27
5.4 Implication for Further Study	27
References-28	
Appendice-30	

LIST OF FIGURES

FIGURES	PAGE NO
Figure1.5.1 Expected Output	4
Figure2.1.1 Highest Number of Hepatitis B Virus Cases in 24 Year	6
Figure 3.3.1 Data Collection	11
Figure 3.4.1 Gender Ratio of Patients	12
Figure 3.4.2 Age Distribution of Patients	13
Figure 3.4.3 Number of Syndrome	14
Figure 3.5.1 Work Methodology	15
Figure 4.2.1 Result using SVM Algorithm	17
Figure 4.2.3 Result using RFC	17
Figure 4.2.7 - 4.7.17 Prediction	19-23
Figure 4.2.18 Hepatitis B and Hepatitis C Screening SSHMCH2018	24
Figure 4.2.18 Hepatitis B and Hepatitis C Screening SSHMCH2018	25

LIST OF TABLES

TABLE	PAGE NO
Figure 3.4.1 Gender Ratio Of Patiens	13
Figure 3.4.26 Accuracy of Algorithms	18

CHAPTER 1

Introduction

1.1 Introduction

Hepatitis B is a viral infection that assaults the liver and can cause both intense and ongoing sickness. The infection is most normally transmitted by needle sharing, unprotective sex, perilous blood bonding, inking and from mother to kid during birth and conveyance, also as through contact with blood or other body liquids. [1][2]

Inking and needle therapy prompted countless cases during the 1980s. Nevertheless, this has gotten more uncommon with improved sterilization. The hepatitis B infections can't be spread by kissing, holding hands, coughing, hugging, sneezing, or breastfeeding. The disease can be analyzed 30 to 60 days after presentation. The analysis is typically affirmed by testing the blood for parts of the infection and for antibodies against the infection. It is one of five primary hepatitis infections: A, B, C, D, and E. [3][4]

The infection is available around the world, however a few populaces in sub-Saharan Africa, Southeast Asia, Eastern Europe, and the Center East, just as in indigenous networks are Hepatitis B transporters. The current audit plans to give a thorough overview of the writing with specific spotlight on various epidemiological inquiries, just as an editorial on the patterns of hepatitis B research as it has occurred in Bangladesh.[5]

The greater part of the 10 million individuals assessed to convey the Hepatitis B infection (HBV) in Bangladesh don't realize they have the contamination and are putting others in danger, and doctor have cautioned. Chronic hepatitis B portrays a range of illness normally described by the presence of distinguishable hepatitis B surface antigen (HBsAg) in the blood or serum for longer than a half year.[6] In certain individuals, ongoing hepatitis B is dormant and doesn't present critical medical conditions, yet others may advance to liver fibrosis, cirrhosis and hepatocellular carcinoma (HCC). The movement of liver illness is related with hepatitis B infection (HBV) DNA levels in the blood. Without antiviral treatment, the five years combined rate of cirrhosis ranges from 8 to 20%. [7]

Individuals with cirrhosis face a critical danger of decompensated liver illness in the event that they stay untreated. Five-year endurance rates among individuals with untreated decompensated cirrhosis can be as low as 15%. Ongoing hepatitis B can be isolated into antigen-(HBeAg) positive or HBeAg-negative malady dependent on the presence or nonattendance of e antigen. The presence of HBeAg is normally connected with higher paces of viral replication and in this way expanded infectivity.[8]

Signs and side effects of hepatitis B extend from mild to extreme. They normally show up around one to four months after you've been contaminated, in spite of the fact that you could consider them to be right on time as about fourteen days post-disease. A few people, normally small kids, might not have any indications. Hepatitis B signs and symptoms may include-

1. Fever
2. Abdominal pain
3. Dark urine
4. Nausea and vomiting
5. Joint pain
6. Weakness and fatigue
7. Joint pain
8. Loss of appetite
9. Yellowing of your skin and the whites of your eyes (jaundice)etc.

Our point is to build up a model for anticipating indications of Hepatitis utilizing some AI Calculations and expanding the consciousness of the individuals of our nation and diminishing the quantity of passings influencing by Hepatitis B. [3][1][4]

1.2 Motivation

The hepatitis B infection was found in 1965 by Dr. Baruch Blumberg who won the Nobel Prize for his disclosure. Initially, the infection was known as the "Australia Antigen" since it was named for an Australian native's blood test that responded with an immunizer in the serum of an American hemophilia tolerant.[9]

Working with Dr. Blumberg, microbiologist Irving Millman assisted with building up a blood test for the hepatitis B infection. Blood donation centers started utilizing the test in 1971 to screen blood donation and the danger of hepatitis B diseases from a blood bonding diminished by 25 percent. Four years after discovering the hepatitis B virus, Blumberg and Millman built up the primary hepatitis B immunization, which was at first a heat-treated form of the virus. [9]

In such case in present, we have figured out how to keep from hepatitis B, so our justification in the hypothesis was to- [10]

- Use of expendable needle and needles, bond screened blood, not to share utensils and utilize safe sex.
- Vaccination (Hepatitis B antibody) in hazard gatherings and relatives of Hepatitis B
- Universal youth vaccination.
- Babies destined to HBsAg positive moms first portion of vaccination inside 24 hours of birth alongside Hepatitis B immunoglobulin. (HBIG)
- Vaccination of Careful, clinical and dental staff.
- Vaccination of emergency clinic and lab staff who interact with blood and blood items regularly.

1.3 Rationale of the Study

The impact of our clinical consideration and treatment on Hepatitis B has a long stretch effect on. Our investigation on clinical consideration and medicines are- [11]

- Ensure supplier admittance to current rules, guidelines and proposals for hepatitis B analysis, treatment and counteraction.
- Establish a referral network for hepatitis B determination, care and treatment.
- Assure ideal admittance to hepatitis B analysis, care, post-presentation prophylaxis and treatment.

- Integrate hepatitis B care, treatment and steady administrations into essential consideration settings.
- Ensure ideal admittance to sexual wellbeing related administrations.
- Address the unpredictable needs connected with hepatitis B through coordination of care.
- Establish projects to help hepatitis B care and treatment for uninsured and underinsured people.

1.4 Research Questions

While doing the research on Hepatitis B, some questions arose my mind. The questions are

1. What causes hepatitis B?
2. Is Hepatitis B curable or not?
2. What are the early signs of hepatitis B?
3. How dangerous is hepatitis B?
4. How is hepatitis B diagnosed?
5. What should not eat in hepatitis B?
6. What is HBeAg test?
7. What can be done to reduce the risk of Hepatitis B?
8. What type of protection must be taken to prevent Hepatitis B?

1.5 Expected Output

In this exploration building up a model for predicting the Hepatitis B is being taken a stab at utilizing some various Calculations of AI. And furthermore, our point is to make individuals more cognizant and make individuals more mindful about Hepatitis B and its danger factor with the goal that individuals can go for an appropriate determination and treatment. Our normal yield is a rundown of some danger factors which can be found in the following years. From that an example of our normal yield is given beneath -

**['Syndrome', 'Surgery', 'Blood-transf', 'Alcohol', 'Sexual-exp', 'Drug-abuse',
'Hap-B-vaccination', 'HBV-vaccine schedule', 'Anti HCB', 'Family-Hap-B', 'Family-Hap-C']**

Figure1.5.1 Expected Output

1.6 Report Layout

Chapter 1: Introduction

In this part, we've talked about the motivation of ours for doing this research and also the rationale of the study, research questions and the expected outcome of the research.

Chapter 2: Background

In this chapter, we will discuss about the related work of our research and comparative studies between our developed model and existing related model as well as the problems and challenges that we've faced.

Chapter 3: Research Methodology

In this chapter, we will discuss about our research methodology that means we will talk about how we did our research and about the terminology.

Chapter 4: Experimental results and Discussion

In this chapter, we will discuss about our research experimental results and outputs. We will talk about which results we have found by implementing the algorithms of machine learning over our collected data.

CHAPTER 2

Literature Review

2.1 Introduction

AI engages IT frameworks to see plans dependent on existing figurings and instructive assortments and to make adequate course of action thoughts. Its algorithm is exceptionally ground-breaking for anticipating any result over the given information or data. That is the reason we have utilized AI calculations for forecast. Hepatitis B has become as a danger for our life. It took an excessive number of lives in this year from our Bangladesh. Not just for our nation it has become a danger for the entire world.

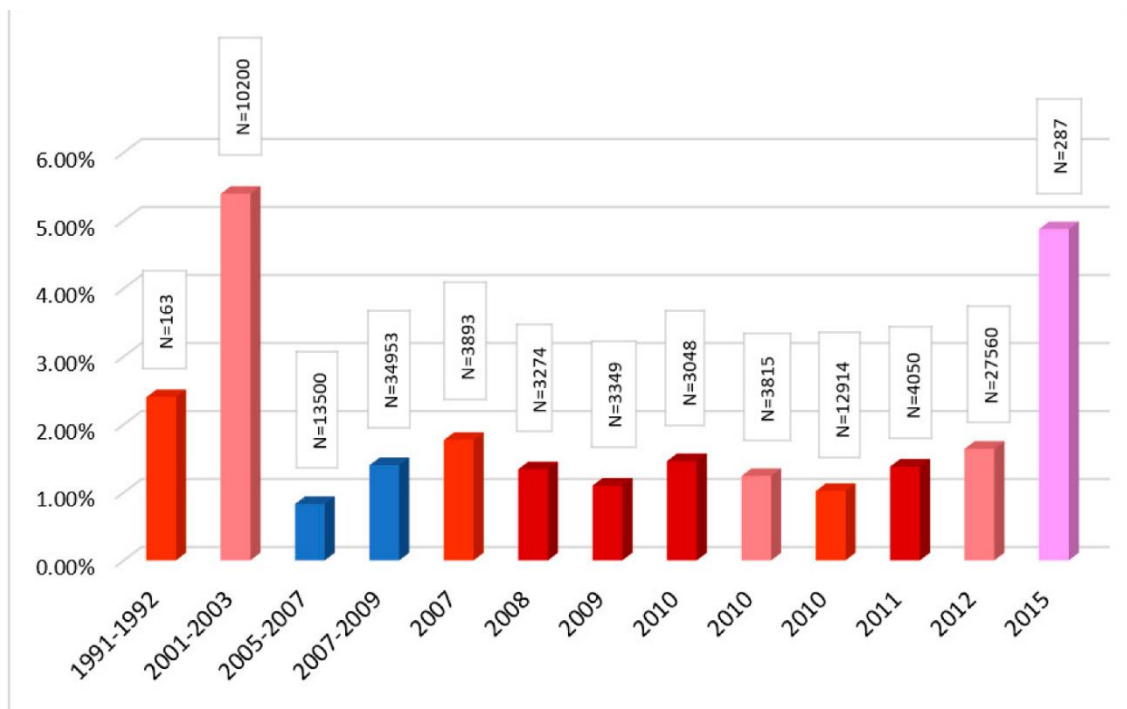


Fig 2.1.1:Hepatitis B surface antigen (HBsAg = prevalence among voluntary and replacement blood donors in Bangladesh. Darker hues of red indicate increasing male: female ratio in the study population (blue bars indicate no gender data was available). The percentage of males in the study population range from 81% [33] to 98% (2011 data in Reference. [12]

2.2 Related Works

For information extraction and stakeholder commitment on 22 and 23 of July 2019, a public workshop for hepatitis surveillance. Vital data and arranging occurred. The members were drawn from the Service of Wellbeing. Medical clinic and research facility staff, and global accomplices. Introductions were made by public specialists and pertinent officials on the status of public hepatitis program, existing DHIS-2 and observation organization, hurt decrease and lab offices for viral hepatitis testing. We evaluated existing information frameworks taking into account the data that should be created to screen the 10 center markers. We followed the techniques proposed in the WHO united vital data rules to choose the most ideal wellspring of data to appraise the benchmark estimation of the center pointers. We validated these data through discussions with key stakeholders and extracted the data elements needed for the Global Reporting System for Hepatitis (GRSH). We reviewed gaps in the data systems with the stakeholders and proposed a way forward that could improve strategic information for viral hepatitis in Bangladesh. At long last, we utilized the gauge estimation of the markers alongside other data accessible to set inclusion focuses for the public arrangement in meeting with the members and public program supervisors. We also met the director of the Institute of Epidemiology, Disease Control and Research (IEDCR) to discuss issues related to an integrated surveillance system for acute hepatitis. [Source:Shaheed Suhrawardy Medical College and Hospital.[13]

2.3 Cost effectiveness assessment

We pilot tested the new WHO tool to estimate the cost of sequelae of HBV and HCV infection at the department of hepatology, Bangabandhu Sheikh Mujib Medical University (BSMMU), an autonomous public-sector hospital which is a premier postgraduate medical institution of the country. The department of hepatology has been established since 1989 and has considerable experience in treating patients with sequelae of hepatitis. We explained the principles underlying the tool to one of the residents. The tools considered different sequelae (example: chronic hepatitis, compensated cirrhosis, hepatocellular carcinoma) and for each, estimated the annual cost by multiplying the unit price) {e.g., cost of an outpatient visit. cost of an ultrasound, by the estimated usage of these resources every year (“q”, the quantity). We collected data on the “q”s (The resident was well aware of the different unit costa) and on the “q”s (The frequency of

outpatient average need for hospitalization for chronic hepatitis. compensated and decompensated cirrhosis and hepatocellular carcinoma). If use of a specific resource was not required each year. we entered a “q” as a fraction to reflect the frequency of usage over the years (example: if hospitalization happened every third year, the yearly “q” was set at 0.33). The tool was further validated during the workshop with CDC where a number of hepatologists were present. There was consensus on most unit costs, but some minor cost corrections were done and some new items related to treatment were added to the tool. It was pointed out that for indoor patients the bed charges will not apply at public sector hospital.

2.4 Cost effectiveness estimates

We entered the expense of conclusion and treatment of HBV and HCV along the expense of sequelae got with the costing instruments in the hepatitis B (www.hepbcalculator.org) and hepatitis C (www.hepcccalculator.org) mini-computers. This prompted evaluations of the steady cost adequacy proportion (ICE, for treatment of people determined to have HBV or HCV disease, including timetable to the mediation getting cost sparing from a cultural viewpoint. at the point when material.[14]

2.5 Research Summary

Developing a model for doing the prediction of Hepatitis B symptoms is being tried which can be seen in future years using some different and effective algorithms of machine learning. There is some related work which were doing using different types of algorithms over different types of collected data and information. Someone has used clinical data; someone has used weather data and another one has used environmental data to predict Hepatitis using different techniques. In May 2016, the World Health Assembly adopted the first "Global health sector strategy on viral hepatitis, 2016-2020". The strategy highlights the critical role of universal health coverage and sets targets that align with those of the Sustainable Development Goals. The strategy has a vision to eliminate viral hepatitis as a public health problem. This is encapsulated in the global targets to reduce new viral hepatitis infections by 90% and reduce deaths due to viral hepatitis by 65% by 2030. Actions to be taken by countries and the WHO Secretariat to reach these targets are outlined in the strategy. [15]

2.6 Scope of the Problem

In every research there can be a scope of problem. So, there is no exception in our work too. Our collected data should be appropriate and our data set should be smooth to do work with. We need to train our collected data properly while working with algorithms. If we can't do this properly our output won't be appropriate and won't be similar to the real-life scenario. If we want to reduce the problem we should work properly and carefully.

2.7 Challenges

No task can be refined without difficulties. While accomplishing any work, difficulties could come in any means of the work. In our examination the primary test we needed to confront while gathering information from the Hepatitis B patients. That implies in our exploration studying information from patients and gathering the data is the fundamental test. The assortment of legitimate and fitting information is required from patients for building a smooth dataset. Thus, we can say that the fundamental test is that the gathered information ought to be more suitable for our examination for finding the better yield.

Chapter 3

Research Methodology

3.1 Introduction

This part gives a system of exploration procedures that were continued in the examination. It gives information about the individuals, that is, the models for the fuse in the assessment, which the members were and how they were tried. We portray the assessment structure that was picked with the ultimate objective of this examination and the reasons behind this choice. The instrument that was used for data amassing is moreover portrayed, and the strategy that were sought after to do this assessment is incorporated. We also talk about the systems used to separate the data. All in all, execution necessity that were continued in the process is similarly discussed.

3.2 Research Subject and Instrumentation

Our research for prediction of Hepatitis B risk factor. We are doing this using some different machine learning algorithms. So, our research subject is - A Comparative Study on Prediction of Hepatitis B Using Machine Learning Algorithm.

In examination of human subjects, an overview is a summary of inquiries got ready for isolating unequivocal data from a particular social affair of people. Reviews are coordinated by phone, mail, by methods for the web, and so on. Overviews are used to extend learning in fields, for instance, social examination and demography. For our research structural questionnaires was first prepared, it was pre-tested. After finalization it was used as research instrument. So, in our research work surveys questionnaires are the main instrument which can be said as research instrument. After surveying the collected data has been generalized for using some preprocessing techniques. After that the preprocessed data has been prepared for our research work. Then the prepare data has been used into machine learning (ML) algorithms for predicting our desired Hepatitis B Symptoms.

3.3.1 Data Collection Procedure

Information assortment is one of the most significant undertakings in our exploration work. Information assortment or data amassing is the path toward the get-together and assessing information on variables of energy, in a developed exact plan that enables one to react to communicated research questions, test theories, and evaluate results. The data collection some portion of exploration is fundamental to all fields of study including physical and humanistic systems, humanities, business, etc. While systems vary by discipline, the highlight on ensuring precise and reasonable aggregation proceeds as in the past. It is perhaps the hardest thing that we have confronted. There are many ways to collect information and data but we have collected our data in two different ways. They are-

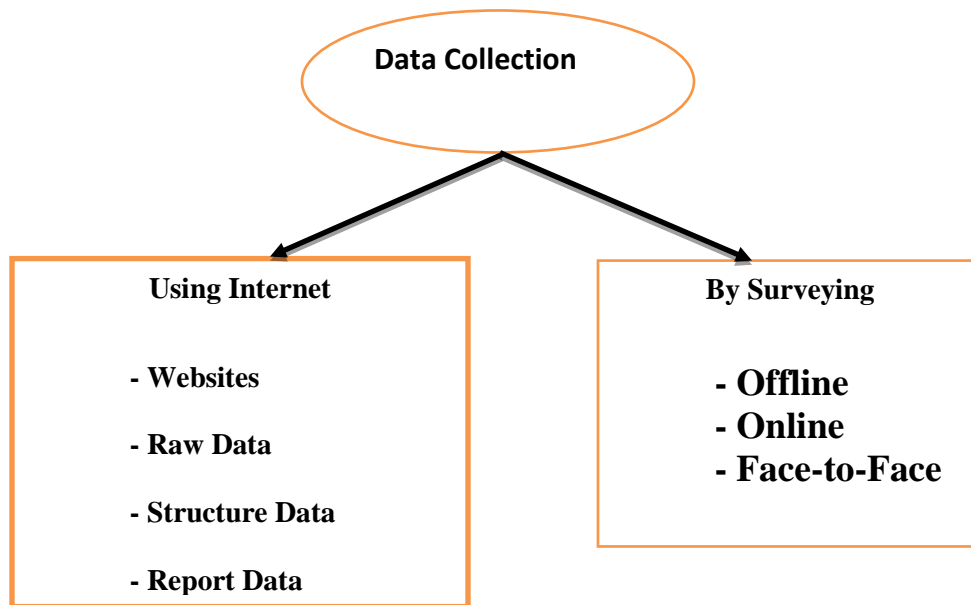


Figure 3.3.1 Data Collection

Surveys:

We have surveyed over more than 500 hundred patients for collecting data and finding Information. We made a question paper including the questions that are necessary for our research work. The question paper was consisting of with questions like — Age, Gender, how many times they have been infected with Hepatitis B, what were the symptoms they faced,

Duration of staying at the hospital, Duration of recovery and some other questions. By collecting their data, we fill it into our dataset.

Internet:

We have taken further support from Internet like websites, magazines, journals, on-line newspapers for enriching our dataset and to make bigger dataset.

3.4 Statistical Analysis

After surveying we have found some analytical statistic data by which some differentiation can be seen and these in below-

Gender Ratio:

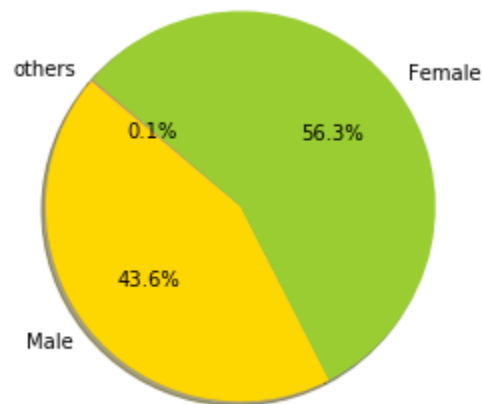


Figure 3.4.1 Gender Ratio of Patients

According to this Hepatitis B & Hepatitis C Virus we can see that female are in danger zone more than male. The pace of both Hepatitis B & Hepatitis C Virus is higher among men, however more ladies are among the individuals who have died from the hepatitis-borne viral ailment in Shaheed Suhrawardy Medical College and Hospital. After our surveying we have found that 43.6% male, 56.3% female and less than 1% other got attacked by hepatitis viral

disease in 2018 of our country. This number is excessively high contrasting with the earlier years.

Many people already been suffering from hepatitis B and female are more in number. It will be more clarify when it will be shown in table format.

Gender	Percentage%
Male	43.6%
Female	56.3%
Other	0.1%

Age Distribution

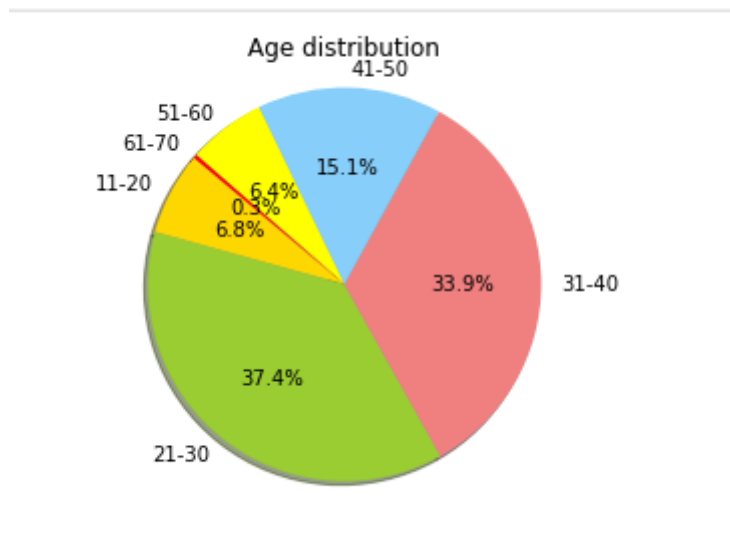


Figure 3.4.2 Age Distribution of Patients

From this figure it can be said that age among 21-30 are the highest number of Hepatitis B 37.4% and 31-40 are second highest number of Hepatitis B 33.9% have been attacked by Hepatitis B in our country according to our survey. Age among 41-50 have taken the 3rd spot and rest of them are significantly lower than 7%. Age between 61-70 are the lowest of all.

Number of Syndrome

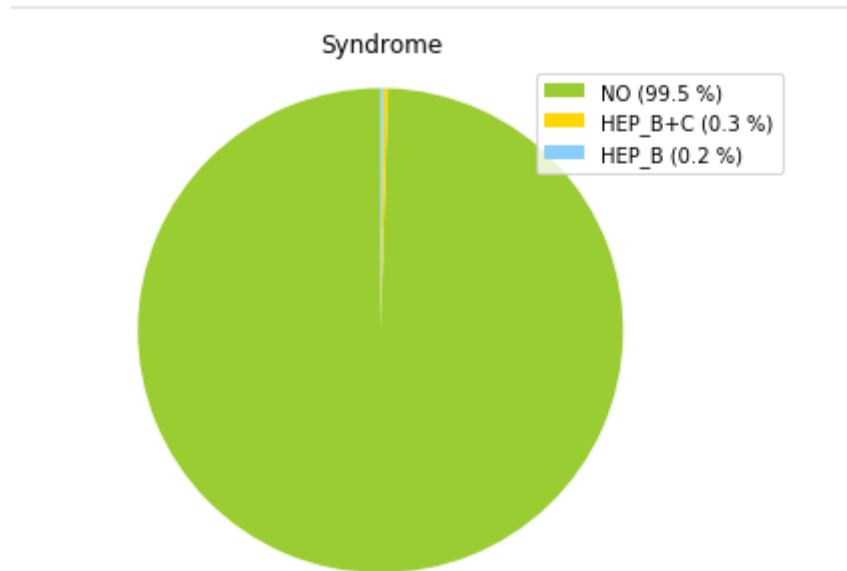


Figure 3.4.3 Number of Syndrome

In our country, 99.5% people had not found the Hepatitis B virus. Remaining 0.2% people who had already attacked by Hepatitis B virus and 0.3% people was attacked by both Hepatitis B and Hepatitis C virus.

3.5 Implementation Requirements

For this examination Hepatitis B side effects are being anticipated for diminishing the future harm of numerous patients. Early side effects forecast can spare a patient life and they can go for better and early analysis. Most importantly, information has been gathered by us by reviewing and utilizing Web. Gathered information have been arranged and cleaned utilizing information pre-preparing procedures. At that point information have been summed up at last a yield has been produced by calculation when it has been given to the calculation for prediction.

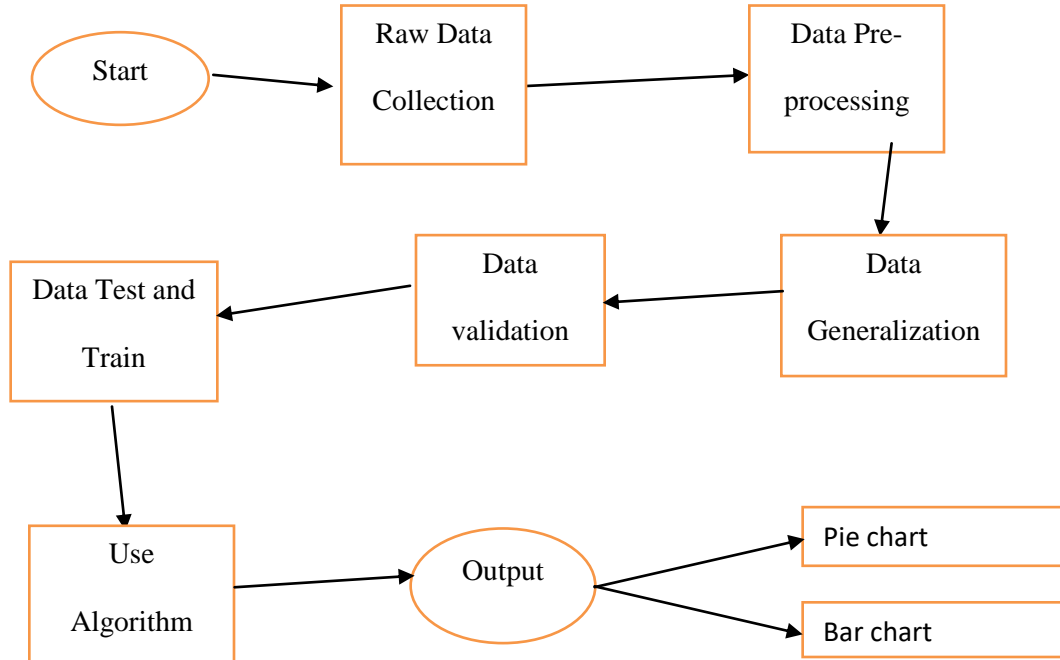


Figure 3.5.1 Work Methodology

Using this methodology our expected output will be predicted. When it comes to algorithm, we must say that two algorithms have been used for prediction of Hepatitis B risk factor. The algorithms are Support Vector Machine (SVM) and Random Forest Algorithm. We are using two algorithms for differentiate the accuracy rate of these two and finally cross validation will be found and confusion matrix also to evaluate the efficiency of the algorithm. For doing these all tasks we need to import to Google Collab. Google Colaboratory is our main working platform for this Hepatitis B prediction which we have used pandas, NumPy library in our project. Google Colaboratory is a Google Research product that allow us to write and execute python codes uninterruptedly. Basically, it is also suited to Machine Learning. Pandas is used for data frame sidearm. Model selection is used for importing the models or algorithm which we want to use and also used for importing train test split function for train and test the dataset.

Chapter 4

Experimental Results and Discussion

4.1 Introduction

The prescient intensity of the determining model-surveyed by Support Vector Machine (SVM) and Random Forest Algorithm are improved by including the new parameters for Hepatitis B fever indications expectation. The restrictive utilization of looking over and gathering of information from numerous areas is the model's primary expectation control. Seasons were emphatically corresponded with Hepatitis B cases. Hence, the Hepatitis B rate patterns gave by this model will help the enhancement of Hepatitis B anticipation. The present work shows the significant jobs contamination rates from the past season and individuals will be increasingly cautious with the Hepatitis B fever side effects. Fusing these two factors in the model essentially improves the prescient intensity of Hepatitis B fever predicting models.

Our model is able to predict the risk factor of Hepatitis B fever which can be seen in the next years. Early prediction is very important in this factor and using that people can go for better treatment in early stage before losing their lives. And our government can also take Proper steps for reducing the damage.

4.2 Experimental Results

Some results have been found after our experiment and that is discussed below,

**['Age_Yr','Sex','Religion','Past_Hep_B','b_Past_Hep_C','a_Family_Hep_B',
'a_Family_Hep_C','Dialysis','CLD','Diabetes','Blood_trans','Surgery','Alcool','Sex_Exp',
','IV_Drug','Vaccination','Vac_Schedule','HBsAg','Anti_HCV']**

At last, the result that array number is “2” which was meant that individual was not infected from the Hepatitis virus.

Also, we can see here that using both algorithm like Support Vector Machine (SVM) and Random forest classifier (RFC) similar results have been found.

Accuracy of Algorithms:

Algorithm	Accuracy
Support Vector machine	96.8345
Random Forest Classifier	97.0

Table 4.2.5 accuracy of algorithms

Here it is clear the accuracy that have been found it is similar in both Support vector Machine and Random Forest classifier.

Prediction:['Surgery','Blood-transf','Alcohol','Sexual-exp','Drug-abuse','Hap-B-vaccination','HBV-vaccine schedule', 'Anti HCB', 'Family-Hap-B', 'Family-Hap-C']

No -1 820
Yes -0 174

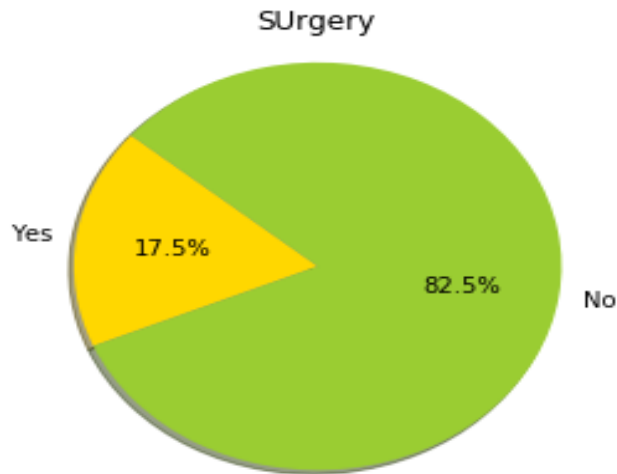


Figure 4.2.7 Surgery

The pie chart showed that 17.5% of patients had been taken previous surgery and the rest of them had not taken surgery before.

No-1 911
Yes-0 84
Others-1

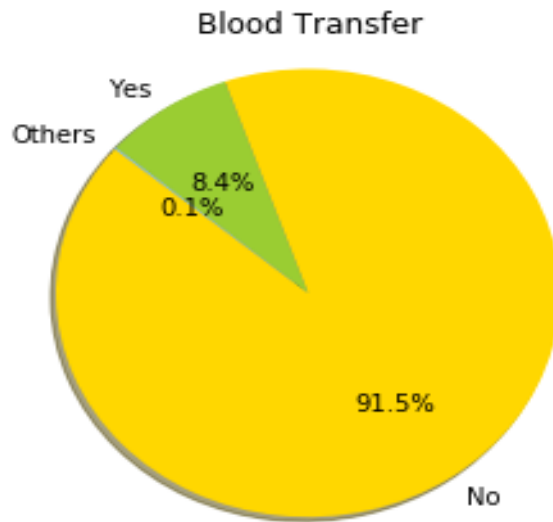


Figure 4.2.8 Blood Transfer

The pie chart showed that 8.4% of patients took a volume of blood that previously taken from a healthy person.

No-1 978
Yes-0 17

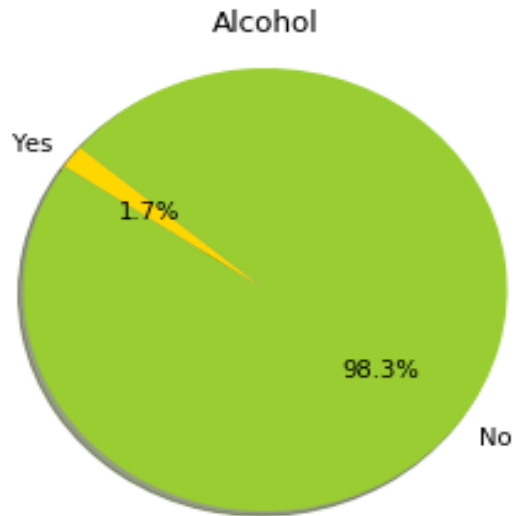


Figure 4.2.9 Alcohol

The pie chart showed that 98.3% of patients are not alcoholic and the rest of patients have a habit of drinking alcohol.

No-1 987
Yes-0 10

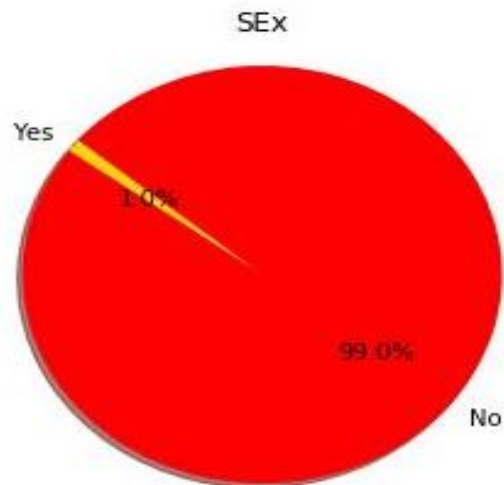


Figure 4.2.10 Sexual-exposure

The pie chart revealed that 99% of patients didn't have sexual activities and only 1% had a sexual intercourse.

No-1 985

Yes-0 11

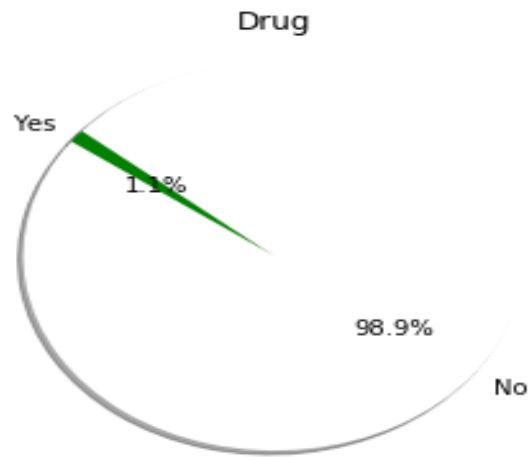


Figure 4.2.11 Drug-abuse

The pie chart showed that 98.9% didn't suffer from illicit drug use disorder and rest of them had been reported to have a drug abuse on purpose.

No-0 900

Yes-1 96

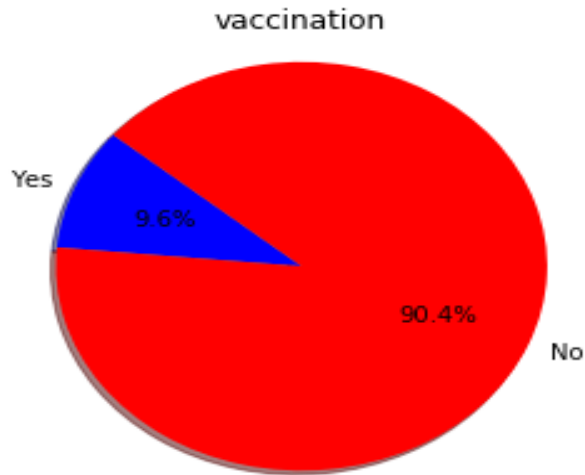


Figure 4.2.12 Hepatitis-B-vaccination

The pie chart showed that 90.4% of patients had been vaccinated and 9.6% didn't take the Vaccine.

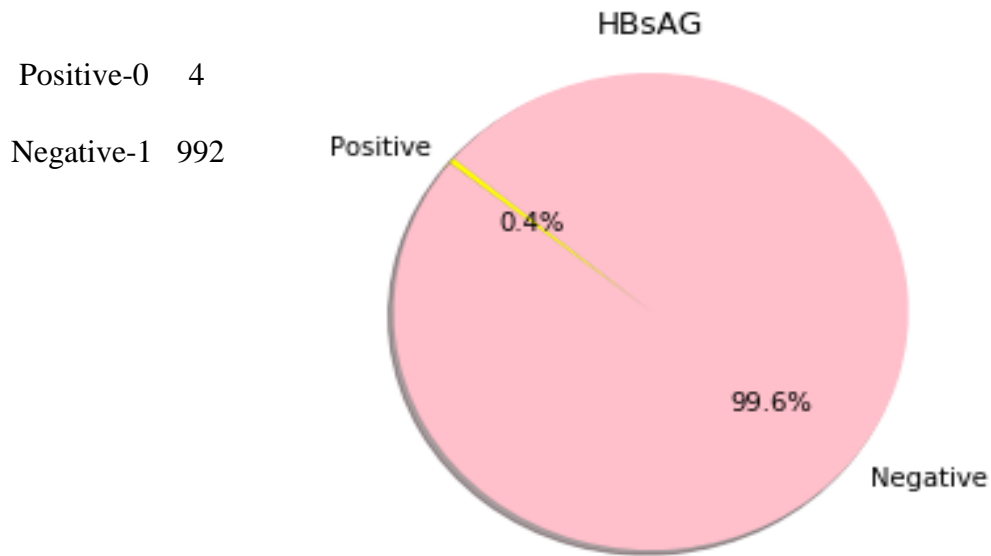


Figure 4.2.13 HBsAG

HBsAG is a test result which means that the person is infected with hepatitis B. So the pie chart showed that only 0.4% found the presence of the Hepatitis B.

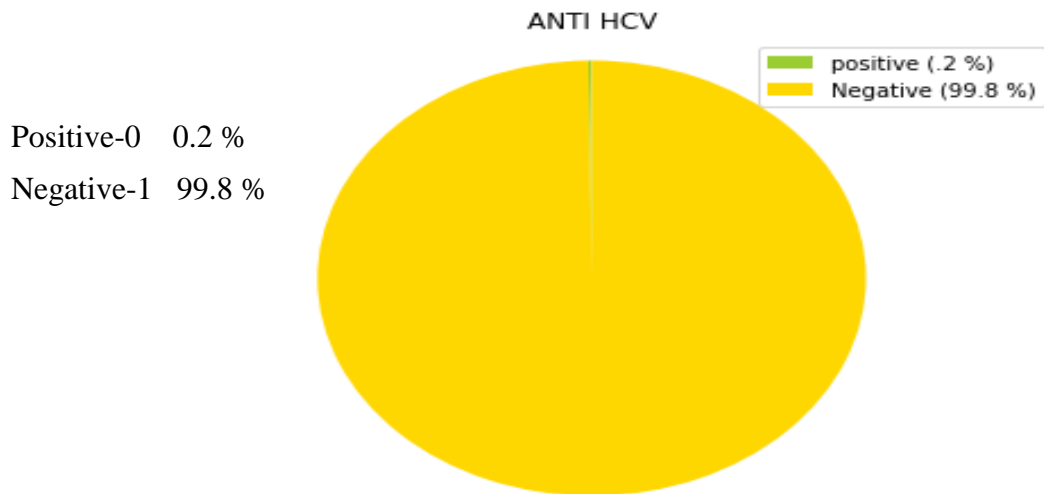


Figure 4.2.15 Anti-HCV

The HCV antibody sometimes called the Anti-HCV test to look for antibodies to the hepatitis C virus in blood. Only 0.2% found positive anti-HCV test and rest of them found negative.

No-1 965
Yes-0 31

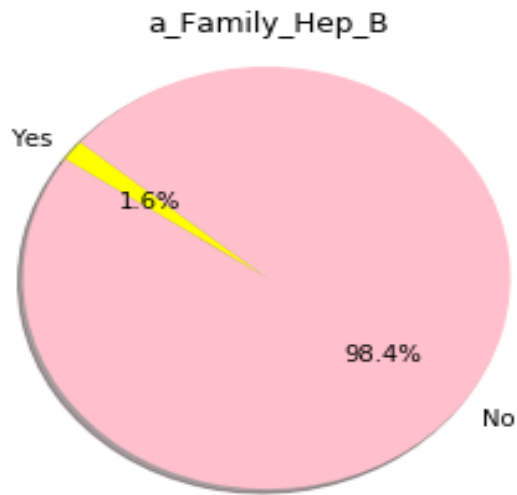


Figure 4.2.16 A _Family_Hepatitis_B

Family Hepatitis B referring to patient's related or family who had been infected by Hepatitis virus. So, pie chart revealed that 1.6% of them got infected by Hepatitis B virus.

No-1 965
Yes-0 31

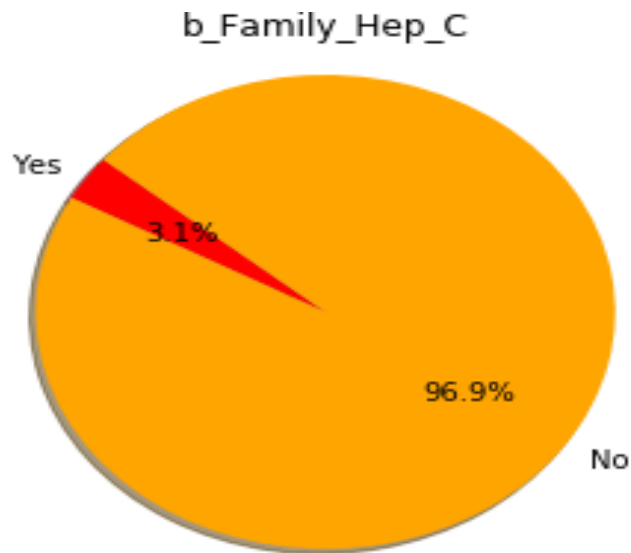


Figure 4.2.17 A _Family_Hepatitis_C

Family Hepatitis C also referring to patient's related or family who had been infected by Hepatitis virus. So, the pie chart exposed that only 3.1% of them infected by Hepatitis C virus.

4.3 Descriptive Analysis

The outbreak will be more prevalent in 2018 among patients have been found in our experiment over our dataset. This prediction is for the year of 2018.

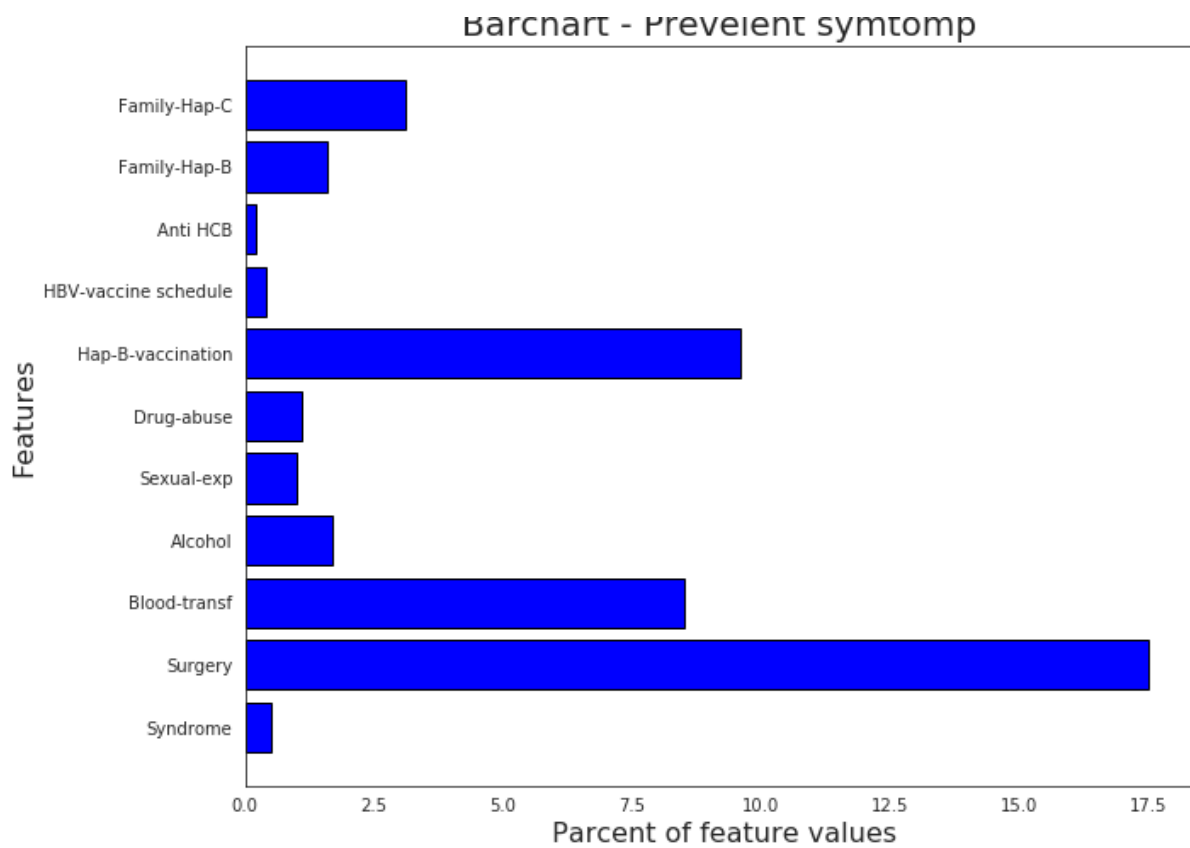


Figure 4.2.18 Hepatitis B and Hepatitis C Screening SSHMCH2018

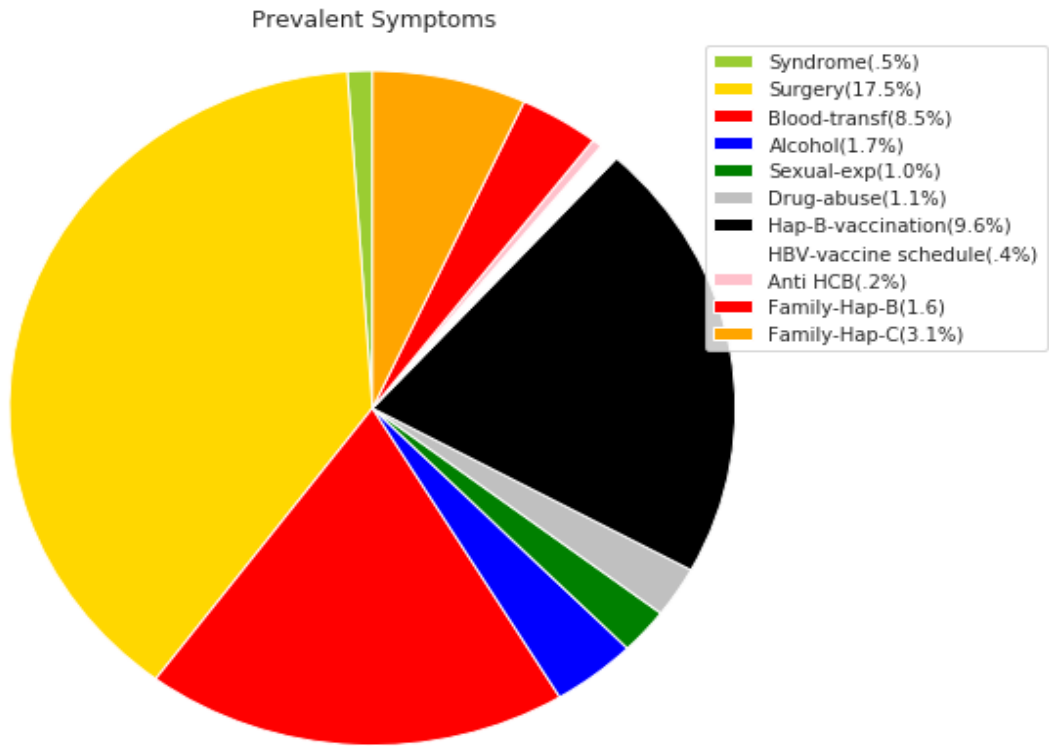


Figure 4.2.19 Hepatitis B and Hepatitis C Screening SSHMCH2018

Chapter 5

SUMMARY, CONCLUSIONS, RECOMMENDATIONS AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

Hepatitis is one of the most commonly known viral diseases in recent days. Bangladesh has drafted a national hepatitis plan (2017-2021) with a goal to reduce 90% of viral hepatitis related morbidity and 65% of hepatitis related mortality from Bangladesh by 2030. In July 2019, a team from WHO that included representatives from India's Institute of Liver & slimy Sciences ((LOS) and the Coalition for Global Hepatitis Elimination travelled to Bangladesh to review national data and strategic information systems for viral hepatitis. introduce the WHO Global Reporting System on Hepatitis (GOSH) and facilitate a workshop on use of strategic information for the national planning process. So, we need to make people more careful and aware about Hepatitis B & C risk factors so that they can go for a quick and better diagnosis and care. How can we make people careful? this question arose when we first thought about this research. Finally, we have come to a conclusion that we could work with Hepatitis B risk factors and that would be so much beneficiary for people. So, a model for predicting Hepatitis B & C risk factors using some different machine learning algorithm has been developed. Support Vector Machine (SVM) model, Random Forest Classifier have been used for prediction the Hepatitis B & C risk factors over our collected data and our dataset. Data collection was the main challenge for our research work.

5.2 Conclusions

The present work shows the significant job of Hepatitis B disease rates from the past season. Joining this factor in the model essentially improves the predict intensity of Hepatitis surveillance determining models. Increments in hepatitis B contaminations are being powered by the narcotic and heroin use pestilences that are grasping numerous networks over the US. Following quite a while of decreases in new cases, progress on hepatitis B avoidance has slowed down in the US overall and, in certain states, hepatitis B diseases have expanded as suggested.

The hepatitis B vaccine programmed in Bangladesh was found highly effective in preventing chronic hepatitis B virus (HBV) infection among children, suggests a new icddr study in collaboration with the Institute of Epidemiology, Disease Control and Research (IEDCR), Bangladesh, the U.S. Centers for Disease Control and Prevention, Atlanta (CDC) the World Health Organization (WHO) and the Stanford University, US. Simply like the admonition includes just as the ones recognized in this research. A model for identification of Hepatitis risk factor has been developed so that they can take quick and proper treatment for perfect and early overcome.

5.3 Recommendations

Machine Learning (ML) is the consistent examination of counts and real models that PC frameworks use to play out a specific task using unequivocal headings, contingent upon models and deriving. It is seen as intelligence. Machine Learning calculations collect a mathematical model reliant on test data, known as training data, to make desires or decisions without being explicitly altered to play out the work. Two different machine learning algorithms have been used to develop our model for predicting Hepatitis risk factor and for showing which is better in prediction over our collected dataset. The three algorithms are- Support vector machine which is also known as SVM, Random forest algorithm. While doing our research work over our dataset similar accuracy rate has been found from these different algorithms. Though it can vary from dataset to dataset of the accuracy of algorithms but for our dataset the accuracy rate is similar for the used algorithms. When we increase or decrease our test value the accuracy of Support Vector Machine gets higher than Random forest. In this way, we have a proposal for you to utilize or attempt Support Vector Machine in anticipating research work.

5.4 Implication for Further Study

In this research work a model for predicting Hepatitis risk factor has been developed using different machine learning algorithms. But in future we have a different plan with it. Like, in future we will predict the pattern of the Hepatitis risk factor using ML algorithms. I think it will be also more beneficiary for people to find the actual reason behind the Hepatitis fever and they will take the proper diagnosis and treatment.

REFERENCES

- [1] Mayo Foundation for Medical Education and Research , "www.mayoclinic.org," [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/hepatitis-b/symptoms-causes/syc-20366802>. [Accessed 5 10 2020].
- [2] WebMD LLC., "www.webmd.com," [Online]. Available: <https://www.webmd.com/hepatitis/digestive-diseases-hepatitis-b>. [Accessed 5 10 2020].
- [3] M. A. G. T. E. M. Khairy H. Morsya, "Clinical, laboratory, and virological characteristics of patients with positive hepatitis B surface antigen in Upper Egypt," *The Egyptian Society of Internal Medicine*, pp. 32-37, 2015.
- [4] healthline media, "www.healthline.com," [Online]. Available: <https://www.healthline.com/health/hepatitis-b>. [Accessed 8 10 2020].
- [5] www.iamat.org, "Mauritania Selective Vaccinations," [Online]. Available: <https://www.iamat.org/country/mauritania/risk/hepatitis-b>. [Accessed 10 10 2020].
- [6] www.dhakatribune.com, "dhakatribune," [Online]. Available: <https://www.dhakatribune.com/health/2017/07/28/10-million-bangladeshis-hepatitis-b>. [Accessed 15 10 2020].
- [7] www.pubmed.com, "Hepatitis B (Chronic): Diagnosis and Management of Chronic Hepatitis B in Children, Young People and Adults," [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25473721/>. [Accessed 15 10 2020].
- [8] www.nice.org.uk, "Guidance," [Online]. Available: <https://www.nice.org.uk/guidance/cg165/chapter/introduction>. [Accessed 20 10 2020].
- [9] W. H. Gerlich, "Medical Virology of Hepatitis B: how it began and where we are now," 2013.
- [10] ww.liver.org.bd, "HEPATITIS B," [Online]. Available: <https://liver.org.bd/hepatitis-b-2/>. [Accessed 2 11 2020].

- [11] health.ny.gov, "Background and Rationale for Hepatitis B," [Online]. Available: https://health.ny.gov/diseases/communicable/hepatitis/strategic/background_rationale_hepatitis_b.htm. [Accessed 8 11 2020].
- [12] www.health.ny.gov, [Online]. Available: https://health.ny.gov/diseases/communicable/hepatitis/strategic/background_rationale_hepatitis_b.htm. [Accessed 13 11 2020].
- [13] www.icddrb.org, "Hepatitis B vaccines for children found effective in Bangladesh," [Online]. Available: <https://www.icddrb.org/news-and-events/news?id=847>. [Accessed 15 11 2020].
- [14] www.who.int, "WHO urges countries to invest in eliminating hepatitis," [Online]. Available: <https://www.who.int/news/item/26-07-2019-who-urges-countries-to-invest-in-eliminating-hepatitis>. [Accessed 17 11 2020].
- [15] "VIRAL HEPATITIS," World health organization, 2016.

APPENDICES

We have put a lot of effort doing this research. While developing the prediction model of Hepatitis symptoms, we have learnt lots of things of which we did not have a proper knowledge before and we are still learning. Now we know how to use technology for saving human lives. We have used different algorithms of machine learning in this research for predicting Hepatitis fever that will help us to take early treatment before having any damages from Hepatitis. We are hopeful that the knowledge and experience we have gained while doing this research work will help us improve a lot in our future work resulting in better output.

Defense 2020

ORIGINALITY REPORT

24%	12%	9%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Saif Mahmud Khan Dourjoy, Abu Mohammed Golam Rabbani Rafi, Zerine Nasrin Tumpa, Mohd. Saifuzzaman. "Chapter 49 A Comparative Study on Prediction of Dengue Fever Using Machine Learning Algorithm", Springer Science and Business Media LLC, 2021 Publication	5%
----------	--	-----------

2	www.who.int Internet Source	2%
----------	---	-----------

3	Submitted to Daffodil International University	0%
----------	--	-----------