

Personalised Vaccination Prediction for Hepatitis B using Machine Learning

Vijayalakshmi. C¹, Dr S. Pakkir Mohideen²

*Department of Computer Science and Engineering*¹

*Department of Computer Applications*²

B.S.A. Crescent Institute of Science and Technology, Chennai, Tamilnadu, India^{1,2}

*vijic.edu@gmail.com*¹

*pakkirmoitheen@crescent.education*²

Abstract

Every newborn baby is mandatorily given three doses of Hepatitis B vaccination. The baby gets its first dose of vaccine at the time of birth. The second dose is given in between 1 to 2 months and the third dosage between 6 to 8 months. The very fact that Hepatitis B vaccination is given at birth specifies the necessity and importance of prevention of diseases caused by Hepatitis B. Infection by Hepatitis B can be the cause of several serious liver diseases. Also its complexity is further increased due to the fact that once vaccinated, the immune system doesn't hold the response to Hepatitis B surface antigens for a long course of time. One also has to consider that there are not many treatments for the diseases caused by the Hepatitis B viruses. For people who are tested to be prone to HBV attack are advised to have a booster vaccination to set right their immune system. Some people with less immunity or immune memory have to take up the entire vaccination cycle of three doses again instead of the booster dose. Predicting if a patient needs a booster dose or revaccination depends upon a lot of factors. This project, with the help of machine learning, aims to achieve perfect prediction of vaccination type the patient has to undergo and it also recommends the amount of dosage that is suitable and sufficient for the person. It aims to achieve personalized recommendation of booster vaccinations for millions of people around the world.

Key Words: Hepatitis B, HBV, HBsAg, anti-HBs, post booster, HBV DNA, vaccination

1. Introduction

Viral infections caused by Hepatitis B can be very dangerous to the liver. Based on the severity of the viral infection Hepatitis B infection can be categorized as acute or chronic. The transmission of Hepatitis B virus takes place through the exchange of infected person's bodily fluids or through blood. Approximately 257 million people throughout the world have a positive Hepatitis B surface antigen and thus are infected by Hepatitis B virus. Mostly from complications, hepatitis B resulted in 878000 deaths in 2015. Particularly, the health workers are more prone to this infection. The currently available precautionary measure is the vaccination against Hepatitis B. No other treatment is available as of now. If careless with the vaccination, it seriously affects the liver and thereby the life of the individual. Due to its severity and its vast spread prevalence, it is considered as one of the most life threatening infection all over the world.

The Hepatitis B infection leads to death by causing liver cirrhosis and liver cancer over the time. All sorts of Hepatitis B infections do lead to highly complicated liver issues which might cause death. Thus, prevention is better than the treatment in the case of Hepatitis B infection. Preventive measures counts down a few methods out of which Vaccination is the best and the most prevalent measure ensured for every individual. These vaccinations do not completely make the human immune system resistant to these harmful viruses for a life time. Thus, making sure humans are not affected by the HBV is much important. The vaccination against this deadly virus was introduced in 1982. Since then, it has been 95% successful in controlling the spread of chronic Hepatitis B disease [10-15].

The Hepatitis B virus has the capacity to stay alive for around 6 to 7 days outside the living organism. If a non-vaccinated person comes across this virus, it has the potential to infect actively. On an average, the incubation period of this virus is around 76 days and varies between 40 to 80 days. Once infected, this virus attack can be detected between 40 to 60 days of infection. During pregnancy, there is a probability to affect the child through the mother. The symptoms of infection would not be felt when it is in an acute stage. However, some may experience symptoms like abdominal pain, dark urine, vomiting, nausea, yellowing of eye and skin and fatigue. Unfortunately, a small proportion of affected people, can undergo liver failure and thereby death. Around 80 to 90 percentages of infants below the age of one are prone to develop chronic infections of Hepatitis B virus if not taken the vaccination cycle at the time of birth. 30 to 50 percentages of children between 1 to 6 years develop this infection. Less than 4 to 5 percent of adults have the risk of infection out of which 20 to 30 % of those adults develop liver cirrhosis and cancer [16-19].

The detection of Hepatitis B cannot be done by clinical people rather a laboratory confirmation is very much essential. Several types of blood tests can be done to the Hepatitis B infected people in order to diagnose and further follow ups. These tests mainly concentrate on the detecting the HBsAg Surface Antigen and Anti-HBs. Apart from vaccination, there is no other potential treatment for this infection. Even though medicines are available for chronically infected patients, preventive measures like vaccines are considered to be the best option against Hepatitis B [20-21]. However, these medicines do not kill the virus. Thus, once infected, the patient has to undergo treatment throughout his life. The Hepatitis B vaccination cycle consists of three doses through injection in the arm. For a newborn baby, the vaccination schedule is taking the first dose immediately at the time of birth, and the second dose during the first month and the final dosage at the sixth month of delivery. There must be at least one month gap between the first and second shot. The gap between the first and third shot must be 4 months. When children are vaccinated in this fashion, they are assured of 95% protection against Hepatitis B infection. Among the adults, there are few types of people, the following group of people will be prone to infection and thus are entitled as high risk group. They are the people taking drugs, people having many sex partners, healthcare workers and people exposed to blood and bodily fluids. People who have deficiency in blood or have undergone blood transfusion,

organ transplantation recipients and dialysis patients are also included in this group of individuals.

2. Related Work

ElkeLeuridan et al [1] mentioned that the long term prevalence of the Hepatitis B vaccination on a human immune system can be measured through four ways. These ways include calculating the endemicity of the vaccinated population in the particular region of the individual, the patient's anamnestic response and the serological test results taken recently. The endemicity of the region is calculated by taking to consideration the seroprevalence of HBsAg (Hepatitis B surface antigen) in that region. If it is greater than 8%, that region can be entitled as a region of high endemicity. If the percentages of seroprevalence vary between 2% to 8%, it is a region of intermediate endemicity. Further regions having less than 2% seroprevalence is termed as low endemicity region. Mei-Chu et al [2] have attempted to determine the dosage and number of vaccinations required for a person identified with negative Hepatitis B surface antigen and anti-HBs. They have mentioned two controversies that have to be effectively determined. One is to predict if the patient needs 1 booster dosage or has to be revaccinated with the complete three dosages. The other one is to predict the range of postbooster anti-HBs status that has to be achieved on receiving a booster dose. People with low immunity power or poor immunity memory are prone to rapid decline in their anti-HBs levels. Such people have to be identified and their booster dosage has to aim at achieving higher postbooster anti-HBs status. Usually boosters aim at achieving positive postbooster anti-HBs status that is $\geq 10\text{mIU/mL}$. But for low immune people, the postbooster anti-HBs status has to be elevated to protective levels that is $\geq 100\text{mIU/mL}$.

They have identified a few predictive factors that help us determine the above mentioned decisions. Those factors include serum glutamate pyruvate transaminase level, Body Mass Index (BMI) and the patient's sex. They have identified that the anti-HBs levels of the prebooster as the perfect predictive tool needed for this study. Using this prebooster anti-HBs status, the booster dosage can be predicted. Apart from this, the duration of effectiveness of the booster dose for that person will be predicted. The result of this study has identified a metrics. For a person who has taken complete Hepatitis B Vaccination of three doses and is found to have negative HBsAg and anti-HBs levels after twenty years of vaccination is supposed to have 1 booster dose to regain their positive postbooster anti-HBs status under prebooster anti-HBs level higher than 1 mIU/mL . HebaElrashidy et al [3], have found that children who are diagnosed with IDDM have and alarming risk of Hepatitis B infections. Thus, HBV vaccines are recommended in countries or areas where HBV is endemic. Diabetic persons have low and compromised immune system and so their responses to the vaccines are less compared to the people who are non-diabetic. This poor immune response may be linked with suppressing of the production of B-cells of the anti-HBs and any kind of defect in antigen uptake. Also the Human Leukocyte antigen is responsible for the unresponsiveness of the diabetic patients for the Hepatitis B vaccines.

Poorolajal et al [4] intended to review and assess the anamnestic responses of the immune system for the booster doses of the Hepatitis B infection for five years after the initial vaccination cycle for the people who initially had anti-HBs that is the hepatitis B surface antigen lesser than 10mIU/mL. From this study it can also be concluded that the vaccination provided by the initial vaccination cycle and one or two booster doses has its effect for around 20 years. The authors also conclude that even when the prebooster Hepatitis B surface antigen levels are undetectable, the postbooster can be an indicator of protection against low anti-HBs levels. Booster doses are a compulsory action for immune compromised patients who are undergoing hemodialysis, chemotherapy and HIV affected patients.

St Juliants et al [5] enquired about the combined hepatitis B vaccine. They are monovalent hepatitis B vaccine and hexavalent hepatitis B vaccine. The main aim is to reduce the number of injections at a visit and to reduce total number of injections as well and spending less time in doctor's visit. A hepatitis B component cannot be used for newborn immunization with combined vaccine since the non-hepatitis B components of combination vaccines have reduced immunogenicity in children less than six weeks of age. The monovalent hepatitis B vaccine must continue to be used as the birth dose. To the further research of theorem, they tend to evaluate when booster vaccinations are needed using hexavalent vaccines with a hepatitis B component, especially in immunization schedules with only three vaccinations in the first year of life with no booster vaccination.

Gavilanes F et al [6], studied about the lipid composition of hepatitis B. The portion of the protein components of HBsAg are exposed to the HBsAg lipid matrix is determined by photoactivable hydrophobic probe and pyrenesulfonylazide is used. It is labelled in both COOH-Terminal and NH₂-Terminal tryptic fragments. These are buried within the HBsAg lipids. The two major HBsAg protein are buried with the lipid matrix of the same particle. Residue 122-150 region is exposed to the aqueous environment with the antigenic residue.

Milich D et al [7] reported about experimental studies and chronic clinical in an attempt for the better understanding the work of HBeAg in natural infection. It is not required for viral assembly, infection or replication because the function of the hepatitis B e antigen is largely unknown. For the serum HBeAg may serve an immunoregulatory role in natural infection is suggested by clinical and experimental data. A target for inflammatory immune response is served by cytosolic HBeAg. The interaction between the HBeAg and the host during HBeAg infection is complexity. Yee JK [8] discovered an 88 base pair fragment in the core promoter of human hepatitis B virus contains a strong liver specific enhancer and a function promoter. It is much more active than the previously described HBV enhancer in expression of the linked bacterial gene expressed from heterologous promoters. The role of the virus is in the pathogenesis of hepatocellular carcinoma and hepatitis. Rehmann B et al [9], investigated the antiviral antibodies and specific cytotoxic T lymphocytes during acute viral hepatitis is completely cleared by hepatitis B. The sterilizing immunity frequently fails to occur after recovery to HBV from acute hepatitis and virus that can maintain traces with the decades of CTL response. The clinical recovery correlates with persistence of HBV DNA followed by

the strength of CTL response to HBV. The blood for many years after clinical recovery from acute hepatitis are often detectable by the traces of HBV.

3. Methodology

The infection of Hepatitis B virus has been a serious problem worldwide and vaccination is the effective and efficient way to curb this virus. So many countries like Japan and China have not yet assigned or following the vaccination programs that are followed worldwide. Therefore Universal vaccination should be the total eradication system. Some of the infected people are categorized as non-responders even after vaccination to the Hepatitis B vaccines. To tackle this problem, many methods are being proposed. To overcome this problem of non-responders, the inclusion of pre-s proteins in the third generation vaccination have been introduced.

For infants it is mandatory to give three doses hepatitis B vaccinations. The first dose is given at the time of birth and the second dose is given after one to two months and third dosage is between six to eight months. Ignoring of Hepatitis B vaccination causes major serious liver disease in future. The problem is people who are tested to be prone to HBV attack are not aware of booster vaccination they must take to set right their immune system. Some people with low immune memory or less immunity have to undergo entire vaccination cycle of three doses instead of the booster dose.

Here the solution is to predict whether the patient needs a revaccination or booster dose depends upon the plenty of factors. So this helps the patient to achieve the perfect prediction of vaccination type they have to undergo and it also say the amount of dosage that is sufficient and suitable for the person. So the amount of dosage and duration of the effect of the booster vaccination is known by the patients. The predicted treatments are identified by serological test result. This is safer and secure way to protect patients from HBV virus, and also to prevent the attacks on them for a greater extent. The solution for this problem is with the help of personal details and medical details the system recommends the type of vaccination treatment. In case of existing patient the medical details are instantly available in the database as reference. If the person is new to the system they have to fill the details for the result of type of vaccination treatment.

4. Proposed System

Hepatitis B virus infection may be either chronic or acute. Chronic is something long standing and acute is something which is self-limiting. Patients with self-limiting infection can do away with the infection gradually from weeks to months spontaneously. To clear the infection children are more prone than adults. To screen the presence of this virus, HBsAg surface antigen is used. It is the latest viral antigen that can be detected which appears during Hepatitis B infection. So the existing system predicts only if a patient is affected with hepatitis B virus or not. It will never show which type of vaccination or booster the patient must undergo. It has been very difficult to keep a track of their own treatments. The system keeps track of patients only during the vaccination cycle. It does not predict every required conditions of the patient who is affecting by

hepatitis B. Structured data is only handled by the existing system. The existing prediction system is ambiguous and broad. First the existing systems are made for the sake of richest people only so that they are able to pay for these kind of prediction systems. These systems recommend a fixed dosage for vaccination for all patients in general. It won't analyse the entire medical details of the patient.

In the proposed system, it aims to predict if the patient needs to take up booster vaccination or revaccination. It shows the exact type of treatments the particular patient must undergo in case they are affected by hepatitis B virus or prone towards it. These steps will take place after the prediction of hepatitis B which do already exists. Immune system might lose its retention capacity against hepatitis B. The system also predicts how long will the booster dosage prevails in the immune system. It is implemented to increase operational efficiency. It will easily find the recommended vaccination treatment with the help of serological tests and medical detail of the individual. It will be the entire pack of testing what exactly the people has to undergo because it includes the lifestyle, smoking habits, drinking habits, any valve replacements, or undergone any blood transfusion. These will help to find the result very accurately. And additionally the system displays remarks. It will show the further future treatments they need to be taken. It lists that the first dosage has to be taken in 1 month, the second dosage has to be taken in 3-4 month and the third dosage has to be taken in 6-8 months.

5. Implementation

The Hepatitis B Vaccination Predictor (HBVP) is a web application that is used by the doctors or medical people who put up camps or run a health care centre to predict the treatment that has to be undertaken by a patient who is affected by the Hepatitis B infection. This application suggests one out of the six vaccination treatments for the patient. For this prediction, the application takes in the details of the patient and provides a personalized recommendation of vaccination. The Prediction of Vaccination is done with the help of advanced and accurate machine learning algorithms. In this application, the Vaccination Prediction algorithm is used to predict the personalized vaccination treatment for the patient. The training dataset for the algorithm is obtained as the input from the patient is given to the model that has been trained with the dataset obtained using HBVP Algorithm.

The web application is designed with a front end that gets the input from the patient using forms. These details of the patient like their personal details, medical history and details, serological test results that confirm the infection of Hepatitis B virus is obtained. These details are stored in the database for further reference of the doctor or medical assistants. Further these details from the database are fetched by the R model to perform the machine learning approach on the input data to predict the personalized vaccination. The machine learning method is called from the web application through the Shiny web app framework. This framework provides a way to invoke the HBVP method from the web application itself without navigating to Rstudio. The Shiny web app framework enables an easy, user friendly web application for the Medical people.

The output of the method is also rendered in the form of a web page that can be invoked from the web application itself to give a continuous flow to the application.

5.1 Architecture Diagram

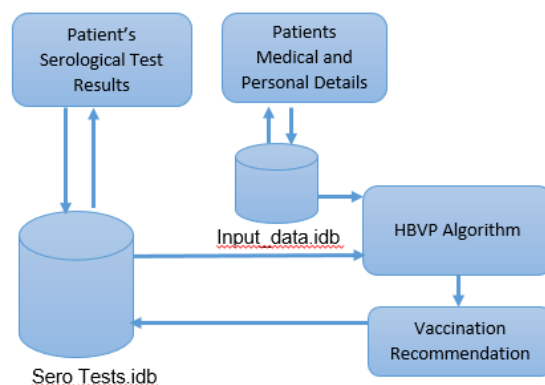


Figure 1: Architecture Diagram

The Architecture Diagram of the Personalised Hepatitis B Vaccination Prediction System is depicted in Figure 1. The Patient's Personal, Medical and Serological test details are requested to be entered by the patient or the medical people in the web page that is designed for this purpose. The Personal and Medical details are stored in the input_data.idb and the Serological test results are stored in SeroTests.idb. The HBVP Algorithm fetches the necessary details that are required for Vaccination Prediction from database and predicts the Treatment that has to be undertaken by that patient.

6. Predictor System

There are three modules in this Personalized Hepatitis B Vaccination Predictor System. They are the input, output and the database module.

6.1 Patient Details

The patient has to first enter his medical and personal details to provide personalized vaccination recommendation for that person. In order to achieve this, the web application presents a form to the patients requesting him/her to enter the details that are asked for in the form.

The following details are prompted by the web application:

- Personal Details like Patient's ID, Name, Age, Sex, Mobile Number and Place of living.
- Medical Details that reveal or track the patient's lifestyle and health related issues and surgeries are gathered.
- The Serological Test Results include the results of the HBsAg surface antigen test, Anti-HBs test, information if the patient has received the initial cycle of vaccination or not and their antibody response.

6.2 Recommended Vaccination

The Prediction of Vaccination Treatment is done through a machine learning approach using the HBVP algorithm. This machine learning method is implemented in R language using Rstudio. The call of the prediction method and the output is displayed in a web app that is integrated with the web application using the Shiny web app framework. The output page displays the Vaccination Treatment that is recommended and the remarks that specifies the interval of test or next dosage. This prediction can also be cross verified using the Decision Tree that is produced from the training dataset by applying the Decision Tree algorithm.

6.3 Sample Data

There are three databases that are employed to support the web application. The databases are managed using XAMPP software that includes Apache server, Mysql database and phpmyadmin.

- The main database is the base for the prediction is the training dataset that is being fed to the machine learning model for learning. This data is stored in the Hepatitis B Vaccination.csv file.
- The inputs of the patient are to be stored in a database for future reference and to retrieve the inputs for the predicted output.
 - The input_data.idb stores the personal and medical data of the patient.
 - The SeroTests.idb stores the Serological Test Results of the patient.

7. Machine Learning Algorithm

7.1 K-Nearest Neighbour

K-Nearest Neighbour Algorithm when compared to any other Machine Learning algorithm is the simplest algorithm for performing Supervised Learning. This algorithm unlike Vaccination Prediction Algorithm models the training data into groups and checks the most similar group that matches the test data. It can be used both for Classification and Regression problems. However it is being widely used for Classification Models.

The KNN algorithm doesn't make assumptions based on the training data that was available during model building and hence is said to be Non - Parametric Algorithm. It is also said to be a Lazy learner algorithm as it doesn't learn about the features and characteristics of the training data immediately instead the dataset is only stored and all the action that is to be performed on the dataset is done during the classification.

This algorithm is very robust in case the dataset is very noisy. Also it is very effective for large training datasets. Although it is the simplest algorithm, it does have complexity while determining the value of K and can be time consuming. While analysing about the cost of the algorithm, it is too costly as for every training data sample the distance has to be calculated between the incoming sample data and modelled training data samples.

7.2 Support Vector Machine

Support Vector Machine Algorithm popularly known as SVM, like KNN is used for both Classification and Regression Models and is highly used Classification related analysis. The

Supervised Learning algorithm plots data points on an n-dimensional space where n is considered to be the feature count and every feature is counted on the coordinates.

SVM is effective for high dimensional feature mapping and for doing non-linear classification. A hyperplane is modelled that classifies the testing dataset based on the training data. This Machine Learning algorithm aims at creating the best decision boundary for classification. The vectors used in this algorithm are mapped to the extreme points and these vectors are called as Support Vectors and hence the name of the Algorithm.

7.3 Hepatitis B Vaccination Prediction (HBVP) Algorithm

Machine learning algorithms are of three types. They are supervised learning, unsupervised learning and Reinforcement learning algorithms. The HBVP algorithm is a type of supervised learning algorithm. Supervised learning algorithms maps the set of inputs to the desired outputs based on the training data set that has been given to the algorithm. The HBVP algorithm builds several decision trees on the given training dataset and merges all the trees produced to create a forest of random decision trees. Such a forest is built with an idea to provide better learnability by combining several learning models. This type of method is called the bagging method.

The HBVP algorithm introduces more randomness to the learning model. This algorithm, by doing so, searches for the best feature instead of the important feature among the set of available features. This diversity makes the model a better one for supervised machine learning.

Decision trees are generated based on the threshold value provided as an input to the model. Randomness of the model can be further increased or decreased by adjusting the threshold value accordingly.

7.3.1 HBVP in R

- Step 1: Import the necessary Libraries.
- Step 2: Load the dataset using pandas.
- Step 3: Extract the dependent and independent variables.
- Step 4: Clean the dataset, Identity missing values and replace then.
- Step 5: Encode the categorical data using Label Encoder from Scikit Learn.
- Step 6: Split the Dataset into 80% training and 20% testing dataset using train_test_split from sklearn.model_selection.
- Step 7: Feature Scale the dataset using StandardScalar from sklearn.preprocessing.
- Step 8: Perform HBVP algorithm on the test and training dataset.
- Step 9: Evaluate the model by calculating the Confusion Matrix, Accuracy, Precision, Recall and F1 Score.
- Step 10: For visual understanding, plot the confusion matrix as a HeatMap using Matplotlib.

7.3.2 Accuracy of HBVP Algorithm

The accuracy of HBVP Algorithm can be viewed by displaying the Confusion Matrix The class error for all the classes is observed to be very minimal. Thus, the HBVP Algorithm achieves highest correctness of Predicted Treatment.

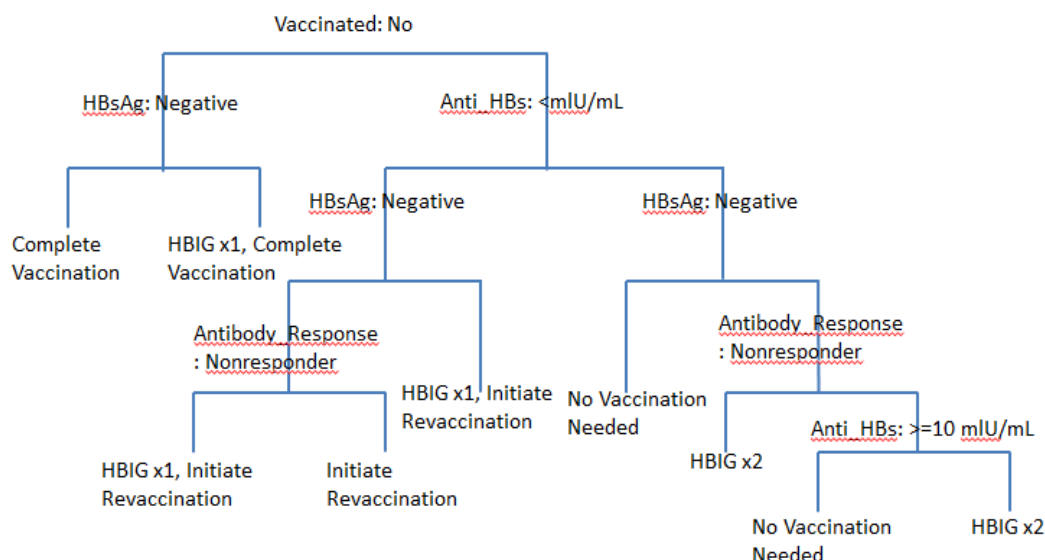


Figure 2: Decision Tree of the Training Dataset

The predictions made by the model can further be cross checked with the Decision Tree that is built using the Decision Tree algorithm on the same training dataset. The Decision Tree that is built using the training dataset is shown in Figure 2. The predictions made by the HBVP Algorithm is verified using this Decision Tree and the predictions are found to be highly precise.

8 .Experimental Results and Discussions

Based on the training dataset that is fed into the HBVP algorithm for supervised learning, the following conclusions and inferences have been made. The age of the patients has been categorized based on their ranges. The categories range from 0 years, 0 to 20 years, 20 to 55 years and beyond 50 years. The number of patients among those who were taken into consideration for the training of the Vaccination Prediction algorithm had received one of the six treatments for their medical details of the patient during their vaccination treatment. The Patient Distribution based on Vaccination Treatment has been depicted in Table 1. 45 Patients and 184 Patients out of the 1064 observations have been given Complete Vaccination and HBIG x1, Complete Vaccination respectively. Around 139 are given HBIG x2 dosage. Initiate Revaccination and HBIG x1, Initiate Revaccination are done for 137 and 106 patients respectively. Out of the 1064 patients, the highest percentage of people is immune to Hepatitis B infection and need not get a booster dose or complete vaccination cycle. These healthy people constitute around 453 of the total 1064 observations.

Table 1: Patient Distribution based on Treatment

Vaccinated	Antibody_Response	HBsAg	Anti_HBs	Treatment
Yes	Responder	Unknown	Unknown	No Vaccination Needed
Yes	Non-Responder	Positive	Unknown	HBIG x2
Yes	Non-Responder	Unknown	Unknown	HBIG x2
Yes	Non-Responder	Negative	Unknown	No Vaccination Needed
Yes	Unknown	Positive	<10 mIU/mL	HBIG x1, Initiate Revaccination
Yes	Unknown	Unknown	<10 mIU/mL	HBIG x1, Initiate Revaccination
Yes	Unknown	Negative	<10 mIU/mL	Initiate Revaccination
Yes	Unknown	Unknown	<10 mIU/mL	Initiate Revaccination
No	Unknown	Positive	Unknown	HBIG x1, Complete Vaccination
No	Unknown	Unknown	Unknown	HBIG x1, Complete Vaccination
No	Unknown	Negative	Unknown	Complete Vaccination

8.1 K-Nearest Neighbour Evaluation

The Confusion Matrix for K-Nearest Neighbour Algorithm is given in Table 2

Table 2: Confusion Matrix for K-Nearest Neighbour Algorithm

	NVN	IRV	HBIG x1, CV	HBIG x1, IRV	CV	HBIG x2
NVN	99	5	0	3	1	5
IRV	5	27	0	2	0	0
HBIG x1, CV	4	3	28	2	2	7
HBIG x1, IRV	7	4	3	13	0	0
CV	4	0	3	1	3	0
HBIG x2	15	1	2	2	0	15

NVN – No Vaccination Needed

IRV – Initiated Re-Vaccination

ICV – Initiated Complete Vaccination

CV – Complete Vaccination

8.2 Support Vector Machine Evaluation

The Confusion Matrix Support Vector Machine is given in Table 3

Table 3: Confusion Matrix for Support Vector Machine

	NVN	IRV	HBIG x1, CV	HBIG x1, IRV	CV	HBIG x2
NVN	110	3	0	0	0	0
IRV	8	26	0	0	0	0
HBIG x1, CV	0	0	46	0	0	0
HBIG x1, IRV	26	1	0	0	0	0
CV	0	0	11	0	0	0
HBIG x2	15	5	0	0	0	15

8.3 HBVP Evaluation

The Correctness of the HBVP Algorithm can be viewed by displaying the Confusion Matrix in Table 4.

Table 4: Confusion Matrix for HBVP Algorithm

	N V N	IR V	HBIG x1, CV	HBIG x1, IRV	C V	HBIG x2
NVN	113	0	0	0	0	0
IRV	0	34	0	0	0	0
HBIG x1, CV	0	0	46	0	0	0
HBIG x1, IRV	0	2	0	24	0	1
CV	0	0	2	0	9	0
HBIG x2	3	0	0	0	0	32

8.4 Result Analysis

The Heatmap that represents the Confusion Matrix for K-Nearest Neighbour in a normalized form is given in Figure 3.

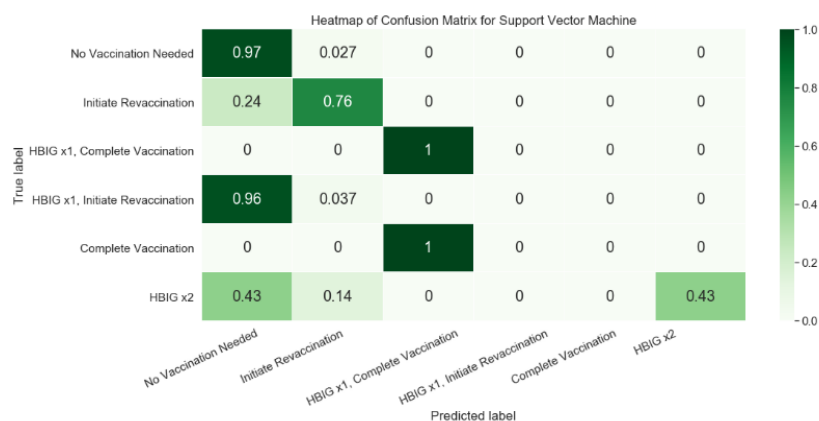
**Figure 3:** Heatmap of Confusion Matrix for K-Nearest Neighbour

Further the Accuracy, Precision, Recall, F1-Score, Support represented in Table 5. Thus, the Accuracy of K-Nearest Neighbour Algorithm is found to be 70%.

Table 5 : Performance Metrics for K-Nearest Neighbour Algorithm

	precision	recall	f1-score	support
1	0.74	0.88	0.80	113
2	0.68	0.79	0.73	34
3	0.78	0.61	0.68	46
4	0.57	0.48	0.52	27
5	0.50	0.27	0.35	11
6	0.56	0.43	0.48	35
accuracy			0.70	266
macro avg	0.64	0.58	0.60	266
weighted avg	0.69	0.70	0.68	266

The Heatmap representation for Support Vector Machine is shown in Figure 4.

**Figure 4:** Heatmap of Confusion Matrix for Support Vector Machine

Further the Accuracy, Precision, Recall, F1-Score, Support are represented in Table 6. Therefore, the Accuracy of Support Vector Machine is found to be 74% and hence it is not advisable to consider this Model for predicting the Personalized Hepatitis B Vaccination dosages for the patients.

Table 6: Performance Metrics for Support Vector Machine

	precision	recall	f1-score	support
1	0.69	0.97	0.81	113
2	0.74	0.76	0.75	34
3	0.81	1.00	0.89	46
4	0.00	0.00	0.00	27
5	0.00	0.00	0.00	11
6	1.00	0.43	0.60	35
accuracy			0.74	266
macro avg	0.54	0.53	0.51	266
weighted avg	0.66	0.74	0.67	266

Further to enhance the readability and understandability, the HBVP Algorithm is represented using Seaborn Heatmap in Figure 5. By doing so, the values represented by Confusion Matrix will also get normalized for a better understanding.

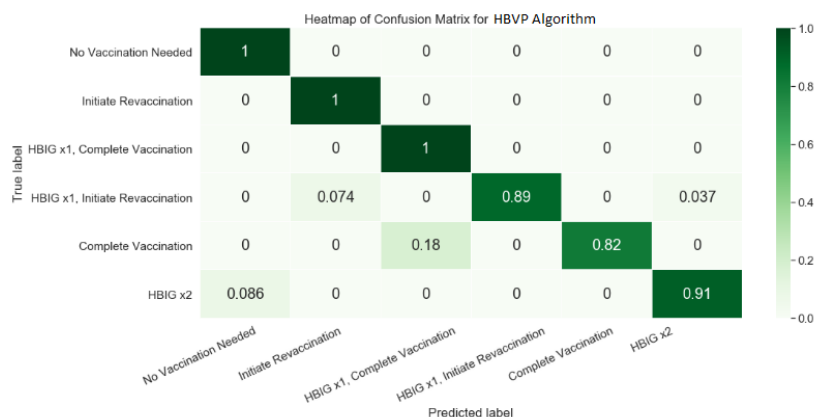


Figure 5: Heatmap of Confusion Matrix for HBVP Algorithm

The Performance of this HBVP Algorithm is measured by calculating the Accuracy, Precision, Recall, F1-Score, Support represented in Table 7. Thus, we can conclude that the Accuracy of HBVP Algorithm for Predicting the Vaccination is 97%.

Table 7: Performance Metrics for HBVP Algorithm

	precision	recall	f1-score	support
1	0.97	1.00	0.99	113
2	0.94	1.00	0.97	34
3	0.96	1.00	0.98	46
4	1.00	0.89	0.94	27
5	1.00	0.82	0.90	11
6	0.97	0.91	0.94	35
accuracy			0.97	266
macro avg	0.97	0.94	0.95	266
weighted avg	0.97	0.97	0.97	266

Several Graphs can be derived from the training dataset for further analysis and inferences. These graphs can be plotted in R using the plot() function that has to be installed from lattice package and imported from the library using library(lattice). The Plot function requires two parameters. These are the columns that have to be taken into consideration for building the bar graph in the X and Y axis. The syntax would be to mention the column that has to be depicted in the Y axis followed by a '~' sign and then the X axis column.

Figure 5.shows the plot against the six treatments and the three age categories namely >55, 0-20 ad 20-55.

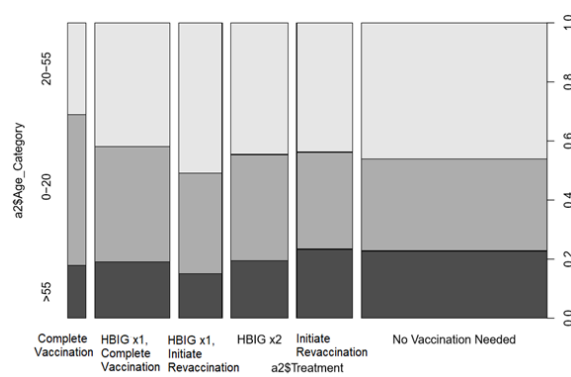


Figure 5: Treatment vs Age Category

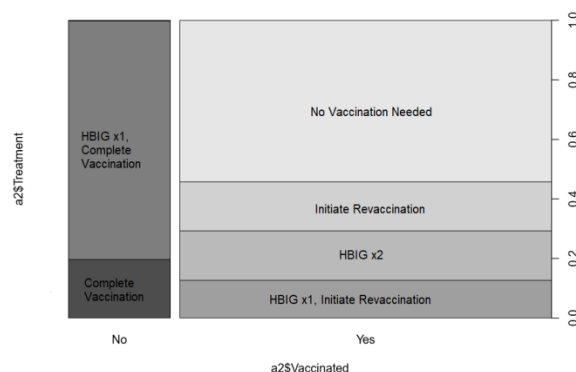


Figure 6 : Treatment vs Vaccinated

The Figure 6 shows that people who were not Vaccinated at the time of their birth, have to compulsorily undergo Complete Vaccination Cycle. Whereas for the people who have completed their three dosage cycle need not undergo Complete Vaccination but based on their Antibody Response, HBsAg and their Anti-HBs levels, they either have to undergo Revaccination with or without HBIG or doesn't have to get Vaccinated.

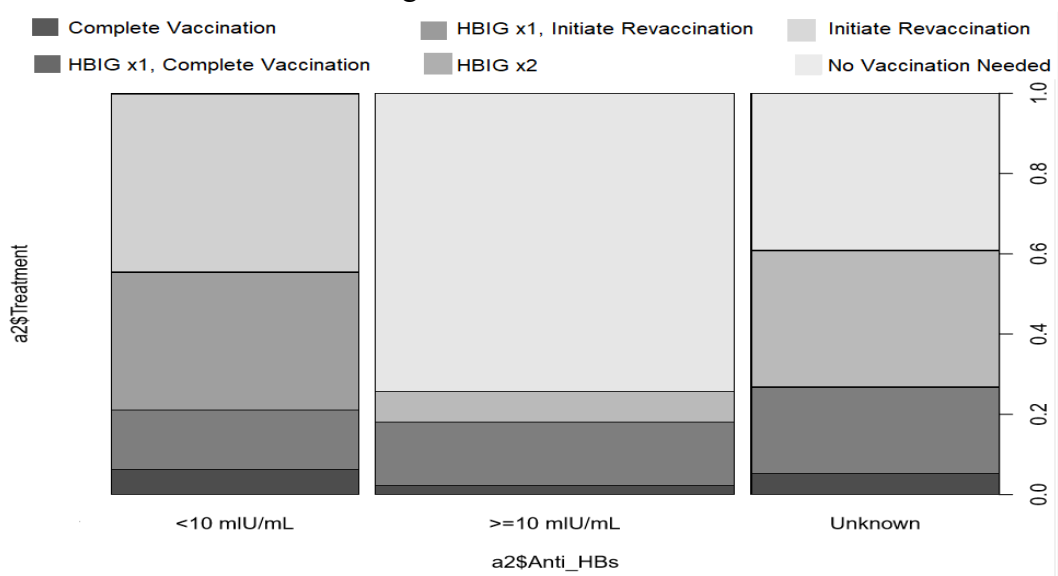


Figure 7: Treatment vs Antibody Response

Figure 7 shows that, a Known Responder to Antibody Response doesn't have to get vaccinated. In case the person is a Known Non-Responder, the patient has to get his HBIG x2 dose or need not be vaccinated by considering his HBsAg and Anti-HBs levels. For a patient whose Antibody_Response is Unknown, then based on the HBsAg and Anti-HBs levels, Revaccination or Complete Vaccination with or without HBIG dose has to be initiated.

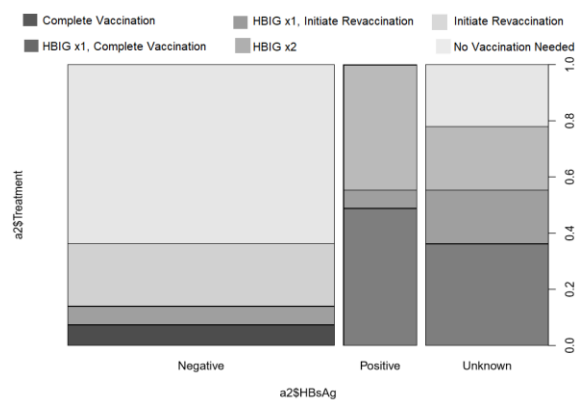


Figure 8: Treatment vs HBsAg

Figure 8 show that, when HBsAg is positive, then that patient has to take HBIG dose. Based on the patients Antibody Response and Anti-HBs levels, Revaccination or Complete Vaccination has to be followed by the patient along with HBIG dosage. When HBsAg is negative, the patient doesn't have to take in HBIG, but based on the other two serological factors, Complete Vaccination or Revaccination or in few cases No Vaccination is required. If their HBsAg is Unknown, other serological factors have to be taken into consideration before deciding the Treatment.

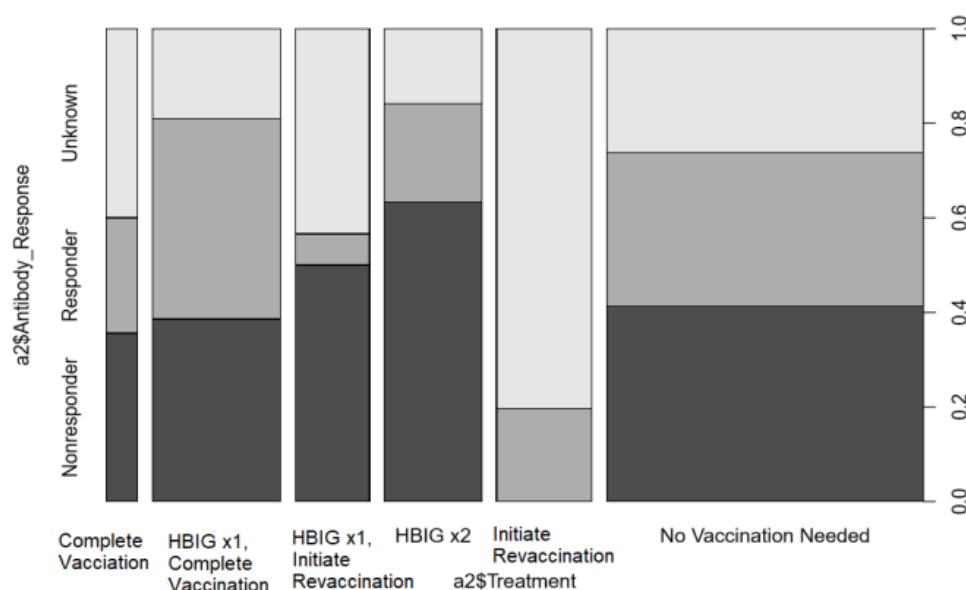


Figure 9: Treatment vs Anti-HBs

Figure 9 show that when the patient's Anti-HBs levels is $<10\text{mIU/mL}$, Revaccination has to be made compulsory. Considering the HBsAg levels of the patient, HBIG dosage should also be given along with the Revaccination dose. If the Anti-HBs levels are Unknown or $\geq 10\text{mIU/mL}$, the Treatment should be predicted based on the other Serological factors.

Table 8 : Comparison Result

		Accu racy	Precision	Recall	F1 Score
K-Nearest Neighbour.	Micro Avg.	70%	64%	58%	60%
	Weighted Avg.		69%	70%	68%
Support Vector Machine	Micro Avg.	74%	54%	53%	51%
	Weighted Avg.		66%	74%	67%
HBVP Algorithm	Micro Avg.	97%	97%	94%	95%
	Weighted Avg.		97%	97%	97%

The comparison of the Algorithm is measured by calculating the Accuracy, Precision, Recall, F1-Score, represented in Table 8.

8. Conclusion and Future Enhancement

Hepatitis B is a significant health problem of public in India, yet there is no proper awareness among the people. The major disease cases progress very silently and suddenly the patients will be seen in advanced or critical stages. With the drugs available currently, complete cure is not at all possible. By prolong therapy, the aim is to long term suppression of the hepatitis B virus, which itself can lead to poor treatment adherence with the prohibitive cost of treatment. Ultimately, this underlies the spread of infection. The present emphasis should be on aggressive vaccination strategies in the population, especially for high risk groups, tribes and health education of general and high risk regarding the lifestyle, early disease detection, preventive measures and proper adherence to drugs. Effective strategies and methodologies are required to improvise efficiency of vaccination for the non-responders. Finally, the rules are ensured about the timing and method booster injection of responders. Response to Anti-HBs gradually decreases after a single vaccination course. Third generation vaccines have been much effective anti HBs response of the non-responders, but the data on the efficacy is limited. The logistics are not yet adequate for proper disease control. The guidelines are much needed for the patients to know about the necessity and timing.

This system can be enhanced further by adding more criteria for predicting even more accurately. Several countries have not yet introduced this system. In the near future, with the expansion of the population in the world, it would be easier to add more criteria to make it more accurate and takes the prediction to the next level to predict the vaccination or booster. The system can also work for the worldwide with huge amount of data sets. Since people around the world may have different skin texture, lifestyle, daily habits which varies from the people in India. And the Africans have very strong immune memory and immunity power they won't get easily prone to hepatitis B virus. They need different sets of vaccination treatments. Thus the system can be made flexible and helpful for the patients to know about the vaccination treatments.

References

- [1] ElkeLeuridan, Pierre Van Damme, Hepatitis B and the Need for a Booster Dose, *Clinical Infectious Diseases*, Volume 53, Issue 1, Pages 68–75, 1 July 2011
- [2] Lu, I-Cheng, Mei-Chu Yen Jean, Chwee Lin, Wei-Hung Chen, D B Perng, Chih-Wen Lin and H Y Chuang. "Predictive factors for anti-HBs status after 1 booster dose of hepatitis B vaccine." *Medicine* (2016).
- [3] HebaElrashidy, Ashraf Elbahrawy, Gamal El-Didamony, Mohamed Mostafa, Nilly M George, Ahmed Elwassief, Abdel-GawadSaeid Mohamed, AmrElmestikawy, Mohamed HanafyMorsy, AlaaHashim& Mohamed Ali Abdelbasseer (2013) "Antibody levels against

- hepatitis B virus after hepatitis B vaccination in Egyptian diabetic children and adolescents, *Human Vaccines & Immunotherapeutics*, 9:9, DOI: 10.4161/hv.25426, 2002-2006
- [4] Poorolajal J, Hooshmand E. “Booster dose vaccination for preventing hepatitis B”. *Cochrane Database of Systematic Reviews* 2016, Issue 6. Art. No.: CD008256. DOI: 10.1002/14651858.CD008256.pub3. (2016)
- [5] Viral Hepatitis Prevention Board. Combined hepatitis B vaccines. *Viral Hepatitis Prevention Board meeting*, St Julians, Malta, October 22-23, 2001. *Viral Hepatitis* 2002;
- [6] Gavilanes F, Gonzales-Ros A, Peterson D. Structure of hepatitis B surface antigen: characterization of the lipid components and their association with the viral proteins. *J Biol Chem.* 1982;257:7770–7777.
- [7] Milich D, Liang TJ. Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. *Hepatology.* 2003;38:1075–1086.
- [8] Yee J. A liver-specific enhancer in the core promoter region of human hepatitis B virus. *Science.* 1989;246:658–670.
- [9] Rehmann B, Ferrari C, Pasquinelli C, Chisari FV. The hepatitis B virus persists for decades after patients’ recovery from acute viral hepatitis despite active maintenance of a cytotoxic T-lymphocyte response. *Nat Med.* 1996;2:1104–1108.
- [10] Govindaraj, Dr. (2021). An Effectual Plant Leaf Disease Detection using Deep Learning Network with IoT Strategies. *Annals of the Romanian Society for Cell Biology.* 25. 8876-8885
- [11] Mast, E.E., Mahoney, F.J., Alter, M.J., Margolis, H.S. Progress toward elimination of Hepatitis B transmission in the United States. *Vaccine.* 1998;16:S48–S51.
- [12] Govindaraj, Dr. (2021). Face Recognition based on Spatio Angular Using Visual Geometric Group- 19 Convolutional Neural Network. *Annals of the Romanian Society for Cell Biology.* 25. 2131-2138
- [13] Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for Trial Sequential Analysis (TSA). ctu.dk/tsa/files/tsa_manual.pdf 2016.
- [14] Govindaraj, Dr. (2021). Skin Lesion Detection Based on Fuzzy Logic. *International Journal of Innovative Technology and Exploring Engineering.* 8. 516. 10.35940/ijitee.J8826.0881019.
- [15] Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JP, Thabane L, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses. *International Journal of Epidemiology*; 38(1):276–86, 2009.
- [16] Lee C, Gong Y, Brok J, Boxall EH, Gluud C. Hepatitis B immunisation for newborn infants of hepatitis B surface antigen-positive mothers. *Cochrane Database of Systematic Reviews*, Issue 2[DOI:10.1002/14651858.CD004790.pub2], 2006.
- [17] Govindaraj, Dr. (2021). Identification of Bone Fragmentation in X-Ray images using CDA. *International Journal of Innovative Technology and Exploring Engineering.* 8. 2121. 10.35940/ijitee.I7851.078919

- [18] M. Tamilselvi and G. Ramkumar, "Non-invasive tracking and monitoring glucose content using near infrared spectroscopy," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2015, pp. 1-3, doi: 10.1109/ICCIC.2015.7435728
- [19] Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. *BMC Medical Research Methodology*;14(120):1–13, 2014.
- [20] Govindaraj, Dr & E, Logashanmugam. (2018). Study on impulsive assessment of chronic pain correlated expressions in facial images. *Biomedical Research*. 29. 10.4066/biomedicalresearch.29-18-886
- [21] Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.