



Computer Science & IT Research Journal  
P-ISSN: 2709-0043, E-ISSN: 2709-0051  
Volume 2, Issue 1, P.16-32, November 2021  
DOI: 10.51594/csitrj.v2i1.275  
Fair East Publishers  
Journal Homepage: [www.fepbl.com/index.php/csitrj](http://www.fepbl.com/index.php/csitrj)



## THE PREDICTION OF HEPATITIS B VIRUS (HBV) USING ARTIFICIAL NEURAL NETWORK (ANN) AND GENETIC ALGORITHM (GA)

Douglas Ibrahim<sup>1</sup> Asabe Sandra Ahmadu<sup>2</sup> Yusuf Musa Malgwi<sup>3</sup> & Bamanga Mahmud Ahmad<sup>4</sup>.

<sup>1</sup>Department of Computer Science, Adamawa State College of Agriculture, P.M.B. 2088 Ganye, Adamawa State, Nigeria.

<sup>2,3</sup> Department of Computer Science, Modibbo Adama University, Yola, P.M.B. 2076 Yola, Adamawa State, Nigeria.

<sup>4</sup>Department of Computer Science, Federal University of Lafiya, Nasarawa State, Nigeria.

\*Corresponding Author: Douglas Ibrahim

Corresponding Author Email: [ddtekk@gmail.com](mailto:ddtekk@gmail.com)

Article Received: 17-09-21

Accepted: 22-10-21

Published: 30-11-21

**Licensing Details:** Author retains the right of this article. The article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licences/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the Journal open access page

### ABSTRACT

The hepatitis B virus causes a liver infection called hepatitis B (HBV). It might be severe and go away on its own. Some kinds, however, can be persistent, leading to cirrhosis and liver cancer. HBV can be transmitted to others without the individual being aware of it; some persons have no symptoms, while others only have the first infection, which later resolves. Others develop a chronic illness as a result of their condition. In chronic cases, the virus attacks the liver for an extended period of time without being detected, causing irreparable liver damage. The manual approach has a high number of errors due to human decision-making, and visual screening is time-consuming, tiresome, and costly in terms of manpower. To predict the occurrence of Hepatitis virus (HBV), this research project thesis suggested an algorithm; Artificial Neural Network (ANN), and genetic algorithm (GA). To develop, evaluate and validate the performance of the model developed using ANN. Medical records of nine hundred patients were collected in the Northern Senatorial District (Mubi South), Central Senatorial District (Hong), and Southern Senatorial District (Ganye) regions

of Adamawa state, Nigeria. Three hundred (300) patient records were collected from each general hospital, for a total of 900 patient records. The success of the proposed technique is demonstrated when ANN is paired with GA, Accuracy (66.30%), Specificity (66.33%), and Sensitivity (77.53%) were discovered. In this study, hepatitis B virus (HBV) was predicted using Artificial Neural Network (ANN) classifier and Genetic algorithm optimization tool were used to select the features that are responsible for hepatitis B virus (Sex, Loss of Appetite, Nausea and vomiting, Yellowish skin and eye, Stomach pain, Pain in muscles and joint). The prediction was found to have acceptable performance measures which will reduce future incidence of the outbreak and aid timely response of medical experts.

**Keywords:** Hepatitis B Virus (HBV), Prediction, Features, Classification.

---

## INTRODUCTION

Hepatitis is a liver inflammation caused by viruses and other factors like alcohol consumption, certain drugs, and trauma. Millions of Nigerians are affected by it in various kinds. Although many cases of hepatitis may not pose a major health risk, infection with some hepatitis viruses can develop chronic (long-term) and lead to liver failure and death in some situations. The liver is an important organ that helps the body digest nutrients, filter blood, and fight infections. The liver's function can be harmed when it is damaged. Excessive alcohol intake, pollution, some medicines, and certain medical conditions can all cause hepatitis. Hepatitis, on the other hand, is almost always caused by a virus (Tseng & Kao, 2013).

The world Hepatitis submit was convened to build on the World Health Assembly 67.6 resolution, which asked WHO member states to develop and implement national viral hepatitis strategies, as well as calling on WHO to examine the feasibility of eliminating hepatitis B and C with view to setting global targets and develop a monitoring mechanism. The submit sought to build momentum towards development of comprehensive national plans within the framework of WHO 2016 – 2021 viral hepatitis strategy. (World Hepatitis Submit, 2015).

Viral hepatitis can be acute or chronic, with serious symptoms and long-term consequences. Hepatitis B virus (HBV) infection is estimated to have infected over 2 billion people worldwide, with 350–400 million chronic carriers (Ganem & Prince, 2014). HBV causes acute and chronic liver disorders and is responsible for an estimated 1 million deaths each year around the world. Its prevalence varies over the world, although it is most prevalent in tropical areas. In Sub-Saharan Africa, it is believed that 5–15% of individuals are chronically infected with HBV. HBV-related cirrhosis and hepatocellular carcinoma provide a 15–25% risk of dying early in adulthood (WHO, 2018).

HBV is predominantly concentrated at birth and early childhood perinatal transmission from an infected mother to her infant in prevalent in areas of high endemicity where at least 8% of the population is chronic HBV carrier. HBV is transmitted to adults by contact with infected blood and body fluids such as sperm, vaginal fluids, and saliva. As a result, unscreened blood and its products, sexual activities, the use of infected or insufficiently sanitized devices, and the sharing of sharp items as may occur during some traditional or cultural rites, such as local circumcision,

are common means of transmission. It could also happen as a result of other iatrogenic (Hou and Liu, 2015).

Hepatitis B virus (HBV) is a DNA virus belonging to the Hepadnaviridae family that causes hepatitis B illness. Hepatitis B is one of the world's most frequent diseases and a serious public health issue. According to the most recent WHO estimate, 2 billion persons worldwide have serologic evidence of previous or current HBV infection, with 350 million chronically infected and at risk of HBV-related liver disease. It is 50 to 100 times more infectious than HIV and ten times more infectious than the hepatitis C virus (HCV), and many carriers are unaware that they are infected. It is a major cause of liver disease, with chronic HBC infection being one of the most common. HBV infection rates range from 2% in developed countries to roughly 8% in underdeveloped countries, with sex, age, and socioeconomic level serving as key risk factors for infection. (Adoga, Bashayi, and Zungwe, 2010).

The major mechanism of transmission frequently correlates with the degree of HBV endemicity. Many countries' health and economies are severely impacted by the sickness, and the intensity of the disease is very changeable and often unanticipated. Because the infectious dosage is so low, actions like sharing both a brush and a razor blade can spread infection. (Odusanya, Wellens, and Weil, 2015).

The hepatitis B virus (HBV) is a bloodborne and sexually transmitted infection that is transmitted by infected blood or other bodily fluids (Saliva, sweat, semen, vaginal secretions, breast milk, urine and faeces). When using the same syringe as an infected individual, from blood transfusions, from mother to child during childbirth, during medical procedures, and during sexual intercourse, transmission can occur. Hepatitis B virus and HIV have comparable transmission methods (Chang, 2017).

There are four recognized modes of transmission currently (Viral Hepatitis Board, 1996):

- i) Prenatally (from mother to kid)
- ii) Through contact with an infected individual
- iii) Through sexual intercourse
- iv) Contact with contaminated blood or bodily fluids etc.

Transmission may occur indirectly via contaminated surfaces and other objects such as tooth brushes, baby bottles, razors, eating utensils, and hospital equipment by contact with mucous membranes for open skin breaks because HBV can remain stable and infections on environmental surfaces for at least 7 days (Willey, Sherwood, and Woolverton, 2018).

Most people who become infected with HBV are able to eliminate the virus from their bloodstream and develop immunity within 6 months of illness. Chronic hepatitis B infection is defined as those who have not eradicated the virus after six months. People who catch chronic infections as children are at a 25% chance of developing liver cancer or cirrhosis as a result of HBV. Furthermore, 8% to 10% of people in underdeveloped nations' general populations become chronically infected, with the majority of infections occurring during childhood (Weinbum, 2018).

Nigeria is one of the countries with a high prevalence of HBV infection. HBV is thought to have been exposed to approximately 75% of the Nigerian population at some point in their lives. Despite

the fact that hepatitis B vaccine is highly effective in avoiding HBV infection and subsequent acute and chronic liver illness, the virus remains a big concern in Nigeria, as numerous workers have documented (Luka, Ibrahim, and Iliya 2018).

### **Objectives and Significance of the Study**

The aim of the study is to extract the optimal feature of hepatitis B using Genetic Algorithm (GA) and Artificial Neural Network (ANN) for classification, and the objectives are: -

- i) To investigate the schematic features and select the relevant feature that is responsible for hepatitis B virus using Genetic Algorithm (GA).
- ii) To preprocess, train and test the developed model with varying dataset.
- iii) To Classify and develop an Artificial Neural Network (ANN) techniques for hepatitis B virus (HBV).
- iv) To evaluate the performance of the model developed using Artificial Neural Network (ANN).
- v) To validate the model using Artificial Neural Network (ANN).

Machine-learning techniques, such as artificial neural networks (ANN), are being used by health-care organizations to improve care delivery while lowering costs. The use of artificial neural networks (ANNs) to forecast outcomes is well-known; nevertheless, ANNs are increasingly being employed to guide health-care management decisions. As developed-country health-care systems shift to a value-based, patient-centered model of care, we confront new challenges in terms of improving the structure and administration of health-care delivery, such as enhancing process integration in care delivery for patient-centered chronic illness management. Artificial intelligence is at the crossroads of emerging technologies, with the ability to provide cost-effective and appropriate health care in real time, as well as manage effective and efficient communication among interdisciplinary stakeholders. (Kuziemy, 2016).

Multiple stakeholders, such as clinicians, managers, and senior executives, must collaborate to make clinical (e.g. diagnosis, prediction, treatment and therapy, medication prescription and administration) and non-clinical (e.g. budget, resource allocation, technology acquisition, service additions/reductions, strategic planning) decisions. Large quantities of frequently obtained digital data are being analyzed by health care companies in order to improve service and cut costs. Analyzing clinical, financial, and operational data to answer concerns about program performance and generate forecasts is one example mentioned (Awowale, 2017).

ANN, as described by Agatonovic-Kustrin and Beresford (2010), collect knowledge by finding patterns and correlations in data and “learn” via experience. According to the authors, an artificial neural network learns by optimizing its inner unit connections to reduce mistakes in predictions and achieve a desired degree of accuracy. After the model has been trained and tested, more data can be entered into it. Policies that encourage transparency and sharing of fundamental datasets across the public and private sectors can help to boost innovation and research productivity.

For training data size and parameter optimization of grey forecasting models, the Genetic Algorithm (GA), which is one of the most frequently utilized meta-heuristic optimization approaches, is used. The application findings show that optimizing parameters and training data

size, as well as using a rolling mechanism, greatly improves the grey model's forecasting ability (IBM Corporation, 2013). This research will show how artificial neural networks and genetic algorithms can be used to improve medical outcomes at various levels.

### **Related Works of the Literature**

Hepatitis B virus infection is endemic in Nigeria, according to Awoleke (2012). Despite this, routine prenatal screening and treatment are not common in the country. The author determines the disease's prevalence and pattern in the obstetric community. Information regarding the patients' socio-demographic characteristics, risk factors such as history of blood transfusion and surgical procedures, and laboratory screening findings were acquired using a pre-structured questionnaire. The SPSS version 16 statistical software tool was used to encode and analyze the data. Simple percentages and when appropriate, the chi-square test were used in the analysis. A p. value of less than 0.05 was considered significant.

Adebola, Aki, and Muhammad (2016) used a cross-sectional survey among the general population and used a multi-stage sampling technique to determine the seroprevalence of hepatitis B infection in Nigeria. Questionnaires were used to collect data on demographic, social, and behavioral factors, and blood samples were tested for HBV. We used descriptive, bivariate, and multivariate analysis.

During the screening for Hepatitis B, Dianne and Brian (2013) demonstrate the process. Chronic hepatitis B virus (HBV) patients are frequently asymptomatic, but they face a variety of health risks including cirrhosis, liver failure, and liver cancer if they have active illness, and they may require antiviral therapy as a result of their infection. Nurse practitioners (NPs) can use a new algorithm to determine who to screen for HBV infection and how to follow up based on screening findings, including what additional tests to request, monitoring needed, and when to refer.

Gregory, John, and Eric (2019) Knowing about Hepatitis B is a critical public health issue that has to be addressed. The author employs a structured questionnaire and also multi stage sampling procedures were utilized, descriptive analysis and proportions were derived for categorical data and to determine prevalence.

An interviewer administered questionnaire was used in Kida, Umar, and Garbati's (2018) cross sectional descriptive study of 240 auxiliary health workers at the university of Maiduguri Teaching Hospital.

For the diagnosis of hepatitis disease, Mohammed and Abdul-Hamid (2014) developed a new classification approach called 3SVM, which combines support vector machines with scatter search approaches. Author employed scatter search to obtain near optimal settings of Support Vector Machine (SVM) parameters and kernel parameters, resulting in superior results and competitive performance.

Mustapha and Jibrin (2004) Used Both Hepatitis Virus B (HBV) and HIV infection are very common in Nigeria and are substantial causes of morbidity and mortality. Co-infection is known to occur since the two viruses have common routes of transmission. The Author employ double ELISA test (Bio-Rad, California, USA) to identify the presence of HIV antibodies. ELISA was used to detect HBsAg in sera (Murex Diagnostic, LTD UK). Following the manufacturer's protocol

was a top priority. In order to examine the significance of the difference between groups, the chi-square test was utilized. P. values less than or equal to 0.05 were regarded as statistically significant.

Useful information from Najla, Eman, and Samy (2018) The author employs artificial neural networks for classification and common regression analysis, in which historical data are best suited to a function, to diagnose hepatitis virus.

The sero-epidemic survey of hepatitis B in Northern Nigeria was conducted by Okoye and Samba in 2006. Chi-square test with Yates correction for small numbers was used to assess the data collected by the author.

Parivash, Parviz, and Seyed (2009) Cirrhosis and its effects can result from chronic liver disorders. The gold standard for measuring the degree of fibrosis and diagnosing cirrhosis has historically been histological diagnosis by liver biopsy, but it is an invasive process with inherent risk and sample uncertainty. Using regular laboratory findings, these scientists evaluate the efficacy of the Artificial Neural Network (ANN) to predict the presence or absence of cirrhosis in individuals with chronic hepatitis B.

Obidike and Uleanya (2015) The authors used a structured interviewer-administered questionnaire to determine the prevalence of Hepatitis B surface antigen (HBsAg) among youngsters for the study. Both parents' biographical information, as well as their occupational and educational standing, were requested. The data was analyzed using SPSS for measures of central tendency and the mean to summarize quantitative variables, as well as Chi-square analysis to create frequency tables.

Shahid, Saumitra, and Vigay (2010) Infection with one of the five (5) hepatitis viruses, which use the liver as their primary site of replication, causes viral hepatitis. The forms of hepatitis were revealed by the author's briefs. Hepatitis A through E viruses (HAV through HEV) are all members of various virus families that generate comparable clinical symptoms during the acute phase of infection but differ in their potential to cause chronic infection.

Yutao and Xiaolu (2019) The logistic regression machine learning classification algorithm was used to create the predictive model. Several crucial parameters are used by the author. Grid search with the training dataset was used to determine the value of each parameter. Confusion matrix, precision, and sensitivity were calculated for assessment purposes.

Yogambigai, Niazlin, and Subramaniam's (2019) study examines household knowledge and awareness of hepatitis B. Adult members of a selected home were questioned face-to-face using a two-stage cluster random sampling technique, and logistic regression was used to evaluate the differences in knowledge and awareness between the groups.

## **MATERIALS AND METHOD**

### **Location of the Study Area**

General Hospital (GNH) Mubi, Northern Senatorial District (**Latitude:** 10°16' 6.89"N and **Longitude:** 13°16'1.24"E), General Hospital (GNH) Hong, Central Senatorial District (**Latitude:** 10°18'37"N and **Longitude:** 12°55'29"E) and General Hospital (GNH) Ganye,

Southern Senatorial District (**Latitude:** 8°26'5"N and **Longitude:** 12°3'3"E) are among the general hospitals in Adamawa State Senatorial Districts where the study is being conducted.

**Data Collection**

This study endeavor will only use quantitative information. Data with a finite and discrete number of values is referred to as quantitative data. Quantitative data is specific; it can be counted and is well defined. Duplicates were found in the data collected for this study. The data is pre-processed by checking for missing data and randomizing the results. The data is extracted from the XML file and stored in a relational database first. The data is extracted, duplicates are removed, and the entire dataset is used to create comma separated value (CSV) files for model training and testing. Data might be quantitative or qualitative, and data collecting accuracy is a key component in sustaining the research's integrity. Despite the fact that there are a variety of data collection strategies available, this project will only collect quantitative data.

The following methods can be used to obtain quantitative data: -

- i. Obtaining data from an organization's or management's information system.
- ii. Conducting research and clinical trials, Medical Laboratory Test Results (MLTR).
- iii. Paying attention to or recording clearly defined events.
- iv. Surveys, interviews, and questionnaires collect data by asking closed-ended questions.
- v. Machine Learning Repository is an example of extracting data from a repository.

The information for this research came from Medical Laboratory Test Results (MLTR), which is a record of patient that can be used in the future.

**RESULTS**

**Research Design**

**Table 1: Properties of the Hepatitis Data Collected**

S/No.	Features	Data type
1.	Sex	Boolean Value
2.	Age	Integer Value
3.	Tiredness (TDS)	Boolean Value
4.	Loss of Appetite (LOA)	Boolean Value
5.	Nausea and Vomiting (NAV)	Boolean Value
6.	Yellowish of Skin and Eyes (YSE)	Boolean Value
7.	Stomach Pain (SPN)	Boolean Value
8.	Pain in the Muscles and Joint (PMJ)	Boolean Value

**Table 2: Samples of the data collected**

S/No	Sex	Age	Tiredness (TDS)	Loss of Appetite (LOA)	Nausea and Vomiting (NAV)	Yellowish of Skin and Eyes (YSE)	Stomach Pain (SPN)	Pain in the Muscles and Joint (PMJ)	HAV	HBV	HCV	HDV	HEV
1.	M	20	0	0	1	0	0	1	0	1	0	0	0
2.	M	35	0	1	0	0	1	0	0	1	0	0	0
3.	F	40	1	0	0	1	0	0	0	0	1	0	0
4.	M	45	0	0	0	0	0	1	0	1	0	0	0
5.	M	13	1	0	0	0	0	0	0	0	0	0	0

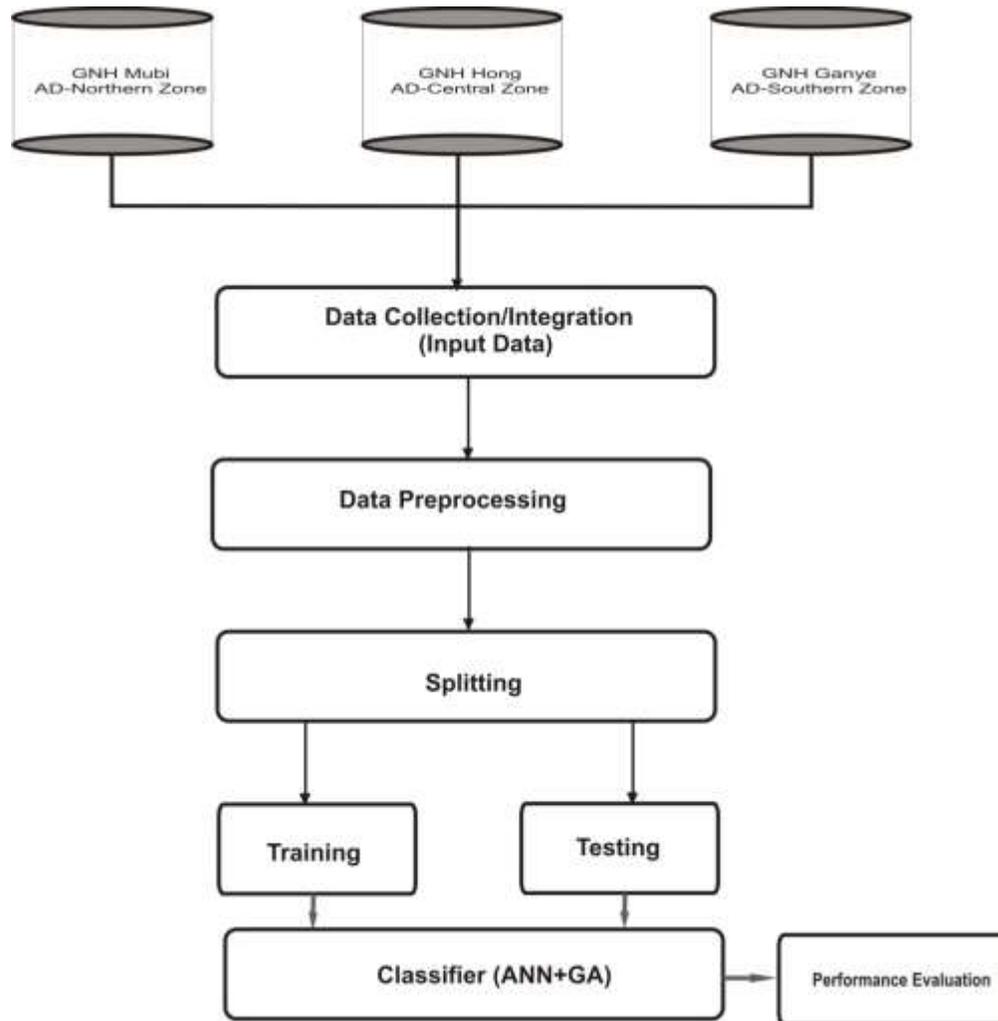


Figure 1: Proposed data classification system



Figure 2: Schematic Block Diagram of Data Preprocessing

**Framework of the Proposed System**

The system's models used Artificial Neural Networks (ANN) for classification and Genetic Algorithms (GA) for feature selection; this setup was used to develop and determine the network's performance.

### Data Preprocessing Algorithm

The initial raw data used by the preprocessor is referred to as a raw input in data preprocessing. A preprocessed input is the transformed data output produced by the preprocessor. The schematic diagram of the hepatitis Neural Network is shown in Figure 3.

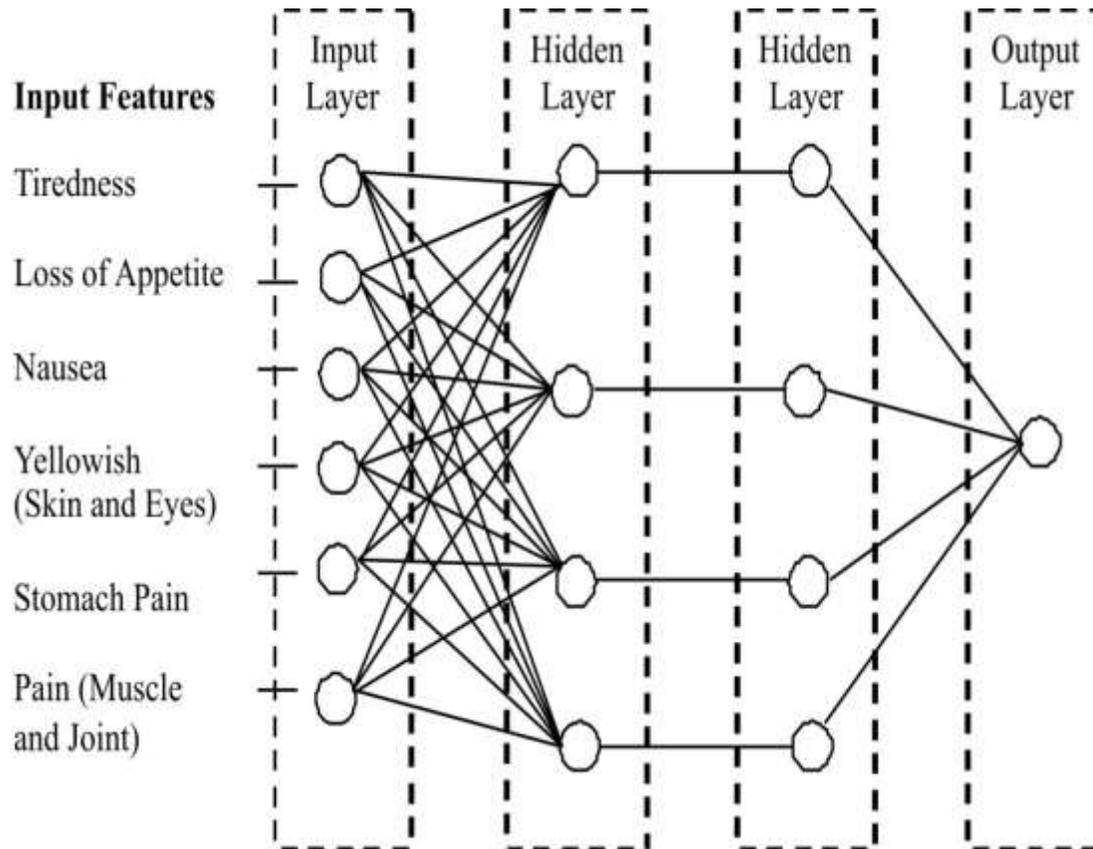


Figure 3: Schematic Diagram of hepatitis Neural Network

### Model Development

To develop the artificial neural network in general, there are five steps:

- i) Data Source/Collection
- ii) Preprocessing
- iii) Building the network
- iv) Training and Testing of the model
- v) Test performance Evaluation of the model

## RESULT AND DISCUSSION

### Data Preprocessing Results

This study used a genetic algorithm to pick features and an artificial neural network (ANN) to classify them in order to predict hepatitis B virus (HBV). The model's outcome was described in this section. Hospital Laboratory provided the hepatitis dataset used in this categorization. The data was cleaned, standardized, trained, and tested before being used. Figure 4 depicts raw unprocessed data from hepatitis test results, while figure 5 shows the hepatitis B virus (HBV) processed data test result.

1	MED. LAB. TEST RESULTS (MLTR)													
2	S/No.	Sex	Age	Tiredness (TDS)	Loss of Appetite (LOA)	Nausea and Vomiting (NAV)	Yellowish of Skin and Eye (YSE)	Stomach Pain (SPN)	Pain in Muscles and Joint (PMJ)	HAV - TEST CONDUCTED	HBV - TEST CONDUCTED	HCV - TEST CONDUCTED	HDV	HEV
3	1	M	20	0	0	1	0	0	1	0	1	0	0	0
4	2	M	35	0	1	0	0	1	0	0	1	0	0	0
5	3	F	40	1	0	0	1	0	0	0	0	1	0	0
6	4	M	45	0	0	0	0	0	1	0	1	0	0	0
7	5	M	13	1	0	0	0	0	0	0	0	0	0	0
8	6	F	48	0	0	0	0	1	0	0	1	0	0	0
9	7	F	30	0	0	1	0	0	1	0	1	0	0	0
10	8	F	25	0	0	0	0	0	1	0	0	0	0	0
11	9	M	15	0	0	1	0	0	1	0	1	0	0	0
12	10	M	50	1	0	1	0	0	0	0	1	0	0	0

Figure 4: Unprocessed samples data of hepatitis collected from various hospitals

### ANN Model Data Set Description

In Adamawa State, Nigeria, patient records were obtained from three (3) hospital laboratories. These records of people who had been diagnosed with hepatitis virus were carefully chosen and analyzed with medical practitioners' explanations.

The first seventy percent (70%) of data sets were used for training, while the remaining thirty percent (30%) were used for testing. The researcher, medical practitioners, and medical experts gave classes to all of the data sets, dividing the hepatitis B virus into four (4) levels based on the availability of blood samples from each patient. Negative and positive (Mild and High) symptoms are caused by a virus infection induced by the hepadaviridae family. Table 2 displays six (6) conditional attributes (symptoms).

### The Design of Predictive Symptoms-Based System Database

Total of nine hundred (900) data was collected which contain a patient's symptoms, and results.

#### i) Data Input

The presence of virus in the body and the patient's test results are the symptoms that a patient noticed. It displays symbols as well as the domains, or ranges, of possible

values. Hepatitis symptoms have four (4) potential values: negative (0), positive (1), mild (2), and high (3).

### Data Transformation

The information is categorized based on the severity of each patient's hepatitis.

#### Hepatitis:

Patients with level zero (0) are defined as having no hepatitis B virus (HBsAb), while those with levels one (1), two (2), and three (3) are classed as having hepatitis B virus (HBsAg), which can be positive, mild, or high.

The hepatitis B viral input and output are summarized in the tables below.

**Table 3: Data Transformation for Hepatitis B virus (HBV).**

Hepatitis	Extent	Class
-ve	Negative	0
+ve	Positive	1
++ve	Mild	2
+++ve	High	3

When a patient's test results are negative (-ve), it signifies that the patient does not have hepatitis B virus (HBV) and is classified as class zero (0). When a patient's test results are positive (+ve), (++ve), or (+++ve), it signifies the patient has hepatitis B virus (HBV) and belongs to one of three classes: (1), (2), or (3).

### ANN training and testing on unprocessed dataset

The same network setup was employed in this step with the 900 preprocessed samples. The training and testing samples are 70 percent and 30 percent, respectively. The network was put to the test to see how well it worked by counting how many samples were correctly identified when compared to the target data.

### ANN Classification

Feed forward back propagation ANN training with four layers, six inputs, two hidden layers, and eight hidden neurons was performed using an unprocessed dataset. The training sample is 70% of the unprocessed dataset, which consists of 900 hepatitis samples. The classifier also trains using back propagation and employs the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) optimization solver to update weights of the inputs at each layer for classification of the dataset using ANN, because the device used to train the model had limited memory, LBFGS was chosen. The Rectified Linear Unit (ReLU) function was employed for the activation function. The activation of a node is determined by changing the sum weighted of that node's activation function. It was chosen because it is simpler to train, performs better, and avoids the vanishing gradient problem. When gradient-based learning methods are employed to train artificial neural networks, the vanishing gradient problem occurs, causing the gradient to be small and preventing the weight from changing value. There were four buried levels, each with eight neurons. The maximum number of iterations was set to 50000 and the L2 regularization value was set to 1e-05. The failure of LBFGS to converge resulted in an increase in the maximum number of iterations. Tables 4, table 5 and 6 shows the results that were achieved. Figure 6 illustrates the confusion matrix.

The training and testing of the dataset were done in a 70:30 ratio. There are 900 cases in the dataset. The transformation of test results on the 900 of hepatitis normalized data is shown in figure 5.

	Sex	Age	Tiredness (TDS)	Loss of Appetite (LOA)	Nausea and Vomiting (NAV)	Yellowish of Skin and Eye (YSE)	Stomach Pain (SPN)	Pain in Muscles and Joint (PMJ)	HBV - TEST CONDUCTED
1	0	20	0	0	1	0	0	1	0
2	0	35	0	1	0	0	1	0	1
3	1	40	1	0	0	1	0	0	0
4	0	45	0	0	0	0	0	1	1
5	0	13	1	0	0	0	0	0	0
6	1	48	0	0	0	0	1	0	1
7	1	30	0	0	1	0	0	1	1
8	1	25	0	0	0	0	0	1	0
9	0	15	0	0	1	0	0	1	1
10	0	50	1	0	1	0	0	0	1

Figure 5: Processed samples data of hepatitis B virus (HBV).

**Table 4: Metrics Test Result on Hepatitis B Virus (HBV) Using ANN and GA.**

TP	FP	TN	FN	Total
76	41	103	50	270

**Hepatitis Infection Class Target Description**

- Hepatitis B Virus (HBV) - Class 0
- Hepatitis B Virus (HBV) + Class 1
- Hepatitis B Virus (HBV) ++ Class 2
- Hepatitis B Virus (HBV) +++ Class 3

The ANN matrix findings for hepatitis samples data are provided in Table 4, with TP, TN, FP, and FN. Table 4. Shows that the classifier made a total of 270 predictions that means 270 patients were being tested for the presence of the hepatitis B virus (HBV), out of these 270 cases the classifier predicted 117 times (patients) in the sample have the disease and 153 times (patients) don't have the disease. In a reality, 126 patients in the sample have the disease, and 144 patients don't have the disease.

Furthermore, Table 4 shows the results of the metrics test result performed just utilizing the ANN paired with GA. Also Table 6 shows the performance metrics for ANN with GA Feature Selection: accuracy (66.30 percent), specificity (63.33 percent), and sensitivity (71.53 percent). As shown in

figure 6, these results were achieved after multiple tests on the ANN, and GA. Table 5 shows the selected features that are responsible for hepatitis B virus (HBV).

**Table 5: The selected and rejected features are presented**

Features	Selected
Sex	Yes
Age	No
Tiredness	No
Loss of Appetite	Yes
Nausea and Vomiting	Yes
Yellowish Skin and Eye	Yes
Stomach Pain	Yes
Pain in Muscles and Joints	Yes

**Table 6: Performance Evaluation of the ANN and GA**

Performance Metrics	Score
Accuracy	66.30%
Sensitivity	63.33%
Specificity	71.53%

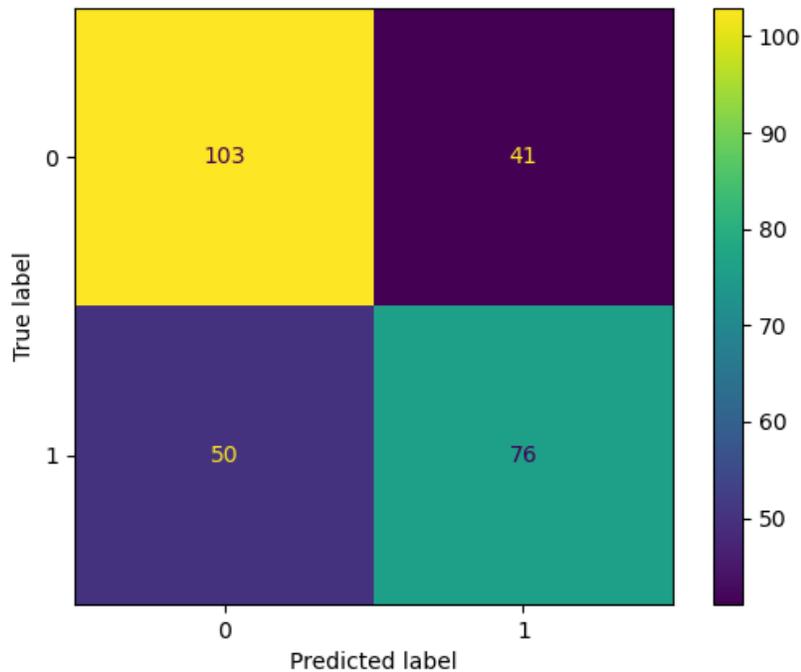


Figure 6: Confusion Matrix for hepatitis test data using ANN and GA.



Figure 7: Performance Metrics for ANN with and without GA Feature Selection

**ANN Classification Test Results**

The ANN matrix test findings for hepatitis B virus (HBV) samples data are summarized in Table 7, with the TP, TN, FP, and FN shown.

**Table 7: Summary of performance evaluation of a classifier**

Performance Metrics	ANN	ANN and GA
Accuracy	61.85%	66.30%
Sensitivity	55.48%	63.33%
Specificity	68.42%	71.53%

**Summary of Genetic Search Result for hepatitis B virus (HBV).**

**Search Method:** Genetic search.

Population size: 50

Number of generations: 40

Probability of crossover: 0.3

Probability of mutation: 0.1

Report frequency: 20

**Result for Hepatitis B virus (HBV) from Classification**

The results of the model's classification of 270 occurrences are presented in this section. According to Table 4. The model detects 63.33 percent of the hepatitis B class instances in the test dataset. The model recognizes 71.53 percent of the hepatitis B cases in the full test dataset, according to

specificity. The model's overall accuracy is 66.30 percent, which is relatively acceptable given the dataset's and Genetic Algorithm (GA) to classify the selected features, yielding accuracy, sensitivity, and specificity as shown in table 7.

### CONCLUSION AND RECOMMENDATION

Machine Learning approaches have proven to be useful in a variety of medical fields. For example, due to the fact that hepatitis is one of the world's worst diseases, it has taken hold in rural regions and poor countries in particular; there have been a lot of research projects in this area. Hepatitis virus (HBV) was predicted using a Genetic algorithm optimization tool and an Artificial Neural Network classifier was use in this research. The forecast was deemed to have acceptable performance measures, which will help medical specialists, respond more quickly in the future and reduce future outbreaks around the world.

The following are the research's recommendations.

- i. A graphical user interface application that would incorporate the model designed to aid the healthcare industry.
- ii. More research on hepatitis is needed to determine the optimum feature selection tool in a genetic algorithm that will allow for efficient optimization of big data sets.
- iii. To show the finest performance evaluation, many data mining tools will be coupled with genetic algorithms.

### References

- Adebola, T. O., & Muhammad, S. B. (2016). Seroprevalence of Hepatitis B Infection in Nigeria: A National Survey. *The American journal of tropical medicine and hygiene. Ahmadu Bello University Teaching*. 95(4), 2016, p. 902–907. from <https://www.researchgate.net/publication/pp.15-0874>
- Adoga, M. P., Bashayi, O. D. & Zungwe, T. (2010). Hepatitis B virus infection in apparently healthy urban Nigerians: data from prevaccination tests. *The Journal of Infection in Developing Countries*, 4(6), 397-400.
- Agatonovic-Kustrin, S., & Beresford, R. (2010). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5), 717–735.
- Awoleke, J. (2012). Hepatitis B. Surface antigenaemia among pregnant women in tertiary health institution in Ekiti state, Nigeria: *Tropical Journal of Obstetrics and Gynaecology*,29(29).
- Awowale, A. (2017). Decision Making in Healthcare Systems: Roles and Responsibilities: University of Maryland; [Google Scholar].
- Cheng, B., & Titterington, D. (2017). Neural Networks: a review from Statistical Perspective. *Statistical Science*, 9, 2-54.
- Dianne, T., & Brian, M. (2013) Screening for Hepatitis B. *The Journal for Nurse Practitioners*, 9(4).
- Ganem, D., & Prince, A. M. (2004). Hepatitis B virus infection—natural history and clinical consequences. *New England Journal of Medicine*, 350(11), 1118-1129.

- Hou, J., & Liu, Z. (2015). Epidemiology and prevention of hepatitis B virus infection. *International Journal of Medical Sciences*, 2(1), 50-57.
- IBM Corporation, (2013). Data-driven healthcare organizations use big data analytics for big gains. Somers, NY: IBM Corporation, [Google Scholar].
- Kida, M., Umar, B. Goni, B. Garbati, M. Adamu, I. Ogah, D. Bakki B. Dayar, A. Ummate, I. & Jusuph, H. (2018). Assessment of knowledge, attitude and practice towards Hepatitis B virus screening and vaccination among auxiliary health workers in university of Maiduguri teaching hospital (UMTH). *Borno Medical Journal*, 15(2).
- Kuziemy, C. (2016). Decision-making in healthcare as a complex adaptive system. *Health Manage Forum*, 29(1), 4-7. [DOI:10.1177/0840470415614842](https://doi.org/10.1177/0840470415614842).
- Luka, S. A., Ibrahim, M. B., & Iliya, S. N. (2018). Seroprevalence of hepatitis B surface antigen among pregnant women attending Ahmadu Bello University Teaching Hospital, Zaria, Nigeria. *Nigerian Journal of Parasitology*, 29(1)38-41.
- Mustapha, S., & Jibrin, Y. (2004). The Prevalence of Hepatitis B Surface Antigenaemia in patients with Human Immunodeficiency Virus (HIV) infection in Gombe, Nigeria. *Annals of African Medicine*, 3(1), 10-12.
- Najla, F., Eman, K. A. & Samy S. A. (2018). Diagnosis of Hepatitis Virus Using Artificial Neural Network. *International Journal of Academic Pedagogical Research (IJAPR)*, 2(11), 1-7. [www.ijeais.org/ijapr](http://www.ijeais.org/ijapr)
- Okoye, I. C., & Samba, S. A. (2006). Sero-epidemic survey of hepatitis B in a population of Northern Nigeria. *Animal Research International*, 3(2), 485-488.
- Odusanya, O. O., & Wellens, R. (2015). Prevalence of hepatitis B surface antigen in vaccinated children and controls in rural Nigeria. *International Journal of Infectious Diseases*, 9(3), 139-143.
- Parviz, V., Seyed, M. & Parivash, E. (2009). Using Artificial Neural Network to predict cirrhosis in patients with chronic Hepatitis B infection with seven Routine Laboratory findings. *Hepatitis Monthly, Autumn* 9(4), 271-275.
- Tahseen, A. J., Huda Y., & Madiha, M. Y. (2011). PCA-ANN for Classification of Hepatitis-C Patients. *International Journal of Computer Applications*, 14(7).
- Tseng, T., & Kao, J. (2013). Clinical utility of quantitative HBsAg in natural history and nucleosanaloque treatment of chronic hepatitis B. *Gastroenterol*, 48, 13-21.
- Uleanya, N., & Obidike, E. (2015). Prevalence and risk factors of Hepatitis B virus transmission among children in Enugu, Nigeria, *Nigerian Journal of Paediatrics*, 42(3), 199-203.
- Weinbaum, C. M., William, I., & Wang, S. A. (2018). Recommendations for identification and public health management of persons with Chronic Hepatitis B. virus Infection. *Morbidity and Mortality Weekly Report*, 57, 1-20.
- WHO, (2018). Hepatitis B. Retrieved from: <http://www.who.int/mediacentre/factsheets/fs204/en/> [Accessed: 23 April 2018].
- Willey, J.M., Sherwood, L.M., & Woolverton, C.J., (2018). The History and scope of Microbiology: Introduction of Microbiology. In: Willey, J.M., Sherwood, L.M and

- Woolverton, C.J.(editors) Prescott's Microbiology.8th Edition, New York. The McGraw Hill Companies. International Edition. P. 1-72.
- World Hepatitis Summit (2015). Building Momentum, making the case, Scottish Exhibition and Conference Center (SECC) Glasgow, United Kingdom 2-4 September, 2015., from: [www.worldhepatitisummit.com](http://www.worldhepatitisummit.com)
- Xiaolu, T., Yutian, C. & Yuantao, H. (2019). Using Machine Learning Algorithms to predict Hepatitis B Surface Antigen Seroclearance. *Computational and Mathematical Methods in Medicine*.<https://doi.org/10.1155/2019/6915850>
- Yogambigai, R., Subramaniam, M., Abram, L., & David, A. (2019) Knowledge and awareness of hepatitis B among households in Malaysia: A Community-based cross-sectional survey. *BMC Public Health*, 1-11.