



THE USE OF A MACHINE LEARNING ALGORITHM, AND ARTIFICIAL NEURAL NETWORK (ANN) TO PREDICT HEPATITIS B VIRUS (HBV)

Douglas Ibrahim¹, Ayuba John², Christopher Richard Kodeke³

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

²Department of Biological Science Technology. Federal Polytechnic,
P. M. B. 35, Mubi, Adamawa State, Nigeria.

³Computer Science Department. Adamawa State Polytechnic
P. M. B. 2146 Yola, Nigeria.

ABSTRACT

Hepatitis B is a liver infection caused by the hepatitis B virus (HBV). It might be acute and resolve on its own. Some kinds, however, can be persistent, leading to cirrhosis and liver cancer. A person can have HBV and spread the virus to others without realizing it; some persons have no symptoms, while others only have the first infection, which then resolves. For others, the condition becomes chronic. In chronic cases, the virus continues to attack the liver over time without detection, resulting in irreversible liver damage. The manual system contain large amount of errors by virtue of human decision, tedious and expensive in terms of labor requirements. This project proposed machine learning algorithm; Artificial Neural Network to predict the occurrence of Hepatitis. Performance Evaluation results of ANN shows the effectiveness of the proposed approach with the overall Accuracy (61.85%), Specificity (55.48%) and Sensitivity (68.42%). In this study, hepatitis B was predicted using Artificial Neural Network (ANN) classifier. The prediction was found to have acceptable performance measures which will aid timely response of medical experts.

Keywords: Prediction, Hepatitis B virus, Machine Learning, Data Mining, Artificial Neural Network.

1. INTRODUCTION

Hepatitis is a liver inflammation caused by viruses and other factors such as alcohol abuse, certain medications, and trauma. Millions of Nigerians are affected in various ways. Although many cases of hepatitis may not pose a major threat to health, infection with certain hepatitis viruses can become chronic (long-lasting), leading to liver failure and death in certain situations. The liver is an essential organ that processes nutrition, filters blood, and fights infections. When the liver is harmed, its function can suffer. Hepatitis can be caused by excessive alcohol consumption, pollutants, certain

drugs, and certain medical disorders. Hepatitis, on the other hand, is almost always caused by a virus (Tseng and Kao, 2013).

The world Hepatitis summit 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 to investigate the feasibility of eliminating hepatitis B and C with the goal of setting global targets and developing a monitoring mechanism. The submission aimed to generate momentum for the establishment of comprehensive national plans within the context of

the WHO viral hepatitis strategy 2016–2021. (World Hepatitis Summit, 2015).

Viral hepatitis causes both acute and chronic infection, with serious complications and long-term consequences. It is believed that more than 2 billion people worldwide have had hepatitis B virus (HBV) infection, with 350 - 400 million being chronic carriers of the virus. (Ganem and Prince, 2014).

HBV causes acute and chronic liver disorders and is responsible for an estimated 1 million deaths worldwide each year. Its prevalence varies over the world, although it is most prevalent in tropical areas. It is estimated that 5–15% of persons in Sub-Saharan Africa have chronic HBV infection (WHO, 2018). Adults are at a 15–25% risk of dying prematurely from HBV-related cirrhosis and hepatocellular carcinoma severe infections may potentially lead to fulminant liver failure. (Ganem and Prince, 2014). About 90% of those infected during the prenatal period, 30% of those infected in early childhood, and 6% of those infected after 5 years of age develop chronic infection (Lok and McMahon, 2017).

In areas of high endemicity where at least 8% of the population are chronic HBV carrier, HBV is mainly concentrated at birth and early childhood perinatal transmission from an infected mother to her baby in common (Hou and Liu, 2015).

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 modes of transmission. Other forms of iatrogenic or horizontal transmission, such as long-term household connections with no sexual involvements in high-endemic areas, could also occur. HBV infection is also acknowledged as a risk to healthcare practitioners' occupational health (Bharai, and Rijal, 2014).

The causal agent of hepatitis B infection is the Hepatitis B virus (HBV), a DNA virus of the family Hepadnaviridae. Hepatitis B is a major health concern and one of the most frequent infections in the globe. According to the most recent World Health Organization (WHO) estimates, 2 billion persons worldwide have serologic evidence of past or present 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 hepatitis C virus (HCV), and many carriers are unaware they are infected. It is an important cause of liver diseases such that chronic infection with HBC is a common cause of death associated with liver failure, cirrhosis and liver cancer (Weinbaum, William, and Wang, 2018). The prevalence of HBV varies between 2% in developed countries where the prevalence is low to about 8% in developing countries where infection is endemic with sex, age and socio-economic status as important risk factors for infection. (Adoga, Bashayi, and Zungwe, 2010).

The degree of endemicity of HBV frequently coincides with the major mechanism of transmission. The disease has a huge influence on the health and national economies of many countries, and its severity is very changeable and often unpredictable. Because the lowest infectious dosage is so low, behaviors like sharing a toothbrush or a razor blade can spread illness (Odusanya, Wellens, and Weil, 2015).

Hepatitis B virus (HBV) is a blood-borne and sexually transmitted infection that spreads via contaminated blood or other body fluids (Saliva, sweat, semen, vaginal secretions, breast milk, urine and faeces). Transmission can occur when using the same syringe as an infected person, during blood transfusions, during childbirth, during medical operations, or during sexual intercourse. Hepatitis B viral transmission channels are very similar to HIV transmission methods. (Chang, 2017).

Currently, there are four recognized modes of transmission (Viral Hepatitis Board 1996) which are:

- a. From mother to child at birth (prenatal)
- b. By contact with infected person
- c. By sexual contact and
- d. Exposure to blood or other infected fluids.

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

Most people who becomes infected with HBV are able to clear the virus from their blood stream within 6 months of post infection and develop immunity. Those who have not cleared the virus after 6 months are considered to have chronic hepatitis B infection. The risk of health from HBV related liver cancer or

cirrhosis is approximately 25% for persons who acquire chronic infections at childhood. Moreover, 8% to 10% of people in the general populations in developing countries become chronically infected and most acquire infections with HBV at childhood (Weinbum, 2018).

Nigeria is one of the countries with a high prevalence of HBV infection. According to reports, around 75% of the Nigerian population has been exposed to HBV at some point in their lives. Despite the fact that hepatitis B vaccine is highly efficient in avoiding HBV infection and the subsequent acute and chronic liver illness, this virus remains a big concern in Nigeria, according to several workers (Luka, Ibrahim, and Iliya 2018).

2. STATEMENT OF THE PROBLEM

The world is experiencing a chronic shortage of well-trained health workers, a crisis felt most acutely in those countries that are experiencing the greatest public health threats. WHO estimates that over 4 million health workers are needed to fill the gaps and the midwives (WHO, 2014).

The manual system contains large amount of errors by virtue of human decision making (Lavanya and Rani, 2011). So it is also difficult to analyse large set of data, train and test the data for clinical decision that will lead to better evaluation of performance of the model developed. The need to predict the transmission rate in community would enable donors and other stakeholders' effective and efficient utilization of resources to minimize the rate of transmission of hepatitis B. Virus.

3. RELATED WORKS

3.1 Hepatitis

Hepatitis B Virus (HBV) is a blood borne and sexually transmitted pathogen that could be acquired through intravenous drug use, sexual intercourse with infected partners, and prenatal transmission from mother to child among others. Hepatitis B is one of the most common infectious diseases reaching hyper-endemic proportions in Africa. In Nigeria, HBV spot surveys amongst people revealed rates ranging between 2.19% and 15.11% (Jonas, 2009). Infection with Hepatitis B Virus (HBV) is a widespread problem. Epidemiological survey showed that about 5% of the world populations are asymptomatic carrier. Chronic HBV infection is a major cause of mortality in men and 50% of chronic carriers can be expected to die from disease due to liver cirrhosis or hepatocellular carcinoma (Omer, 2018).

Hepatitis is serious and grouping threat to the health of Nigerians. Millions of Nigeria from all walks of life are already living with chronic hepatitis B virus (HBV) infection. All of these persons are at increased risk for liver disease, cancer and death. They are from every state in the nation and from all social economic and ethnic groups. The HAV and HEV are entirely transmitted and cause an acute, self-limited infection, with complete resolution, except in rare situations in which fulminate disease with high mortality is observed. The associated clinical illness, named as hepatitis A and hepatitis E respectively can each occur either as epidemics, or as sporadic cases in the absence of a recognized outbreak. The HBV, HDV and HCV are parenterally transmitted, causes acute infection with a high propensity to become chronic with long-term sequence such as cirrhosis and hepatocellular carcinoma called liver cancer (Seeger and Zoulim, 2017).

Unfortunately, many people are unaware of the viral hepatitis status and too few have accessed viral hepatitis care and treatments. Too many people are falling through the cracks and too many are continuing to die from causes related to viral hepatitis. Public health and health care system are missing key opportunities to prevent infections, diagnose and treat people, prevent serious disease and save lives (WHO, 2018).

3.2 Types of Hepatitis

Five viruses are responsible for most cases of viral hepatitis, which is an inflammation of their liver due to a viral infection. These are the hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV) and hepatitis E virus (HEV) All the hepatitis viruses can cause acute hepatitis. The HAV and HEV are enterally transmitted (Seeger and Zonlin, 2017).

Hepatitis B virus infection is associated with a considerable burden of disease. In high income countries, transmission of hepatitis B (HBV) may occur through injecting drug use, but in most cases occur through heterosexual transmission of transmission between men who have sex with men. Chronic infection with HBV is associated with an increased risk of liver cirrhosis (between 8% and 20% within five years of diagnosis) and liver failure (EASL, 2012).

3.2.1 Hepatitis A virus (HAV)

HAV is present in the faeces of infected persons and is most often transmitted through consumption of contaminated water or food. Certain sex practices can also spread HAV. Infections are in many cases mild,

with most people making a full recovery and remaining immune from further HAV infections. However, HAV infections can also be severe and life threatening. Most people in areas of the world with poor sanitation have been infected with this virus. Safe and effective vaccines are available to prevent HAV.

3.2.2 Hepatitis B virus (HBV)

HBV is spread through contact with infected blood, sperm, and other bodily fluids. HBV can be passed on from infected moms to infants at birth or from a family member to a newborn in early infancy. Transfusions of HBV-infected blood and blood products, contaminated injections during medical operations, and injectable medication use can all result in transmission. HBV also offers a risk to healthcare personnel who suffer accidental needle stick injuries while caring for patients infected with HBV. To prevent HBV, vaccinations that are both safe and efficient are available.

The hepatitis B virus (HBV) chronically infects 350 million people worldwide, causing illnesses ranging from acute hepatitis to chronic hepatitis, cirrhosis, and cellular carcinoma (HCC), and is responsible for over one million deaths each year. (Lemon and Walker, 2017).

Chronic HBV infection with cirrhotic liver has been identified as the single most important risk factor for the development of HCC, one of the most deadly malignancies and the third highest cause of death in males.

3.2.3 Hepatitis C virus (HCV)

HCV is mostly transmitted through exposure to infective blood. This may happen through transfusions of HCV-contaminated blood and blood products, contaminated injections during medical procedures, and through injection drug use. Sexual transmission is also possible, but is much less common. There is no vaccine for HCV.

3.2.4 Hepatitis D virus (HDV)

This infection occurs only in those who are infected with HBV. The dual infection of HDV and HBV can result in a more serious disease and worse outcome. Hepatitis B vaccines provide protection from HDV infection.

3.2.5 Hepatitis E virus (HEV)

HEV is mostly transmitted through consumption of contaminated water or food. HEV is a common cause of hepatitis outbreaks in developing parts of the world and is increasingly recognized as an important

cause of disease in developed countries. Safe and effective vaccines to prevent HEV infection have been developed but are not widely available (Parkin, 2012).

4. CAUSES OF HEPATITIS B (HBV)

Infection with one of the five known viruses that mostly affect the liver causes viral hepatitis. The viruses that cause hepatitis A, B, C, D, and E (HAV, HBV, HCV, HDV and HEV). Despite significant commonality in clinical manifestations, clinical manifestations differ significantly in their morphology, genetic organization, taxonomic classification, and processes of replication. These viruses enter the host through one of two main routes: enteral or parenteral. The age at which a person becomes infected determines the likelihood of developing a chronic infection. Up to 90% of babies infected with the hepatitis B virus develop chronic infection. In comparison, approximately 5% of individuals will acquire chronic hepatitis B. Chronic hepatitis B can lead to major health complications such as liver damage and cirrhosis over time (Hollinger and Emerson, 2007).

According to Jill (2019), Hepatitis B virus is spread by direct contact with body fluids (blood, semen, or vagina fluids) of an infected person. This can happen during unprotected sex or while sharing needles used to inject drugs. A baby can be infected during birth if the mother has the hepatitis B virus. The hepatitis B virus also can be spread if you live with an infected person and share household items that may come in contact with body fluids, such as toothbrushes or razor. The virus can be spread through the following: -

- a) **Sex with an infected person:** Among adults, hepatitis B is often spread through sexual contact.
- b) **Injection Drug use:** Sharing needles, syringe and any other equipment to inject drugs with someone infected with hepatitis B can spread the virus.
- c) **Outbreaks:** While uncommon, poor infection control has resulted in outbreaks of hepatitis B in healthcare settings.
- d) **Birth:** Hepatitis B can be passed from an infected mother to her baby at birth. Worldwide, most people with hepatitis B were infected with the virus as an infant.

5. SIGNS AND SYMPTOMS OF HEPATITIS

There are many types of hepatitis and the signs and symptoms of hepatitis. Acute infection is a short-term illness that happens in the first 6 months after a person is infected with the hepatitis B virus. Acute

infection may cause only mild symptoms or no symptoms at all. When symptoms do occur, they may include the following (Jill, 2019).

- a) Tiredness
- b) Loss of appetite
- c) Nausea and vomiting
- d) Jaundice (yellowish of the skin and eyes)
- e) Stomach pain
- f) Pain in the muscles and joint

6. HEPATITIS TEST

There are several blood tests available to detect the hepatitis B virus. Antibodies to the virus are looked for in the testing. Hepatitis B virus tests can determine whether you have recently been infected or if you are a carrier. They can also tell you if you've had the hepatitis B virus in the past and are now immune to it, or if you've had the hepatitis B vaccine. The following individuals should be checked for the hepatitis virus.

- a) Pregnant women
- b) Infants born to infected mothers
- c) Sex partners and those who live with an infected person.
- d) People with Human Immunodeficiency Virus (HIV) infection.
- e) Users of injected illegal drugs
- f) Men who have sex with men
- g) People who are the source of blood or other body fluid exposures (for example, when a healthcare worker has been stuck by a needle).
- h) People born in a country with a high rate of hepatitis virus infection or people with parents having hepatitis.
- i) People receiving dialysis, cancer treatment, or treatment with drugs that suppress the immune system (Schweitzer, 2015).

6.1 Treatment, Prevention and Protection of Hepatitis B Virus (HBV)

People with chronic HBV infection should be evaluated for liver damage every 6 – 12 months. Several antiviral medicines are presently approved for the treatment of chronic HBV infection. Although these medications are successful in preventing significant liver issues in up to 40% of individuals, they do not eliminate the virus. The last recourse is a liver transplant, but livers are not always accessible. Avoid alcohol because it can aggravate liver illness. Although there is no medication available to treat newly acquired HBV infection, symptoms can be managed. A vaccine is the most effective way to protect against the hepatitis B virus. When you are exposed to the

virus, the vaccine activates your immune system to combat it.

It is typically administered in three doses during a six-month period. People who have recently been exposed to the hepatitis B virus but have not been vaccinated are frequently given the vaccination combined with a brief course of Hepatitis B Immune Globulin (HBIG), which includes antibodies to the virus. In some cases, it can provide further protection against infection. Some of the liver problems caused by the infection can be treated. Hepatitis B vaccine provides the best protection; two or three doses are administered over a one to six-month period, depending on the brand.

Whenever a woman is pregnant, she should be tested for hepatitis B (HBsAg) blood test); infants born to HBV-infected mothers should be given HBIG (hepatitis B Immune globulin) and vaccine with 12 months of birth. The sex partners to get vaccinated too, and always follow safer sex practices example: using condom (Parkin, 2012).

Tahseen, Huda and Madiha (2011) Research work on PCA-ANN for classification of hepatitis, the author's uses ANN for classification and trimmed down the huge data by using principal component analysis which is considered as one of the most prevalent and useful statistical method.

Uleanya and Obidike (2015) The authors used a structured interviewer delivered questionnaire to determine the prevalence of Hepatitis B surface antigen (HBsAg) in children for the study. Bio-data, occupational and educational level of both parents were among the items sought. SPSS was used to analyze the data, and measures of central tendency and the mean were used to summarize quantitative variables. Chi-square analysis was performed to create frequency tables.

Vigay, Saumitra and Shahid (2010) Viral hepatitis is caused mainly by infection with one of the five (5) hepatitis virus, which uses the liver as their primary site of replication. The author briefs revealed the types of hepatitis. Each of these type, known as hepatitis A through E viruses (HAV to HEV), belong to different virus families, these viruses cause similar clinical manifestations during the acute phase of infection but vary in their ability to cause chronic infection.

Xiaolu and Yutao (2019) The predictive model was created using the logistic regression machine learning classification algorithm. Several crucial parameters are used by the author. The grid search

with the training dataset was used to determine the value of each parameter. The confusion matrix, accuracy, and sensitivity were calculated for evaluation purposes.

The research work by Yogambigai, Niazhlin and Subramaniam (2019) explore knowledge and awareness of hepatitis B among households. A two stage cluster random sampling design was used and adult member of selected household was interviewed face-to-face logistic regression was used to estimate the differences in knowledge and awareness between the groups.

7. DATA MINING

Data mining is a logical process that searches through enormous amounts of info to identify usable data. The purpose of these strategies is to discover previously unknown patterns. Once these patterns are discovered, they may be leveraged to make specific decisions for the growth of their organization. There are three steps involved: Exploration, Pattern identification, Deployment.

- a) **Exploration:** In the first step of data exploration, data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.
- b) **Pattern Identification:** Once data is explored refined and defined for the specific variables, the second step is to form pattern identification. Identify and choose the patterns which make the best prediction.
- c) **Deployment:** Patterns are deployed for desired outcome.

Data mining techniques are used to build methods for discovering knowledge from data, which is then used to uncover hidden or undiscovered information that is not obvious but potentially useful. Data mining relies heavily on classification and clustering techniques. Classification classifies data according to a classifier model, whereas clustering groups data according to distance or similarity (Sandhya and Kangalammal, 2013).

Data mining classification algorithms are a type of machine learning technique in which each class has an instance that is identified by its distinct properties. It is an important part of data mining and a valuable forecasting tool for models generated from an input data set. It is also a method for forecasting the class of labels that have never been discovered and are used to classify future data patterns (Shazmeen, Baig, and Pawar, 2013).

7.1 Data Mining Classification Tools

Data mining tool for extracting information from huge sets of data. It is a machine learning tools available for data analysis. Over the years many algorithms have been created to extract what is called nuggets and knowledge from large sets of data (Nikam and Meshram, 2014).

7.2 Data Mining Classification Tools of Hepatitis B

One of the most difficult tasks is to choose the right data mining techniques (Gibert, Sanchez-Marre and Codina, 2010). As the commercial software tools provide more and more possibilities together technical know-how and management capabilities is required to have effective and efficient information from data gathered (Gibert et al., 2010). Consider the available tools below: -

7.3 Python

Python is a general-purpose programming language, which means it may be used for a variety of purposes. Python is used in web development, artificial intelligence, machine learning, operating systems, mobile application development, and video game creation. Because it was designed for statistical computation and graphics, it provides a plethora of statistical packages to its users. Python, on the other hand, is a general-purpose programming language with several uses. Python, on the other hand, can be used for statistics. Python statistics is a general-purpose language with statistics components. However, when it comes to creating complex research pipelines that combine statistics with, for example, image analysis, text mining, or physical experiment control, Python's richness is an important tool (Lutz, 2011).

8. MACHINE LEARNING

Machine learning is a technique for learning a new algorithm from experience. It entails the investigation of algorithms capable of autonomously extracting information. The advancement of information technology has resulted in a significant number of databases and massive amounts of data in a variety of fields. Database and information technology research has resulted in a method for storing and manipulating earlier data for future decision making. Data mining is the process of collecting usable information and patterns from massive amounts of data. It is also known as the knowledge discovery process, data mining, knowledge extraction, and data/pattern analysis (Malli and Nandyal, 2017).

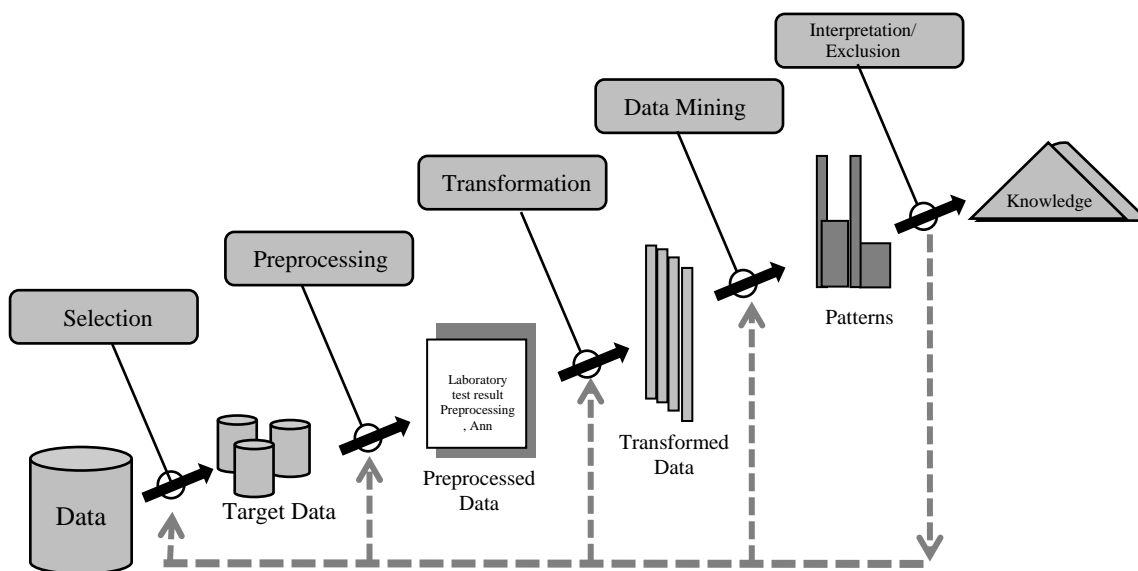


Figure 1: Knowledge discovery process (Bharati, 2010).

Artificial intelligence and machine learning shares mutual threads with various fields, they include mathematics, optimization, physics, theory of games, and theoretical computers. Below are some problems solved with machine learning in various industry (Vinyals, 2015): -

- a) Medical Diagnosis
- b) Predictive modeling
- c) Topic spotting
- d) Face Detection
- e) Pattern Recognition
- f) Spoken language understanding
- g) Prediction of weather

Machine learning programs unlike static programs provide a higher degree of accuracy of result. Static programs usually use imprecise impression or examine only few numbers of data. Machine learning has three set of data namely the domain set, the label set, and the training set.

8.1 Types of Machine Learning

According to Jeanmonod, 2018 they are:

a) Supervised learning

This machine learning is a learning strategy in which the datasets are labeled and the algorithm attempts to discover the relationship between the labeled set and the features set. The retrieved knowledge is used for prediction and recognition in supervised learning. Supervised learning is divided into two types:

classification and regression. Machine learning categorizes labels or targets as classes in classification. It predicts the class to which each instance belongs, whereas regression predicts numeric values as labels or targets for each dataset instance. Prediction of labels or targets in supervised learning is based on the experience obtained from learning the training dataset (Chao, 2011).

b) Unsupervised learning

The objective of this machine learning is to have the computer learn how to do things that it was not informed to do or taught how to go do it. Unlike the supervised learning, the goal is not to classify but to ensure decision making that will yield maximum rewards. Unlike supervised learning there are no labels but observation commonly known as outputs in which, the first approach to this system is not to explicitly feed it with categorizations

c) Reinforcement learning

Reinforcement learning as an impact in an environment based on the action it encounters. Hence it feedback is based on how much it has learned from observation. This learning is considered a trial and error model. It learns a particular way how to act given a particular pattern (Chao, 2011).

9. NEURAL NETWORK

ANNs are prominent machine learning algorithms that have seen widespread adoption in recent years. Multilayer Perception (MLP) is the most basic version of ANN, which is a neural network that

adjusts its weights during training via back propagation. Pattern Recognition Network (PRN) and Convolution Neural Network (CoNN) are two other neural network types that have lately gained popularity in categorization. Artificial Neural Networks (ANNs) are massively parallel, distributed processing systems that represent a computing technology based on the human Information Processing System. They are a massively connected network of Neurons, which are Simple Processing Elements. They have a natural proclivity to absorb and save information so that it might be used later.

ANNs can be used for classification, pattern recognition and function approximation Neural Networks are often used for statistical analysis and data modeling, in which their role is perceived as an alternative to standard nonlinear regression or cluster analysis techniques (Cheng and Titterington 1994). Thus, they are typically used in problems that may be couched in terms of classification, or forecasting. A Neural Network is a computational system inspired by the structure and learning ability. A Neural Network has the following features:

- a) A large number of very simple processing Neuron-like processing elements
- b) Large number of weight connections between the elements.
- c) Distributed representation of knowledge over the connections knowledge is acquired by network through a learning process.

It is an Information Processing System consisting of large number of Interconnected Processing Units. It works in a way our nervous system processes the information. It is basically a dense interconnection of simple non-linear computation elements. It has been successfully applied to problems involving pattern classification, function approximation, regression, prediction and other. (Ke-Lin and Swamy, 2014).

9.1 Artificial Neural Network Layers

There are multiple layers of neurons (nodes) in an artificial neural network, each layer receives multiple inputs to its nodes and, after applying some mathematical calculations, passes the information to the next layer. Thus, each layer feeds the next layer with information. This is the basic concept of a feed-forward neural network. There are three main

categories of layers in an artificial neural network: input, hidden, and output layers.

- a) **Input Layer:** It is the layer in which input data groups are introduced to the network. Parameters in input layers have to be selected before analysis. The number of neurons in an input layer is equal to the number of input data; every input neuron is transmitted to the next layer which is the hidden layer.
- b) ii) **Hidden Layer:** The hidden layer is the basic function of the network. In this layer, data received from the input layer is processed properly and then transmitted to the output layer.
- c) iii) **Output Layer:** Learning takes place in the output layer. Linear units are connected to the output consisting of hidden layers (Abdi, 2003). It is the final layer in the network and it processes the data received from the hidden layer and creates the output. The number of neurons is equal to the number of outputs received by the network. Values obtained are the output values for the problem in the artificial neural network (Dag, 2012).

10. METHODOLOGY

The goal of this study is to forecast hepatitis B virus (HBV) using an Artificial Neural Network (ANN) for classification. This chapter describes the approach employed for this research, including how it was pre-processed, how suitable characteristics were selected from the data classification, and how an Artificial Neural Network (ANN) was used in training and testing with hepatitis B virus-prepared data (HBV).

10.1 Data Collection

The data for this study was obtained from the UCI Machine Learning Repository. The UCI Machine Learning Repository is a database or data generator that collects and stores machine learning and statistics-related data. Also in Adamawa state, Nigeria, are the Northern Senatorial District (Mubi South) General Hospital, the Central Senatorial District (Hong) General Hospital, and the Southern Senatorial District (Ganye) General Hospital. Table 1 shows the properties of the hepatitis data set; it is regarded as a center for machine learning and intelligent systems.

Table 1: Properties of the hepatitis data collected

S/No.	Features	Data type
1.	Sex	Boolean Value
2.	Age	Boolean 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 (PMT)	Boolean Value

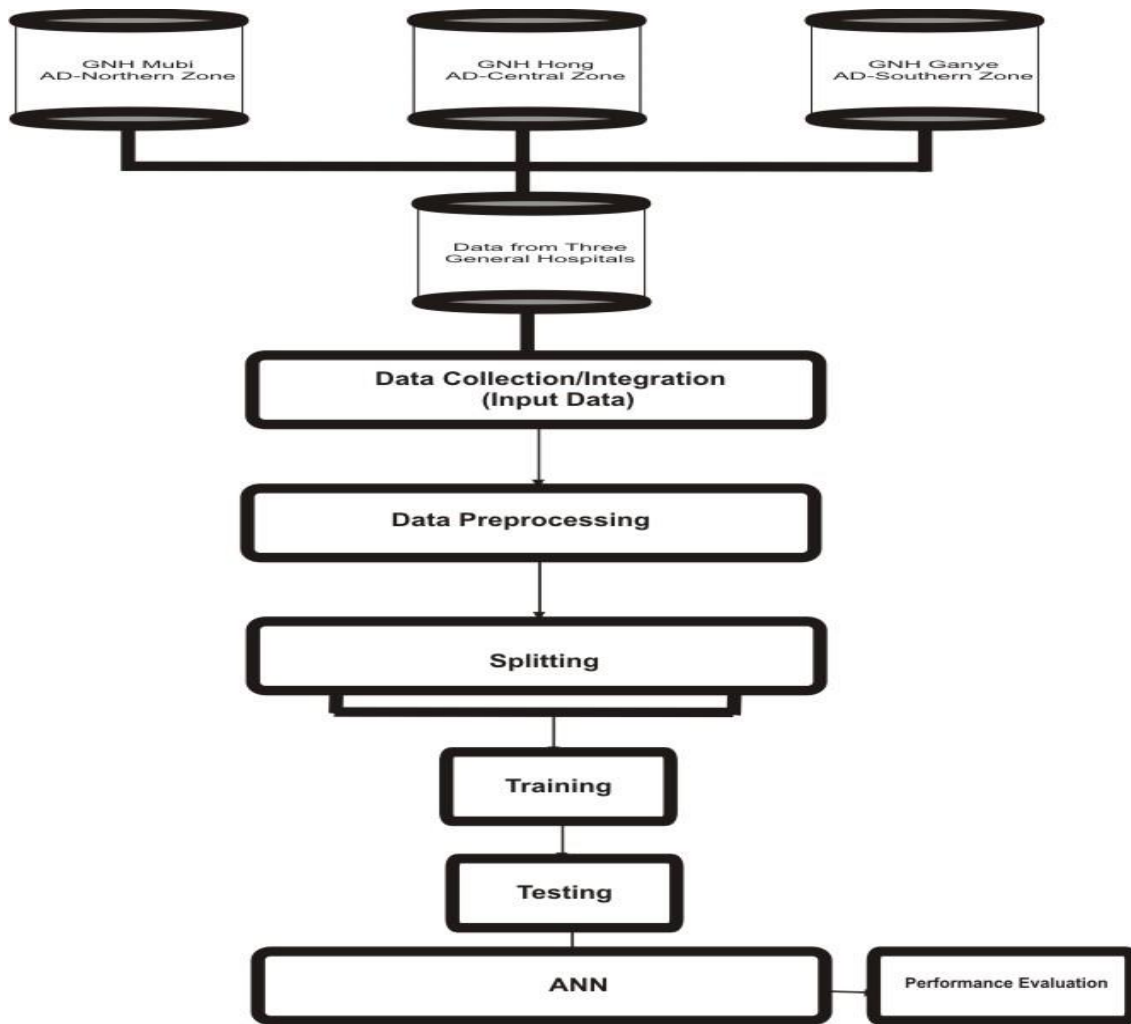


Figure 2: Data classification system

11. DATA NORMALIZATION

Data normalization is a methodical way of deconstructing tables in order to minimize data redundancy and undesired qualities such as outliers. It is a multistep process that converts data from relation tables into tabular form by deleting

duplicate data. Data preparation is critical since the quality of the data determines the outcome of the study. This is due to data being proven to be noisy due to the large size of the database, complexity, and various heterogeneous data sources. The preparation and production of high-quality data is accomplished

by preprocessing the data to remove noise, irrelevant features, redundant features, and incorrectly categorized samples. These irrelevant features

diminish classification accuracy, necessitating data preparation to prepare the data for mining activities.

Table 2: Features Description of Hepatitis Data Transformation

Rough Set	Symptoms	Domain 1	Domain 2	Domain 3	Domain 4
1.	Tiredness	No=0	Yes=1		
2.	Loss of Appetite	No=0	Yes=1		
3.	Nausea/Vomiting	No=0	Yes=1		
4.	Yellowish (skin/eye)	No=0	Yes=1		
5.	Stomach Pain	No=0	Yes=1		
6.	Pain (Muscles/Joint)	No=0	Yes=1		
			(Positive=1)	Mild=2	High=3

12. TRAINING AND TESTING THE MODEL

The features that have been chosen will be used during the training and testing phases. Artificial Neural Network (ANN) will be used to build a model that will categorize a trained dataset. A classifier's accuracy is the percentage of correctly identified cases. That is the example or instances that have been placed in the appropriate classes. The related operational characteristic curve, also known as the receiving operating characteristic curve, is a curve that depicts the relationship between the real positive and the false positive. Tables are used to express the outcome. The model's performance is displayed for analysis. Analyses that demonstrate how accurately the models classify hepatitis outcomes.

13. CONFUSION MATRIX

A confusion matrix is a table that displays a classifier's accuracy on a given problem. The confusion matrix is a m by n matrix given a data set with n classes, where cij is the number of tuples allocated to class cij and correct class ci. A confusion matrix with the optimal solution having zero at its diagonal. The confusion matrix, which will be used to describe the performance of a classification model, contains both the predicted and the actual classification.

14. RESULTS AND DISCUSSION

14.1 Data Preprocessing Results

This study predict hepatitis B virus (HBV), using ANN for classifying. This section discussed the result of the model.

a) Data Input

The symptoms that a patient observed are the present of virus in the body and the test results of the patient. It shows symbols and the range of their possible values called domains. Those symptoms for hepatitis have four (4) possible values which can be negative (0), positive (1), Mild (2) and High (3).

14.2 Data Transformation

The data are classified based on the level of hepatitis of each patient.

Hepatitis:

A patient with level zero (0) is classified to be a victim of no hepatitis B virus (HBsAb), a patient with level one (1), two (2), and three (3) is classified to be a victim with hepatitis B virus (HBsAg) which can either be positive, mild or high. Table 3 summarize the hepatitis B virus input and output.

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

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

Table 4: ANN Matrix test result on hepatitis B virus (HBV).

TP	TN	FP	FN	TOTAL
76	91	42	61	270

N=270	Predicted No	Predicted Yes	
Actual No	TN (91)	FP (42)	133
Actual Yes	FN (61)	TP (76)	137
	152	118	

Figure 3: Confusion Matrix of the Result of the ANN Classifier

Table 5: Performance Evaluation of the Classifier

Performance Metric	Score
Accuracy	61.85%
Sensitivity	55.48%
Specificity	68.42%

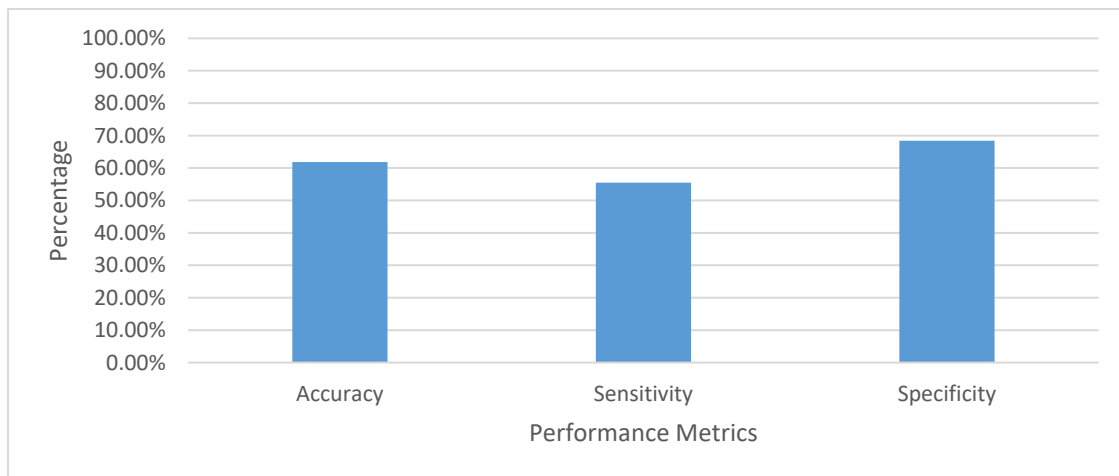


Figure 4: Performance metrics of ANN

Table 4 shows the matrix results of ANN for hepatitis samples data indicating TP, TN, FP and FN. The classifier made a total of 270 predictions that means 270 patients were being tested for the presence of hepatitis B virus (HBV). Out of these 270 cases, the classifier predicted 118 patients have the disease and also predicted 152 patients have no disease. The performance evaluation and confusion matrix obtained using the ANN is given in table 5 and figure 3 respectively.

15. CONCLUSION

Machine Learning techniques have proven to be of great tools in various sects, there has been quite a number of research works in hepatitis owing to the fact that is one of the deadliest disease in the world that have taken hold of rural areas and developing countries particularly. In this study hepatitis was predicted using Artificial Neural Network classifier. The prediction was found to have acceptable performance measures which will aid timely response of medical experts in making decisions.

16. RECOMMENDATION

The following are the recommendations for this research work.

- a. A graphical user interface application that would integrate the model developed which will help the health sector.
- b. More research should be carried out on hepatitis to explore the best machine learning algorithm so as to enable efficient optimization of a large data.

REFERENCES

- Abdi, H. (2003). Factor rotations in factor analysis: In Encyclopedia for Research Methods for the Social Sciences (Page 792–795). Retrieved from <https://www.utdallasherve/Abdi-rotations-pretty.pdf> (9th January, 2019).
- Adoga, M. P., Bashayi, O. D. & Zungwe, T. (2010). Hepatitis B virus infection in apparently healthy urban Nigerians: data from prevaccination tests. *J. Infect in Dev. Countries* 4 (6): page. 397-400.
- Bharati, S. & Lama, S. (2014). Hepatitis B. vaccination status and needle stick and sharps-related injuries among medical school students in Nepal: A cross sectional study. *BMC Notes* 7: 774.
- Chan, H. Wong, V. Won, G. Tse, C. Chan, H. & Sung, J. (2010). A longitudinal study on the natural history of serum hepatitis B surface antigen changes in chronic hepatitis B. *HEPATOLOGY* 2010;52:1232-1241.
- Cheng, B. & Titterington, D. (1994). Neural Networks: a review from Statistical Perspective. *Statistical Science* 9, page 2 – 54.
- Chao, W. (2011). Machine Learning Tutorial. *Disp.Ee.Ntu.Edu.Tw*, 56. Retrieved from <http://disp.ee.ntu.edu.tw/~pujols/Machine Learning Tutorial.pdf>
- Dag, Z. (2012). Determination of the Relationship between Palm Terriers and Blood Glucose HbA1C Parameters: *Using Artificial Neural Networks*. Unpublished PhD Thesis Sakarya University, Institute of Science and Technology; 7(13).
- European Association for the study of Disease of liver (EASL) (2012). Clinical Practice Guidelines: Management of chronic hepatitis B virus (HBV) infection. *Journal of Hepatology*, 57; page 167-185.
- Ganem, D. & Prince, A. (2014). Hepatitis B virus infection, natural history and clinical consequences. *N Engl J. Med* 350:p.1118-1129.
- Gibert, K. Sanchez, M. & Codina, V. (2010). Choosing the right Data Mining Technique: Classification of Methods and Intelligent Recommendation. In Proceedings of the IEMS, Fifth Biennial Meeting: International Congress on Environmental Modelling and Software page 1933-1940.
- Hollinger, F. B., & Emerson, S. U. (2007). Hepatitis A virus. In *Field's Virology* (eds Knipe, D. M. and Howley, P. M.), Lippincott, Williams and Wilkins, Philadelphia, 2007, page 911–948.
- Hou, J. & Liu, Z. (2015). Epidemiology and prevention of hepatitis B virus infection. *Int J. Med Sci.* 2:page 5-57.
- Jeanmonod, D. J., & R. (2018). We are In tech Open , the world ' s leading publisher of Open Access books Built by scientists , for scientists TOP 1% Control of a Proportional

- Hydraulic System. *Intech Open*, 2, 64. <https://doi.org/10.5772/32009>
- Jill, S. (2019). Medical Review on November 7, 2019. Written by Kathleen Davis, FNP <https://www.medicalnewstoday.com/articles/306288#vaccine>
- Ke-Lin, D. & Swamy, M. (2014). *Neural Networks and Statistical Learning*. Springer London Heidelberg New York Dordrecht (4&19).DOI 10.1007/978-1-4471-5571-3Ke-Lin & Swamy 2014.
- Lok, A. & McMahon, B. (2017). Chronic hepatitis B. *Hepatology* 45:507-539.
- Lutz, M. (2011) *Programming Python*, O'Reilly Media, *more advanced book* 4th ed. *the Tkinter chapters*.
- Luka, S. A., Ibrahim, M. B., & Iliya, S. N. (2018). Seroprevalence of hepatitis B surface antigen among pregnant women attending Ahmadu Bello Univeristy Teaching Hospital, Zaria, Nigeria. *Nig. J. Parasit* 29 (1):38-41.
- Malli, P. K., & Nandyal, S. (2017). Machine learning Technique for detection of Cervical Cancer using k-NN and Artificial Neural Network, 6(4).
- Nikam, V. B. & Meshram, B. B; (2014). Parallel and Scalable rules based classifier using map-reduce paradigm on Hadoop cloud. *International Journal of Advance Technology in Engineering and Science*, 2(8), 558-568.
- Omer, E. E. (2018) Clinical Significance of makers of hepatitis B. *Medicine digest* 2, 10 – 15.
- Odusanya, O. O; & Wellens, R. (2015). Prevalence of hepatitis B surface antigen in vaccinated children and controls in rural Nigeria. *Intl J Infect Dis*. 9(3):p.139-143.
- Parkin, D. M. (2012). Global cancer statistics, *CA Cancer J. Clinol*; page 74 – 106.
- Sandhya, M., Kangalammal A. & Senthamarai, C. (2013) "A Comparative study f Decision Rule Induction for incomplete data using Roughset and Ramdom Tree Approaches", *ISOR Journal of Computer Engineering* Vol. 9, Issue 3, page 06-10, www.isorjournals.org, Retrieved 13/01/14.
- Seeger, C., Zoulim, F. & Mason, W. S. (2017). Hepadnaviruses. In *Field's Virology* (edsKnipe, D. M. and Howley, P. M.), Lippincott, Williams and Wilkins, Philadelphia, page. 2977–3030.
- Shazmeen, S. F; Braig, M. M. & Pawar, M. R. (2013). Performance Evaluation of difference Data Mining classification Algorithm and Predictive analysis. *Journal of Computer Engineering*, 10(6), 01-06.
- Schweitzer, A. & Horn, J. Krause, G; (2015) Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *Lancet* 2015; 386:1546–1555.
- Tahseen, A. J., Huda Y. & Madiha, M. Y. (2011). PCA-ANN for Classification of Hepatitis-C Patients *International Journal of Computer Applications* (0975 – 8887)Volume 14– No.7, February 2011.
- Tseng, T. & Kao, J. (2013). Clinical utility of quantitative HBsAg in natural history and nucleos analogue treatment of chronic hepatitis B. *Gastroenterol* 2013;48:13-21.
- Uleanya, N. & Obidike, E. (2015). Prevelance and risk factors of Hepatitis B virus transmission among children in Enugu, Nigeria, *Niger J. Paed* 2015; 42(3):199-203.
- Vinyals, O. & Le, Q. V. (2015). A neural conversational model. *Proceedings of the 31st International Conference in Machine Learning*, Vol.37, *arXiv:1506.05869v3*, 2015.
- Vijay, K. Saumitra, D. & Shahid, J. (2010). Biology and Pathogenesis of Viruses, *journal of computer science Research gate* Vol. 98, No. 3. <http://www.researchgate.net/pub>.
- 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. Available from:
<http://www.who.int/mediacentre/factsheets/fs204/en/> [Accessed: 23 April 2018].
https://www.who.int/health-topics/hepatitis#tab=tab_1
<https://www.who.int/news-room/q-a-detail/hepatitis> 1 September 2019 |

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
<https://www.who.int/news-room/q-a-detail/hepatitis> 1 September 2019.

Xiaolu, T. Yutian, C. & Yuantao, H. (2019). Using Machine Learning Algorithms to predict Hepatitis B Surface Antigen Seroclearance. *Computational and Mathematical Methods in Medicine*. Article ID 6915850. Received 6 January 2019; Accepted 2 April 2019, published 11 June 2019., from: <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.