International Journal of Academic Pedagogical Research (IJAPR)
ISSN: 2643-9603
Vol. 4, Issue 5, May – 2020, Pages: 19-35

Detection of Hepatitis (A, B, C, D and E) Viruses Using Machine Learning

[1]Alile Solomon Osarumwense and [2]Bello Moses Eromosele

[1,2]Department of Computer Science, Faculty of Physical Sciences, University of Benin, Benin City, Edo State. P.M.B. 1154, Nigeria.
Email: [1]solomon.alile@physci.uniben.edu,[2]moseseromo@yahoo.com

Abstract**:** *Hepatitis Disease is a life threatening inflammatory condition of the liver cells that causes damage to the liver which is usually caused by a viral infection which affects persons ranging from infants, older children and adults in respective of age. This disease is caused by Single Stranded RNA Virus (ssRNA Virus family), Symmetrical RNA Virus (sRNA Virus family) and Double-Stranded DNA Virus (dsDNA Virus family). Furthermore, the transmitted Hepatitis Virus is of five types namely: Hepatitis A Virus (HAV), Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), Hepatitis D Virus (HDV) and Hepatitis E Virus (HEV) respectively. The symptoms of this disease are jaundice, fever, fatigue, loss of appetite, nausea, vomiting, abdominal pain, joint pain, dark urine, clay colored faeces and diarrhea just to name a few. Additionally, WHO declared that Hepatitis virus occurs sporadically and categorized as an epidemic worldwide, with a tendency for repeated recurrences. This Hepatitis infection has caused millions of death worldwide yearly due to lack of early diagnosis of the ailment. In recent past, several systems have been developed to diagnose this endemic disease, but they generated a lot of false negative during testing and were unable to detect Hepatitis Disease, its overlapping symptoms and various types. Hence, in this paper, we proposed and simulated a model to predict Hepatitis (A, B, C, D and E) using a machine learning technique called Bayesian Belief Network. The model was designed using Bayes Server and tested with data collected from Hepatitis UCI medical repository. The model had a 99.97% prediction accuracy and 96.98%, 95.08%, 97.32%, 98.11%, 97.71% and 95.71% sensitivity of Hepatitis Disease, HAV, HBV, HCV, HDV and HEV in that order.*

Keywords: Viral Hepatitis Disease; Hepatitis A Virus; Hepatitis B Virus; Hepatitis C Virus; Hepatitis D Virus; Hepatitis E Virus; Machine Learning; Bayesian Belief Network; Prediction; Detection.

1.0 Introduction

The human body is the structure of a human which consists of cells, tissues, organs, and organ systems that is structured from the simplest to the most complex as stated by [1]. Furthermore, the human body consists of several systems that regulate all parts of the body from the cells, tissues and organs resident in the body. Examples of such systems are circulatory system; endocrine system, respiratory system and digestive system just to name a few. Of all these systems, the digestive system is very important due to its obligation in the breakdown of food intake into essential nutrients that are required by the body to stay healthy.

In [2], it was stated that the human digestive system comprises of the gastrointestinal tract and the accessory organs of digestion namely the tongue, salivary glands, pancreas, gallbladder and liver. On the other hand, all aforementioned accessory organs work in synergy with one another for optimal functionality of the human digestive system. Of these accessory organs, the liver is the most important due to the many functions it provides, some of which are so important to digestion in the human body.

[3] defined liver as an accessory digestive gland that plays a major role in the chemical processes that occur within the human body to sustain life and classified as the second largest organ in the human body; it is located in the right upper area of the abdomen. The liver has much functionality that aids digestion in the human; some of which are detoxification of different metabolites, amalgamate proteins and manufacture biochemicals required for digestion and control the storage of glycogen which is derived from glucose. Despite the operational functionalities the liver exhibit, it is subject to anomalies caused by disease causing organisms.

In [4], disease was defined as disorder in organization or function in humans, animals and plant, especially one that generates unequivocal indications affecting a specific part within the body of the living organism. Conversely, diseases can be categorized into two major type namely communicable and non-communicable diseases.

Non-communicable diseases (NCD) are diseases that are not transmitted from one person to another such as Parkinson's disease, autoimmune diseases, strokes, most heart diseases, just to name a few which may be chronic or acute but are non-infectious as affirmed by [5].

Communicable disease (CD) is a type of disease that is transmitted from one person to another. The transmission of this disease takes place via bacteria or viruses that are airborne, through contact of blood and different types of body fluids from an infected person [6]. This class of diseases are regarded as Bubonic plague due to its mode of spread; hence, termed as an infectious disease. Examples of such disease are Hepatitis, tuberculosis etc.

In [7], it was stated that the main cause of infectious diseases (CDs) are pathogenic microorganisms such as bacteria, fungi, parasites and viruses, with the disease spreading from one person to another in a direct or indirect mode. Of all the CDs causing organisms, diseases caused by viruses are the most terrifying of all.

Viruses are submicroscopic communicable agents that replicate itself inside living cells of an organism as avowed by [8]. Viruses have the ability to communicate a disease to all forms of living organisms such as plants, animals, microorganisms and humans just to name a few. However, viruses are found in every ecosystem on planet earth and are classified as the most copious form of biological entity.

In [9], it was declared that viruses are also located in every ecological unit on planet earth and are categorized as the most abundant form of organic entity. However, viruses can be classified into 7 main classes namely double stranded DNA virus (e.g. Adenoviruses, Herpesviruses, Poxviruses), single stranded DNA virus (e.g.Parvoviruses), dsRNA virus (e.g.Reoviruses), double stranded DNA-RT viruses (e.g. Hepadnaviruses),(-)single stranded RNA virus (e.g.Orthomyxoviruses, Rhabdoviruses), single stranded RNA-RT viruses (e.g. Retroviruses) and (+)single stranded RNA virus (e.g., Picornaviruses, Togaviruses) respectively.

Of all the above-listed virus families, RNA and DNA virus families are the most dreadful of all with both causing different forms of a communicable disease called viral Hepatitis disease.

Viral Hepatitis is a life threatening rabble-rousing (inflammatory) condition of the liver cells and damage to the liver which is usually caused by a viral infection which affects persons ranging from infants, older children and adults in respective of age as stated by [10]. The symptoms of this disease are jaundice, fever, fatigue, loss of appetite, nausea, vomiting, abdominal pain, joint pain, dark urine, clay colored faeces and diarrhea respectively. Additionally, the level of severity of Hepatitis Disease ranges from mild, moderate, severe, and critical level with the symptoms of the disease varying in people, with some persons showing no symptoms whatsoever, hence being categorized as been asymptomatic while individuals with the infection are classified as symptomatic persons who are the major carrier of the infection.

However, there are five types of viruses that can cause Hepatitis infection with similarities in their individual symptoms namely: Hepatitis A, Hepatitis B, Hepatitis C, Hepatitis D and Hepatitis E respectively.

The aforementioned types of Hepatitis are caused by Hepatitis A virus (HAV), Hepatitis B virus (HBV), Hepatitis C virus (HCV), Hepatitis D virus (HDV) and Hepatitis E virus (HEV) in that order. The three most common are Hepatitis A, B, and C. However, infection with any of these three viruses can lead to life threatening complications.

Furthermore, each type of virus has different characteristics and their transmission modes are different, but their symptoms tend to be alike. However, the incubation period of these viruses in humans differs ranging from 14 to 28 days in respective of the type of virus contacted. Besides, Hepatitis E virus is subdivided into 4 different types namely: Genotypes 1, 2, 3 and 4 in that order. Genotypes 1 and 2 have been found only in humans while genotypes 3 and 4 move continuously in several animals such as pigs, wild boars, and deer just to name a few without causing any disease, and seldom infect humans. Between 2015 and 2016, World Health Organization (WHO) estimated that Hepatitis A caused approximately 7,134 deaths which was 0.5% of the mortality due to viral Hepatitis; it also estimated that Hepatitis B and Hepatitis C resulted in an estimated 887 000 and 399,000 deaths mostly from cirrhosis and primary liver cancer and Hepatitis E caused approximately 44,000 deaths worldwide [11, 12, 13, 14 and 15]. Additionally, WHO avowed that Hepatitis virus occurs sporadically and categorized as an epidemic worldwide, with an inclination for repeated recurrences.

Due to the sporadic and epidemic nature of Hepatitis virus, clinical methods have been utilized in diagnosing this disease such as reverse transcriptase polymerase chain reaction (RT-PCR), serological test, a nucleic acid test and HDV RNA test respectively. However, the usage of the aforementioned diagnostic methods has the following issues: RT-PCR tests are able to identify transcripts and but not able to detect functional proteins, the problem of bleeding and infection arises with either method using serological test with the patients prone to fainting as a result of drawing blood from their bodies, nucleic acid test are very expensive and not affordable for the common man and HDV RNA test are complex tests that are only done in few laboratories around the world. Hence, a lot of false positives are produced as a result of usage of the above-listed methods due to the overlapping symptoms Hepatitis and its various types has with other liver and flu-like diseases.

Moreover, with the steady growth in artificial intelligence, several machine learning techniques has been utilized in diagnosing Hepatitis disease in the works of [16, 17, 18, 19, 20, 21,22,23,24 and 25] but they generated a lot of false negative during testing and were unable to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases.

In this paper, a supervised machine learning technique called Bayesian Belief Network (BBN) was utilized in diagnosing Hepatitis disease with its symptoms. BBN is a flexible probabilistic network that merges expert knowledge and observed datasets. It plans a route for cause and effect associations between variables and trains them with probability that indicates the level in which one variable is likely to sway another. In this paper, BBN was our technique of choice because of its capability to make predictive inference.

2. Related Works

Several studies have been conducted on diagnosing Hepatitis Disease using Artificial Intelligence. In [16], a knowledge based system for diagnosing Hepatitis B virus which utilized Generalized Regression Neural Network (GRNN) was developed. The system diagnosed Hepatitis B virus with high detection accuracy and assisted in reducing extra time consumed in treatment of

Hepatitis B. Despite the high detection accuracy, the system failed to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases. The system neural network is capital and time intensive; and it also has a slow convergence speed.

In [17], an intelligent system for diagnosing Hepatitis B using Generalized Regression Neural Network (GRNN) was developed. The system demonstrated the ability to detect Hepatitis B with high detection accuracy. However, the system had the following issues: the system neural network learning process requires a lot of time and quite expensive; it also has a slow convergence speed and less generalizing performance issue. Additionally, the system was unable to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases

In [18], a system for diagnosing hepatitis virus that employed Artificial Neural Network (ANN) was developed. The system showed its ability to detect Hepatitis virus with 98.44% detection accuracy. Regardless of the high detection accuracy, the system had the following issues: The system neural network is capital and time intensive; it does not offer information about the relative importance of the various parameters, the neural network convergence speed is relatively slow, it has less generalizing performance issue, it has concern of arriving at local minimum and has over-fitting problems. Also, the system was unable to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases.

In [19], a system for diagnosing Hepatitis disease using Case-Based Reasoning (CBR) and Particle Swarm Optimization (PSO) was developed. The system detected Hepatitis disease with 94.58% detection accuracy. In spite of the high detection accuracy, the system had the following issues: CBR Libraries are biased, the most suitable cases derived using CBR might not be retrieved, it requires high adaptation and retrieval knowledge; PSO algorithm can be difficult to classify initial design parameters, the iterative process has a low convergence rate; it also has a problem of converging prematurely as a result of been trapped within local minimum especially when dealing with complex problems. Also, the system was unable to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases.

In [20], an expert system that utilized Fuzzy Logic to diagnose Hepatitis B was developed. The expert system diagnosed Hepatitis B with 92.2% detection accuracy. Regardless of the system high detection accuracy, the system had the following issues: the system reasoning module could not handle issue of uncertainty and make bi-directional inferences. Also, the system was unable to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases.

In [21], a system that predicted the outcome of Hepatitis B Virus (HBV) using Artificial Neural Network (ANN) was developed. The system predicted Hepatitis B virus with a high prediction accuracy. Despite the high detection accuracy, the system failed to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases. However, the system had the following issues: The system neural network is time consuming and capital intensive; it does not provide information about the relative significance of the various parameters utilized, it also has a slow convergence speed and less generalizing performance issue.

In [22], an expert system for detecting Hepatitis diseases using SL5 object programming language was developed. The proposed system detected Hepatitis disease with high detection accuracy in patients. This system was evaluated and tested by a group of medical practitioners and patients with liver issues and they were contented with its performance. Regardless of the expert system high detection accuracy, it had the following issues: the expert system development is capital-intensive and comes with a high maintenance cost. Also, the system failed to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases.

In [23], an intelligent learning system for predicting hepatitis using Adaptive Neuro Fuzzy Inference System was developed. The proposed system predicted hepatitis disease with 93.06% detection accuracy. Despite the high detection accuracy, the system failed to detect Hepatitis and its various types due to the overlapping symptoms the disease shares with liver and flu-like diseases; the system neural network learning process requires a lot of time and quite expensive; it also has a slow convergence speed and less generalizing performance issue, the fuzzy logic reasoning module could not handle issue of uncertainty and make bi-directional inferences.

3. Machine Learning

Machine learning is a set of methods for creating models that describe or predict using example data or past experience. However, there are several types of machine learning namely Supervised Learning: it trains data and includes desired outputs (e.g. Bayesian Belief Networks, Neural Networks, Deep learning etc.), Unsupervised Learning: it trains data and does not include desired outputs (e.g. Clustering, Dimensionality Reduction), Semi-Supervised Learning: it trains data and includes few desired outputs and Reinforcement Learning: it gains from sequence of actions (Temporal Difference Learning, Q-learning) [24].

In this paper, we intend to employ supervised machine learning technique called Bayesian Belief Network due to its predictive capability based on past experience and example data at its disposal during training and testing of observed datasets.

Bayesian Belief Network (BBN) is directed acyclic graphical model that employs probability to illustrate conditional dependencies that prevail amongst nodes on a graph [25]. It is a complex probabilistic network that merges expert knowledge and experimental

datasets. It plans out route of cause and effect relationships between variables and encodes them with probability that signify the amount in which one variable is probable to sway another. Bayesian Belief Network is based on the Bayes theorem which relies on probability.

The Bayes theorem is represented in the mathematical equation below:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)} \quad (1)$$

Where,
P(a) is the probability of event "a" happening without any information about event "b". It is called the "Prior".
P(a/b) is the conditional probability of event "a" happening given that event "b" has already occurred. It is otherwise called the "Posterior".
P(b/a) is the conditional probability of event "b" happening given that event "a" has already occurred. It is called the "Likelihood".
P(b) is the probability of event "b" happening without any information about event "a". It is called the "Marginal Likelihood".
The Naive Bayes classifiers are often represented as a type of directed acyclic graph (DAG). The Directed Acyclic Graph (DAG) comprises of vertices representing random variables and arrows connecting pairs of nodes. Figure 1 shows a pictorial representation of a Bayesian Belief Network.

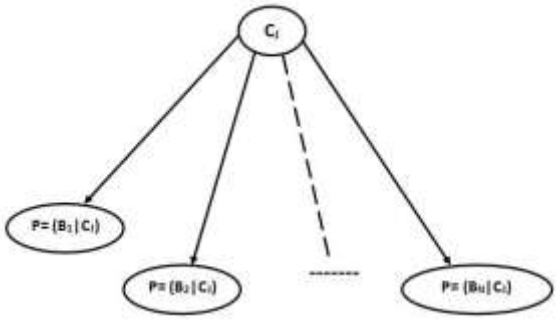

Figure 1: *A Pictorial Representation of a Bayesian Belief Network*

Some advantages of this model are: it is quite quick in making inferences, the resulting probabilities are easy to interpret, the learning algorithm is quite easy to understand and the model adequately combines with utility functions to make optimal inferences. In this paper, we intend to detect Hepatitis Diseases and its various types using a supervised machine learning technique called Bayesian Belief Network (BBN). A model consisting of 39 nodes where some nodes represent a form of disease ailment or factors that influence diagnosis of Hepatitis Disease and its various types will be designed using Bayes Server. A Hepatitis Disease dataset will be used to train and test the system. Using the Pareto Principle, 80% of the dataset will be used to train the model while the remainder will be used in testing the model. The aim of the model is to achieve a high level of detection accuracy with the use of the overlapping symptoms of Hepatitis disease and its various types.

4. Methodology

Simulation, Result and Discussion

The dataset used in training, testing and predicting Hepatitis Disease was retrieved from [26]. The dataset consist of mixture of disease ailments and factors taken into consideration in the detection of Hepatitis disease amounting to 39 with each ailment and factor having a value which represents the probability of such disease ailment and factor causing Hepatitis Disease. The ailments and factors are Abdominal Pain, Clay Colored Faeces, Continuous Fever, Dark Urine, Diarrhea, Double-Stranded DNA Virus (dsDNA Virus), Duration, Fatigue, Fever, Genotype 1, Genotype 2, Genotype 3, Genotype 4, Hepatitis A Virus, Hepatitis B Virus, Hepatitis C Virus, Hepatitis Disease, Hepatitis D Virus, Hepatitis E Virus, Incubation Period, Intermittent Fever, Joint Pain, Loss of Appetite, Malaria Fever, Nausea, Neutropenic Fever, Pel-Ebstein Fever, Quartan Fever, Quotidian Fever, Remittent Fever, Single Stranded RNA Virus (ssRNA Virus), Symmetrical RNA Virus (sRNA Virus), Tertian Fever, Typhoid Fever, Vomiting, 1Week, 24 Hours,48 Hours and 72 Hours respectively. Figure 2 below shows a sample the dataset.

International Journal of Academic Pedagogical Research (IJAPR)
ISSN: 2643-9603
Vol. 4, Issue 5, May – 2020, Pages: 19-35

HAV	HBV	HCV	HDV	Hepatitis	HEV	Incubation Period	Intermittent Fever	Joint Pain	Loss of Appetite	Malaria Fever	Nausea	Neutropenic Fever
1.26	-2.02	0.425	1.1	0.0277	-0.642	-0.281	-0.00617	-0.206	-2.21	0.341	0.593	0.42
0.353	-0.604	1.41	0.407	-0.47	-0.756	-1.57	1.34	1.43	0.0745	0.319	-0.742	-0.928
-1.49	0.712	0.723	-0.48	1.49	-0.619	1.59	1.27	-1.15	-1.66	-2.22	0.824	0.639
1.38	-0.55	0.918	-0.148	-0.534	-0.326	-0.182	0.317	1.25	-0.0227	-1.43	0.0448	-1.26
-0.0715	-0.502	-0.651	-0.578	1.33	-0.32	-1.18	-1.5	0.753	1.74	-0.984	-1.51	-0.332
-1.83	0.282	-0.675	-0.0315	0.169	-1.35	-0.216	-0.685	-0.719	-1.04	-0.77	1.09	-0.372
0.0592	-0.116	-0.454	0.96	-0.0924	-1.1	0.926	-1.52	-0.523	-1.6	1.31	1.47	0.832
-0.601	-0.000515	-0.519	-0.178	-1.92	2.28	-1.24	0.668	-1.44	1.43	-0.139	0.76	0.408
0.248	-1.85	-1.73	-1.01	-0.303	1.35	-0.648	0.283	0.501	0.798	0.0979	0.686	-0.109
-0.3	-0.902	-0.933	-0.194	1.69	-1.65	1.13	0.563	0.853	0.242	-1.49	0.467	-0.443
0.469	0.208	-0.00786	0.458	-0.932	0.706	-2.44	0.158	-1.06	-1.26	0.0852	-1.11	-0.131
0.732	-0.237	0.854	1.55	0.492	0.343	-1.31	0.0418	-0.602	0.108	-1.33	0.204	0.515
-1.18	-0.632	1.55	0.193	0.419	-0.191	-0.745	-0.345	0.849	1.76	0.014	-0.422	1.76
-0.364	0.0432	-0.0599	0.957	0.305	-0.339	-0.0531	-0.447	-0.522	-1.03	0.434	1.17	-1.02
0.73	-0.619	0.768	0.649	1.56	0.499	-1.34	-0.0513	-1.13	0.727	1.53	0.0394	-1.56
-0.862	0.614	-0.073	0.533	-0.229	-0.454	0.334	0.901	1.17	0.174	1.2	1.52	1.36
0.338	0.303	0.119	1.84	-0.835	0.406	0.7	0.813	1	0.984	0.356	0.0627	0.454
0.792	0.785	-1.23	0.295	-0.927	-1.28	-0.666	0.112	2.46	-0.348	-0.977	1.44	-0.00154
-1.29	-1.02	0.489	1.4	0.158	-0.0208	0.668	0.115	0.431	1.17	-1.41	0.419	0.437
0.228	0.444	0.297	1.88	2.56	0.798	0.577	-0.0335	1.06	0.644	-0.319	0.976	0.59
0.526	-0.28	-1.39	-0.398	-1.72	-2.04	1.51	-1.4	-0.883	-1.16	0.602	-0.853	0.56
-1.38	0.666	1.71	-1.24	-0.165	0.669	-0.493	-1.31	-0.497	0.739	0.373	0.651	-1.56
0.902	0.603	-0.727	-0.201	-2.2	0.279	-0.998	0.387	-1.95	0.295	0.21	0.0797	0.3
-0.169	-0.656	1.81	-1.2	1.5	0.333	0.366	-0.907	0.19	-0.546	-0.313	0.794	0.528

Figure 2: *Snapshot of Dataset*

The Bayesian Belief Network model was designed using Bayes-Server platform. The Bayesian Belief Network (BBN) for predicting Hepatitis Disease was designed such that the nodes on the network are linked based on the probability of a disease ailment resulting to another and factor influencing another factor. In our model for a case to be denoted as a Hepatitis case, The ailments, disease causing agents and other factors taken into cognizance in the diagnosis of Hepatitis Disease are: Abdominal Pain, Clay Colored Faeces, Continuous Fever, Dark Urine, Diarrhea, Double-Stranded DNA Virus (dsDNA Virus), Duration, Fatigue, Fever, Genotype 1, Genotype 2, Genotype 3, Genotype 4, Hepatitis A Virus, Hepatitis B Virus, Hepatitis C Virus, Hepatitis Disease, Hepatitis D Virus, Hepatitis E Virus, Incubation Period, Intermittent Fever, Joint Pain, Loss of Appetite, Malaria Fever, Nausea, Neutropenic Fever, Pel-Ebstein Fever, Quartan Fever, Quotidian Fever, Remittent Fever, Single Stranded RNA Virus (ssRNA Virus), Symmetrical RNA Virus (sRNA Virus), Tertian Fever, Typhoid Fever, Vomiting, 1Week, 24 Hours,48 Hours and 72 Hours respectively.
Figure 3 shows the BBN model for detecting Hepatitis Disease, its symptoms and various types.

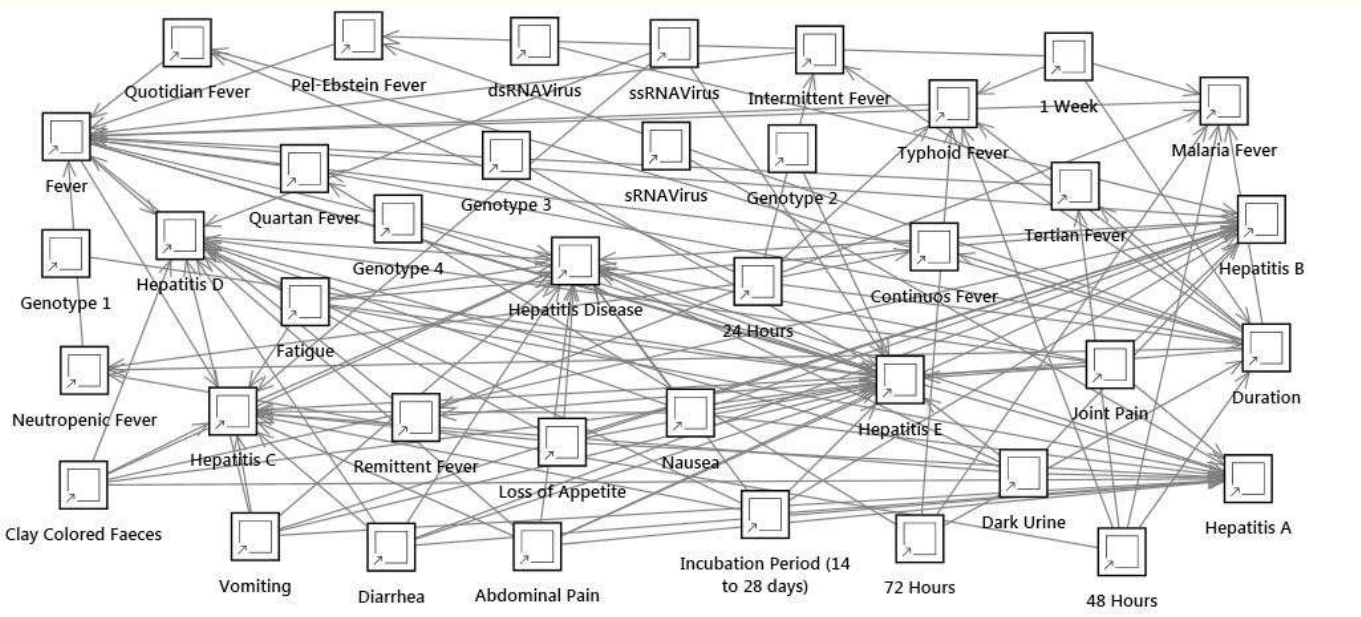

Figure 3: *Bayesian Belief Network Model for Detecting Hepatitis Disease with its symptoms and various types.*

So, to mathematically represent our model we have:

Hepatitis Disease

$$= \prod_{i=1}^{39} P(\text{Disease}_i|\text{Parents}(\text{Disease}_i) \quad (2)$$

Where,
Disease: Node with a Disease Ailment
Parents (Disease_i) = Nodes that converge on Disease $\text{Ailment}_{i..}$
The dataset was used to train and test the model. Upon completion of training and testing the BBN model, the test data converged at time series 2. The log likelihood value for each case was recorded. Figure 4 shows the BBN model convergence of Hepatitis and its symptoms at Iteration Count 2.

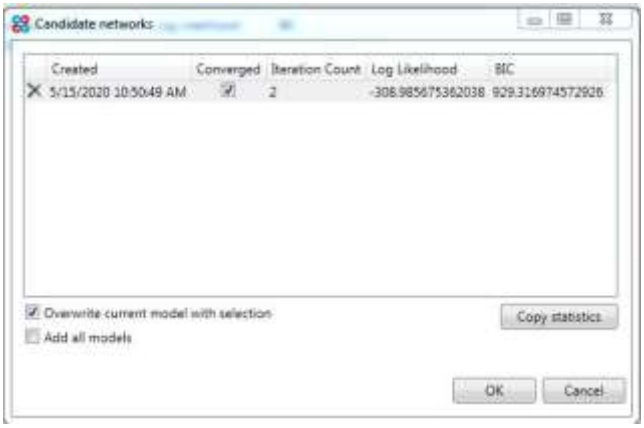

Figure 4: *Bayesian Belief Network Model for Detecting Hepatitis Disease Converging at Time Series 2.*

Figures 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 shows log likelihood batch query chart for predicting Hepatitis Disease with its symptoms, feature importance chart for nodes in the model, the in-sample anomaly detection chart, the mesh query plot for the loglikelihood of a disease causing ailment (sRNAVirus) being the reason for a Hepatitis Infection, the likelihood plot showing relation of sRNAVirus leading to a Hepatitis A infection case and its probabilities; the likelihood plot showing relation of dsRNAVirus leading to a Hepatitis B infection case and its probabilities; the likelihood plot showing relation of ssRNAVirus leading to a Hepatitis C infection case and its probabilities; the likelihood plot showing relation of ssRNAVirus leading to a Hepatitis D infection case and its probabilities; the likelihood plot showing relation of ssRNAVirus leading to a Hepatitis E infection case and its probabilities; likelihood plot showing relation of sRNAVirus leads to a Hepatitis Disease infection case and its probabilities and the loglikelihood graph for detecting Hepatitis Disease with its symptoms respectively. The result generated from the simulation indicated that the network was able to predict 99% Hepatitis on the dataset accurately and it had a loglikelihood of 12.52 on the test dataset. The figure 5 below shows the loglikelihood batch query chart for predicting Hepatitis Disease with its symptoms and various types.

International Journal of Academic Pedagogical Research (IJAPR)
ISSN: 2643-9603
Vol. 4, Issue 5, May – 2020, Pages: 19-35

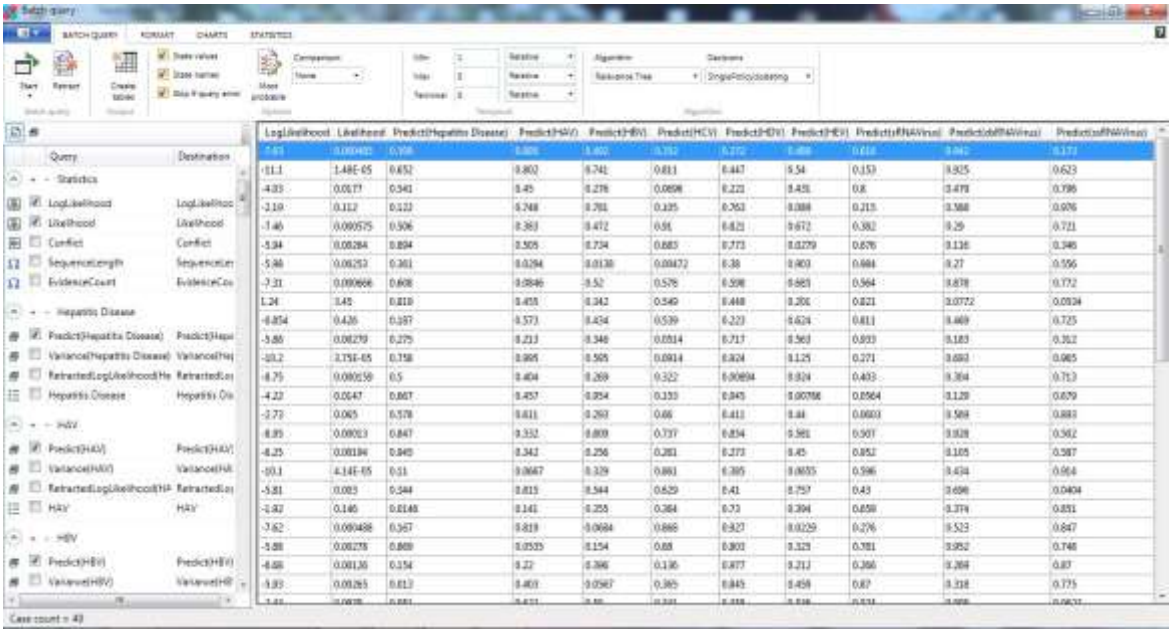

Figure 5: *The Loglikelihood Chart Batch Query for Predicting Hepatitis Disease with Its Symptoms and Various Types*

This loglikelihood chart batch shows the result of the test data. Here, 50 experimental cases were conducted and the analysis of the result generated from the test data is shown below:
In Experiment 1: The value of Predict(Hepatitis Disease) was 0.509 compared to 0.50910010145,Predict (HAV) was 0.895 compared to 0.08946247109, Predict(HBV) was 0.492 compared to 0.491508143, Predict(HCV) was 0.352 compared to 0.351631476, Predict(HDV) was 0.272 compared to 0.2716111486, Predict(HEV) was 0.489 compared to 0. 4890021654, Predict(sRNAVirus) was 0.618 compared to 0.6181001478, Predict(dsRNAVirus) was 0.942 compared to 0.941724894, Predict(ssRNAVirus) was 0.173 compared to 0.1726331092 in Experiment 1.
In Experiment 2: The value of Predict(Hepatitis Disease) was 0.652 compared to 0.651722189,Predict (HAV) was 0.802 compared to 0.801811294, Predict(HBV) was 0.741 compared to 0.740644189, Predict(HCV) was 0.811 compared to 0.810821773, Predict(HDV) was 0.447 compared to 0.446910009 Predict(HEV) was 0.54 compared to 0.5360001012, Predict(sRNAVirus) was 0.153 compared to 0.152612093, Predict(dsRNAVirus) was 0.925 compared to 0.924820091, Predict(ssRNAVirus) was 0.623 compared to 0.62281091482 in Experiment 2.
In Experiment 3: The value of Predict(Hepatitis Disease) was 0.541 compared to 0.540622189,Predict (HAV) was 0.45 compared to 0.450000194, Predict(HBV) was 0.276 compared to 0.275841012, Predict(HCV) was 0.0696 compared to 0.0696112293, Predict(HDV) was 0.221 compared to 0.22096542, Predict(HEV) was 0.431 compared to 0.430654129, Predict(sRNAVirus) was 0.8 compared to 0. 800012109, Predict(dsRNAVirus) was 0.479 compared to 0.4790011034, Predict(ssRNAVirus) was 0.796 compared to 0.795411022 in Experiment 3.
Furthermore, this experiment continues up to Experiment number 50. Hence, the system results showed a 0.0295 value difference between the prediction results and original test data of 100% resulting to 99% prediction accuracy.

The figure 6 below shows the feature importance Chart for nodes in the Bayesian Belief Network model

International Journal of Academic Pedagogical Research (IJAPR)
ISSN: 2643-9603
Vol. 4, Issue 5, May – 2020, Pages: 19-35

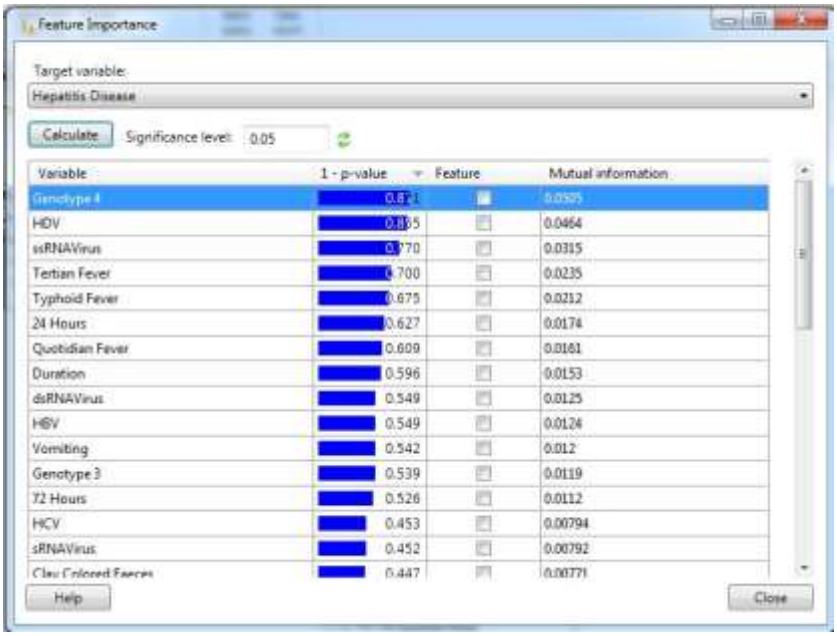

Figure 6: *The Feature Importance Chart for Nodes in the BBN Model*

The Feature Importance Chart shows p-value of the variable (nodes), Feature and Mutual information in reference to the Hepatitis Disease Node.
The p-value signifies the likelihood (probability) of the nodes being the cause of a Hepatitis Disease infection.
The Feature box is checked if that particular node is fully involved in the cause of a Hepatitis Disease infection.
The Mutual Information shows the relationship with nodes directly connected to one another (i.e. in this case the direct relationship of the nodes with the Hepatitis Disease) and assigned a value.
The Significance Level signifies the margin of error in the detection of Hepatitis Disease and its symptoms.
The figure 7 below shows the in-sample anomaly detection chart for the Bayesian Belief Network Model.

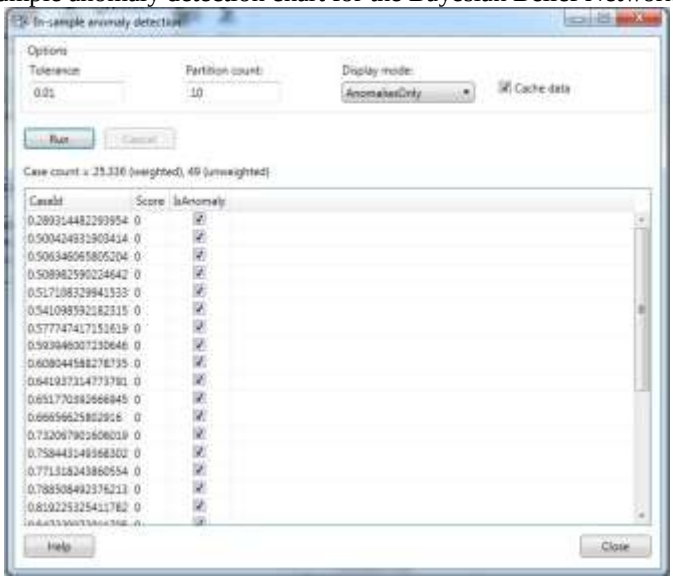

Figure 7: *The In-Sample Anomaly Detection Chart Chart for Nodes in the BBN Model*

The In-sample Anomaly Detection Chart shows 50 experimental results of detecting Hepatitis Disease. Each Case is assigned an ID(Identification value) which is the value of the Predict(Hepatitis Disease) in Figure 5 above. The IsAnomaly checkbox is checked to identify that each case is a confirmed case of Hepatitis Disease infection. The 50 cases of Hepatitis Disease has a case count value of 25.336 (weighted) which signifies the importance of the cases leading to a Hepatitis Disease infection and 49 case

International Journal of Academic Pedagogical Research (IJAPR)
ISSN: 2643-9603
Vol. 4, Issue 5, May – 2020, Pages: 19-35

count value signifies the number of cases in the pool of data available to the system for detection of Hepatitis Disease excluding the Hepatitis Disease column in the dataset pool. The tolerance is the margin of error that could be encoutered as regards to the detection of the Hepatitis Disease and its symptoms. The figure 8 below shows the mesh query plot for the loglikelihood of a disease causing ailment (sRNAVirus) being the reason for a Hepatitis Disease Infection.

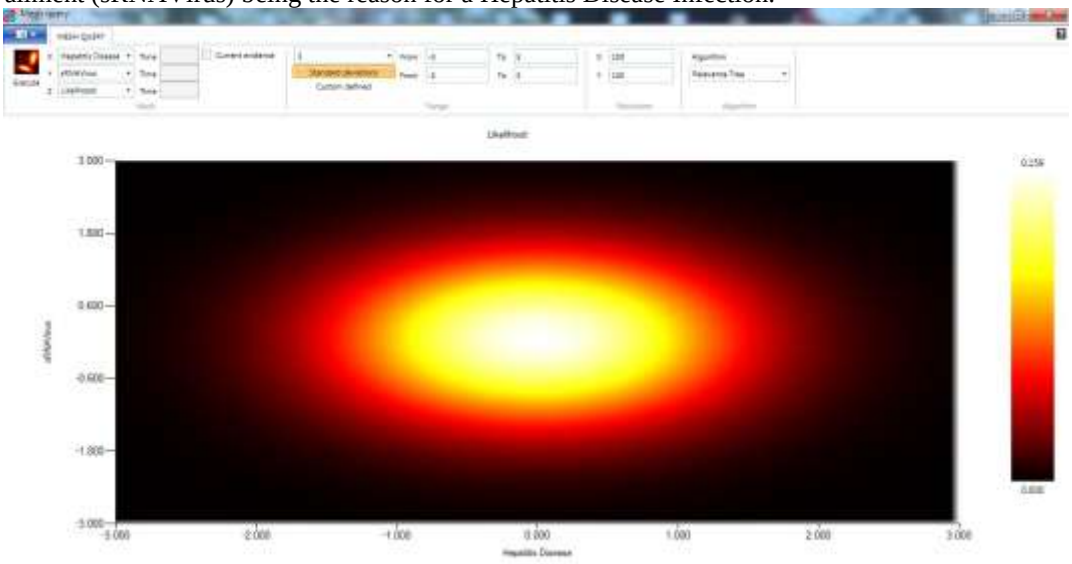

Figure 8: *The Mesh Query Plot for the Loglikelihood of a Disease Causing Virus (sRNAVirus) Being The Reason For A Hepatitis Disease Infection.*

The mesh query plot shows the loglikelihood/likelihood of a node in this case sRNAVirus being a cause of a confirmed Hepatitis infection. The Node (sRNAVirus) is plotted on the Y-axis and the other node Hepatitis Disease plotted along the X-axis.
In this context, the Red contour signifies the likelihood of a sRNAVirus being a major of cause of Hepatitis Disease infection with the contour ranging from interval (-1.600 to 1.600) on the Y-axis and interval (-1.500 to1.500) on the X-axis.
The Yellow contour shows the loglikelihood of a sRNAVirus being the reason of a Hepatitis Disease infection with the contour ranging from interval (-0.730 to 0.650) on the Y-axis and interval (-1.350 to 0.980) on the X-axis.
Figure 9 below shows the likelihood plot showing relation of sRNAVirus leading to a Hepatitis A Virus infection case and its probabilities.

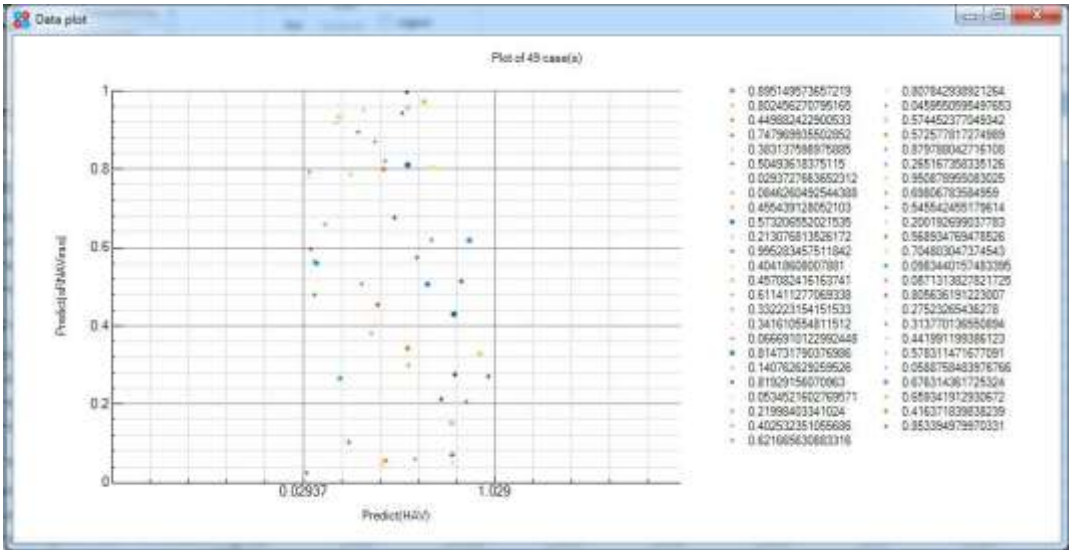

Figure 9: *The Likelihood Plot Showing Relation of sRNAVirus Leading to a Hepatitis A Virus Infection Case And Its Probabilities In The BBN Model*

The likelihood plot shows the possibility of how contact with sRNAVirus leads a Hepatitis A Virus infection case. In this plot, 50 experimental cases were taken into consideration with each colored point in the graph classified as a case and assigned a probability which is stationed on the right of the graph. The sRNAVirus on the Y-axis is plotted against Hepatitis A Virus (HAV) on the X-axis.

However, from this graph, there are five diagnostic classes of Hepatitis A Virus cases which our system was able to detect; they are asymptomatic, mild, moderate, severe, and critical classes respectively.

Asymptomatic Class: This class ranges from 0 to 0.2 on Y-axis and 0.02937 to 1.029 on X-axis. This region has 8 colored points (cases). This signifies that the 8 colored points in this region represent 8 cases of no Hepatitis A Virus infection whatsoever, hence this category of patients are categorized as being Asymptomatic.

Mild Class: This class ranges from 0.2 to 0.4 on Y-axis and 0.02937 to 1.029 on X-axis. This region has 9 colored points (cases). This signifies that the 9 colored points in this region represent 9 cases of patients with Hepatitis A Virus infection with the severity level categorized as being Mild.

Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.02937 to 1.029 on X-axis. This region has 10 colored points (cases).This signifies that the 10 colored points in this region represent 10 cases of patients with Hepatitis Disease infection with the severity level categorized as being Moderate.

Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.02937 to 1.029 on X-axis. This region has 7 colored points (cases). This signifies that the 7 colored points in this region represent 7 cases of patients with Hepatitis A virus infection with the severity level categorized as being Severe.

Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.02937 to 1.029 on X-axis. This region has 15 colored points (cases). This signifies that the 15 colored points in this region represent 15 cases of patients with Hepatitis A virus infection with the severity level categorized as being Critical.

All the 49 cases in figure 9 had a probability value less than 1; with the highest probability value of sRNAVirus causing a Hepatitis A Virus Infection reported to be 0.950878955083025 which is less than 1.

Of the 50 experimental cases, the system predicted 49 cases of Hepatitis A Virus ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 95.08% sensitivity of Hepatitis A Virus (HAV) Infection.

Figure 10 below shows the likelihood plot showing relation of dsRNAVirus leading to a Hepatitis B Virus infection case and its probabilities.

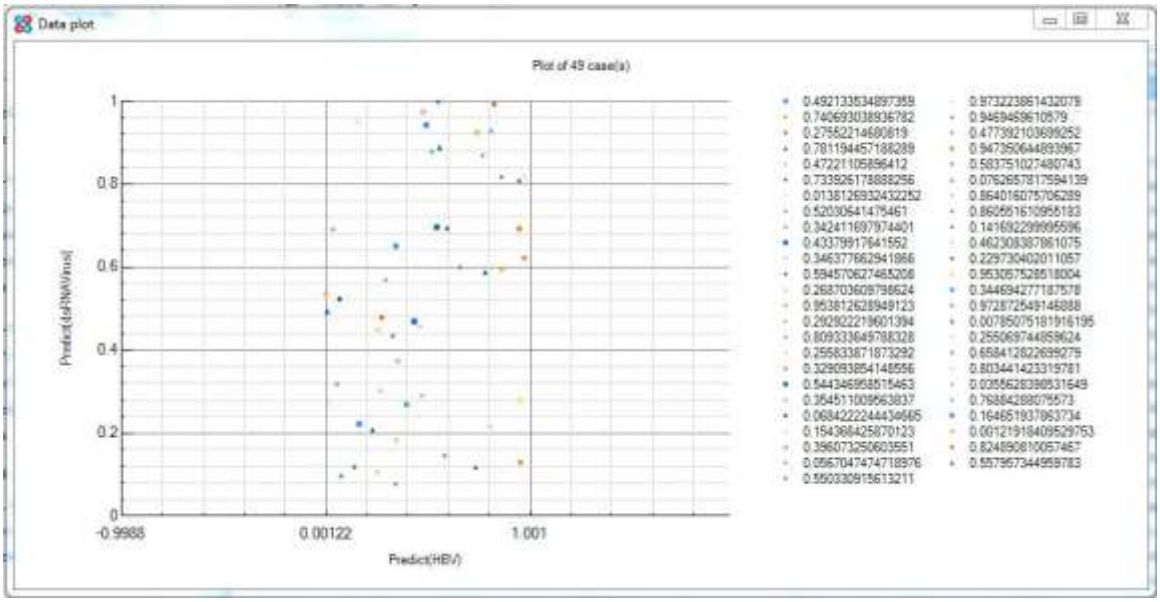

Figure 10: *The Likelihood Plot Showing Relation of dsRNAVirus Leading to a Hepatitis B Virus Infection Case And Its Probabilities In The BBN Model*

The likelihood plot shows the possibility of how contact with sRNAVirus leads a Hepatitis B Virus infection case. In this plot, 50 experimental cases were taken into consideration with each colored point in the graph classified as a case and assigned a

probability which is stationed on the right of the graph. The sRNAVirus on the Y-axis is plotted against Hepatitis B Virus (HBV) on the X-axis.
However, from this graph, there are five diagnostic classes of Hepatitis B Virus cases which our system was able to detect; they are asymptomatic, mild, moderate, severe, and critical classes respectively.
Asymptomatic Class: This class ranges from 0 to 0.2 on Y-axis and 0.00122 to 1.001 on X-axis. This region has 8 colored points (cases). This signifies that the 8 colored points in this region represent 8 cases of no Hepatitis B Virus infection whatsoever, hence this category of patients are categorized as being Asymptomatic.
Mild Class: This class ranges from 0.2 to 0.4 on Y-axis and 0.00122 to 1.001 on X-axis. This region has 10 colored points (cases). This signifies that the 10 colored points in this region represent 10 cases of patients with Hepatitis B Virus infection with the severity level categorized as being Mild.
Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.00122 to 1.001 on X-axis. This region has 11 colored points (cases).This signifies that the 11 colored points in this region represent 11 cases of patients with Hepatitis B Virus infection with the severity level categorized as being Moderate.
Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.00122 to 1.001 on X-axis. This region has 7 colored points (cases). This signifies that the 7 colored points in this region represent 7 cases of patients with Hepatitis B Virus infection with the severity level categorized as being Severe.
Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.00122 to 1.001 on X-axis. This region has 13 colored points (cases). This signifies that the 13 colored points in this region represent 13 cases of patients with Hepatitis B Virus infection with the severity level categorized as being Critical.
All the 49 cases in figure 10 had a probability value less than 1; with the highest probability value of sRNAVirus causing a Hepatitis B Virus Infection reported to be 0.973223861432079 which is less than 1.
Of the 50 experimental cases, the system predicted 49 cases of Hepatitis B Virus ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 97.32% sensitivity of Hepatitis B Virus (HBV) infection.

Figure 11 below shows the likelihood plot showing relation of ssRNAVirus leading to a Hepatitis C Virus infection case and its probabilities.

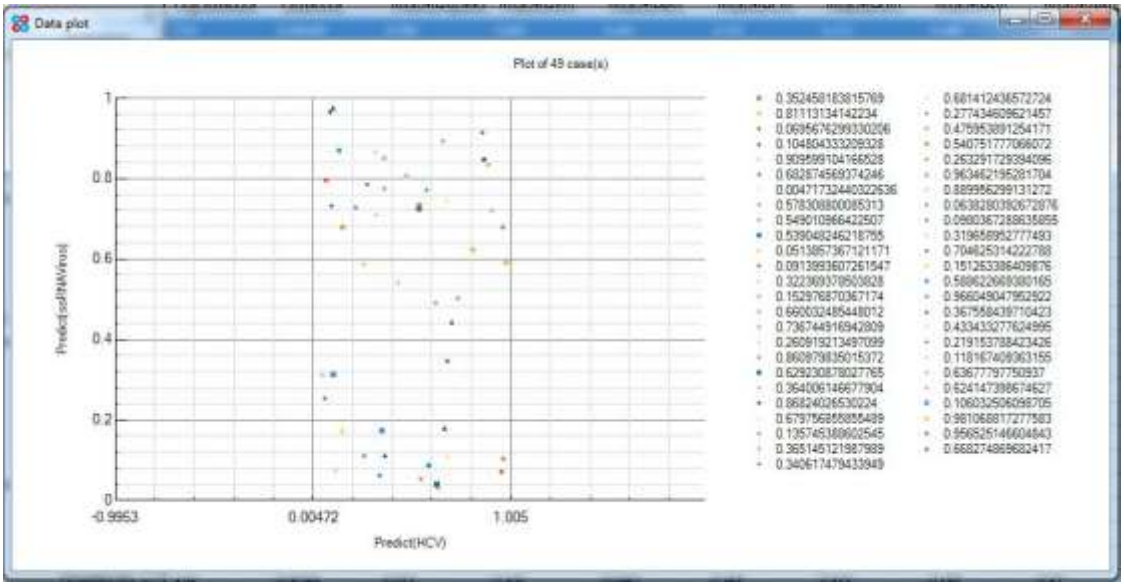

Figure 11: *The Likelihood Plot Showing Relation of ssRNAVirus Leading to a Hepatitis C Virus Disease Infection Case And Its Probabilities In The BBN Model*

The likelihood plot shows the possibility of how contact with ssRNAVirus leads a Hepatitis C Virus infection case. In this plot, 50 experimental cases were taken into consideration with each colored point in the graph classified as a case and assigned a probability which is stationed on the right of the graph. The ssRNAVirus on the Y-axis is plotted against Hepatitis C Virus (HCV) on the X-axis.
However, from this graph, there are five diagnostic classes of Hepatitis C Virus cases which our system was able to detect; they are asymptomatic, mild, moderate, severe, and critical classes respectively.

Asymptomatic Class: This class ranges from 0 to 0.2 on Y-axis and 0.00472 to 1.005 on X-axis. This region has 14 colored points (cases). This signifies that the 14 colored points in this region represent 14 cases of no Hepatitis C Virus infection whatsoever, hence this category of patients are categorized as being Asymptomatic.
Mild Class: This class ranges from 0.2 to 0.4 on Y-axis and 0.00472 to 1.005 on X-axis. This region has 4 colored points (cases). This signifies that the 4 colored points in this region represent 4 cases of patients with Hepatitis C Virus infection with the severity level categorized as being Mild.
Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.00472 to 1.005 on X-axis. This region has 7 colored points (cases).This signifies that the 7 colored points in this region represent 7 cases of patients with Hepatitis C Virus infection with the severity level categorized as being Moderate.
Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.00472 to 1.005 on X-axis. This region has 14 colored points (cases). This signifies that the 14 colored points in this region represent 14 cases of patients with Hepatitis C Virus infection with the severity level categorized as being Severe.
Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.00472 to 1.005 on X-axis. This region has 10 colored points (cases). This signifies that the 10 colored points in this region represent 10 cases of patients with Hepatitis C Virus infection with the severity level categorized as being Critical.
All the 49 cases in figure 10 had a probability value less than 1; with the highest probability value of ssRNAVirus causing a Hepatitis C Virus Infection reported to be 0.981068817277583 which is less than 1.
Of the 50 experimental cases, the system predicted 49 cases of Hepatitis C Virus ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 98.11% sensitivity of Hepatitis C Virus (HCV) Infection.

Figure 12 below shows the likelihood plot showing relation of ssRNAVirus leading to a Hepatitis D Virus infection case and its probabilities.

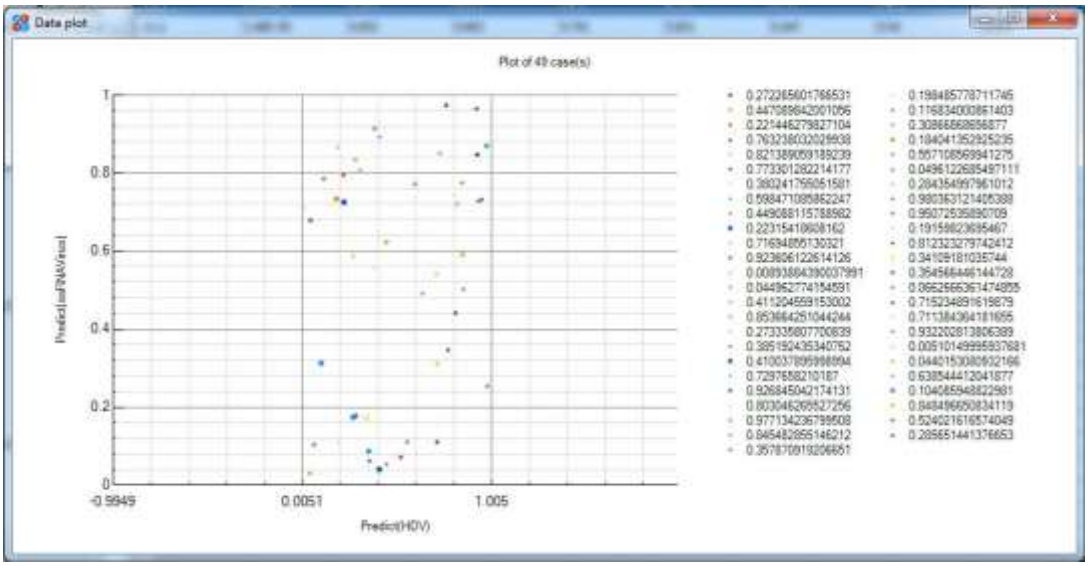

Figure 12: *The Likelihood Plot Showing Relation of Predict ssRNAVirus Leading to a Hepatitis D Virus Infection Case And Its Probabilities In The BBN Model*

The likelihood plot shows the possibility of how contact with ssRNAVirus leads a Hepatitis D Virus infection case. In this plot, 50 experimental cases were taken into consideration with each colored point in the graph classified as a case and assigned a probability which is stationed on the right of the graph. The ssRNAVirus on the Y-axis is plotted against Hepatitis D Virus (HDV) on the X-axis.
However, from this graph, there are five diagnostic classes of Hepatitis D Virus cases which our system was able to detect; they are asymptomatic, mild, moderate, severe, and critical classes respectively.
Asymptomatic Class: This class ranges from 0 to 0.2 on Y-axis and 0.0051 to 1.005 on X-axis. This region has 14 colored points (cases). This signifies that the 14 colored points in this region represent 14 cases of no Hepatitis D Virus infection whatsoever, hence this category of patients are categorized as being Asymptomatic.

Mild Class: This class ranges from 0.2 to 0.4 on Y-axis and 0.0051 to 1.005 on X-axis. This region has 4 colored points (cases). This signifies that the 4 colored points in this region represent 4 cases of patients with Hepatitis D Virus infection with the severity level categorized as being Mild.
Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.0051 to 1.005 on X-axis. This region has 7 colored points (cases).This signifies that the 7 colored points in this region represent 7 cases of patients with Hepatitis D Virus infection with the severity level categorized as being Moderate.
Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.0051 to 1.005 on X-axis. This region has 13 colored points (cases). This signifies that the 13 colored points in this region represent 13 cases of patients with Hepatitis D Virus infection with the severity level categorized as being Severe.
Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.0051 to 1.005 on X-axis. This region has 11 colored points (cases). This signifies that the 11 colored points in this region represent 11 cases of patients with Hepatitis D Virus infection with the severity level categorized as being Critical.
All the 49 cases in figure 10 had a probability value less than 1; with the highest probability value of ssRNAVirus causing a Hepatitis D Virus Infection reported to be 0.977134236799508 which is less than 1.
Of the 50 experimental cases, the system predicted 49 cases of Hepatitis D Virus ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 97.71% sensitivity of Hepatitis D Virus (HDV) infection.

Figure 13 below shows the likelihood plot showing relation of ssRNAVirus leading to a Hepatitis E Virus infection case and its probabilities.

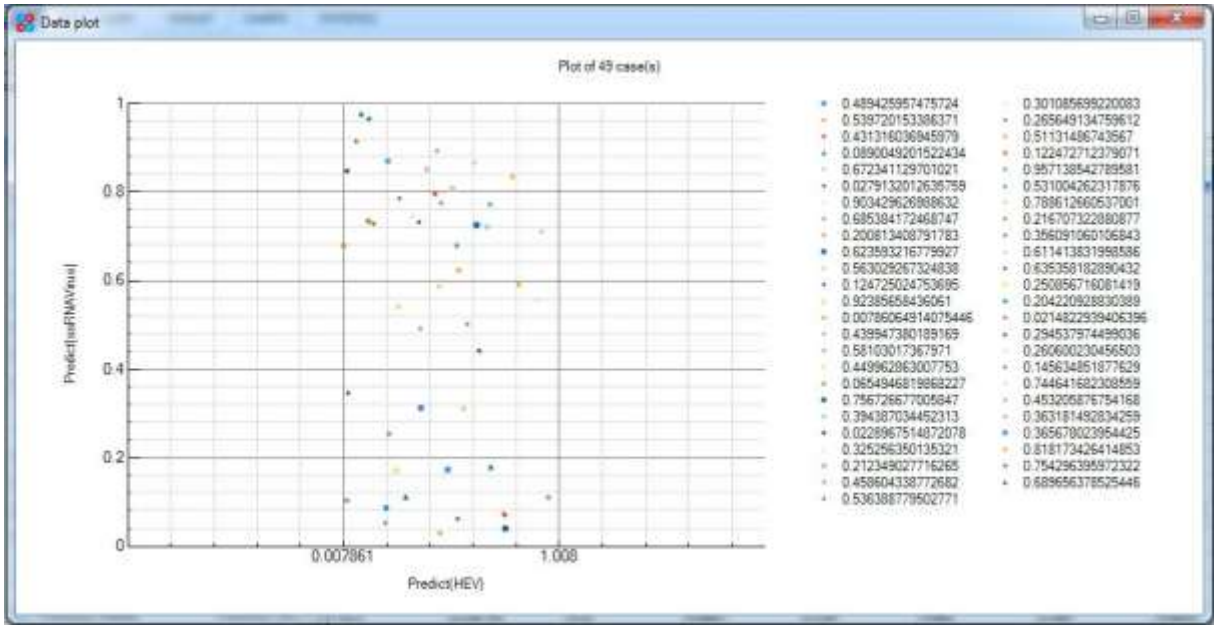

Figure 13: *The Likelihood Plot Showing Relation of ssRNAVirus Leading to a Hepatitis E Virus Infection Case And Its Probabilities In The BBN Model*

The likelihood plot shows the possibility of how contact with ssRNAVirus leads a Hepatitis E Virus infection case. In this plot, 50 experimental cases were taken into consideration with each colored point in the graph classified as a case and assigned a probability which is stationed on the right of the graph. The ssRNAVirus on the Y-axis is plotted against Hepatitis E Virus (HEV) on the X-axis.
However, from this graph, there are five diagnostic classes of Hepatitis E Virus cases which our system was able to detect; they are asymptomatic, mild, moderate, severe, and critical classes respectively.
Asymptomatic Class: This class ranges from 0 to 0.2 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 13 colored points (cases). This signifies that the 13 colored points in this region represent 13 cases of no Hepatitis E Virus infection whatsoever, hence this category of patients are categorized as being Asymptomatic.
Mild Class: This class ranges from 0.2 to 0.4 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 4 colored points (cases). This signifies that the 4 colored points in this region represent 4 cases of patients with Hepatitis E Virus infection with the severity level categorized as being Mild.

Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 7 colored points (cases).This signifies that the 7 colored points in this region represent 7 cases of patients with Hepatitis E Virus infection with the severity level categorized as being Moderate.
Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 15 colored points (cases). This signifies that the 15 colored points in this region represent 15 cases of patients with Hepatitis E Virus infection with the severity level categorized as being Severe.
Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 10 colored points (cases). This signifies that the 10 colored points in this region represent 10 cases of patients with Hepatitis E Virus infection with the severity level categorized as being Critical.
All the 49 cases in figure 10 had a probability value less than 1; with the highest probability value of ssRNAVirus causing a Hepatitis E virus Infection reported to be 0.957138542379071 which is less than 1.
Of the 50 experimental cases, the system predicted 49 cases of Hepatitis E Virus ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 95.71% sensitivity of Hepatitis E Virus (HEV) Infection.

Figure 14 below shows the likelihood plot showing relation of sRNAVirus Leading to a Hepatitis Disease Infection Case And Its Probabilities In The BBN Model

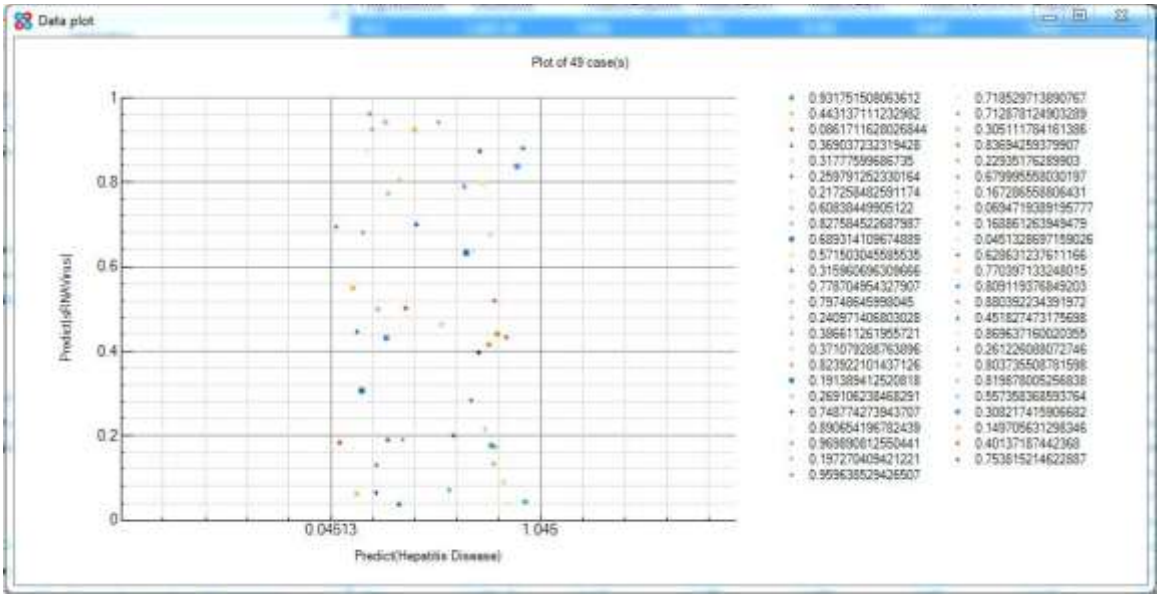

Figure 14: *The Likelihood Plot Showing Relation of sRNAVirus Leading to a Hepatitis Disease Infection Case And Its Probabilities In The BBN Model*

The likelihood plot shows the possibility of how contact with sRNAVirus leads a Hepatitis E Virus infection case. In this plot, 50 experimental cases were taken into consideration with each colored point in the graph classified as a case and assigned a probability which is stationed on the right of the graph. The sRNAVirus on the Y-axis is plotted against Hepatitis Disease on the X-axis.
However, from this graph, there are five diagnostic classes of Hepatitis Disease cases which our system was able to detect; they are asymptomatic, mild, moderate, severe, and critical classes respectively.
Asymptomatic Class: This class ranges from 0 to 0.2 on Y-axis and 0.04513 to 1.045 on X-axis. This region has 14 colored points (cases). This signifies that the 14 colored points in this region represent 14 cases of no Hepatitis Disease infection whatsoever, hence this category of patients are categorized as being Asymptomatic.
Mild Class: This class ranges from 0.2 to 0.4 on Y-axis and 0.04513 to 1.045 on X-axis. This region has 4 colored points (cases). This signifies that the 4 colored points in this region represent 4 cases of patients with Hepatitis Disease infection with the severity level categorized as being Mild.
Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.04513 to 1.045 on X-axis. This region has 11 colored points (cases).This signifies that the 11 colored points in this region represent 11 cases of patients with Hepatitis Disease infection with the severity level categorized as being Moderate.

Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 19 colored points (cases). This signifies that the 19 colored points in this region represent 19 cases of patients with Hepatitis Disease infection with the severity level categorized as being Severe.
Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.007861 to 1.008 on X-axis. This region has 11 colored points (cases). This signifies that the 11 colored points in this region represent 11 cases of patients with Hepatitis Disease infection with the severity level categorized as being Critical.
All the 49 cases in figure 10 had a probability value less than 1; with the highest probability value of sRNAVirus causing a Hepatitis Disease Infection reported to be 0.969890812550441 which is less than 1.
Of the 50 experimental cases, the system predicted 49 cases of Hepatitis Disease ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 96.98% sensitivity of Hepatitis Disease infection.
The figure 15 shows the Loglikelihood Graph for Detecting Hepatitis Disease with its symptoms.

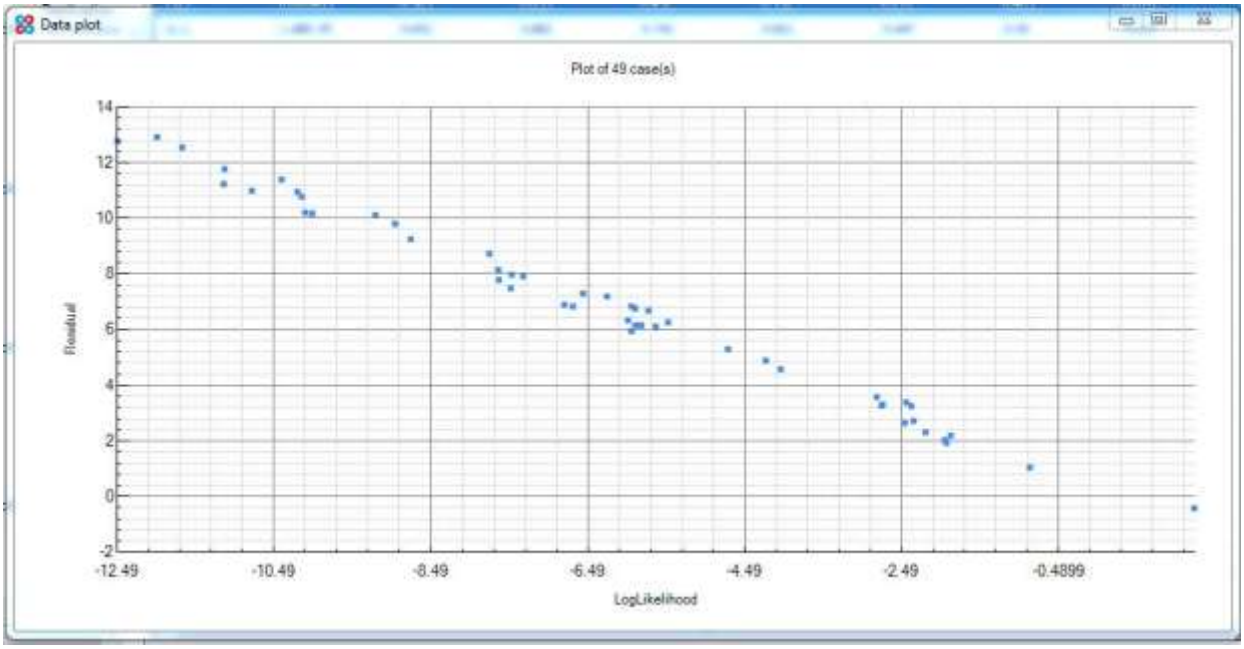

Figure 15: *The Loglikelihood Graph for Detecting Hepatitis Disease with its Symptoms*

This loglikelihood graph for detecting Hepatitis Disease shows the residual values on the vertical axis plotted against the loglikelihood values on the horizontal axis which are independent variables.
A residual value is a measure of how much a regression line vertically misses a data point. Regression lines are the best fit of a set of data. The lines are categorized as averages; a few data points will fit the line and others will miss.
In this graph, it shows that 50 experimental cases resulted in value of 12.52, 12.50, 11.65, 11.25, 11.05, 11.00…….. and -0.25 respectively.
Ideally, residual values should be equally and randomly spaced around the horizontal lines. Taking a view of the system' experimental results values obtained from the horizontal lines on the graph, it can be seen that the point where the highest residual value and the loglikelihood independent variable attained meets at -12.49 on the horizontal line with 14 being the highest value that can be reached on the vertical line.
The residual value attained is 12.52 and loglikelihood independent value is -12.49, the difference between both values is 0.03 which is the difference between the values of the prediction results and original test data of 100% in figure 5.
Hence, in this system the highest residual value, a loglikelihood independent value can attain is 14. With 14, being the 100 % residual value mark, to get our prediction accuracy percentage, we have predicted value subtracted from highest residual value i.e. 100% -0.03= 99.97% residual loglikelihood percentage.
Furthermore, the Likelihood graph results in figure 9, 10,11,12,13 and 14 above showed all classes of severity status of Hepatitis Disease Infection cases ranging from asymptomatic, mild, moderate, severe, and critical classes respectively with their probabilities while the Loglikelihood graph in figure 15 showed the 99.97% prediction accuracy of the system.
Thus, the probability of having Hepatitis infection given there is evidence of ailments and factors that influence diagnosis of the aforesaid disease is denoted as:

P(Hepatitis Disease| Abdominal Pain, Clay Colored Faeces, Continuous Fever, Dark Urine, Diarrhea, Double-Stranded DNA Virus (dsDNA Virus), Duration, Fatigue, Fever, Genotype 1, Genotype 2, Genotype 3, Genotype 4, Hepatitis A Virus, Hepatitis B Virus, Hepatitis C Virus, Hepatitis D Virus, Hepatitis E Virus, Incubation Period, Intermittent Fever, Joint Pain, Loss of Appetite, Malaria Fever, Nausea, Neutropenic Fever, Pel-Ebstein Fever, Quartan Fever, Quotidian Fever, Remittent Fever, Single Stranded RNA Virus (ssRNA Virus), Symmetrical RNA Virus (sRNA Virus), Tertian Fever, Typhoid Fever, Vomiting, 1Week, 24 Hours,48 Hours and 72 Hours) = 0.969890812550441.

From the experiment, it can be seen that our model has a higher residual log likelihood value which is 12.52, an overall prediction accuracy of 99.97%; 96.98%, 95.08%, 97.32%, 98.11%, 97.71% and 95.71% sensitivity of Hepatitis Disease, HAV, HBV, HCV, HDV and HEV in that order.
Finally, comparing the 99.97% prediction accuracy of our model with the experiments conducted by [18, 19, 20 and 23] which has 98.44%, 94.58%,92.2% and 93.06% prediction accuracy respectively, it is obvious our model has a better prediction accuracy. The higher prediction accuracy achieved by our model could be due to the range of the dataset used in training and testing the model as well as its ability to predict the overlapping symptoms Hepatitis disease shares with liver and flu-like diseases, hence aiding the high detection accuracy of the aforesaid disease.

5. Conclusion

Hepatitis Disease is a contagious disease that is quite difficult to detect due to the overlapping symptoms the disease shares with other liver and flu-like diseases. In time past, several methods have been employed in detecting Hepatitis Disease and its various types with the aim of curbing untimely deaths of patients due to lack of early diagnosis of the said disease which medical professionals need to improve on.
In this paper, we utilized a supervised machine learning technique called Bayesian Belief Network model to predict Hepatitis Disease, its symptoms and various types. The network had 39 nodes with each node representing an exclusive ailment and factor that influence diagnosis of Hepatitis Disease. The model was trained and tested and had an overall accuracy of 99.97% in predicting Hepatitis Disease with its symptoms; with 95.08%, 97.32%, 98.11%, 97.71% and 95.71% sensitivity of HAV, HBV, HCV, HDV and HEV in that order. The system can be deployed in health facilities to help provide information which will be used to detect Hepatitis Disease and its symptoms and various types. It will also bring about improvement in the following areas: Prediction of Hepatitis Disease and various its types, Detection of Hepatitis Disease and Diagnosis of liver diseases with similar symptoms as Hepatitis Disease and its various types.

REFERENCES

1. Zoroddu, M.A., Aashet, J., Crisponi, G., Medici, S., Peana,M., and Nurchi, V.M. (2019): "The Essentials Metals for Humans: A Brief Overview". Journal of Inorganic Biochemistry. 195:.DOI:10.1016/j.jinorgbio.2019.pp:120-129.
2. Kong F. and Singh R.P. (2008):"Disintegration of Solid Foods in Human Stomach". J. Food Sci. 73(5): R67–80.doi:10.1111/j.1750-3841.2008.00766.x.PMID 18577009.
3. Saladin, K. (2011): "Human Anatomy". McGraw Hill. ISBN 9780071222075, pp: 674–679.
4. Scully, J.L. (2004): "What is a disease?". European Molecular Biology Organization 2004. EMBO reports Vol 5, No 7, 2004. DOI:10.1038/sj.embor.7400195. PMCID:PMC1299105, PMID:15229637. Retrieved 2nd April 2020 from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1229105/. pp:650-653
5. World Health Organization (2018): "Non-Communicable Diseases". Retrieved 1st April 2020 from URL: https//www.who.int/news-room/facts-sheets/detail/noncommunicable-diseases/
6. Egwari, L.O (2015): "Communicable and Non-Communicable Diseases in Nigeria: A Synthesis". Retrieved from URL: https://www.nimr.ng/wp-content/uploads/2015/11/3. pp: 13-16
7. World Health Organization (2020):"Infectious Diseases". Retrieved 11th April, 2020 from URL:https//www.who.int/topics/infectious_diseases/en/
8. Koonin, E.V., Senkevich, T.G. and Dolja, V.V. (2006): "The Ancient Virus World and Evolution of Cells". Biology Direct. 1 (1): 29. doi:10.1186/1745-6150-1-29 .PMC 1594570 . PMID 16984643. Retrieved 3rd April 2020 from URL: https://www.pubmed.ncbi.nlm.nih.gov/16984643/
9. Lawrence, C.M., Menon, S., Eilers, B.J., Bothner, B., Khayat, R., Douglas, T. and Young, M.J. (2009): "Structural and Functional Studies of Archaeal Viruses". The Journal of Biological Chemistry. 284 (19): 12599–603. DOI:10.1074/jbc.R800078200 . PMC 2675988 .PMID 19158076. Retrieved 9th April 2020 from URL: https://www.ncbi,nlm.nih.pmc/articles/PMC2675988.
10. WHO (2016):"Global Health Sector Strategy On Viral Hepatitis; Towards Ending Viral Hepatitis 2016-2021".Retrieved from URL: Retrieved from 12th May, 2020 from URL: https://www.apps.who.int/iris/bitstream/handle/10665/246177/WHO-HIV-2016-eng.pdf/

11. WHO (2019):"Hepatitis A". Retrieved 12th May, 2020 from URL: https://www.who.int/news-room/fact-sheets/detail/hepatitis-a
12. WHO (2019):"Hepatitis B". Retrieved 12th May, 2020 from URL: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b
13. WHO (2019):"Hepatitis C". Retrieved 12th May, 2020 from URL: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c
14. WHO (2020):"Hepatitis D". Retrieved 12th May, 2020 from URL: https://www.who.int/news-room/fact-sheets/detail/hepatitis-d
15. WHO(2019):"Hepatitis E". Retrieved 12th May, 2020 from URL: https://www.who.int/news-room/fact-sheets/detail/hepatitis-e
16. Ogah U.S, Zirra, P.B. and Sarjiyus,O. (2017): "Knowledge Based System Design For Diagnosis of Hepatitis B Virus (HBV) Using Generalized Regression Neural Network (GRNN)". American Journal of Computing and Engineering, Vol.1, Issue No. 1, 2017, pp:1-19.
17. Dakshata, P. and Seema S. (2011): "Artificial Intelligence Based Expert System For Hepatitis B Diagnosis". International Journal of Modeling and Optimization, Vol. 1, No. 4, October 2011. pp: 362-366.
18. Metwally, N.F., AbuSharekh, E.K. and Abu-Naser, S.S. (2018): "Diagnosis of Hepatitis Virus Using Artificial Neural Network".International Journal of Academic Pedagogical Research (IJAPR), ISSN: 2000-004X, Vol. 2 Issue 11, November – 2018, pp: 1-7.
19. Neshat, M., Sargolzaei, M., Toosi, A.N. and Masoumi, A. (2012): "Hepatitis Disease Diagnosis Using Hybrid Case Based Reasoning and Particle Swarm Optimization". International Scholarly Research Network, ISRN Artificial Intelligence,Volume 2012, Article ID 609718,doi:10.5402/2012/609718, pp:1-6.
20. Ahmad, G., Khan, M.A., Abbas, S., Athar, A., Khan, B.S. and Aslam, M.S. (2019): "Automated Diagnosis of Hepatitis B Using Multilayer Mamdani Fuzzy Inference System". Hindawi Journal of Healthcare Engineering,Volume 2019, Article ID 6361318,https:/doi.org/10.1155/2019/6361318. pp:1-11.
21. Ravanshad, M., Sabahi, F., Falahi, S., Kenarkoohi, A., Amini-Bavil-Olyaee, S., Hosseini, S.Y., Madvar, H.R. and Khanizade, S. (2011): "Prediction of Hepatitis B Virus Lamivudine Resistance Based on YMDD Sequence Data Using an Artificial Neural Network Model". Kowsar M.P.Co,Hepatitis Monthly, 2011;11(2): pp:108-113.
22. Elsharif, A.A., Al-qumboz, M.N.A., Alshawwa, I.A.., AbuMettleq, A.S., Dheir, I.M. and Abu-Naser, S.S. (2019): "Hepatitis Expert System Diagnosis Using Sl5 Object". International Journal of Academic Information Systems Research (IJAISR), ISSN: 2000-002X, Vol. 3 Issue 4, April – 2019, pp: 10-18.
23. Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O. and Akbari, E. (2019): "A Predictive Method for Hepatitis Disease Diagnosis Using Ensembles of Neuro-Fuzzy Technique". Journal of Infection and Public Health 12 (2019), pp: 13–20.
24. Simon, A., Deo, M.S., Venkatesan, S. and Babu, D.R.R. (2015): "Machine Learning and its Applications". International Journal of Electrical Sciences & Engineering (IJESE), Volume 1, Issue 1; 2015 pp: 22-24.
25. Ben-Gal, I. (2007): "Bayesian Networks". Encyclopedia of Statistics in Quality and Reliability. John Wiley and Sons, Ltd. Retrieved April 15th 2020 from URL: https://www.eng.tau.ac.il/bengal/BN.pdf/. pp:1-3
26. UCI Machine Learning Dataset for Hepatitis (2020): "Hepatitis Dataset for Machine Learning". Retrieved 14th May, 2020 from URL: https//www.archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/hepatitis.data".