A Machine Learning Based Approach for Detecting Dengue Haemorrhagic Fever

[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: *Dengue Haemorrhagic Fever is a mosquito-borne disease caused by the arthropod genus mosquito which is found in the tropics and sub-tropics region of the world; transmitting the dengue virus otherwise called serotypes which is a member of the RNA virus group and Flaviviridae virus family from one host to another. Hence, it is categorized as an endemic ailment due to its mode of transmission. Conversely, the symptoms begin to manifest in host after three to fourteen days after infection. Some of its symptoms might include loss of appetite, headache, high fever, metallic taste, vomiting, rash and joint pains just to name a few. This dengue 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 Dengue Haemorrhagic Fever and its overlapping symptoms. Hence, in this paper, we proposed and simulated a model to predict Dengue Haemorrhagic Fever using a machine learning technique called Bayesian Belief Network. The model was designed using Bayes Server and tested with data collected from Dengue Haemorrhagic Fever medical repository. The model had a 99.84% prediction accuracy.*

Keywords: Dengue Fever; Dengue Haemorrhagic Fever; Machine Learning; Bayesian Belief Network; Prediction; Detection.

1.0 Introduction

Fever is defined as the rise in body temperature which surpasses the normal temperature of 37° Celcius or 98.6° Farenheint.Conversely, fever is categorized as one of the natural defenses of the body's fight against pathogenic microorganisms such as bacteria and viruses which find it difficult to exist in environment of very high temperature. This ailment affects persons of all ages ranging from children to adults. However, stern fever is a symptom of a severe condition that needs prompt medical attention [1].

Fever, otherwise known as pyrexia and febrile response can be uncomfortable in adults when the temperature rises to 103^oF equivalent to 39.4^oCelcius or even higher. However, in the case of infants and toddlers, a slight increase in temperature may signify a serious illness.

In [2], it was stated that fever may be a result of numerous medical conditions varying from non-serious to life-threatening which are characterized by increase in body temperature.

Nevertheless, fever with increment in body temperature are of different types with respect to their duration such as continuous fever (a fever where the temperature stays above normal throughout the day), quotidian fever (a fever that has occurs at intervals within 24 hours), intermittent fever (fever that the temperature rises for a certain period and later returns back to normal), quartan fever (fever with high temperature which exceeds 72 hours), Tertian fever (fever with high temperature of a 48 hours period), Pel-Ebstein fever (fever of more than a week fluctuating between high and low), malaria fever, typhoid fever and dengue fever just to name a few. Of the aforementioned fevers, dengue fever is the most dreaded of all due to its mode of transmission and contagiousness of the dengue virus, hence categorized as an endemic ailment.

Viruses are submicroscopic communicable agents that replicates itself solely inside living cells of an organism as avowed by [3]. Viruses have the capability to infect all forms of living organisms such as plants, animals, microorganisms and humans. Examples of such viruses are dsDNA Virus (e.g. Adenoviruses, Herpesviruses, Poxviruses), ssDNA Virus (e.g.Parvoviruses), dsDNA-RT virus (e.g. Hepadnaviruses) and RNA virus (Flaviviridae or Flavivirus) respectively just to name a few. Of all the aforesaid viruses, Flaviviridae which is a member of the RNA virus family is the most dreadful of all. Furthermore, Flaviviridae causes a deadly disease called Dengue Haemorrhagic Fever virus which is ruthless if not diagnosed early as stated by [4].

In [5], Dengue Haemorrhagic Fever Virus (DHF) was defined as a ruthless and life threatening ailment that affects persons ranging from infants, older children and adults in respective of age which has similarities with symptoms of influenza caused by breeding of arthropod genus mosquito. The dengue viruses is of four types namely DenV1, DenV2, DenV3 and DenV4 respectively and are each called serotypes. This difference between each of these serotypes solely relies on the ability to stimulate the production of specific antibodies. The viruses are transmissible by a kind of mosquito called Aedes aegypti that bites humans during the day in tropical and subtropical (very hot and humid) regions of the world. In addition, the symptom of this virus starts within three to fourteen days after infection and varies in humans according to age. Children below the age of 5 might be asymptomatic or have similar type of fever and rashes on their bodies, in comparison with much older children and adults; they are prone to have harsh

symptoms such as high fever that commences swiftly usually hitting its climax with a short period of time, severe headache, muscular and joint pains, pain at the rear of the eyes, nausea, vomiting and rashes on the body.

Furthermore, World Health Organization acknowledged that DHF as a virus and deadly mosquito-borne disease is the most hazardous of all mosquito-borne diseases worldwide with the estimate of over 2.5 billion people being at risk of this dengue infection and tens of millions resulting to death yearly due to lack of early diagnosis of the disease [5,6].

In [6], it was affirmed that due to lack of early diagnosis of this disease in persons, lots of death of patients were recorded. Hence, there is need for system to aid early diagnosis of this ailment to curb the untimely death of patients.

In time past, artificial intelligence has been applied in diagnosing Dengue Haemorrhagic Fever in the works of [6, 7,8,9,10,11,12,13,14, 15 and 16] but they generated a lot of false negative during testing and were unable to detect Dengue Haemorrhagic fever due to the overlapping symptoms the disease shares with some viral and mosquito borne diseases.

In this paper, a supervised machine learning technique called Bayesian Belief Network (BBN) was utilized in diagnosing Dengue Haemorrhagic Fever with its symptoms. BBN is a versatile probabilistic network that merges expert knowledge and investigational 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 influence 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 Dengue Haemorrhagic Fever using Artificial Intelligence.

In [6], an expert system to predict dengue fever which utilized Data Mining and Artificial Intelligence was developed based on symptoms provided by the users of the expert system. The system demonstrated the ability to predict the said ailment with high prediction accuracy. However, the system failed to detect to mosquito-borne diseases with overlapping symptoms with Dengue Haemorrhagic Fever and other types of fevers. Also, usage of Data Mining technique violates system user's privacy, it has performance issues as regards results generated using this technique; expert systems have implementation and high maintenance cost and difficulty in creating inference rules.

In [7], an expert system to diagnose dengue fever that employed SL5 Object Language was developed. The system detected the said disease with high detection accuracy. The system was user-friendly and quite easy to use. Moreover, the expert system development is capital-intensive with a high maintenance cost. The system also failed to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers.

In [8], an expert system that employed Fuzzy Logic to diagnose dengue fever early was developed. The system diagnosed dengue fever early with high detection accuracy. Despite the high detection accuracy, the system failed to make bi-directional inference on the dataset; it could not handle the problem of uncertainty and it failed to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers.

In [9], a soft expert system that relied on soft set theory and Fuzzy Logic to diagnose dengue fever was designed. The expert system demonstrated the ability to measure risk level of dengue fever in patients with a high detection accuracy based on input variables from dataset. Nevertheless, the system failed to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers; the system reasoning module could not handle issue of uncertainty and make bi-directional inferences.

In [10], a decision support system to diagnose malaria and dengue fever which employed Fuzzy Logic was developed. The system diagnosed malaria and dengue fever based on symptoms with high detection accuracy with 91.3% accuracy. Conversely, the system reasoning module could not handle problem of uncertainty and make bi-directional inferences; with the system unable to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers.

In [11], a system that predicted dengue infection which utilized WEKA Data Mining tool. The system showed the capability of detecting dengue infections in patients with 92% detection accuracy. On the other hand, the Data mining technique utilized by the system is resource and capital intensive; there is a high risk of unauthorized disclosure of sensitive information and has performance issues as regards results generated using this technique. Also, the system failed to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers.

In [12], an expert system that detected dengue fever using early signs and symptoms based on Fuzzy Logic and Data Mining was developed. The system detected the dengue fever with high detection accuracy. However, the system reasoning module could not handle issue of uncertainty and make bi-directional inferences; Data mining technique utilized by the system is resource and capital intensive. Also, the system failed to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers.

In [13], a decision support expert system which employed Fuzzy Logic and Simple Additive Weighting was designed. The system detected dengue fever with a high detection accuracy based on dataset. On the other hand, the system failed to make bi-directional

inferences and was not able to handle the issue of uncertainty. Furthermore, the system was unable to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers.

In [14], a system that utilized Artificial Neural Network (ANN) to detect dengue fever was developed. The system classified dengue fever with 95% accuracy based on the dataset. However, the system failed to detect to mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers. The system neural network is capital and time intensive; it does not provide information about the relative significance of the various parameters, slow convergence speed, less generalizing performance, issue of arriving at local minimum and has over-fitting problems.

In [15], a hybrid system that combined Multi-Swarm Optimization (MSO) and Multi-Layer Perceptron (MLP) classifier to detect serotypes of dengue fever was developed. The system showcased its ability to detect the serotypes of dengue fever with 85.18% detection accuracy. Conversely, the system failed to detect mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers. With the use of MSO, it is not easy to define preliminary design constraints; MLP training is time-consuming and convergence speed is very slow.

In [16], a system that utilized Alternating Decision Trees (ADT) for early diagnosis of dengue fever was developed. The system demonstrated its ability to detect dengue fever with 84% detection accuracy based on the dataset. Nevertheless, the system was unable to detect mosquito-borne diseases with overlapping symptoms with dengue fever and other types of fevers. ADT are usually unstable, relatively inaccurate and results derived from the decision trees are sometimes biased which are usually in support of attributes that have gained more levels in the generated decision tree.

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) [17].

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 [18]. 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:

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

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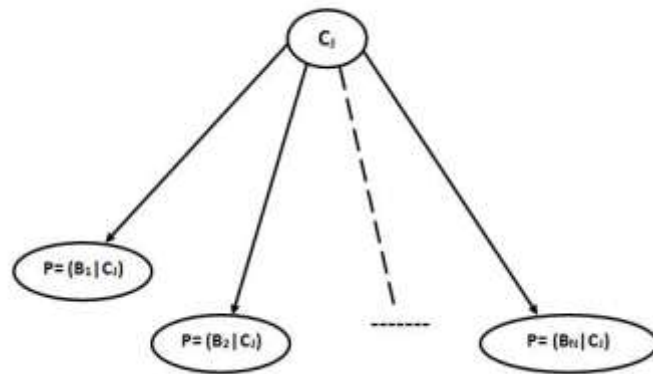
Figure 1: *A Pictorial Representation of a Bayesian Belief Network*

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 Dengue Haemorrhagic Fever using a supervised machine learning technique called Bayesian Belief Network (BBN). A model consisting of 58 nodes where each node represents a form of disease ailment will be designed using Bayes Server. A Dengue Haemorrhagic Fever Virus 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 Dengue Haemorrhagic Fever.

4. Methodology

Simulation, Result and Discussion

The dataset used in training, testing and predicting Dengue Haemorrhagic Fever was retrieved from [19]. The dataset consist of 58 disease ailments and each ailment has a value which represents the probability of such disease ailment causing Dengue Haemorrhagic Fever (DHF). The ailments, disease causing agents and other factors taken into cognizance in the diagnosis of Dengue Haemorrhagic Fever are Abdominal Pain, Anopheles Mosquito, Arthropod Genus Mosquito, Chest Pain, Chills, Continuous Fever, Constipation, Cough, Dengue Haemorrhagic Fever, DenV1, DenV2, DenV3, DenV4, Diarrhea, Fatigue, Fever, Flaviviridae, Haemoglobin, Headache, Hematocrit, High Temperature, Hot Weather, Humid Weather, Influenza, Intermittent Fever, Joint Pain, Loss of Appetite, Malaria, Metallic Taste, Muscle Pain, Nausea, Neutropenic Fever, Pel-Ebstein Fever, Platelets, Plasmodium Falciparum, Plasmodium Malariae, Plasmodium Knowlesi, Plasmodium Ovale, Plasmodium Vivax, Quartan Fever, Quotidian Fever, Rash, Region, Remittent Fever, RNA Virus, Serotypes, Sore Throat, Sweating, Swollen Abdomen, Sub-Tropical Region, Tertian Fever, Tropical Region, Typhoid Fever, Vomiting, Weakness, Weather, and Weight Loss respectively. Figure 2 below shows a sample the dataset.

Cough	DenV1	DenV2	DenV3	DenV4	DHF	Diarrhea	Fatigue	Fever	Flaviviridae	Haemoglobin	Headache	Hematocrit
-0.12	0.232	-0.991	0.3	-0.769	0.38	1.29	-1.26	-0.0351	0.158	0.057	0.64	-0.474
1.22	0.549	0.974	0.519	1.24	-2.41	-1.38	0.812	0.449	-1.19	-0.425	1.27	0.492
1.22	-2.61	-0.0263	0.416	0.0584	-0.767	0.409	0.595	-0.971	1.51	-1.09	-1.6	-1.72
1.04	0.284	-1.2	0.996	-2.13	0.113	-0.203	-0.272	1.21	0.0367	0.528	0.569	-1.99
0.93	-0.624	-0.0509	1.31	-0.507	-0.544	-1.61	-3.07	-0.827	-0.37	-1.09	0.687	1.53
-0.163	-1.14	-0.276	1.78	-0.155	-0.226	-0.0497	-0.35	0.485	0.127	-0.843	1.49	-0.592
-1.81	-0.716	0.58	-0.344	-1.5	1.52	1.37	-1.87	-0.75	0.843	-0.0955	-0.16	0.181
0.384	0.808	1.81	0.598	-0.751	0.271	-1.35	-0.871	-1.27	0.458	0.466	0.798	0.192
0.255	-2.44	-0.606	-0.878	1.19	0.0961	-0.196	-0.0646	0.253	-1.33	0.77	-0.606	0.67
0.142	0.744	-1.02	0.846	-0.224	0.28	-1.06	-0.278	1.27	0.422	-0.0851	-0.527	-0.828
0.333	0.765	0.00742	0.216	-0.000588	0.658	0.0986	-1.41	-0.381	-0.785	-0.134	-0.0873	-0.61
0.756	0.269	0.532	-1.3	0.247	1.58	0.0185	-0.336	0.781	-0.096	-0.0288	-0.544	0.23
0.236	-0.024	0.787	0.757	-0.306	-0.343	0.0491	-1.58	-1.95	0.972	0.0734	0.474	-0.0138
-0.997	1.18	-1.02	0.00501	2.25	0.468	-2.52	0.0799	0.867	0.302	-0.0656	-2.08	-0.151
-0.355	0.175	0.846	0.245	1.2	-0.101	2.48	0.244	0.757	0.0869	-0.462	-0.02	-0.339
2.36	-1.84	-0.219	-0.613	1.25	1.54	0.655	-1.14	1.62	-2.06	0.353	-2.04	-1.59
0.274	1.02	-0.566	1.42	-0.447	1.79	-1.94	-0.729	0.0059	0.396	0.131	-0.993	-1.16
-0.577	-0.356	-0.869	1.83	-1.47	-0.415	2.06	-0.253	0.0913	-0.303	1.31	0.94	0.51
-0.967	-0.134	-0.0103	0.861	1.76	0.755	0.346	-0.517	-0.333	-1.67	0.272	1.05	-0.74
0.256	-0.157	0.738	-0.591	-1.22	0.801	-1.02	-2.06	0.284	-0.366	1.04	-0.561	1.9
0.701	-1.64	-1.07	-1.45	0.378	0.742	0.0638	0.478	1.16	-1.53	0.141	0.858	0.41
-1.49	-2.52	-1.63	-0.193	0.577	0.516	-0.744	0.488	0.506	-0.292	-0.376	-1.57	0.125
-1.39	0.973	-1.41	0.802	0.507	-1.18	1.17	-1.22	-0.189	1.91	-0.786	-1.62	1.01
-1.28	-0.0243	9.23E-06	-0.532	-0.182	0.646	-0.523	-1.85	-1.96	-0.00857	0.142	0.0853	-0.463

Figure 2: *Snapshot of Dataset*

The Bayesian Belief Network model was designed using Bayes-Server platform. The Bayesian Belief Network (BBN) for predicting Dengue Haemorrhagic Fever (DHF) was designed such that the nodes on the network are linked based on the probability of a disease ailment resulting to another or a factor influences another factor.

In our model for a case to be diagnosed as a Dengue Haemorrhagic Fever case, there will be presence of several or all of the following disease ailments and factors that sways the diagnosis of Dengue Haemorrhagic Fever such as Abdominal Pain, Anopheles Mosquito, Arthropod Genus Mosquito (AGM), Chest Pain, Chills, Continuous Fever, Constipation, Cough, Dengue Haemorrhagic Fever (DHF), DenV1, DenV2, DenV3, DenV4, Diarrhea, Fatigue, Fever, Flaviviridae, Haemoglobin, Headache, Hematocrit, High Temperature, Hot Weather, Humid Weather, Influenza, Intermittent Fever, Joint Pain, Loss of Appetite, Malaria, Metallic Taste, Muscle Pain, Nausea, Neutropenic Fever, Pel-Ebstein Fever, Platelets, Plasmodium Falciparum, Plasmodium Malariae, Plasmodium Knowlesi, Plasmodium Ovale, Plasmodium Vivax, Quartan Fever, Quotidian Fever, Rash, Region, Remittent Fever, RNA Virus, Serotypes, Sore Throat, Sweating, Swollen Abdomen, Sub-Tropical Region, Tertian Fever, Tropical Region, Typhoid Fever, Vomiting, Weakness, Weather, and Weight Loss respectively. Figure 3 shows the BBN model for detecting Dengue Haemorrhagic Fever and its symptoms.
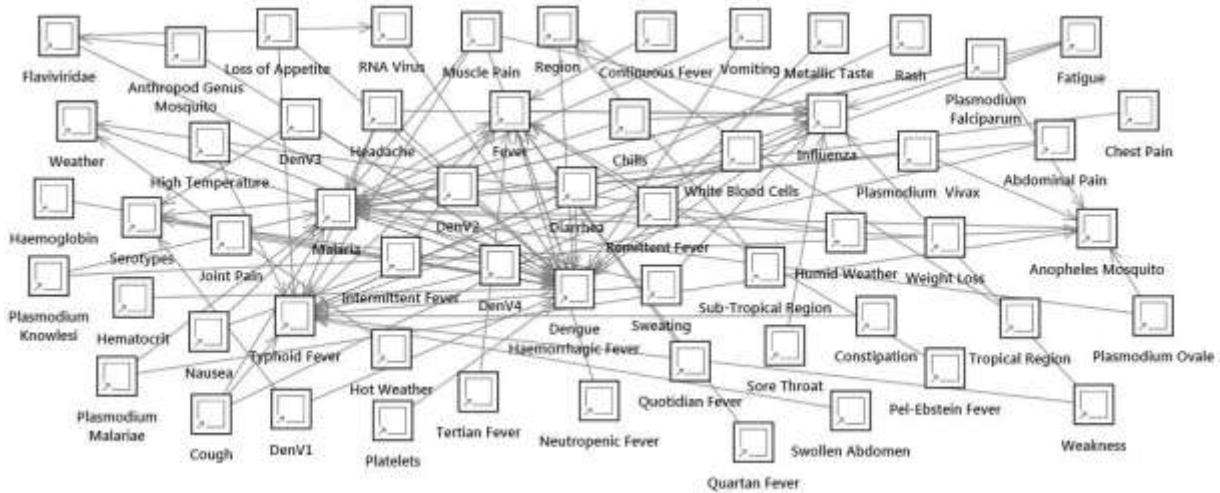

Figure 3: *Bayesian Belief Network Model for Detecting Dengue Haemorrhagic Fever with its symptoms.*

So, to mathematically represent our model we have:

Dengue Haemorrhagic Fever

$$= \prod_{i=1}^{58} P(Disease_i | Parents(Disease_i)) \hspace{4cm} (2)$$

Where,

Disease: Node with a Disease Ailment

Parents (Disease$_i$) = Nodes that converge on Disease 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 Dengue Haemorrhagic Fever and its symptoms at Iteration Count 2.
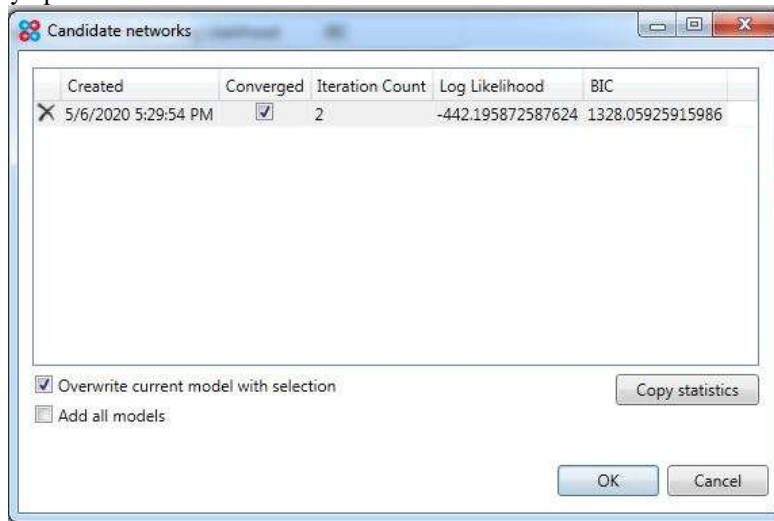

Figure 4: *Bayesian Belief Network Model for Dengue Haemorrhagic Fever Convergence at Time Series 2.*

Figures 5, 6, 7, 8, 9, and 10 shows log likelihood batch query chart for predicting Dengue Haemorrhagic Fever 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 (Flaviviridae) being the reason for the cause of a Dengue Haemorrhagic Fever Infection, association plot showing relation of Predict (Flaviviridae) leading to a Dengue Haemorrhagic Fever infection case and its probabilities and the log likelihood graph for detecting Dengue Haemorrhagic Fever with its symptoms respectively. The result generated from the simulation indicated that the network was able to predict 99% Dengue Haemorrhagic Fever on the dataset accurately and it had a loglikelihood of 64.48 on the test dataset.

The figure 5 below shows the loglikelihood batch query chart for predicting Dengue Haemorrhagic Fever with its symptoms
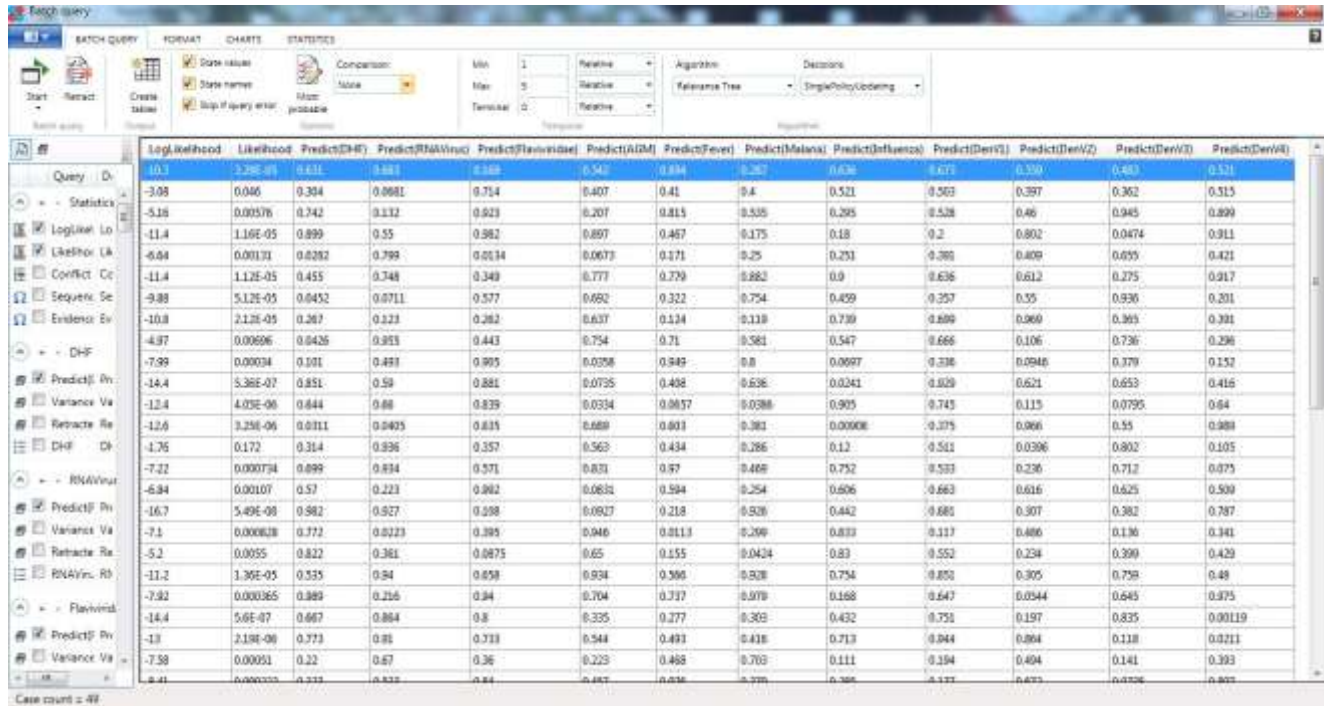
Figure 5: *The Loglikelihood Chart Batch Query for Predicting Dengue Haemorrhagic Fever with Its Symptoms*

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(DHF) was 0.631 compared to 0.63068517474, Predict (RNA Virus) was 0.683 compared to 0.68251045678, Predict(Flaviviridae) was 0.169 compared to 0.1692315303861, Predict(AGM) was 0.542 compared to 0.54150365897675, Predict(Fever) was 0.894 compared to 0.894023521131117, Predict(Malaria) was 0.287 compared to 0. 28761597712371279, Predict(Influenza) was 0.636 compared to 0.63590472768411, Predict(DenV1) was 0.673 compared to 0.673231032271144, Predict(DenV2) was 0.359 compared to 0.359231032261024, Predict(DenV3) was 0.483 compared to 0.483221032271104 and Predict(DenV4) was 0.521 compared to 0.521211032241164 in Experiment 1.

In Experiment 2: The value of Predict(DHF) was 0.304 compared to 0.3038316321, Predict (RNA Virus) was 0.0681 compared to 0.06805103245, Predict(Flaviviridae) was 0.714 compared to 0.7135315353762, Predict(AGM) was 0.407 compared to 0.40690365793234, Predict(Fever) was 0.41 compared to 0.4053521131117, Predict(Malaria) was 0.4 compared to 0.40012154312141272, Predict(Influenza) was 0.521 compared to 0.52050432761410, Predict(DenV1) was 0.503 compared to 0.50262231032241121, Predict(DenV2) was 0.397 compared to 0.3970231030061012, Predict(DenV3) was 0.362 compared to 0.361611032230303 and Predict(DenV4) was 0.515 compared to 0.515101031021003 in Experiment 2.

In Experiment 3: The value of Predict(DHF) was 0.742 compared to 0.7416103300, Predict (RNA Virus) was 0.132 compared to 0.1317002123, Predict(Flaviviridae) was 0.923 compared to 0.9228434321007, Predict(AGM) was 0.207 compared to 0.20691364743143, Predict(Fever) was 0.815 compared to 0.8147521210563, Predict(Malaria) was 0.535 compared to 0.53462154210012345, Predict(Influenza) was 0.295 compared to 0.29470432721239, Predict(DenV1) was 0.528 compared to 0.5279231036241023, Predict(DenV2) was 0.46 compared to 0.46015031040161000, Predict(DenV3) was 0.945 compared to 0.944760322011128 and Predict(DenV4) was 0.899 compared to 0.899111500211104 in Experiment 3.

Furthermore, this experiment continues up to Experiment number 50. Hence, the system results showed a 0.0159% 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
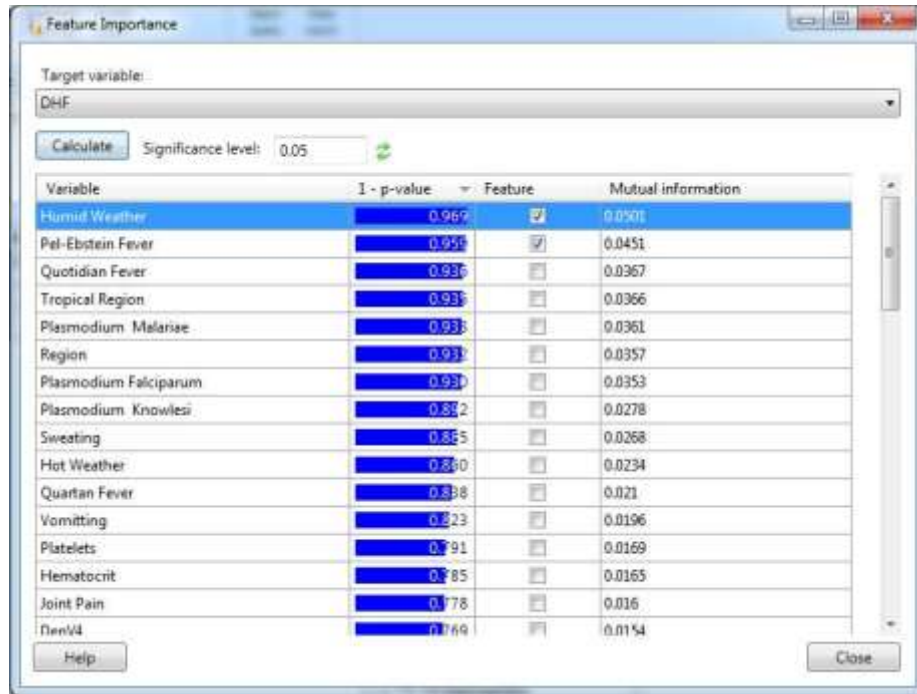
Figure 6: *The Feature Importance Chart for Nodes in the BBN Model*

The Feature Importance Chart shows p-value of the variable (nodes), Feature and Mutual information in reference to the Dengue Haemorrhagic Fever (DHF) Node.

The p-value signifies the likelihood (probability) of the nodes being the cause of a Dengue Haemorrhagic Fever infection.

The Feature box is checked if that particular node is fully involved in the cause of a Dengue Haemorrhagic Fever 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 Dengue Haemorrhagic Fever node) and assigned a value.

The Significance Level signifies the margin of error in the detection of Dengue Haemorrhagic Fever and its symptoms.

The figure 7 below shows the in-sample anomaly detection chart for the Bayesian Belief Network Model.
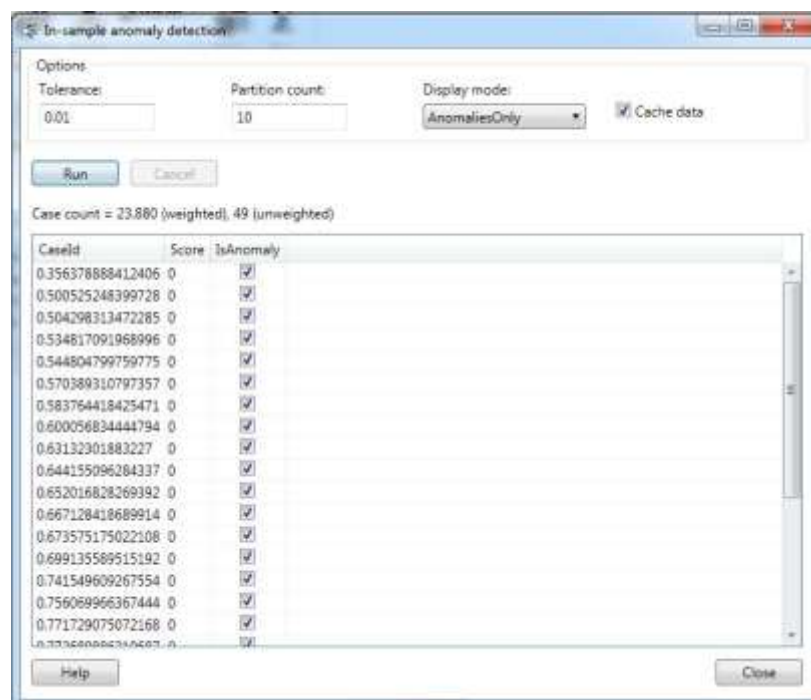
Figure 7: *The In-Sample Anomaly Detection Chart Chart for Nodes in the BBN Model*

The In-sample Anomaly Detection Chart shows 50 experimental results of detecting Dengue Haemorrhagic Fever. Each Case is assigned an ID(Identification value) which is the value of the Predict(Dengue Haemorrhagic Fever) in Figure 5 above. The IsAnomaly checkbox is checked to identify that each case is a confirmed case of Dengue Haemorrhagic Fever infection. The 50 cases of Dengue Haemorrhagic Fever has a case count value of 23.880 (weighted) which signifies the importance of the cases leading to a Dengue Haemorrhagic Fever infection and 49 case count value signifies the number of cases in the pool of data available to the system for detection of Dengue Haemorrhagic Fever excluding the Dengue Haemorrhagic Fever column in the dataset pool. The tolerance is the margin of error that could be encoutered as regards to the detection of the Dengue Haemorrhagic Fever and its symptoms.

The figure 8 below shows the mesh query plot for the loglikelihood of a disease causing ailment (Flaviviridae) being the reason for a Dengue Haemorrhagic Fever Infection.
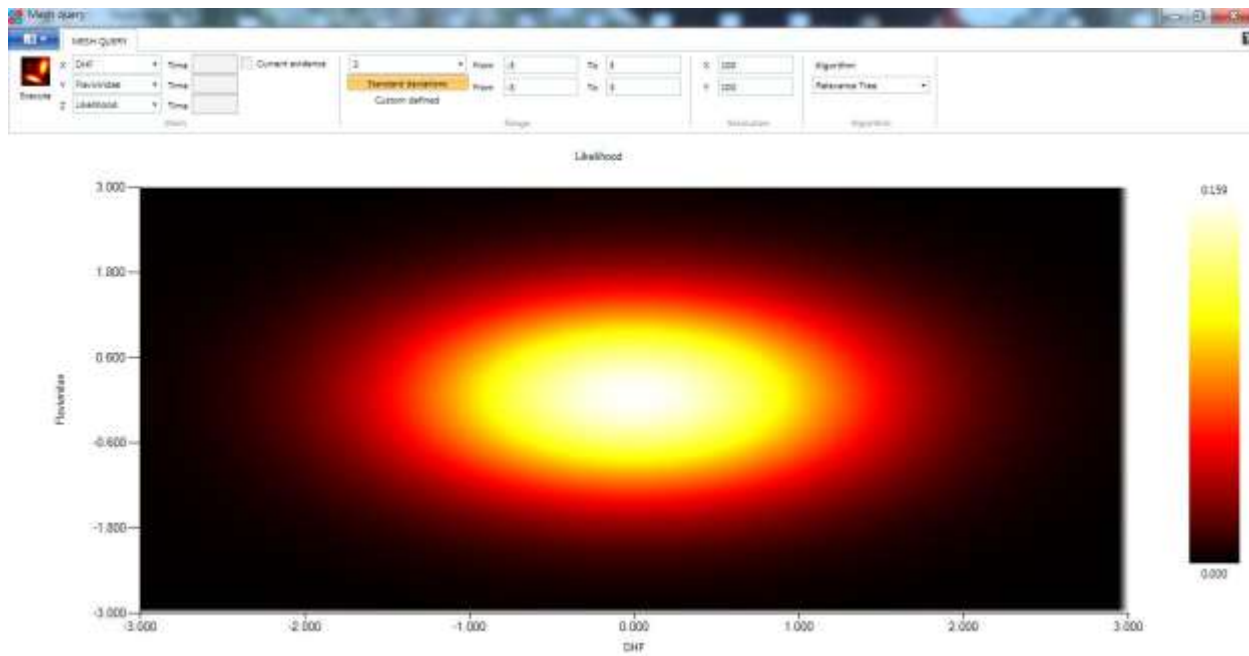

Figure 8: *The Mesh Query Plot for The Loglikelihood of A Disease Causing Ailment (Flaviviridae) Being The Reason For A Dengue Haemorrhagic Fever (DHF) Infection.*

The mesh query plot shows the loglikelihood/likelihood of a node in this case Flaviviridae being a cause of a confirmed Dengue Haemorrhagic Fever infection. The Node (Flaviviridae) is plotted on the Y-axis and the other node Dengue Haemorrhagic Fever (DHF) plotted along the X-axis.
In this context, the Red contour signifies the likelihood of a Flaviviridae being a major of cause of Dengue Haemorrhagic Fever 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 Flaviviridae being the reason of a Dengue Haemorrhagic Fever infection with the contour ranging from interval (-0.620 to 0.580) on the Y-axis and interval (-1.110 to 0.995) on the X-axis.

The figure 9 below shows the association plot showing relation of Predict (Flaviviridae) leading to a Dengue Haemorrhagic Fever infection case and its probabilities.
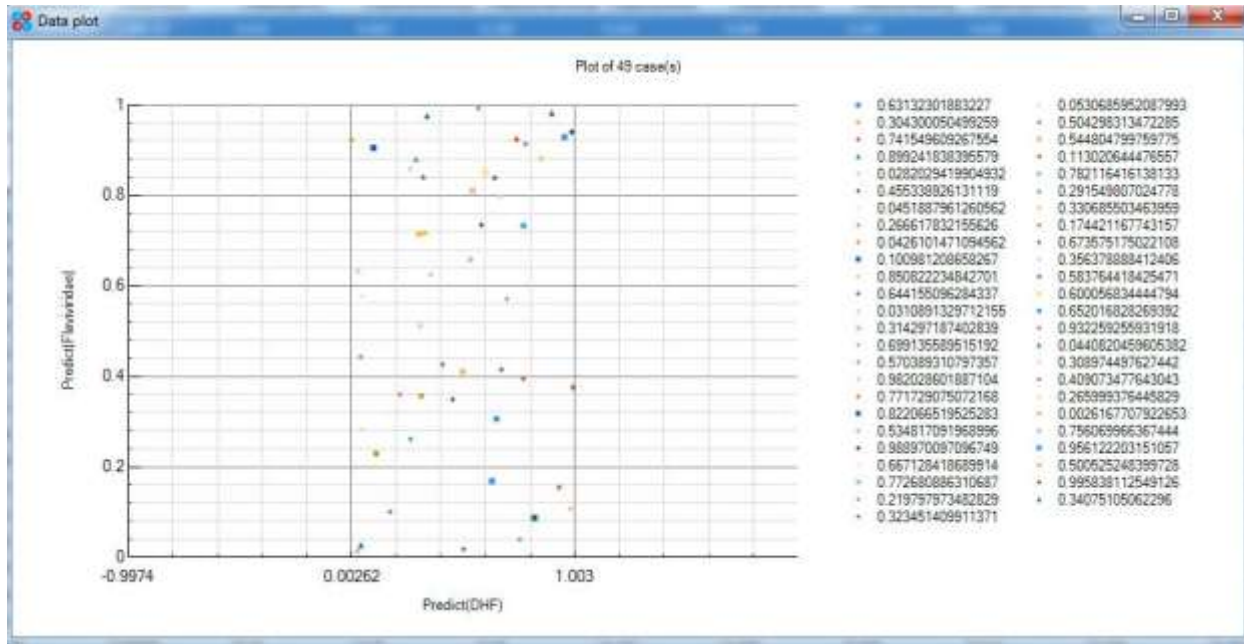
Figure 9: *The Association Plot Showing Relation of Predict (Flaviviridae) Leading to A Dengue Haemorrhagic Fever Infection Case And Its Probabilities In The BBN Model*

The association plot shows the relation of how contact with Flaviviridae Virus leads a Dengue Haemorrhagic Fever (DHF) 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 Flaviviridae on the Y-axis is plotted against DHF on the X-axis.

However, from this graph, there are five diagnostic classes of Dengue Haemorrhagic Fever 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 0.00262 to 1.003 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 Dengue Haemorrhagic Fever 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 0.00262 to 1.003 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 Dengue Haemorrhagic Fever infection with the severity level categorized as being Mild.

Moderate Class: This class ranges from 0.4 to 0.6 on Y-axis and 0.00262 to 1.003 on X-axis. This region has 8 colored points (cases).This signifies that the 8 colored points in this region represent 8 cases of patients with Dengue Haemorrhagic Fever infection with the severity level categorized as being Moderate.

Severe Case: This class ranges from 0.6 to 0.8 on Y-axis and 0.00262 to 1.003 on X-axis. This region has 8 colored points (cases). This signifies that the 8 colored points in this region represent 8 cases of patients with Dengue Haemorrhagic Fever infection with the severity level categorized as being Severe.

Critical Class: This level ranges from 0.8 to 1 on Y-axis and 0.00262 to 1.003 on X-axis. This region has 16 colored points (cases). This signifies that the 16 colored points in this region represent 16 cases of patients with Dengue Haemorrhagic Fever infection with the severity level categorized as being Critical.

Of the 50 experimental cases, the system predicted 49 cases ranging from asymptomatic, mild, moderate, severe, and critical classes correctly from the test data with 99% sensitivity accuracy.

All the 49 cases in figure 9 had a probability value less than 1; with the highest probability value of Flaviviridae causing a Dengue Haemorrhagic Fever Infection reported to be 0.995838112549126 which is less than 1.

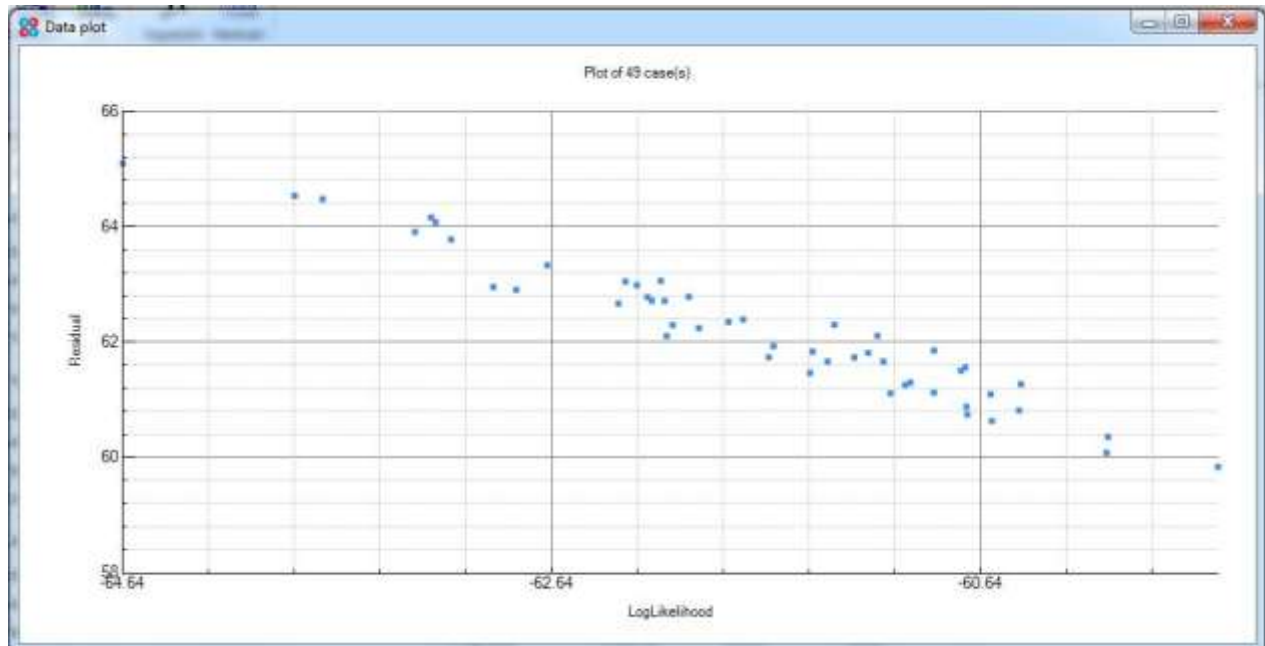The figure 10 shows the Loglikelihood Graph for Detecting Dengue Haemorrhagic Fever with its Symptoms.

Figure 10: *The Loglikelihood Graph for Detecting Dengue Haemorrhagic Fever with its Symptoms*

This loglikelihood graph for detecting Dengue Haemorrhagic Fever 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 64.48, 64.35, 64.28, 64.15, 64.05, 64.…….. and 58.97 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 -64.64 on the horizontal line with 66 being the highest value that can be reached on the vertical line.

The residual value attained is 64.48 and loglikelihood independent value is -64.64, the difference between both values is 0.16 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 66. With 66, being the 100 % residual value mark, to get our prediction accuracy percentage, we have predicted value subtracted from highest residual value i.e. 100% -0.16= 99.84% residual loglikelihood percentage.

Furthermore, the detection results in figure 9 showed all classes of severity status of Dengue Haemorrhagic Fever Infection cases with their probabilities while figure 10 showed 99.84% prediction accuracy of the system.

Thus, the probability of having Dengue Haemorrhagic Fever infection given there is evidence of ailments and factors that influence diagnosis of the aforesaid disease is denoted as:

P(Dengue Haemorrhagic Fever| Abdominal Pain, Arthropod Genus Mosquito, Chest Pain, Chills, Continuous Fever, Constipation, Cough, DenV1, DenV2, DenV3, DenV4, Diarrhea, Fatigue, Fever, Flaviviridae, Headache, Hematocrit, High Temperature, Hot Weather, Humid Weather, Influenza, Intermittent Fever, Joint Pain, Loss of Appetite, Malaria, Metallic Taste, Muscle Pain, Nausea, Neutropenic Fever, Pel-Ebstein Fever, Plasmodium Falciparum, Plasmodium Malariae, Plasmodium Knowlesi, Plasmodium Ovale, Plasmodium Vivax, Quartan Fever, Quotidian Fever, Rash, Remittent Fever, Sore Throat, Sweating, Swollen Abdomen, Sub-Tropical Region, Tertian Fever, Tropical Region, Typhoid Fever, Vomiting, Weakness and Weight Loss) = 0.995838112549126.

From the experiment, it can be seen that our model has a higher residual log likelihood value which is 64.48 and a prediction accuracy of 99.84%.

Finally, comparing the 99.84% prediction accuracy of our model with the experiments conducted by [10, 11, 14, 15 and 16] which has 91.3%, 92%,95%,85.18% and 84% 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 other mosquito-borne diseases have in common with Dengue Haemorrhagic Fever, hence aiding the high detection accuracy of the aforesaid disease.

5. Conclusion

Dengue Haemorrhagic Fever is a contagious disease that is quite tough to detect due to the overlapping symptoms the disease shares with other mosquito-borne diseases. To diagnose this disease, medical professionals need to advance on the existing methods employed in detecting Dengue Haemorrhagic Fever to curb untimely deaths of patients due to lack of early diagnosis of the said disease.

In this paper, we utilized a supervised machine learning technique called Bayesian Belief Network model to predict Dengue Haemorrhagic Fever and its symptoms. The network had 58 nodes with each node representing an exclusive ailment. The model was trained and tested and it had an accuracy of 99.84% in predicting Dengue Haemorrhagic Fever with its symptoms. The system can be deployed in health facilities to help provide information which will be used to detect Dengue Haemorrhagic Fever and its symptoms. It will also bring about improvement in the following areas: Dengue Haemorrhagic Fever prediction, Dengue Haemorrhagic Fever detection and Diagnosis of mosquito-borne diseases with similar symptoms as Dengue Haemorrhagic Fever.

REFERENCES

[1] Krista O'Connell (2016): "Causes, Symptoms and Diagnosis?". Retrieved from URL:
https://www.healthline.com/symptom/fever
[2] Garmel GM, Mahadevan S.V. (2012): "An Introduction to Clinical Emergency Medicine" (2nd ed.). Cambridge: Cambridge University Press. ISBN 9780521747769. pp 5.
[3] 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/
[4] Gould, E.A and Solomon, T. (2008): "Pathogenic Flaviviruses". Lancet. 371 (9611): pp 500-509. doi:10.1016/S0140-6736(08)60238-X .PMID 18262042.
[5] World Health Organization (2020): "WHO Report on Global Surveillance of Epidemic-Prone Infectious Diseases-Dengue and Dengue Haemorrhagic Fever". Retrieved from URL:
https://www.who.int/csr/resources/publications/dengue/CSR_ISR_2000_1/en/
[6] Chandra, D., Shaziya I. and Pandey, D. (2019): "Prediction of Dengue with the Use of AI and Data Mining: An Expert System". International Journal of Computer Engineering in Research Trends, Volume-6, Issue-7, 2019 Regular Edition E-ISSN: 2349-7084 pp 340-346
[7] Aysha I. Mansour, Samy S. and Abu-Naser (2019): "Knowledge Based System for the Diagnosis of Dengue Disease". International Journal of Academic Health and Medical Research (IJAHMR), ISSN: 2000-007X, Vol. 3 Issue 4, April – 2019, pp 12-19.
[8] D. Saikia and J. C. Dutta (2016): "Early Diagnosis of Dengue Disease using Fuzzy Inference System", 2016 International Conference on Microelectronics, Computing and Communications (MicroCom), Durgapur, 2016, DOI: 10.1109/MicroCom.2016.7522513, pp. 1-6.
[9] Amjad, M., Bukhari, F., Ameer, I. and Gelbukh, A. (2018):"A Distinct Approach To Diagnose Dengue Fever with the Help of Soft Set Theory". DOI:10.26782/jmcms.2019.10.00032, Corpus ID: 43925305. pp 1-19.
[10] Sharma, P., Singh, D.B.V., Bandil, M.K. and Mishra, N. (2013): "Decision Support System for Malaria and Dengue Disease Diagnosis (DSSMD)". International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 3, Number 7 (2013), pp. 633-640
[11] Shaukat, K., Masood, N., Mehreen, S. and Azmeen, U. (2015):"Dengue Fever Prediction: A Data Mining Problem". J Data Mining Genomics Proteomics 6:181. doi:10.4172/2153-0602.1000181. pp 1-5.
[12] Husin, N.A., Alharogi, A., Mustapha, N., Hamdan, H. and Husin, U.A. (2018): "Early Self-Diagnosis of Dengue Symptoms Using Fuzzy and Data Mining Approach". Proceedings of the 3rd International Conference on Applied Science and Technology (ICAST'18), AIP Conf. Proc. 2016, https://doi.org/10.1063/1.5055450, pp 020048-1–020048-7.
[13] Fauzi, Nungsiyati, Noviarti, T., Muslihudin, M., Irviani, R. and Maseleno, A. (2018): "Optimal Dengue Endemic Region Prediction using Fuzzy Simple Additive Weighting based Algorithm". International Journal of Pure and Applied Mathematics, Volume 118 No. 7 2018, ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version), Retrieved from URL: http://www.ijpam.eu, Special Issue, ijpam.eu, pp 473-478.
[14] Balasaravanan, K. and Prakash, M. (2018): "Detection of Dengue Disease Using Artificial Neural Network Based Classification Technique". International Journal of Engineering & Technology, 7 (1.3) (2018), pp 13-15.

[15] Kumar, K.N.,Vigneswari D. and Devi, B.A.S.R (2018): "MSO – MLP Diagnostic Approach for Detecting DENV Serotypes". International Journal of Pure and Applied Mathematics, Vol.118, No.5, 2018, pp 1-6.

[16] Kumar, M.N. (2013): "Alternating Decision Trees for Early Diagnosis of Dengue Fever".arXiv:1305.7331v2 [cs.LG] 5 Jun 2013. pp 1-13

[17] 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.

[18] 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/. pp1-3

[19] Datadryad(2020): "Dengue Hemorrhagic Fever(DHF) Dataset" .Retrieved 3rd May, 2020 from URL: http://datadryad.org/review?doi=doi:10.5061/dryad.078bn.