Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021)
IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

A Study on Dengue Fever in Bangladesh: Predicting the Probability of Dengue Infection with External Behavior with Machine Learning

Md. Sanzidul Islam
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
sanzid.cse@diu.edu.bd

Sharun Akter Khushbu
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
sharun.cse@diu.edu.bd

Akm Shahariar Azad Rabby
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
shahariar.cse@diu.edu.bd

Dr. Touhid Bhuiyan
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
headcse@daffodilvarsity.edu.bd

2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) | 978-1-6654-1272-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICCS51141.2021.9432288

***Abstract* — The "2019 Dengue Outbreak" was a nationwide pandemic situation in Bangladesh, particularly in Dhaka city. About 179 people died and 101,354 confirmed dengue cases were found all over the country. The developing countries like Bangladesh have some limitations in the medical sector and many people don't get proper treatment in time. Henceforth, this research work has attempted to predict the chances to get infected with dengue fever from some external behaviors, like- fever, pain, sitophobia, headache etc. This article has demonstrated a model to predict the probability of dengue fever before taking the pathological test. So, the suspective patient may get some initial diagnosis by giving their anatomical symptoms as input and further this will decrease the dependency on the pathological test for acquiring the primary treatment. Different machine learning models are used to predict the probability and an accuracy near to 100% has been achieved finally.**

Keywords — Machine Learning model, Dengue disease, anatomical symptoms.

I. INTRODUCTION

Dengue fever (DF) and dengue hemorrhagic fever(DHF) a re-transforming mosquito-borne disease, acute feverish sickness [1]. Dengue's more severe form comes when it is transmitted by the bite of Aedes mosquitoes in any one of the serotypes [2]. This public endemic spread across the tropical and subtropical countries. In the 19th century in the era of 1950s, the Philippines has discovered and renowned it as a dengue fever [3]. During the long 50 years, dengue has been intimidating as a global threat in the South-East Asia region, which is caused by 30 fold regained forces [4]. Dengue fever affects around 128 countries, which is half of the population, where children mostly face endangerment [3]. In a single year 390 million people stuck with fever and 96 million people experience hospitalized misery with dengue fever. In 1964 Bangladesh was first attempted in dengue fever detection as called 'Dacca fever' [5] later this endemic discovery caused a place of horrible situation in Dhaka. In Dhaka the first official outbreak through the disease in 2000 by around 6000 cases and below 100 deaths [6-7] after that in 2008 to 2011 this disease reformed as the another form [8-9] still now people have been suffering most in Dhaka, Bangladesh [10]. Current situation in 2020 DF also highly attacks as many as 3 million people in south asia region, but with determination by making thought of earlier precaution to reduce the number of disease [11]. Therefore, economic defeat has come for the country economically, family impacted by money issues, treatment cost, physically absent and so on have to suffer unless they maintain health in early stages [12-13]. Due to this unsustainable stage in Bangladesh should take few precautions about knowing dengue disease before the test. Accordingly, prevention and control is only possible by generating a forecasting surveillance system on behalf of some external abnormal issues. Dengue virus illness includes the occurrence of some non-specific symptoms in a very first stage and continuously it gets started to remain as high form. Globally, no vaccine has been saturated against dengue for this emergence of dengue people. By the illness of dengue initially the victim of illustration issues as high temperature (up to 106 degrees Fahrenheit), rehydration, low plasma rate, severe headache, swollen lymph glands, severe muscle and joint pains, rashes, vomiting, mild bleeding from the nose or gums, mild bruising on the skin and irritable mood swing.

To the best of our knowledge, it has been attempted to focus on the available 400 patient data of dengue collected from Dhaka city in recent data. Among all 70% data of Dengue positive with symptoms and about 30% of dengue negative patients with several symptoms. Receiving 10 to 12 external behaviour of each patient and will try to predict the dengue positive or negative according to their external behaviour. The proposed system mentions public health in Bangladesh because in recent years our approaches are quite different from the others. Prediction easily defeats the unpredictable hospitalize situation from severe and non severe dengue. In the rural area it makes it more convenient in early diagnosis.

II. LITERATURE REVIEW

Machine learning reflection has a wide response in any field. Among them is the biomedical field also covered with ML analysis. Dengue Fever prediction several approaches done by the research field shown as an ML prediction based on some attributable behavioural responses. Several countries during the situation they made analysis on ML in dengue cases in recent years. Before starting the processing

there must be needed the understanding of the scope behind the analysis as well as refill the research section with a new perspective in Bangladesh.

Hani M. Aburas et al. [14] assumed by the analysis over dengue prediction on confirmed cases through ANN model estimation. In Malaysia, there has been a huge size of affected area by the disease since January,2001 to 2007. Concerned with dengue cases from NEA which is public domain from Malaysia nearby 14209 cases were attempted towards ANN with high correlation coefficient rate is 0.96. Neural networks require input sequence training and a test dataset covered by the 70% and 30%. Hidden neurons exist by the 4 features like temperature, humidity etc. produce correlation coefficient and RMSE error.

James A. Potts [15] explored a biomedical research dengue disease which required machine learning classification. Over the 100 countries being affected by the dengue thus Thailand has been suffered by the dengue throughout this concept taking input as two hospitals statistical data who are infected in 72 hours. Machine learning classification is Logistic Regression applied with 5 features of dengue referred to as platelets, white blood cells differentials and more. 1230 patients are conducted to analyze suspected illness pre notice in 72 hours or 3 days. Logistic Regression classified accurately 100% for febrile illnesses.

Nazmul Karim et al. [16] introduced the circumstances of Bangladesh in Dhaka city regarding 2000 to 2008 long 8 years of statistical data proclaimed by the machine learning algorithm is linear regression. Dengue disease likely features by means of model validation claims climate change have a great effect in Dhaka causing an increased dengue.

Benjamin M. Althouse et al. [17] exploited linear regression and two classification of Logistic regression and SVM of forecasted over exceeding dengue incidence in Bankok and Singapore region. Dengue dataset outlook comes to the internet data source for forecasting a surveillance system. Dengue incidents predicted by the binary classification almost took 5 years of information which generated 95% level of accuracy. Due to classification analysis both LR and SVR compared between r2 value and correlation fitted value. Nirosha Sumanasinghe et al. [18] reflected by the gist analysis of SVR machine learning algorithm covered with an efficient prediction on three vectors which is rainfall, temperature and population density in Thailand above four regions. Due to public awareness among five years the dengue scenario had represented whether it is fever or not. Added a turning point was 71% accuracy from the SVR machine learning algorithm on behalf of correlation coefficient scores and MSE scores. During Dengue prediction on Thailand in 4 provinces shows correlation coefficient rate where temperature and population rate was detached with dengue rather than rainfall may cause dengue for further checking have used 10 fold cross validation.

Arul Earnest et al. [19] forecasted on Singapore DF and DHF situation with further notification by the analysis of ARIMA model and K-H model. Dengue fever infected dataset was 2001 to 2006 which means long 6 years based on dataset claimed prediction on 2007 to june of 2008 decrease positive cases. Though illustration of the K-H model and ARIMA expose a dilemma between models which outcome is best. After that ARIMA time series model flexible but unsuccessful to define a surveillance system the reason behind is MAPE scores is 19.96. Knorr-Held(K-H) model average moving was great with a sensitive analysis whereas MAPE value is 17.21. Below the section showed a comparative DF prediction on a weekly basis.

Howell T. Ho et al. [20] demonstrated dengue incidence correctly with 6 behavioural features in Metropolitan Manila of Philippines. Purpose of research showed a graphical representation defined correlation coefficient on 3 several studies is Google Dengue Trend(GDT), Dengue Incidence(DI) last one is dengue related search query was taken by per year of 2009 to 2014 long 6 years related information. National Epidemiology Center delivered the behavioural features for weekly update besides GDT with related search query prediction defined the future increase or decrease dengue patterns.

III. PROPOSED METHODOLOGY

The section of methodology such a combination of theoretical and statistical analysis among the algorithms or model specification. Therefore, in this section illustrated by the working approach from start with data process to machine learning model output also briefing with how convenient ML classifier is. The proposed "Dengue positive and negative prediction" has been necessary steps are mentioned below.

*A. **Gist of Research***

Prediction of research topics is necessary for better understanding, future development and introducing models. A gist of research shows transferation process of data preprocessing by the attributable response after that run into the model at the end of research deliberate introduction of required instruments being with its uses. Here dengue fever attributes belonging to external behaviour after preprocessing we evaluated its DF occurrences of prediction by several ML classification experiments.

*B. **Data Collection***

Due to the research of Dengue Fever future probability prediction needs relevant data of dengue disease. Hence, dengue disease prediction is quite challenging against the external symptoms. Thereby the reason for there has been necessary to make a dataset where both types of patients include which have positive in dengue and which have not but symptoms likely same. To keep it sense, we have prepared a dataset on 400 cases where 70% data consist of the positive dengue cases on the other hand 30% data for negative cases in Bangladesh respectively. According to the preparation of the dataset both positive and negative cases stand for prediction with a good knowledge what are the dengue symptoms or not next classify into them for dengue prediction. Added with around 12 external behaviours chosen for attributes, which are clinically approved as dengue-like symptoms, which enrich further ensurity and similarity. The number of cases chosen by the different ages from child to adult or older. The collected features convert into numerical value by the label of encoding where it designates the label with positive and negative as true and false respectively [21].

Sl.	Features Name	Data Type	Limit
1	High Temperature(Fever)	Numeric	<160° F
2	Rehydration	Binary	N/Y
3	Low Plasma Rate	Numeric	1.5-4.5(Lakh/ mm^3)
4	Headache	Binary	N/Y
5	Swolen lymph glands	Binary	N/Y
6	Rashes	Binary	N/Y
7	Vomiting	Binary	N/Y
8	Bleeding Symptoms	Binary	N/Y
9	Irritable Mood Swing	Binary	N/Y
10	Body ache	Binary	N/Y
11	WBC	Numeric	4000-11000 (/cumm)
12	Haemoglobin	Numeric	12.0-17.6 (g/dL)
Predict Dengue		Binary	Positive/ Negative

Table_1: Clinical Parameter of Dengue Disease

C. *Data Preprocessing & Model*

A numerical representation needs to preprocess by raw data collecting from the dengue health cases. According to the data collection we have 30% of data for negative dengue cases and more than 70% cases positive for training sets. Therefore, our overall dataset has a lack of value which comes from the test set. At the beginning of data analysis there are some necessary steps, the first one is to preprocess the data. Trained data is accurately applicable for running into the algorithm later the value of the test set after filtering with splitting data and encoding data transfer into new clean data. The cases of dataset for DF prediction now apparel accurately to classify the dengue cases and classification of output more existing.

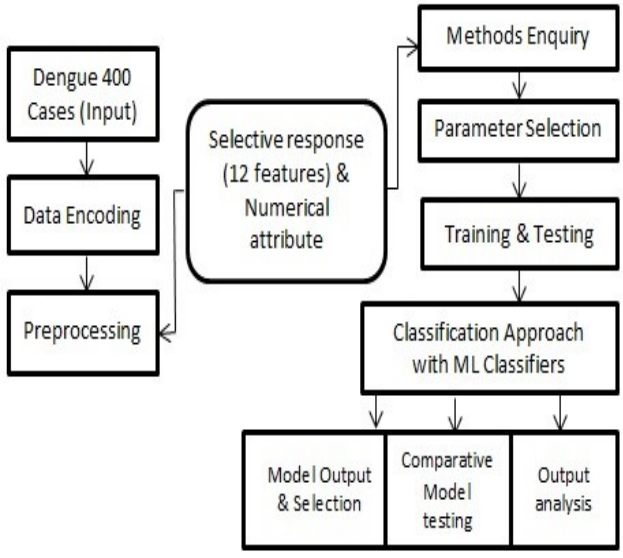

Figure_1: Model of dengue cases for prediction

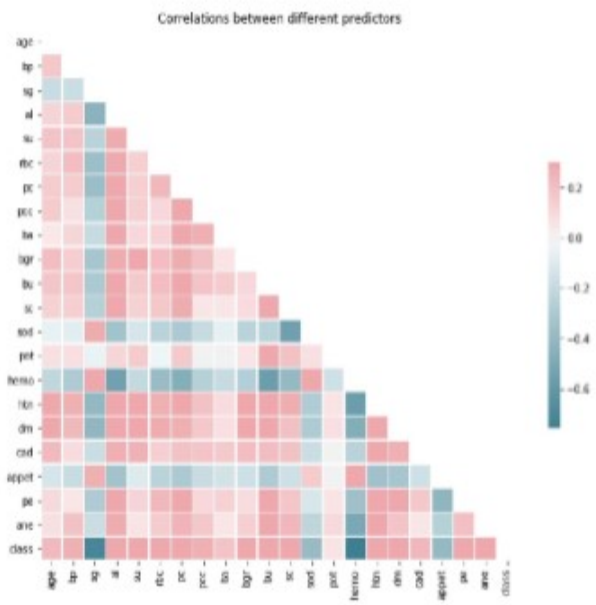

Figure_2: Correlation among the ML classifier

The Dengue cases preprocess under the label encoding with 12 features segment a one statement either true or false by the prediction when the next case matches similarity or not. Within the method enquiry whole numerical statistics of CSV depart into the training and testing section. Then the ML classification procedure surely executes through the dengue prediction over the dengue dataset and classified the accuracy model with a best comparison. Based on the output accuracy we develop an optimum model measured by the crucial parameter such as confusion matrix, Performance, Precision, and other related scores. As we use different classifiers to evaluate the performance of Dengue prediction the comparative performance with accuracy parameters are shown in the next section. Applying with Multivariate ML classification both supervised and unsupervised learning a healthy well known entresol. Supported with background study of dengue till data preprocess to prepare data a feeding form then gradually applied with KNN [22], SVM [23], RF [24], DT [25], NB [26], LR [27], ANN [28]. According to this ML classifiers highly recommended for any disease prediction because of previous study highly enriched into these classifiers.

IV. Experiment and Discussion

A. *Model Training*

ML classifiers are remaining as a specific excellence because of the descriptive analysis performed for obtaining the experimental results. Since the data is divided into two parts on train and test set, the dengue cases analysis goes among 400 cases with 12 features, which are covered by the several ML classifications. During the process of dengue dataset by the use of library in python to get value whereas classification is explored by the number of each sequence where differentiate one parameter to another parameter that

shows the actual predictor and highest output generator. Based on the 400 cases covered with a 12 dengue external behaviour as feature selection we prepare our model using the ratio of 5:1. Dataset contained train set and test set of four portions to learn the system of actual prediction. The cases of dengue collected from Bangladesh public health association have two categories both positive and negative cases. Thus ML algorithms abstract with input form and generate one output form for each classification. Similarly training data and testing data predict high achievement with 400 cases where 80% training data around 320 data have both positive and negative cases. In order to predict DF another test set consists of 80 data for both. 7 classifiers classify with two terminals when the case has Dengue it classifies positive classification in case no dengue then it will go into the negative classification. In here RF, DT, LR is true for positive dengue in specificity and sensitivity percentage.

B. Experimental Results between different ML classifiers

Over the dengue disease prediction whether it can be positive or negative showed a tremendous percentage in all ML classifiers by evaluating the training and test set cases. On behalf of the result performance measures some value as Accuracy, Precision, Sensitivity, Specificity, Confusion matrix and Loss of ML classifiers.

In table_3 where DT and Naive Bayes both reflect with the highest priority model on high accuracy 100% where the dengue prediction model should effectively answer if NB and DT model will assist. NB and DT figure the same value of confusion matrix such as (28 00) (00, 52) for train and test set. Specificity and sensitivity score is actually 1 where correlation coefficient value is 1 for both cases. After the prediction means DT and NB will be the best classifier to detect dengue for human external behaviour.

Another two classifiers in table_3 where SVM and Logistic Regression both reflect with the third highest priority model on high accuracy 97.50% where the dengue prediction model should effectively answer if SVM and LR model will assist. NB and DT figure the same value of confusion matrix such as (27 01) (01, 51) for X and Y train and X and Y test. Specificity and sensitivity score is actually 0.96 and 0.98 where correlation coefficient value is 1 for both cases. After the prediction SVM and LR properly classify accurately near to 98% to detect dengue for human external behaviours.

RF classification determines the value how accurately to predict the cases is 98.75%. Where confusion matrix expressed such as (28 01) (01 51) and specificity and sensitivity score is 1 and 0.9655 that means it accurately 99% classify the positive or non-positive.

At last two ML classifiers predict little bit wrong way thus ANN and KNN predict incorrectly on dengue 400 cases where accuracy is below 80% and the accuracy is 65% and 71.25% confidence level is low with specificity and sensitivity score is 0.85 and 0.56 for KNN besides ANN sensitivity and specificity score is 0 and 0.65. Correlation coefficient value 0 for ANN and 0.43 for KNN where confusion matrix fitted with (22 06) (17 35) and for ANN fitted with (00 28) (00 52).

Classifiers	Prediction of Models
KNN	71%
SVM	98%
Naive Bayes	100%
Logistic Regression	98%
Random Forest	99%
ANN	65%
Decision Tree	100%

Table_2: Model Prediction in different Classifiers On Dengue Disease

V. CONCLUSION

The study of dengue disease now outbreak in bangladesh where affected scenario large in a day by day. In urban areas intimately adopt with dengue. All ages start from children to younger or older people being infected daily. Public hospital care attempts daily a bundle of positive dengue cases who come for their treatment. Many cases recover besides many other cases not much long run for late come into the hospital to test dengue positive or negative. Due to this endemic situation we try to develop a classification system based approach where people who get sick their cases identify whether dengue is positive or negative by the ensured 12 symptoms. Our CSV dataset hold by the 400 dengue cases in Bangladesh from the public domain, most cases are positive and few cases are negative. After cleaning the dataset ML classifiers applied on those cases and perform accurately with prominent accuracy. NB, DT and RF classifiers performance is very lucrative such as 100% and 99% respectively. Rather, SVM and LR also give impressive output by the accuracy of 98%. According to the analysis using this several ML classification predictions in here only two models predict inaccurately which is ANN and KNN regarded as 65% and 71% accuracy. In a sense, Dengue disease prediction for positive cases can accurately predict the results with five ML models such as NB, DT, RF, SVM, LR.

VI. ACKNOWLEDGMENT

DIU_NLP & Machine learning Research Lab deliberately support to fulfil our approach. We are grateful to give us support together with facilities and guidance besides we are thankful to our Department of Computer Science & Engineering.

REFERENCES

[1] Morens DM, Fauci AS, “Dengue and hemorrhagic fever: a potential threat to public health in the United States”, JAMA 299: 214–216, 2018

[2] Halstead SB, "Pathogenesis of dengue: challenges to molecular biology", Science 239: 476–481, 1988.
[3] Rosario Z.Capeding, Imelda A.Luna, Emilly Bomasang, Sacorro Lupisan, Jean Lang, Remi Forrat, Anh Wartel, Denis Crevat, "Live-attenuated, tetravalent dengue vaccine in children, adolescents and adults in a dengue endemic country: Randomized controlled phase I trial in the Philippines", volume 29, issue 22, pages 3863-3872, 2011.
[4] World Health Organization, "Comprehensive guidelines for prevention and control of dengue and dengue haemorrhagic fever, 2011.
[5] Russell PK, Buescher EL, McCown JM, Ordonez J, "Recovery of dengue viruses from patients during epidemics in Puerto Rico and East Pakistan", 573–579, 1966.
[6] Shahera Banu, Wenbiao Hu, Yuming Guo, Cameron Hurst, Shilu Tong, "Projecting the impact of climate change on dengue transmission in Dhaka, Bangladesh", pp. 137-142, 2014.
[7] Shifat Sharmin, Kathryn Glass, Elvina Viennet, David Harley, "Interaction of Mean Temperature and Daily Fluctuation Influences Dengue Incidence in Dhaka , Bangladesh", 2015.
[8] Rashedul Hasan, Md Mujibur Rahman, Md Moniruzzaman, Abdur Rahim, Satyajit Barua, Rajib Biswas, Pijous Biswas, Syed Ghulam Mogni Mowla, MA Jalil Chowdhury, " Chikungunya-An Emerging Infection in Bangladesh: A case Series", 2014.
[9] Faijul Islam Chowdhury, Ahmedul Kabir, Aparna Das, Shegufta Mishket Mukerrama, Shahin Masud, "Chikungunya Fever: An Emerging Threat to Bangladesh", Volume 13, pp: 60-64, 2012.
[10] Paul, K. K., Dhar-Chowdhury, P., Haque, C. E., Al-Amin, H. M., Goswami, D. R., Kafi, M. A. H., Brooks, W. A., "*Risk factors for the presence of dengue vector mosquitoes, and determinants of their prevalence and larval site selection in Dhaka, Bangladesh*", 2018.
[11] Althouse, B.M., Ng, Y.Y., Cummings, D.A., "Prediction of dengue incidence using search query surveillance", 2011,
[12] Duane J. Gubler, "The Economic Burden of Dengue", Volume 86(5): 743–744, 2012.
[13] Luis R. Carrasco, Linda K. Lee, Eng Eong Ooi, Donald S. Shepard, Tun L. Thein, Victor Gan, Alex R. Cook, David Lye, Lee Ching Ng, Yee Sin Leo, "Economic Impact of Dengue Illness and the Cost-Effectiveness of Future Vaccination Programs in Singapore", 2011.
[14] Hani M. Aburas, B. Gultekin Cetiner, Murat Sari, "Dengue-Confirmed Cases Prediction: A Neural Network Model".
[15] James A. Potts, Robert V. Gibbons, Alan L. Rothman, Anon Srikiatkhachorn, Stephen J. Thomas, Pra-on Supradish, Stephenie C. Lemon, Daniel H. Libraty, Sharone Green, Siripen Kalayanarooj, "Prediction of Dengue Disease Severity among Pediatric Thai Patients Using Early Clinical Laboratory Indicators", 2010.
[16] MD Nazmul Karim, Saif Ullah Munshi, Nazneen Anwar, Md. Shah Alam, "Climatic factors influencing dengue cases in Dhaka city: A model for dengue prediction", Indian Journal of Medical Research, Jul; 136(1): 32–39, 2012.
[17] Benjamin M. Althouse, Yih Yng Ng, Derek A. T. Cummings, "Prediction of Dengue Incidence Using Search Query Surveillance", 2011.
[18] Nirosha Sumanasinghe, Armin R. Mikler, Jayantha Muthukudage, Chetan Tiwari, Reynaldo Quiroz, "Data Driven Prediction of Dengue Incidence in Thailand".
[19] Arul Earnest, Say Beng Tan, Annelies Wilder-Smith, David Machin, "Comparing Statistical Models to Predict Dengue Fever Notifications", Computational and Mathematical Methods in Medicine Volume, doi:10.1155/2012/758674, 2012.
[20] Howell T. Ho, Thaddeus M. Carvajal, John Robert Bautista, Jayson Dale R. Capistrano, Katherine M. Viacrusis, Lara Fides T. Hernandez, Kozo Watanabe, "Using Google Trends to Examine the Spatio-Temporal Incidence and Behavioral Patterns of Dengue Disease: A Case Study in Metropolitan Manila, Philippines", 2018.
[21] Naiyar Iqbal, Mohammad Islam, "Machine Learning for Dengue Outbreak Prediction: A Performance Evaluation of Different Prominent Classifiers", 363–371, 2019.
[22] Iqbal, N. and Islam, M., "Machine learning for Dengue outbreak prediction: An outlook, International Journal of Advanced Research in Computer Science", 8(1):93-102, 2017
[23] Fathima, A., & Manimegalai, D. "Predictive analysis for the arbovirus-dengue using svm classification", International Journal of Engineering and Technology, 2(3), 521-7, 2012.
[24] Tin Kam Ho, "The random subspace method for constructing decision forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832–844, Aug 1998.
[25] Kumar, M. N., "Alternating decision trees for early diagnosis of dengue fever", arXiv:1305.7331, 2013.
[26] A. Cano, J. G. Castellano, A. R. Masegosa, and S. Moral, "Selective gaussian naive bayes model for diffuse large-b-cell lymphoma classification: Some improvements in preprocessing and variable elimination", pp. 908–920, 2005.
[27] Chadwick, D., Arch, B., Wilder-Smith, A., & Paton, N., "Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis", Journal of Clinical Virology, 35(2), 147-153, 2006
[28] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," The bulletin of mathematical biophysics, vol. 5, no. 4, pp. 115–133, 1943.