Machine Learning Algorithms for Prediction of Dengue Disease

[1] Yalla Venu Madhavi, [2] S. Aruna

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

venumadhaviyalla1@gmail.com

[2], Assistant Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280

***Abstract**: Dengue fever is a global trouble, especially in Yemen. Although early detection is essential to reduce the mortality price of dengue, the best diagnosis of dengue requires extra time because of the multiplicity of diagnoses. This trouble therefore requires the improvement of latest diagnostic methods. The purpose of this mission is to increase a diagnostic version for the early analysis of dengue the usage of effective getting to know techniques (EMLT). This article gives a dengue prediction model based totally on EMLT. Five different styles of mastering models are K-Nearest Neighbour (KNN), Gradient Boosting Classifier (GBC), Extra Tree Classifier (ETC), extreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (Light GBM). All classifiers are skilled and tested on the information using 10-fold go-validation and Holdout go-validation tactics. At the check, every version is evaluated the use of distinct metrics: precision, F1 situation, do not forget, precision, AUC, and run time. According to the results, the ETC model achieves the highest cross-validation and 10-fold go-validation accuracy, with 99.12% and 99.03%, respectively. In the Holdout go-validation method, we concluded that the best classifier with excessive people is ETC, which done 99.12%. Finally, the experimental results display that the overall performance of the goods inside the insurance opposition outperforms 10-fold go-validation. As an end result, the dengue prediction approach proved to be useful and effective in supporting fitness specialists to appropriately be expecting dengue.*

Keywords: Dengue Disease; Machine Learning; Extra Tree; SMOTE+ENN; balanced dataset

I. INTRODUCTION

Dengue is an ailment transmitted by mosquitoes which is easily spread

ISSN: 2366-1313

during hot weather. The vector of transmission is the female mosquito, referred to by the name of "Aedes aegypti". Temperature, precipitation as well as rapid, unplanned urbanization is among the major causes of dengue's rapid spread throughout the tropical. Dengue instances have grown dramatically in recent years worldwide. The actual amount of cases of dengue has been by never recorded or calculated correctly. According to a WHO document the WHO estimates that 390 million cases of dengue fever are recorded worldwide each year, of which 96% are confirmed medically grave. Today, a variety of screening systems as well as diagnostic models have been developed to help doctors pick off and recognize diseases. Recently, artificial intelligence has become widely utilized to analyze medical records and a wide range of systems for selection assistance have been developed through the application of device mastering as well as deep learning. Artificial intelligence technology can improve the health and quality of living for thousands and millions of humans in the coming decades. The techniques also function well for predicting cases of dengue. But, using methods to master the gadgets in order to predict dengue outbreaks based mainly on unbalanced and often contradictory scientific data and, for the most part, remains an endeavour. Many of these elements influence the precision of forecasting dengue models.

II LITERATURE REVIEW

Subitha, N., and colleagues [7] of 2013 used the okay-manner set of rules that randomly picks a good variety of objects, each that is first the clusters imply or central. In each of the closing instruments, an object is assigned to the group that is the most equivalent, in accordance with the distance between the object and the clusters imply. It then computes a new suggestion for each cluster. This process continues until the characteristic of the criterion converges. This concept can be applied within the realm of photo segmentation, where you apply the blood micro-image as it enters and then the signal is removed by a neural network to get an accurate result regarding dengue fever.

ISSN: 2366-1313

Bhavani, M., and colleagues [3] in the dataset used in 2016 comprised of attributes such as bleeding, fever, flavour, fatigue. The main objective of the study is to evaluate the effectiveness of different types of techniques and to compare their performance overall. The classes used in this research are REP Tree, J48, SMO, Zero R and Random Tree. The effectiveness of these techniques was compared against using charting graphs and tables. Weak the device for mining records is employed to create the types.

III System Analysis

EXISTING SYSTEM:

The present device to aid in earlier analysis of Dengue disorder detection makes use effective techniques for studying gadgets that are based on the medical information typically involves the gathering of information about patients, including symptoms and signs along with medical records as well as laboratory test results. Machine learning algorithms are applied to study the data, hoping to identify patterns, and anticipate that there will be a chance for Dengue infection. The systems typically employ features like decision bushes and assist vector machines or neural networks, which provide an early warning system and aid health professionals with timely interventions and control of patients.

DISADVANTAGES OF EXISTING SYSTEM:

A lack of, inaccurate documents from a medical professional may result in inaccurate predictions which can hinder the device's normal performance.

PROPOSED SYSTEM:

The system proposed for the earlier diagnosis of Dengue illness prediction by making use of a green system to be aware of techniques that are based entirely on medical records has been created to deal with the present issues thoroughly. It requires the establishment of a solid information collection system to gather sure that there is a wide array of data from science, as well as various affected-person information and laboratory tests. Additionally, the system is able to prioritize explanation for each version and also the

Device that is gaining information about trends or presenting clean reasons in order to improve healthcare experts' faith and knowledge of forecasts. Updates and continuous

ISSN: 2366-1313

version education may be integrated in order to adjust into new Dengue virus strains and changing health information. Secure and strict privacy procedures could be put in place to safeguard patient information and meet moral standards. Studies of rigorous validation and assessment for various populations of patients could be conducted to determine its generalizability as well as accuracy. In addition, a feedback loop between healthcare professionals will allow constant improvement and refinement of its capabilities to create an improved medical decision aid.

ADVANTAGES OF PROPOSED SYSTEM:

The new device to aid in an early Dengue disease prediction using the latest methods of studying machines provides an integrated approach to solving the challenges of today. It is able to provide high-quality accuracy diagnosing by obtaining full medical records, and employing solid modelling strategies. It also emphasizes trust and openness through the ability to explain versions and thus gaining the trust from healthcare professionals.

IV Data Set Description

The data used in the training of our model for dengue prediction came from data on patients' clinical indications and symptoms, as well as physical modifications that are observed through the infection, in addition to the body temperature, pain in the eyes and fever. The entire set of attributes is provided by our algorithm so that training could be done by using this device to determine the likelihood of catching dengue. The information was collected from private and public hospitals. The data set contained stats about patients from five South Indian states, particularly Tamil Nadu, Andhra Pradesh, Telangana, Kerala and Karnataka and Karnataka, states that suffer from an overly high rate of mortality, a dangerously high threat, and deadly health outcome of the condition. It contains 23 attribute, and 1640 data in the database. The description of the data is provided below.

ISSN: 2366-1313

```
...       Age  Sex  Fever  Headache  Arthralgia  Myalgia  \
0          14    1      0         1           1        1
1          27    1      1         1           1        1
2          28    1      1         0           0        1
3          61    0      1         0           1        1
4          37    1      0         0           1        0
...       ...  ...    ...       ...         ...      ...
1635       37    0      1         1           1        1
1636       27    0      0         1           0        1
1637       23    0      0         0           1        1
1638       30    0      0         0           1        0
1639       30    0      1         0           1        1

      Conjunctivitis or Pain behind eyes  Skin rash  Generalized weakness  \
0                                      0          1                     0
1                                      1          1                     1
2                                      0          1                     0
3                                      1          1                     1
4                                      0          1                     0
...                                  ...        ...                   ...
1635                                   0          1                     1
1636                                   1          0                     1
1637                                   0          1                     0
1638                                   0          0                     1
1639                                   1          1                     1
...
1638     0              0    0    0      0
1639     1              0    1    1      0

[1640 rows x 23 columns]
```

Dataset source: Kaggle

Dataset Format: Excel Format

Rows: 1640

Columns: 23

Age - Age of a patient in years

Gender -sex of patient

Fever - severity of fever (High / low)

Vomiting - if patient had feeling of nausea (Y / N)

Headache - if the patient had headache or not

Arthralgia -if the patient had a joint pain(Y/N)

Myalgia - if the patient had a muscle pains(Y/N)

Jaundice - yellowing of a skin or eye(Y/N)

Abdominal Pain - pain anywhere between your chest and groin (Y/N)

Watery Diarrhea - a patient has infections in the intestinal tract(Y/N)

Ecchymosis - blood pooled under the skin(Y/N)

Meningites - viral infection of patient(Y/N)

Kidney Failure - infection of kidney(Y/N)

IgM - test the Immunoglobulin M in a patient(Y/N)

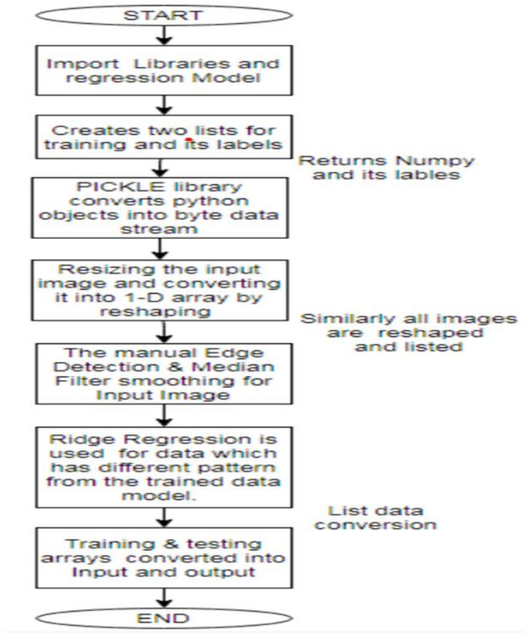
IgG - abundant antibody in the blood (Y/N)

SYSTEM DESIGN

SYSTEM ARCHITECTURE

System structure diagrams are an abstract representation of the design of devices components. It gives a short explanation of the components' structure to aid in understanding things-element relations and system operations. Diagrams of the gadget's structure are an image of the structure of the device. It reveals the relationships between each of the distinct components in the machine and also outlines what functions each element does. The most well-known image of the device reveals that the main capabilities of the device and also its relationships with the extraordinary components of the machine.

ISSN: 2366-1313

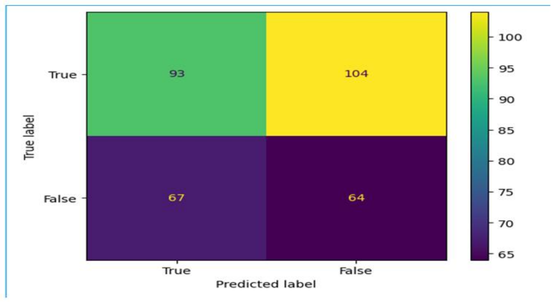


V MACHINE LEARNING ALGORITHMS

This Dengue Disease Prediction undertaking for studying the impact of the machine learning models include Decision Tree Classifier and Random Forest Classifier. Utilizing these methods, we are forecast the results of someone having dengue, or not. If we receive fee 1, i.e... If he does have dengue but in other situation, he's not always. In the event that we provide all of the indications and symptoms, it is possible to determine the outcome.

Confusion Matrix:

The confusion matrix can be which is used to evaluate the how well the models perform in a collection of test information. It is simplest to determine when the true values to the records you look at can be identified. The structure of the matrix is unproblematic to comprehend, however the terminology used to describe it can be confusing. Because it reveals the mistakes in the overall performance of the version in the form of the matrix, it is also called a errors matrix.



True Positive (TP): The model has predicted yes while the real price is actual.

True Negative (TN) This model provides prediction with no information about costs, which are real and is also fabricated.

False Positive (FP) A false positive is a prediction that said True. The real or true version is forecasting False.

False Negative (FN) A prediction model that predicts False, and costs that are real as well as false.

Accuracy:

ISSN: 2366-1313

It's one of the primary parameters that determine the precision of problem of classification. It determines how often the model predicts a correct output. It can be determined in terms of the ratio between the ranges of predictions that are correct that are made using the classifier, to the array of predictions that are made by the classifiers. The formula can be found below:

Accuracy= TP+TN/TP+TN+FT+FN

= 93+64/93+67+104+64

=0.47

Precision:

The reason for this is the amount of exact outputs that are provided by the model or the numerous fantastical instructions that have been successfully predicted by the software, and the amount of which are almost genuine. The number of accurate outputs can be determined with the following system:

Precision = TP/TP+FP

=93/93+67

=0.58

Recall:

It's defined as an out-of-general top quality instruction, which is how our model has predicted accurately. The amount of consideration should be as large as it is possible.

Recall =TP/TP+FN

=93/93+104

=0.47

F1_Score:

When two models are of very low precision, but have high recall, or vice versa it can be difficult to assess the quality of two models. For this reason you can make use of F-score. This score helps us assess the accuracy and recall while at the same time. The F-score is the highest score in the event that recall is comparable to accuracy. It is calculated by with the formula below:

F1_Score = 2* recall*precision/ recall + precision

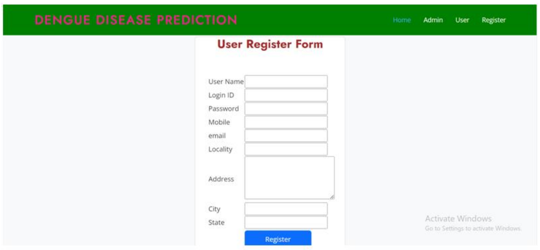
=2*0.47*0.58/0.47+0.58

=0.52

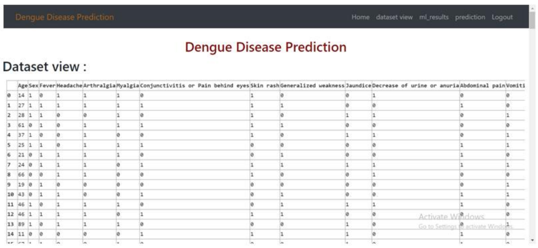
OUTPUT SCREENS

Home page

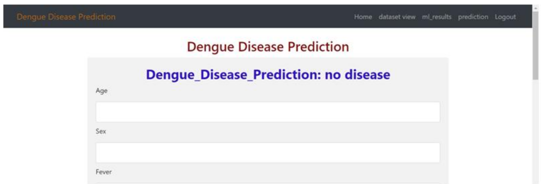
Register Form



Dataset view



Prediction results



VI CONCLUSION

Dengue disease is a worldwide issue today. The early detection and treatment of dengue may help avoid migraines and even save lives at the grocery store. In this study we propose dengue prediction tools and reviewed the general efficiency of various gadget mastering techniques that can predict dengue (the use for Decision Tree Classifier and Random Forest Classifier). The primary portion of the process, i.e. The initial part of the study the values that were not present had been determined using the technique of averaging. Techniques for determining features have been employed to identify crucial aspects. Statistics have been normalized in line with Z-Score. To address the issue of imbalance in the dataset we employed the method of SMOTE+ENN that is hybrid. The pass-validation strategy we used was 10 fold move-validation, and holdout to separate the dataset into schooling and examining models. The device learning about models been developed and their effectiveness was assessed in terms of precision, F1 score, precision and keeping in mind.

REFERENCES

1. I.S.Abubakar, S.B.Abubakar, A.G.Habib, A.Nasidi, N.Durfa,P. O.Yusuf, S.Larnyang, J.Garnvwa, E.Sokomba, "Randomised Controlled Double-Blind Non-Inferiority Trial of Two Antivenins for Saw-Scaled or Carpet Viper (Echis ocelots) Envenoming in Nigeria," PLoS Neglected Tropical Diseases, vol. 4, no. 7, p. e767,Jul. 2010. https://doi.org/10.1371/journal.pntd.0000767.

2. A.A.H. Nassar, A.A.Torbosh, Y.A. Mahyoub, and M.A.A. Amad, "Risk

ISSN: 2366-1313

Factors Associated with Dengue Fever Outbreak in Taiz Governorate, Yemen, 2018: Case-control Study," Jul. 2021,https://doi.org/10.21203/rs.3.rs-624873/v1.

3. S. S. Nimmannitya, "Dengue and Dengue Haemorrhagic Fever, "Manson's Tropical Diseases, pp. 753–761, 2009,https://doi.org/10.1016/b978-1-4160-4470-3.50045-8.

4. D.J.GUBLER, "Dengue and Dengue Hemorrhagic Fever,"Tropical Infectious Marimuthu, T., and V. Balamurugan. "A novel bio computational model for mining the dengue gene sequences, "International Journal of Computer Engineering & Technology, vol. 6, no. 10, pp. 17-33, Oct.2015.

5. Rao, NK Kameswara, GP Saradhi Varma, D. Rao, and P. Cse."Classification rules using decision tree for dengue disease, "International Journal of Research in Computer and Communication Technology, vol. 3, no. 3, pp. 340-343, Mar.2014.

6. P. Manivannan and P. I. Devi, "Dengue fever prediction using K-means clustering algorithm, "2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Mar. 2017, https://doi.org/10.1109/itcosp.2017.8303126.

7. K. S. Ahmed Bin and S. Kamran Jabbar, "Dengue Fever in Perspective of Clustering Algorithms," Journal of Data Mining in Genomics & Proteomics, vol. 06, no. 03, 2015, https://doi.org/10.4172/2153-0602. 1000176.

8. N. A. Husin, N. Salim, and A. R. Ahmad, "Modelling of dengue out break prediction in Malaysia: A comparison of Neural Network and Nonlinear Regression Model,"2008 International Symposium on Information Technology,://doi.org/10.1109/itsim.2008.4632022.

9. A. Padmapriya and N. Subitha, "Clustering Algorithm for Spatial Data Mining: An Overview," International Journal of Computer Applications, vol. 68, no. 10, pp. 28–33, Apr. 2013, https://doi.org/10.5120/11617-7014.

10. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.