International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper Volume-6, Issue-7 ,2019 Regular Edition E-ISSN: 2349-7084

Prediction of Dengue with the Use of AI and Data Mining: An Expert System

Dinesh Chandra[1*], Shaziya Islam[2], Deepak pandey[3]

[1*]*Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai, India*
[2] *Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai, India*
[3]*Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai, India*

e-mail:dinmax007@gmail.com, secondauthor_shaziya.islam26@gmail.com, thirdauthor_14deepak1993@gmail.com

**Corresponding Author: dinmax007@gmail.com ,*

Available online at: http://www.ijcert.org

Received: 30/07/2019, Revised: 13/08/2019, Accepted: 28/August/2019, Published: 02/September/2019

Abstract:- **Background/Objectives**: Dengue fever is a mosquito-borne tropical disease caused by the dengue virus. It is a life-threatening disease lots of people died due to dengue because its symptoms are not detected at early stages many parsons thought that it was a normal fever or headache so that they ignore it which cause there are in dangerous situations and worst case they lose their life.

Methods/Statistical analysis: We applied data mining techniques along with artificial intelligence technique to create an expert system which can diagnose dengue with the help of symptoms provided by the users. In data mining portion, we use data filtering, data cleaning and clustering, and some other technique to enhance our dataset. Moreover, in AI portion we create an expert system where we create a knowledge base, fact base and GUI portion through user enter their symptoms and our system work is to predict dengue based on symptoms that user feed in GUI as input.

Findings: With the implementation of this project, we expect that our expert system is capable of predicting dengue based on person symptom's that we take as a Dataset and saves lots of life of various persons. Main aim this project is maccuracy easure accuracy and efficiency also because there is lots of work is pending in this area, and some researchers are searching for new methods.

Improvements/Applications: our proposed work will apply in the field of the medical area where a person is capable of checking their dengue symptoms and analyzing their disease.

Keywords: Dengue, Malaria, Data Mining, Machine Learning, Artificial Intelligence, Expert System

1. Introduction

Necessity is the mother of invention. Since the earlier period, our ancestors are sorting out essential data from information by hand. However, with the speedily increasing volume of knowledge nowadays, a lot of automatic and practical mining approaches are needed. New strategies like Bayes theorem within the 1700s and multivariate analysis within the 1800s were a number of the primary techniques accustomed establish patterns in information. After the decade, with the proliferation, ubiquity, and incessantly developing the power of engineering, information assortment, and information storage were remarkably enlarged. As data sets have grown up in size and complexness, direct active information analysis has progressively been increased with indirect, automatic data processing. This has been assisted by alternative discoveries in engineering science, like neural networks, clustering, genetic algorithms within the Fifties, call trees within the

Nineteen Sixties, and support vector machines within the Nineteen Eighties. Data mining is the process of applying these ways to information to uncover hidden patterns.

Data mining or data processing technology has been used several for several years by many fields like businesses, scientists, and governments. it's accustomed sift through volumes of knowledge} like airline traveler trip information, population knowledge and promoting knowledge to come up with marketing research reports, though that news is usually not considered to be data mining

Dengue could be a life-threatening sickness prevailing in many developed still as developing countries like Asian country. This is often a virus and deadly disease or mosquito-borne disease caused by breeding of arthropod genus mosquito. Datasets that are accessible for dengue describe data regarding the patients suffering with dengue disease and without breakbone fever disease together with their symptoms like Fever Temperature, WBC, Platelets, Severe Headache, Vomiting, metallic taste, Joint Pain, Appetite, Diarrhea, Hematocrit, Hemoglobin, and the way several days suffer in numerous town.

Dengue is presently considered the most relevant arthropod-borne viral disease in terms of morbidity and mortality, affecting tens of millions of people annually worldwide. Its main vector is the mosquito Aedes aegypti, usually found in urbanized areas in tropical and subtropical countries. There are four serotypes of dengue virus [2]. Dengue fever is a mosquito-borne tropical disease caused by the dengue virus.

It is a life-threatening disease prevalent in several developed as well as developing countries like India. This is a virus born disease caused by breeding of Aedes mosquito. Symptoms of dengue are: Fever Temperature, WBC, Platelets, Severe Headache, Vomiting, Metallic Taste, Joint Pain, Appetite, Diarrhea, Hematocrit, Hemoglobin, and how many days suffer in the different city [2].

Lots of people died due to dengue because its symptoms are not detected at early stages many parsons thought that it was a normal fever or headache so that they ignore it which cause there are in dangerous situations and the worst case they lose their life[3]. In this system, we applied data mining techniques along with artificial intelligence technique to create an expert system which can diagnose dengue with the help of symptoms provided by the users. In data mining portion, we use data filtering, data cleaning, and clustering, and some other technique to enhance our dataset.

Moreover, in AI portion we create an expert system where we create a knowledge base, fact base and GUI portion through user enter their symptoms and our system work is to predict dengue based on symptoms that user feed in GUI as input[4].

2. Related Work

In 2018 P. Siriyasatien1, S. Chadsuthi2, K. Jampachaisri3, K. Kesorn*,4 proposed a Dengue Epidemics Prediction: A Survey of the State-of-the-Art based on Data Science Processes where analyzes the significant components that can be used in a dengue prediction model. They have attempted to identify the factors directly related to the probability of a dengue epidemic occurring, particularly climate factors, the rate of mosquito bites, rainfall, and the rate of dengue infection in mosquitoes as the essential factors contributing to severe outbreaks. Another surveillance method for dengue outbreaks surveillance is to ascertain information posted by tourists on social networks. The immediacy of this information quickly alerts the international community about the prevalence of the disease in an area. Dengue outbreaks are a risk for all travelers who should, therefore, study the seasonal risk factors at their intended destinations, and seek and receive accurate advice about preventive measures during their travel[4].

In 2018 Abdul Mahatir Najar, Nopember Surabaya, Mohammad Isa Irwan proposed an Extreme Learning Machine Method for Dengue Hemorrhagic Fever Outbreak Risk Level Prediction where they apply extreme learning machine (ELM) method to predict the risk of the outbreak based on weather condition. They Develop ELM architecture with weather variables as input nodes and risk level of DHF outbreak as the target. They use binary sigmoid activation function and bipolar sigmoid with several hidden neurons between 5 - 200 nodes. The results show that ELM can predict the level of risk of DHF with the best performance of ELM network using a binary sigmoid activation function with 50 hidden neurons[5].

In 2018 Virginia Ortiz Andersson, Marco A. Ferreira Bircky and Ricardo Matsumura Araujoz proposed a methodology on Towards Predicting Dengue Fever Rates Using Convolutional Neural Networks and Street-Level Images where they investigate the use of street-level images, such as those from Google Street View, along with Convolutional Neural Networks to predict Dengue Fever (DF) and Dengue Hemorrhagic Fever (DHF) rates in urban locations. They conduct a case study in the city of Rio De Janeiro, Brazil, using the proposed methodology and DF/DHF data between the years of 2010 to 2014. We compare two Siamesebased CNNs, yielding an overall accuracy of 67% for 20,400 different locations. We conclude that street-level images are useful for the problem[6].

In 2018 Ria Arafiyah1 and Fariani Hermin1 proposed a Data mining for dengue hemorrhagic fever (DHF) prediction with naive Bayes method where The use of data mining method, able to build database support in decision-makers diagnose DHF disease This study predicts DHF with the method of Naive Bayes. Parameter of the input variable is the patient's medical data (temperature, spotting, bleeding, and tornuine test) and the output variable suffers from DBD or not while the system output is a diagnosis of the patient suffering from DHF or not. Result of the model test by using tools of Orange 3.4.5 obtained level of the precision model is 77,3% [7].

In 2018 P. Sathya [1], Mrs. A.Sumathi [2] proposed a method for Predicting Dengue Fever Using Data Mining Techniques where they are trying to find out some of the characteristics on Dengue illness so that they can rightly categorize patients because different patients require different types of treatment. Pakistan has been the target of Dengue illness from last few years. Dengue fever is used in clustering techniques to evaluate their actions. The dataset was gathered from lotus and 24 care hospitals. For accurately classify our dataset, various clustering techniques are used. Appraise the performance of all the techniques individual based on tables and graphs depending upon dataset [8].

In 2017 Kashish was Shakil, Sharma Anis and Mansaf Alam proposed a DENGUE Disease Prediction Using WEKA DATA-MINING TOOL where they have firstly classified the dengue data set and then compared the different data mining techniques in weka through Explorer, knowledge flow and Experimenter interfaces. Furthermore, to validate our approach, we have used a dengue dataset with 108 instances, but weka used 99 rows and 18 attributes to determine the prediction of disease and their accuracy using classifications of different algorithms to find out the best performance. Their main objective is to classify data and assist the users in extracting useful information from data and quickly identify a suitable algorithm for the accurate predictive model from it. From the findings of this paper it can be concluded that Naïve Bayes and J48 are the best performing algorithms for classified accuracy because they achieved maximum accuracy= 100% with 99 correctly classified instances, maximum ROC = 1, had least mean absolute error and it took minimum time for building this model through Explorer and Knowledge flow results[2].

In 2017 H. Abdul Rahiml, F. Ibrahim proposed A NOVEL PREDICTION SYSTEM IN DENGUE FEVER USING NARMAX MODEL where the development of nonlinear autoregressive moving average with exogenous input (NARMAX) models in diagnosing dengue infection. The developed system bases its prediction solely on the bioelectrical impedance parameters and physiological data. Three different NARMAX model order selection criteria, namely FPE, AIC, and Lipschitz have been evaluated and analyzed. This model is divided into two approaches, which are the unregularized approach and regularized approach. The results show that using the Lipschitz number with regularized approach yield better accuracy by 88.40% to diagnose the dengue infectious disease. Furthermore, this analysis show that the NARMAX model yield better accuracy as compared to autoregressive moving average with exogenous input (ARMAX) model in diagnosis intelligent system based on the input variables namely gender, weight, vomiting, reactance and the day of the fever as recommended by the outcomes of statistical tests with 76.70% accuracy[9].

In 2017 Abdul Mahatir Najar, Nopember Surabaya, Mohammad Isa Irawan proposed an Extreme Learning Machine Method for Dengue Hemorrhagic Fever Outbreak Risk Level Prediction where they apply extreme learning machine (ELM) method to predict the risk of the outbreak based on weather condition. They Develop ELM architecture with weather variables as input nodes and risk level of DHF outbreak as the target. They use binary sigmoid activation function and bipolar sigmoid with some hidden neurons between 5 - 200 nodes. The results show that ELM can predict the level of risk of DHF with the best performance of ELM network using a binary sigmoid activation function with 50 hidden neurons[1].

In 2017 Ms.S.Freeda Jebamalar, Dr.A.Anitha proposed A Survey on Prediction of Dengue Fever Using Data Mining Techniques where they survey on dengue disease and it symptoms, where they survey the importance of data mining techniques and them, concluded that data mining techniques are more effective in predicting the dengue disease with accuracy ranging from 70% to 100% [10]..

In 2017 P. Manivannan, Dr. P. Isakki Devi proposed a method for Predicting Dengue Fever Prediction Using K-Means Clustering Algorithm where has been focused four stages namely preprocessing, attribute selection, clustering and predicting the dengue fever. R 3.3.2 Tool is used for preprocessing the household of dengue dataset. D win's method has been applied to generate filled dataset by substituting all missing values for nominal and numeric attributes with mode and mean value. Dengue virus can be predicted by applying different data mining techniques. The main goal of research work is to predict the people who are affected by dengue depending upon the categorization of the age group using the K-means clustering algorithm has been implemented[11].

3. Methodology

The Purpose of Our project is to detect dengue based on symptoms of users that are provided by them. For detecting, we collect Data set of various persons from various

sources like internet, twitter, and some previous author's references, and we collect some data from the hospital also. We aim to create an expert system which detects dengue with the help of Data mining and Artificial Intelligence Technique [3].

We implement Data mining technique in Data set pre-processing, filtering, cleaning, clustering, and classification phase, as well as we, implement Artificial Intelligence technique to create knowledge acquisition, fact-based knowledge base, etc.

Our goal is to create an expert system which detects dengue diseases based on systems of a person with the help of Data mining and Artificial Intelligence.

Algorithm 4.1: Proposed Algorithm to Detect Dengue

Step 1: Creation of Data set.

Step 2: Apply Data mining techniques, such as filtering, pre-processing, and cleaning.

Step 3: We use a classifier to classify dengue data with respect the others i.e., between infected and non-infected persons.

Step 4: From the Data set, we acquire knowledge about the system of dengue.

Step 5: Now, from the above information, we create a knowledge base and collect facts from it.

Step 6: Our next task is to create an expert system through which a user can input their diseases symptom, and our system can predict dengue.

Step 7: Now, we create an interface engine which helps to interact uses from the knowledge base.

Step 8: Then, we create a GUI that is a user-friendly environment which helps the user to enter their systems.

1. Data set

A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity.

A data set is organized into some data structure. In a database, for example, a data set might contain a collection of business data (names, salaries, contact information, sales figures, and so forth). The database itself can be considered a data set, as can bodies of data within it related to a particular type of information, such as sales data for a particular corporate department[4].

"For the creation of the data set, we collect records of various patient's from internet and nearby hospital and other sources."

2. Pre-processing

Data pre-processing describes any processing performed on raw data to prepare it for another processing procedure. Commonly used as a preliminary data mining practice, data preprocessing transform the data into a format that will be more easily and effectively processed for the user -- for example, in a neural network.

There square measure variety of various tools and strategies used for preprocessing, including: sampling, that selects a representative set from an outsized population of data; transformation, that manipulates information to provide one input; denoising, that removes noise from data; standardization, that organizes knowledge for additional economical access; and have extraction, that pulls out specific knowledge that's vital in some explicit context

"In the pre-processing phase, we collect dataset from various sources, and we apply some pre-processing techniques then we use it to data cleaning phase."[5]

3. Data Cleaning

Data cleansing is that the method of neutering information during a given storage resource to create positive that it's correct and proper. There is a unit some ways to pursue information cleansing in varied computer code and information storage architectures; most of them center on the careful review {of information of knowledge of information} sets and also the protocols related to any explicit data storage technology.

• Data cleaning: fill in missing values, smooth noisy data, identify or remove

Outliers and resolve inconsistencies.

• Data integration: using multiple databases, data cubes, or files.

• Data transformation: normalization and aggregation.

• Data reduction: reducing the volume but producing the same or similar analytical results.

• Data discretization: part of data reduction, replacing numerical attributes with simple ones.[3]

4. Filtering

Data filtering in IT will seek advice from a good vary of methods or solutions for processing information sets. This implies the information sets square measure refined into merely what a user (or set of users) wants, while not together with different information which will be repetitive, unsuitable or perhaps sensitive. Differing kinds of knowledge filters are accustomed amend reports, question results, or other forms of knowledge results.[2]

5. Clustering

Clustering is that the task of dividing the population or knowledge points into a variety of teams specified knowledge points within the same team's ar a lot of almost like alternative knowledge points within the same cluster than those in alternative teams. In straightforward words, the aim is to segregate teams with similar traits and assign them into clusters.[2]

6. Components of Expert Systems

The strength of an Expert System (ES) derives from its knowledge domain - AN organized assortment of facts and heuristics regarding the system's domain.

An ES is made in a method referred to as knowledge engineering, throughout that data regarding the domain is acquired from human specialists and different sources by knowledge engineers that is the build-up of data in knowledge bases, from that conclusion square measure to be drawn by the abstract thought engine, is that the hallmark of an expert system.[3]

7. Knowledge Base

It contains domain-specific and high-quality data. Data is needed to exhibit intelligence. The success of any Es majorly depends upon the gathering of extraordinarily correct and precise information. What's Knowledge? Data} is an assortment of facts. The {data} is organized as data and facts concerning the task domain. Data, information, and past expertise combined along are termed as knowledge.[4] elements of information Base The mental object of associate Es may be a store of each, factual and heuristic information.

- **Factual Knowledge** − it's the data widely accepted by the knowledge Engineers and students within the task domain.
- **Heuristic knowledge** − it concerns practice, correct judgment, one's ability of analysis, and approximation.

Information representation it's the strategy accustomed to organize and formalize the knowledge within the knowledge base. It's within the variety of IF-THEN-ELSE rules.[6]

8. Knowledge Acquisition

The success of any expert system majorly depends on the standard, completeness, and accuracy of the knowledge hold on within the knowledge base.

Readings make the knowledge base from numerous specialists, scholars, and therefore the knowledge Engineers. The knowledge engineer could be a person with the qualities of fellow feeling, fast learning, and case analyzing skills.

He acquires data from subject skilled by recording, interviewing, and observant him at work, etc. He then categorizes and organizes the knowledge during a purposeful means, within the kind of IF-THEN-ELSE rules, to be utilized by interference machine. The knowledge engineer conjointly monitors the development of the Es.[6]

9. Inference Engine

The inference engine: Combines the facts of a selected case with the information contained within the knowledge base to return up with a recommendation. In a rule-based expert system, the inference engine controls the order during which production rules are applied (Afired@) and resolves conflicts if over one rule is applicable at a given time. This can be what Areasoning@ amounts to in rule-based systems.

Directs the user interface to query the user for any info it wants for additional inferencing. The facts of the given case are entered into the working memory, which acts as a sheet, accumulating the data concerning the case at hand. The abstract thought engine repeatedly applies the foundations to the memory, adding new info (obtained from the rules conclusions) to it, till a goal state is created or confirmed.[6]

10. GUI

A graphical user interface (GUI) is an interface through that a user interacts with electronic devices like computers, hand-held devices, and alternative appliances. This interface uses icons, menus, and alternative visual indicator (graphics) representations to show data and connected user controls, in contrast to text-based interfaces, wherever knowledge and commands area unit in text. GUI representations area unit manipulated by an inform device like a mouse, trackball, stylus, or a finger on a touch screen. .[7]

11. End-User

They're the final or ultimate user of a computer system. The end-user is that the individual who uses the merchandise after it's been developed and marketed. The term is helpful as a result of it distinguishes 2 categories of users, users who need a bug -free and finished product (end users), and users who might use a similar product for development functions. The term end-user sometimes implies an individual with a comparatively low level of computer expertise.

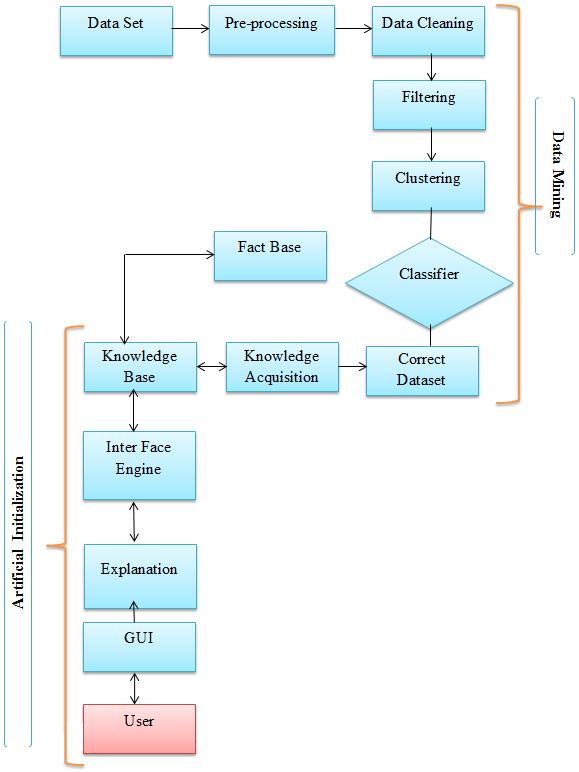

Figure 1.Proposed flow model

4. Results and Discussion

In the proposed work, we expect that our expert system is capable of predicting dengue based on person symptom's that we take as a Dataset. We have studied several research papers related to dengue and its prediction. Many authors have implemented various techniques on the prediction of dengue. Some techniques give good results, and some are useful information. In this proposed work, we study the literature of various authors and take a basic idea about the implementation of this project. Our necessary work in this project is to create an expert system which uses data mining techniques, Artificial Intelligence, Machine Learning, and Natural Language processing. With the help of these techniques, we are capable of creating an expert system which helps us to predict dengue diseases based on infected person symptoms.

Dengue fever is a mosquito-borne tropical disease caused by the dengue virus. It is a life-threatening disease prevalent in several developed as well as developing countries like India. This is a virus born disease caused by breeding of Aedes mosquito. Symptoms of dengue are: Fever Temperature, WBC, Platelets, Severe Headache, Vomiting, Metallic Taste, Joint Pain, Appetite, Diarrhea, Hematocrit, Hemoglobin, and how many days suffer in a different city.

Lots of people died due to dengue because its symptoms are not detected at early stages many parsons thought that it was a normal fever or headache so that they ignore it which cause there are in dangerous situations and in the worst case they lose their life.

With the implementation of this project, we expect that our expert system is capable of predicting dengue based on person symptom's that we take as a Dataset and saves lots of life of various persons. Our aspect of this project is in terms of accuracy and efficiency also because there is lots of work is pending in this area, and some researchers are searching for new methods.

5. Conclusion and Future Scope

The main conclusions of the study may be presented in a short Conclusion Section. In this section, the author(s) should also briefly discuss the limitations of the research and Future Scope for improvement.

Component The proposed work can be analyzed in terms of feasibility and accuracy, so identification and prediction of dengue play an essential role in saving a person's life because lots of personal life are in danger and lots of people died due to dengue because dengue is not diagnosed or person having not so much equipment's.

Our work in this project is to create an expert system which uses data mining techniques, Artificial Intelligence, Machine Learning, and Natural Language processing. With the help of these techniques, we are capable of creating an expert system which helps us to predict dengue diseases based on infected person symptoms.

Prediction of dengue is a life savior project, and with the help of this project, we can save many people's life. There is lots of work is pending in this area, and some researchers are searching for new methods. In this proposed work, we are collecting a data set of 100 persons. Our future work is to enhance our project and expand the dataset.

References

[1] Md Younis Md Alzarrous and Mr. Surya Prakash Mishra, "A novel data mining technique to discover pattern from a huge text corpus," *international journal of Modern Engineering Research*, vol. 4, no. 5, p. 6, May 2014.

[2] kashish, Anis, Shadma, Alam, Mansaf Ara Shakil, "Dengue Disease prediction Using Weka Data Mining tool," *Elsevier*, p. 26, Feb 2015.

[3] Dr. Dinesh singh, "An Empirical study of Techniques and Various Domains in Data Mining for Efficient approach in Various Fields," *International Jounral of New Inventions in Engineering and Technology* , vol. 8, no. 1, p. 7, April 2018.

[4] S. Chadsuthi2, K. Jampachaisri3, K. Kesorn* P. Siriyasatien1, "Dengue Epidemics Prediction: A Survey of the state of the art based on data science process," *ieee*, vol. 20, p. 36, october 2018.

[5] Nopember Surabaya, Mohammad Isa Irawan n Abdul Mahatir Najar, "Extreme Learning Machine Method for Dengue Hemorrhagic Fever Outbreak Risk Level Predictio," *ieee*, vol. 4, p. 5, sep 2018.

[6] Marco A. Ferreira Bircky and Ricardo Matsumura Araujoz Virginia Ortiz Andersson, "Towards Predicting Dengue Fever Rates Using Convolutional Neural Networks and Street-Level Images ," *ieee*, p. 8, June 2018.

[7] Ria Arafiyah1 and Fariani Hermin1, "Data mining for dengue hemorrhagic fever (DHF) prediction with naive Bayes method," *Jounral of Physics*, p. 5, July 2018.

[8] Mrs. A.Sumathi [2] P. Sathya [1], "Predicting Dengue Fever Using Data Mining Techniques," *IJCST*, vol. 6, no. 2, p. 3, March-April3 2018.

[9] F. Ibrahim proposed H. Abdul Rahiml, "A NOVEL PREDICTION SYSTEM IN DENGUE FEVER USING NARMAX MODEL ," *Ieee*, p. 5, october 2017.

[10] Dr.A.Anitha proposed Ms.S.Freeda Jebamalar, "A Survey on Prediction of Dengue Fever Using Data Mining Techniques," *IJSEM*, vol. 2, no. 12, p. 3, Dec 2017.

[11] Dr. P. Isakki Devi P. Manivannan, "Dengue Fever Prediction Using K-Means Clustering Algorithm," *Ieee*, p. 5, Sep. 2017.

