International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Malaria and Dengue Classification System using Machine Learning

Mrs. Shamal J. Tavandkar and Prof. Manisha Darak

Department of Computer Engineering Siddhant College Of Engineering-[SCOE] Sudumbare, Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

The health care environment is found to be rich in information, but poor in extracting knowledge from the information. This is because we don't having effective and efficient tools and methods. By applying the latest technology like machine learning algorithms and techniques and methods valuable knowledge can be extracted from the health care system can be very useful for further enhancement. Malaria and Dengue have a lots of bad syndromes which can harm human body very badly. We are using Deep Learning algorithms to increase the accuracy of Malaria and Dengue Disease prediction System. It is implemented as desktop application in which user submits the heterogeneous data like text and image of blood cells symptoms. It retrieves hidden data from stored database and deep learning model and compares the user values with trained data set.

Keywords: *Machine Learning, Disease prediction, Malaria, Dengue.*

Introduction

Diseases are caused due to various reasons. They can be transmitted through various viruses or due to some chemical reactions in our body. Among various life-threatening diseases, diseases which have similar symptoms have gathered a great deal of attention in medical research. The diagnosis of diseases with similar symptoms is a challenging task, which can offer automated prediction about the disease of patient so that further treatment can be made effective. The diagnosis of such diseases is usually based on signs, symptoms and cell image of the patient.A major challenge faced by healthcare organizations, such as hospitals and medical centers, is the lack of resources at affordable costs and accuracy with less time.

[1] In healthcare, quality service depends on diagnosing patients properly and administering effective treatments with affordable cost. The available disease database consists of both numerical and cell image data. Before applying any algorithm or any operation on available dataset we first need to check its authentication and consistency. We need to first perform data preprocessing operation on dataset to make it clean and proper.

[2] By applying the algorithm, proposed system can identify and extract the hidden knowledge from dataset, i.e. patterns and relationships associated with disease from a database. The Healthcare prediction system is an end user and an online consultation project which can help in decreasing the total time period required for accessing each patient by doctors on time. Here we propose a system that allows users to get instant guidance on their health issues having similar symptoms through a predictive health care system online.

This system is also responsible for classifying diseases having similar symptoms.

Scope

1. Ensure fast disease prediction.
2. Ensure accurate disease prediction.
3. Require less manpower
4. Ensures low cost 5. Easily accessible

Literature survey

Name-: Automatic Diagnosis With Efficient Medical Case Searching Based on EvolvingGraphs.(Xiaoli Wang 1, Yuan Wang2, Chuchu Gao1, Kunhui lin1, and yadi li3),IEEE,2018

Description-: The knowledge graph-based method to build the linkage between various types of multimodal data builds a semantic rich knowledge base using both medical dictionaries and practical clinical data collected from hospitals and proposes a graph modeling method to bridge the gap between different types of data, and the multimodal clinical data of each patient are fused and modeled as one unified profile graph and also develop a lazy learning algorithm for automatic diagnosis based on graph similarity search.

Limitation-: GP Model is not employed in graph sequence of different form of typical algorithms attempting to eagerly train a global GP model on the entire dataset.

Name-: Using Electronic Health Records and Machine Learning to Make Medical Related Predictions from Non-Medical Data (Stavros Pitoglou, Yiannis Koumpouros and Athanasios Anastasiou) ,International Conference on Machine Learning and Data Engineering,2018
Description-: The hypothesis that the application of machine learning techniques on data of this nature can be used to address prediction/forecasting problems in the Health IT domain.
Limitation-: The novelty of this approach consists in that medical data (test results, diagnoses, doctors' notes etc.) are not included in the predictors' dataset.

Name-: Monitoring Mobile Patients Using Predictive Analysis By Data From Wearable Sensors, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) 2018.
Description-: In this paper system uses sensors, the data acquisition unit, microcontroller and software.
This system is able to send alarm messages about the patient's critical health data by text messages or by email reports. By using this information the healthcare professional can provide necessary medical advising.
Limitation-: Sensors and controller, camera is costly.

Name-: Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges, Center for Applied Autonomous Sensor Systems, Orebro University, SE-70182, Orebro, Sweden;2017.
Description-: This paper provides a recent review of the latest methods and algorithms used to analyze data from wearable sensors used for physiological monitoring of vital signs in healthcare services. In particular, the paper outlines the more common data mining tasks that have been applied such as anomaly detection, prediction and decision making when considering in particular continuous time series measurements.
Limitation-: The selected data mining technique is highly dependent on the data mining task to be performed. According to the considered data mining tasks in Section 3, for anomaly detection task, SVM, HMM, statistical tools and frequency analysis are more commonly applied.

Name-: An Artificial Neural Network approach for classification of Vector-Borne diseases. (Prajwal Shimpi, Sanskruti Shah, Maitri Shroff , Anand Godbole),(ICEEOT) 2018.
Description-: Three diseases prevalent in India: malaria, dengue and chikungunya. The proposed method uses an Artificial Neural Network (ANN) based backpropagation algorithm for training and testing. A number of gradient optimization techniques are used like Adaptive Moment Estimation, RMS Prop, Adagrad, Classical Momentum and Nesterov accelerated gradient.
Limitation-: Apart from the 3 possible diseases that were taken in consideration in this paper, different set of diseases, that is, outside of vector-borne diseases are not detected.

Name-: Analytical study of heart disease diagnosis using classification Techniques. (C.Sowmiya; P.Sumitra), International Conference on Intelligent Techniques in Control, Optimization and Signal Processing,2017
Description-: In this paper the potential of nine classification techniques was evaluated of prediction of heart disease.Using medical profiles such as a age, sex, blood pressure, chest pain type, fasting blood sugar. It can predict like of patients getting heart disease Based on this, medical society takes part interest in detecting and preventing the heart disease.
Limitation-: It can only experiment a priori algorithm. Classification of diseases is not accurate.

Name-: An ensemble based on distances for a kNN method for heart disease diagnosis. (Alberto Palacios Pawlovsky), IEEE, 2018
Description-: This paper introduces an ensemble based on distances for a kNN (k Nearest Neighbor) method and shows results of its application to heart disease diagnosis.
Limitation-: Ensemble only gives an average accuracy of nearly 85% for any of the configurations and versions tested with the UCI heart disease Cleveland data set.

Name-: Predictive Analytics in Health Care Using Machine Learning Tools and Techniques.
(B. Nithya, Dr.V.Ilango), ICICCS, 2017
Description-: It offers a variety of alerting and risk management decision support tools, targeted at improving patient's safety and healthcare quality and disease predictions **Limitation-:** Massive amounts of heterogeneous, distributed, diverse, highly dynamic data sets and increasingly large amounts of unstructured and non-standardized information with respect to varied types of cancers

Name-: Non-invasive method for bronchopulmonary diseases diagnosis in patients of all ages based on the microwave technologies.(Ivan V. Semernik, Alexander V. Dem'yanenko, Olga E. Semernik, Alexander A. Lebedenko), IEEE , 2017
Description-: In this paper the method of bronchial asthma diagnostics based on analysis of microwave band electromagnetic radiation propagation through the patient chest is described. The suggested method allows realizing innocuous non-invasive diagnostics of respirator system diseases among patients of all ages. It also allows monitoring the patient's condition and the disease progression during the whole period of treatment.
Limitation-: The disadvantages of the suggested method are absence of necessity of doing breathing

manoeuvres by the patient, absence of non-influence on the patient etc.

Name-: The Ethical Challenges of Applying Machine Learning and Artificial Intelligence in Cancer Care.(Rima Hajjo),IEEE,2018
Description-: This article examines the ethical issues of applying ML and AI in cancer care and classifies them into three major categories: bias, the societal implementation of the technology, and the effects of big data analytics on cancer patients.
Limitation-: Algorithms trained on data sets with these characteristics are adopted in healthcare; they have the potential to exacerbate health disparities.

Proposed Methodology

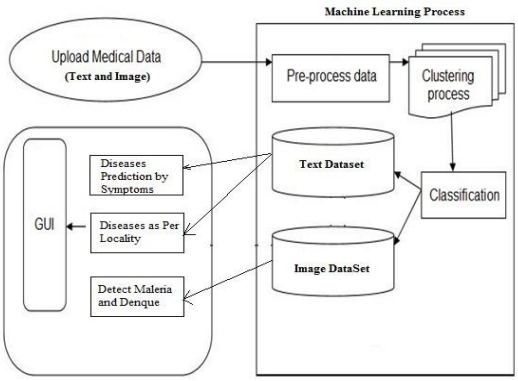

Fig 1. System architecture

The people are suffering from the many viral diseases like Dengue, Malaria. This information is collected from the various hospitals and the analysis of data is done and prediction of some diseases can be made. This system gives the prediction as per locality of the area.

Description:
Module 1:
In this system we detect the malaria and dengue diseases based blood cell dataset and apply image processing with the use of machine learning technique.
Module 2:
After that we collect the patient data from every hospital in particular area for identify the exactly which area disease will spread out more based on clustering algorithm.
Module 3:
Here provide the module for detecting diseases based on symptoms. **A. Algorithm:**
CNN is the basic algorithm used in the project which is as follows,
1. Classify dataset under labeled folders with blood samples images as CNN is supervised algorithm
2. Read dataset and prepare dataset in one file as pickle or numpy.
3. Read features of all images and label (here name of dataset folder) of it using following functions,
    a. Conv2D
    b. Maxpool2D
    c. Relu activation for layers
    d. Sigmoid activation for dense layer
    e. Binary Crossentropy for loss calculation
4. Store it in model file
5. Get input image
6. Read features of input image
7. Compare features of stored features
8. Show label as prediction of nearly matched features.

Now days, to analyzation of the imaginary data in deep learning, most important and frequently used algorithm is Convolutional neural network (CNN, or ConvNet). For the minimal preprocessing of a data CNN have different types of the variations. Along with the translation invariance characteristics and the architecture related with the shared-weight, CNN also known as shift invariant or space invariant artificial neural networks (SIANN). Along with the animal visual cortex, the connectivity between the neurons resembles the organization; networks were inspired by biological processes. Only in restricted visual fields, Respond to stimuli of Individual cortical neurons is known as the receptive field. The receptive fields of various neurons part overlap specified they cowl the whole field of regard. CNN's use comparatively very little preprocessing compared to alternative image classification algorithms. This implies that the network learns the filters that in ancient algorithms were handengineered. This independence from previous information and human effort in a feature style may be a major advantage. They need applications in image and video recognition, recommended systems, image classification, medical image analysis, and linguistic communication process. A CNN consists of associate input associated an output layer, likewise as multiple hidden layers. The hidden layers of a CNN generally encompass convolutional layers, pooling layers, totally connected layers, and social control layers.

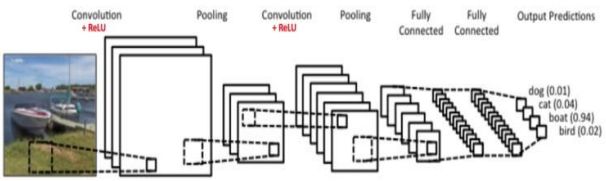

Fig 2. Simple ConvNet

The Convolutional Neural Network in Fig. is similar in architecture to the original LeNet and classifies an input image into four categories: dog, cat, boat or bird. There are four main operations in the ConvNet shown in fig. above:
1. Convolution
2. Non Linearity (ReLU)
3. Pooling or Sub Sampling
4. Classification (Fully Connected Layer)

An Image could be a matrix of picture element values. Basically, each image will be diagrammatic as a matrix

of picture element worth Channel could be a typical term won't to talk to a precise element of a picture. A picture from a customary photographic camera can have 3 channels – red, inexperienced and blue – you'll imagine those as 3 2d-matrices stacked over one another (one for every color), every having picture element values within the vary zero to 255. After training with CNN on blood cell images dataset, a model file with weights will be generated and will be used for classification or prediction of presence of parasites of malaria or dengue. If result is positive then we will be using image processing techniques such as colour heat map and contours to predict the perecentage of severity of parasites in blood cells. The overall system will perform as stated in below algorithm

Overall Algorithms of the System:

1. Start
2. Input Image of blood cell
3. Input 3 symptoms
4. Map image with trained model of CNN
5. Map symptoms with trained symptoms
6. Show results
7. Predict percentage of infection if present in cell images
8. Show severity percentage

B. Mathematical Model:

Let 'S' be the system

- Where,
  - S= {I, O, P, Fs, Ss} o Where,
  - I = Set of input Set of output o P = Set of technical processes o Fs = Set of Failure state
  - Ss = Set of Success state

- Identify the input data I1, I2, , In

I = {(Input Data (Text, Image), Dataset (Dengue, Malaria))}

- Identify the output applications as O1, O2,,On

{(Malaria Detection, Dengue Detection)}

- Identify the Process as P

P = {(Image pre-processing, Image Processing, Grey-scale, smoothing, Edging, segmentation, feature extraction, classification, show result)}

- Identify the Failure state as Fs

Fs = {(If data set not loaded, If not predicted, if more time required to predict}

- Identify the Success state as Ss

P = {(Correct prediction within time)}

Result and Discussion

In the proposed system, we will be using supervised CNN approach is used to predict the results from images as well as text dataset is used for symptoms classification. CNN gives accuracy than other algorithms. Also text dataset and symptoms classification works in compliment to the CNN to get more precise results.

Comparative results of existing and proposed system is as follow,

Table 1: Comparative Results

Parameters	Existing System	Proposed System
Image Dataset	Somewhat	Yes
Text Dataset	Somewhat	Yes
Symptoms Classification	Many Diseases	Focused on Dengue and Maleria
Prediction with Images and Symptoms	No	Yes
New Symptoms or Image Addition	No	Yes
Deep Learning	Somewhat	Yes
Execution Requirements	Mostly Heavy with Softwares like Matlab	Lightweight with Python, OpenCV and Tensorflow
Time of prediction	More	Less
Accuracy	Less	High

With reference to Table 1 it is clear that we overcome various problems in existing system and our approach works efficiently.

Conclusion

A robust and novel method by using machine learning for diagnoses malaria and dengue has been implemented in this paper. By using this method we obtain the less than 60 seconds time to give a diagnosis as compared to other clinical laboratories. The prediction algorithm is design to predict the area in danger zone of particular disease by considering the locality from the database to calculate the results.The results have to be the same as the Python output, as well as keeping to an acceptable processing speed and duration. The research will focus on the benefits it can provide for the successful diagnosis of malaria, dengue and the supportive treatment.The system prediction is very important in the awareness about the viral diseasespreading in the locality as people get alerted by the system about any particular disease so they take precautions about that.

References

[1]. Xiaoli Wang , Yuan Wang, Chuchu Gao, Kunhui lin, and yadili "Automatic Diagnosis With Efficient Medical Case Searching Based on Evolving Graphs",IEEE,2018.

[2]. Stavros Pitoglou , Yiannis Koumpouros and Athanasios Anastasiou "Using Electronic Health Records and Machine

[3]. Learning to Make Medical Related Predictions from NonMedical Data" ,International Conference on Machine Learning and Data Engineering,2018

[4]. Prajwal Shimpi, Sanskruti Shah, Maitri Shro_ , Anand Godbole "An Artificial Neural Network approach for classification of Vector-Borne diseases.",(ICEEOT) 2018.

[5]. C.Sowmiya;P.Sumitra "Analytical study of heart disease diagnosis using classification Techniques",International Conference on Intelligent Techniques in Control, Optimization and Signal Processing,2017.
[6]. Sourabh Dudakiya, Heren Galani, Suvarna E. Pawar "Monitoring Mobile Patients Using Predictive Analysis By Data From Wearable Sensors", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) 2018.
[7]. Alberto Palacios Pawlovsky "An ensemble based on distances for a kNN method for heart disease diagnosis", IEEE, 2018
[8]. Rima Hajjo "The Ethical Challenges of Applying Machine Learning and Artificial Intelligence in Cancer Care",IEEE,2018
[9]. Dengue guidelines for diagnosis, treatment, prevention and control, A joint publication of the World Health Organization (WHO) and the Special Programme for Research and Training in Tropical Diseases (TDR), 2009.
[10]. Handbook for clinical management of dengue; World Health Organization, 2012.
[11]. Mathur, A., Tripathi, A.S. and Kuse, M. Scalable system for classification of white blood cells from Leishman stained blood stain images. Journal of pathology informatics.4(Suppl) (2013).
[12]. Nayak, S.K. and Sampathila, N. Development of a protocol for screening Leukemia from the microscopic images acquired from blood smear. International Conference on Signal, Power, Communication, Security, and Computing Applications (ICSPCSCA-2016), College of Engineering Vatakara, Kozikodu, Kerala, 2016, 62-64.
[13]. Tulsani, H., Saxena, S. and Yadav, N. Segmentation using morphological watershed transformation for counting blood cells. IJCAIT 2 (3) (2013) 28-36.