International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

A SURVEY ON BLOOD DISEASE DETECTION USING MACHINE LEARNING

Senthil Kumar V[1], Ragul S S[2], Ramya Shree[3], Agustus A[4]

[1]*Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology [anonymous], Coimbatore, Tamilnadu, India*

[2-4] *Department of Computer Science and Engineering, Kumaraguru College of Technology [anonymous], Coimbatore, Tamilnadu, India*

Abstract— *Technology is changing the world to live in easy way, but not all people in the world experiencing technology. To them this world is still giving more complex problems. We believe this proposal will make them to live easier. Now a days, people are spending lot of money in laboratories and in hospitals. In laboratories, they are using high level software which requires highly configured and specified PCs, hardware. Some laboratories in towns and village does not have PCS with high specification. This project will be used by entry level PCs. By running .py, the user will be taken to local host app UI and then one will promptly enter the required values to predict blood deficiency disease (required fields varies depends on diseases) And if client want to view and use the project through online mode they can access the endpoints provided by Streamlit. And the machine learning model will classify the disease and show result with nutrition chart .Using machine learning with python, to create model and streamlit for User Interface.*

Keywords—Model selection, Random forest, Gaussian, SVC.

1. INTRODUCTION

Machine learning has undergone vital development over the past decade and is already used with success in several bright applications covering a good array of knowledge connected issues. One in all the foremost interesting queries is whether or not it will be with success applied to the sphere of medical medicine and what reasonably knowledge is required. Laboratory tests area unit familiar ensure, exclude, classify or monitor diseases and to guide treatment. However, verity power of laboratory check results is often underestimated, since clinical laboratories tend to report check results as individual numerical or categorical values, with physicians concentrating primarily on those values that fall outside a given reference vary.

2. PROJECT OBJECTIVE

The objective of the project is to classify and predict diseases using machine learning algorithms with the help of blood sample results provided by the laboratory**.** The most important questions is how this technologies and software can be applied to medical field successfully.Blood data analysis, only machine learning classification algorithms were used.
They were as follows:

- Gaussian
- Random forest.
- Support vector classification (SVC).

A. Gaussian classification:

Gaussian Processes square measure a generalization of the mathematician chance distribution and may be used because the basis for stylish non-parametric machine learning algorithms for classification and regression. They're a kind of kernel model, like SVMs, and in contrast to SVMs, they're capable of predicting extremely tag category membership chances, though the selection and configuration of the kernel used at the center of the strategy will be difficult. Since mathematician works well for

multiclass classification we are able to think about alternative algorithms random forest, SVC will be tried. to look at the comparison table confer with fig.2

B .Random forest:

A random Forest could be a common machine learning algorithmic rule that belongs to the supervised learning technique. It will be used for each Classification and Regression issues in mil. it's supported the construct of ensemble learning, that could be a method of mixing multiple classifiers to resolve a fancy downside and to boost the performance of the model. Random Forest could be a classifier that contains variety of call trees on numerous subsets of the given dataset and takes the common to boost the prophetic accuracy of that dataset. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting, to classify whether the patient has leukemia or not, SVC model was used. The problem with SVC was over fitting, the result accuracy was 1.0. So to overcome over fitting problem, random forest classifier was used to classify patients with anemia. Initially Patient gender, age and hemoglobin count is considered. The accuracy percent observed was 93%.

C. Support vector classification (SVC):

Support vector machines (SVMs) area unit powerful nevertheless versatile supervised machine learning strategies used for classification, regression, and, outliers' detection. SVMs are very efficient in high dimensional spaces and generally are used in classification problems. SVMs are popular and memory efficient because they use a subset of training points in the decision function. For anemia classification, the hemoglobin column from the dataset consisting of 11 columns were taken to classify the patients whether they have anemia or not. The accuracy observed in SVC is 98%. Plotting confusion matrix, result is [[51 1] [1 68]] (ref fig.1) which denotes true positive, true negative, false positive and false negative values. For the same model linear classification was performed in which very less accuracy was obtained.

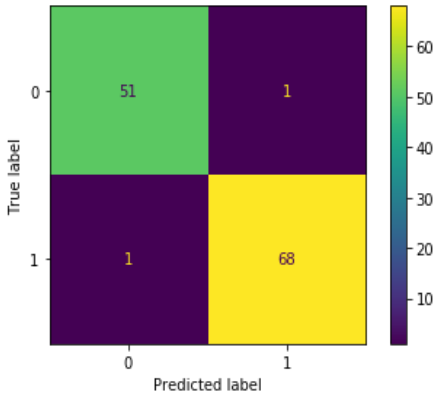

Fig-1.1-Confusion matrix-SVC

Comparison Table for anemia:

Algorithms used	Accuracy found
Gaussian	94%
Random forest	93%
SupportVector Classification	98%

Table-1.1

3. DEPLOYEMENT OF MODELS

Streamlit is a open source platform that data science team and machine learning team can use to deploy the model and we used the Streamlit to deploy our model here. In first step of deployment we converted our models which is anaemia model and leukaemia model into pickle file and stored our pickle file in a corresponding repository and we should have the Python file in same repository so they can access our pickle file easily or you can specify the file path of the pickle file directly while loading the pickle file in Python file. On the next step the Python file will have Streamlit code which we can create our user interface so that the client can access our model and use the model through this UI which is created using Streamlit and python. Nowadays creating a model is bit easier which means many people familiar with that but deploying the model is bit complicated so we chose Streamlit to deploy the model correctly .So, we can use our model efficiently for this disease detection. By using this deployment method, people can use this project in local and through Internet also.

And nutrition chart is also displayed when if the diagnosis shows positive result. And we are providing nutrition chart in weekly basis so that patient can check the nutrition chart next week and they can see the next result. And analyse if they are having any improvement in their diagnosis compared to previous week.

4. END USER BENEFITS:

a) People can use this blood deficiency detection through both online and offline mode.(ref fig.2)

b) Computers or laptops with low specification can run this project (locally).

c) This can be used or viewed through mobile phones too as website.

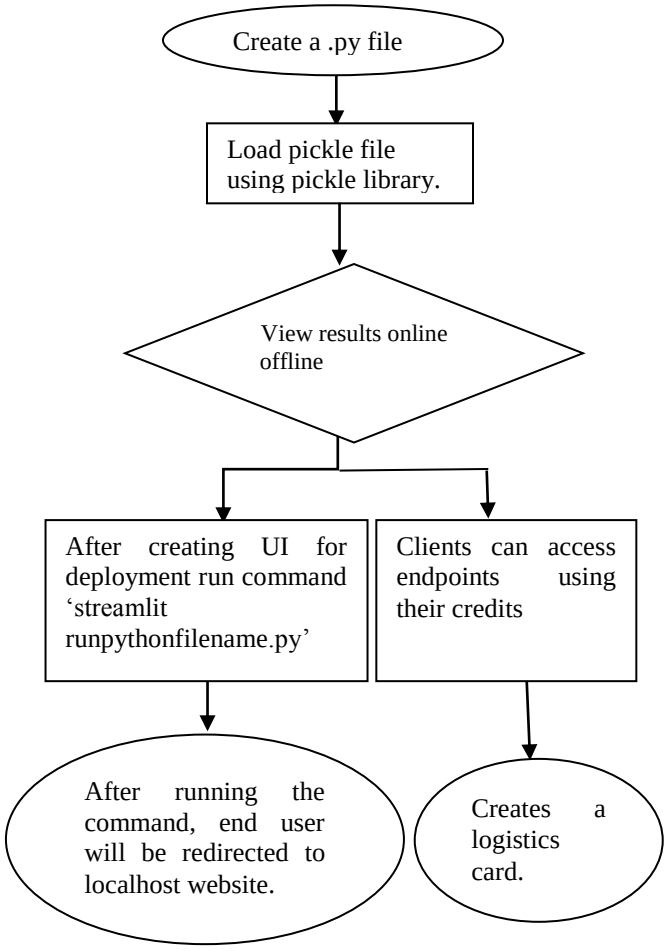

Fig-1.2-Flow chart for end user online/offline

5. INFERENCES:

The surveys and papers mentioned in references really gave more accurate information related to this blood deficiency detection. A survey on blood disease detection using classical machine learning by Lewontin, Richard C gave us conclusion for selecting schema for dataset. An application of machine learning to haematological diagnosis by Jordan, M. I. & Mitchell provide information to deal and detect more than one disease using very minimal amount of data. Machine Learning in Detection and Classification of Leukaemia Using Smear Blood Images: A Systematic Review,25Jun 2021 , in this paper they used images as input to predict leukaemia disease result, and it helped to predict Leukaemia using normal data. Department of Haematology and Blood Banking, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. From this paper we obtained some major terms related to blood diseases like Haemoglobin, platelets. And we referred more papers to achieve implementing this blood disease detection web app using Streamlit

6. CONCLUSION:

Detection of blood deficiency using laboratory results without doctor intervention could be time saving for the patient as well as doctors. In this way most essential/ emergency cases could be handled by the doctors to save patients. The model built is to classify patients who had anaemia and leukemia. In future, predictions for few other diseases/deficiency related to blood will

be done using machine learning models. At the end, the models will be uploaded in streamlit to display the output when input is being given. It will help the hospitals to get result within fraction of seconds. Food chart for deficient components in the blood will be displayed after classification.

7. REFERENCES:

[1] Blood diseases detection using classical machine learning algorithms (2019) Lewontin, Richard C. It ain't necessarily so: The dream of the human genome and other illusions. New York Review of Books, 2001. Feldman, Eric A., Eric Feldman, and Ronald Bayer, eds. Blood feuds: AIDS, blood, and the politics of medical disaster. Oxford University

[2]An application of machine learning to haematological diagnosis Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. Science 349, 255–260 Luo, Y., Szolovits, P., Dighe, A. S. & Baron, J. M. Using Machine Learning to Predict Laboratory Test Results. American journal of clinical pathology 145, 778–788

[3]Development of Machine Learning Model for Diagnostic Disease Prediction Based on Laboratory Tests- April 16th, 2021 Esteva A, et al. A guide to deep learning in healthcare. Nat. Med. 2019; 25:24–29. doi: 10.1038/s41591-018-0316-z Zhang Y, et al. Computational analysis and prediction of lysine malonylation sites by exploiting informative features in an integrative machine-learning framework. Brief. Bio inform.

[4]Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review,25Jun 2021 Mustafa Ghaderzadeh ,1 Farkhondeh Asadi ,1 Azamossadat Hosseini ,1 Davood Bashash ,2 Hassan Abolghasemi ,3 and Arash Roshanpour4

1Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2Department of Hematology and Blood Banking, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3Pediatric Congenital Hematologic Disorders Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

4Department of Computer Science, Sama Technical and Vocational Training College, Tehran Branch (Tehran), Islamic Azad University (IAU), Tehran, Iran

[5] Automatic Pattern Recognition for Detection of Disease from Blood Drop Stain Obtained with Microfluidic Device Guthrie, R., Susi, A.: A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. Pediatrics 32(3), 338–343

[6] Parkinson's Disease Detection And Classification Using Machine Learning And Deep Learning Algorithms– A Survey

[7] Deep Learning Based Automatic Malaria Parasite Detection from Blood Smear and Its Smartphone Based Application

[8] Laboratory diagnosis of thalassemia V. Brancaleoni,E. Di Pierro,I. Motta,M. D. Cappellini

[9] Stroke Disease Detection and Prediction Using Robust Learning Approaches Tahia Tazin ,1 Md Nur Alam,1 Nahian Nakiba Dola,1 Mohammad Sajibul Bari,1 Sami Bourouis ,2 and Mohammad Monirujjaman Khan 1

1Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka 1229, Bangladesh

2Department of Information Technology, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

[10] Effective Heart Disease Detection Based on Quantitative Computerized Traditional Chinese Medicine Using Representation Based Classifiers Ting Shu,1 Bob Zhang ,1 and Yuan Yan Tang 1 1Department of Computer and Information Science, University of Macau, Taipa, Macau

[11] Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review

[12] Disease Prediction From Various Symptoms Using Machine Learning

[13]Web Based Disease Detection System,Authors : Hasnuhana Mazumder, Sayantan Saha, Argha Roy Chowdhuri, Anindita Dey

[14] Classification of Diseases Using Machine Learning Algorithms: A Comparative Study by Marco-Antonio Moreno-Ibarra,YennyVilluendas-Rey

[15] Deep Learning Methods for Predicting Disease Status Using GenomicData.

[16] Use of artificial intelligence in infectious disease Said Agrebia and Anis Larbi

[17] Early and accurate detection and diagnosis of heart disease using intelligent computational model

[18] Efficient Automated Disease Diagnosis Using Machine Learning Models Naresh Kumar ,1 Nripendra Narayan Das,2 Deepali Gupta ,3 Kamali Gupta,3 and Jatin Bindra1

[19] Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction Milad Mirbabaie, Stefan Stieglitz & Nicholas R. J. Frick

[20] Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development Authors:Arkadip Raya,Avijit Kumar Chaudhurib