Machine Learning Based Heart Disease Prediction

T. Divya[1], K. Meenatchi[2]

[1]M.C.A, A.R.J College of Engineering and Technology, Mannargudi.
[2]M.C.A., M.Phil., A.R.J College of Engineering and Technology, Mannargudi.

Abstract
Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you're born with (congenital heart defects), among others. According to World Health Organization (WHO), cardiovascular disease (CVD) is one of the lethal diseases leads to the most number of deaths worldwide. Cardiovascular disease prediction aids practitioners in making more accurate health decisions for their patients. Early detection can aid people in making lifestyle changes and, if necessary, ensuring effective medical care. Machine learning (ML) is a plausible option for reducing and understanding heart symptoms of disease. This project proposes a Support Vector Machine (SVM) technique as the backbone of computer-aided diagnostic tools for more accurately forecasting heart disease risk levels. SVM modelling is a promising classification approach for predicting medication adherence in CVD patients. This predictive model helps stratify the patients so that evidence-based decisions can be made and patients managed appropriately. The chi-square statistical test is performed to select specific attributes from the Cleveland heart disease (HD) dataset. The data visualization has been generated to illustrate the relationship between the features. According to the findings of the experiments, the random forest algorithm achieves 88.5% accuracy during validation for 303 data instances with 13 selected features of the Cleveland HD dataset.

I. INTRODUCTION

The cardiovascular system is sometimes called the blood-vascular, or simply the circulatory, system. It consists of the heart, which is a muscular pumping device, and a closed system of vessels called arteries, veins, and capillaries. As the name implies, blood contained in the circulatory system is pumped by the heart around a closed circle or circuit of vessels as it passes again and again through the various "circulations" of the body. As in the adult, survival of the circulation of blood to maintain homeostasis and a favourable cellular environment. In response to this need, the cardiovascular system makes its appearance early in development and reaches a functional state long before any other major organ system. Incredible as it seems, the primitive heart begins to beat regularly early in the fourth week following fertilization. The vital role of the cardiovascular system in maintaining homeostasis depends on the continuous and controlled movement of blood through the thousands of miles of capillaries that permeate every tissue and reach every cell in the body. It is in the microscopic capillaries that blood performs its ultimate transport function. Nutrients and other essential materials pass from capillary blood into fluids surrounding the cells as waste products are removed. The cardiovascular system is sometimes called the blood-vascular, or simply the circulatory, system. It consists of the heart, which is a muscular pumping device.

II. MACHINE LEARNING

Machine learning is a branch of AI. Other tools for reaching AI include rule-based engines, evolutionary algorithms, and Bayesian statistics. While many early AI programs, like IBM's Deep Blue, which defeated Garry Kasparov in chess in 1997, were rule-based and dependent on human programming, machine learning is a tool through which computers have the ability to teach themselves, and set their own rules. In 2016,

Google's DeepMind beat the world champion in Go by using machine learning–training itself on a large data set of expert moves. Machine Learning algorithm is trained using a training data set to create a model. When new input data is introduced to the ML algorithm, it makes a prediction on the basis of the model.

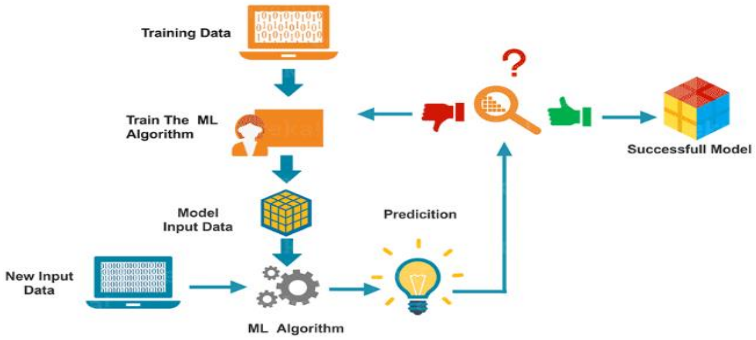

Fig.2a Machine Learning

The prediction is evaluated for accuracy and if the accuracy is acceptable, the Machine Learning algorithm is deployed. If the accuracy is not acceptable, the Machine Learning algorithm is trained again and again with an augmented training data set. This is just a very high-level example as there are many factors and other steps involved.

III. LITERATURE SURVEY

The proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms. It developed heart disease prediction using KStar, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than KStar, Multilayer perception and J48 techniques using kfold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory. Therefore, the accuracy's performance is improved more to give better decision to diagnosis disease.

It focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy. The recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms.

IV. SOFTWARE DESCRIPTION

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL). This tutorial gives enough understanding on Python programming language. Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages. Python is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development

Domain. Pandas is mainly used for data analysis and associated manipulation of tabular data in Data frames. Pandas allows importing data from various file formats such as comma-separated values, JSON, Parquet, SQL database tables or queries, and Microsoft Excel. Pandas allows various data manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features. The development of pandas introduced into Python many comparable features of working with Data frames that were established in the R programming language. The panda's library is built upon another library NumPy, which is oriented to efficiently working with arrays instead of the features of working on Data frames. NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

References

[1] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.

[2] Marjia Sultana, Afrin Haider, "Heart Disease Prediction using WEKA tool and 10-Fold cross-validation", The Institute of Electrical and Electronics Engineers, March 2017.

[3] Dr.S.Seema Shedole, Kumari Deepika, "Predictive analytics to prevent and control chronic disease", https://www.researchgate.net/punlication/316530782, January 2016.

[4] Ashok kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation", Springer, 17 September 2016.

[5] Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining Techniques", Orient J. Computer Science Technology, vol.6 2017, pp.457-466.

[6] Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018. [7] R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018.

[8] Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, "Heart disease Prediction using Machine Learning and Data mining Technique", March 2017.

[9] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System", 2016, pp.962-969.