ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 1, 2024

NAVIGATING HEALTH DATA: BUILDING PREDICTIVE MODELS FOR DISEASE DIAGNOSIS WITH MACHINE LEARNING

Mrs. V .Kalyani ,Assistant Professor, Department Of IOT, SICET, Hyderabad
Varkur Anwesha Reddy, Rachamalla Saisheran, Kottam Meghana,
Surakanti Dharma Teja
Department Of IOT, SICET, Hyderabad

ABSTRACT

This project is an attempt to help one to predict the disease he/she is having through the symptoms and the correct readings of the bodily vitals needed. There are times when people keep on ignoring health issues due to high medical fees. This may lead to severe issues later and even death. If not covered by insurance, medical bills can be a menace. This website is an approach in reducing the effort of a normal person by estimating the kind of disease one has and its severity. We have designed a disease prediction system using multiple machine learning algorithms. Based on the symptoms, age, and gender of an individual, the diagnosis system gives the output giving the information about whether the user is suffering from that particular disease or not. According to the severity, some diet plans and some exercises which can minimize the effects of the disease to some extent are also provided. It provides a simple yet effective approach for predicting the disease, if the provided values of vitals are accurate. The user will experience a simple yet effective User Interface and pleasing design.

Keywords: Disease Prediction, Machine Learning Algorithm, Supervised Learning, Diagnosis System, User Interface.

I. INTRODUCTION

Human life is evolving every single day, but is the health of the generation improving or declining? Life is full of uncertainty. Every now and then we come across many people suffering from fatal health issues due to late identification of diseases. The study says, One in two Indian diabetics are unaware of their condition. Nearly 463 million people in the world have diabetes. One in four deaths in India are now because of CVDs with ischemic heart disease and stroke responsible for more than 80% of this burden. The study estimates more than 50 million people in the world, considering the adult population, would be affected with chronic liver disease. But, it can be prevented by identifying the disease in its early stage. The project "Disease Prediction using Machine Learning" is developed to identify general disease in earlier stages. Now-a-days, people put health as a secondary priority, which leads to various problems. According to research, 40% of people ignore the symptoms, due to fear of facing financial issues or other generic reasons. Many cannot afford to consult a doctor or some are very busy and have a tight schedule, but ignoring the recurring symptoms for a long period of time may have severe consequences to their health. According to research 70% of people in India suffer from common diseases and the mortality rate is 25%, mostly due to ignorance in early stages. The main motive to develop this project is that a user can conveniently have a check-up of their health, if they have any of the symptoms.

Due to an increased amount of data growth in the medical and healthcare field the accurate analysis on medical data which has been benefited from early patient care. With the help of disease data, data mining finds hidden pattern information in the huge medical data of the data set.. We proposed a disease prediction platform, based on the vitals of the patient.Our DisEase web application predicts the occurrence of heart disease, diabetes & liver disease. We have also provided a proper diet plan according to the diseases. Along with it, we have also provided an about page which gives information about the symptoms & information about the diseases.

II. LITERATURE REVIEW

In this paper [1] the author has presented, "Disease prediction using Machine Learning over Big Data". Big data is the fastest concept in the current trend, so this concept is applied in more fields. Big data is most widely used in every field because it is very large. Big data is applied in the medical field. Both sides develop better growth in both fields, that is, big data is applied in medical fields and the medical fields at the same time increases the growth in the big data field. Big data helps to achieve better growth in the medical and health care sectors. It

additionally, provides more merits gives, (i) medical data analysis with accuracy, (ii) early prediction for disease, (iii) patient-oriented data with accuracy, (iv) The medical data, is securely stored and used in many places, (v) incomplete regional data are reduced and give the accurate result. The goal of the concept is to choose the region and collect the hospital data or medical data of the particular selected region, this process is using the machine learning algorithm. Then, finding the missing data based on latent factors gets the incomplete data and it is reduced. The previous system uses the CNN-UDRP (Unimodal unwellness Risk Prediction), then endlessly implements consequent level victimization the CNN-MDRP (Multimodal unwellness Risk Prediction). The CNN-MDRP overcomes the drawback of CNN-UDRP. The CNN-MDRP uses the hospital data, that is structured and unstructured data. The CNN-MDRP algorithm based prediction is produced more accurately, this accuracy is compared with previous systems. The advantages of the concept is, better feature description and better accuracy, and the disadvantages of this system is, this feature is only applicable for the structured data so it is not good in disease description.

Authors in this paper [2] have proposed the concept of machine learning-based disease prediction using big data to overcome the machine learning drawbacks. The smooth progress of big data is moved in the biomedical and healthcare communities in hospitals for accurate results in any experiment result. This concept is (a) reduces the unfinished data and (b) effective disease prediction. The proposed concept is tested or experimented with real-life hospital data collections such as hospital-oriented information like daily updated data- doctor data prefer doctor details, patient data prefer patient details, disease data prefer disease-oriented data, etc. This technique overcomes main two difficulties in the existing system are, (i) incomplete data, (ii) missing data. To rebuild the latent factor model. The concept is to get the information from a hospital that collected information from a forum called “structured and unstructured data”, and by using the Machine Learning Decision Tree algorithm and Map Reduce (MR) algorithm. The MR algorithm is employed for data partitioning. It reaches 94.8% with the traditional speed but it's quicker than CNN-UDRP then , it reports the knowledge of disease occurrences possibilities.

The paper [3] author has presented the info mining concept “Disease Prediction by using Machine Learning”. The best growth of the stage is developing that technique into the healthcare basis, the data analysis is an important part of every field. Data mining predicts the information for healthcare is called rapid growth of the medical care field. The existing one is designed for the purpose of (i) analyzing, (ii) managing, (iii) predicting healthcare data, it is to describe the overall healthcare systems. The concept of machine learning is applied to disease-related information retrievals and the treatment processes in these types of processes are achieved by using data analysis. The predictions of outbreaks in diseases are using the decision tree because it is very effective. This concept-based experiment shows that the result is related to the disease symptoms so that data is described using a modified prediction model. If the concept chooses the training set like medical patient symptoms, then, use the decision tree, then, predict, finally give the symptoms of the patient and get the accurate result for disease prediction. This concept is only performed, that is it predicts only the patient-related information with low time and low cost.

Authors, presents the survey paper [4] for “prediction of disease using machine learning over big data”. Can develop the medical specialty basis this concept is applied to produce the medical data into mass medical data, which means the data which is enlarged. The goal of this concept is targeted: the simplest data is stored into the space of medical massive data analysis, called “medical data analysis in massive collection”. It produces the accuracy and it reaches the 4.8% speed faster than the CNN-UDRP. It only focuses on these three data, (a) structured data, (b) text data, (c) structured and text data. In this proposed system it improves the medical data oriented term.

In this paper [5] the author has presented, “personalized disease prediction care from harm using big data”, for healthcare analysis. This concept describes the medical field as a rich data industry because it holds healthcare records, also. The daily treatment records are increased every day, that is it includes the number of transactions, and the patient information is stored and retrieved from the database. The medical treatment records are updated every day because every day improves the patient's health improvements based on treatment. It gives the right solutions for various sorts of diseases. This system is to change medical records, which means manually noting every medical oriented record into the electronic record, that is, digitizing the

medical care. This technology is simply called, “e-healthcare”. The medical data is stored within the database. Big data methods and the logic are used to analyze statistical analytics. The proposed system is understood as, “disease recommendation system”, and this technique holds the specialized tool, this tool is creating the profile. The profile-making needs some information from the personalized persons, that's doctors, patients, etc. If entering the required field of the system, finally get the personalized model health profile, but this personalization includes a huge amount of profiling information and other data. This personalized profile is based on current treatment and any other treatment it takes, if it can use the same profile, cannot require the profiling again. It increased the computational time, so frame the time for clinical purposes. This concept is extracted and applies to applications like Collaborative Assessment and Recommendation Engine (CARE). The CARE analyzes the performance limitation, and it improves personalized disease prediction. The concept improved application CARE is classified into two types namely, (A) Current CARE Architecture, (B) Parallel CARE Architecture. The CARE expresses the overall performance for patient-oriented big data. It takes more time.

Enthusiastically presents the author, give the knowledge are collected by the paper [6] namely, “Use the Weighted Ensemble to Neural Network based Multimodal Disease Risk Prediction (WENN-MDRP) and have selection of Ant colony improved classifier for disease prediction over the large data concepts”. This concept of feature selection performance gives the dataset, this data set makes it one of the significant tasks. The feature selection task is split into levels by level, (i) first, structuring the normal and more explainable models, (ii) applying the concept knowledge and learning its performance, (iii) finally, ready to prepare the clean, that is clear the data. Then, the proposed concept analyzes the feature selection difficulties for giant data-based data analytics, resolving this complexity by using the Improved Ant Colony Optimization (IACO) technique. This technique early solves the missing data problem in incomplete data, which means it before setting the latent factor model, also. But it is not easy to select the best feature from the medical data. The second technique WENN-MDRP is called the unheard technique, in this technique, it helps to select the best features from medical data. These two methods are combined and give the special merit of improved prediction with accuracy if this accuracy is evaluated to compare the experimental techniques. This concept works only when the time is full to fill the needed instances like, (i) accuracy, (ii) precision, (iii) recall. It selects the best feasible, but not previously checked the possibility.

Authors, the survey Paper [7] for "Disease prediction in big data healthcare" using extended CNN. This concept is applied in the medical field to implement the hospital. It provides (i) high accuracy, (ii) high performance, (iii) high convergence speed. To select the particular region and then analyze the chronic diseases that hold the structured data (extracted useful features), the unstructured data is used the CNN technique, so automatically selects the features. The novel CNN proposed the medical data, and disease risk the model combines this data. The characteristic behavior of this system is selecting the data via the previous term. This term is previously applied if possible but not satisfied the disease changes, because disease level is not standard, it is changed every second. To take the selected data from a large number of data and improve the accuracy by using risk classification terms. The proposed system aims to predict the risk of liver-oriented disease. So, the hospital dataset is related to liver-oriented disease and it collects only the structured data from liver disease information. In the proposed system it uses disease risk modeling and gets the accuracy. But the risk prediction depends on the different features of medical data with higher accuracy.

This paper [8] author has presented big data techniques in public health like, “Terminology, Machine Learning, Privacy”. The digitized world is day by day, increasing the huge amount of data and increasing the data rate, so meet the staggering, but this problem is solved by entering the new and fantastic forum and a clever concept called the “Big Data”. Big data cannot lock the novel approach chances it understands public health. The concept is expressed step by step but it is very forced. Firstly, it takes the classification of sources of the data like big data, and then to clarify the terminology and then, identify its threads. The medical field-oriented research takes the big data, and it includes the protection, hypothesis-related generating research takes the big data information. The interpretability is not aimed at this proposed system using machine learning techniques.

This paper [9] author has presented the concept “Improving disease prediction by machine learning”, that is using machine learning and improving the disease prediction. Big data is expanding medical data, so improving

this type of information. This concept uses the genetic algorithm, it is utilizing the recovery data, that is the missing data, then, it dataset includes the medical data. This system uses the two calculation terms namely, (i) KNN, (ii) SVM. The chronic diseases increase the data CNN-MDRP technique uses the medical data. The database includes the medical data, and personal data, and detailed history of patients is stored. The RNN based techniques easily find out the logical data. This system uses online and offline methods.

In the paper [10] the author has presented the concept "Competitor Mining and Unstructured Dataset Handling Technique", which is used in healthcare communities. This paper competitive mining is described with its related works. Finally gave competitor mining algorithms with its advantages and drawbacks. This paper experimental result shows CMiner++ yielded least computation time when comparing others

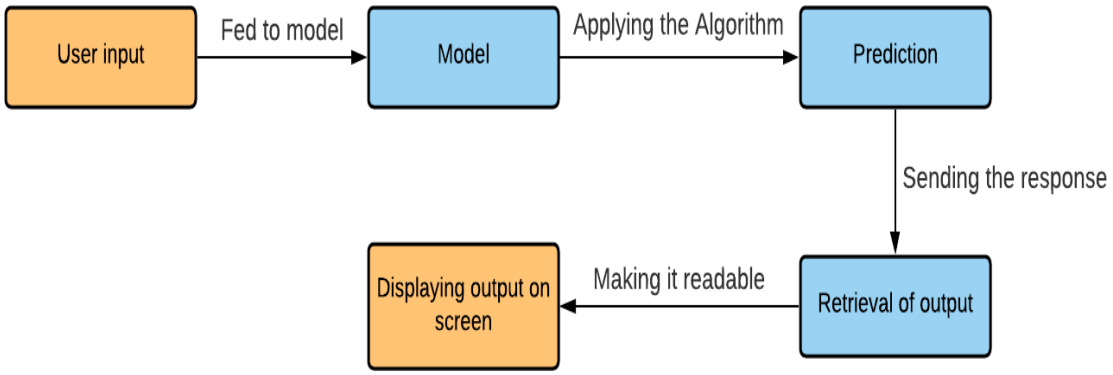
III. OBJECTIVES

- To predict the likelihood of contracting the disease.
- To give information about the diseases that are predicted.
- To provide the diet & exercise information
- To provide no expense disease diagnosis.

IV. METHODOLOGY

In this website, we have made it easy for the user to predict whether he/she has a particular disease or not. In the homepage, the brief overview of our website is shown, which will help the users to figure out the contents inside the website. There are three dropdown menus, for heart, liver and diabetes, each containing the "Predict", "About the disease", and "Exercise to follow" sections. In the "About the disease" section, information of the disease is given, and in the "Exercise to follow" section, the methods to prevent that particular disease is given. The "Predict" section contains the main highlight of our website. When the user enters this section, he/she is prompted to fill a series of input fields, based on which, our models intelligently predict whether that person has a particular disease or not. This is done by using the Machine Learning algorithms for Classification. We have used Gradient Boosting Classifiers for prediction of Diabetes and Random Forest Classifiers for prediction of Liver Disease and Heart Disease. These models are trained to have very high accuracy on the given dataset, which enables them to predict the outcome correctly for most part.

When the user enters the data, it is fed as an input to the respective model that we are using for that disease. Based on that input, the output is calculated, which is a binary number (0 or 1, where 0 indicates that the person is healthy and 1 indicates that the patient is suffering from the disease). This output is retrieved from the model, and is transferred into a statement of the patient being healthy or not, and is rendered on the screen. Along with the outcome, the graph is also displayed, which shows the trend of occurrence of the respective disease with respect to age group. Here is a brief outline of how our model works:



V. CONCLUSION

- In this way, with the help of AI, there can be early diagnosis of the patient.
- Boon to common people and sometimes, doctors.
- Saves time required for diagnosing the problem.

VI. REFERENCES

[1] Shraddha Subhash Shirsath, Prof. Shubhangi Patil Disease Prediction Using Machine Learn.Over Big Data". International Journal of Innovative Research in Science, Engineering and Technology, [2018]. ISSN (Online) : 2319-8753, ISSN (Print) : 2347-6710.

[2] Vinitha S, Sweetlin S, Vinusha H, Sajini S. "Disease Prediction Using Machine Learning Over Big Data". Computer Science & Engineering: An International Journal (CSEIJ), Vol.8, No.1, [2018]. DOI:10.5121/cseij.2018.8101.

[3] Sayali Ambekar and Dr.Rashmi Phalnikar. "Disease Prediction by using Machine Learning". International journal of computer engineering and applications, Volume XII, special issue, May 18. ISSN: 2321-3469.

[4] Lohith S Y, Dr. Mohamed Rafi. "Prediction of Disease Using Learning over Big Data - Survey". International Journal on Future Revolution in Computer Science & Communication Engineering. ISSN:2454-4248.

[5] J. Senthil Kumar, S. Appavu. "The Personalized Disease Prediction Care from Harm using Big Data Analytics in Healthcare". Indian Journal of Science and Technology, vol 9(8), DOI:10.17485/ijst/2016/v9i8/87846, [2016]. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.

[6] A Survey on Disease Prediction by Machine Learning over Big Data from Healthcare Communities International organization of Scientific Research 59 | Page

[7] Gakwaya Nkundimana Joel, S. Manju Priya. "Improved Ant Colony on Feature Selection and Weighted Ensemble to Neural Network Based Multimodal Disease Risk Prediction (WENN-MDRP) Classifier for Disease Prediction Over Big Data". International Journal of Engineering & Technology, 7(3.27) (2018) 56-61.

[8] Asadi Srinivasulu, S.Amrutha Valli, P.Hussain Khan, and P.Anitha. "A Survey on Disease Prediction in Big Data Healthcare using an Extended Convolutional Neural Network". National conference on Emerging Trends in information, management and Engineering Sciences, [2018].

[9] Stephen J.Mooney and Vikas Pejaver. "Big data in public health: Terminology, Machine Learning, and Privacy", Annual Review of Public Health [2018].

[10] Smriti Mukesh Singh, Dr. Dinesh B. Hanchate. "Improving Disease Prediction by Machine Learning". eISSN: 2395-0056, p-ISSN:2395-0072.

[11] Joseph, Nisha, and B. Senthil Kumar. "Top-K Competitor Trust Mining and Customer Behavior

[12] Investigation Using Data Mining Technique."Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.2 (2018).

[13] Kumar, B. Senthil. "Adaptive Personalized Clinical Decision Support System Using Effective Data Mining Algorithms." Journal of Network Communications and Emerging Technologies (JNCET) www.jncet. org 8.1 (2018).

[14] Unnikrishnan, Asha, and B. Senthil Kumar. "Biosearch: A Domain Specific Energy Efficient Query Processing and Search Optimization in Healthcare Search Engine." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.1 (2017).

[15] Kumar, B. Senthil. "Adaptive Personalized Clinical Decision Support System Using Effective Data Mining Algorithms." Journal of Network Communications and Emerging Technologies (JNCET) www.jncet. org 8.1 (2017).

[16] Kumar, B. Senthil. "Data Mining Methods and Techniques for Clinical Decision Support Systems." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 7.8 (2017).

[17] Sreejith, B. Senthil. "Identification of Diabetes Risk Using Machine Learning Approaches." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 7.8 (2017).

[18] Bhavitha Varma, B. Senthil. " A Different Type of Feature Selection Methods for Text Categorization on Imbalanced Data." Journal of Network Communications and Emerging Technologies (JNCET) www.jncet. org 8.1 (2017)