International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Survey on Machine Learning based E-Health System for Disease Prediction

Ankush Gadge[1], Vishal Awate[2], Nisha Bankar[3], Monika Biradar[4], Prof. S. V. Shinde[5]

[1-4]Science and computer department, Savitribai Bai Phule Pune University, Pune, Maharashtra, India

[5]PDEA's College of Engineering, Manjari BK, Pune, Maharashtra, India

Article Info
Volume 9, Issue 1
Page Number : 86-91

Publication Issue :
January-February-2022

Article History
Accepted : 20 Jan 2022
Published: 01 Feb 2022

ABSTRACT

The health reports of the people including diagnostics information and medical prescriptions are provided in the form of test-based case notes due to this the previous health conditions and the medicines used by the person are not known when they visit the hospital later. But storing all the health information of a person in the cloud as the soft copy reduces this problem. To achieve every hospital, dispensary, laboratory must have an internet connection for registration of patient's data, each patient will be identified by the unique Health ID and all the data of the patient will be stored in the cloud and the data can be accessed by only the particular patient. Accurate and on-time analysis of any health-related problem is important for the prevention and treatment of the illness. To diagnose the disease by accessing all information from linked Health ID with Machine Learning algorithm will boost the system in detection of diseases. Here the work presents review of previous researcher's techniques used for the prediction of diseases and number of parameters used.

Keywords : Cloud Computing, Disease Prediction, Health Id Generation, Machine Learning Algorithm.

I. INTRODUCTION

For the prevention and treatment of health problems, accurate and timely investigation of any health-related issue is critical. In the event of a critical illness, the standard method of diagnosis may not be adequate. The development of a medical diagnostic system based on machine learning (ML) algorithms for illness prediction may aid in a more accurate diagnosis than the current technique.

To create and build a specific framework for a smart health system based on the Internet of Things. The framework takes a tiered approach to addressing critical difficulties with IoT-based smart health systems, including a comprehensive data gathering method from the patient to cloud storage that can be accessed remotely. The numerous illnesses and their factors that are utilised to forecast diseases are described in Section A. Section B describes the

machine learning algorithms used. And the system's objectives are described in section C.

[1] Hypothesis behind selection of different parameters

The following clinical parameters were discovered for each of the diseases while observing the various aspects for the detection of heart disease, cancer, and diabetes throughout the process of monitoring the various features for the detection of these diseases:

Heart disease Parameters

Age, blood pressure, gender, smoking habit, any prevalent stroke, diabetes, total cholesterol level, systolic BP, diastolic BP, prevalent hyper-tension, body mass index, glucose levels, heart rate, etc.

Cancer Parameters

Age, air pollution, gender, alcohol use, occupational hazards, dust allergy, genetic risk, balanced diet, chronic lung disease, obesity, passive smoker, smoking, chest pain and coughing of blood, etc.

Diabetes Parameters

glucose before fasting, glucose anytime, age, gender, blood pressure, family member with diabetes past or present, bmi, pregnancies, percentage of occurring diabetes, polyphagia, polydipsia, weakness, polyuria, genital thrush, sudden weight loss, visual blurring, irritability, itching, delayed healing, glucose, pregnancies, blood pressure, insulin, skin thickness, diabetes pedigree function, etc.

Based on these characteristics, it is determined that some characteristics, such as BMI, glucose levels, genetic history, weight loss, smoking behaviours, and so on, are shared by a variety of diseases.

[2] Machine Learning Algorithms

RF [2021]: Random Forest is a low-classification-error ensemble classifier that comprises of numerous uncorrelated individual decision trees that function as a single set with the ability to handle big data sets, outliers, and noise in the data. Each tree generates a categorization that aids in improving accuracy. Due to the construction of multiple trees, this strategy may result in sluggish real-time prediction.

SVM [2021]: SVM is a supervised learning model that can do non-linear classification. This model employs a technique known as the kernel, which translates its inputs into high-dimensional feature spaces without explicitly doing so. With insufficient training data, SVM may identify illness.

KNN [2021] is a supervised ML technique that is used to solve classification issues. This model looks at the closest K neighbours to identify the class of new data. To determine the distance between two data, KNN employs a variety of distance functions, including Manhattan, Minkowski, and Euclidean.

[3] Objectives

I. Detail survey on disease prediction system by various researchers.

II. To diagnose the disease by accessing all information from linked Aadhar Card ID.

III. Use of Machine Learning algorithm for accurate diagnosis of disease.

IV. To provide privacy and security for all sensitive data.

V. Use of cloud database for 100 percentages of availability of data at any time.

II. LITERATURE REVIEW

A.U. Haq, et al (2018) proposed system that uses a database that contained information about patients having heart disease. Authors extracted features using three selection algorithms which are relief, minimum

redundancy, and maximum relevance (mRMR), and least absolute shrinkage and selection operator which was cross-verified by the K-fold method. The extracted features were sent to 6 different machine learning algorithms and then it was classified based on the presence or absence of heart disease. On the entire characteristics of the Cleveland heart disease dataset, the authors tested the performance of several machine learning techniques such as logistic regression, k-nearest neighbour, artificial neural network, support vector machine, Naive Bayes, and decision tree. In the second, we selected key features using the feature selection algorithms Relief, mRMR, and LASSO. The performance of the third classifiers was evaluated based on a set of features. Also employed was the k-fold cross-validation approach. Performance assessment measures were used to examine the performance of the classifiers. Before using classifiers, all characteristics were normalised and standardised.

Tülay Karaylan et al (2017) presented an artificial neural network back propagation algorithm-based heart disease prediction system. The neural network was fed 13 clinical characteristics as input, and then trained with the backpropagation algorithm to predict the presence or absence of heart disease with a 95% accuracy rate. There are input variables, which are disease risk factors derived from a dataset, and output variables, which are a category, such as "disease absence" and "disease presence," in the heart disease prediction system. The term "supervised learning challenge" refers to the difficulty of predicting cardiac disease. The prediction heart disease is "classification kind of supervised learning" because the output variables are of the category type. Heart disease prediction system was developed using the Backpropagation Algorithm, which is a standard ANN learning approach. The Backpropagation Algorithm is the best classification algorithm for heart disease prediction since it is the only technology that can handle nonlinear connections.

Michalis V. Karamouzis Dimitrios I. Fotiadis et al (2015) proposed A variety of ML techniques such as artificial neural networks (ANNs), bayesian networks (BNs), SVM, DT have been widely applied in cancer research for the development of predictive models, resulting in effective and accurate decision making.

Authors R. Delshi howsalya devi1 et al (2012) proposed a general disease prediction based on symptoms of the patient. They employed a RF algorithm for disease prediction, which provided accurate results. A collection of illness symptoms is necessary for disease prediction. The person's living habits and check-up information are taken into account in this general illness prognosis. We begin by downloading the illness dataset from the UCI machine learning website, which is organised as a disease list with symptoms. The dataset is then pre-processed for cleaning, which includes eliminating commas, punctuations, and white spaces. As a result, the dataset is utilised as a training dataset. After that, the feature was retrieved and chosen. Then, using classification methods such as the random forest algorithm, we categorise the data.

Authors Naipeng Dong et al. (2011) conducted a review of eHealth Challenges. Nowadays, medical records may be accessed at any time and from any location. In healthcare, the cloud computing paradigm promotes the exchange and integration of medical records. The cloud computing paradigm, on the other hand, has some advantages, but it also presents privacy and security risks to health data. To improve the degree of confidence between patients and healthcare, cloud service providers should address security risks in the cloud. Cloud Hybrid "Host" IaaS "Build" PaaS "Consume" SaaS inside the premises Intranet High-level Security Behind the iron curtain Internet access outside of the office Low Level of Security Outside the perimeter of the firewall Relationship between delivery and service models in the private and public clouds.

The use of cloud computing in e-health is explained by the authors Sahandi, et al (2009). The authors suggest a review on the usage of cloud computing in e-health. Electronic health records (EHRs), laboratory information systems, pharmaceutical information systems, and medical photographs may all be stored, managed, protected, shared, and archived using the cloud. Patients will get better treatment overall as a result of up-to-date health records and ongoing interactions between healthcare professionals. Aside from a lack of standards, legislation, and interoperability challenges, the biggest roadblocks to widespread cloud use by healthcare providers are security, confidentiality, and trust concerns.

Authors Sathya, et al (2018) provide a service model for patient comfort from a technical and economic standpoint, as well as open hurdles in applying IoT in the real world medical industry.

Study of Heart Disease Prediction System Using Data Mining Classification Techniques was suggested by authors Chaitrali S. et al (2012). Many hospitals currently handle healthcare data using healthcare information systems, according to them, since the systems include large amounts of data that may be utilised to extract hidden information for generating intelligent medical diagnoses. The major goal of this study is to create an Intelligent Heart Disease Prediction System that uses a historical heart database to diagnose heart disease. Medical words including sex, blood pressure, and cholesterol are employed as 13 input qualities in the development of this system. Two extra traits, obesity and smoking, are employed to produce more suitable findings, since they are regarded essential features for heart disease. The classification methods utilised in data mining include Neural Networks, Decision Trees, and Naive Bayes.

Authors for a full examination of the subject's health, Mostafa Haghi et al (2020) employed simultaneous monitoring. Diseases such as cardiovascular, lung cancer, and respiratory are influenced by air pollutants, which include poisonous, dangerous gases and specific materials. The dangers of being exposed to such contaminants have long been recognised and documented. The impact of air pollution on the brain and mental disease has been studied in new study (e.g., depression). The study found a link between individuals' psychological well-being and air pollution. To put it another way, greater levels of air pollution lower the amount of time a subject spends outside. As a result of reducing physical activity outdoors, psychological anguish worsens, including reduced exposure to sunshine and social isolation.

The authors Mohammadzadeh et al. (2020) conduct a survey on the use of wearable smart sensors to monitor patients' vital signs during epidemics. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) technique was used to conduct this systematic review. This approach was developed by Moher et al. and is one of the most effective ways for writers to conduct systematic reviews.

III. PROPOSED SYSTEM

Fig. 1 shows the proposed system for disease prediction.

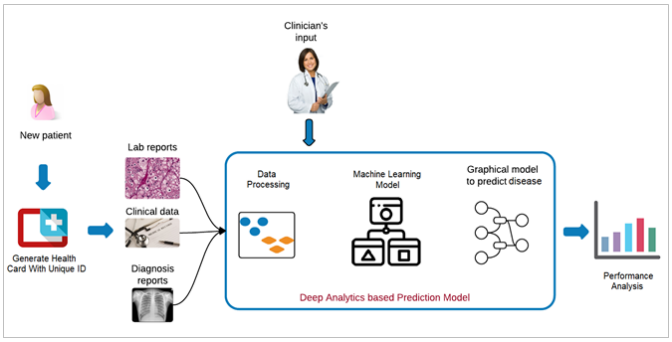


Figure 1: Disease prediction system model using machine learning model

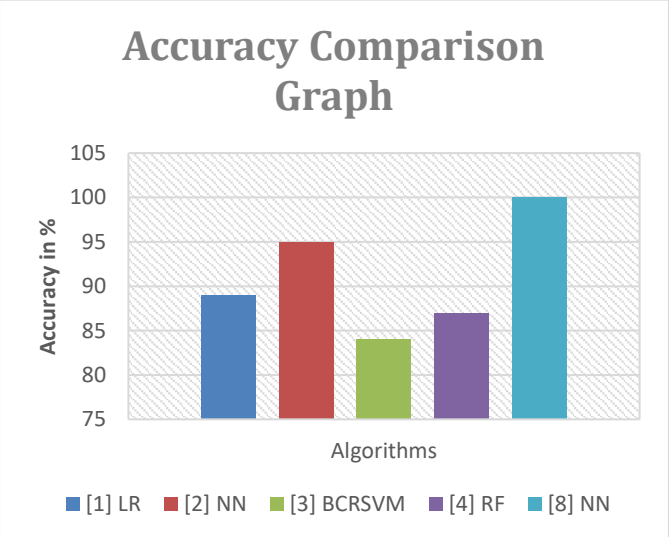
Initially Unique Health Id is generated for each patient and all data and reports can be search using that Health ID. The doctor may not be available always when needed. But, in the modern time scenario, according to

necessity one can always use this prediction system anytime.

The symptoms of an individual along with the age and gender can be given to the ML model to further process. After preliminary processing of the data, the ML model uses the current input, trains and tests the algorithm resulting in the predicted disease. The proposed system uses dataset consisting of gender, symptoms, and age of an individual was pre-processed and fed as an input to different ML algorithms for the prediction of the disease. The different ML models used for prediction of disease. The outcome of the models is the disease as per the symptoms, age, and gender is given to the processing model. Following are the important modules of proposed system.

IV. RESULT ANALYSIS

Table 1 shows the accuracy reading of algorithms used by researchers (Maintained in reference papers) [1-11]. Graph 1 shows the accuracy comparison graph of various algorithms and techniques used by researchers for disease prediction. From survey we can say that neural network (NN) based algorithms outperforms other ML algorithms in terms of accuracy.



Graph 1: Accuracy graph comparison

Table 1: Accuracy comparison of system.

	[1] LR	[2] NN	[3] SVM	[4] RF	[8] NN
ACC.	89	95	84	87	100

V. CONCLUSION

The work presents the detail review of predicting the disease based on the symptoms, age, and gender of an individual patient. The various Machine Learning, deep learning models are applied for disease prediction by researchers. As some models were dependent on the parameters, they couldn't predict the disease and the accuracy percentage was quite low. This work would help in lowering the cost required in dealing with the disease and would also improve the recovery process.

More complicated ML algorithms are required in future work to improve disease prediction. For greater performance, learning models should be adjusted more often after training. To minimise overfitting and improve model accuracy, datasets should be enlarged to include more demographics. Finally, better feature selection approaches should be utilised to improve learning model performance.

The system can be further expanded. It can use more number of input attributes. Other data mining / machine learning / deep learning / hybrid algorithms or transfer learning techniques can also be used for predication. Text mining may be used to extract unstructured data from healthcare databases.

VI. REFERENCES

[1]. A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, a hybrid intelligent system framework for the prediction of heart disease 2018.

[2]. T. Karayılan, O. Kılı¸c, in "Prediction of heart disease using neural network" 2017 International Conference on Computer Science and Engineering (UBMK) (IEEE, 2017), pp. 719–723

[3]. K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and structural biotechnology journal 13, 8 (2015).

[4]. R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm, 2012.

[5]. N. Dong, H. Jonker, and J. Pang, “Challenges in eHealth: from enabling to enforcing privacy,” in Foundations of Health Informatics Engineering and Systems. FHIES 2011. Lecture Notes in Computer Science, Z. Liu and A. Wassyng, Eds., pp. 195–206, Springer, 2011.

[6]. European Network Information Security Agency, An SME Perspective on Cloud Computing, European Network Information Security Agency, Heraklion, Greece, 2009.

[7]. Sathya, M. & Madhan, S. & Jayanthi, K. (2018). Internet of things (IoT) based health monitoring system and challenges. International Journal of Engineering and Technology (UAE). 7. 175-178. 10.14419/ijet. v7i1.7.10645.

[8]. Chaitrali S. Dangare and Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications, 2012, 44-48.

[9]. Seshadri Dhruv R., Davies Evan V., Harlow Ethan R., Hsu Jeffrey J., Knighton Shanina C., Walker Timothy A., Voos James E., Drummond Colin K. “Wearable Sensors for COVID-19: A Call to Action to Harness Our Digital Infrastructure for Remote Patient Monitoring and Virtual Assessments”, Frontiers in Digital Health,2020,10.3389/fdgth.2020.00008,2673-253X.

[10]. Mohammadzadeh, N., Gholamzadeh, M., Saeedi, S. et al. The application of wearable smart sensors for monitoring the vital signs of patients in epidemics: a systematic literature review. J Ambient Intell Human Comput (2020).

[11]. N. M. Yungaicela-Naula, C. Vargas-Rosales and J. A. Perez-Diaz, "SDN-Based Architecture for Transport and Application Layer DDoS Attack Detection by Using Machine and Deep Learning," in IEEE Access, vol. 9, pp. 108495-108512, 2021.

Cite this article as :

Ankush Gadge, Vishal Awate, Nisha Bankar, Monika Biradar, Prof. S. V. Shinde, " Survey on Machine Learning based E-Health System for Disease Prediction, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 9, Issue 1, pp.86-91, January-February-2022.
Journal URL : https://ijsrset.com/IJSRSET229134