International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

PATIENT DISEASES PREDICTION USING MACHINE LEARNING

Ms.Bharathi Krishna L[1], Ganesan P[2], Hari Krishna S[3] , Hariharan S[4]

Assistant Professor, Department of Computer Science and Engineering[1]
UG Scholar, Department of Computer Science and Engineering[2]
UG Scholar, Department of Computer Science and Engineering[3]
UG Scholar, Department of Computer Science and Engineering[4]
SRM Valliammai Engineering College, Chengalpattu, Tamil Nadu, India

---

Abstract -The wide adaptation of computer-based technology in the health care industry resulted in the accumulation of electronic data. Due to the substantial amounts of data, medical doctors are facing challenges to analyze symptoms accurately and identify diseases at an early stage. However, supervised machine learning (ML) algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and aiding medical experts in the early detection of high-risk diseases. In this literature, the aim is to recognize trends across various types of supervised ML models in disease detection through the examination of performance metrics. The most prominently discussed supervised ML algorithms were Naive Bayes (NB),Decision Trees (DT), K-Nearest Neighbor (KNN). As per findings, Support Vector Machine (SVM) is the most adequate at detecting kidney diseases and Parkinson's disease. The Logistic Regression(LR) performed highly at the prediction of heart diseases. Finally , Random Forest (RF), and Convolutional Neural Networks (CNN)predicted in precision breast diseases and common diseases , respectively.

INTRODUCTION

The emergence of Artificial Intelligence (AI) enabled computerized systems to perceive, think and operate in an intelligent manner like humans [1]. AI is a multi disciplinary concept of ML, Computer Vision, Deep Learning, and Natural Language Processing . ML algorithms apply various optimization, statistical ,and probabilistic techniques to learn from data that was generated from past experiences, and deploy it indecision making . These algorithms deemed to be applied in many disciplines including network intrusion recognition , customer purchase behavior detection, process manufacturing optimization, credit card fraud detection, and disease modulation. Many of these applications have been designed using the supervised learning approach. In this approach, datasets with known labels are induced to prediction models to predict unlabeled examples . This presents the hypothesis that medical doctors can utilize supervised learning as a powerful tool to conduct diseases diagnosis more efficiently.

EXISTING SYSTEM

The purpose of the problem statement is to identify the issue that is a concern and focus it in a way that allows it to be studied in a systematic way. It defines the problem and proposes a way to research a solution, or demonstrates why further information is needed in order for a solution to become possible. Prediction using traditional disease risk model usually involves a machine learning and supervised learning algorithm which uses training data with the labels for the training of the models. High-risk and Low-risk patient classification is done in groups test sets. But these models are only valuable in clinical situations and are widely studied. A system for sustainable health monitoring using smart clothing by Chen et.al. He thoroughly studied heterogeneous systems and was able to achieve the best results for cost minimization on the tree and simple path cases for heterogeneous systems.

1. PROPOSED SYSTEM

The aim of this study is to test the proposed hypothesis that supervised ML algorithms can improve health care by the accurate and early detection of diseases. In this study, we investigate studies that utilize more than one supervised ML model for each disease recognition problem. This approach renders more comprehensiveness and precision because the evaluation of the performance of a single algorithm over various study settings induces bias which generates imprecise results. The analysis of ML models will be conducted on few diseases located at heart, kidney, breast, and brain. For the detection of the disease, numerous methodologies will be evaluated ML.

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

1.1 RELATED WORKS

COLLECTING DATASET:

Data Collection is one of the most important tasks in building a machine learning model.We collect the specific dataset based on requirements from internet.The dataset contains some unwanted data also. So first we need to pre-process the data and obtain perfect data set for algorithm.

Data pre-processing:

Data pre-processing can be done by data cleaning, data transformation, data selection.

Data cleaning:

Fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.Data transformation may include smoothing, aggregation, generalization, transformation which improves the quality of the data.Data selection includes some methods or functions which allow us to select the useful data for our system.

1.2 LOGISTIC ALGORITHM:

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression).

1.3 SVM(Support Vector Machine):

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

1.3 RANDOM FOREST

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction

a) Predicting Heart Disease at Early Stages using Machine Learning: A Survey, Predicting and detection of heart disease has always been a critical and challenging task for healthcare practitioners. Hospitals and other clinics are offering expensive therapies and operations to treat heart diseases. So, predicting heart disease at the early stages will be useful to the people around the world so that they will take necessary actions before getting severe. Heart disease is a significant problem in recent times; the main reason for this disease is the intake of alcohol, tobacco, and lack of physical exercise. Over the years, machine learning shows effective results in making decisions and predictions from the broad set of data produced by the health care industry. Among them, Support vector machine gives the best accuracy of 95.55%,Rahul Katarya; Polipireddy Srinivas, 2020.

b) Heart Disease Prediction Using Supervised Machine Learning Algorithms, Predicting and detecting cardiac disease has always been a difficult and time-consuming undertaking for doctors. To treat cardiac disorders, hospitals and other clinics are giving costly therapies and operations. As a result, anticipating cardiac disease in its early stages will be beneficial to people all around the world, allowing them to take required treatment before it becomes serious. Heart disease has been a major issue in recent years, with the primary causes being excessive alcohol use, tobacco use, and a lack of physical activity. Machine learning methods are utilized to forecast cardiac illnesses in this article. For training and testing, a data collection containing diverse human health parameters is used. Among them, algorithms gives the best accuracy of 90.96%,Narendra Mohan; Vinod Jain,2021c) Ethereum Smart Contract Analysis a Systematic Review, Satpal Singh, 2022.

C) Liver Disease Prediction using Semi Supervised based Machine Learning Algorithm, Recently, liver disease is emerging as one of the most common diseases with high fatality rate. The number of liver affected patients are steadily increasing due to various factors such as excessive alcohol consumption, inhalation of hazardous fumes, eating tainted food, pickles, and narcotics. Liver disease can also lead to a variety of serious illnesses, including liver cancer. To enhance the process of liver disease classification, a semi supervised machine learning algorithm has been presented in this research work. The use of liver patient datasets in the development of classification algorithms to predict liver disease is being investigated. Among them, algorithm gives the best accuracy of 89.30%, A.Jaya Mabel Rani; S. Nishanthini,2022.

d) Designing Disease Prediction Model Using Machine Learning Approach, Now-a-days, people face various diseases due to the environmental condition and their living habits. So the prediction of disease at earlier stage becomes important task. But the accurate prediction on the basis of

sym pto ms becomes too difficult for doctor. The correct prediction of disease is the most challenging task. To overcome this problem data mining plays an important role to predict the disease. Medical science has large amount of data growth per year. Due to increase amount of data growth in medical and healthcare field the accurate analysis on medical data which has been benefits from early patient care. The accuracy of general disease prediction by using CNN is 84.5% which is more than KNN algorithm, Dhiraj Dahiwade; Gajanan Patle,2019.

CONCLUSION

The use of different ML algorithms enabled the early detection of many maladies such as heart, kidney, breast, and brain diseases. Furthermore, ML model showed superiority in ac-curacy at most times for kidney diseases and PD because of its reliability in handling high-dimensional, semi-structured and unstructured data. For Breast cancer prediction, RF showed more superiority in the probability of correct classification of the diseases because of its ability to scale well for large datasets and its susceptibility to avoid overfitting.

REFERENCES

[1] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction ofheart disease using machine learning," in 2018 Second InternationalConference on Electronics, Communication and Aerospace Technology(ICECA), 2021, pp. 1275–1278.

[2] Y. Hasija, N. Garg, and S. Sourav, "Automated detection of dermatolog-ical disorders through image-processing and machine learning," in 2017International Conference on Intelligent Sustainable Systems (ICISS),2020, pp. 1047–1051.

[3] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparingdifferent supervised machine learning algorithms for disease prediction,"BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1–16, 2022.

[4] R. Katarya and P. Srinivas, "Predicting heart disease at early stagesusing machine learning: A survey," in 2020 International Conferenceon Electronics and Sustainable Communication Systems (ICESC), 2021,pp. 302–305.

[5] P. S. Kohli and S. Arora, "Application of machine learning in diseaseprediction," in 2018 4th International Conference on Computing Com-munication and Automation (ICCCA), 2022, pp. 1–4.