Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Multiple Disease Prediction System Using Machine Learning

Mohammed Asrarulhaq Khadir, Abdulla Mohd,Muqtadir Ali,

Dr. Pathan Ahmed Khan

mah2khadir@gmail.comBE Student – Dept. Of CSE, ISL Engineering College, TS, India
bingoabdulla@gmail.com BE Student – Dept. Of CSE, ISL Engineering College, TS, India

Princemuqtadir535@gmail.comBE Student – Dept. Of CSE, ISL Engineering College, TS, India

pathanahmed@islec.edu.inAssociate Professor – Dept. Of CSE, ISL Engineering College, TS, India

Article Info
Page Number: 1435-1445
Publication Issue:
Vol. 72 No. 1 (2023)

ABSTRACT
The field of healthcare is facing increasing challenges due to the rise in the number of diseases and the growing population. Medical practitioners are finding it difficult to diagnose and predict the onset of diseases with the required level of accuracy, especially with a rapidly growing population. To address this need, a dynamic healthcare system is required that can process data in a smart, efficient, and precise manner. Machine learning and artificial intelligence have proven to be powerful tools in this regard. This project focuses on developing a prediction system that can identify multiple diseases simultaneously, for ex, predicting heart disease by using parameters such as Pulse Rate, Cholesterol, Blood Pressure, Heart Rate, etc. The user has to enter various parameters of the disease and the system would display the output whether he/she has the disease or not. This project can help a lot of people as one can monitor the persons' condition and take the necessary precautions. The proposed model can identify diseases such as Heart disease, Diabetes, Liver disease, hepatitis, jaundice, and Parkinson, and also determine the risk factors associated with them with good accuracy and precision. The proposed system can detect more than one disease at a time.

Keywords:Heart, Diabetes, Liver, Hepatitis, Jaundice, Parkinson, Multiple Diseases, Prediction System.

Article History
Article Received: 15 October 2022
Revised: 24 November 2022
Accepted: 18 December 2022

1. INTRODUCTION

In this digital world, data is an asset, and enormous data was generated in all the fields. Data in the healthcare industry consists of all the information related to patients(1). Here a general architecture has been proposed for predicting the disease in the healthcare industry. Many of the existing models are concentrating on one disease per analysis(2). Like one analysis for diabetes analysis, one for heart analysis, one for liver diseases like that. There is no common system present that can analyze more than one disease at a time. Thus, we are concentrating on providing immediate and accurate disease predictions to the users about the symptoms they enter along with the disease predicted(3). So, we are proposing a system which used to predict multiple diseases by using Streamlit. In this system, we are going to analyze Diabetes, Heart, and Liver, Hepatitis, Parkinson, Jaundice disease analysis. Later many more diseases can be included. To implement multiple disease prediction systems we are going to use

machine learning algorithms, and Streamlit. Python pickling is used to save the behaviour of the model(4). The importance of this system analysis is that while analysing the diseases all the parameters which cause the disease is included so it is possible to detect the disease efficiently and more accurately. The final model's behaviour will be saved as a python pickle file.

Multiple disease prediction model is the system in which the users are allowed to input their health records that they have obtained from various prescribed tests' reports. Disease predictions with this model can be considered as an intermediate step between taking up the prescribed health check-up test and consultation with the medical practitioners(5). Thereupon, not all the patients will have a need to go for consultation. Compared with the existing system, the model that has been proposed is more flexible and efficient, as the prediction of all six diseases is done with single set of inputs and the users are given a choice, either to predict with the features they are most interested in or with the system recommended features. The benefit of multiple disease prediction model is that it can predict the occurrence probability of various diseases in advance, thereby reducing the mortality ratio. The proposed model has overcome one of the demerits of the existing model, which is prediction can be done with single set of input data with higher flexibility and it is fixed with 6 types of diseases. A lot of analysis over existing systems in the health care industry considered only one disease at a time. For example, one system is used to analyse diabetes, another is used to analyse heart, and another system is used to predict Parkinson disease. Maximum systems focus on a particular disease. When an organization wants to analyse their patient's health reports then they have to deploy many models.

The approach in the existing system is useful to analyse only particular diseases. In multiple diseases prediction system a user can analyse more than one disease on a single website. The user doesn't need to traverse different places in order to predict whether he/she has a particular disease or not. In multiple diseases prediction system, the user needs to select the name of the particular disease, enter its parameters and just click on submit. The corresponding machine learning model will be invoked and it would predict the output and display it on the screen(6).

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. For example first is for liver analysis, one for diabetes analysis, one for jaundice diseases like that. If a user wants to predict more than one disease, he/she has to go through different sites. There is no common system where one analysis can perform more than one disease prediction. Some of the models have lower accuracy which can seriously affect patients' health. When an organization wants to analyse their patient's health reports, they have to deploy many models which in turn increases the cost as well as time some of the existing systems consider very few parameters which can yield false results(7).The proposed system is a powerful disease prediction tool that enables users to predict a multiple disease at a one place with great accuracy and ease. The system is designed to run on a website, making it easily accessible to users from all over the world. By entering various parameters of the disease, the system can generate a prediction of whether the user has the disease or not. This can help users to monitor their health status and take necessary precautions to avoid or

manage the disease. The system uses machine learning and predictive modelling algorithms to generate its predictions, making it highly efficient and reliable. The system is trained to predict the disease using multiple algorithm to achieve better accuracy, Logistic Regression, K-Nearest Neighbour, Random Forest Algorithm, and Support Vector Machine and Streamlit framework to deploy our model(8).

2. LITERATURE SURVEY:

The vast majority of current studies focused on a common illness. When a user wants to analyze diabetes, they must use one model, and when they try to analyze heart disease, they must use another model. This is a lengthy procedure. Furthermore, if a user has several illnesses but the current method can only predict one of them, then there is a risk of death. The key goal of this article is to create a multiple disease prediction model, so the machine learning approaches employed are briefly discussed. Different machine learning techniques are employed to analyze diabetes, heart disease prediction, and liver disease, Jaundice, Hepatitis, Parkinson.

Several studies have investigated the use of machine learning algorithms for predicting heart disease. For instance, a study by Kaur and Singh (2020) proposed a prediction model for heart disease using SVM and K-nearest neighbour algorithms. The results showed that the SVM algorithm outperformed the K-nearest neighbour algorithm in terms of accuracy(9)(10).

Similarly, liver disease prediction models have also been developed using machine learning algorithms. In a study by Khan et al. (2019), a Random Forest algorithm was employed to predict liver disease with high accuracy. The study also highlighted the importance of using feature selection techniques to improve the performance of the prediction model.

Diabetes is another disease that has been extensively studied using machine learning algorithms. A study by Patil et al. (2021) developed a prediction model for diabetes using logistic regression and K-nearest neighbour algorithms. The results showed that the logistic regression algorithm performed better than the K-nearest neighbour algorithm in terms of accuracy(11).

Parkinson's disease prediction models have also been developed using machine learning algorithms. In a study by Acharya et al. (2020), a Random Forest algorithm was employed to predict Parkinson's disease with high accuracy. The study also highlighted the importance of using feature selection techniques to improve the performance of the prediction model.

Hepatitis is a viral infection that affects the liver and can lead to severe complications if not detected and treated promptly. Machine learning techniques have been employed to develop predictive models for hepatitis. For example, a study by Wang et al. (2018) utilized a Support Vector Machine algorithm to predict hepatitis B infection. The study demonstrated high accuracy in distinguishing between hepatitis B-positive and -negative cases, facilitating early intervention and appropriate medical care.

Jaundice, characterized by yellowing of the skin and eyes, can be caused by various underlying conditions affecting the liver. Prediction models using machine learning

algorithms have been developed to aid in the diagnosis and management of jaundice. In a study by Das et al. (2020), Random Forest algorithm was employed to predict neonatal jaundice. The model integrated clinical parameters, bilirubin levels, and demographic data to accurately identify infants at risk of developing severe jaundice, enabling timely intervention and monitoring(12)(13).

3.PROPOSED METHODOLOGY

The proposed method uses the Logistic Regression and K-Nearest Neighbour and Random Forest and SVM algorithms. The main idea behind the proposed system after reviewing the existing system is to create a multiple prediction system based on the required inputs, with high accuracy(14).

Purpose of Proposed System:

1.Developing a user-friendly web-based system for users and remote hospitals, to diagnose the disease and take correct treatment plan.

2. Recognizing different diseases accurately from required inputs.

MODULES:

User module:

In this module user have access to all the diseases (Diabetes, heart, hepatitis, jaundice, liver, Parkinson). User can open any disease to predict the disease. User have to enter required parameters necessary to predict the disease(16)

Diabetes Module:

When user select diabetes prediction to predict the disease, Diabetes module include required parameter such as no of pregnancies, glucose level, blood pressure, skin thickness, insulin, bmi, diabetespedigreefunction, age.

Heart Module:

When user select heart prediction to predict the disease, Heart module include required parameterssuch as age and gender then, chest pain type, resting bp, serum cholesterol, fasting blood sugar, resting ECG, max heart rate achieved, ST depression induced by exercise relative to rest, Peak exercise ST segment, no of major vessel(0-3) coloured by fluoroscopy, thalassemia(17).

Hepatitis Module:

When user select hepatitis prediction to predict the disease, Hepatitis module include required parameters such as age, gender. Total bilirubin direct bilirubin, alkaline phosphatase alanine aminotransferase, total proteins. Albumin.

Jaundice Module:

When user select jaundice prediction to predict the disease, jaundice prediction includes required parameter such as age, gender. Total bilirubin direct bilirubin, alkaline phosphatase alanine aminotransferase, total proteins. Albumin(18).

Liver Module:

When user select liver prediction to predict the liver disease, liver prediction includes required parameter such as age, gender. Total bilirubin direct bilirubin, alkaline phosphatase alanine aminotransferase, total proteins. Albumin.

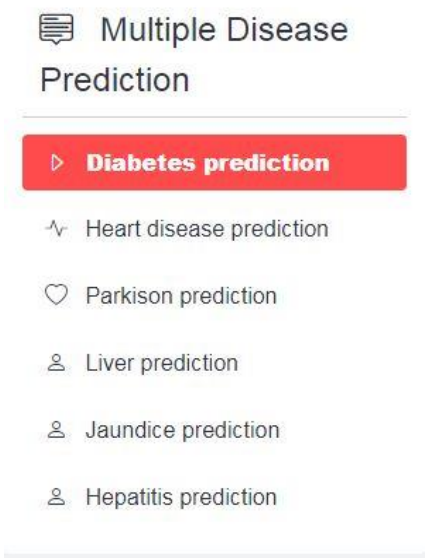
Parkinson Module:

When user select the Parkinson prediction to predict Parkinson disease, Parkinson prediction includes required parameter such as Average vocal fundamental frequency, Maximum vocal fundamental frequency, Minimum vocal fundamental frequency, Jitter(%), Jitter(Abs) MDVP:RAP, MDVP:PPQ, Jitter:DDP, MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:DDA , NHR, HNR, status, RPDE, DFA, spread1, spread2,D2, PPE(19).

Graphical User Interface (GUI)

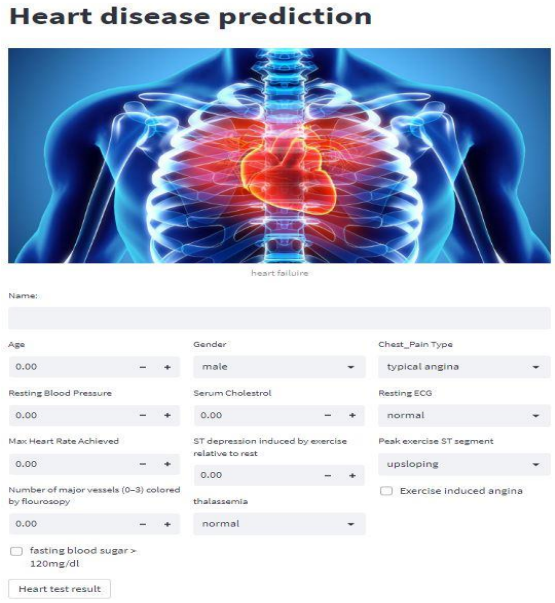
A graphical user interface or GUI, is a type of interface that allows users to interact withelectronicdevicesthroughgraphicaliconsandvisualindicatorssuchassecondarynotation, as opposed to text-based interfaces, typed command labels or text navigation(20).

Menu: Fig 1: GUI of Menu



Diabetes : Fig 2: GUI of Diabetes

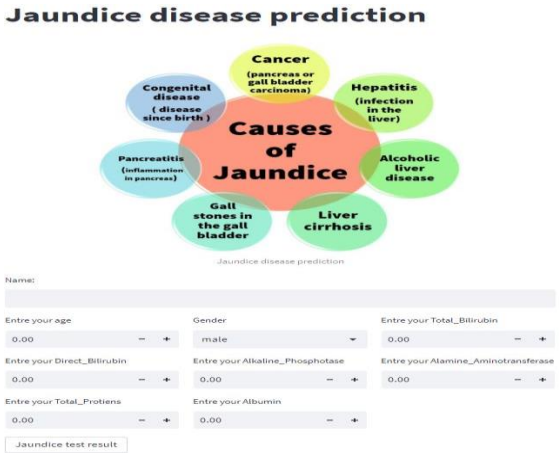
Heart Disease : Fig 3 : GUI of Heart



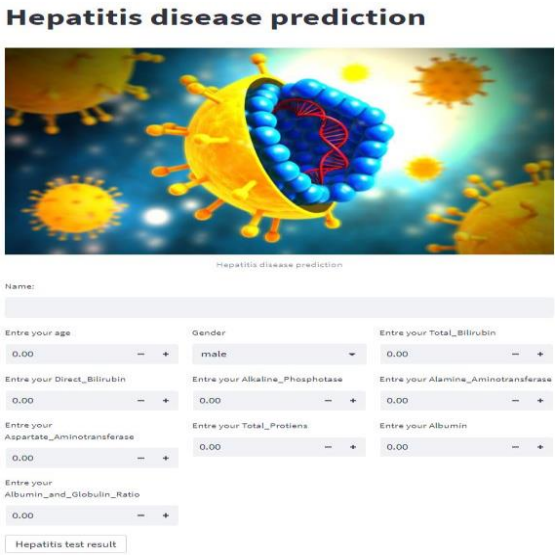
Parkinson : Fig 4 : GUI of Parkinson

Liver: Fig 5 : GUI of Liver

Jaundice : Fig 6: GUI of Jaundice



Hepatitis : Fig 7 : GUI of Hepatitis



4. CONCLUSION and FUTURE WORK:

Our project focused on the development of a Multiple Disease Prediction System using machine learning algorithms. The aim was to create a robust system capable of predicting

various diseases simultaneously, namely Heart disease, Liver disease, Diabetes, Jaundice, Hepatitis, and Parkinson's disease.

By implementing popular machine learning algorithms such as Support Vector Machine (SVM), K-nearest Neighbors (KNN), Random Forest, and Logistic Regression, we were able to provide accurate disease predictions based on user-provided parameters. The system allowed users to input specific disease-related parameters, and the model would predict whether they have the particular disease or not.

By providing users with an accessible web interface, our project aimed to enhance the healthcare system by enabling individuals to monitor their health conditions and take necessary precautions. Early disease detection can lead to timely interventions and improved life expectancy. Overall, our Multiple Disease Prediction System serves as a valuable tool in the field of healthcare, leveraging the power of machine learning to assist in disease prediction and contribute to proactive healthcare management.

There are several potential areas for future work and enhancements to further improve the Multiple Disease Prediction System:

Expansion of Disease Coverage: The system can be expanded to include more diseases, allowing for a broader range of predictions. Incorporating additional diseases based on available datasets and medical research can enhance the system's usefulness and provide users with a more comprehensive disease prediction capability.

Integration of Real-Time Data: Currently, the system relies on static datasets for predictions. Integrating real-time data from various sources such as wearable devices, electronic health records, and health monitoring systems can enable dynamic and up-to-date disease prediction. This would enhance the accuracy and relevance of the predictions, providing users with timely information about their health status.

Refinement of Machine Learning Algorithms: Future work can involve exploring and implementing advanced machine learning algorithms or ensemble methods to further enhance the accuracy and robustness of disease prediction. Techniques such as deep learning, gradient boosting, and neural networks can be explored to improve the model's performance and handle complex disease patterns.

User Feedback and Personalization: Gathering feedback from users and healthcare professionals can help identify areas for improvement and tailor the system to specific user needs. Incorporating user feedback and preferences can enhance the user experience and increase the system's usability and effectiveness.

Mobile Application Development: Creating a mobile application for the Multiple Disease Prediction System can enhance accessibility and convenience for users. The application can provide features such as personalized health monitoring, reminders for medical check-ups, and notifications for disease risk factors based on user profiles.

Collaboration with Healthcare Providers: Collaborating with healthcare providers and professionals can facilitate the integration of the prediction system into existing healthcare systems. This collaboration can lead to the development of decision support tools that assist healthcare providers in making accurate diagnoses and treatment plans based on the predictions generated by the system.

REFERENCES:

1. 2022 International Mobile and Embedded Technology Conference (MECON) | 978-1-6654-2020-4/22/$
2. Kaur, R., & Singh, D. (2020). Comparative analysis of machine learning algorithms for heart disease prediction. Journal of Healthcare Engineering, 2020, 1-10.
3. Khan, S. A., Khan, A. A., Asif, A., & Ishaq, M. (2019). Prediction of liver disease using machine learning algorithms. Journal of Medical Imaging and Health Informatics, 9(4), 625-631.
4. Kumar, A., Dhabliya, D., Agarwal, P., Aneja, N., Dadheech, P., Jamal, S. S., & Antwi, O. A. (2022). Cyber-internet security framework to conquer energy-related attacks on the internet of things with machine learning techniques. Computational Intelligence and Neuroscience, 2022 doi:10.1155/2022/8803586
5. Kumar, S. A. S., Naveen, R., Dhabliya, D., Shankar, B. M., & Rajesh, B. N. (2020). Electronic currency note sterilizer machine. Materials Today: Proceedings, 37(Part 2), 1442-1444. doi:10.1016/j.matpr.2020.07.064
6. Mandal, D., Shukla, A., Ghosh, A., Gupta, A., & Dhabliya, D. (2022). Molecular dynamics simulation for serial and parallel computation using leaf frog algorithm. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 552-557. doi:10.1109/PDGC56933.2022.10053161 Retrieved from www.scopus.com
7. Mehraj, H., Jayadevappa, D., Haleem, S. L. A., Parveen, R., Madduri, A., Ayyagari, M. R., & Dhabliya, D. (2021). Protection motivation theory using multi-factor authentication for providing security over social networking sites. Pattern Recognition Letters, 152, 218-224. doi:10.1016/j.patrec.2021.10.002
8. Patil, D., Khatri, R., & Saha, S. (2021). A comparative analysis of machine learning algorithms for diabetes prediction. Journal of Ambient Intelligence and Humanized Computing, 1-14.
9. Acharya, D. P., Adeli, H., & Nguyen, T. K. (2020). Application of machine learning in Parkinson's disease diagnosis. Brain Sciences, 10(4), 212.
10. Wang, H., Ding, Y., Tang, H., Wang, L., & Xia, J. (2018). Prediction of hepatitis B infection among chronic hepatitis B carriers using machine learning algorithms. Frontiers in Public Health, 6, 230.
11. Das, A. P., Saini, A., & Arora, S. (2020). Machine learning approach for prediction of neonatal jaundice. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

12. Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020) IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1
13. M. R. Al-Mahdi and H. R. Al-Samarraie, "Predictive analytics for diabetes using machine learning algorithms," 2019 6th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), Granada, Spain, 2019, pp. 698-703.
14. R. Sinha, K. Gupta, M. Gupta, A. Singla, and D. K. Mishra, "A Survey on Machine Learning Algorithms for Heart Disease Prediction," 2019 2nd International Conference on Computing, Communication, Control and Automation (ICCUBEA), Greater Noida, India, 2019, pp. 1-6.
15. J. Jain and N. Gupta, "Prediction of Parkinson's disease using machine learning algorithms: A review," 2019 IEEE 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Greater Noida, India, 2019, pp. 1-5.
16. S. K. Tripathy and S. Nayak, "A Review on Predictive Analytics for Hepatitis B Using Machine Learning Algorithms," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-6.
17. A. Sharma, R. K. Yadav, and S. Jain, "A Review on Machine Learning Based Approaches for Jaundice Detection," 2020 2nd International Conference on Computing, Communication, Control and Automation (ICCUBEA), Jaipur, India, 2020, pp. 1-6.
18. Khan, P. A., & Yadav, S. K. (2021). A machine learning-based approach for classification of heart disease in health care domain through data mining. *Design Engineering (Toronto)*, 742–753. http://www.thedesignengineering.com/index.php/DE/article/view/4915
19. Fayaz, N., Fatima, H., Fatima, S., Ahmed Khan, P., & Student, B. E. (n.d.). Diagnosis of diabetic patient and analyzing insulin dosage by using gradient boosting and logistic regression algorithm. Jetir.org. Retrieved May 10, 2023, from https://www.jetir.org/papers/JETIR2205237.pdf
20. Khan, P. A., & Sharma, Y. K. (2021). Prediction of Effective Heart Disease in Health care domain using Data Mining Techniques. International Journal of Grid and Distributed Computing, 14(1), 1400–1408. http://sersc.org/journals/index.php/IJGDC/article/view/36761
21. Khan, P. A. (n.d.). DISEASE PREDICTION USING BINARY TREE ALGORITHM BY COMPARING ITS ACCURACY WITH VARIOUS OTHER ALGORITHMS AND DRUG RECOMMENDATION BASED ON PREDICTED DISEASE CONSIDERING AGE FACTOR. Retrieved May 10, 2023, from http://journalstd.com/gallery/9-may2022.pdf
22. Khan, P. A., Ali, M. R., Assistant Professor, & Assistant Professor. (n.d.). Methodological implementation of academic organizational operated data with clustering mechanism using an improved k- means algorithm. Ijmec.com. Retrieved May 10, 2023, from https://doi-ds.org/doilink/10.2022-22255487

23. Khan, P. A., Mir, M., Zakir, M. D., & Khan, A. (n.d.). Crop yield prediction using machine learning algorithms. Jespublication.com. Retrieved May 10, 2023, from https://jespublication.com/upload/2022-V13I5017.pdf