e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

PROGNOSISAI: AN AI-ENABLED DISEASE PREDICTIONSYSTEM FOR EARLY DETECTION AND PREVENTION

Shivam Pathak[*1], Vaishnavi Tiwari[*2], Er. Prabhat Kumar Yadav[*3], Dr. Vibha Srivastava[*4], Er. Kaushlendra Yadav[*5]

[*1,2,3,4,5]Department Of Information Technology, Shri Ramswaroop Memorial College Of Engineering AndManagement, Lucknow, Uttar Pradesh, India.

DOI : https://www.doi.org/10.56726/IRJMETS52202

ABSTRACT

This study aims to assist a person in predicting the sickness they are experiencing based on their symptoms and accurate measurements of their vital signs. Sometimes, people choose to ignore health conditions because the cost of care is so high. In the ever-changing world of today, where human existence is always changing, population health becomes a critical issue. When health conditions are diagnosed too late, there might be serious repercussions, such as higher death rates. Taking note of this urgent problem, we suggest a unique machine learning-based method for disease prediction. Our research intends to enable early disease identification through vital sign evaluations and symptom analysis, empowering individuals to take proactive measures for their health.

Keywords: Random Forest, Gaussian Naïve Classifier, Support Vector Classifier, Machine Learning Algorithm, Desease Prediction.

I. INTRODUCTION

The use of machine learning in disease prediction systems has triggered a paradigm shift in the healthcare industry. Early disease prediction of a human is an important step in the treatment of disease. Since the very beginning, a doctor has handled it almost exclusively. Thus, the healthcare industry thrives on innovation to make logistics efficient [1]. The current general paradigm is mostly dependent on the existence of symptoms and interpersonal cummunication[2].The rapid expansion of medical data necessitates the development of effective analytic methods in order to glean insightful information and promote early disease identification. In this regard, the ultimate objective of our research is to create a comprehensive illness prediction platform in order to overcome the difficulties related to late disease diagnosis.

One of the most pressing issues in the medical industry is the workload on the doctors [3] and the unaffordable consultation cost[4]. The current methodology of the medical industry consists of the patient visiting a generalist doctor and explaining to the doctor the conditions, and symptoms faced by the patient upon which the doctor infers possible diseases and then channels them to a specialist doctor [5]. The use of artificial intelligence (AI)-based methods for the detection of medical diseases is becoming more and more popular as a result of the quick development of AI technology. Large volumes of medical data may be sifted through by these algorithms, which can then be used to identify trends and predict prognoses, treatment options, and illness diagnoses. The promise of artificial intelligence (AI) in medical illness diagnosis is its capacity to enhance and supplement established diagnostic techniques, improving precision, efficacy, and eventually patient outcomes. By automating repetitive processes, Significant cost savings and resource optimisation within healthcare systems could result from this automation, increasing access to high-quality care and lowering its cost on a worldwide basis. There is much promise for the future of AI- based medical disease identification. These algorithms have the potential to completely transform healthcare delivery as technology advances by enabling more precise and timely diagnoses, individualised treatment regimens, and enhanced patient monitoring. Moreover, AIdriven insights can reveal brand-new therapeutic targets or biomarkers, expanding our knowledge of how diseases work and making it easier to create tailored treatments. Interest in applying AI-based techniques for medical disease diagnosis has increased due to the quick development of AI technologies. The core of the medical sector is innovation. It is what propels novel therapeutics, remedies and treatments[6]. AI-based disease identification has promise as a complement to current techniques, with the potential to enhance patient outcomes, speed, andaccuracy.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

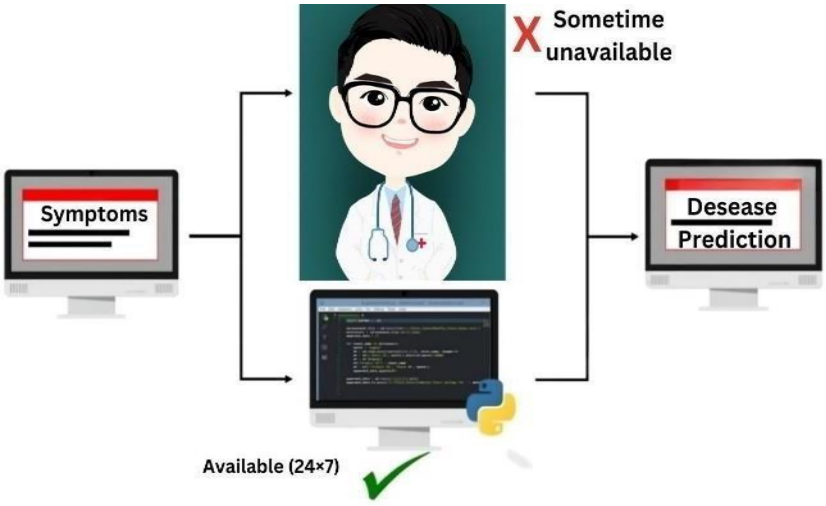

Fig. 1 Proposed system for disease prediction

When needed, the doctor might not always be available. But as things are right now, this prediction technique can be applied anytime it's needed. The age, gender, and symptoms of a person can be input into the ML model for further processing. Following initial data processing, the machine learning model trains and tests the algorithm that produces the anticipated disease using the current input.

II. LITERATURE REVIEW

The body of research on the application of machine learning algorithms for disease prediction is extensive, including a broad spectrum of diseases and approaches. Several research studies have exhibited the effectiveness of machine learning algorithms in forecasting a range of health issues.

To get beyond the limitations of machine learning, the authors of this work [7] have put up the idea of large data-driven machine learning-based illness prediction. Big data is being smoothly implemented in hospitals' biomedical and healthcare communities to ensure reliable findings in every experiment. This idea is (a)minimizes incomplete data and (b) improves the accuracy of disease prediction. The suggested idea is put to the test or tested using actual hospital data sets, including daily updated hospital-oriented information, patient and doctor preferences for patient details, disease preferences for disease-oriented data, and so on. The two primary issues with the current system that this technique addresses are (i) missing data and (ii) incomplete data. to reassemble the model of latent factors. The idea is to use the Machine Learning Decision Tree algorithm and the Map Reduce (MR) technique to obtain data from a hospital that gathered data from a forum known as "structured and unstructured data." For data, the MR algorithm is utilized. dividing. With the conventional pace, it achieves 94.8%, although it does so more quickly than CNN-UDRP. After that, it reports on the likelihood of disease occurrences.

The survey study [8] on "prediction of disease using machine learning over big data" is presented by the authors. can create the foundation for a medical specialization This idea is used to create mass medical data—that is, data that has been expanded—from medical data. This concept's intended outcome is that the most basic data is kept in the "Medical data analysis in massive collection" is the term for the field of huge medical data analysis. Compared to CNN-UDRP, it generates correctness and reaches 4.8% speed faster. Only these three types of data are covered: (a) text data, (b) structured data, and (c) combined structured and text data. The word "medical data oriented" is improved in this suggested method.

The author of this study introduced the idea of "Improving disease forecasting by machine learning," which entails enhancing disease prediction through the application of machine learning[9]. Big data is improving this kind of knowledge by increasing the amount of medical data. This idea makes use of a genetic algorithm and recovers data—that is, missing data—from a dataset that also contains medical data. The two calculation terms used by this system are (i) KNN and (ii) SVM. The number of chronic illnesses rises. The medical data is used in the CNN-MDRP approach. The database contains personal and medical information as well as a thorough history of each patient. The logical data may be found with ease using RNN-based approaches. Both online and offline techniques are used in this system. The idea of "Competitor Mining and Unstructured Dataset Handling

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Technique," which is applied in the healthcare industry, isprovided by the author in the paper [10]. Competitive mining is discussed in this studyalong with relevant studies. at last provided rival mining algorithms along with their benefits and cons. This work Comparing the experimental results with other methods, CMiner++ produced the lowest computation time.

III. METHODOLOGY

1. Data Collection:

Identify relevant datasets encompassing diverse medical domains, including clinical records, imaging data, genetic information, and patient demographics. Ensure the quality and integrity of the collected data by performing data validation and preprocessing procedures, including cleaning, normalisation, and handling missing values.

2. Algorithm Selection:

Explore a range of machine learning algorithms suitable for disease prediction tasks:

- Decision Trees
- Random Forests
- Support Vector Machines (SVM) Consider the characteristics of the dataset and the complexity of the prediction taskwhen selecting appropriate algorithms.

3. Model Training:

Split the dataset into training, validation, andtesting sets to evaluate model performance effectively. Implement the selected machinelearning algorithms using appropriate libraries and frameworks. Train the models on the training data using cross-validation techniques to optimise hyperparameters and prevent overfitting.

4. Model Evaluation:

Evaluate the performance of the trained models using relevant evaluation metrics, including accuracy and precision. Conduct comparative analysis between different algorithms to identify the most effective model for disease prediction. Validate the robustness and generalizability of the models using the validation and testing datasets.

5. Implementation:

Develop a disease prediction system integrating the trained machine learning models into a user friendly interface. Utilise scalable and efficient implementation techniques, leveraging parallel computing and distributed processing, to handle large datasets effectively.

6. Validation:

Validate the disease prediction system through real-world case studies and application scenarios involving diverse patient populations and healthcare settings. Solicit feedback from healthcare practitioners and domain experts to assess the usability, reliability, and clinical relevance of the system.

IV. RESULT AND DISCUSSION

The experimental results demonstrated promising findings, indicating the effectiveness of machine learning algorithms in disease prediction across various medical domains. Interpreting the data revealed important information about illness processes and predicted traits. Furthermore, the discourse emphasised the consequences for healthcare professionals and legislators, stressing the revolutionary capacity of machine learning driven illness forecastingsystems to enhance healthcare provision and patient results.

V. CONCLUSION

In conclusion, a major development in healthcare technology is the creation of a disease prediction system utilising machine learning techniques. Such systems have the potential to revolutionise illness diagnosis, treatment, and outcomes by utilising big databases and sophisticated algorithms. The results highlight how crucial machine learning-driven strategies are for enhancing conventional diagnostic techniques and patient care. Prolonged investigation and creativity in this domain have the potential to revolutionise healthcare provision worldwide.

VI. REFERENCES

[1] Zhou, S.-M., Fernandez-Gutierrez, F.,Kennedy, J., Cooksey, R., Atkinson, M., Denaxas,S., Siebert, S., Dixon, W.G., O'Neill, T.W. and Choy, E., "Defining disease phenotypes in primary care electronic health records by a machine learning approach: A case study in identifying rheumatoid arthritis", PloS One, Vol. 11, No. 5,(2016),e0154515.https://doi.org/10.1371/journal.pone.0154515

[2] Mobeen, A., Shafiq, M., Aziz, M.H. and Mohsin, M.J., "Impact of workflow interruptions on baseline activities of the doctors working in theemergency department", BMJ Open Quality, Vol.11, No. 3, (2022), e001813. doi: 10.1136/bmjoq- 2022- 001813. 7.

[3] Ahmed, S., Szabo, S. and Nilsen, K., "Catastrophic healthcare expenditure and impoverishment in tropical deltas: Evidence fromthe mekong delta region", International Journal for Equity in Health, Vol. 17, No. 1, (2018), 1-13.doi: 10.1186/s12939-018-0757-5.

[4] Roberts, M.A. and Abery, B.H., "A person- centered approach to home and community-based services outcome measurement", Frontiers in rehabilitation Sciences, Vol. 4, (2023).
doi:10.3389/fresc.2023.1056530.

[5] Littell, C.L., "Innovation in medical technology: Reading the indicators", Health Affairs, Vol. 13, No. 3, (1994),226-235.https://doi.org/10.1377/hlthaff.13.3.226.

[6] 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.

[7] 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.

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

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

[10] Roberts, M.A. and Abery, B.H., "A person centered approach to home and community based services outcome measurement", Frontiers in rehabilitation Sciences, Vol. 4, (2023).
doi:10.3389/fresc.2023.1056530.