DISEASE PROGNOSIS : DEMYSTIFIED WITH LINEAR REGRESSION

Ms. R. Punitha Gowri[1], Mr. K. Hareesh Kumar[2], Mr. T. Jagadeesh[3], Mr. S. Karthik[4]

Assistant Professor[1], Students[2 3 4], Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Perundurai, Erode, Tamil Nadu, India.

ABSTRACT

The proactive approach of disease prediction is a crucial aspect of healthcare and medical research, aimed at identifying the likelihood of an individual developing a specific medical condition or ailment in the future. This crucial project uses a variety of data sources, such as genetic information, lifestyle characteristics, medical histories, and newly developed technology, to develop prediction models that can offer early warning indicators and insights into possible health problems. The ultimate objective of disease prediction is to equip healthcare professionals and individuals with the knowledge and tools necessary to take preventive measures, make informed decisions, and ultimately improve overall health and well-being.A dependable approach to disease prediction, our suggested Linear Regression methodology has multiple benefits over conventional techniques, including high accuracy, the capacity to extract intricate information from photos, and platform portability.

Keywords: Disease Prediction, Early Disease Prediction, Healthcare, Machine Learning, Prognosis, Semantic Word Analysis.

1. INTRODUCTION

The potential of disease prediction to revolutionize healthcare lies in its ability to enable early intervention, enhance patient outcomes, and alleviate the strain on healthcare systems. This, in turn, will lead to a more proactive and healthier approach to wellness and healthcare management. The objective of this project is to examine the changing landscape of disease prediction, emphasizing the significance of data-driven approaches and their potential impact on public health. We will explore the fundamental methodologies, technological advancements, and ethical considerations related to disease prediction, and analyze how this emerging field is positioned to revolutionize healthcare. The emphasis will be on the importance of proactive prevention and personalized care.

1.1 HEALTHCARE

The healthcare industry is an essential component of modern society, comprising a diverse network of professionals, institutions, and technologies dedicated to promoting the well-being and treatment of individuals. It is a complex and constantly evolving ecosystem that addresses a wide range of medical needs, from preventative care and diagnosis to treatment and rehabilitation. Healthcare not only plays a

fundamental role in preserving human health and reducing suffering, but it is also a significant driver of scientific and technological advancement. In this overview, we will explore the intricate world of healthcare, examining its vital role in society, the obstacles it faces, and the innovations that are revolutionizing the way we access, provide, and experience healthcare services. As we navigate the complex terrain of healthcare, we will emphasize the importance of collaboration, technology, and patient-centered care in shaping the future of this critical sector.

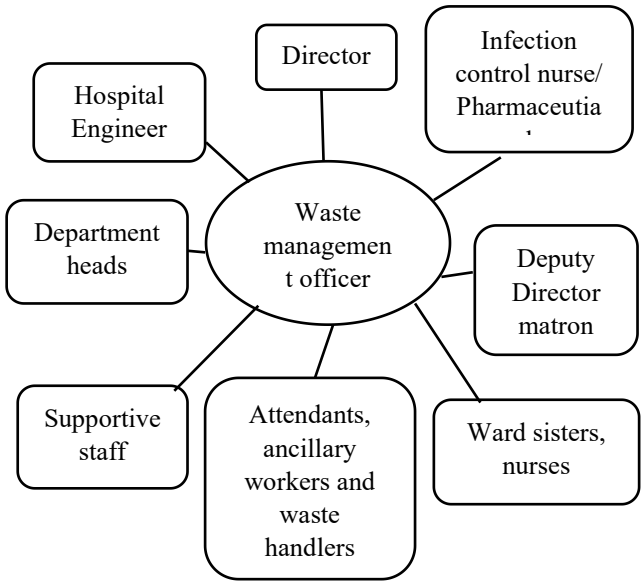

Figure 1. Healthcare

1.2 MACHINE LEARNING

In an increasingly data-driven society, the artificial intelligence subfield of machine learning has become a disruptive force. With the help of this field, computers can learn from enormous volumes of data, recognising patterns, forming hypotheses, and adapting to new knowledge all without the need for explicit programming. Machine learning has several uses in a variety of industries, including finance, healthcare, transportation, and entertainment. This technology completely transforms how we approach difficult issues and put decision-making procedures into motion. We will examine the enormous influence of machine learning in this introduction, covering everything from its fundamental ideas to real-world applications. We will also explore how it is influencing industries, improving our everyday lives, and laying the foundation for a world where data-driven, intelligent solutions are the norm. We will explore machine learning's promise, face its difficulties, and seize the fascinating opportunities it presents for fostering creativity and improving our comprehension of the outside world as we proceed on this adventure.

1.3 DISEASE PREDICTION

Disease prediction, which use data-driven strategies and state-of-the-art technologies to forecast and identify potential health issues before they become clinically manifest, is a crucial component of healthcare.Disease prediction utilizes artificial intelligence, machine learning, and extensive health data analysis to improve early detection, prevention, and intervention measures. Through the examination of various patient data sets, such as genetic markers, lifestyle variables, and past medical records, predictive models may be created to estimate the probability of particular illnesses or ailments. In addition to enabling healthcare providers to create individualized preventive programs, this dynamic strategy makes resource allocation easier and eventually improves public health outcomes overall.

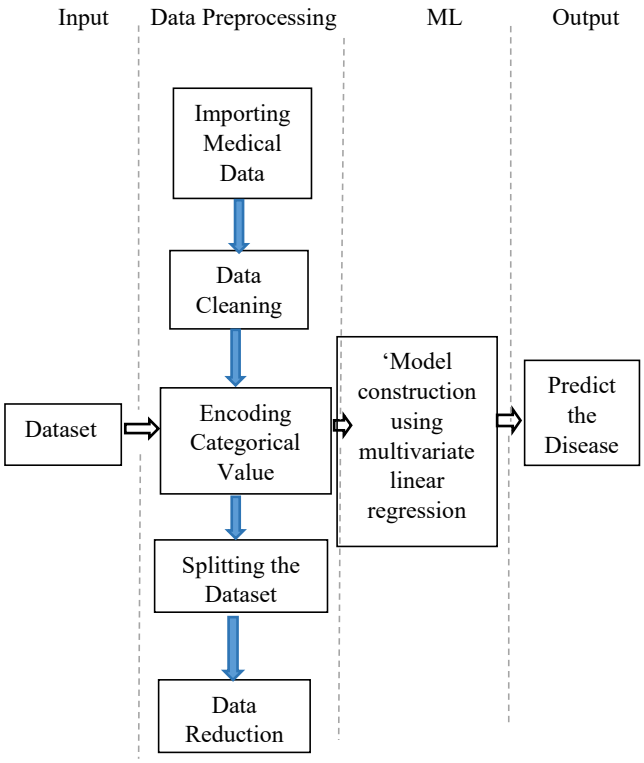

Figure 2. Disease Prediction

2. LITERATURE REVIEW

[1] Noncommunicable illness, particularly chronic disease, is the most prevalent cause of deteriorating physical and mental health complications. It is also a major source of sickness and death all across the world. Chronic illness is primarily prevented at a certain stage, despite the fact that its existence is crucial. These sickness prediction models were developed to help physicians and patients make clinical decisions. The study focuses on the Stacking Classifier in order to evaluate the performance of the best prediction model. With scores of 98.84% on PIMA, 98% on Synthetic after validation, 97.3% on ADRC, and 96.20% on FHD, the Proposed Model has the highest F1-Score. The suggested technique performs better, according to the findings of the comparative experiment conducted on the PIMA dataset. The proposed model is explained in the paper as well. It does an ethical analysis of explainability in relation to the application of machine learning models in therapeutic settings. The authors may determine that, in comparison to the other classification models, the suggested technique is the best one based on evaluation variables like accuracy and F1-score.

[2] A difficult subject that could improve the lives of the elderly and disabled who reside in remote places is the Digital Twin-based Healthcare System. The rise in popularity of smart city establishments and the evolution of various healthcare services optimised for smartphones has led to an increased interest in and need for modern digital twin-based healthcare systems. The proposed DTHS is constructed over the cloud platform for Parkinson disease prediction using the Optimised Fuzzy based k-Nearest Neighbour (OF-k-NN) classifier model.. It outperforms the existing Neural Network and Kernel-based SVM classifiers in terms of prediction accuracy for DS1 of 97.95% and 91.48%, prediction time for DS1 of 0.00127 seconds and DS2 of 0.00105 seconds, F1-Score of 0.98 for DS1 and 0.91 for DS2 and Matthews Correlation Coefficient of DS1 of 0.93675 and DS2 of 0.79816.

[3] In medical informatics, deep learning-based models are being used to forecast and identify cardiovascular diseases (CVDs). These models are able to describe phenotypes, pinpoint clinical symptoms, and offer solutions for managing complicated illnesses. A deep learning model for CVD prediction might be constructed using an enormous dataset of patient medical information. In order to investigate CVDs for early prediction, this study employed deep learning-based regression analysis on a dataset of 2621 medical records from UAE hospitals, including age, symptoms, and

CVD information. The authors suggest a long short-term memory-based deep neural network for early CVD prediction using regression analysis. Because of the overlapping symptoms, the accuracy level of the diseases enhanced when they were simulated in pairs of one disease with another. The study's findings indicate that coronary heart disease may be predicted with 71.5% accuracy, with 84.4% overlap with dyspnea. The authors demonstrated the efficacy of our suggested strategy on a variety of evaluation criteria.

[4] Asymptomatic Chronic Kidney Disease is widespread, and guideline-directed monitoring to predict CKD by many indicators is underused. Computer-aided automated diagnostics (CAD) can help forecast CKD. Because of their excellent classification accuracy, CAD systems such as deep learning algorithms are crucial in illness diagnosis. The study advocated integrating the highest performing DL models into the IoMT. With this approach, deep learning will be used more effectively and efficiently to aid predictive analytics in increasing CKD prediction.To fill in the missing variables, the author used multiple imputations (MI). The imputation procedure used linear regression to predict continuous data and logistic regression to identify categorical variables. In multiple imputations (MI), missing values in the dataset are replaced by n times, where n is usually a small number. To narrow down the data to a subset with a plausible range of values, the authors chosen the dataset that had the nearest means and standard deviations for its variables to the original dataset.

[5] A novel computing paradigm called fog computing enables geographically scattered end users to get extremely scalable, latency-aware services. Since data is swiftly stored and evaluated closer to data sources on local fog nodes, it is safer than cloud computing. In recent years, block chain (BC) technology has become one of the most amazing, revolutionary, and quickly growing inventions. The emphasis of BT's open platform is on anonymity and data security. By using the consensus process, it also guarantees the veracity and security of the data. Nowadays, BC is used in many industries, including healthcare. This study offers safe healthcare services based on Block chains that are effective for sickness prediction in fog computing.Compared to existing approaches, this suggested work efficiently groups and predicts the disease. To improve the effectiveness of the prediction findings, more hybrid clustering and classification models, as well as security and privacy for patient medical data access, may be implemented in the future.

3. EXISITING SYSTEM

Medical analysis's fast expanding and broad field of research is computer-aided diagnosis (CAD). Since medical diagnosis errors can result in extremely misleading therapies, there has been a strong push in recent years to develop computer-aided diagnostic tools. The use of machine learning (ML) in computer-aided diagnostic testing is essential. Simple equations are insufficient for accurately identifying objects such as body organs, thus necessitating training from real-life examples for effective pattern recognition. Pattern recognition and machine learning (ML) have enormous potential to improve the accuracy of disease diagnosis and treatment strategies in the realm of biomedical research. Furthermore, ML approaches ensure an impartial decision-making process. ML offers a reputable

method for developing superior and automated algorithms to analyse high-dimensional and multi-modal biomedical data. This survey paper focuses on the comparative study of various ML algorithms for detecting diseases such as heart disease and diabetes. It emphasizes the collection of algorithms and techniques used in ML for disease detection and decision-making processes.

4. PROPOSED SYSTEM

Modern healthcare has changed how it predicts disease thanks to data-driven techniques and sophisticated tools. Predictive models are gradually being used to identify patterns and risk factors that will help healthcare professionals, researchers, and individuals make educated decisions about disease prevention and management. These decisions are made possible by the abundance of health-related data that is now available, including genetic information, electronic health records, lifestyle choices, and environmental factors. Users can search for illness diagnostic and treatment information in this proposed work by entering symptoms as a query into the search engine. The symptoms are pre-processed to aid in the discovery of keywords that aid in the rapid diagnosis of the condition. To forecast the disease, a linear regression technique is given, in which the system receives a medical dataset as input and produces a prediction of whether or not the patient has a certain condition. The system is trained using a huge dataset and illness labels, with features taken from medical images and used to train a Linear Regression model. The Linear Regression model learns to recognize patterns in characteristics linked with various diseases, and once trained, it may be used to forecast disease.

5. MODULE DESCRIPTIONS

5.1 DATA PREPROCESSING SEMANTIC WORD ANALYSIS

The initial stage in disease prediction is data pre-processing, which encompasses the purification and organization of data for analysis. This may entail eliminating extraneous information, rectifying inaccuracies, and standardizing the data. Semantic word analysis is a form of data pre-processing that concentrates on the significance of words and phrases. It can be employed to recognize and extract crucial keywords from the data, which can subsequently enhance the precision of disease prediction.

5.2 DISEASE IDENTIFICATION

The second stage of predicting a disease is to identify the disease itself by analysing the patient's medical history and symptoms. This can be accomplished through various techniques, including machine learning, natural language processing, and knowledge-based systems.

5.3 WEIGHT ASSIGNING

The third stage in predicting diseases is the allocation of weight. This process entails assigning weights to the various diseases that have been detected to indicate the probability of each disease. The weights are generally determined by the patient's symptoms, medical history, and other relevant factors.

5.4 PATTERN MATCHING

The fourth stage in disease prediction is pattern matching, which entails comparing the symptoms exhibited by the patient with a database containing established disease patterns. Various techniques can be employed for this purpose, including string

matching, fuzzy matching, and graph matching.

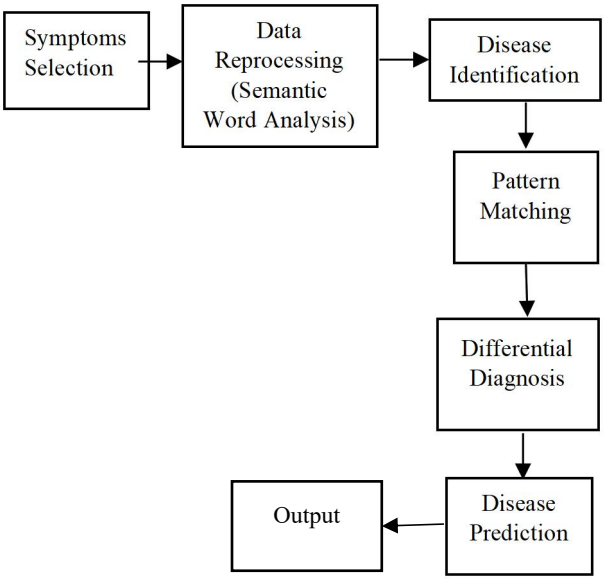

Figure 3. Block Diagram

5.5 DIFFERENTIAL DIAGNOSIS

Differential diagnosis is the final step in disease prediction. This involves narrowing down the list of possible diseases to the most likely one. This is done by considering the patient's symptoms, medical history, and the results of the previous steps.

6. RESULT ANALYSIS

Comparing two methods, convolutional neural network (CNN) and linear regression (LR), the accuracy of the CNN algorithm is 75%, while the LR method's is 81%. This demonstrates that within the given dataset and problem domain, LR outperforms CNN in terms of prediction accuracy. The outcomes emphasise how crucial it is to select the appropriate algorithm, with LR exhibiting a higher ability to recognise and capture the underlying patterns in the data, leading to more accurate predictions. It's important to keep in mind that accuracy might not be sufficient to provide a comprehensive assessment of model performance on its alone. Further examination, considering variables such as accuracy, recall, and other relevant measures, could offer a more complex view of the algorithms' capacity to manage the specific challenges of the assigned task.

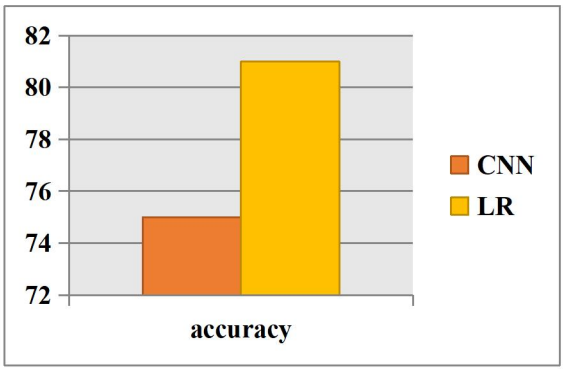

Figure 4. Comparison graph

CONCLUSION

Disease prediction systems have emerged as powerful tools in healthcare, revolutionizing the industry through early detection, personalized medicine, resource allocation, and research acceleration. By utilizing advanced technologies and data-driven approaches, these systems can detect patterns and risk factors that contribute to disease development, allowing for timely interventions, improved patient outcomes, and reduced healthcare costs. To guarantee dependability, accuracy, and efficacy in clinical settings, implementing these systems necessitates careful consideration of input design, output design, system testing, and implementation techniques. As these systems continue to evolve and incorporate more advanced machine learning techniques, their impact on healthcare will inevitably expand, empowering individuals to take control of their health and fostering a healthier and more resilient population.

FUTURE ENHANCEMENT

To enhance the precision and dependability of learning models, it is imperative to

obtain access to superior quality and larger datasets. Additionally, novel machine learning algorithms that cater to healthcare applications must be developed. Ensuring the interpretability of machine learning models is crucial to enable clinicians to comprehend the reasoning behind the predictions and hmachineave faith in the outcomes. Personalized medicine can be achieved by developing machine learning models that are customized to individual patients, taking into account their medical history, genetic profile, and lifestyle

REFERENCES

[1] Pooja Yadhav, S.C. Sharma, Rajesh Mahadeva, Shashikant P. Patole, "Examining Hyperparameters And Feature Selection Methods for Anticipating Non-Communicable Chronic Diseases With a Stacking Classifier", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no.11 , Iss no. 2, pp. 80030-80055, January, 2023.

[2] L. Abirami, J. Karthikeyan, "A Digital Twin-Based Healthcare System (DTHS) For Early Parkinson's Disease Detection And Diagnosis Through An Optimized Fuzzy-Based K-Nearest Neighbor Classifier Model", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 11, Iss no. 5, pp. 96661–96672, September, 2023.

[3] SaraGhorashi, Khunsa Rehman, Anam Riaz, H.K. Alkahtani, Amna Parveen, Ahmed H. Samak, "Utilizing Regression Analysis For Anticipating Concurrent Symptoms Of Cardiovascular Diseases", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 11, Iss no. 5, pp. 60254-60266, June, 2023.

[4] Vijay Panwar, Naved Choudhary, Sonam Mittal, Gaurav Sahu, "Streamlined Medical Diagnosis Of Human Heart Diseases Through Machine Learning Techniques With And Without Gridsearchcv", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 10, Iss no. 4, pp. 80151-80173, April, 2022.

[5] P.G. Shynu, Varun G.Menon, R.L. Kumar, Yunyoung Nam, "Blockchain Enabled Secure Healthcare Application for Diabetic-Cardio Disease Prediction In Fog Computing Environment", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 9, Iss no. 9, pp. 45706-45720, March, 2021.

[6] A. Noor, N. Javaid, N. Alrajeh, B. Mansoor, A. Khaqan and S. H. Bouk, "Heart Disease Prediction Using Stacking Model With Balancing Techniques and Dimensionality Reduction, IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol. 11, pp. 116026-116045, October, 2023.

[7] A. Rahim, Y. Rasheed, F. Azam, M.W. Anwar, M.U. Rahim, Abdul W. Muzaffar, "A Unified Machine Learning Framework For Enhanced Prediction Of Cardiovascular Diseases", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 9, Iss no. 1, pp. 106575-106588, July, 2021.

[8] P. Chittora et al., "Prediction of Chronic Kidney Disease - A Machine Learning Perspective", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol. 9, Iss no. 2, pp. 17312-17334, January, 2021.

[9] S. Akter, Ahsan Habib, MD. Ashiqul Islam, W.A. Fahim, Manik Ahmed,

"Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Diseases", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 9, Iss no. 1, pp. 165184-165206, November, 2021.

[10] T. Rahman et al., "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol. 9, Iss no. 1, pp. 120422-120441, August, 2021.

[11] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol. 8, Iss no. 4, pp. 107562-107582, June, 2020.

[12] Syed J. Pasha, E. Syed Mohamed, "Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 8, Iss no. 3, pp. 184087-184108, September, 2020.

[13] W. Hong, Z. Xiong, Nannan Zheng, Y. Weng, "A Medical History-Based Potential Disease Prediction Algorithm", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 7, Iss no. 4, pp. 131094-131101, September, 2019.

[14] K. Yang et al., "HerGePred: Heterogeneous Network Embedding Representation for Disease Gene Prediction", IEEE Journal of Biomedical and Health Informatics, vol. 23, Iss no. 4, pp. 1805-1815, July 2019.

[15] Rajeswari P, And Reena G.S, "Analysis of Liver Disorders Using Data Mining Algorithms", Global Journal of Computer Science and Technology, vol no. 10, Iss no. 2, pp. 48-52, January, 2019.

[16] Iyer A, Jeyalatha S, and Sumbaly R, "Diagnosis of Diabetes Using Classification Mining Technique", International Journal of Data Mining and Knowledge Management Process, vol no. 5,Iss no. 1, pp. 1-14, Feburary, 2019.

[17] Sanjay Kumar, Sarthak Katyal, "Effective Analysis And Diagnosis Of Liver Disorders Through Data Mining", IEEE Access : Multidisciplinary : Rapid Review : Open Access Journal, vol no. 5, Iss no. 2, pp. 1047–1051, September, 2018.