Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1719

Predictive Modeling for Multifaceted Diseases: A Comprehensive Review

Kshama S B[1]; Ananya Dixit[1]; Azra Rumana[1]; and Harshini K[2]
Information Science and Engineering , BMS Institute of Technology and Management, Bengaluru, India

Abstract:- Electronic data has accumulated due to the rising incidence of chronic illnesses, the complexity of the relationships between various diseases, and also the widespread use of computer-based technologies in sector of health care. Doctors are encountering challenges in accurately diagnosing illnesses and analysing symptoms due to extensive volumes of data. In many of the reviews of the present medical service frameworks, the focus was on considering one disease at a time. The majority of severe articles focus on a certain illness. These days, the inability to identify the precise infection has led to an increase in mortality. Indeed, a previously recovered patient might experience reinfection with another illness. Algorithms in machine learning (ML) have demonstrated substantial capability in outperforming traditional systems for diagnosing diseases, playing a pivotal role in assisting medical professionals in the early identification of elevated-risk diseases. In this literature, the intention is to identify patterns across different types of supervised and unsupervised ML models in disease detection by assessing performance metrics.

Keywords:- *Chronic Diseases, Disease Detection, Machine Learning, Performance Metrics.*

I. INTRODUCTİON

The capacity to accurately and efficiently forecast and diagnose diseases is still a major concern in the rapidly changing field of healthcare. Conventional diagnosis techniques often rely heavily on clinical expertise and subjective assessment making them time-consuming, error-prone, and sometimes fail to capture the intricacies of many illness situations. With its capacity to analyse massive volumes of data and spot patterns, machine learning (ML) has become a potent tool for overcoming these constraints. ML-powered multiple disease prediction holds immense potential to revolutionise healthcare by enabling early identification where ML models can examine various data sources, including medical records, genetic information, and environmental factors, to pinpoint individuals who are prone to developing various health conditions. This advance notice enables for proactive interventions, potentially preventing disease onset or delaying its progression, culminating in improved patient outcomes and decreased healthcare costs.

Personalized Treatment by considering individual risk profiles and specific disease combinations, ML models can help tailor treatment plans to optimise effectiveness and minimise adverse effects. This personalised approach to healthcare ensures that patients obtain the most suitable care for themselves in unique situations, resulting in improving health results and enhancing living standards.

The ability to forecast multiple diseases simultaneously brings about numerous advantages across various facets of healthcare. Currently, many individuals Experience numerous chronic conditions, creating a complex landscape for diagnosis, treatment, and resource allocation.

Multiple disease prediction can address this challenge in several ways:

- Tackling the Burden of Multiple Diseases: By predicting multiple diseases early, healthcare systems can intervene proactively, potentially preventing the onset or delaying the progression of illnesses. This can enhance patient concequences and decrease healthcare costs, and alleviate the burden on individuals and families.
- Understanding Disease Comorbidities: Multiple diseases regularly interact with each other, making diagnosis and treatment more complex. By analysing relationships between different diseases, we can develop a deeper understanding of these interactions and design more effective treatment strategies.
- Personalized Medicine and Precision Healthcare: Predicting multiple diseases allows for a more personalised approach to healthcare. By considering an individual's specific risk factors and disease profile, we can tailor treatment plans to optimise outcomes and minimise adverse effects.
- Resource Optimization and Cost-Effectiveness: Predicting multiple diseases can help optimise resource allocation within healthcare systems. By identifying individuals at high risk for multiple illnesses, resources should be directed at those who require them the most., maximising efficiency and minimising costs.
- Public Health Surveillance and Disease Prevention: Early prediction of multiple diseases can contribute significantly to public health efforts. By monitoring trends and identifying populations at high risk, interventions can be implemented to prevent outbreaks and improve population health outcomes.
- Drug Discovery and Development: Understanding the connection between different diseases can lead to the development of more potent drugs and therapies. By targeting multiple pathways simultaneously, we can improve treatment efficacy and address the complex challenges of multi-morbidity.
- Systems for Supporting Clinical Decision-making: The integration of prediction models assists healthcare

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1719

professionals in making clinical decisions in diagnosing, planning treatments, and assessing risks. This integration promotes more informed decision-making and enhances the quality of patient care.

- Patient Empowerment and Self-Management: Empowering individuals with information about potential health risks through multiple disease predictions enables them to take a more proactive approach in overseeing their well-being. This can lead to healthier lifestyles, improved self-care, and reduced healthcare utilisation.
- Connecting Research and Clinical Practice: Multiple disease prediction models can serve as a link between research and clinical practice. By transforming research findings into practical tools, we can accelerate the implementation of new knowledge into the clinical setting and improve patient care.
- Ethical Considerations and Future Directions: While multiple disease prediction offers immense potential, ethical considerations must be addressed. These include data privacy, informed consent, and potential biases in algorithms. As we progress, it is essential to develop ethical guidelines and frameworks to guarantee the responsible and fair utilization of this technology.

Multiple disease prediction holds immense promise for revolutionising healthcare by improving diagnosis, treatment, prevention, and resource allocation. By addressing the ethical considerations and fostering collaboration between researchers, clinicians, and policymakers, we can unlock the full potential of this technology to improve health outcomes for all.

Machine Learning is a Promising Solution. Machine learning algorithms have exhibited exceptional ability in examining intricate datasets, identifying significant patterns, and formulating forecasts. Multiple disease prediction benefits greatly from their capacity to handle vast amounts of patient data, incorporating medical history, symptoms, test outcomes, and genetic information. ML systems can detect subtle correlations between diseases and their underlying risk factors by gaining knowledge from these data sources.

II. APPROACHES TO MULTİPLE DİSEASE PREDİCTİON

ML-based multiple illness prediction encompasses various methodologies, each with unique advantages and disadvantages. Utilising algorithms for supervised learning is one typical strategy. These algorithms utilise the patterns they have learnt to forecast the disease status of new patients. They first learn from labelled data, where the disease condition of each patient is known.

Algorithms are employed for unsupervised learning in an alternative approach, operating without the need for labelled samples. They identify patterns and clusters within the data, making unsupervised learning particularly valuable for exploring data and uncovering potential disease clusters.

III. LİTERATURE SURVEY

Sl No.	Title of the paper	Description	Methodology	Observation
[1]	Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare(IEEE)	The goal of the paper is to create a model that uses machine learning techniques to predict diabetes.	● Naive Bayes ● SVM ● Random Forest ● Simple CART	Compared to the other algorithms that are employed, the SVM method produced the best results, according to the authors.
[2]	Disease Prediction using Machine Learning Algorithms (IEEE)	The paper demonstrates the prediction of diseases such as Malaria, Dengue, Impetigo, etc	● Decision Tree ● Random Forest ● Naive Bayes	The description of patient's symptoms may lack accuracy, indicating the presence of overfitting.
[3]	Comparison of Machine Learning Models for Parkinson's Disease Prediction	An evaluation of the performance analysis includes five models designed for Parkinson's disease. performance analysis.	● Logistic Regression ● Naive Bayes ● Decision Tree ● Random forest	The bagging classifier shows signs of overfitting, as it achieves a high training accuracy of 98.5%, but its test accuracy is notably lower at 91.5%.
[4]	Heart Disease Prediction using Machine Learning	The precision of ML techniques in forecasting heart disease was demonstrated in this paper.	● KNN ● Decision tree ● Linear Regression ● SVM	Scope limited to Heart disease.
[5]	Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS	This paper examined health sensor data using EPS and AI techniques, concentrating on the early-stage prediction of risks associated with heart and diabetes diseases.	● Naive Bayes ● Random Forest ● Decision tree ● Logistic Regression	As SVM showed lower accuracy, Regression and Random forest were considered to give the precise results.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1719

[6]	Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings	This paper has an accurate diagnosis of PD using ML based prediction system. It also uses feature selection and classification using Voice recording data.	● L1-norm ● SVM	The analysis of experimental results effectively demonstrates the system's ability to classify between individuals with Parkinson's disease and those who are healthy.
[7]	Multiple Disease Prediction	A web-based application where different machine learning algorithms are utilized for predicting probability of multiple diseases based on user-provided medical information. Their scope of diseases included diabetes, parkinsons , heart and liver	● KNN ● Random forest ● XGBoost	The application offers a user-friendly interface, multiple disease prediction, and accessibility.
[8]	Early Warning Signs Of Parkinson's Disease Prediction Using Machine Learning Technique	The paper proposes an ML based approach for temely identification of Parkinson's disease (PD) using non-motor symptoms.	● SVM ● Decision Tree	The proposed approach offers early detection of PD
[9]	Multiple Disease Prediction using Machine Learning and Deep Learning with the implementation of Web Technology	The paper proposes a web application that utilises Machine learning and deep learning techniques utilized to forecast the probability of various illnesses based on user-provided medical information. The system predicted for diabetes, heart and kidney	● Random Forest, ● Naive Bayes, ● SVM	User-friendly interface, multiple disease prediction, and accessibility
[10]	An effective Parkinson's disease prediction using logistic decision regression and machine learning with big data	The paper discusses the anticipation of Parkinson's disease. The authors proposed a method that uses a dataset to train the model.	● LDR	The authors found that the LDR model was able to predict PD and also suggest that the method has the potential to be used to accurately diagnose the disease.
[11]	Diabetes Disease Prediction Using Machine Learning	This research introduces the approach for diabetes prediction using ML algorithms. Using a patient dataset, these algorithms' performance was assessed.	● SVM ● Logistic Regression ● KNN	It was found that the SVM algorithm achieved the best performance than alternative algorithms in use.
[12]	Logistic regression technique for prediction of cardiovascular disease	The research examines the application of logistic regression (LR), a statistical approach used to forecast outcomes that are either yes or no in the presence or absence of cardiovascular disease (CVD).	● Logistic Regression	The authors used LR to examine the connections between a quantity of risk factors, blood pressure, cholesterol, smoking status, age, gender, and the existence or non-existence of CVD.
[13]	Heart Disease Prediction Using Logistic Regression	To assess how well logistic regression (LR), a statistical technique for forecasting binary outcomes, predicts the probability of heart disease with different heart disease risk variables.	● Logistic Regression	The authors used LR to examine the connections between a quantity of risk factors encompassing blood pressure, cholesterol, blood sugar, age, gender, and chest discomfort.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1719

[14]	Logistic Regression and SVM-based Diabetes Prediction System	In this study, researchers propose the use of machine learning algorithms to assess how accurately certain risk factors predict the likelihood of an individual developing diabetes.	● Logistic Regression ● SVM	They attribute this difference to SVM's ability to handle non-linear relationships between risk factors and diabetes.

The reviews have explored the numerous studies which illustrate the utilisation of various ML algorithms, including Logistic Regression, K-Nearest Neighbours, and employing Support Vector Machines (SVM) and Deep Learning models for predictionan extensive array of ailments like diabetes, cancer, cardiovascular disease, and Parkinson's disease. These studies have demonstrated the promise of ML in achieving high prediction accuracy, with some models achieving performance exceeding conventional diagnostic methods. However, challenges remain, including data quality and accessibility, etc.

The study focuses on the integration of machine learning in the healthcare domain. It utilizes four classifiers employing MLalgorithms, including Simple CART, SVM, Naive Bayes and Random Forest. The research conducts experiments using the WEKA which predicts the occurrence of Diabetes disease.The classifiers are compared based on the time taken for training and testing, and accuracy values. The dataset comprises 9 features and 768 instances, resulting in accuracies of 77% for Naive Bayes, 79% for SVM, 76.9% for Random Forest, and 76.5% for Simple CART. Notably, the analysis of overall performance suggests that the Support Vector Machine outperforms Naive Bayes, Random Forest, and Simple CART in forecasting diabetes disease [1]. Challenges encountered during the use of SVM and Simple CART included sensitivity to high dimensionality, difficulties with non-linear datasets, the black box nature of the models, hyper parameter tuning, overfitting, high variance, limited feature importance interpretation, and unsuitability for continuous data. These challenges also encompassed computational expenses, underscoring the various factors that need to be considered while applying these algorithms in the sector of health care.

In addressing the challenges previously mentioned, an exhaustive examination of Disease Prediction using ML Algorithms was undertaken [2]. This paper proposes the training of a machine learning model using Decision Trees, a technique that involves systematically partitioning the dataset into progressively smaller subsets to predict the target value, i.e., the disease. Furthermore, the study incorporates the application of the Naive Bayes algorithm, recognized for its simplicity in implementation, speed, efficiency, and versatility in handling both continuous and discrete data. In this study, the dataset consists of 132 symptoms and 4920 instances, covering 41 diseases. Impressively, employing this approach resulted in an accuracy of 95.12% for Decision Trees, 95% for Random Forest, and 95.12% for Naive Bayes. This highlights the efficacy of the methodology in predicting diseases.

A study akin to [2], titled "Comparison of Machine Learning Models for Parkinson's Disease Prediction" [3], was critically examined. This research introduced an additional methodology involving Logistic Regression, recognized as a potent and versatile tool for disease prediction, offering advantages regarding interpretability, efficiency, and robustness. The dataset utilised in this investigation comprised 31 features and 4290 instances. The outcomes demonstrated notable accuracies, with Logistic Regression achieving 89.83%, Naive Bayes also at 89.83%, Decision Tree at 93.22%, and Random Forest, which gave an accuracy rate of 94.92%[3]. This comparison underscores the diverse strengths and performance nuances in the context of machine learning models predicting Parkinson's disease.

Researchers have delved into data mining applications to predict heart diseases, as evidenced by their work on "Heart Disease Prediction Using Machine Learning Algorithms" [4]. This investigation elucidates the extraction of intriguing patterns and knowledge from extensive datasets. The study meticulously compares and evaluates the precision of various data mining and machine learning approaches to find the most efficient one. Ultimately, the results favour the KNN (K-Nearest Neighbors) method as the optimal choice for predicting heart diseases. The dataset included 14 features, this gave a level of accuracy 87% for KNN, 79% for Decision tree, 78% for linear regression and 83% for SVM.

The paper [5] focuses on the relatively underexplored realm of healthcare—specifically, the early-stage risk prediction of Non-Communicable Disease (NCD) through wearable technology within the Healthcare Common Procedure Coding System. Employing methodologies akin to those in reference [3], the dataset incorporated 17 features and comprised 520 instances. Notably, the Naive Bayes algorithm demonstrated a remarkable accuracy of 81%.

Researchers conducted an examination of Parkinson's disease, referenced as [3], focusing on L1-Norm SVM and an Effective Recognition System utilising voice recordings [6]. The system's development involved employing a SVM as a machine learning classifier to differentiate between individuals with Parkinson's disease (PD) and those who are healthy. Feature selection utilised the L1-Norm SVM to identify pertinent and characteristics that exhibit a strong correlation essential for accurate PD and healthy classification. The dataset encompassed 22 features, yielding an accuracy of 94% for DBN (Deep Belief Network), 99% for L1 Norm, and 99% for SVM.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1719

The study titled "Multiple Disease Prediction Using Machine Learning Algorithms" [7] incorporates the KNN approach [4], Random Forest [5], and XGBoost. While XGBoost stands out as a resilient and widely applied algorithm in supervised learning, it does come with certain drawbacks, such as computational complexity, susceptibility to imbalanced datasets, and a reliance on meticulous pre-processing. The dataset used in the study comprises 2359 features and 1683 instances. The predictive accuracies achieved were 85% for Diabetes, 82% for Heart disease, 78% for Parkinson's, and 75% for Liver disease.

In the study documented in [8], the authors directed their attention towards predicting Parkinson's disease, extending the groundwork established in a preceding investigation [6]. Their exploration delved into assessing the effectiveness of two ML algorithms—specifically, the SVM and Decision Tree—previously examined in another study [4]. The dataset employed for Parkinson's disease analysis in this research encompassed 44 distinct features, drawing from an extensive collection of 240 speech measurements. Notably, the achieved predictive accuracy was remarkable, with the SVM algorithm reaching an impressive 90%, and the Tree classifiers achieving 84%. This research provides great perspectives to the domain of applying ML in predicting Parkinson's disease, highlighting the effectiveness of SVM and Decision Tree models in this particular domain.

In another comprehensive review, the primary objective was to develop a user-friendly web application enabling accurate and simultaneous prediction of multiple diseases. The approach ingeniously integrates various disease detection techniques, eliminating the necessity for additional websites or software [9]. Employing Random Forest, Naive Bayes, and SVM [1][2][3][4][5], the dataset encompassed varying features, ranging from 9 to 24, and instances ranging from 195 to 5110. The researchers demonstrated notable success, achieving an accuracy of 88% for Diabetes, 85% for Heart disease, and 82% for Kidney disease.

For diagnosing Parkinson's disease, specific tests like blood tests or ECGs are typically required. To address the complexity of classifying Parkinson's disease, the recommended approach involves utilising ML based algorithm called Logistic Decision Regression (LDR) [10]. LDR amalgamates the advantages of Logistic Regression (LR) and Decision Trees, offering interpretability, the ability to handle both continuous and discrete data, and efficiency for large datasets. The researchers proposed a dataset consisting of 22 features and 31 instances to train the LDR model. Notably, the study revealed that the LDR model achieved a remarkable accuracy of 93.5% in predicting Parkinson's disease.

In our extensive literature review [11], we explored various algorithms such as SVM, LR, and KNN, noting a limitation in their application owing to the absence of a varied dataset. To address this, we enhanced prediction capabilities by incorporating two additional algorithms, Naïve Bayes and K-Nearest Neighbors, and conducted an evaluation based on their success factors to determine which algorithm exhibited superior accuracy. This study involved the analysis of 768 instances, with 268 classified as positive and the rest as negative. The SVM algorithm demonstrated an impressive accuracy of 92.13%, while the LR, KNN, and DT algorithms achieved accuracies of 89.55%, 87.92%, and 85.05%, respectively.

The study [12] investigates the utilisation of logistic regression (LR), a statistical technique also examined in previous works [3][5][11], for predicting binary outcomes in the context of determining the existence or nonexistence of cardiovascular disease (CVD). The dataset comprised 303 instances. LR demonstrated an accuracy of 87.10% in successfully predicting the presence or absence of CVD in the subjects.

In the contemporary healthcare sector, machine learning is commonly employed to detect diseases and predict their occurrences through data modelling. In studies focusing on risk assessment in intricate scenarios, such as heart disease, the Logistic Regression ML algorithm is notably prevalent [13]. The authors utilised the Cleveland Heart Disease Dataset, encompassing 303 cases with diverse factors contributing to the risk of heart disease. The achieved accuracy in their analysis was 85.47%.

Evidence has demonstrated that simpler systems, such as Logistic Regression and SVM, has the potential to generate more precise outcomes than their more complex counterparts. Another recent review [14] parallels the investigation in [11], incorporating the widely recognized Pima Indian Diabetes Dataset comprising 768 instances involving diverse risk factors for diabetes. In the realm of diabetes prediction, researchers found that SVM outperformed LR, achieving an accuracy rate of 79% compared to LR's 75%.

IV. CONCLUSİON

Early identification of diseases not only extends life expectancy but also helps avoid financial hardships. A multi-disease prediction model facilitates the simultaneous prediction of various illnesses.

The application of diverse utilization of machine learning algorithms has facilitated the precaucisous detection of numerous conditions, including diabetes, Parkinson's, breast, brain, and kidney diseases. Amidst the widely recognized algorithms for predictive modelling in the literature, SVM and Logistic Regression (LR) emerge as the best prominent, with accuracy being a pivotal performance metric.

The SVM model distinguishes itself for its consistent superiority in accuracy, especially in handling high-dimensional, semi-structured, and unstructured data, making it particularly effective in predicting Parkinson's and diabetes. Notably, upon concluding the assesment, regression stands out as the algorithm with the utmost accuracy in predicting heart disease.. Early identification,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1719

facilitated by these predictive models, proves crucial in enhancing life expectancy and averting financial challenges associated with disease management.

REFERENCES

[1]. Ayman Mir, Sudhir N Dhage "Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare" (IEEE) 2019

[2]. Amin Ul Haq, Jian Ping Li,Moahmmad Hammad Memon,Jalaluddin Khan, Asad Malik, Tanvir Ahmed "Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings" (IEEE) 2019

[3]. Sneha GramPurohit, Chetan Sagarnal "Disease Prediction using Machine Learning Algorithms" (IEEE) 2020

[4]. Tapan Kumar, Pradyumn Sharma, Nupur Prakash "Comparison of Machine learning models for Parkinson's Disease prediction" (IEEE) 2020

[5]. Archana Singh, Rakesh Kumar "Heart Disease Prediction using Machine Learning" (IEEE) 2020

[6]. Mohammed Juned Shaikh, Soham Manjrekar "Multiple Disease Prediction" (IEEE) 2020

[7]. Kranthi Kumar Singamaneni Dr.G.Putlibai ,Dr.P,Sagaya Aurelia, P Gopala Krishna, Dr.D.StalinDavid "An effective Parkinson's disease prediction using logistic decision regression and machine learning with big data" 2021.

[8]. Rahatara Ferdousi, M Anwar Hossain, AbdulMotaleb El Saddik "Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS" (IEEE) 2021

[9]. Pawan Kumar Mall , Rajesh Kumar Yadav, Arun Kumar Rai , Vipul Narayan , Swapnita Srivastava "Early Warning Signs Of Parkinson's Disease Prediction Using Machine Learning Technique" 2022

[10]. Mostafizur Rahman; Saiful Islam; Sadia Binta Sarowar; Meem Tasfia Zaman "Multiple Disease Prediction using Machine Learning and Deep Learning with the Implementation of Web Technology" (IEEE) 2023

[11]. D. Bertsimas, L. Mingardi and B. Stellato " Machine Learning for Real-Time Heart Disease Prediction" (IEEE) 2021

[12]. M. A..Sarwar, N. Kamal, W. Hamid and M. A. Shah "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare". (IEEE) 2019

[13]. T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi, "Parkinson's Disease Diagnosis Using Machine Learning and Voice". (IEEE) 2019

[14]. 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) 2020

[15]. K.G.Dinesh, K.Arumugaraj, K.D. Santhosh and V. Mareeswari ,"Prediction of Cardiovascular isease Using Machine Learning Algorithms," (IEEE) 2019

[16]. Mehrbakhsh Nilashi, Othman bin Ibrahim, Hossein Ahmadi, Leila Shahmoradi "An analytical method for diseases prediction using machine learning techniques" 2019

[17]. Prashant Kumbharkar, Deepak Mane, Santosh Borde Sunil Sangve "Diabetes Disease Prediction Using Machine Learning Algorithms" 2022

[18]. A. L. Yadav, K. Soni and S. Khare, "Heart Diseases Prediction using Machine Learning" (IEEE) 2023

[19]. Rubini P. E., Dr. C. A. Subasini, Dr. A. Vanitha Katharine, V. Kumaresan, S. Gowdham Kumar, T. M. Nithya "A Cardiovascular Disease Prediction using Machine Learning Algorithms" 2021

[20]. Richa Mathur, Vibhakar Pathak Devesh Bandil "Parkinson Disease Prediction Using Machine Learning Algorithm" 2019