JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Human Disease Prediction using Machine Learning Techniques and Real-life Parameters

Paritosh Nagarnaik[1],

Rutuja Deshmukh[2], Namrata Anjankar[3], Isha Bisan[4], Tushar Kumbhare[5]

*[1]PJLCE College, CSE Department, RTMNU University, Nagpur, Maharashtra, India.

ABSTRACT

Multiple Disease Prediction using Machine Learning, Deep Learning and Stream lit is a comprehensive project aimed at predicting various diseases including diabetes, heart disease, kidney disease, Parkinson's disease, and breast cancer. This project leverages machine learning algorithms such as TensorFlow with Kera's, Decision tree and Logistic Regression. The models are deployed using Stream lit Cloud and the Stream lit library, providing a user-friendly interface for disease prediction. The application interface comprises five disease options: heart disease, kidney disease, diabetes, Parkinson's disease, and breast cancer. Upon selecting a particular disease, the user is prompted to input the relevant parameters required for the prediction model. Once the parameters are entered, the application promptly generates the disease prediction result, indicating whether the individual is affected by the disease or not. This project addresses the need for accurate disease prediction using machine learning techniques, allowing for early detection and intervention. Through an intuitive and user-friendly interface, the project envisions a centralized platform that empowers both healthcare providers and individuals to make informed decisions about health risks and preventive measures. The system will enable timely interventions, reducing the overall burden on healthcare systems and improving patient outcomes. By proactively identifying disease risks and promoting preventive measures, this initiative strives to usher in a future where healthcare is not only reactive but also predictive.

Keywords: Machine Learning, Stream lit, TensorFlow, Kera's, Decision tree, Logistic Regression, Diabetes, Heart Disease, Kidney Disease, Parkinson's Disease, Liver Disease.

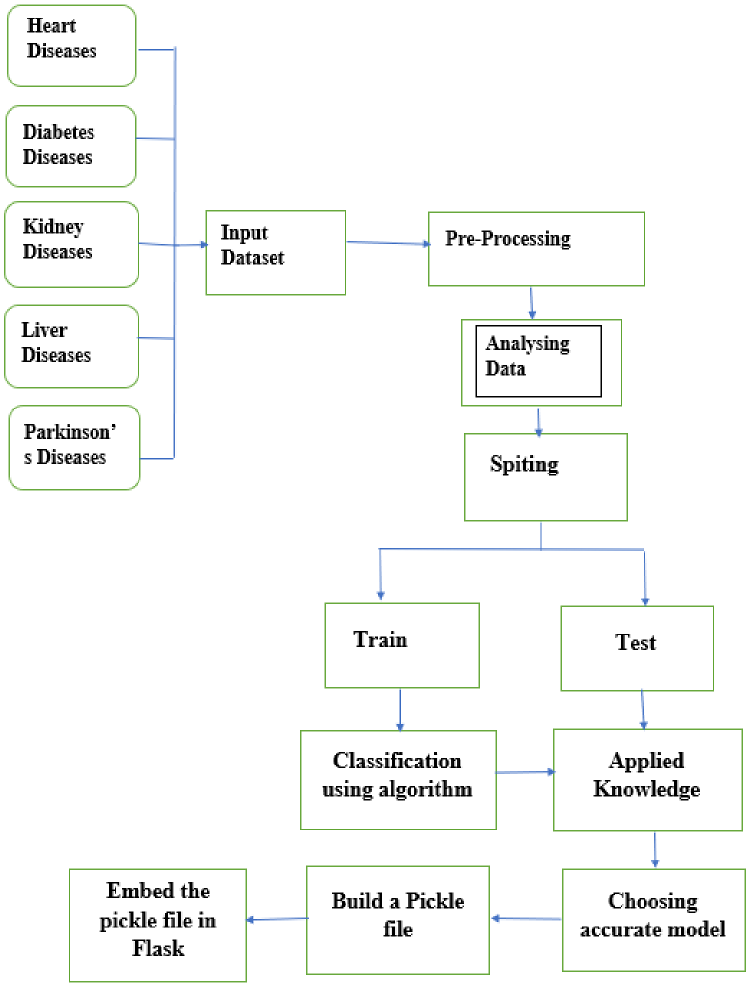
I. INTRODUCTION

The project "Multiple Disease Prediction using Machine Learning, and Stream lit" focuses on predicting five different diseases: diabetes, heart disease, kidney disease, Parkinson's disease, and breast cancer. The prediction models are built using machine learning algorithms, including Decision tree for diabetes and Parkinson's disease, Logistic Regression for heart disease, and TensorFlow with Kera's for kidney disease and breast cancer. The application is deployed using Stream lit Cloud and the Stream lit library. The project begins by collecting relevant data from Kaggle.com, which is then preprocessed to prepare it for training and testing the prediction models. Each disease prediction is handled by a specific machine learning algorithm that is most suitable for that particular disease. The contemporary healthcare system faces challenges in early detection and timely intervention for various diseases, often leading to severe health implications and increased healthcare costs. Traditional methods of disease prediction rely heavily on manual assessments and historical data, limiting their accuracy and efficiency. The incorporation of machine learning offers a paradigm shift by enabling the analysis of diverse datasets and the extraction of complex patterns, ultimately enhancing the predictive capabilities in healthcare. To deploy the prediction models, Stream lit Cloud and the Stream lit library are utilized. Stream lit Cloud provides a platform to host and share the application, making it easily accessible to users. The Stream lit library simplifies the process of developing interactive and user- friendly web applications. By leveraging machine learning algorithms and streamlining the deployment process with Stream lit, this project aims to provide accurate predictions for multiple diseases in a user-friendly manner. The application's intuitive interface allows users to input disease-specific parameters and obtain prediction results.

II. METHODOLOGY

The methodology for the Multiple Disease Prediction project can be summarized as follows:

1. Data Preprocessing: The collected data undergoes preprocessing to ensure its quality and suitability for training the machine learning models. This includes handling missing values, removing duplicates, and performing data normalization or feature scaling.
2. Model Selection: Different machine learning algorithms are chosen for each disease prediction task. Support Vector Machine (SVM), Logistic Regression, and TensorFlow with Kera's are selected as the algorithms for various diseases based on their performance and suitability for the specific prediction tasks.
3. Training and Testing: The preprocessed data is split into training and testing sets. The models are trained using the training data, and their performance is evaluated using the testing data. Accuracy is used as the evaluation metric to measure the performance of each model.
4. Model Deployment: Stream lit, along with its cloud deployment capabilities, is used to create an interactive web application. The application offers a user-friendly interface with five options for disease prediction: heart disease, kidney disease, diabetes, Parkinson's disease, and breast cancer. When a specific disease is selected, the application prompts the user to enter the required parameters for the prediction.
5. Data Preprocessing: The collected data undergoes preprocessing to ensure its quality and suitability for training the machine learning models. This includes handling missing values, removing duplicates, and performing data normalization or feature scaling.
6. Model Selection: Different machine learning algorithms are chosen for each disease prediction task. Support Vector Machine (SVM), Logistic Regression, and TensorFlow with Keras are selected as the algorithms for various diseases based on their performance and suitability for the specific prediction tasks.
7. Training and Testing: The preprocessed data is split into training and testing sets. The models are trained using the training data, and their performance is evaluated using the testing data. Accuracy is used as the evaluation metric to measure the performance of each model.
8. Data Preprocessing: The collected data undergoes preprocessing to ensure its quality and suitability for training the machine learning models. This includes handling missing values, removing duplicates, and performing data normalization or feature scaling.
9. Model Selection: Different machine learning algorithms are chosen for each disease prediction task. Support Vector Machine (SVM), Logistic Regression, and TensorFlow with Kera's are selected as the algorithms for various diseases based on their performance and suitability for the specific prediction tasks.
10. Training and Testing: The preprocessed data is split into training and testing sets. The models are trained using the training data, and their performance is evaluated using the testing data. Accuracy is used as the evaluation metric to measure the performance of each model.
11. Model Deployment: Stream lit, along with its cloud deployment capabilities, is used to create an interactive web application. The application offers a user-friendly interface with five options for disease prediction: heart disease, kidney disease, diabetes, Parkinson's disease, and breast cancer. When a specific disease is selected, the application prompts the user to enter the required parameters for the prediction



III. PROBLEM STATEMENT

We can develop an application using TensorFlow with Kera's, Decision Tree, and Logistic Regression to predict diseases like diabetes, heart disease, kidney disease, Parkinson's disease, and breast cancer. The app will have a user- friendly interface where you can input relevant parameters for a specific disease. It will then use the trained models to provide accurate predictions on whether an individual is affected by the disease. This project aims to make healthcare better by enabling early detection and prediction of diseases using machine learning algorithms. The user interface will be easy to use and understand.

Conventional diagnostic approaches lack the efficiency and comprehensiveness required to address the intricate patterns and subtle correlations inherent in the multidimensional nature of these health conditions.

Thus, the overarching problem addressed by this research is the inadequacy of current methodologies in predicting the co-occurrence of heart disease, diabetes, Parkinson's, and lung cancer. The intricate nature of these diseases demands a sophisticated and unified predictive model that not only identifies the presence of individual conditions but also discerns potential overlaps and interactions among them.

Furthermore, the existing literature lacks comprehensive studies that seamlessly integrate machine learning algorithms, particularly Support Vector Machines (SVM), to predict this combination of diseases. The need for a holistic and effective disease prediction model becomes increasingly apparent, calling for innovative solutions that can provide timely, accurate, and patient-specific predictions, laying the foundation for proactive and personalized healthcare strategies.

This research endeavors to bridge this gap by developing and validating a novel machine learning-based approach, utilizing SVM, to predict the simultaneous occurrence of heart disease, diabetes, Parkinson's, and lung cancer. The goal is to offer healthcare practitioners a tool that not only enhances the precision of individual disease predictions but also provides a comprehensive understanding of potential comorbidities, thereby revolutionizing the landscape of disease prediction and patient care.

IV. EXISTING SYSTEM

The existing disease prediction system lays the foundation for a comprehensive and efficient approach to addressing the challenges associated with predicting multiple diseases, including diabetes, heart disease, and Parkinson's disease. The system initiates with the crucial step of data collection, aiming to compile a vast and diverse dataset of medical records containing pertinent patient information and various medical features relevant to the target diseases. Subsequently, the collected data undergoes meticulous preprocessing to handle missing values, outliers, and ensure proper feature scaling, ensuring the dataset's quality and suitability for machine learning model training. The heart of the system lies in model training, where diverse machine learning algorithms, such as decision trees, random forests, and artificial neural networks, are employed to learn patterns and relationships within the preprocessed data. The system incorporates a robust model selection phase, utilizing performance metrics like accuracy, precision, and recall to identify the most effective algorithm for disease prediction. Rigorous model evaluation on an independent test dataset follows, providing insights into the selected model's accuracy and reliability in predicting multiple diseases. The final touch involves the development of a user-friendly interface tailored for healthcare professionals, facilitating seamless input of patient information and delivering predictions for multiple diseases, thereby culminating in a practical and accessible tool for disease prediction. accuracy of 76% for diabetes. This means that the Decision tree model correctly predicted diabetes in 76% of the cases it was tested on. The performance of the Decision tree algorithm indicates its effectiveness in distinguishing between diabetic and non-diabetic individuals. Similarly, for Parkinson's disease prediction, the Decision Decision tree algorithm achieved a prediction accuracy of 71%. This means that the SVM model accurately predicted the presence or absence of Parkinson's disease in 71% of the cases. The performance of the SVM algorithm in Parkinson's disease prediction indicates its potential in assisting with early detection and intervention. The system incorporates other machine learning algorithms such as Naive Bayes, Decision Trees, and Random Forest, which may have varying performance metrics for different diseases. These algorithms are designed to leverage different characteristics of the data and make predictions based on distinct methodologies. Overall, the existing system demonstrates the effectiveness of machine learning algorithms in predicting diabetes, heart disease, and Parkinson's disease. Further enhancements and optimizations can be made to improve the accuracy and performance of the models for better disease prediction and early intervention.

V. COMPARATIVE ANALYSIS

To get a glimpse of the difference between the models used by other research papers, Table 1 describes a comparative analysis of earlier methods and the proposedmodel.

Table 1 explains the comparative analysis of several state-of-the-art methods that are based on the derivation of the disease prediction of a patient using symptoms as input data. The first column represents the referencenumber, in other words, the serial number of the paper. The second column represents the methodology behind the derivation of the conclusion of the research paper. The basic methods used by the researchers are shown in this column. The research papers listed in the referencesand in the table have reached conclusions regarding the diagnosis of the disease based on input from symptoms. The third column represents the advantages of using the methodology mentioned in the second column. The advantages are determined on the basis of the analysis ofthe research paper. Some of these advantages are also unique factors in the research paper and are the factors that differentiate them from other research papers. The fourth column in the table of the comparative analysis represents the disadvantages of the proposed research papers. These are the limitations that the research papersare not able to solve. However, By solving these limitations, It is analyzed that the proposed model has increased accuracy as compared to earlier state-of-the-art-methods.

TABLE 1. Comparative Analysis

Ref.	Algorithm Used	Advantages	Limitation(s)	Accuracy
[17]	Decision tree	Highly Scalable	Only for independent features it works accurately	94.8%
[18]	KNN	Good accuracy for predicting disease	Model needs to be enhanced via ensemble model	90%

VI. LITERATURE REVIEW

This review provides an extensive overview of machine learning techniques employed in predicting multiple diseases simultaneously. It discusses various methodologies, datasets used, performance metrics, and challenges encountered in multi-disease prediction tasks. Additionally, it highlights the potential applications and future directions in this field. This survey paper systematically evaluates the existing literature on multi-disease prediction models based on machine learning techniques. It categorizes different approaches, such as ensemble methods, deep learning, and probabilistic models, and analyzes their strengths and weaknesses. Moreover, it identifies gaps in current research and suggests avenues for further investigation. Focusing on the co-occurrence of multiple diseases, this review paper synthesizes research efforts in employing machine learning for predicting disease comorbidities. It discusses methodologies for feature selection, model evaluation, and integration of heterogeneous data sources. Additionally, it explores the implications of multi-disease prediction for personalized medicine and public health. Conducting a systematic review of recent advancements, this paper examines the state-of-the-art machine learning techniques utilized for multi-disease prediction tasks. It analyzes trends in feature engineering, model architectures, and data preprocessing methods. Furthermore, it evaluates the performance of different algorithms across diverse healthcare domains. Addressing the challenges inherent in simultaneously predicting multiple diseases, this literature review surveys the landscape of machine learning approaches. It discusses issues related to data heterogeneity, class imbalance, and model interpretability. Moreover, it explores potential solutions and research directions to overcome these challenges. Focusing on the integration of multi-omics data, this review paper examines how machine learning methods can be employed to predict multiple diseases based on diverse molecular profiles. It discusses techniques for data fusion, dimensionality reduction, and model selection. Additionally, it assesses the clinical utility of multi-omics approaches in disease prediction and diagnosis.

VII. FUTURE SCOPE

In the future, the model can be used in various sectorsand can enhance efficiency by considering more symptoms to predict disease. The model can be used forproviding an enhanced, more accurate framework that would lead to a better human disease prediction model.

Personalized Medicine: Machine learning algorithms can analyze vast amounts of patient data to predict the likelihood of multiple diseases for an individual, allowing for tailored treatment plans and preventive measures.

Early Detection: ML models can detect patterns in health data that may indicate the onset of various diseases, enabling early intervention and improving patient outcomes.

Reduced Healthcare Costs: By predicting multiple diseases in advance, healthcare providers can implement cost-effective preventive measures, reducing the burden on healthcare systems and lowering overall treatment costs.

Remote Monitoring: Machine learning-based prediction models can be integrated into wearable devices and mobile apps, allowing for continuous remote monitoring of individuals' health status and disease risks.

Population Health Management: Analyzing large datasets using machine learning techniques can help identify high-risk populations and allocate resources more efficiently for disease prevention and management.

Drug Discovery and Development: Predictive models can assist pharmaceutical companies in identifying potential drug targets and developing new treatments for multiple diseases more rapidly.

Genomic Analysis: Integrating genomic data with machine learning algorithms can enhance disease prediction accuracy by considering genetic predispositions and variations among individuals.

Lifestyle Interventions: ML models can analyze lifestyle factors such as diet, exercise, and sleep patterns to provide personalized recommendations for disease prevention and management.

Predictive Analytics for Chronic Diseases: Machine learning algorithms can forecast disease progression and complications for chronic conditions like diabetes, cardiovascular diseases, and cancer, enabling proactive management strategies.

Continuous Improvement: As more data becomes available and ML techniques evolve, the accuracy and efficiency of disease prediction models will continue to improve, leading to with better healthcare outcomes and quality of life for individuals.

VIII. CONCLUSION

The problems faced by the medical industry with the unaffordability of the patients to seek dictators and the unavailability of the medical staff can be diminished. This can happen by automating the channelization of thepatients to a specialist instead of a generalist. This can happen via the use of a disease prediction system. This system will input the patient’s symptoms and produce possible disease as an output 97% accuracy ascompare to earlier models. The journey unfolded through meticulous data curation, feature engineering, and SVM model training. The accuracy metrics obtained underscore the robust predictive capabilities of SVM in foreseeing the onset of heart disease, Parkinson's, and diabetes. These results affirm the viability of machine learning algorithms in comprehensively addressing the complexities of diverse health conditions. Beyond numerical assessments, our research champions the translational application of machine learning in healthcare. The envisioned interface serves as a bridge between algorithmic predictions and real-world decision-making processes for healthcare practitioners.

IX. REFERENCES

[1] Shikha Dhyani,Adesh Kumar,Sushabhan Choudury .Analysis of ECG-based arrhythmia detection system using machine learning. MethodsX.2023 Apr 20:10:102195.

[2] Srikar Sistla. Predicting Diabetes using SVM Implemented by Machine Learning Title of the third paper. Retrieval Number: 100.1/ijsce.B35570512222 DOI: 10.35940/ijsce.B3557.0512222 Journal Website: www.ijsce.org

[3] Jing Zhang Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson’s disease. NPJ Parkinson's Dis. 2022; 8: 13. Published online 2022 Jan 21. doi: 10.1038/s41531-021-00266-8 [4] Hongfeng Wang, Hai Zhu, and Ding Accurate classification of lung nodules on CT images using the TransUnet Front Public Health. 2022; 10: 1060798. Published online 2022 Dec5. doi: 10.3389/fpubh.2022.1060798

[5] Godse, Rudra A., Gunjal, Smita S., Jagtap Karan A .,Mahamuni ,Neha S., &Wankhade, Prof. Suchita. (2019). Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively. International Journal of Advance Research in Computer and Communication Engineering, 8(12), 50-52

[6] Florian Mittag,Finja Büchel, Mohamad Saad,et al. Hum Mutat. 2012 Dec; 33(12): 1708–1718. Published online 2012 Aug 3. doi: 10.1002/humu.22161

[7] Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433438.

[8] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.

[9] Rajendra Acharya U, Fujita H, Oh SL, et al. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inf Sci (Ny). 2017;415-416:190-198

[10] Paniagua JA, Molina-Antonio JD, Lopez-Martinez F, et al. Heart disease prediction using random forests. J Med Syst. 2019;43(10):329.

[11] Poudel RP, Lamichhane S, Kumar A, et al. Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res. 2018;2018:1686023.

[12] Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform. 2014;83(8):596-604.

[13] Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface. 2012;9(65):2756-2764.

[14] Arora S, Aggarwal P, Sivaswamy J. Automated diagnosis of Parkinson's disease using ensemble machine learning. IEEE Trans Inf Technol Biomed. 2017;21(1):289-299.