A Hybrid Approach to Detect and Diagnosis of Disease Using Machine Learning Models.

Sabari V E
sabarivedhagiri100@gmail.com
Dept. of AIML,
AMC Engineering College
Bangalore, Karnataka

C. Subhashri
subhashrimc@gmail.com
Dept. of AIML,
AMC Engineering College
Bangalore, Karnataka

ABSTRACT

In recent times, predicting multiple diseases has become increasingly challenging in the medical field. Diseases such as heart disease, diabetes, and Parkinson's disease pose significant threats to global health. Some of these diseases progress rapidly, leading in creating symptoms like diziness, weakness and shivering. Our research highlights significant advances in the medical field by utilizing machine learning techniques to predict multiple diseases through a web application on patient test datasets. This application can predict diseases.

The research employs in algorithms with Logistic regression to determine whether one is affected by a disease, enhancing the accuracy and efficiency of prediction. This study predicts an individual's health status by identifying unknown values from the provided input datasets. The application reliably predicts these un- known values from user-submitted data to determine the presence of a disease. One of the primary features of this application is its ability to identify multiple diseases simultaneously within a web app, allowing for the input of extensive data for disease prediction.

Stream lit Cloud and the Stream lit library improve usability, enabling individuals to easily predict the risk of various diseases. The web app creates an accessible interface for users by offering selection options for different diseases. By selecting a specific disease within the web app, users can predict the disease based on the provided data. This study demonstrates that multiple diseases can be identified quickly, and the process is simplified through a user- friendly web app interface. High accuracy rates, ranging from 70% to 85%, can be achieved using these machine learning techniques.

KEYWORDS

ML (Machine Learning), ICMR (Indian Council for Medical Research), Logistic Regression, SVM (Support Vector Machine)

INTRODUCTION

Chronic diseases such as Diabetes, Heart disease, and Parkinson's disease have

presented substantial difficulties to global health systems in recent years, resulting in increased morbidity and death. These illnesses have a significant influence not only on individuals' quality of life but also on healthcare infrastructure and their resources.

Early detection and accurate prediction of these disorders creates an successful disease management and intervention techniques, which can greatly improve patient health, while lowering healthcare expenditures. With the introduction of improved computing tools, machine learning has emerged as a strong tool in medicine, presenting efficient answers for disease prediction and diagnosis.

Machine learning algorithms can evaluate massive volumes of healthcare data, revealing intricate patterns and relationships that might otherwise go undetected. By using these skills, machine learning can provide predictive insights that allow for early diagnosis and individualized treatment strategies. This work investigates the use of machine learning approaches to forecast the start and progression of diabetes, heart disease, and Parkinson's disease. We have explored numerous machine learning models, and evaluated their performance, by determining the most effective routes to accurate prediction. Our study makes use of a large dataset that includes ICMR (Indian Council Of Medical Research), clinical, demographic, and lifestyle information, ensuring a thorough and holistic analysis of the factors that contribute to these disorders. Through this effort, this research demonstrate machine learning's potential to revolutionize the landscape of chronic disease management, emphasizing its role in facilitating early intervention and improving patient care.

By improving machine learning's predictive capabilities, expecting to contribute to continuing efforts to reduce the worldwide burden of chronic diseases and improve the overall well-being of those affected.

Contemporary to the medical landscape to the early detection and accurate diagnosis creates an effective treatment and management. With advancing technique of ML create an exponential growth on the health care with an significant opportunity on enhancing in predicting accuracy and efficiency.

Diabetes is characterized by certain aspects like giving number of pregnancies, glucose level, BP, skin thickness, insulin level, BMI, age which captures the progression of disease. The ML model integrates multiple factor of data and predicts the likelihood of diabetes with an greater precision rate.

Heart Disease creating an mortality world wide the millions of individuals affected, with an early detection would prevent by not letting one to lead into severe complications which further improves the patients outcome, its characterized by certain aspects such as resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic and heart rate.

The cardiovascular challenges in certain aspects like accuracy and timely diagnosis, which aims in identifying patterns relative to heart disease, thereby leads up in facilitating prompt intervention.

Parkinson’s Disease an neurogenerative disorder which affects movement. The early detection would subtle the symptoms of the neurological conditions, using ML technique and algorithms it enhances in early detection. These are characterized by MDVP, SHIMMER, HNR.

In this successful implementation it holds an potential preventive health care by

offering an non invasive, cost effective, which makes an augments decisionmaking by empowering patients by fostering proactive management. Through this research we have aimed in evolving an AI driven health care solution paving an sophisticated accessible diagnostic technology.

LITREATURE REVIEW

Examining and diagnosing of disease using machine learning, with the goal of improving healthcare outcomes and reduction in costs through early prediction and prevention. This review covers diabetes, heart disease and Parkinson's disease providing insights from both operational and technical perspectives.

Diabetes:

Diabetes is an metabolic disease characterized by elevated blood glucose levels, which can cause severe damage to the heart, blood vessels, eyes, nerves, and kidneys leading in providing an early result at beginning stage by using hybrid approch technique.

Patients with low health latency rate had higher retention for diabetes. Participants in randomized group who engage in significant exercise are made an informed food choices, able to better control their diabetes.

In diabetes Support Vecotr Machine (SVM) has provided an highest accuracy rate and measuring of about 77% accuracy rate.

Heart Disease:

The heart is a crucial organ that pumps blood to all other organs in the human body. This study aims to predict heart disease with high accuracy, determining whether a person is at risk. Data mining technique in machine learning plays a vital role in context to multiple disease prediction.

Chronic diseases using certain Machine Learning techniques and algorithms such as Naive Bayes an supervised algorithm used for classifying, Logistic Regression which classifies binary and predicts the probablity of its outcome, observation or the event and Support Vector Machine (SVM) algorithm its associated in analyzing data for classification, and Artificial Neural Network (ANN) which helps in connecting up nodes of the hidden layer. These
algorithms provided an better performance rate and accuracy in prediction.

By using ANN for heart disease diagnosis yielded an accuracy of 81% with 20 neurons in the hidden layer, indicating significant potential of heart disease predictor

Parkinson's Disease:

In the disorder of parkinson’s which affects the central nervous system and movements. It results from nerve cell damage in the brain, leading to low dopamine levels, causing uncontrolled balance, walking difficulties, shaking hands, and memory loss, which gives an analysis about the disease using the predictor.

By using SVM for parkinson yielded an accuracy rate of 87%, indicating significant potential ability with respect to the disease predictor

TEHNICAL REQUIRMENTS

1. Data Collection and Preprocessing:

- Data Sources: Gathered data from reliable sources such as ICMR, public health databases, hospitals, or clinical studies.
- Data Cleaning: Handling missing values, outliers, and data inconsistencies.
- Feature Engineering: Selected and created relevant features for the prediction model.

2. Machine Learning Models:

- Model Selection: Chooses an appropriate models based on the type of diseases and the nature of the data. Common models includes Logistic Regression, Decision Trees, Random Forests, SVM (Support Vector Machines), and Neural Networks.
- Model Training: Trained the models using a portion of the dataset and validated their performance on a separate validation dataset.
- Model Evaluation: Using certain metrics such as accuracy, precision, recall, F1 score, and ROC-AUC to evaluate model performance.

3. Implementation:

- Programming Languages: Used languages like Python for implementing machine learning algorithms.
- Libraries and Frameworks: Utilised libraries such as scikit-learn, TensorFlow, Keras, or PyTorch for building and training models.

4. Deployment:

- Environment Setup: Built up an suitable environment for deploying our model, which could be cloud-based (AWS, Google Cloud, Azure) or on-premise.
- API Development: Developed an APIs to allow easy access to the prediction model for applications or for the end-users.

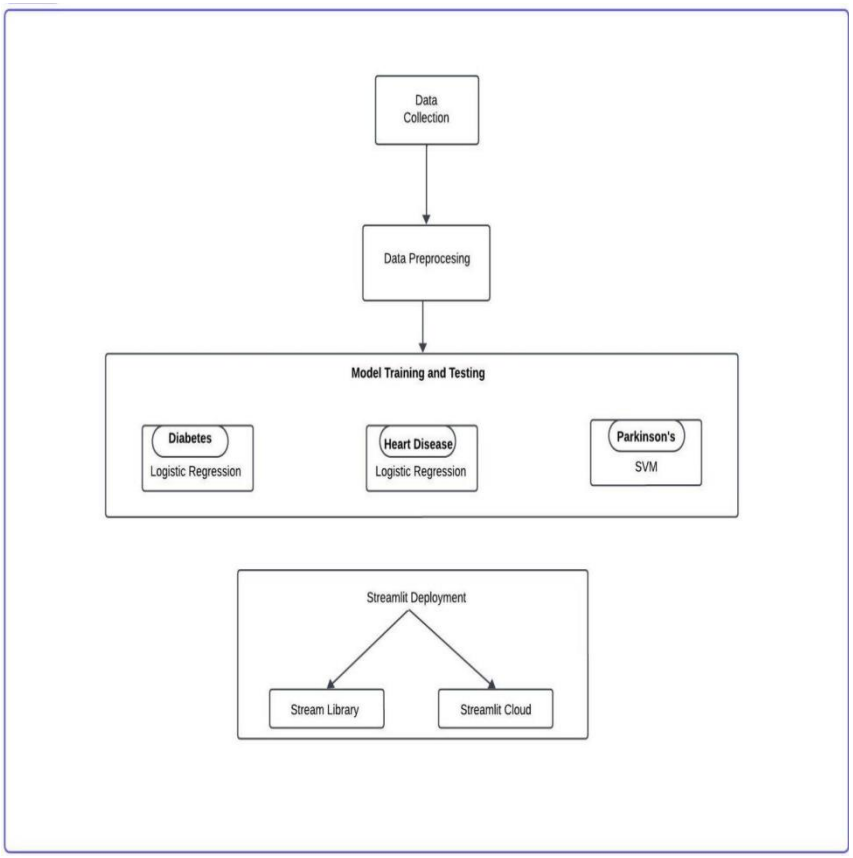
5. Documentation:

- Writing Clear Descriptions: This have documented each step of our research and methodology clearly.
- Code Repositories: This have maintained an repository for our code, ensuring it is well-documented and reproducible.

6. Ethical Considerations:

- Data Privacy: Ensures that the patient data is anonymized and complies with data protection regulations like HIPAA or GDPR.
- Bias and Fairness: It has address certain potential biases in our data and model to ensure fair and unbiased predictions.

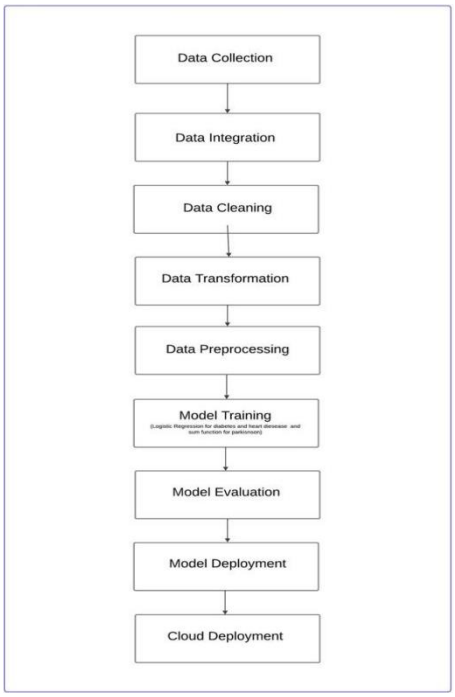
ARCHITECTURE



The data is collected from ICMR or from public health data base which are in form of demographic, medical history and lab results. The data integration merges up the data from various sources and creates an unified data by using certain tools, and the data cleaning process handles up the missing value and remove the duplicate corrected inconsistencies by using python and certain libraries, further the data is transformed by numerical standard range and converts the categorical data into an numerical format and creates up an feature existing an improved performance model.

Data preprocessing splits up the data as training set of about 80% and 20% as testing and validates the set of the training data, it balances up the data by using certain technique and feature is selected using certain technique like recursive feature. The model is trained and tested with diabetes, where the

model is trained with logistic regression algorithm by using Scikit learn in python by processing the trained model using the datasets by using validation set the hyper parameter



optimizes using certain techniques like random search, the model is evaluated with accuracy, precision, recall and F1-score, AUC-ROC.

Heart disease model is trained with logistic regression and uses certain library as of diabetes as Scikit-learn in python and hyperparameter gets optimized accordingly.

Parkinson's disease uses sum function approach like algorithms such as decision tree, SVM (Support Vector Machine) by using library like Scikit-learn and pickle which would further train the datasets

Its deployed with certain frameworks for creating interactive web application, the process builds and its developed by user friendly interface using stream lit with the prediction model and its hosted by applications on cloud platform. Stream lit is integrated with trained model into stream lit application, and the features are uploaded with patient data for prediction and displays results and interpretability insights, it visualizes data and model performance metrics.

Its deployed in cloud environment by setting up an virtual machine with an server-less environment further its installed with certain libraries and dependencies, it implements continuous integration and continuous deployment pipeline for automatic updates and maintenance. And the scalability ensures that it handles multiple users of large datasets.

The model is trained and tested with diabetes, where the model is trained with

logistic regression algorithm by using Scikit learn in python by processing the trained model using the datasets by using validation set the hyper parameter optimizes using certain techniques like random search, the model is evaluated with accuracy, precision, recall and F1-score, AUC-ROC.

Heart disease model is trained with logistic regression and uses certain library as of diabetes as Scikit-learn in python and hyperparameter gets optimized accordingly.

Parkinson's disease uses sum function approach like algorithms such as decision tree, SVM (Support Vector Machine) by using library like Scikit-learn and pickle which would further train the datasets

Its deployed with certain frameworks for creating interactive web application, the process builds and its developed by user friendly interface using stream lit with the prediction model and its hosted by applications on cloud platform. Stream lit is integrated with trained model into stream lit application, and the features are uploaded with patient data for prediction and displays results and interpretability insights, it visualizes data and model performance metrics.

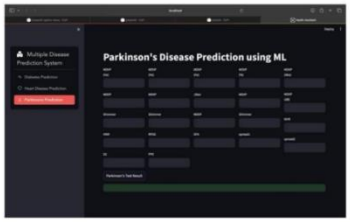
Its deployed in cloud environment by setting up an virtual machine with an serverless environment further its installed with certain libraries and dependencies, it implements continuous integration and continuous deployment pipeline for automatic updates and maintenance. And the scalability ensures that it handles multiple users of large datasets.

RESULTS:

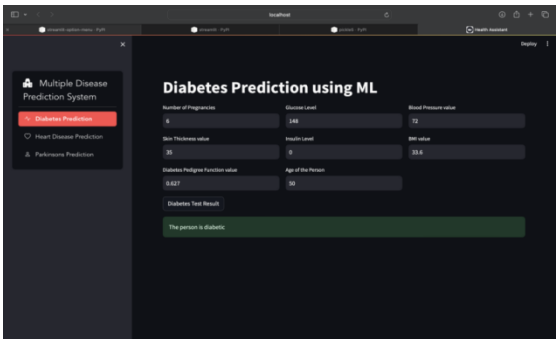
The structure approch in ensuring research process with clear findings, making it to undestand easier and replicate.

Here below is the accuracy rates of the disease using diffrent algorithms:

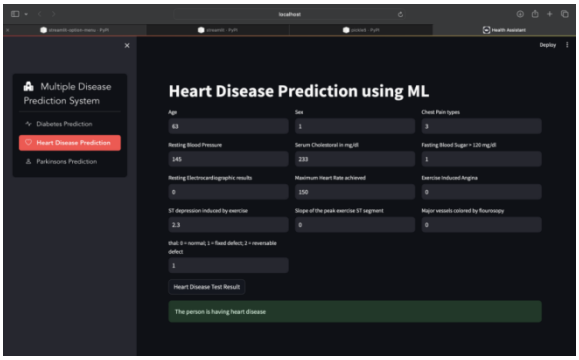
Disease	Logistic Regression	SVM	KNN	Accuracy
Diabetes	77.60	75.64	75.52	77.60
Heart Disease	82.50	78.87	83.84	83.84
Parkinson's	86.42	84.63	87.17	87.17



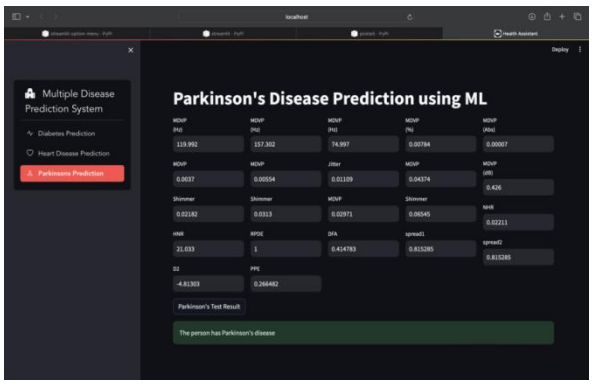
1. Diabetes



2.Heart Disease



3.Parkinson



CONCLUSION

Diabetes is a sizable persistent circumstance that impacts many people international, often by way of disrupting the frame's capacity to modify blood sugar ranges. With out right control, it could result in critical headaches like cardiovascular sickness, nerve harm, and kidney problems. No matter

development in treatments and technology, diabetes remains a tremendous fit-ness situation. Encouraging healthy existence, making sure early prognosis, and advancing medical remedies are critical in handling the disease. Persisted research and public education are crucial to improving the lives of these withdiabetes and lowering its worldwide impact.

Heart disease is a essential worldwide fitness difficulty, accounting for a substantial variety of deaths every year. This huge category consists of conditions like coronary artery ailment, coronary heart failure, and arrhythmias, often influenced by means of way of life factors, genetics, and different fitness situations. Notwithstanding advancements in medical generation and treatment, heart disorder remains a major project. Efforts to promote coronary heart-wholesome behaviour, enhance remedy alternatives, and lift consciousness are essential. Ongoing research and public fitness tasks play a important position in preventing heart ailment and improving the fine of life for the ones affected.

Parkinson's disease, in this research demonstrates an potential ML model, particularly SVM (Support Vector Machine) in predicting the disease with high accuracy rate. By its significant feature and symptoms the model promises for early diagnosis and an improved management, it should focus towards the future on broader validation by enhancing model interpretability and by integrating the tool for clinical and widespread practices. In this treatment it create an widespread quality of life with patient. This research bought us an imperative and support towards innovation in treatment.

REFERENCES

1. Arumugam, K., Naved, M., Shinde, P.P., Leiva-Chauca, O., Huaman-Osorio, A. and Gonzales-Yanac, T., 2023. Multiple disease prediction usingMachine learning algorithms. *Materials Today: Proceedings*, *80*, pp.3682-3685.

2. Harimoorthy, K. and Thangavelu, M., 2021. Retracted article: Multi-diseaseprediction model using improved SVM-radial bias technique in healthcaremonitoring system. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), pp.3715-3723.

3. Xie, S., Yu, Z. and Lv, Z., 2021. Multi-disease prediction based on deeplearning: a survey. *Computer Modeling in Engineering & Sciences*, *128*(2),pp.489-522.

4. Men, L., Ilk, N., Tang, X. and Liu, Y., 2021. Multi-disease prediction usingLSTM recurrent neural networks. *Expert Systems with Applications*, *177*,p.114905.

5. Yang Q, Khoury MJ, Botto L, Friedman JM, Flanders WD. Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. The American Journal of Human Genetics. 2003 Mar 1;72(3):636-49.

6. Pham, T., Tao, X., Zhang, J., Yong, J., Li, Y. and Xie, H., 2022. Graph-based multi-label disease prediction model learning from medical data anddomain knowledge. *Knowledge-based systems*, *235*, p.107662.

Kunjir A, Sawant H, Shaikh NF. Data mining and visualization for predictionof multiple diseases in healthcare. In2017 International Conference on BigDataAnalytics and Computational Intelligence (ICBDAC) 2017 Mar 23 (pp.329-334). IEEE.

7. Dubey, A.K., 2021. Optimized hybrid learning for multi disease predictionenabled by lion with butterfly optimization algorithm. *Sādhanā*, *46*(2), p.63.

8. Nori, N., Kashima, H., Yamashita, K., Ikai, H. and Imanaka, Y., 2015,August. Simultaneous modeling of multiple diseases for mortality prediction inacute hospital care. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 855-864).

9. Yuan X, Chen J, Zhang K, Wu Y, Yang T. A stable AI-based binary andmultiple class heart disease prediction model for IoMT. IEEE Transactions onIndustrial Informatics. 2021 Aug 18;18(3):2032-40.

10. Weng CH, Huang TC, Han RP. Disease prediction with different types ofneural network classifiers. Telematics and Informatics. 2016 May 1;33(2):277-92.

11. Singh SM, Hanchate DB. Improving disease prediction by machinelearning. Int. J. Res. Eng. Technol. 2018 Jun;5:1542-8.

12. Jain D, Singh V. Feature selection and classification systems for chronicdisease prediction: A review. Egyptian Informatics Journal. 2018 Nov1;19(3):179-89.

13. Deepthi Y, Kalyan KP, Vyas M, Radhika K, Babu DK, Krishna Rao NV.Disease prediction based on symptoms using machine learning. InEnergySystems, Drives and Automations: Proceedings of ESDA 2019 2020 Sep 1(pp. 561-569). Singapore: Springer Singapore.

14. Wang X, Wang F, Hu J. A multi-task learning framework for joint diseaserisk prediction and comorbidity discovery. In2014 22nd International Conference on Pattern Recognition 2014 Aug 24 (pp. 220-225). IEEE.

15. Epstein FH. Multiple risk factors and the prediction of coronary heartdisease. Bulletin of the New York Academy of Medicine. 1968 Aug;44(8):916.

16. Wang L, Wong L, Li Z, Huang Y, Su X, Zhao B, You Z. A machine learningframework based on multi-source feature fusion for circRNA-disease association prediction. Briefings in Bioinformatics. 2022 Sep;23(5):bbac388.

17.