RESEARCH ARTICLE OPEN ACCESS

Enhanced Healthcare Risk Assessment: Multi-Disease Prediction Using State-of-the-Art Machine Learning Algorithms

Mrs. Vijay Sree Dhulipalla[1],Mr. Kantheti Raju Mithra[2],Mr. John Saida Mohammad[3], Mr. Ramisetty P V S R Sathwik[4], Mr. Kantheti Pedda ManiKanta[5], Mr. Katta Jaya Prakash[6]

[1]Assistant Professor,Department of Artificial Intelligence, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES, and Vinjanampadu, Guntur, A.P, India
Email: vijayasreevignesh@gmail.com

[2]STUDENT, Department of Artificial Intelligence, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES, and Vinjanampadu, Guntur, A.P, India
Email: rajumithra2002@gmail.com

[3]Assistant Professor, Department of Artificial Intelligence, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES, and Vinjanampadu, Guntur, A.P, India
Email: md.saida@gmail.com

[4]STUDENT, Department of Artificial Intelligence, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES, and Vinjanampadu, Guntur, A.P, India
Email: sathwik3721@gmail.com

[5]STUDENT, Department of Artificial Intelligence, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES, and Vinjanampadu, Guntur, A.P, India
Email:manikantheti96@gmail.com

[6]STUDENT, Department of Artificial Intelligence, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES, and Vinjanampadu, Guntur, A.P, India
Email: jayaprakash545@gmail.com

--************************----------------------------------

Abstract:

The increasing prevalence of electronic health data has prompted a shift towards supervised machine learning (ML) algorithms for enhanced disease detection in healthcare. This study investigates the performance trends of these algorithms, highlighting the proficiency of Support Vector Machine (SVM) in detecting kidney diseases and Parkinson's disease. Logistic Regression (LR) excels in predicting heart diseases, while Random Forest (RF) and Convolutional Neural Networks (CNN) show promise in forecasting breast diseases and common ailments, respectively. This research contributes valuable insights for leveraging ML models in disease diagnosis, signifying a potential paradigm shift in healthcare methodologies.

Keywords **—Health care, Supervised Machine Learning, Disease Prediction, Artificial Intelligence.**

--************************----------------------------------

I. INTRODUCTION

A. Motivation

The advent of Artificial Intelligence (AI) has ushered in a new era where computerized systems can emulate human-like perception, cognition, and operation [1]. AI, a multidisciplinary field encompassing Machine Learning (ML), Computer Vision, Deep Learning, and Natural Language Processing [2], utilizes ML algorithms that leverage optimization, statistical, and probabilistic techniques to glean insights from data. This acquired knowledge is then applied to decision-making processes [3]. The extensive applications of ML algorithms span various domains such as network intrusion recognition, customer purchase behavior detection, process manufacturing optimization, credit card fraud detection, and disease modulation. The supervised learning approach, involving the training of prediction models on datasets with known labels to infer unlabeled examples, holds promise for more efficient disease diagnosis in the medical field [4].

Reports from Medicaid services and centers for Medicare reveal that 50% of Americans grapple with multiple chronic diseases, resulting in a staggering $3.3 trillion expenditure on U.S. healthcare in 2016, equating to $10,348 per person [5]. Additionally, the World Health Organization and the World Economic Forum highlight a significant economic loss of $236.6 billion in India by 2015 due to fatal diseases stemming from malnutrition and unhealthy lifestyles [6]. These statistics underscore the vulnerability of individuals to a spectrum of diseases, emphasizing the critical need for early disease detection to reduce fatality rates. Early prediction not only alleviates the financial burden on the economy but also ensures better overall community well-being [5], [6].

Yuan identifies two critical factors contributing to errors in ML algorithms. Firstly, the accuracy and impartiality of decisions hinge on the quality and selection of datasets. Secondly, the success of ML algorithms heavily relies on the proper extraction of features from datasets, a process known to be challenging, time-consuming, and computationally demanding. These challenges can compromise the performance of learning models, potentially leading to fatal errors that pose risks to patients' lives. In contrast, Ismaeel argues that conventional statistical techniques, coupled with the experience and intuition of medical professionals, introduce biases and errors in risk detection associated with diseases. The surge in electronic health data has intensified the challenges faced by medical practitioners in accurately identifying diseases early. To address these challenges, advanced computational methodologies, such as ML algorithms, have been introduced to uncover meaningful patterns and hidden information from data, thereby aiding critical decision-making. Consequently, this has reduced the burden on medical staff and improved patient survival rates [3], [8].

B. Aim

The primary objective of this study is to empirically test the hypothesis that supervised ML algorithms can enhance healthcare by enabling accurate and early disease detection. The investigation focuses on studies employing multiple supervised ML models for each disease recognition problem. This comprehensive approach ensures precision by evaluating the performance of various algorithms in diverse study settings, mitigating potential biases associated with relying on a single algorithm. The analysis encompasses diseases related to the heart, kidneys, breasts, and brain. Multiple methodologies, including KNN, NB, DT, CNN, SVM, and LR, will be evaluated for disease detection. The literature concludes by identifying the best-performing ML models for each specific disease, contributing valuable insights to the field of healthcare.

II. LITERATURE REVIEW

A. Common Diseases

Dahiwade et al. [9] introduced an ML-based system for predicting common diseases by leveraging a dataset from the UCI ML repository. The study utilized CNN and KNN, showcasing CNN's superiority in both accuracy and processing time. The emphasis was on CNN's ability to discern intricate relationships within feature spaces. However, a notable drawback was the lack of detailed parameters for neural networks, and the oversight of considering bias in the analysis.

B. Kidney Diseases

In the realm of Chronic Kidney Disease (CKD) detection, Serek et al. [12] conducted a comparative study on classifier performance. Their findings favored RF, demonstrating excellence in F-measure and accuracy, while NB excelled in precision. SVM emerged as a preferred choice in various studies on kidney diseases due to its adaptability in handling intricate datasets. However, a critical point of weakness surfaced concerning the insufficient exploration of hyper-parameters.

C. Heart Diseases

The prediction of heart diseases was a focus for Marimuthu et al. [16], with LR standing out for its remarkable accuracy. Dwivedi [17] elevated precision by incorporating additional parameters, and studies by Polaraju [18] and Vahid et al. [19] consistently favored Logistic Regression over alternative techniques. Despite the thorough analyses conducted in these studies, the constraints imposed by small dataset sizes hindered the ability to target diseases with higher accuracy and precision.

D. Breast Disease

Shubair [20] delved into breast cancer detection using ML algorithms, favoring SVM due to its superior recall, accuracy, and precision. Yao [21] echoed support for Random Forest, emphasizing its scalability for large datasets. While these studies presented multiple metrics, the omission of data preprocessing details introduced a potential disadvantage.

E. Parkinson's Disease

In Parkinson's disease (PD) diagnosis, Chen et al. [22] proposed an effective system using FKNN, surpassing SVM in sensitivity, accuracy, and specificity. Behroozi [23] introduced a novel classification framework for PD, enhancing accuracy through a filter-based feature selection algorithm. Eskidere [24] tracked PD progression, highlighting LS-SVM as the highest-performing model. The suggestion of calibrating ML models before evaluation emerged as a key consideration for further improvement.

III. EXISTING SOLUTION

One existing solution for multiple disease prediction using machine learning is the use of ensemble learning techniques. Ensemble learning combines multiple machine learning models to improve predictive performance. In the context of disease prediction, an ensemble approach could involve combining the predictions of multiple individual models, such as decision trees, support vector machines (SVM), logistic regression, or neural networks.

For example, researchers might develop an ensemble model for disease prediction by training multiple individual classifiers on a dataset containing features related to various diseases. Each classifier could specialize in predicting a specific disease or group of diseases based on different sets of input features. Then, the predictions from each individual classifier are combined using techniques such as averaging, voting, or stacking to produce a final prediction for each patient

IV. PROPOSED SYSTEM

The proposed system aims to enhance the accuracy and efficiency of disease diagnosis through the integration of ensemble learning techniques. The focus is on predicting diseases related to the heart, kidney, breast, and brain using a combination of machine learning algorithms. The system will utilize a diverse set of algorithms to address the complexity and variability of different diseases. These algorithms will be implemented to collectively predict diseases across different organs. The system will utilize the strengths of each algorithm to provide comprehensive and accurate predictions for early disease detection. Model evaluation will involve assessing performance metrics such as accuracy, sensitivity, precision, and F-measure to ensure the effectiveness of the proposed system. Additionally, the ensemble approach helps mitigate the limitations of individual algorithms and enhances the overall robustness of disease prediction.

V. MODEL

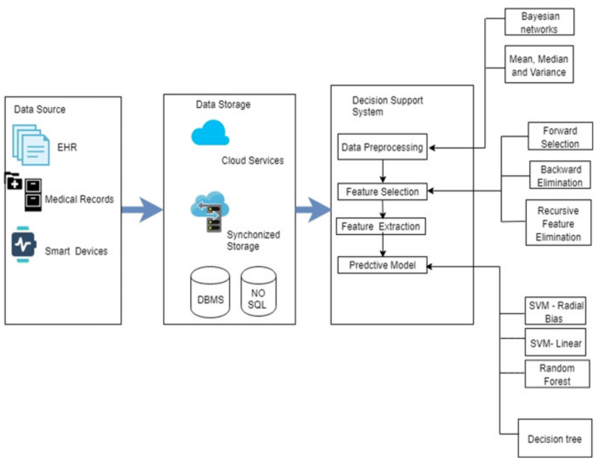


Fig a. model

VI. IMPLEMENTATION AND RESULT

Raw data from different healthcare datasets:

Healthcare datasets contain a wide range of information, including patient demographics, medical history, laboratory test results, imaging data, and more. These datasets are collected from various sources such as hospitals, clinics, research institutions, and wearable devices.

Data Gathering:

Data gathering involves collecting healthcare data from diverse sources, ensuring it is representative and comprehensive. This process may include obtaining permissions, accessing databases, extracting relevant information, and consolidating data from multiple sources into a unified dataset.

Data Preprocessing:

Data preprocessing is a crucial step to ensure the quality and usability of the collected healthcare data. It involves tasks such as handling missing values, dealing with outliers, standardizing data formats, and converting categorical variables into numerical representations. Additionally, preprocessing may include data normalization to ensure consistency and improve model performance.

Data Cleaning:

Data cleaning aims to identify and correct errors, inconsistencies, and inaccuracies in the healthcare data. This process involves tasks such as removing duplicates, correcting typos, resolving discrepancies, and validating data against predefined criteria or rules. Data cleaning is essential to maintain data integrity and reliability for downstream analysis.

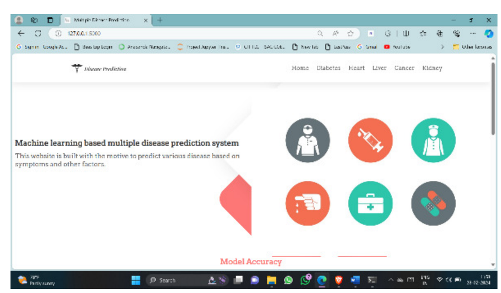
Feature Scaling:

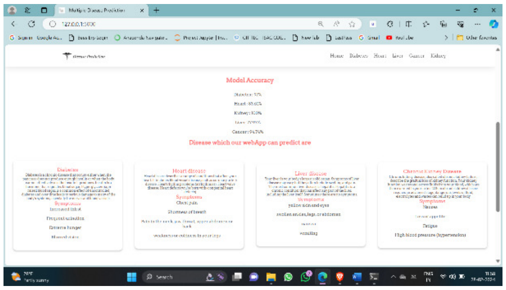
Feature scaling is a preprocessing technique used to standardize the range of features in the dataset. In healthcare data, features may have different scales and units of measurement, which can affect the performance of machine learning models. Feature scaling methods such as normalization or standardization are applied to ensure that all features contribute equally to model training and improve convergence speed.

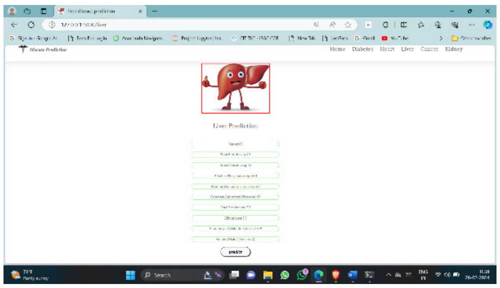
Modeling:

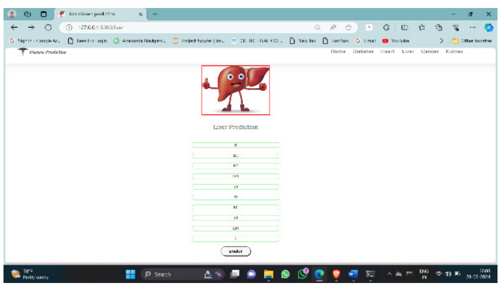
Modeling involves building machine learning models using the preprocessed healthcare data to predict outcomes, classify patients into different groups, or derive insights from the data. Various machine learning algorithms such as decision trees, support vector machines, neural networks, and ensemble methods can be applied depending on the specific healthcare task and the nature of the data. Model selection, evaluation, and optimization are iterative processes aimed at improving model performance and generalizability.

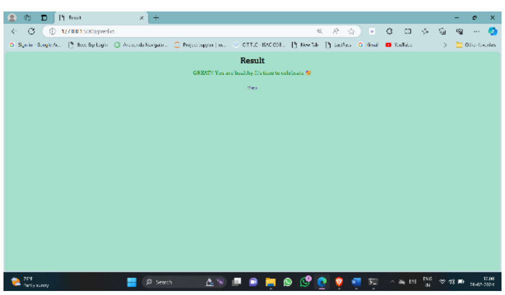
fig b. Home Page of Multiple Disease Prediction System











ACKNOWLEDGMENT

We extend our gratitude to healthcare institutions for providing diverse datasets. Special thanks to contributors involved in data gathering, preprocessing, and cleaning, crucial for robust analysis. Recognition is given to efforts in feature scaling, enhancing model accuracy. Lastly, we appreciate the broader data science community for advancing methodologies that underpin this research.

CONCLUSION AND FUTURE SCOPE

Future efforts will focus on expanding the API to include more diseases, improving prediction accuracy, exploring advanced machine learning models, integrating multi-omics data, embracing personalized medicine, implementing real-time monitoring through wearables, and enhancing model interpretability with Explainable AI methods.

REFERENCES

[1] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, “Comparingdifferent supervised machine learning algorithms for disease prediction,”BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1–16, 2019.

[2] R. Katarya and P. Srinivas, “Predicting heart disease at early stagesusing machine learning: A survey,” in 2020 International Conferenceon Electronics and Sustainable Communication Systems (ICESC), 2020,pp. 302–305.

[3] P. S. Kohli and S. Arora, “Application of machine learning in diseaseprediction,” in 2018 4th International Conference on Computing Com-munication and Automation (ICCCA), 2018, pp. 1–4. M. Patil, V. B. Lobo, P. Puranik, A. Pawaskar, A. Pai, and R. Mishra,“A proposed model for lifestyle disease prediction using support vectormachine,” in 2018 9th International Conference on Computing, Com-munication and Networking Technologies (ICCCNT), 2018, pp. 1–6.

[4] F. Q. Yuan, “Critical issues of applying machine learning to conditionmonitoring for failure diagnosis,” in 2016 IEEE International Confer-ence on Industrial Engineering and Engineering Management (IEEM),2016, pp. 1903–1907.

[5] S. Ismaeel, A. Miri, and D. Chourishi, "Using the extreme learning ma-chine (elm) technique for heart disease diagnosis," in 2015 IEEE CanadaInternational Humanitarian Technology Conference (IHTC2015), 2015,pp. 1–3.

[6] D. Dahiwade, G. Patle, and E. Meshram, "Designing disease predictionmodel using machine learning approach," Proceedings of the 3rd Inter-national Conference on Computing Methodologies and Communication,ICCMC 2019, no. Iccmc, pp. 1211–1215, 2019.

[6] S. Jadhav, R. Kasar, N. Lade, M. Patil, and S. Kolte, "Disease Predictionby Machine Learning from Healthcare Communities," InternationalJournal of Scientific Research in Science and Technology, pp. 29–35,2019.

[7] R. Saravanan and P. Sujatha, "A state of art techniques on machinelearning algorithms: A perspective of supervised learning approachesin data classification," in 2018 Second International Conference onIntelligent Computing and Control Systems (ICICCS), 2018, pp. 945–949.

[8] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronickidney disease dataset by applying machine learning methods," in 2018IEEE 12th International Conference on Application of Information andCommunication Technologies (AICT), 2018, pp. 1–4.

[9] V. S and D. S, "Data Mining Classification Algorithms for Kidney Dis-ease Prediction," International Journal on Cybernetics &Informatics,vol. 4, no. 4, pp. 13–25, 2015.

[10] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit,S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronickidney disease using machine learning techniques," 2016 Managementand Innovation Technology International Conference, MITiCON 2016,pp. MIT80–MIT83, 2017.

[11] P. Kotturu, V. V. Sasank, G. Supriya, C. S. Manoj, and M. V. Mahesh-warredy, "Prediction of chronic kidney disease using machine learningtechniques," International Journal of Advanced Science and Technology,vol. 28, no. 16, pp. 1436–1443, 2019.M. Marimuthu, M. Abinaya, K. S., K. Madhankumar, and V. Pavithra,"A Review on Heart Disease Prediction using Machine Learning andData Analytics Approach," International Journal of Computer Applica-tions, vol. 181, no. 18, pp. 20–25, 2018.
[12] A. K. Dwivedi, "Performance evaluation of different machine learningtechniques for prediction of heart disease," Neural Computing andApplications, vol. 29, no. 10, pp. 685–693, 2018.

[13]K. Polaraju, D. Durga Prasad, and M. Tech Scholar, "Prediction ofHeart Disease using Multiple Linear Regression Model," InternationalJournal of Engineering Development and Research, vol. 5, no. 4, pp.2321–9939, 2017. [Online]. Available: www.ijedr.org

