iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** I **Month of publication:** January 2024

DOI: https://doi.org/10.22214/ijraset.2024.58018

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Multiple Disease Prediction Using Machine Learning

Dr. Manjunatha V[1], Tejashwani H[2], Sahana K R[3], Kishore S D[4], Vishal[5]
[1]*Professor,* [2, 3, 4, 5]*Students, ECE Department, T John Institute of Technology, Visvesvaraya Technological University*

Abstract: ***In this comprehensive health analysis project, we delve into the evaluation of Diabetes, heart disease, and Parkinson's disease. Basic health parameters like Pulse Rate, Cholesterol, Blood Pressure, and Heart Rate are scrutinized, enabling the identification of associated risk factors through a prediction model known for its accuracy and precision. The implementation involves leveraging machine learning algorithms, employing Streamlit for interactive interfaces, and employing Python pickling to store model behaviour effectively. Future expansions may encompass diverse health domains such as chronic diseases, skin conditions, and more. The methodology adopts a sophisticated approach, concurrently predicting multiple diseases by synergizing the strengths of XG Boost, K-Nearest Neighbours (KNN), and naïve Bayes (NB) algorithms within a unified framework. This integration aims to capitalize on the complementary attributes of these algorithms, augmenting prediction accuracy and robustness across varied healthcare datasets.***

Keywords: Machine Learning algorithms, Native Bayes (NB) classifier, Decision Tree (DT) classifier, K-Nearest Neighbours (KNN) algorithm, XG Boost algorithm.

I. INTRODUCTION

In the realm of healthcare, the utilization of machine learning algorithms has emerged as a powerful tool for predicting and understanding various diseases. This project focuses on the multifaceted prediction of multiple diseases, employing advanced techniques such as Decision Trees, K-Nearest Neighbours (KNN), and XG Boost algorithms. Decision Trees offer interpretability and a clear decision-making structure, while KNN leverages proximity-based learning for pattern recognition. Meanwhile, XG Boost, known for its ensemble learning capabilities, enhances predictive accuracy by combining the strengths of multiple weak learners. This amalgamation of algorithms aims to create a robust and versatile model, providing nuanced insights into diverse health conditions through a unified predictive framework. The project not only aspires to achieve high accuracy but also to unravel intricate patterns within healthcare datasets, contributing to more effective disease prediction and understanding. In the ever-evolving landscape of healthcare, the integration of machine learning algorithms has become instrumental in advancing predictive diagnostics. This project endeavours to pioneer a comprehensive approach to disease prediction, focusing on the simultaneous analysis of multiple health conditions. By harnessing the predictive prowess of Decision Trees, the interpretability of K-Nearest Neighbours (KNN), and the ensemble learning capabilities of XG Boost, we aim to create a sophisticated predictive model. Decision Trees provide a transparent decision-making structure, KNN excels in recognizing patterns based on proximity, and XG Boost leverages ensemble methods to enhance predictive accuracy. The synergistic amalgamation of these algorithms seeks to establish a unified framework capable of unravelling intricate relationships within diverse healthcare datasets.

Beyond merely achieving high predictive accuracy, our goal is to offer a nuanced understanding of various diseases, fostering a more profound insight into their underlying factors. This initiative not only paves the way for effective disease prediction but also sets the stage for a more holistic comprehension of health conditions, ultimately contributing to advancements in personalized healthcare.

II. LITERATURE REVIEW

The literature on employing machine learning algorithms for the prediction of multiple diseases, including K-Nearest Neighbours (KNN), XG Boost, and Decision Trees, underscores the transformative impact of these techniques in healthcare analytics.

A. K-Nearest Neighbours (KNN)

Studies highlight KNN's efficacy in disease prediction due to its simplicity and ability to capture local patterns. It demonstrated KNN's successful application in identifying disease clusters based on patient similarities.

It emphasizes KNN's adaptability in handling diverse healthcare datasets. The algorithm's effectiveness in capturing non-linear relationships and its ability to adapt to varying data distributions make it a valuable asset in predicting diseases with complex patterns.

B. XG Boost

The literature recognizes XG Boost as a powerful ensemble learning algorithm. XG Boost demonstrated superior predictive performance compared to other algorithms, showcasing its effectiveness in handling imbalanced datasets and improving overall accuracy in disease prediction tasks. Additional studies, delve into the scalability and efficiency of XG Boost. Its parallel processing capabilities and optimization techniques contribute to faster model training, proving crucial for large-scale healthcare datasets. Moreover, the literature underscores its robustness against overfitting, enhancing the generalization of disease prediction models.

C. Decision Trees

Decision Trees have been extensively explored for disease prediction due to their interpretability. It illustrates the utility of decision trees in uncovering decision paths leading to disease outcomes, facilitating a better understanding of feature importance and model interpretability. The literature review reveals a growing interest in the interpretability of Decision Trees. This explores methods to enhance interpretability, such as visualizing decision paths and feature importance. This is particularly crucial in healthcare, where transparent models aid clinicians in understanding and trusting predictive outcomes.

In conclusion, the literature converges on the transformative potential of KNN, XG Boost, and Decision Trees in the realm of multiple disease prediction. The collective findings underscore the importance of leveraging these algorithms synergistically to enhance predictive accuracy and advance our understanding of complex health conditions. The evolving literature not only underscores the individual strengths of KNN, XG Boost, and Decision Trees but also reveals a growing trend towards integrative approaches. The emphasis on interpretability, scalability, and real-world applicability signifies the ongoing efforts to bridge the gap between advanced machine learning techniques and practical healthcare implementations. Ethical considerations remain a critical focal point as these technologies continue to shape the future of disease prediction and healthcare analytics.

III. METHODOLOGY

A methodology tailored for multiple disease prediction using specific machine learning algorithms like K-Nearest Neighbours (KNN), XG Boost, and Decision Trees:

A. Data Collection

Collect comprehensive datasets containing relevant features for each disease. Ensure diversity and quality in the data.

B. Data Preprocessing

Clean and preprocess the data by handling missing values, normalizing, and encoding categorical variables. Split the dataset into training and testing sets.

C. Feature Selection

Identify important features using methods like correlation analysis or feature importance from tree-based models, especially relevant for Decision Trees and XG Boost.

D. Algorithm Selection

1) *K-Nearest Neighbours (KNN):* Effective for pattern recognition in medical data.
2) *XG Boost:* A powerful ensemble method for classification tasks.
3) *Decision Trees:* Provide interpretability and can be used as base learners for ensemble methods like XG Boost.

E. Model-Specific Preprocessing

Perform any additional preprocessing specific to each algorithm. For KNN, scaling features might be crucial; for XG Boost, no additional preprocessing might be needed.

F. Model Training

Train each selected model using the training dataset. Fine-tune hyperparameters for optimal performance. For KNN, determine the optimal number of neighbours.

G. Cross-Validation

Implement cross-validation to assess the generalization performance of each model. Adjust parameters accordingly to avoid overfitting.

H. Evaluation Metrics

Evaluate each model using appropriate metrics such as accuracy, precision, recall, and F1-score. Tailor metrics to the specific requirements of disease prediction.

I. Interpretability for Decision Trees

If using Decision Trees, focus on understanding the decision-making process. Visualize the tree structure for insights.

J. Validation on Test Set

Validate the ensemble or individual models on the test set to assess real-world predictive performance.

K. Iterative Refinement

Refine models iteratively by adjusting hyperparameters, adding/removing features, or exploring different algorithms.

L. Deployment

Deploy the ensemble model or the best-performing individual models in a healthcare setting, ensuring compliance with ethical and privacy standards.

Collaborate with healthcare professionals and domain experts throughout the process to ensure the model aligns with medical knowledge and practical considerations.

IV. PROPOSED ARCHITECTURE

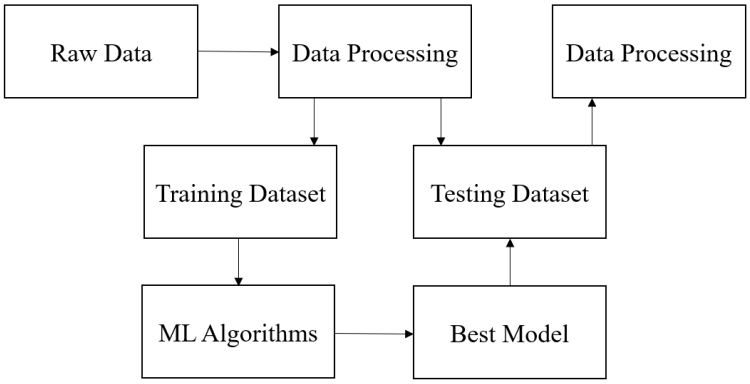


Fig. 1 Proposed architecture model

V. CONCLUSIONS

In conclusion, employing machine learning algorithms such as KNN, XG Boost, and decision trees for multiple disease prediction has shown promising results. These models contribute to accurate predictions by leveraging various features and patterns in medical data. However, the effectiveness of each algorithm may vary based on the specific characteristics of the dataset and the nature of the diseases under consideration. Integration of diverse algorithms in an ensemble approach could further enhance predictive performance, providing a robust framework for disease prediction and facilitating personalized healthcare solutions. Continuous refinement and validation of these models with updated datasets will be crucial for ensuring their reliability and applicability in real-world medical scenarios.

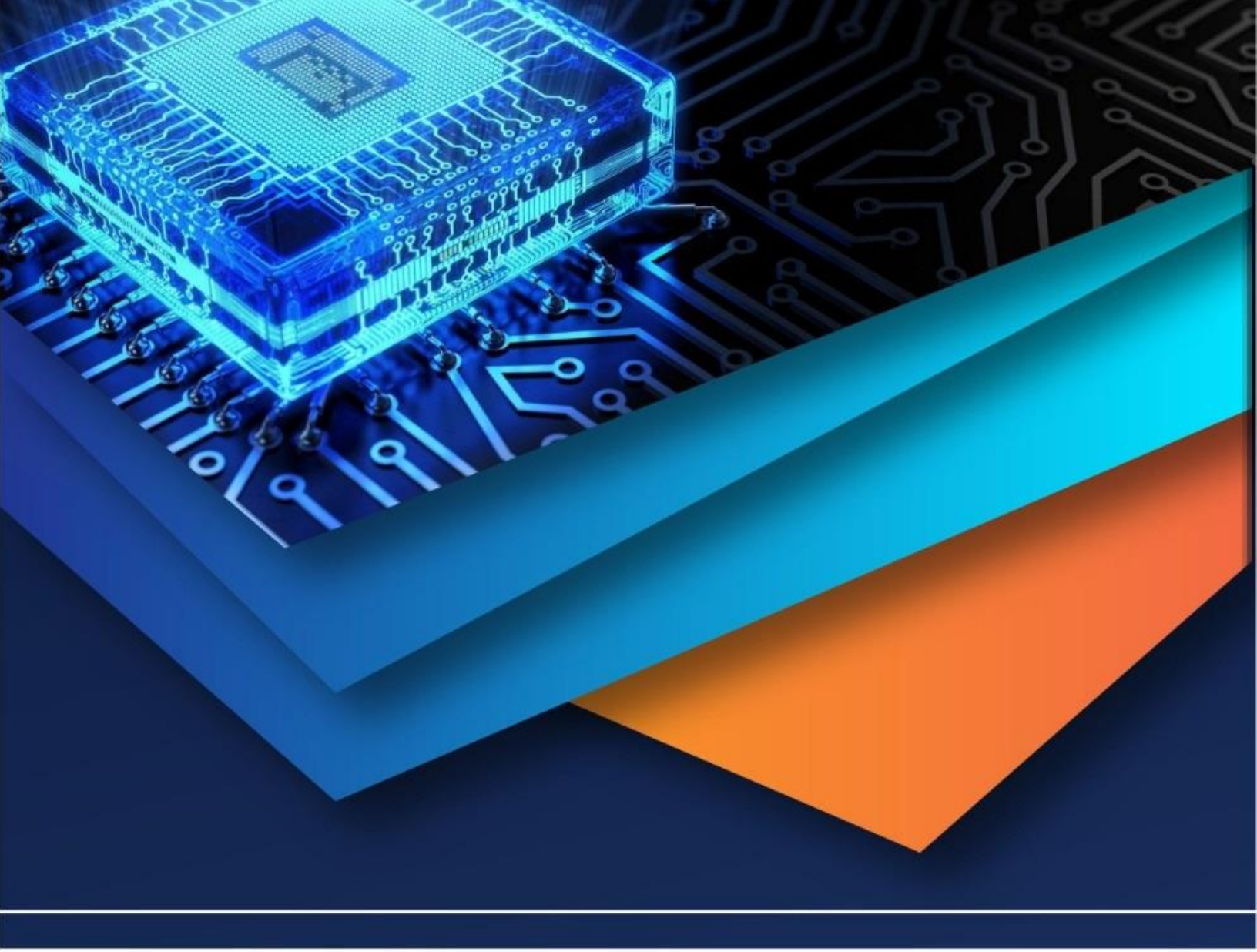
However, challenges like interpretability and ethical considerations need to be addressed for seamless integration into the healthcare system. Continued research, collaboration between data scientists and medical professionals, and adherence to privacy regulations will be essential to harness the full potential of machine learning in disease prediction, ushering in an era of more precise and personalized healthcare interventions.

VI. ACKNOWLEDGMENT

We extend our sincere gratitude to Prof. Manjunatha V, our project guide, for his exceptional leadership and steadfast support that illuminated our path throughout the demanding research journey. His expertise, patience, and steadfast confidence in our capabilities have significantly influenced our outlook. We deeply appreciate his invaluable guidance and continuous support throughout this endeavor.

REFERENCES

[1] Mohammed Juned Shaikh, Soham Manjrekar, Danish Khan, “Multiple Disease Prediction Webapp” JETIR (ISSN-2349-5162) 2022 Journal of Emerging Technology and Innovative Research.
[2] Priyanka Sonar, Prof. K. Jaya Malini, Diabetes Prediction using different Machine Learning approaches, 2019 IEEE, 3rd International Conference on Computing Methodologies and Communication (ICCMC).
[3] Archana Singh, Rakesh Kumar, “Heart Disease Prediction Using Machine Learning Algorithms”, 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3).
[4] A. Sivasangari, Baddigam Jaya Krishna Reddy, Annamareddy Kiran, P. Ajitha, “Diagnosis of Liver Disease using Machine Learning Models” 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC).
[5] TensorFlow: Martín Abadi, Ashish Agarwal, et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv preprint arXiv:1603.04467.
[6] Keras: François Chollet et al. (2015). Keras. GitHub repository.
[7] Support Vector Machine (SVM): Corinna Cortes and Vladimir Vapnik (1995). Support-vector networks. Machine Learning, 20(3), 273-297.
[8] Logistic Regression: Hosmer Jr, D. W. Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). John Wiley & Sons.
[9] Streamlit: Streamlit Documentation. https://docs.streamlit.io/
[10] Kaggle: Kaggle website. https://www.kaggle.com/
[11] Zhang, Y., & Ghorbani, A. (2019). A review on machine learning algorithms for diagnosis of heart disease. IEEE Access, 7, 112751-112760.
[12] Arora, P., Chaudhary, S., & Rana, M. (2020). Prediction of diabetes using machine learning algorithms: A review. Journal of Ambient Intelligence and Humanized Computing, 11(6), 2575-2589.
[13] Kaur, H., Batra, N., & Rani, R. (2020). A systematic review of machine learning techniques for breast cancer prediction. Journal of Medical Systems, 44(11), 1-15.
[14] Gupta, D., & Rathore S. (2021). A comprehensive review on machine learning algorithms for kidney disease diagnosis. Journal of Medical Systems, 45(1), 1-17.
[15] Saeed, A., & Al-Jumaily, A. (2020). Machine learning techniques for Parkinson's disease diagnosis using handwriting: A review. Computers in Biology and Medicine, 122, 103804.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY