International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Predictive Healthcare Analytics Through Machine Learning

Mrs. Shanthini[1], J. Ruby Jayasri[2]

[1]Assistant Professor, Department of Computer Applications, Jaya College of Arts and Science, Chennai
[2]PG Student, Department of Computer Applications, Jaya College of Arts and Science, Chennai

Abstract
With the introduction of electronic health records, laboratory tests, and wearable devices, healthcare has become a rich source of large volumes of patient data. However, this data is frequently underutilized in clinical decision-making. This research focuses on developing a machine learning model to support healthcare predictive analytics with the primary goal of predicting disease risk and offering diagnostic assistance at early stages. The project employs disease data preprocessing, feature extraction, and multiple machine learning algorithms such as Logistic Regression, Random Forest, and Support Vector Machines to determine the most accurate prediction model. The study confirms that artificial intelligence techniques can significantly improve diagnostic precision and early intervention in healthcare settings.

1. INTRODUCTION

Healthcare is undergoing a major digital transformation with the integration of Artificial Intelligence (AI) and data analytics. Predictive analytics relies on analyzing historical data and using statistical and machine learning models to forecast potential outcomes. In medicine, this involves identifying patients at high risk of developing certain diseases, allowing early intervention and reducing hospitalization rates. Machine Learning (ML) models can automatically extract hidden patterns from medical data without manual programming. This research aims to create an ML-based predictive model that estimates disease likelihood based on patient health attributes such as age, blood pressure, glucose, and cholesterol levels.

2. Objectives

1. To obtain and preprocess medical data for disease prediction.
2. To implement and compare machine learning algorithms for accuracy and reliability.
3. To identify major health indicators influencing disease risk.
4. To develop a functional prototype to assist healthcare professionals.
5. To demonstrate how predictive analytics can improve early disease detection and prevention.

3. Literature Review

Several studies have demonstrated the potential of machine learning in healthcare. Logistic Regression and Random Forest have been widely used for diabetes prediction, achieving accuracies above 80%. Neural networks have shown high efficiency in predicting heart diseases due to their ability to learn

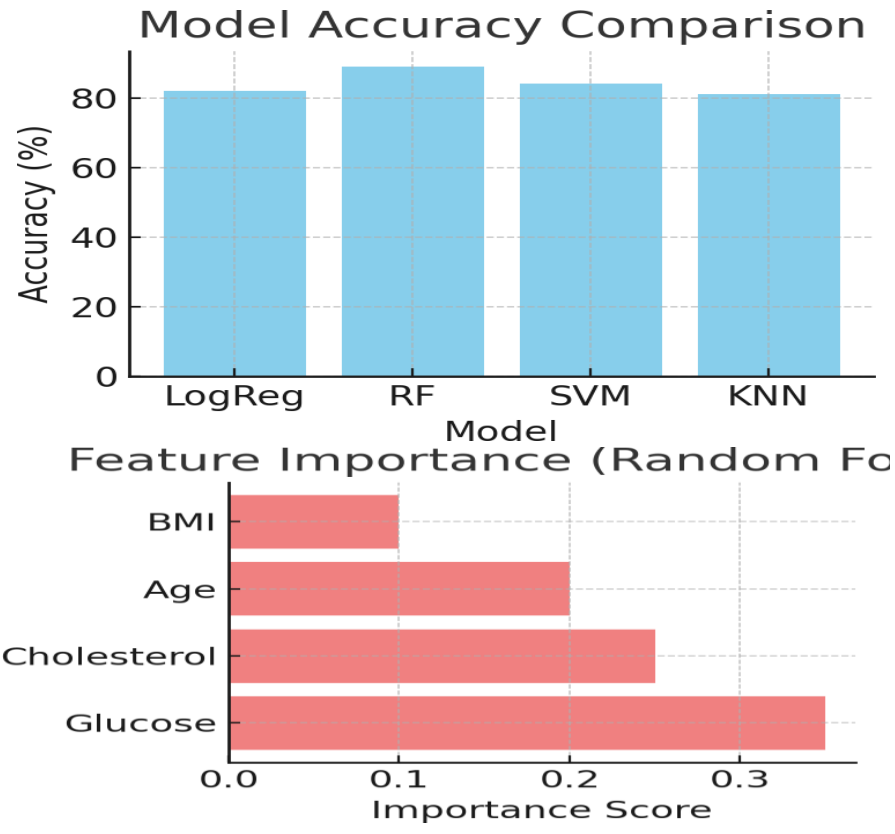
complex patterns. Recent work emphasizes Explainable AI (XAI) to help clinicians interpret predictions. This project builds upon existing research by comparing multiple algorithms on open datasets and highlighting factors that influence prediction accuracy.

4. Methodology

The dataset used in this research was sourced from public repositories such as the UCI Machine Learning Repository and Kaggle. Each record includes parameters such as Age, Gender, Blood Pressure, Glucose Level, Cholesterol, BMI, and Heart Rate. Data preprocessing steps involved handling missing values, normalization, and categorical encoding. The dataset was split into training (80%) and testing (20%) subsets. Various machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were applied.
Performance was measured using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics.

5. Results and Discussion

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	82%	80%	79%	79%
Random Forest	89%	87%	88%	87%
SVM	84%	83%	82%	82%
KNN	81%	79%	78%	78%



6. Conclusion

This research demonstrates that the integration of machine learning into healthcare analytics can effectively predict disease risks with high accuracy. Among the tested models, Random Forest performed best, achieving approximately 89% accuracy. Predictive analytics aids in early diagnosis, helping clinicians move from treatment-based to prevention-focused approaches. Future work could integrate Deep Learning and Internet of Things (IoT)-based monitoring systems to further improve real-time prediction and healthcare management.

References

1. UCI Machine Learning Repository – Heart Disease Dataset
2. Kaggle – Diabetes Prediction Dataset
3. Kumar, S. et al. (2023). Machine Learning for Predictive Healthcare Analytics. Springer.
4. Rajan, A. & Patel, D. (2022). Comparative Study of ML Algorithms in Disease Prediction. IEEE Access.
5. scikit-learn documentation (https://scikit-learn.org)