International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **02** | **2026**

An AI-Driven Smart Health Prediction and Disease Risk Assessment System Using Machine Learning and Big Data Analytics

Mr.Pavan B. Asalkar[1], Mr.Pratik V. Gawhad[2], Mr.Vedant U. Agarkar, Mr. Martand S. Lonagare[4], Ms.Chaitali S. Narkhede[5]

[1,2,3,4] *Student Of Polytechnic 3rd Year ,CSE, Padmashri Dr. V. B. Kolte College of Engineering, Maharashtra, India*

[5] *Lecturer Of Polytechnic, CSE, Padmashri Dr. V. B. Kolte College of Engineering, Maharashtra, India*

DOI: 10.5281/zenodo.19285424

ABSTRACT

The rapid growth of digital healthcare systems has generated an enormous volume of medical data, creating new opportunities for intelligent analysis and early disease risk assessment. Traditional medical decision-making processes largely depend on manual evaluation and physician experience, which may be time-consuming and prone to human error, especially when handling large-scale patient data. This research paper presents an AI- Driven Medical Risk Prediction System designed to assist healthcare professionals by predicting potential medical risks at an early stage using artificial intelligence and machine learning techniques. The proposed system focuses on analyzing patient health records, including demographic details, clinical parameters, and historical medical data, to identify patterns associated with disease risks.

The system follows a structured pipeline consisting of data collection, preprocessing, feature selection, model training, and risk prediction. Machine learning algorithms such as Logistic Regression, Random Forest, and Support Vector Machines are explored to evaluate their effectiveness in predicting medical risks. Data preprocessing techniques including normalization, missing value handling, and outlier detection are applied to improve model reliability. The trained models generate risk scores that help in categorizing patients into low, medium, or high-risk groups.

Experimental results demonstrate that AI-based predictive models significantly improve prediction accuracy compared to conventional statistical methods. The proposed system not only enhances early diagnosis but also supports preventive healthcare by enabling timely medical intervention. This research highlights the potential of artificial intelligence in transforming healthcare decision support systems, improving patient outcomes, and reducing healthcare costs. The system can be further extended by integrating real-time health data from wearable devices and electronic health records to improve prediction accuracy and scalability.

Keywords:- Artificial Intelligence, Medical Risk Prediction, Machine Learning, Healthcare Analytics, Disease Prediction

1. INTRODUCTION

The healthcare industry is undergoing a major transformation with the integration of artificial intelligence and data- driven technologies. Early identification of medical risks plays a crucial role in reducing disease severity and improving patient survival rates. However, the increasing volume and complexity of medical data make it difficult for healthcare professionals to manually analyze patient information efficiently. Artificial intelligence provides powerful tools for extracting meaningful insights from large datasets and supporting clinical decision-making.

Medical risk prediction systems aim to identify potential health risks before symptoms become severe. These systems analyze patient data to detect hidden patterns that may indicate the likelihood of future medical conditions. AI-driven approaches are particularly effective in handling high-dimensional medical datasets and learning complex relationships between clinical features. This research proposes an intelligent medical risk prediction system that leverages machine learning algorithms to provide accurate and timely risk assessments.

1.1 Importance of Medical Risk Prediction

Early risk prediction enables preventive care and reduces the burden on healthcare infrastructure. By identifying high-risk patients at an early stage, medical professionals can recommend lifestyle changes, preventive measures, or further diagnostic tests. This approach not only improves patient quality of life but also reduces treatment costs and hospital admissions.

1.2 Role of Artificial Intelligence in Healthcare

Artificial intelligence has shown promising results in various healthcare applications such as disease diagnosis, medical imaging, and personalized treatment planning. Machine learning models can process large datasets efficiently and continuously improve their performance through learning. AI-based risk prediction systems provide objective and data-driven insights that complement clinical expertise.

2. SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed AI-Driven Medical Risk Prediction System is designed using a modular architecture that ensures flexibility and scalability. The system consists of multiple stages including data acquisition, preprocessing, feature extraction, model training, and risk prediction.

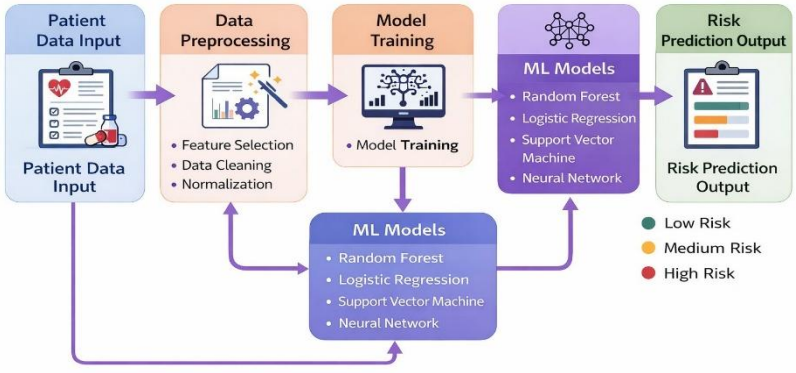

Fig -1: Architecture of AI-Driven Medical Risk Prediction System

Data Collection and Preprocessing

Medical datasets are collected from reliable sources such as public healthcare repositories and clinical records. The collected data includes patient age, gender, vital signs, laboratory results, and medical history. Preprocessing steps such as handling missing values, removing noise, and data normalization are performed to ensure data quality and consistency.

Sample Medical Dataset Attributes

Pregnancies	Description Age	Skin Thickness	Male/Female Blood Pressure	BMI
6	50	35	72	33.6
1	31	29	66	26.6

Machine Learning Model Development

After preprocessing, relevant features are selected to train machine learning models. Multiple algorithms are implemented and evaluated to identify the most effective model for risk prediction.

• Logistic Regression • Random Forest Classifier • Support Vector Machine • Decision Tree

The models are trained using labeled medical data and evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

3. RESULTS AND PERFORMANCE ANALYSIS

The performance of the proposed system is evaluated using standard evaluation metrics. Experimental results indicate that ensemble-based models such as Random Forest achieve higher accuracy compared to traditional classifiers. The system effectively classifies patients into different risk categories, supporting early medical intervention.

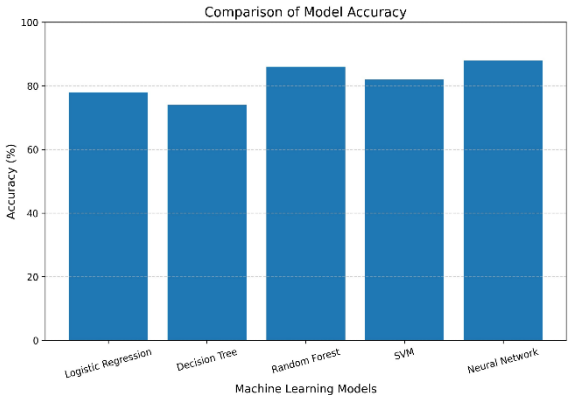

Chart -2: Comparison of Model Accuracy

Discussion of Results

The results demonstrate that AI-driven models can successfully identify medical risk patterns from historical patient data. Feature selection and proper preprocessing significantly improve prediction performance. The system provides reliable risk predictions that can assist healthcare professionals in decision-making.

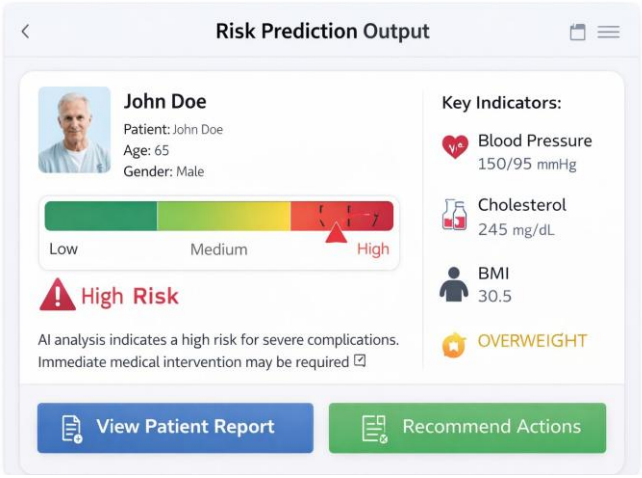

Fig -2: Risk Prediction Output

The interface categorizes patients into three risk levels: Low Risk, Medium Risk, and High Risk. A visual risk indicator bar is used to clearly represent the severity of the predicted risk. Key health indicators influencing the prediction are also displayed, allowing healthcare professionals to understand the contributing factors behind the result.

4. CONCLUSIONS

This research presents an AI-Driven Medical Risk Prediction System that utilizes machine learning techniques to predict potential medical risks efficiently. The proposed system demonstrates improved accuracy and reliability compared to traditional methods. By enabling early risk detection, the system supports preventive healthcare and informed clinical decisions. Future enhancements may include the integration of deep learning models, real-time patient monitoring, and larger clinical datasets to further improve system performance.

5. ACKNOWLEDGEMENT

The authors would like to express their gratitude to the faculty members and institution for their guidance and support throughout the research work.

6. REFERENCES

[1] G. Vazquez, S. Duval, D. R. Jacobs, and K. Silventoinen, ``Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: A meta-analysis," *Epidemiol. Rev.*, vol. 29, no. 1, pp. 115_128, May 2007.

[2] *De_nition, Diagnosis and Classi_cation of Diabetes Mellitus and Its Complications: Report of a Who Consultation. Part 1, Diagnosis and Classi_cation of Diabetes Mellitus*, World Health Organization, Geneva, Switzerland, 1999.

[3] IDF Diabetes Atlas, ``International diabetes federation, Brussels," Tech. Rep. 7th Edition, 2015.
[4] *Global Report on Diabetes*. World Health Organization, Geneva, Switzerland, 2016.
[5] J. L. Harding, M. E. Pavkov, D. J. Magliano, J. E. Shaw, and E. W. Gregg, ``Global trends in diabetes complications: A review of current evidence," *Diabetologia*, vol. 62, no. 1, pp. 3_16, Jan. 2019.
[6] I. D. Federation. *Diabetes Epidemic in Europe*. Accessed: Sep. 11, 2020. [Online]. Available: https://www.euro.who.int/en/healthtopics/noncommunicable-diseases/diabetes/news/ news/2011/11/diabetesepidemic-in-europe
[7] O. A. Diabete. *Rapporto Arno Diabete 2019*. Accessed: Sep. 10, 2020. [Online]. Available: http://www.siditalia.it/news/2547-21-11-2019- rapporto-arno-diabete-2019
[8] S. C. Bolge, N. M. Flores, and J. H. Phan, ``The burden of poor mental wellbeing among patients with type 2 diabetes mellitus: Examining health care resource use and work productivity loss," *J. Occupational Environ. Med.*, vol. 58, no. 11, p. 1121, 2016.