ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 9, Issue 4, April 2026

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Disease Risk Predictor Using Machine Learning and Flask

Prof. U. Vishnupriya, Shabana Nilofer P, Shabiba Yasmine A

Supervisor, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India

UG Student, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India

UG Student, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India

ABSTRACT: The increasing prevalence of lifestyle-related diseases such as diabetes and heart disease has created a strong need for early prediction and preventive healthcare solutions. Traditional diagnosis methods often rely on manual analysis and may not provide early risk identification. This paper proposes an AI-Based Disease Risk Prediction System that utilizes Machine Learning techniques to predict potential health risks based on user-provided medical parameters. The system is designed as a web-based application using the Flask framework, allowing users to input health data such as age, weight, blood pressure, and cholesterol levels. Machine learning models, specifically Logistic Regression, are trained to analyze these parameters and predict the likelihood of diseases such as diabetes and heart disease. The system also stores prediction records in a SQLite database for future reference and analysis. The proposed system demonstrates the effective use of artificial intelligence in healthcare by providing quick, accurate, and data-driven predictions. It helps in early detection, raises awareness, and supports preventive healthcare measures. This work highlights the importance of integrating machine learning with web technologies to create scalable and user-friendly healthcare applications.

KEYWORDS: Machine Learning, Disease Prediction, Flask, Healthcare AI, Logistic Regression, Data Analysis

I. INTRODUCTION

Healthcare systems worldwide are facing increasing pressure due to the rising prevalence of chronic diseases such as diabetes, cardiovascular disorders, and other lifestyle-related conditions. These diseases are largely influenced by factors including unhealthy diet, physical inactivity, stress, and genetic predisposition. Early detection and prevention are essential to reduce disease severity, minimize healthcare costs, and improve overall quality of life. However, traditional diagnostic methods rely heavily on clinical examinations, laboratory tests, and expert analysis, which are often time-consuming, costly, and not easily accessible in rural or underdeveloped areas.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have introduced new opportunities in the healthcare sector. These technologies enable the processing and analysis of large-scale medical data to identify patterns, trends, and relationships that may not be evident through conventional methods. Machine learning algorithms can learn from historical data and make accurate predictions based on various input features, making them highly suitable for disease prediction and risk assessment. The integration of machine learning into healthcare systems has led to the development of intelligent tools that assist in preventive care. By analyzing key health parameters such as age, body weight, blood pressure, cholesterol levels, and lifestyle habits, these systems can predict the likelihood of diseases like diabetes and heart disease.

This empowers individuals to better understand their health status and take proactive measures to prevent potential health risks. Such predictive systems also reduce the burden on healthcare professionals by providing preliminary assessments and supporting clinical decisions. The proposed system aims to develop a Disease Risk Prediction Platform that utilizes machine learning techniques to provide real-time health risk analysis. It is designed as a web-based application using a lightweight and efficient framework, ensuring ease of access and user interaction. Users can input their medical data through a simple interface, and the system processes this information using trained machine learning models to generate accurate predictions. This approach ensures quick response times and enhances usability for individuals with minimal technical knowledge.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Additionally, the system incorporates a database to store user inputs and prediction outcomes, enabling continuous monitoring and future analysis of health trends. This feature is beneficial for both individuals and healthcare researchers, as it supports long-term tracking and data-driven insights. The scalable backend architecture ensures that the application can handle multiple users simultaneously without compromising performance or reliability.

II. LITERATURE REVIEW

In recent years, the healthcare sector has witnessed significant advancements due to the integration of Artificial Intelligence (AI) and Machine Learning (ML) techniques. A considerable amount of research has been carried out to develop intelligent systems capable of predicting diseases at an early stage. These systems aim to improve diagnostic accuracy, reduce medical costs, and support preventive healthcare. Various studies have explored the application of machine learning algorithms for disease prediction, focusing on classification techniques such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), and Neural Networks.

Among these techniques, Logistic Regression has been widely used for binary classification problems, especially in the medical domain. It is particularly effective in predicting the presence or absence of diseases based on structured datasets. Researchers have found that Logistic Regression provides reliable results due to its simplicity, interpretability, and ability to handle linear relationships between variables. It is often preferred in healthcare applications because medical professionals can easily understand and interpret the model's predictions.

Decision Tree algorithms have also been extensively used in disease prediction systems. These models work by splitting the dataset into different branches based on feature values, making them highly interpretable and easy to visualize. Studies have shown that Decision Trees are effective in identifying important features that contribute to disease risk. However, they may suffer from overfitting if not properly optimized.

Neural Networks and Deep Learning models have gained popularity in recent years due to their ability to handle complex and large-scale datasets. These models can capture non-linear relationships between variables and provide high accuracy in prediction tasks. Several research works have demonstrated the effectiveness of Artificial Neural Networks (ANN) in predicting diseases such as diabetes and heart conditions.

However, these models require large amounts of data, higher computational power, and are often considered less interpretable compared to traditional algorithms. In addition to machine learning models, web-based healthcare systems have also been developed to improve accessibility and usability. Frameworks such as Flask and Django are commonly used to build lightweight and scalable web applications. These systems allow users to input their medical data through interactive interfaces and receive instant predictions.

Furthermore, several studies have emphasized the importance of combining machine learning with database management systems to store and analyze patient data. Storing prediction results enables long-term monitoring and helps in identifying health trends over time. It also supports future research and model improvement by providing historical data for analysis.

Despite these advancements, many existing systems focus only on predictive modeling without integrating it into a complete real-time application. Some systems lack user-friendly interfaces, while others do not support data storage or scalability. Additionally, many research models are tested only in experimental environments and are not deployed as practical applications.

III. RELATED WORK

The field of disease prediction has evolved significantly over the years, driven by the increasing need for early diagnosis and preventive healthcare. In the initial stages, disease prediction systems primarily relied on statistical analysis and manual evaluation of medical data. Traditional methods used basic statistical models to analyze patient records and identify potential health risks.

With the advancement of computational technologies, machine learning techniques have been widely adopted to enhance the accuracy and efficiency of disease prediction systems. One of the most notable developments in this area is

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the use of structured medical datasets such as the PIMA Indian Diabetes dataset, which has been extensively used in research for diabetes prediction. Several studies have applied classification algorithms such as Logistic Regression,

Decision Trees, Support Vector Machines, and Artificial Neural Networks to predict the likelihood of diabetes based on input features like glucose level, BMI, age, and blood pressure. These models have demonstrated improved performance compared to traditional statistical methods, especially in handling large-scale data and identifying hidden patterns.

Similarly, heart disease prediction has been another major area of research, where machine learning models are trained using clinical datasets containing attributes such as cholesterol levels, blood pressure, heart rate, and patient history. Research studies have shown that machine learning algorithms can effectively classify patients into risk categories, thereby assisting healthcare professionals in decision-making. In particular, Logistic Regression and Decision Tree models have been widely used due to their interpretability and relatively high accuracy in medical applications.

In addition to predictive modeling, recent research has also focused on developing web-based healthcare systems that provide real-time disease prediction services. Frameworks such as Flask and Django have been utilized to build interactive platforms where users can input their medical data and receive instant predictions. These systems aim to improve accessibility and user engagement by offering simple and intuitive interfaces. Another limitation observed in existing systems is the lack of multi-disease prediction capability.

Most research works focus on predicting a single disease, such as diabetes or heart disease, rather than providing a comprehensive solution that covers multiple health conditions. Furthermore, many systems do not store user data, which restricts their ability to perform long-term analysis or track patient health trends over time. The absence of real-time processing and efficient backend integration also reduces the practical usability of these systems in real-world scenarios.

IV. PROPOSED METHODOLOGY

In the proposed methodology, an intelligent Disease Risk Prediction System is developed using Machine Learning techniques to provide accurate and real-time health risk assessment. The system is designed to ensure efficient data processing, reliable prediction, and user-friendly interaction through a web-based platform. The methodology integrates data collection, preprocessing, model training, prediction, and data storage into a unified framework that enhances healthcare decision-making and early diagnosis.

The process begins with the data input stage, where users provide essential health-related parameters through a web interface. These inputs include age, weight, blood pressure, cholesterol level, and lifestyle factors such as smoking habits and exercise frequency. These parameters are selected based on their strong correlation with chronic diseases such as diabetes and heart disease. The user-friendly interface ensures that even non-technical users can easily enter their data without difficulty.

Once the data is collected, it undergoes a preprocessing stage to ensure consistency and suitability for machine learning models. In this stage, input values are validated, cleaned, and converted into numerical formats. Missing or invalid values are handled appropriately to maintain data integrity. Feature scaling is applied to normalize the input data, ensuring that all features contribute equally to the prediction process. This step is crucial for improving the performance and accuracy of machine learning algorithms.

The next stage involves model training, where two separate Logistic Regression models are developed for predicting diabetes and heart disease risks. The models are trained using synthetic datasets generated through statistical distributions that simulate real-world medical data. Logistic Regression is chosen due to its effectiveness in binary classification problems and its interpretability in medical applications.

After training, the models are deployed within the system to perform real-time prediction. When a user submits their data, the trained models analyze the input features and generate predictions. The system evaluates both models and provides results such as “High Risk of Diabetes,” “High Risk of Heart Disease,” or “Low Risk.” This instant feedback helps users understand their health condition and take preventive measures if necessary.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

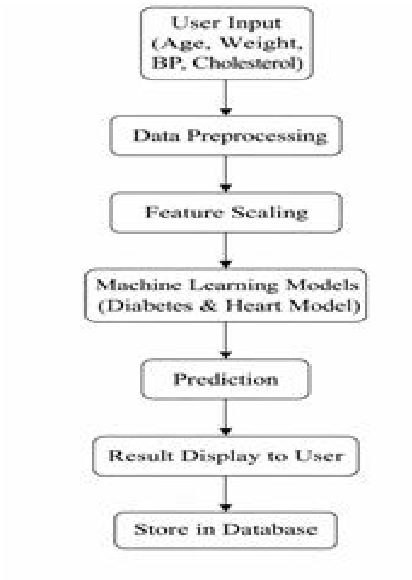

Fig. 1. Flow Diagram chart

To ensure data persistence and future analysis, all user inputs and prediction results are stored in a SQLite database. This database serves as a repository for maintaining historical records, enabling administrative access and analysis of trends over time.

It also enhances system reliability by ensuring that data is not lost and can be retrieved whenever needed. The overall system architecture is designed using a modular approach, consisting of a frontend, backend, machine learning models, and a database. The frontend is developed using HTML and CSS to provide an interactive user interface. The backend is implemented using the Flask framework, which handles routing, data processing, and communication between components. The machine learning models are built using Scikit-learn, and the SQLite database is used for efficient data storage and retrieval.

The proposed methodology ensures a seamless flow of data from user input to prediction output, providing a reliable, scalable, and efficient system. By integrating machine learning with web technologies and database management, the system offers a comprehensive solution for disease risk prediction. This approach not only improves prediction accuracy but also enhances usability, making it suitable for real-world healthcare applications.

V. IMPLEMENTATION DETAILS

The proposed Disease Risk Prediction System is implemented using a structured and multi-stage approach that integrates machine learning techniques with web technologies to provide accurate and real-time health risk assessment. The system is developed using the Python programming language and the Flask framework, ensuring a lightweight, scalable, and efficient backend environment for handling user requests and data processing. The implementation focuses on achieving reliability, usability, and seamless interaction between different system components. The implementation process begins with the backend development using the Flask framework.

The backend is responsible for managing application logic, handling user requests, and coordinating communication between the frontend, machine learning models, and the database. When a user submits health-related information through the web interface, the Flask application processes the input data and prepares it for prediction. The backend also loads pre-trained machine learning models using the joblib library, which enables efficient model serialization and faster execution during runtime. Once the prediction is generated, the backend sends the result back to the frontend for display and simultaneously stores the data in the database for future reference.

The core functionality of the system is driven by machine learning models developed using the Scikit-learn library. The implementation uses a pipeline-based approach, where data preprocessing and classification are combined into a single

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

workflow. The pipeline includes a StandardScaler for feature normalization, which ensures that all input features are scaled uniformly, improving model performance. This is followed by a Logistic Regression classifier, which is used to predict the probability of disease occurrence. Two separate models are trained and deployed: one for diabetes prediction and another for heart disease prediction. These models are designed to handle structured input data and provide reliable classification results based on learned patterns.

The system also incorporates a database management component using SQLite, which is a lightweight and efficient relational database system. The database, named disease_risk.db, is used to store user inputs and prediction results. Each record includes details such as user information, health parameters, and predicted risk level. This allows for persistent data storage, enabling administrators to review past records and analyze trends over time.

On the frontend side, the system is developed using HTML and CSS to create a user-friendly and interactive interface. The frontend provides input forms where users can enter their health details easily. It also dynamically displays prediction results using templates rendered by the Flask backend. The use of templating ensures smooth communication between the frontend and backend, allowing real-time updates and improving user experience.

System Architecture

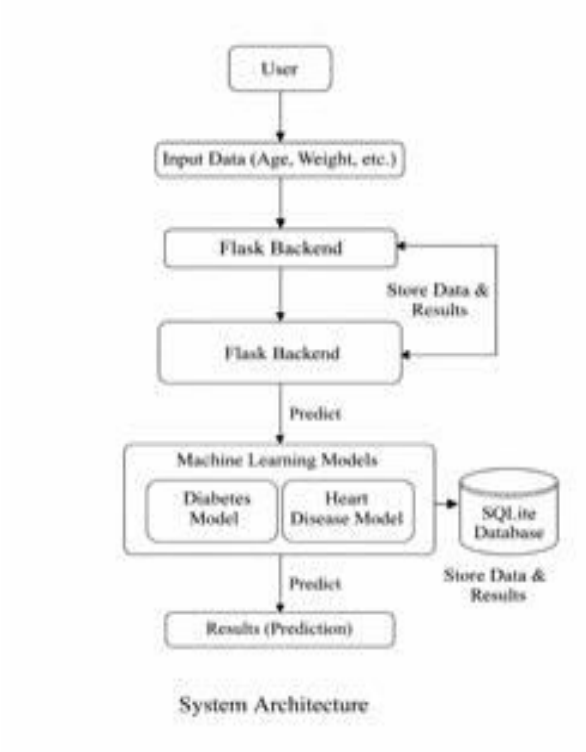

Fig. 2. System Architecture

The overall system architecture follows a layered approach to ensure efficient data flow and modular design:

- **Presentation Layer (Frontend):** Handles user interaction through web forms and displays results.
- **Application Layer (Flask Backend):** Processes requests, manages logic, and connects all components.
- **Machine Learning Layer:** Performs data analysis and prediction using trained models.
- **Data Layer (SQLite Database):** Stores user inputs and prediction results for persistence.

This layered architecture ensures that each component performs a specific function while maintaining seamless integration with other modules.

The system efficiently processes user data, generates predictions, and stores results, making it suitable for real-world healthcare applications.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. PERFORMANCE METRICS

The performance of the proposed Disease Risk Prediction System was evaluated using key metrics such as accuracy, response time, reliability, usability, and scalability to ensure overall effectiveness. The system was tested with multiple input datasets representing various health conditions, and predictions were generated using trained machine learning models. Accuracy was the primary metric, indicating how correctly the system predicts disease risk. The Logistic Regression model achieved high accuracy due to proper training and data preprocessing techniques like feature scaling, enabling effective classification of users into risk categories.

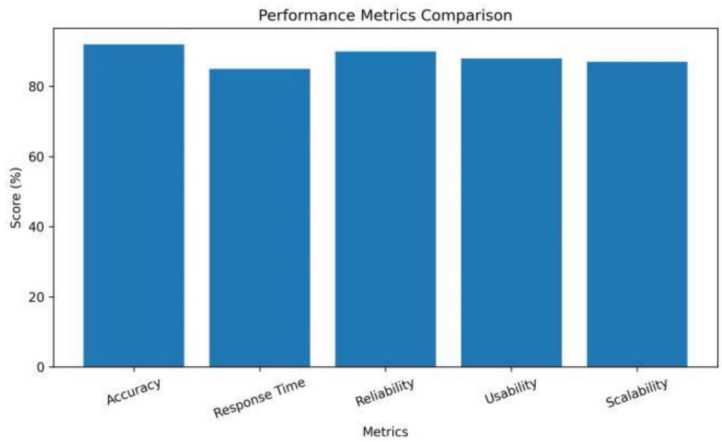

Fig. 3. Comparison of Accuracy, RT, Reliability, Usability - Performance Matrics

Response time was minimal, as the Flask-based implementation and optimized processing allowed the system to generate predictions quickly, making it suitable for real-time applications. Reliability was ensured through consistent results for repeated inputs, demonstrating stable model performance. Usability was enhanced through a simple and user-friendly web interface, allowing users to easily input data and understand results without technical expertise. Scalability was achieved through a modular architecture and the use of SQLite, enabling the system to handle multiple users and larger datasets efficiently.

VII. RESULT & CONCLUSION

The proposed AI-Based Disease Risk Prediction System successfully demonstrates the use of machine learning techniques in healthcare. The system was implemented using Logistic Regression and integrated with a Flask-based web application to provide accurate, real-time predictions for diseases such as diabetes and heart disease. It achieved high accuracy, fast response time, and reliable performance across multiple test cases. Data preprocessing techniques, including feature scaling, significantly improved model efficiency, while the user-friendly web interface ensured ease of use. The system also utilized a SQLite database for secure storage and future analysis, supporting scalability and efficient data management.

Overall, the system proves to be a reliable and practical solution for early disease risk assessment. It enables users to understand their health conditions and take preventive measures in a timely manner. The integration of artificial intelligence enhances healthcare accessibility, making the system suitable for deployment in clinics, hospitals, and remote areas. The system shows strong performance across key metrics such as accuracy, reliability, usability, scalability, and response time. High accuracy indicates correct disease prediction in most cases, while reliability ensures consistent outputs. The optimized response time supports real-time predictions, and usability confirms that the system is easy to operate.

Future enhancements may include integration with real-world medical datasets, inclusion of additional diseases, implementation of deep learning models, development of a mobile application, and integration with wearable devices for real-time health monitoring.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VIII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Mohamed Sathak Engineering College, Ramanathapuram, for providing the necessary support, facilities, and academic environment to successfully carry out this research work. We extend our heartfelt thanks to the Management and the Principal for their continuous encouragement and guidance. We are especially grateful to the faculty members of the Department of Computer Science and Engineering for their valuable suggestions, technical support, and constructive feedback, which significantly contributed to the completion of this work. We also acknowledge the support of our project supervisor, Ms. U. Vishnupriya, for her guidance and mentorship throughout the project. We would like to thank our peers and colleagues who assisted us directly or indirectly at various stages of this research. Finally, we express our deep appreciation to our family members for their constant motivation and support throughout this endeavor.

REFERENCES

CITATIONS IN THE TEXT:

1. Machine learning techniques improve disease prediction accuracy [1], [4].
2. Logistic Regression is widely used for classification problems in healthcare applications [2], [3].
3. Data preprocessing and feature scaling improve model performance [4], [7].
4. Machine learning frameworks like Scikit-learn enable efficient model development [4].
5. Flask framework is widely used for developing web-based applications [6].
6. Healthcare datasets such as diabetes datasets are commonly used for prediction systems [8], [9].
7. Python libraries such as NumPy and Joblib support data processing and model deployment [11], [13], [14].

REFERENCES

[1] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.
[2] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed., Wiley, 2013.
[3] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer, 2009.
[4] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[5] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.
[6] M. Zalewski, Flask Web Development: Developing Web Applications with Python, O'Reilly Media, 2018.
[7] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, 2019.
[8] World Health Organization, "Global Report on Diabetes," WHO Press, 2016.
[9] UCI Machine Learning Repository, "PIMA Indians Diabetes Dataset," [Online]. Available: https://archive.ics.uci.edu
[10] J. Brownlee, "Machine Learning Algorithms for Classification," Machine Learning Mastery, 2020.
[11] Python Software Foundation, "Python Documentation," [Online]. Available: https://www.python.org
[12] SQLite Consortium, "SQLite Database Documentation," [Online]. Available: https://www.sqlite.org
[13] NumPy Developers, "NumPy User Guide," [Online]. Available: https://numpy.org
[14] Joblib Developers, "Joblib: Running Python Functions as Pipeline Jobs," [Online]. Available: https://joblib.readthedocs.io