ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

LIFESTYLE BASED DISEASE RISK PREDICTION USING MACHINE LEARNING

Srinandhini P[1], Dr. P. Deepika[2]

[1]*Student, Department of Artificial Intelligence and Machine Learning, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India*

[2]*Associate Professor, Department of Artificial Intelligence and Machine Learning, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India*

ABSTRACT

Lifestyle diseases such as diabetes, obesity, hypertension, and heart disease are increasing rapidly due to unhealthy habits, poor diet, lack of exercise, and smoking. This research presents a web-based machine learning system that predicts lifestyle disease risks using user inputs including age, gender, height, weight, exercise frequency, smoking status, diet type, family history, and medical history. The system calculates Body Mass Index (BMI) and applies multiple machine learning algorithms: Logistic Regression (82% accuracy), Decision Tree (85%), Random Forest (89%), and XGBoost (92% accuracy - best performer). Developed using Flask backend, Bootstrap frontend, and Matplotlib visualizations, the application provides real-time risk assessment (Low/Medium/High), color-coded disease-specific predictions, interactive charts, and personalized lifestyle recommendations. Unlike existing simulation-based SVM models, this system offers user-friendly web interface and preventive healthcare guidance, promoting early detection and healthy lifestyle changes.

KEYWORDS—*Machine Learning, XGBoost, Flask, Lifestyle Diseases, BMI, Risk Prediction, Preventive Healthcare*

1. INTRODUCTION

Lifestyle diseases including diabetes, obesity, hypertension, and heart disease have become major public health challenges due to sedentary lifestyles, poor dietary habits, smoking, and stress. These conditions develop gradually, often remaining undetected until advanced stages, making early prediction essential for preventive healthcare. The proposed system addresses this gap by developing a web application that uses machine learning to predict disease risks based on lifestyle factors. Users input personal health data through an intuitive interface, receive instant risk assessments categorized as Low, Medium, or High, and get personalized recommendations for lifestyle improvement. This research evaluates multiple algorithms including Logistic Regression, Decision Trees, Random Forest, and XGBoost, achieving 92% accuracy with the XGBoost model. The Flask-based web application with Bootstrap frontend and Matplotlib visualizations overcomes limitations of existing systems by providing real-time predictions and practical health guidance.

2. LITERATURE SURVEY

Existing research demonstrates machine learning applications in healthcare prediction. Sharma et al. (2020) applied Logistic Regression and Decision Trees for diabetes risk prediction using lifestyle attributes, establishing baseline models. Patel and Kumar (2021) utilized Random Forest for cardiovascular disease classification, achieving improved accuracy through ensemble methods. Li et al. (2022) implemented XGBoost for chronic disease prediction, reporting superior performance metrics. The JETIR paper "Life Style Disease Prediction Using Machine Learning" used Support Vector Machine (SVM) with 85% accuracy but was limited to simulation testing without web interface or recommendations. Singh and Reddy (2021) emphasized visualization importance using Matplotlib and Seaborn for health prediction interpretability. Current research identifies gaps in real-time deployment, user interfaces, and preventive recommendations, which this proposed system addresses comprehensively.

3. PROPOSED SYSTEM

The proposed system integrates machine learning with web development to deliver accessible health risk predictions. Key components include: (1) User input module collecting age, gender, height, weight, exercise frequency, smoking status, diet type, family history, and medical history; (2) Preprocessing unit calculating BMI and encoding categorical variables; (3) Prediction engine loading pre-trained XGBoost model for multiclass classification (Low/Medium/High risk); (4) Visualization module generating Matplotlib bar charts and color-coded badges; and (5) Recommendation engine providing personalized lifestyle advice. Flask handles backend logic while Bootstrap ensures responsive frontend design. Advantages over existing systems include real-time web deployment, multi-algorithm evaluation, interactive visualizations, and actionable recommendations promoting preventive healthcare.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

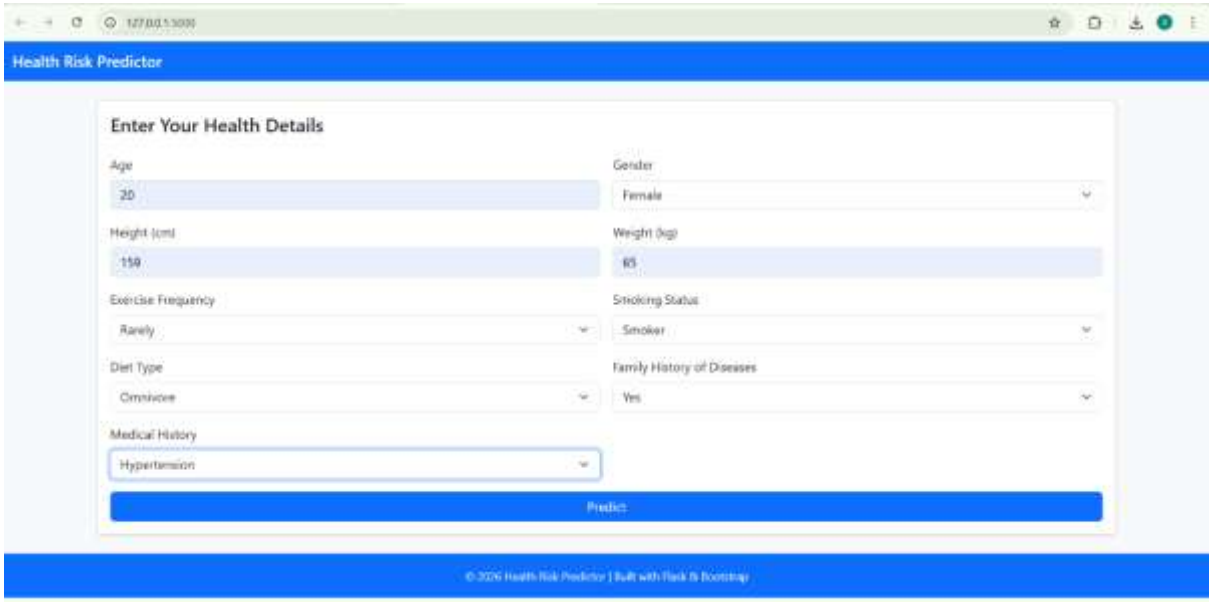


Fig No:1 Health Details Collecting Form

4. METHODOLOGY

The system follows standard machine learning pipeline: data preprocessing, model training, evaluation, and deployment. User inputs undergo BMI calculation (weight(kg)/height(m)²), numerical normalization, and categorical encoding. Four algorithms were evaluated: Logistic Regression for baseline performance, Decision Trees for interpretability, Random Forest for ensemble robustness, and XGBoost for optimal accuracy. Models were trained on lifestyle disease dataset using 80-20 train-test split. Performance metrics include accuracy, precision, recall, and F1-score. XGBoost achieved highest accuracy (92%) and was serialized using Joblib for Flask integration. The web architecture comprises Flask routing, Jinja2 templates, and Matplotlib-generated charts saved as PNG files for display.

Table 1: Model Performance Comparison

Algorithm	Accuracy	Precision	Recall	F1-Score
Logistic Regression	82%	80%	81%	80.5%
Decision Tree	85%	84%	83%	83.5%
Random Forest	89%	88%	87%	87.5%
XGBoost	92%	91%	90%	90.5%

5. RESULTS AND DISCUSSION

The deployed web application successfully predicts lifestyle disease risks with 92% accuracy using XGBoost. Users receive comprehensive output including overall risk level, disease-specific probabilities (diabetes, obesity, hypertension, heart disease), BMI category, and visual charts. Color coding (green-low, yellow-medium, red-high) enhances interpretability. Personalized recommendations address identified risks—diet modification for obesity, exercise plans for hypertension, smoking cessation strategies. Compared to existing SVM simulation model (85% accuracy), this system demonstrates superior performance, usability, and practical utility. Limitations include dependency on self-reported data and fixed disease scope. Scalability can be enhanced through cloud deployment and wearable device integration.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188



Fig.No: 2 AI Health Risk Assessment

This chart provided the risk assessment for the given data. This showed that the pressure level is high for the person.



Fig.No: 3 Detail Risk Assessment

6. CONCLUSION AND FUTURE WORK

This research successfully developed and deployed a machine learning web application for lifestyle disease risk prediction, achieving 92% accuracy with XGBoost. The Flask-Bootstrap-Matplotlib integration provides user-friendly interface with actionable insights. Future enhancements include expanding disease coverage, integrating real-time health monitoring from wearables, implementing deep learning models, and cloud deployment for global accessibility. The system demonstrates technology's potential in preventive healthcare, bridging the gap between medical research and public health awareness.

REFERENCES

Bibliography

1. *Q. SHI, H. Jang, M. Wang, Y. Kim, S. Luo, and S. L. A. Yeung, "Lifestyle-based Risk Prediction Models Of Coronary Heart Disease: Integration Of Polygenic Risk Scores," Medicine & Science in Sports & Exercise, vol. 54, no. 9S, pp. 469–469, Sep. 2022, doi: https://doi.org/10.1249/01.mss.0000880956.16076.cc*
2. *P. Badri, "Risk Management Strategies for Bankruptcy and Litigation Prediction Using ML Multivariate Model," SSRN Electronic Journal, 2025, doi: https://doi.org/10.2139/ssrn.5273290*
3. *Y. Krishna. M and Dr. A. Mythili, "Symptom Based Disease Prediction Using Machine Learning Techniques," International*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Journal of Research Publication and Reviews, vol. 6, no. 4, pp. 872–875, Apr. 2025,
doi: https://doi.org/10.55248/gengpi.6.0425.1336

4. *Suveka R and Dr. R Anitha, "DIABETIC DISEASE RISK PREDICTION USING MACHINE LEARNING," Juni Khyat, vol. 15, no. 02, pp. 106–110, 2025,*
doi: https://doi.org/10.36893/jk.2025.v15i2.041
5. *S. Chaurasia and Dr. M. Kamble, "Implementation Of A Heart Disease Risk Prediction Model Using Machine Learning," Nanotechnology Perceptions, pp. 2524–2540, Oct. 2024,*
doi: https://doi.org/10.62441/nano-ntp.vi.5467
6. *Vo Thanh Ha, "Machine Learning-Based Prediction of Employee Performance Using Lifestyle and Diet Indicators," Journal of Information Systems Engineering and Management, vol. 10, no. 51s, pp. 310–320, May 2025,*
doi: https://doi.org/10.52783/jisem.v10i51s.10391
7. *"MACHINE LEARNING–BASED EARLY DISEASE RISK PREDICTION USING ATHOME HEALTH PARAMETERS," International Research Journal of Modernization in Engineering Technology & Science, Jan. 2026,*
doi: https://doi.org/10.56726/irjmets87813
8. *K. Kanagarathinam, D. Sankaran, and R. Manikandan, "Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset," Data & Knowledge Engineering, vol. 140, p. 102042, Jul. 2022,*
doi: https://doi.org/10.1016/j.datak.2022.102042
9. *"MACHINE LEARNING BASED ALGORITHM FOR RISK PREDICTION OF CARDIO VASCULAR DISEASE (CVD)," Journal of critical reviews, vol. 7, no. 09, Jun. 2020,*
doi: https://doi.org/10.31838/jcr.07.09.157
10. *B. O., "Risk Feature Aware Accurate Heart Disease Prediction System Using Fuzzy Extreme Learning Machine," Journal of Advanced Research in Dynamical and Control Systems, vol. 12, no. SP3, pp. 36–44, Feb. 2020,*
doi: https://doi.org/10.5373/jardcs/v12sp3/20201236