Lifestyle Disease Prediction and Recommendation Using Machine Learning

Ms. Rashmi G S[1], Prof. Naseema C A[2], Dr. Bhagyajyothi K L[3], Prof. Tajunnisa N M [4]
[1]*M.Tech Student, Department of CSE, KVG College of Engineering, Sullia, India*
[2, 4]*Assistant Professor, Department of CSE, KVG College of Engineering, Sullia, India*
[3] *Associate Professor, Department of CSE, KVG College of Engineering, Sullia, India*

Abstract: Poor dietary habits, sedentary lifestyle, and lack of physical activity often lead to lifestyle diseases such as diabetes, hypertension, and obesity. Early detection is vital to minimize their health impact and reduce strain on healthcare systems. This study presents a Lifestyle Disease Prediction and Recommendation System that analyzes individual health and lifestyle attributes to assess risk and offer personalized guidance using machine learning.

Four supervised algorithms—Support Vector Machine (SVM), Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN)—are used to classify individuals into low, medium, or high-risk categories. A structured preprocessing pipeline ensures data quality through cleaning, encoding, and feature scaling. Model performance is evaluated using metrics like accuracy, precision, recall, F1-score, and confusion matrix.

A lightweight, rule-based recommendation module provides lifestyle suggestions based on parameters like BMI, blood pressure, and blood sugar. By integrating prediction and guidance, this system highlights the role of machine learning in preventive healthcare and supports the development of accessible, user-friendly tools for proactive health management.

Keywords: Disease Risk Prediction, Lifestyle Diseases, Personalized Recommendations, Preventive Healthcare, Supervised Learning Algorithms

I. INTRODUCTION

Lifestyle diseases such as diabetes, hypertension, and obesity are growing global health issues, primarily driven by poor diet, inactivity, stress, and sleep problems (Rezaee et al., 2020; Ganie et al., 2024). Recognized by the WHO as a leading cause of mortality, these conditions highlight the importance of early detection and prevention. Traditional diagnosis often relies on retrospective clinical analysis, potentially delaying risk identification (Mishra et al., 2018).

Machine learning (ML) provides effective tools for early intervention by analyzing health data to identify risk patterns (Uddin et al., 2019). ML models can evaluate factors like BMI, blood pressure, and activity levels to estimate disease risk.

This study introduces a Lifestyle Disease Prediction and Recommendation System using four ML algorithms: SVM, Logistic Regression, Random Forest, and KNN. The system categorizes users by risk and offers personalized, rule-based recommendations (Abdulhadi & Al-Mousa, 2021).

Custom suggestions on diet, exercise, and stress management promote preventive care. The system uses structured preprocessing and standard evaluation metrics to ensure reliable performance. It aims to bridge disease prediction and actionable self-care, advancing personalized healthcare (Ren et al., 2023).

II. LITERATURE SURVEY

The application of ML in lifestyle disease prediction has gained momentum in recent years due to its capacity to extract hidden patterns from complex datasets and facilitate early diagnosis. Several studies have proposed ML-based diagnostic models across various health domains.

Rezaee et al. (2020) and Ganie et al. (2024) developed multi-disease prediction models focusing on cardiovascular conditions and obesity using ensemble learning, showcasing high accuracy but often requiring extensive feature sets and computational resources. Similarly, Mishra et al. (2018) proposed a healthcare decision support system that highlighted the effectiveness of automated reasoning in clinical predictions, though it lacked personalized lifestyle guidance.

Uddin et al. (2019) performed a comparative study of supervised algorithms for disease prediction, emphasizing model interpretability and generalization. Their findings established a strong foundation for model selection in healthcare contexts—guiding the inclusion of SVM, Logistic Regression, Random Forest, and KNN in the present study.

Patil et al. (2018) focused on lifestyle disease prediction using Support Vector Machine, achieving promising results with limited features. Shorewala (2021) explored ensemble methods for early detection of coronary heart disease, advocating for Random Forest due to its superior robustness and interpretability—echoed in the findings of this research.

For hybrid and adaptive frameworks, Ren et al. (2023) suggested integrating multiple ML models for higher predictive accuracy, while Tiwari et al. (2022) recommended ensemble approaches for cardiovascular disease diagnosis. These studies support the relevance of combining classic ML methods with rule-based personalization, as done in this project.

Furthermore, Abdulhadi and Al-Mousa (2021) evaluated multiple classifiers for diabetes detection and highlighted the importance of preprocessing and feature scaling, which was closely followed in the proposed work.

Lastly, the importance of incorporating real-time and lightweight systems for preventive healthcare in resource-limited environments was discussed by Sai Ganesh et al. (2024). Their recommendation to deploy ML-based platforms with minimal hardware demands strongly supports the direction of this study.

In summary, while previous works have demonstrated the strength of various ML models in disease prediction, a gap remains in integrating risk classification with personalized, rule-based recommendations and scalable deployment platforms. This study addresses that gap by presenting a unified, interpretable, and accessible solution for lifestyle disease prediction and prevention.

III. METHODOLOGY

This study aims to design and implement a machine learning-based system for the prediction of lifestyle diseases, along with a rule-based module for generating health recommendations. The methodology comprises data collection and preprocessing, model development using multiple supervised learning algorithms, performance evaluation, and integration of a rule-based recommendation system. The proposed system classifies individuals into lifestyle disease risk categories and suggests preventive lifestyle modifications based on key health indicators.

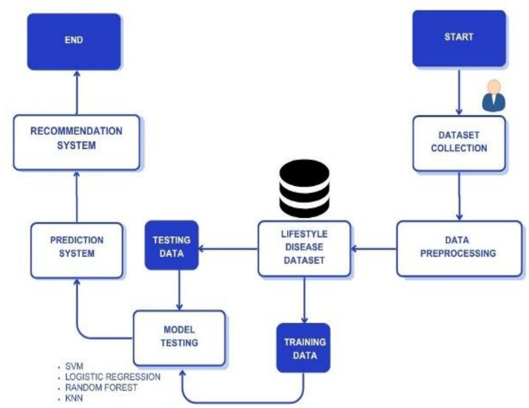

Fig. 1: Block diagram of the proposed system

A. Data Preprocessing

The dataset used in this study includes individual health and lifestyle attributes such as age, BMI, dietary habits, sleep quality, and stress levels, along with a class label indicating the lifestyle disease risk level (low, medium, or high). Initial data cleaning involved the removal of records with missing values to ensure data integrity and consistency. All variables were pre-encoded as integers, including categorical features like gender, smoking, and junk food intake, eliminating the need for further encoding. To enhance model performance—especially for algorithms sensitive to scale such as KNN and SVM—standardization was applied to all numerical features, centering values around zero mean with unit variance.

B. Model Evaluation

The models were evaluated using accuracy, precision, recall, F1-score, and a confusion matrix to assess prediction correctness and class-wise performance.

C. Recommendation System

A rule-based recommendation module was implemented to complement the risk prediction component. Based on the predicted risk level and selected input features, the system suggests tailored interventions including:

- Dietary modifications
- Exercise routines
- Sleep and stress management tips

The logic was implemented using a set of structured if–else conditions derived from general health guidelines and preventive care principles.

D. System Integration

To enhance usability and accessibility, the machine learning models and the recommendation engine were integrated into an interactive web-based system. The integration involved:

- A backend service for exposing the trained models via RESTful APIs
- A user-facing interface for data input, prediction display, and recommendation delivery
- Real-time interaction between the user and the system components
- Validation to ensure functional correctness, responsiveness, and ease of use

This architecture enables a seamless and efficient workflow from data input to actionable health insights, supporting proactive lifestyle management.

IV. RESULTS

A. Performance Comparison

The performance of the machine learning models—Support Vector Machine (SVM), Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN)—was evaluated on the test dataset using standard classification metrics. The results presented here highlight each model's ability to predict the risk of lifestyle diseases accurately.

Model	Train Accuracy	Test Accuracy
SVM	96.00%	92.00%
LOGISTIC	95.00%	93.00%
RANDOM FOREST	100.00%	93.00%
KNN	88.00%	79.00%

Fig. 2: Comparison of Performance

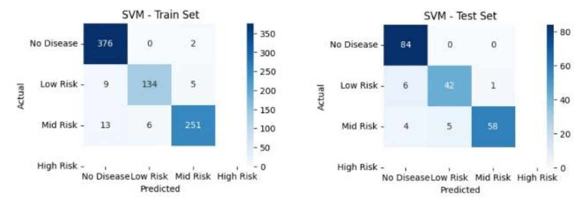

Fig. 3: Confusion Matrix of SVM Classifier on Training and Test Sets

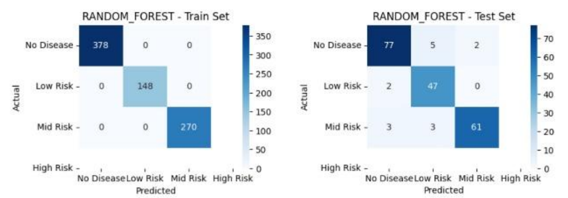

Fig. 4: Confusion Matrix of Random Forest Algorithm on Training and Test Sets

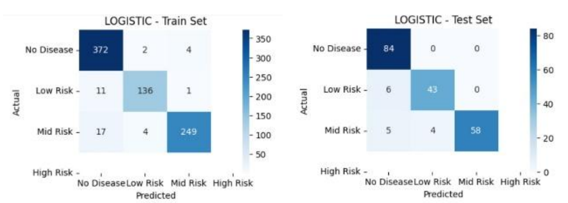

Fig. 5 Confusion Matrix of Logistic Regression Algorithm on Training and Test Sets

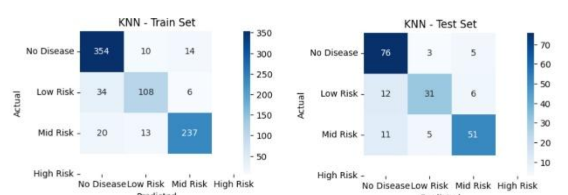

Fig. 6: Confusion Matrix of KNN Algorithm on Training and Test Sets

B. Recommendation Module Output

In addition to prediction, the system successfully generated tailored recommendations based on both the predicted risk level and specific user attributes such as

BMI, diabetes, and hypertension levels. Suggestions include:

- *High Risk*: "Consult a healthcare professional immediately. Strictly avoid processed food and sugar. Follow a medical diet and regular monitoring."
- *Medium Risk*: "Reduce intake of sugar and salty foods. Increase physical activity (30 mins daily). Manage stress through yoga/meditation."
- *Low Risk*: "Maintain a balanced diet and regular exercise. Continue monitoring your health periodically."

Further personalized tips were provided based on individual health parameters, focusing on diet control, stress reduction, and regular exercise.

C. Summary

Among the four models tested, the Random Forest algorithm consistently outperformed the others across all evaluation metrics. The integration of a rule-based recommendation system further enhanced the system's utility by providing users with immediate, personalized preventive suggestions based on their predicted risk level. Overall, the system demonstrated strong robustness, interpretability, and practical applicability for early lifestyle disease risk assessment.

V. CONCLUSION

This study introduces a ML-based system for predicting lifestyle disease risks and delivering preventive health recommendations. Using four supervised algorithms—SVM, Logistic Regression, Random Forest, and KNN—the system classifies users into low, medium, or high-risk categories based on health and lifestyle inputs. Random Forest outperformed the others in classification metrics, highlighting its effectiveness in healthcare prediction tasks.

To complement prediction, a rule-based recommendation engine provides users with personalized lifestyle advice aligned with their risk level. These suggestions are simple, actionable, and grounded in standard health guidelines, making the system practical for broader use, especially in resource-limited settings.

The findings support the role of machine learning in enabling early detection, personalization, and proactive health management. However, limitations such as the static nature of the recommendations and limited dataset diversity are acknowledged.

Future enhancements could include real-time data integration from wearables, exploration of deep learning or hybrid models, and deployment via mobile or telehealth platforms. Such advancements could boost accuracy, adaptability, and user reach, contributing to more accessible and preventive healthcare solutions.

REFERENCES

[1] Abdulhadi N, Al-Mousa A (2021) Diabetes detection using machine learning classification methods. In: 2021 International Conference on Information Technology (ICIT), pp 350–354. IEEE. https://doi.org/10.1109/ICIT52682.2021.9491788

[2] Ganie SM, Reddy BB, Hemachandran K, Rege M (2024) An investigation of ensemble learning techniques for obesity risk prediction using lifestyle data. *Decision Analytics J* 100539 https://doi.org/10.1016/j.dajour.2024.100539

[3] Mishra AK, Keserwani PK, Samaddar SG, Lamichaney HB, Mishra AK (2018) A decision support system in healthcare prediction. In: *Advanced Computational and Communication Paradigms*. Springer, Singapore, pp 156–167 https://doi.org/10.1007/978-981-10-8240-5_18

[4] Patil M, Lobo VB, Puranik P, Pawaskar A, Pai A, Mishra R (2018) A proposed model for lifestyle disease prediction using support vector machine. In: 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp 1–6. IEEE https://doi.org/10.1109/ICCCNT.2018.8493897

[5] Ren L, Wang T, Zhang H, Bouras A (2023) A prediction framework for lifestyle-related disease prediction using healthcare data. In: 2023 International Conference on Computer Applications Technology (CCAT), pp 94–99. IEEE https://doi.org/10.1109/CCAT59108.2023.00042

[6] Rezaee M, Putrenko I, Grolinger K, Capretz MAM (2020) Development and validation of risk prediction models for multiple cardiovascular diseases and type 2 diabetes. *PLoS ONE* 15(7):e0235758 https://doi.org/10.1371/journal.pone.0235758

[7] Sai Ganesh C, Sai Kishore R, Reddy BK, Rout SK (2024) Data-driven disease prediction and

lifestyle monitoring system. In: 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU). IEEE https://doi.org/10.1109/IC-CGU58078.2024.10530759

[8] Shorewala V (2021) Early detection of coronary heart disease using ensemble techniques. *Informatics Med Unlocked* 26:100655 https://doi.org/10.1016/j.imu.2021.100655

[9] Tiwari A, Chugh A, Sharma A (2022) Ensemble framework for cardiovascular disease prediction. *Comput Biol Med* 146:105624 https://doi.org/10.1016/j.compbiomed.2022.105624

[10] Uddin S, Khan A, Hossain ME, Moni MA (2019) Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak* 19(1):16 https://doi.org/10.1186/s12911-019-1004-8

[11] Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ (2010) Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak* 10:16 https://doi.org/10.1186/1472-6947-10-16