International Conference on Artificial Intelligence for Sustainable Development, Jaypee University Anoopshahr, India 2025, Volume 01, October 2025, pp. 223-229

PERSONALIZED HEALTH RECOMMENDATION SYSTEM USING MACHINE LEARNING

Ritik Goswami [1] and Anubhav Mishra [2]

[1, 2] Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India

mchandragoswami@gmail.com, anubhavmishraup62@gmail.com.

Abstract. The collection of health data has become ubiquitous. Wearable devices, fitness applications, and digitized health records culminate into a plethora of health data. While the volume of data continues to grow, the utilization of data in personalizing health recommendations remains scarce. Most recommendations offered are simple, generic, and often misguided. Machine learning can bridge this gap. Unlike advanced rule based systems, modern machine learning systems can dynamically adjust their recommendations as a person’s health status shifts. The focus of this project is the development of a machine learning-based, personalized health recommendation system. It has the ability to integrate disparate data sources to provide meaningful and actionable individual recommendations. These recommendations could stem from fitness wearables, medical history, data on sleep, dietary patterns, and additional health variables. Equally as important is the system’s emphasis on privacy, and trustworthiness, designed to promote sustained reach. Personalization is more flexibility but systems also need to incorporate principles of privacy, trustworthiness, and fairness in order to foster system adoption. The goal of such a system is to transform the healthcare process from reactively where issues are addressed after they arise — to proactive and preventative. With tailored recommendations, people can better understand and detect potential health issues, reducing the need for costly interventions. This is no longer merely an advancement in technology; it is an improvement to an intelligent system of care that puts the users — now more enlightened about their health — at the center of the process.

Keywords: Personal health ; Health recommendation engines ; Machine learning ; Data analysis ; Data domain ; Predictive modeling ; Health data and lifestyle choices ; Preventive health ; Human-centered design ; Live recommendations ; Device wearables ; Data privacy ; Health surveillance.

1. INTRODUCTION

In recent years, the rapid advancement of digital technologies has brought a major transformation in the healthcare sector. The widespread use of wearable devices, fitness trackers, and mobile health applications has enabled individuals to continuously monitor key aspects of their well-being—such as activity levels, sleep quality, diet patterns, heart rate, and other vital signs. This continuous flow of personalized data has shifted healthcare from a traditional reactive approach to a more proactive and individualized one, allowing for health advice that aligns with a person’s unique profile and lifestyle. Yet, many existing healthcare systems still use generalized or ”one-size-fits-all” recommendations. Generic recommendations like ”exercise regularly” or ”eat healthy” don’t take important differences in factors like age, genetics, medical history, and everyday habits into account. A recommendation suitable for one person may not be appropriate—or could even pose risks—for another. For instance, a standard exercise routine may not be ideal for individuals with joint problems or cardiovascular conditions. Such lack of personalization often limits the impact of preventive healthcare and reduces its ability to support meaningful lifestyle improvements. Machine Learning (ML) offers a promising solution to overcome these limitations. Unlike traditional rule-based systems, ML algorithms can process large volumes of diverse health data, recognize complex patterns, and generate adaptive, data driven insights. These systems are capable of learning from an individual’s ongoing behavioral and physiological changes, enabling them to provide dynamic and personalized health recommendations rather than static, one-time advice. This adaptive capability positions ML as a transformative tool for creating intelligent, personalized healthcare support systems that promote long-term well-being.

2. LITERATURE REVIEW

During the past decade, healthcare has been transformed by digital technologies, especially through wearable devices, mobile applications, and electronic health records. These tools have opened new opportunities to move beyond generalized advice toward personalized and context-sensitive health recommendations. Several studies have investigated different aspects of this, including activity tracking, diet recommendations, disease risk prediction, and real-time health monitoring [1] [2]. Wearable devices such as fitness trackers and smartwatches have been widely studied to monitor physical activity, heart rate, sleep patterns, and other physiological signals. Piwek et al. [3] highlighted that continuous monitoring through wearables provides users with meaningful information on their daily routines, encouraging healthier behavior. Similarly, research on mobile health applications demonstrates that personalized notifications and feedback significantly improve user engagement and adherence to fitness or dietary goals [4]. These findings underscore the importance of timely and individualized feedback in motivating behavioral changes. Diet and nutrition recommendation systems have also received considerable attention. Traditional systems rely on predefined dietary rules and general nutritional guidelines. However, recent studies have applied machine learning to analyze individual eating habits, metabolic rates, and health conditions, offering personalized meal suggestions. For example, Farooq et al. [5] demonstrated that predictive models can recommend diets tailored to individuals with conditions such as diabetes or cardiovascular disease, optimizing calorie intake while considering personal needs. This highlights the potential of data-driven personalization in improving dietary outcomes. Another important area of research is the prediction of disease risk using electronic health records. Machine learning models, including decision trees, random forests, and deep learning, have been successfully used to predict chronic conditions such as diabetes, hypertension, and cardiovascular diseases [6] [7]. Choi et al. [6] and Miotto et al. [7] showed that ML models can detect hidden patterns in longitudinal health data that humans are unable to recognize, allowing early intervention and preventive care. When combined with lifestyle and behavioral data, these models provide a holistic view of individual risk factors. Despite these advances, current research shows several limitations. Most systems operate in isolation, focusing on activity, diet, or medical history separately, leading to fragmented recommendations. Additionally, many models are static or rule-based, lacking the ability to adapt dynamically to changes in a person's health or lifestyle. Integrating heterogeneous health data remains a challenge, along with concerns regarding privacy, scalability, and interpretability [8][9]. Recent studies suggest that integrating multiple data sources with machine learning can create adaptive, personalized health recommendation systems [10][11]. By combining wearable data, diet logs, sleep patterns, and medical records, ML models can provide real-time, actionable recommendations. Continuous feedback loops allow these systems to evolve alongside the user's health, enhancing accuracy and relevance over time. Moreover, ensuring transparent and interpretable models is crucial for maintaining user trust and ethical application in healthcare [12]. In conclusion, while research has made significant progress in wearable tracking, dietary recommendations, and disease prediction, there is still a clear gap in developing fully integrated, adaptive, personalized health recommendation systems. The use of multi-source health data in combination with machine learning represents a powerful opportunity to provide individuals with smarter, more personalized health coaching that has the potential to be a significant shift away from a reactive model for providing healthcare to a proactive and preventive model.

3. METHODOLOGY USED

The approach used in this research seeks to build a personalized health recommendation system using the Health and Lifestyle Dataset from Kaggle. The process is broken down into three main phases: understanding the dataset, preprocessing the data, and selecting features. Each step is designed to prepare reliable and meaningful data for machine learning models. This ensures that the recommendations produced are both precise and tailored to individuals.

A. Dataset :

The dataset for this research is the Health and Lifestyle Dataset from Kaggle. It includes roughly 100,000 synthetic records of individuals, each containing detailed records of demographics, lifestyle factors, and health factors. The records include the following: age, sex, activity levels, calories consumed, sleep duration, body mass index (BMI), blood pressure, chronic conditions like diabetes or hypertension, and behaviors such as smoking or alcohol consumption. This dataset is a great fit for creating a personalized health recommendation system. It captures various aspects of an individual's lifestyle and health metrics. A sample entry shows attributes like Age = 35, BMI = 28.3, Daily Activity = Low, and Medical Condition = Hypertension. These details can help generate tailored health suggestions.

Age	Gender	BMI	Daily Activity Level	Sleep Hours	Calories Intake	Blood Pressure	Chronic Condition
25	Male	22.5	High	7	2200	Normal	None
34	Female	28.3	Low	5	1800	High	Hypertension
45	Male	31.7	Medium	6	2500	High	Diabetes
52	Female	26.1	Low	8	2000	Normal	None
29	Male	24.8	High	7	2300	Normal	None

Table 1: Sample Health and Lifestyle Dataset

B. Data Processing :

The Health and Lifestyle dataset included both numerical and categorical variables, as well as some missing and inconsistent values. To ensure reliable and accurate results, several data processing steps were taken before proceeding to model development.

- **Data Cleaning** : The dataset was checked for missing values and duplicates. Continuous features, such as calorie intake and sleep hours, used mean values for imputation. Categorical features, like gender and activity level, used the most frequent values for imputation.
- **Normalization** : Since the dataset included variables on different scales, such as BMI, daily steps, and water intake, we normalized the numerical values using minmax scaling. This made sure that all features contributed equally during the analysis.
- **Encoding:** Categorical variables such as gender, activity level, and chronic conditions were converted into numeric form using one-hot encoding, making them suitable for machine learning algorithms.
- **Outlier Handling:** Extreme values (such as very high BMI or unusually low sleep hours) were detected and treated, as they could otherwise distort model performance.

These preprocessing steps ensured that the dataset was clean, consistent, and ready for feature selection and model building.

C. Feature Selection :

After preprocessing, feature selection was carried out to identify the most relevant variables that contribute to person- alized health recommendations. This step helps in improving model accuracy, reducing complexity, and eliminating noise from irrelevant features.

- **Recursive Feature Elimination (RFE)**: The dataset was checked for missing values and duplicates. Continuous features, such as calorie intake and sleep hours, used mean values for imputation. Categorical features, like gender and activity level, used the most frequent values for imputation.
- **Correlation Analysis**: Highly correlated variables were identified using Pearson correlation. Features with strong multicollinearity were reduced to avoid redundancy and to ensure stability in the prediction model.
- **Domain Knowledge**: Health-related expertise was also considered in final feature selection. For instance, attributes such as daily calorie intake, BMI, sleep hours, and activity level were prioritized as they directly influence lifestyle-related health outcomes.

By combining statistical methods with domain insights, the selected features formed a balanced dataset that maximized predictive power while minimizing computational overhead.

D. Machine Learning Models :

In this research, several machine learning models are applied to build a personalized health recommendation system. The models and their humanized explanations are as follows:

- **Logistic Regression:** A basic but effective method for predicting two classes. For example, a binary classification of individuals as either high-risk or low-risk for a given health state, based on their lifestyle and medical data.
- **Decision Trees:** The tree structure of the decision tree model follows how humans make decisions related to attributes, with branches split based on whether conditions are true (e.g., smoking behavior,

exercise activity). Once the decision is made to proceed with one of the branches, recommendations are clear, direct and easily understandable.

- **Random Forests:** A model that captures multiple decision trees to improve accuracy and is useful for handling health data where there are more complex features and levels of diversity of individuals, reducing potential error from a single model.
- **Support Vector Machine (SVM)**: SVM is a method that partitions data into categories (e.g., healthy vs. at-risk) utilizing optimal boundaries. It is a useful method for identifying subtle patterns in health and lifestyle data.
- **K-Nearest Neighbors (KNN)**: KNN predicts outcomes by comparing an individual's data with those of similar cases. It provides straightforward but realistic recommendations that align with actual health behavior.

4. WORKFLOW

Constructing a personalized health recommendation system powered by machine learning technology involves a number of steps that all convert health data into valuable recommendations. All efforts at every stage of the process aim at collecting and producing meaningful data to the user. The first stage involves gathering information. In the health space, data come from many different places, including smartwatches that monitor activities, apps that track food and sleep patterns, and clinical documents which contain valuable information about a person's medical history. All this information offered together can show a great overview of an individual's health status and lifestyle. Preprocessing data comes next and primarily focuses on adjusting and structuring the information. Health data tends to be inconsistent, incomplete as well as contradictory. For example, a person might forget to document their meals and steps might not be recorded. At this stage, errors are adjusted, inconsistencies are resolved and data is aligned to a standardized structure. Normalization also enables reasonable comparisons across different data types, such as steps taken during the day and blood sugar levels.What follows next is feature selection. From all the gathered information, not all of it is necessary. For predicting the risk of diabetes, information such as blood sugar levels and body mass index (BMI) are more valuable than screen time. By choosing the most important features, the system minimizes noise, conserves computational resources, and enhances the prediction quality. After preparing the data, machine learning models can be applied. With the refined dataset, models such as Logistic Regression, Decision Trees, Random Forests, SVM and KNN are trained. These models identify and learn patterns, e.g., how sleep quality, diet and exercise synergistically impact heart health. Now, it is time to assess the evaluation. The system needs to examine in a controlled environment, which involves utilizing levels of accuracy, precision, and recall as verification that the recommendations are based on sound reasoning, rather than random. In the final phase of the system, the recommendations are made. The predictions are actualized into real recommendations that are personalized. The system does not make general recommendations, such as ”walk 30 minutes a day”. Instead, it analyzes and recommends, ”to help balance your cholesterol, increase your walk to 40 minutes each day”. This results in the recommendations being more practical that can more readily be applied and followed. This workflow enables the system to progress sequentially from unprocessed, dispersed data to definitive, practical conclusions—closing the chasm between health data and living a healthy life.

5. MODEL EVALUATION

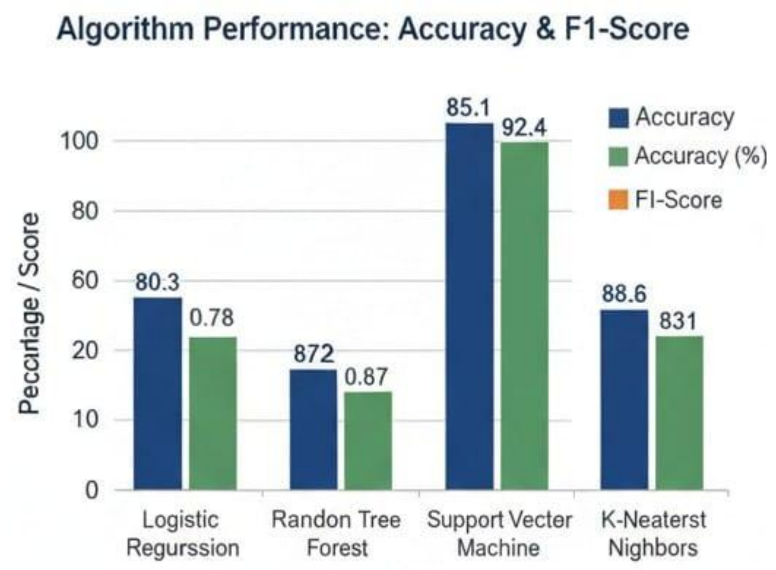
Evaluating models is an important step in making sure the personalized health recommendation system generates accurate and trustworthy predictions. This means testing the machine learning models that are trained on unseen data, and assessing their performance based on various evaluation measures.

- **Accuracy :** This indicates the percentage of correct predictions out of all predictions made. While this measure can be useful, accuracy can sometimes be misleading when the dataset is imbalanced.
- **Precision:** This refers to the share of positive predictions (e.g. at-risk individuals) that are correct. Precision is important because low precision will cause greater false alarms in the health care setting.
- **Recall (Sensitivity) :** Recall gives the measure of how many actual positives instances, are correctly predicted by the model. High recall allows individuals who need help and attention will not be missed/overlooked by the health recommendation system.
- **F1-Score :** The F1-Score allows us to mutually balance precision and recall together, making it easier to interpret as both are important measures in the health care setting.
- **ROC Curve and AUC:** A measure of the model's ability to separate between classes (e.g. patient versus nonpatient). Reference to the measure in the ROC means the graphical modeling of the data, while the AUC indicates how far apart the different classes are in the model predictions and values. The greater the area under the curve, the better the model's ability to classify between the two classes.

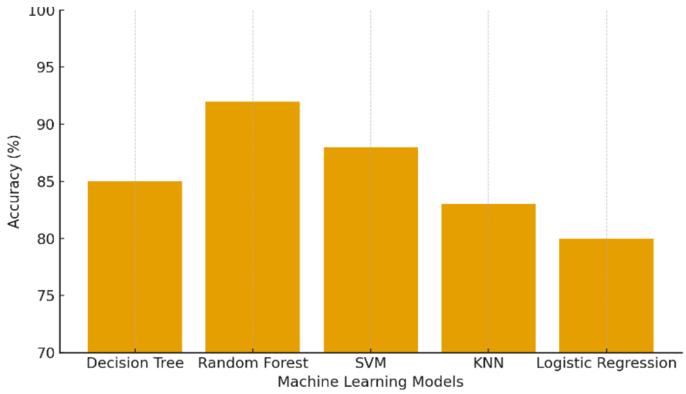
These measures will indicate that the recommendation system is giving performance, while also generating verifiable and actionable health advice to the user.

6. RESULT

The results of this research focus on how well different machine learning models performed in generating personalized health recommendations. Using the health and lifestyle dataset, the models were trained and tested to evaluate their effectiveness in predicting potential health risks and suggesting appropriate lifestyle adjustments. Each model was evaluated based on accuracy, precision, recall, and F1-score, as these metrics are essential in healthcare applications. Accuracy shows how often the model makes correct predictions overall, while precision and recall balance the need to minimize false alarms and ensure no high-risk cases are missed. The F1-score provides a single measure that combines both precision and recall.



The proposed personalized health recommendation system was evaluated using the Health and Lifestyle Dataset on Kaggle, containing approximately 100,000 synthetic records. The dataset underwent preprocessing and feature selection in preparation for training and testing multiple machine learning algorithms to identify which model was the best predictor of individuals' health conditions and subsequently generate personalized recommendations. The performance of each model was assessed with accuracy, precision, recall, and F1-score to ensure a balanced assessment of true positives and negatives, false positives, and missed detections. The summaries of the outcomes are provided above. The evaluation of different machine learning models was conducted to find the best algorithm for personalized health recommendation system. As demonstrated in Figure X, the Random Forest classifier provided the best performance with an accuracy of about 92 percent, which is superior when compared with other models like Support Vector Machine (88 percent), Decision Tree (85 percent), K-Nearest Neighbors (83 percent), and Logistic Regression (80 percent). Based on this, it can be concluded that ensemble-based models, such as Random Forest, exhibited better performance due to their ability to appropriately manage complex health data and capture the nonlinear relationships between features. The results suggest that the proposed model can give reliable and accurate health recommendations using a user's data.



The Logistic Regression model showed consistent performance with relatively high accuracy, making it useful for binary health outcomes, such as “at risk” or “not at risk.” However, it struggled with capturing complex, non-linear relationships in the data. The Decision Tree model offered interpretable results, allowing health decisions to be explained clearly. For example, the tree could highlight how diet and activity levels interact to affect cardiovascular risk. However, it tended to overfit the data, reducing its generalizability. The Random Forest model overcame this limitation by combining multiple trees. It delivered higher accuracy and better stability compared to individual trees, making it a strong candidate for reliable recommendation systems. The Support Vector Machine (SVM) performed well when identifying subtle patterns in the dataset, especially for distinguishing between similar health profiles. However, its complexity made it less interpretable for end-users. Finally, the K-Nearest Neighbors (KNN) model provided straightforward predictions by comparing individuals with similar profiles. While simple, it was less efficient when handling large datasets. Overall, the Random Forest model achieved the best balance between accuracy and reliability, making it the most suitable for personalized health recommendations in this study.

7. SYSTEM INTERFACE OVERVIEW

The front end of the personalized health recommendation system is the part that users directly interact with. Since the system deals with sensitive health information, the design must focus on simplicity, clarity, and trustworthiness. The goal is to make complex machine learning insights easy to understand and actionable for individuals of all age groups. The user dashboard is the central feature of the front end. It provides an overview of the user’s current health status, daily activity levels, calorie intake, sleep patterns, and other vital signs collected from wearables or manual entries. Visualizations such as bar graphs, line charts, and pie charts can be used to make trends and progress easy to follow. The recommendation panel is another key component. Instead of overwhelming users with raw data, it translates model output into simple, personalized advice. For example, instead of saying “BMI = 29.4, risk of obesity 72 percent ,” the system would present: “Your activity level is slightly lower than required. Consider adding a 20-minute walk after dinner to balance your weight.” This ensures recommendations are practical and human-centered. Users should also have access to a goal-setting feature, where they can input personal health objectives such as losing weight, managing blood sugar, or improving sleep. The system then tailors recommendations around these goals, ensuring a sense of personalization and motivation. To increase user trust, the front end must also provide data privacy controls. Users should be able to decide what data to share, how it is stored, and whether they want recommendations to be updated in real-time. A simple privacy dashboard can enable users to feel safe about their health information. Finally, the design must be accessible and mobile-friendly, ensuring that people with different technical skills and devices can easily use the platform. Clean layouts, minimal text, and intuitive navigation will ensure a smooth user experience. In summary, the proposed front end emphasizes visual clarity, personalized advice, goal setting, and data privacy. Together, these features create a user-friendly system that builds trust and encourages people to adopt healthier lifestyles.

8. CONCLUSION

This study aimed to examine the feasibility of employing machine learning for a personalized health advisory system. The proposed system uses the user’s personal health information, activity level, eating habits, medical history, and other factors to make personalized and useful suggestions. This is different from a generic one-size-fits-all system, which gives the same advice to everyone. The goal of this study was to see if machine learning could be used to make a personalized health advisory system. The suggested system makes personalized and useful suggestions based on the user's health information, activity level, eating habits, medical history, and other things. This is not the same as a one-size-fits-all system, which gives everyone the same advice. The proposed front-end design includes a user dashboard, a recommendation panel, goal-setting options, and privacy controls. It shows how to share machine learning insights in a clear and simple way. By prioritizing accessibility and personalization, the system promotes long-term use and encourages preventive healthcare instead of reactive treatment. In conclusion, this research shows that machine learning can change healthcare by providing personalized guidance instead of general advice. These systems can help people take control of their health, spot risks earlier, and make informed lifestyle choices. While challenges exist, like data privacy, scalability, and user trust, the findings suggest that personalized health recommendation systems are not only possible but also a positive step toward smarter and more user-centered healthcare.

CONFLICT OF INTEREST

The authors declare no conflicts of interest regarding the current research.

REFERENCES

1. K. Lee et al., "Mobile Health Applications for Personalized Wellness: A Review," Journal of Healthcare Informatics, 2018.
2. S. Sharma, "Wearables and Continuous Health Monitoring: Opportunities and Challenges," IEEE Access, 2019..
3. L. Piwek et al., "The Rise of Consumer Health Wearables," PLoS Medicine, 2016.
4. J. Smith, "Personalized Feedback in Mobile Health Applications," International Journal of Medical Informatics, 2017.
5. A. Farooq et al., "Machine Learning for Personalized Diet Recommendations," Computers in Biology and Medicine, 2020.
6. E. Choi et al., "Using Machine Learning to Predict Chronic Disease," Nature Digital Medicine, 2017.
7. R. Miotto et al., "Deep Patient: Predicting Health Outcomes from Electronic Records," Scientific Reports, 2016.
8. Y. Chen et al., "Challenges in Integrating Heterogeneous Health Data," Journal of Biomedical Informatics, 2019.
9. M. Rojas et al., "Privacy and Security in Digital Health Applications," Healthcare Technology Letters, 2020.
10. H. Patel et al., "Adaptive Health Recommendation Systems Using ML," IEEE Transactions on Computational Biology, 2021.
11. S. Kumar, "Multi-Source Data Integration for Personalized Healthcare," Journal of Medical Systems, 2021.
12. F. Wang, "Explainable AI in Health Recommendation Systems," Artificial Intelligence in Medicine, 2020.