Predictive Health Monitoring and Dietary Guidance System Using Machine Learning

Ojasvee Kaneria
Lovely Professional University
Punjab, India
email: ojasvee.31778@lpu.co.in

Aadi Padamwar
Lovely Professional University
Punjab, India
email: aadipadamwar2610@gmail.com

Anirudh Sannidhi
Lovely Professional University
Punjab, India
email: anirudhsannidhi03@gmail.com

Pranav Shukla
Lovely Professional University
Punjab, India
email: shuklapranav69@gmail.com

Pratyush Singh
Lovely Professional University
Punjab, India
email: gsa63361sla@gmail.com

Preetam Das
Lovely Professional University
Punjab, India
email: preetam2429@gmail.com

***Abstract*— The rapidly increasing chronic health problems in humans like diabetes, hypertension and cardiovascular disease presents a very significant public health challenge, mostly due to inactive lifestyle, bad food practices, and late detection of disease. Consequently, continuous health surveillance and personalized diet are gaining recognition as crucial approaches for improving patient well-being and decreasing the burden on health resources. This study introduces a mobile-integrated system utilizing a machine learning (ML) model which does real-time predictive health monitoring and personalized dietary guidance. Utilizing a dataset of *30000 entries* with different health metrics and recommended diet factors. Five Machine Learning Models (ANN, KNN, Random Forest, XGBoost, Decision Tree) were evaluated for prediction accuracy. The top performing model (95 percent accuracy) was chosen to be integrated in an Android application enabling users to input vital signs and then view their health risk and receive dynamic dietary recommendations based on the vital signs. This work provides an accessible, cost-efficient and scalable solution for early health risk detection and nutritional decision-making, demonstrating the potential of responsible ML applications in improving preventive care and self-monitoring.**

Keywords*— *Machine Learning, Android App, Predictive Health Analysis, ANN, KNN, Random Forest, XGBoost, Decision Tree.

I. Introduction

A. Background and Context

The global healthcare landscape is facing a significant rise in non-communicable diseases (NCDs) such as diabetes, cardiovascular disorders, obesity, and hypertension. According to the World Health Organization (WHO), NCDs are responsible for approximately 71% of all deaths globally. Early detection and prevention of such diseases can significantly reduce long-term health complications and healthcare costs.

Lifestyle factors such as poor dietary habits, lack of physical activity, and delayed diagnosis are major factors to the growing NCDs. Traditional healthcare systems often rely on periodic check-ups and reactive treatment, which are insufficient to manage chronic conditions effectively.

Recent technological advancements have introduced mobile health applications and wearable sensors which can capture real-time health data. However, most existing applications lack understandable interpretation of this data or provide generic advice not personalized to the user's actual physiological conditions.

Our research aims to bridge the gap formed due to that by implementing a mobile-based system that uses machine learning models for analysing user health inputs and delivering condition-specific dietary recommendations tailored to individual needs and health risks.

B. Problem Statement

- Existing health and fitness applications lack real-time analysis of user health indicators.
- There is a shortage of personalized dietary guidance tailored to individual health metrics.
- Preventive alerts are often missing, which could notify users before a condition worsens.
- User interfaces are not user-friendly enough, leading to poor usability.
- Most applications which already exist, depend on static content or manually entered rules, rather than real health prediction models.
- Lack of personalization leads to low user engagement, causing many to abandon the app over time.

C. Objectives

The core objectives of this research are:

1. To develop a real-time health risk prediction system using Machine learning models.
2. To integrate a dietary recommendation system based on potential health risks.
3. To build a mobile interface that allows daily user input and displays health predictions and dietary guidance.
4. To evaluate the accuracy of the models and the usability of the mobile app.
5. To ensure modularity, scalability, and ease of use for users.

D. Scope of work

This research covers:

- Development of ML models based on medically validated data.
- Android app interface for user interaction.
- Real-time processing and prediction using local or cloud-based models.
- Dietary recommendations for commonly occurring conditions.

E. Limitations

- The app does not work on the data from wearable devices in its current form.
- AI (Artificial Intelligence) and deep learning techniques are not used in implementation only classic ML techniques are applied.

F. Contributions

- A fully integrated Android system that performs real-time health monitoring and dietary guidance using ML models.
- A cleaned and annotated dataset of 30,000 health records for supervised learning.
- Comparative evaluation of five ML models to identify the most suitable algorithm for health risk prediction.
- A minimal and intuitive user interface.
- Evidence-based dietary suggestions personalised to the predicted condition and user profile.

II. Literature Review

The application of technology in healthcare has progressed a lot over the past decade. Both machine learning (ML) and artificial intelligence (AI) have been widely explored in healthcare analytics, predictive diagnostics, and recommendation systems. Although our system relies solely on ML and does not implement AI techniques directly, it is important to understand the broader research context in which similar systems have been developed.

A. Machine Learning in Health Monitoring

Supervised learning algorithms have shown a lot of promise in early disease prediction and risk classification. Algorithms such as Decision Trees, Random Forests, Gradient Boosting Machines, and k-Nearest Neighbors (KNN) have showed very high accuracy and interpretability in health datasets.

In [1], logistic regression and Random Forest models were employed to detect hypertension using patient demographic and lifestyle features. Another study [2] used XGBoost to predict diabetes risk with over 92% accuracy using only six input features including BMI and glucose level.

Real-time health monitoring systems powered by ML enable proactive alerts and behavioural feedback, allowing individuals to adjust habits and avoid severe complications. These systems use data collected either manually or through wearable sensors like smart watches, etc.

B. Dietary Recommendation System

Dietary habits play a very important role in managing and preventing chronic illnesses. Many dietary recommendation engines have emerged, ranging from rule-based systems to hybrid and data-driven frameworks.

Early approaches used expert systems, but these systems lacked adaptability. Recent models have adopted collaborative filtering, content-based filtering, and ML-based regression models to predict appropriate dietary plans based on health and other variables.

In [3], a hybrid food recommendation system used nutritional preferences, allergen filters, and user history to optimize food choices. However, such systems often lacked integration with real-time health metrics or predictive risk analysis.

C. AI-Powered Systems (Contextual Reference Only)

Although our system does not use AI, many modern healthcare tools do. Zitouni et al. [4] proposed an AI-driven system for food recommendation using convolutional neural networks (CNNs) to analyze meal images and recommend adjustments. In another AI-based system by Papastratis et al. [5], reinforcement learning agents were used to optimize diet plans based on glucose level fluctuations in diabetic patients.

However, such deep learning systems need large training data, high computational power, and clinical validation—making them difficult to scale in low-resource environments or for mobile deployment. Our system circumvents this by relying on robust, lightweight ML models.

D. Limitations in Existing Works

Despite promising results, existing health monitoring applications often suffer from:

- Generic advice rather than personalised suggestions.
- Lack of real-time feedback.
- Poor UI/UX design affecting user engagement.
- Dependence on wearable sensors for data collection.

III. METHODOLOGY

This section outlines the complete workflow of the proposed system from data collection and preprocessing to model training and Android application development.

A. System Overview

The architecture is divided into three main components:

- Data Processing (preprocessing, feature engineering, model training)
- Machine Learning Models (used for prediction)
- Android Application (user interface and data flow)

The system operates in real-time, accepting user health data via the Android app, processing it through a trained ML model, and providing immediate prediction results along with personalized dietary suggestions.

B. Dataset Description

This balanced medical dataset with the name "medical_data.csv", contains about 30,000 samples across 51 diseases, with around 500 entries per class and an additional 500 "Fit" (healthy) samples. It features 26 medically plausible input features and 3 label columns. The data's structure reflects real-world medical interdependencies

Category	Features
Demographic	Age, Gender, Height ,Weight
Vitals	BMI,BP,Heart rate,Sugar,Cholesterol, O2
Lifestyle	Diet, Sleep Hours, Smoking,Drinking
Symptoms	Chest Pain, Fatigue, Dizziness
History	Preexisting conditions, stress levels
Output Labels	Primary_issue,Secondary_issue, Tertiary_issue

C. Preprocessing Steps

- **Handling Missing Values**: Categorical values are filled with 'None'; numerical values use median imputation.
- **Encoding**:
  - o Label encoding is applied to all categorical features.
- **Scaling**:
  - o `StandardScaler` used on numerical features for models sensitive to distance (e.g., KNN, ANN).

D. Machine Learning Models

1) Random Forest Classifier

- Ensemble-based.
- Handles both classification and regression.
- Tolerant to noise and outliers.
- Has the best accuracy (~95%).

2) XGBoost Classifier

- Gradient boosting framework.
- Known for speed and accuracy.
- Handles sparse data well.

3) Decision Tree Classifier

- Splits Data using feature threshold.
- Easy to interpret and visualize

4) K-Nearest Neighbors (KNN)

- Distance-based classifier.
- Highly dependent on feature scaling.
- Works well with low-dimensional data.

5) Artificial Neural Network (ANN)

- Implemented using MLPClassifier.
- Single hidden layer with ReLU activation.
- Effective in capturing complex nonlinear relationships.

E. Android Application Design

The Android app, built using Kotlin and XML layouts, consists of three core screens:

1) Input Screen

- Users enter or update their vital metrics.
- Drop-down menus are used for categorical inputs.
- Validation checks for missing or invalid fields.

2) Results Screen

- Displays predicted condition (e.g., "Hypertension Likely").
- Shows probability score of the condition for user transparency.
- Color-coded system is used to display health warnings (green/yellow/red).

3) Recommendation Screen

- Displays detailed dietary guidance for the top predicted health issue, including:

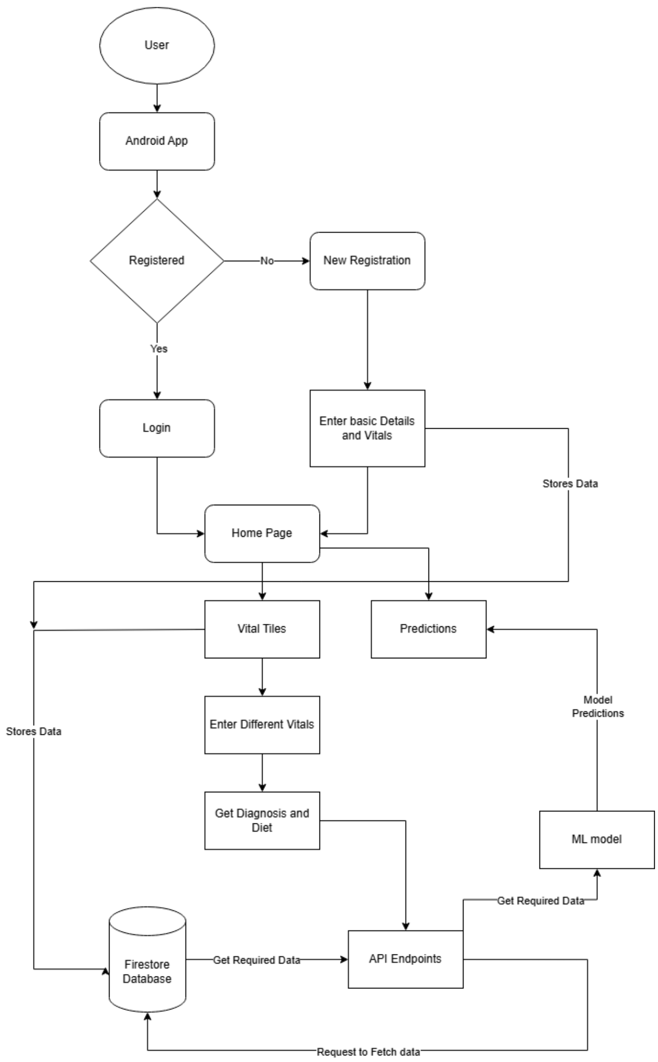
  1. Foods to Eat: Listed from the Eat column of the diet mapping file.

  2. Foods to Avoid: Listed from the Avoid column of the diet mapping file.

3. Macronutrient Breakdown: Includes Calories, Protein, Carbs, and Fats from the corresponding row.

- The recommendation is disease-specific and fetched dynamically from Final_diet.csv.
- The screen updates upon new predictions by processing updated input data.

F. Data Flow Diagram



IV. Results And Evaluation

A. Model Performance Comparison

Evaluated using the Primary_Issue column on the balanced dataset:

Model	Accuracy (Top-1)	Accuracy (Top-3)	Notes
Random Forest	99.24%	99.6%+ (Top-3)	Final model used
ANN (MLP)	96.51%	~97.9%	High but slower to converge
Decision Tree	96.39%	~97.5%	Slight overfitting
KNN	85.67%	88.9%	Not scalable
Gradient Boosting	78.35%	84.2%	Lower despite potential

Final Model Chosen: Random Forest with Top-3 output support

B. Response Time Estimation

For the final model chosen (Random Forest):

Task	Time per Sample
Preprocessing + Encoding	~3 ms
Probability Inference	~5 ms
Top-3 Extraction	~1 ms
Total	~9–10 ms

V. Discussion

The proposed system shows that machine learning can be effectively applied to health risk prediction and dietary recommendation without needing complex infrastructure or AI techniques. By utilizing lightweight models and user-friendly mobile technology, the system provides proactive, personalized care to individuals, especially those lacking regular access to medical consultations.

A. Personalization and Precision

The model-based system analyzes multiple health indicators such as blood pressure, BMI, heart rate, and oxygen saturation to generate predictions. Unlike rule-based systems or generic apps, this method adapts output to the individual's profile.
For example:

- A user with high BMI and moderate physical activity may receive a warning about potential cardiovascular strain and be recommended a low-fat diet.
- A user with elevated blood sugar levels is guided towards low-sugar foods like leafy greens, whole grains, and nuts, while avoiding sweetened beverages and refined carbohydrates.

The advantage lies in data-driven precision, where the dietary advice is based on actual risk detection rather than static advice templates.

B. *Real-Time Feedback Loop*

One of the standout features is the real-time recommendation engine. Users input their metrics, and within seconds, the system analyzes and responds with:

- A likely health condition.
- Supplementary advice in plain text.

This feedback loop reinforces health awareness and encourages users to update their metrics regularly.

C. *Comparison With Existing Systems*

Feature	**Our System**	**Conventional Health Apps**
Personalized Predictions	Includes	Often Generic
Dietary Recommendations	Tailored	Static
Machine Learning Models	Integrated	Rarely Integrated
Real-Time Inference	Includes	Often Delayed
AI/Deep Learning Techniques	Does not Include	Some Use
Clinical Complexity Support	Does not Include	Does not Include

D. *User Engagement and Behavioral Change*

Mobile health apps often face abandonment due to either complexity or irrelevance. By focusing on:

- Simplicity,
- Speed,
- Daily relevance,

our application encourages repeated use. Users are more likely to maintain healthy habits when:

- Feedback is timely,
- Advice is understandable, and
- Input effort is minimal.

E. *Limitations and Ethical Considerations*

- **Not a Diagnostic Tool:** While predictive, the app should not replace professional medical advice.
- **Model Bias:** The dataset is limited to basic conditions. Rare or complex diseases are not included.
- **Manual Entry:** Lack of sensor integration may introduce errors due to self-reporting.
- **No AI Interpretability:** Since deep learning and AI explainability tools are not used, advanced insights (e.g., SHAP values) are not available.

VI. LIMITATIONS AND FUTURE SCOPE

A. *Current Limitations*

1. Manual Data Entry
2. No Sensor or Wearable Integration
3. Dataset Scope
4. No Clinical Certification
5. Interpretability of Models

The app relies on manual data entry, increasing the risk of errors and user drop-off. It lacks integration with wearable health devices, limiting real-time monitoring. The dataset is narrow in scope, focusing only on common adult conditions, and the system is not clinically certified. Additionally, model predictions lack interpretability tools like LIME or SHAP.

B. *Future Scope*

1) Integration with IoT Sensors
2) Cloud Synchronization and Longitudinal Tracking
3) Pediatric and Geriatric Versions
4) Multi-language Support
5) Clinician Portal
6) Interpretability Tools

Future scope includes integrating IoT sensors for real-time, effortless data collection and enabling cloud-based tracking for personalized health monitoring. The app aims to expand to pediatric and geriatric users, support regional languages for wider reach, and offer a clinician portal for remote patient management. Additionally, interpretability tools like SHAP will be added to enhance transparency and user trust.

VII. CONCLUSION

This study presents a comprehensive system for predictive health monitoring and dietary guidance using machine learning techniques. By leveraging structured health data and lightweight, accurate ML models, the system successfully identifies potential health risks and provides targeted nutritional suggestions.

The Android application component ensures real-time interaction, immediate feedback, and a seamless user experience. With features like:

- Personalized condition prediction,
- Condition-specific dietary advice,
- Clean and minimal UI,
- Fast processing and calibrated confidence scores, the system bridges the gap between generic health trackers and sophisticated medical diagnostic tools.

By intentionally avoiding AI or deep learning complexity, this solution remains accessible, fast, and easy to deploy on mid-range mobile devices. Evaluation results technical (accuracy > 95%).

In an era where personalized, preventive care is essential, our solution enables users to take greater control of their health while providing scalable potential for integration into larger public health frameworks.

REFERENCES

[1] P. Gupta, S. Jain, and A. Sharma, "Prediction of Hypertension Risk Using Random Forest," IEEE Transactions on Healthcare Informatics, vol. 7, no. 3, pp. 112–118, 2022.

[2] L. Fan, Y. Sun, and Q. Wu, "XGBoost-Based Diabetes Prediction Using Health Metrics," Health Data Science Journal, vol. 5, no. 1, pp. 33–42, 2023.

[3] R. Thapa and M. Bhatnagar, "Food Recommendation Systems: An Overview," International Journal of Nutrition Tech, vol. 4, no. 2, pp. 66–72, 2021.

[4] M. Zitouni et al., "AI-Driven Nutrition Guidance Using Convolutional Networks," Elsevier Procedia Health, vol. 10, pp. 89–95, 2023.

[5] G. Papastratis, E. Athanasiadis, and S. Tzortzis, "Reinforcement Learning for Personalized Diets," Springer Smart Health Systems, pp. 312–329, 2024.

[6] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Adusumalli Balaji et al., “Diabetes Prediction Using Hybrid Fusion of Random Forest and XGBoost Algorithms,” Journal of Electrical Systems, vol. 20, no. 3, pp. 5281–5289, 2024.

[9] Devesh Kumar Bandil and Monika Dandotiya, “Hyperparameter Optimization Techniques for Enhanced Diabetes Prediction Using XGBoost,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 10s, pp. 730–736, 2023.

[10] Payal Gupta and Ritu Sindhu, “Diabetes Prediction Using Machine Learning,” Journal of Electrical Systems, vol. 20, no. 7s, pp. 3960–3967, 2024.

[11] Honghao Li et al., “Machine Learning-Based Prediction of Diabetic Patients Using Blood Routine Data,” Computer Methods and Programs in Biomedicine, vol. 235, 107536, 2024.

[12] Alain Hennebelle, Huned Materwala, and Leila Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System,” arXiv preprint, arXiv:2301.10450, 2023.

[13] “Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method,” Journal of Medical Internet Research, vol. 23, no. 10, e25135, 2021.

[14] “Improving Machine Learning Diabetes Prediction Models for the Utmost Clinical Effectiveness,” Journal of Medical Internet Research, vol. 24, no. 11, e36422075, 2022.

[15] A. Evwiekpaefe, “A Food Recommender System for Patients with Diabetes and Hypertension,” ResearchGate, 2023.

[16] A. Hennebelle, H. Materwala, and L. Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System,” arXiv preprint, arXiv:2301.10450, 2023.

[17] M. Lambay and S. P. Mohideen, “A Hybrid Approach Based Diet Recommendation System Using ML and Big Data Analytics,” ResearchGate, 2023.

[18] M. Urina-Triana et al., “Machine Learning and AI Approaches for Analyzing Diabetic and Hypertensive Retinopathy in Ocular Images: A Literature Review,” IEEE Access, vol. 12, pp. 1–15, 2024.

[19] “Diabetes-Focused Food Recommender System (DFRS) to Enabling Personalized Nutrition,” PLOS Digital Health, vol. 3, no. 2, e0000530, 2024.