Title: Intelligent Cholesterol-Based Diet and Exercise Recommendation System Using Machine Learning and OCR

**Abstract:**

Maintaining a healthy lifestyle is essential for preventing cardiovascular diseases, obesity, and other metabolic disorders. Traditional diet plans often fail due to their generic nature and lack of real-time adaptability. This project proposes an Intelligent Diet and Exercise Recommendation System that utilizes OCR, NLP, and Machine Learning to provide personalized dietary recommendations based on cholesterol levels extracted from medical reports.

Additionally, the system integrates real-time wearable data (step count, heart rate, activity levels) to further refine recommendations. The approach involves automated cholesterol extraction using OCR/NLP techniques, diet classification using ML models, and exercise recommendation via decision trees and KNN. An interactive Spring Boot-based web app with visual dashboards provides users with real-time health insights. The system ensures continuous learning by allowing users to provide feedback, refining recommendations over time.

By leveraging real-time data fusion, predictive analytics, and AI-driven insights, this project bridges the gap between static diet plans and personalized, dynamic health recommendations.

**Introduction:**

In today’s fast-paced lifestyle, maintaining a healthy diet and regular exercise routine is essential yet challenging. With the rise in health issues such as obesity, diabetes, and cardiovascular diseases, people are increasingly looking for personalized fitness guidance. However, generic diet plans and workout routines often fail to meet individual needs.

To address this, we propose an **Advanced Diet Recommendation System** that leverages **machine learning and data extraction techniques** to provide personalized diet plans based on an individual's health parameters.

Our system focuses on extracting critical health metrics, particularly **cholesterol and its components (Total Cholesterol, HDL, LDL, and Triglycerides)**, from medical reports in **PDF format**. By analysing these values along with user inputs such as **age and gender**, the system can provide **customized dietary recommendations** that align with an individual’s cardiovascular health and nutritional needs.

The system integrates **PDF data extraction techniques (using PyMuPDF and OCR for scanned reports)** to efficiently retrieve health metrics from medical reports. These extracted values are then processed through a **machine learning model**, which classifies users into appropriate dietary categories (e.g., **Low Fat Diet, Mediterranean Diet, High Protein Diet**) based on medical guidelines and best practices for heart health.

**Key Features of the System:**

* **Automated Extraction** – Extracts cholesterol-related data from medical reports using PDF processing techniques.
* **Personalized Diet Plans** – Suggests a diet tailored to the user’s cholesterol levels, triglycerides, and overall lipid profile.
* **Machine Learning Model** – Analyses health data to predict the most suitable diet based on scientific research.
* **User-Friendly Interface** – Built using **Spring Boot for the UI**, ensuring a seamless experience for users.

With this system, individuals can gain **valuable insights into their cardiovascular health** and receive **evidence-based dietary recommendations**, ultimately helping them manage their health more effectively.

**Literature Review:**

High levels of cholesterol are generally considered to be associated with atherosclerosis. In the past two decades, however, a number of studies have shown that excess cholesterol accumulation in various tissues and organs plays a critical role in the pathogenesis of multiple diseases. Here, we summarize the effects of excess cholesterol on disease pathogenesis, including liver diseases, diabetes, chronic kidney disease, Alzheimer’s disease, osteoporosis, osteoarthritis, pituitary-thyroid axis dysfunction, immune disorders, and COVID-19, while proposing that excess cholesterol-induced toxicity is ubiquitous. We believe this concept will help broaden the appreciation of the toxic effect of excess cholesterol, and thus potentially expand the therapeutic use of cholesterol-lowering medications.[1]

[Genetic](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/genetics), observational, and clinical intervention studies indicate that circulating levels of [triglycerides](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/triglyceride) and cholesterol transported in triglyceride-rich [lipoproteins](https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/lipoprotein) (remnant cholesterol) can predict cardiovascular events[2]

The study aimed to determine effects of a ketogenic diet on metabolic dysfunction, testicular antioxidant capacity, apoptosis, inflammation, and spermatogenesis in a high-fat and high-cholesterol diet-induced obese mice model. Forty-two male C57BL/6 mice were fed either a normal diet (NC group) or a high-fat and high-cholesterol (HFC) diet (HFC group) for 16 weeks, and mice from the HFC group were later randomly divided into two groups: the first were maintained on the original HFC diet, and the second were fed a medium-chain triacylglycerol (MCT)-based ketogenic diet for 8 weeks (KD group). A poor semen quality was observed in the HFC group, but this was eliminated by the ketogenic diet. Both the HFC and KD groups exhibited enhanced apoptosis protein expressions in testis tissue, including caspase 3 and cleaved PARP, and higher inflammation protein expressions, including TNF-α and NF-κB. However, the KD group exhibited a statistically-significant reduction in lipid peroxidation and an increased glutathione peroxidase level as compared with the HFC group. The HFC diet induced obesity in mice, which developed body weight gain, abnormal relative organ weights, metabolic dysfunction, and liver injury. Overall, the results showed that a ketogenic diet attenuated oxidative stress and improved the semen quality reduced by the HFC diet.[3]

Both hypercholesterolemia and aging are related to cognitive decline or Alzheimer’s disease. However, their interactive influence on the neurodegenerative progress remains unclear. To address this issue, 6-month-old and 16-month-old female mice were fed a 3% cholesterol diet for 8 weeks, followed by hippocampus-related functional, pathological, biochemical and molecular analyses. The high cholesterol diet did not exacerbate age-dependent cognitive decline and hippocampal neuronal death, and even greatly mitigated decreases of synaptophysin and growth associated protein 43 expression in the hippocampus of aged mice. Compared with young controls, aged mice fed normal diet showed mild activation of hippocampal microglia with increased expression of CD68, a marker of the microglial M1 phenotype, and decreased expression of CD206, a marker of the microglial M2 phenotype. More interestingly, the high cholesterol diet not only improved NLRP3 inflammasome activation and IL-1β expression, but also increased levels of anti-inflammatory cytokines IL-4 and IL-6 in the hippocampus of old mice, suggesting playing pro- and anti-neuroinflammatory effects. In addition, the cholesterol rich diet resulted in a defect of the blood-brain barrier of aged hippocampus, as revealed by increased brain albumin content. These results have revealed both harmful and protective effects of high cholesterol diet on aged brain, which helps us to understand that hypercholesterolemia in the aged population is not associated with dementia and cognitive impairment.[4]

Changes from the traditional diet to the western diet such as high-calorie fast food and fat, thus impacting the high prevalence of cholesterol in the community. One effort to reduce blood cholesterol can be done through consultation or counselling with a nutritionist. The study aimed to measure the effect of nutritional counselling on the decrease of cholesterol. This quasi-experimental study has used a sample of 48 patients of health centre in the Banda Aceh City, divided into case and control groups through non-random assignment. Cholesterol data obtained through laboratory blood tests, and the tool used for its testing is Autocheck Â with scale ratio, while the variable method of counselling is done face to face with a tool leaflet. Analysis using a t-test. The results showed a decrease in blood cholesterol in patients given nutritional counselling using a leaflet medium of 20,2 mg/dl (p=0,000), while in the control group (counselling without media) also showed a decrease in cholesterol levels of 6,9 mg/dl (p=0.001). Counselling methods using leaflets have a better value of effectiveness than without the media. Conclusion, the provision of nutrition counselling is very influential on the decline in blood cholesterol, and the use of media leaflets better in nutrition counselling. Suggestion, every health services should cooperate with universities to conduct nutritional counselling routinely and become one of program priority in the handling of degenerative diseases.[5]

Aging is a process that occurs continuously and causes several changes in the body that will affect functions that can increase cholesterol levels in the body. This research aims to determine whether there is a relationship between diet and cholesterol levels in the elderly in Tenggela Village, Tilango District. The research design used an analytic survey method with a cross-sectional approach. The sampling technique used was purposive sampling. The sample was the elderly who lived in the village of Tenggela. Methods of data collection used a questionnaire. The results of the Chi-Square analysis test obtained a p value of 0.000 which is smaller than a value of 0.05, which means that H0 is rejected and Ha is accepted or there is a relationship between diet and the incidence of cholesterol in the elderly in Tenggela Village, Tilango District. Suggestions for the village government It is necessary to improve education in the form of socialization for the elderly or families about how to prevent hypercholesterolemia in the elderly [6]

Documents in PDF format are nowadays called the Universal document format. PDF to speech converter systems involves many steps to achieve. Text extraction is the primary step From PDF to do further processing. In this paper we start with the brief discussion about the steps involved in extracting the text from PDF documents. The aim of this paper is to give the introduction with some basic concepts on PDF, and with text extraction concepts, which will be useful for the readers who are less familiar in this area of research.[7]

In conclusion, we are developing an intelligent system that analyses users’ cholesterol reports and provides personalized dietary recommendations to promote a healthier lifestyle. By leveraging machine learning, our system will extract key health indicators such as cholesterol levels, triglycerides, HDL, and LDL cholesterol from medical reports, along with additional factors like age and gender. This data-driven approach ensures that users receive tailored diet plans designed to improve their cardiovascular health and overall well-being.

Our platform integrates seamlessly with a user-friendly Spring Boot-based interface, allowing individuals to upload their reports effortlessly and receive real-time insights. Additionally, by incorporating a step count feature, our system can further refine recommendations based on physical activity levels, offering a holistic approach to health management.

By automating the analysis process, our system eliminates the need for manual interpretation of complex medical reports, making healthcare more accessible and actionable. In the long run, this solution has the potential to help users make informed dietary choices, reduce the risk of heart disease, and encourage a healthier lifestyle. Through continuous improvements and possible integration with wearable health devices, our system aims to become an essential tool for proactive health monitoring and management.

|  |  |  |  |
| --- | --- | --- | --- |
| Cholesterol Type | Low Heart Disease Risk | Borderline Heart Disease Risk | High Heart Disease Risk |
| Total Cholesterol | Less than 200 | 200 - 239 | 240 and higher |
| LDL Cholesterol | Less than 130 | 130 - 159 | 160 and higher |
| HDL Cholesterol | 60 and higher | 50 - 59 | Less than 50 |
| Triglycerides | Less than 150 | 150 - 199 | 200 and higher |

Measurements in **mg/dL**.

**Proposed Methodology:**

**1. Data Extraction from Medical Reports (OCR & NLP)**

* Text-based PDFs: Extracted using PyMuPDF (Fitz) or PDFMiner
* Scanned PDFs: Processed using Tesseract OCR / EasyOCR
* Text Cleaning & NLP:
  + Named Entity Recognition (NER) for medical terms using spaCy/BERT
  + Regular Expressions (RegEx) to extract cholesterol levels

**2. User Profile & Additional Inputs**

* Manual Inputs: Age, gender, weight, BMI, dietary preferences (vegetarian/non-vegetarian)
* Wearable Integration *(Optional)*: Step count, heart rate, exercise history from Google Fit/Apple Health/Fitbit API

**3. Machine Learning for Personalized Diet Recommendation**

* Feature Engineering:
  + Health Metrics: Total Cholesterol, HDL, LDL, Triglycerides
  + User Data: Age, gender, BMI, activity level
  + Wearable Data: Steps per day, heart rate trends
* ML Algorithms Used for Diet Classification:
  + Random Forest / Decision Trees → If data is limited
  + Neural Networks (Deep Learning) → If large datasets are available
  + Collaborative & Content-Based Filtering → Personalizes diet over time
* Output: Categorizes users into Low Fat, Mediterranean, High Protein, or Custom Diet Plans

**4. Exercise Recommendation System**

* K-Nearest Neighbors (KNN): Suggests workout intensity based on cholesterol levels and step count
* Decision Trees: Recommends exercises (e.g., brisk walking, yoga, cardio) based on heart health

**5. User Interface & Web App Development**

* Backend: Spring Boot (REST APIs to process ML recommendations)
* Frontend: React.js (Interactive dashboards, progress tracking, charts)
* Visualizations:
  + D3.js / Chart.js for cholesterol trends & diet effectiveness
  + Wearable Data Graphs (Steps, Heart Rate)

**6. Continuous Learning & Feedback System**

* User Feedback Mechanism: Users can rate diet recommendations, refining ML models over time
* Adaptive Model: Improves accuracy with new user data & dietary outcomes
* **Key Objectives of the Web Application:**

1. **User-Friendly Report Upload** – Provide an easy-to-use interface for users to **upload their cholesterol reports (PDF format)** for analysis.
2. **Seamless Integration with ML Model** – Ensure smooth backend processing where the **ML model extracts cholesterol data and predicts a suitable diet**.
3. **Real-Time Dietary Recommendations** – Display **personalized diet suggestions** instantly after processing the report.
4. **Interactive Dashboard for Users** – Show extracted health data (Total Cholesterol, HDL, LDL, Triglycerides) with visualizations (graphs, charts).
5. **Secure Data Handling** – Ensure **data privacy** and security while processing sensitive medical information.
6. **Scalability & Performance Optimization** – Optimize API calls and database interactions for **fast processing and real-time responses**.
7. **Mobile Responsiveness** – Design a responsive UI that works across **desktop and mobile devices** for accessibility.
8. **User Authentication & History Tracking (Optional)** – Allow users to create accounts, **save past reports**, and track dietary recommendations over time.

**Tools & Technologies Used:**

| **Component** | **Technology Used** |
| --- | --- |
| **Backend** | Python |
| **Frontend** | HTML/CSS, JavaScript |
| **Database** | MongoDB |
| **OCR & NLP** | PyMuPDF, Tesseract OCR, spaCy, BERT |
| **Machine Learning** | Random Forest, SVM, Neural Networks |

**Certified Dietary Guidelines Based on BMI, Gender, Medical Conditions, and Age:**

| **Category** | **Sub-Category** | **Guidelines** |
| --- | --- | --- |
| **BMI-Based Categorization** | Underweight (BMI < 18.5) | - Increase calorie intake with nutrient-dense foods (healthy fats, protein, complex carbs).  - Include protein-rich foods like lean meats, eggs, legumes.  - Consume whole grains, nuts, and dairy products[10]˒[11]. |
|  | Normal Weight (BMI 18.5–24.9) | - Maintain a balanced diet with all essential nutrients.  - Focus on whole foods, fruits, vegetables, lean proteins, and whole grains.  - Moderate sugar, salt, and unhealthy fats intake[12]. |
|  | Overweight (BMI 25–29.9) | - Reduce portion sizes and calorie-dense foods.  - Increase fiber intake with vegetables, fruits, and whole grains.  - Engage in regular physical activity[10]˒[13]. |
|  | Obesity (BMI ≥ 30) | - Follow a calorie-deficit diet focusing on low-carb, high-fiber foods.  - Limit sugar, trans fats, and processed foods.  - Include lean proteins and healthy fats.  - Regular exercise is essential[10]˒[11]. |
| **Gender-Based Recommendations** | Male | - Higher protein intake to maintain muscle mass.  - More calories due to higher metabolism.  - Increase omega-3 fatty acids for heart health.  - Adequate intake of zinc, magnesium, and vitamin D[12]˒[14]. |
|  | Female | - Higher iron intake (especially pre-menopause).  - Increased calcium and vitamin D for bone health.  - Balanced carb and protein intake to regulate hormones.  - Folic acid for reproductive health[10]˒[15]. |
| **Medical Condition-Based Guidelines** | Hypertension (High Blood Pressure) | - Follow the **DASH diet** (low sodium, high potassium and magnesium).  - Reduce salt intake (< 2,300 mg/day).  - Increase whole grains, fruits, and vegetables.  - Avoid processed foods and excess caffeine[14]˒[16]. |
|  | Diabetes | - Carbohydrate counting and glycemic index control.  - Increase fiber intake (whole grains, vegetables, legumes).  - Avoid refined sugars and processed carbs.  - Regular meal timing for blood sugar stability[11]˒[17]. |
|  | Hyperlipidemia (High Cholesterol) | - Reduce saturated and trans fats.  - Increase omega-3 fatty acids (fish, flaxseeds, walnuts).  - Eat more soluble fiber (oats, legumes, fruits).  - Limit red meat and high-fat dairy products[16]˒[18]. |
|  | PCOS (Polycystic Ovary Syndrome) | - Low glycemic index diet with whole foods.  - Increase protein and healthy fats for hormone balance.  - Reduce processed carbs and sugary foods.  - Exercise regularly for insulin sensitivity[15]˒[19]. |
|  | Osteoporosis | - High calcium (dairy, leafy greens, fortified foods).  - Sufficient vitamin D (sunlight, fatty fish, supplements).  - Weight-bearing exercises to strengthen bones.  - Limit caffeine and alcohol intake[20]. |
|  | Kidney Disease | - Reduce sodium, phosphorus, and potassium intake.  - Moderate protein intake (plant-based sources preferred).  - Stay hydrated but avoid excessive fluid intake if advised.  - Avoid processed and high-phosphorus foods[15]˒[21]. |
| **Age-Based Dietary Recommendations** | Children (2–12 years) | - Balanced diet rich in fruits, vegetables, whole grains, and lean protein.  - Limit sugary snacks and processed foods.  - Encourage hydration and active play.  - Adequate calcium intake for bone development[12]˒[23]. |
|  | Adolescents (13–18 years) | - Increased protein and calcium for muscle and bone growth.  - Limit fast food and high-sugar drinks.  - Ensure sufficient iron intake (especially for girls).  - Encourage healthy eating habits and physical activity[14]˒[23]. |
|  | Adults (19–50 years) | - Maintain a well-balanced diet with lean proteins, whole grains, and healthy fats.  - Moderate alcohol intake and avoid processed foods.  - Regular physical activity to prevent weight gain and disease.  - Focus on portion control and mindful eating[12]˒[15]. |
|  | Seniors (50+ years) | - Increase protein intake to prevent muscle loss (sarcopenia).  - Higher fiber intake for digestion and heart health.  - Adequate calcium and vitamin D to maintain bone density.  - Limit sodium and processed foods to manage blood pressure[20]˒[21]. |

**References:**

[1] Song, Y., Liu, J., Zhao, K., Gao, L., & Zhao, J. (2021). Cholesterol-induced toxicity: An integrated view of the role of cholesterol in multiple diseases. In *Cell Metabolism* (Vol. 33, Issue 10). <https://doi.org/10.1016/j.cmet.2021.09.001>

[2] Castañer, O., Pintó, X., Subirana, I., Amor, A. J., Ros, E., Hernáez, Á., Martínez-González, M. Á., Corella, D., Salas-Salvadó, J., Estruch, R., Lapetra, J., Gómez-Gracia, E., Alonso-Gomez, A. M., Fiol, M., Serra-Majem, L., Corbella, E., Benaiges, D., Sorli, J. v., Ruiz-Canela, M., … Fitó, M. (2020). Remnant Cholesterol, Not LDL Cholesterol, Is Associated with Incident Cardiovascular Disease. *Journal of the American College of Cardiology*, *76*(23). <https://doi.org/10.1016/j.jacc.2020.10.008>

[3] Liu, C. Y., Chang, T. C., Lin, S. H., & Tsao, C. W. (2022). Is a Ketogenic Diet Superior to a High-Fat, High-Cholesterol Diet Regarding Testicular Function and Spermatogenesis? *Frontiers in Nutrition*, *9*. <https://doi.org/10.3389/fnut.2022.805794>

[4] Chen, Y., Yin, M., Cao, X., Hu, G., & Xiao, M. (2018). Pro- and anti-inflammatory effects of high cholesterol diet on aged brain. *Aging and Disease*, *9*(3).

<https://doi.org/10.14336/AD.2017.0706>

[5]al Rahmad, A. H. (2018). Pengaruh Pemberian Konseling Gizi terhadap Penurunan Kadar Kolesterol Darah. *Jurnal Kesehatan*, *9*(2). <https://doi.org/10.26630/jk.v9i2.947>

[6] Hamna Vonny Lasanuddin, Rosmin Ilham, & Rianti P. Umani. (2022). HUBUNGAN POLA MAKAN DENGAN PENINGKATAN KADAR KOLESTEROL LANSIA DI DESA TENGGELA KECAMATAN TILANGO. *Jurnal Ilmu Kedokteran Dan Kesehatan Indonesia*, *2*(1). <https://doi.org/10.55606/jikki.v2i1.566>

[7] Sasirekha, D., Chandra, E., & Ph, D. (2013). Text Extraction from PDF document. *Amrita International Conference of Women in Computing*.

[8] Sáez, T., Pageé, A., Kirschenman, R., Quon, A., Spaans, F., & Davidge, S. T. (2023). A High Cholesterol Diet During Late Pregnancy Impairs Long-Term Maternal Vascular Function in Mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *43*(1). <https://doi.org/10.1161/ATVBAHA.122.318421>

[9] Hamna Vonny Lasanuddin, Rosmin Ilham, & Rianti P. Umani. (2022). HUBUNGAN POLA MAKAN DENGAN PENINGKATAN KADAR KOLESTEROL LANSIA DI DESA TENGGELA KECAMATAN TILANGO. *Jurnal Ilmu Kedokteran Dan Kesehatan Indonesia*, *2*(1). <https://doi.org/10.55606/jikki.v2i1.566>

[10] World Health Organization (WHO) (2020). Healthy Diet. Retrieved from: <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>

[11] Centers for Disease Control and Prevention (CDC) (2023). Nutrition. Retrieved from: <https://www.cdc.gov/nutrition/>

[12] Dietary Guidelines for Americans, 2020-2025. U.S. Department of Health and Human Services. Retrieved from: <https://www.dietaryguidelines.gov/>

[13] American Heart Association (AHA) (2022). DASH Eating Plan. Retrieved from: <https://www.heart.org/en/healthy-living/healthy-eating>

[14] American Diabetes Association (ADA) (2023). Diabetes Nutrition. Retrieved from: <https://www.diabetes.org/healthy-living/food-nutrition>

[15] National Kidney Foundation (NKF) (2023). Kidney Disease Diet. Retrieved from: <https://www.kidney.org/nutrition>

[16] National Osteoporosis Foundation (NOF) (2021). Bone Health Guidelines. Retrieved from: <https://www.nof.org/>

[17] Academy of Nutrition and Dietetics (2022). Nutrition and Diet Recommendations. Retrieved from: <https://www.eatright.org/>

[18] National Institute on Aging (2022). Healthy Eating for Older Adults. Retrieved from: <https://www.nia.nih.gov/health/healthy-eating>

[19] Harvard T.H. Chan School of Public Health (2022). PCOS Diet and Nutrition. Retrieved from: <https://www.hsph.harvard.edu/nutritionsource/>

[20] Mayo Clinic (2023). Osteoporosis Prevention. Retrieved from: <https://www.mayoclinic.org/diseases-conditions/osteoporosis/symptoms-causes/syc-20351968>

[21] National Institutes of Health (NIH) (2021). Kidney Disease and Nutrition. Retrieved from: <https://www.niddk.nih.gov/health-information/kidney-disease/chronic-disease-nutrition>

[22] World Health Organization (WHO) (2020). Infant and Young Child Feeding. Retrieved from: <https://www.who.int/health-topics/infant-and-young-child-feeding>

[23] American Academy of Pediatrics (AAP) (2023). Nutrition in Childhood and Adolescence. Retrieved from:

<https://www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/Nutrition>