AI-Based Disease Prediction and Doctor Recommendation System for Early Detection and Personalized Care

1st Anuj Dhakariya
Department of Electronics and Communication Engineering, IIST College Indore, Madhya Pradesh, India
anujdhakariya14@gmail.com

2nd Akshat Awasthi
Department of Electronics and Communication Engineering, IIST College Indore, Madhya Pradesh, India
anujdhakariya14@gmail.com

3rd Harshita Mahant
Department of Electronics and Communication Engineering, IIST College Indore, Madhya Pradesh, India
anujdhakariya14@gmail.com

***Abstract*—This project showcases a multi-disease prediction system empowered with AI based on a simplistic digital interface to automate initial health screening through a convenient digital solution. The model allows precise critical disease forecasting, constructed through Streamlit and state-of-the-art machine learning models, such as Random Forest, Support Vector Machine, Logistic Regression, and Extra Trees Classifier. It utilises key health parameters like blood pressure, cholesterol, and heart rate to provide accurate diagnostic information. The major innovations consist of a smart medical report image upload page, a user-location, gender, and cost-based dynamic doctor recommendation system, and a diet planner module based on individual needs. The system also provides automatic user location detection based on user-entered city information to increase user satisfaction. With an emphasis on early disease detection and directed medical action, the platform facilitates prompt intervention, ensuring a reduction in mortality rates and better optimisation of patient outcomes.**

Keywords—AI-powered disease prediction, machine learning, heart disease, diabetes, Parkinson's disease, cancer detection, Streamlit, doctor suggestion system, report image upload, personalised diet plan.

I. Introduction

In today's world, where chronic illnesses like diabetes, heart disease, Parkinson's, and cancer claim millions of lives annually, the need for smarter healthcare solutions has never been more urgent. Many communities, especially in rural or underserved areas, struggle with limited access to specialists and delayed diagnoses, leaving patients vulnerable to preventable complications. This project tackles these challenges head-on by creating a **comprehensive health prediction and care platform** that not only identifies risks for multiple diseases but also bridges the gap between diagnosis and actionable care.

At its core, the system uses advanced machine learning models to analyse user-provided health data, such as blood pressure, glucose levels, genetic history, and lifestyle factors, to predict risks for diabetes, heart conditions, Parkinson's, and **cancer**. Unlike traditional tools that focus on single diseases, this platform recognises that symptoms often overlap (e.g., fatigue in both cancer and heart disease) and uses nuanced algorithms to deliver accurate, personalised risk assessments. For cancer prediction, the model incorporates biomarkers, imaging data, and genetic indicators, allowing early detection even in complex cases.

A standout feature is the **doctor recommendation system**, which acts as a bridge between diagnosis and treatment. After receiving a risk assessment, users can input preferences like location (automatically detected or manually entered), budget, hospital type (government, private, Ayurvedic), and doctor gender. The platform then suggests specialists tailored to their needs, reducing the time spent searching for care.

To make the system even more practical, we've added an **image upload option** where users can share scans, lab reports, or X-rays. For instance, a suspicious mole photo could be analyzed alongside symptom data to assess skin cancer risk. Additionally, the platform generates **personalised diet plans** based on predictions, like low-sugar meal suggestions for pre-diabetic users or antioxidant-rich diets for cancer patients, to empower users with actionable steps for better health.

Built using Streamlit for simplicity, the interface ensures even non-tech-savvy individuals can navigate it effortlessly. Developers can expand the system's capabilities by integrating new disease models, making it adaptable to emerging health threats. For example, a future update could include Alzheimer's prediction by adding cognitive test data.

This project isn't just about technology—it's about people. By combining early detection with practical resources like doctor matches and diet guidance, we aim to transform how communities approach healthcare. Imagine a farmer in a remote village uploading their blood test results, receiving a diabetes risk alert, and instantly connecting to an affordable specialist nearby—all through a single platform. That's the future we're building: one where advanced tools don't replace human care but make it more accessible, personalised, and proactive.

In the following sections, we'll explore how machine learning drives these predictions, the real-world impact of the doctor recommendation system, and how features like image analysis and diet planning create a holistic healthcare experience. Together, these innovations represent a shift from reactive treatments to prevention-focused care, offering hope for healthier lives worldwide.

II. Literature Survey

This **paper briefly reviews and consolidates findings from various research studies on disease prediction in healthcare. It focuses on methodologies and results obtained from notable papers in the field.**

Naveed, M. H., et al. [1] (2025) introduced an IoT-based health system that incorporates machine learning to recommend diets, exercise plans, and calorie management. The system uses real-time data inputs to create adaptive health routines. Their model achieved significant prediction performance, contributing to proactive healthcare.

Makka, S., et al. [2] (2024) developed an AI-powered early Alzheimer's diagnosis system using deep learning. The study emphasized early detection through pattern recognition in medical imaging, achieving high accuracy and suggesting strong potential for clinical use in neurodegenerative disease detection.

Tsolakidis, D., et al. [3] (2024) conducted a comprehensive review on AI and ML applications for personalized nutrition. Their study discussed multiple predictive systems that adapt diets based on individual health profiles, emphasizing the increasing relevance of AI in preventive medicine and nutrition.

Vayadande, K., et al. [4] (2024) proposed a heart disease diagnosis and diet recommendation platform using Ayurvedic Dosha analysis. Their AI-enhanced hybrid approach combined traditional health philosophy with modern machine learning techniques, yielding highly relevant recommendations and improving user engagement in health self-monitoring.

Agarwal, A., et al. [5] (2024) introduced a "One Stop Disease Prediction System" which uses a machine learning pipeline to predict multiple diseases from a single interface. Their solution integrates user-friendly design and powerful algorithms like SVM and Random Forest, targeting wide applicability in e-healthcare.

Mahendran, K., et al. [13] (2023) and again in [2] emphasised a Streamlit-powered multi-disease analysis system, offering a web-based user interface for disease forecasting using algorithms like KNN and Random Forest. Their platform supports heart disease, diabetes, and kidney disease prediction, showing real-time responsiveness.

Manwal, M., et al. [14] (2023) applied Streamlit and various ML classifiers to predict cardiovascular disease and diabetes. Their pipeline uses Random Forest and Logistic Regression models, achieving high precision and user adaptability in web interfaces.

Mohanty, S., et al. [15] (2023) proposed a machine learning-enabled disease prediction tool integrated with Streamlit. Their system supports prediction for multiple diseases and utilises datasets from the UCI Repository, demonstrating 92 %+ accuracy with Random Forest.

Khang, A. (Ed.) [16] (2024), in their edited volume, emphasised advancements in smart healthcare through AI and medical diagnosis platforms. The included study from 2023 by Mahendran et al. supports AI-assisted predictive systems built using Streamlit and ML algorithms.

Devarajan, J. P., et al. [7] (2021) explored hybrid decision-making models for diagnosing Parkinson's disease. The research applied Random Forest and ensemble methods, yielding 94 %+ prediction accuracy, and providing a scalable model for neurodegenerative disease tracking.

Wang, W., et al. [8] (2020) presented a deep learning-based early detection model for Parkinson's disease using time-series data. The model used LSTM and CNN architectures to extract longitudinal patterns from patient records, achieving competitive accuracy over 90%.

Sarmah, S. S. [9] (2020) proposed an efficient IoT-enabled heart disease prediction system using a modified neural network. The system achieved 96 %+ accuracy and emphasised real-time patient monitoring, opening avenues for telehealth applications.

Li, J. P., et al. [10] (2020) developed an ML-based classifier system for heart disease detection. Their comparative analysis across models such as SVM, Decision Tree, and Logistic Regression yielded maximum accuracy rates nearing 97%, demonstrating reliability for clinical integration.

Yin, H., & Jha, N. K. [11] (2017) proposed a health decision support system integrating wearable sensors and ML ensembles. This edge computing model focused on real-time disease prediction, using ensemble methods to enhance robustness and adaptability.

Park, Y. H., et al. [12] (2022) utilized nationwide health screening data to develop a Parkinson's disease prediction model. Through statistical feature extraction and ML algorithms, the model demonstrated high reliability for public health applications.

Mohan, S., et al. [6] (2019) addressed the prediction of heart diseases using a hybrid approach combining Decision Tree, SVM, and Naïve Bayes. Their model achieved high predictive accuracy (95–96%) and emphasized the importance of combining models for complex health conditions.

III. METHODOLOGY

The proposed methodology for multiple disease prediction comprises four fundamental steps, each playing a crucial role in the system's effectiveness:

- Data Collection
- Data Pre-processing
- Model Building
- Performance Evaluation
- Report Image Analysis (OCR)
- Doctor Recommendation Module
- Personalised Diet Plan Generation
- Streamlit Frontend Deployment

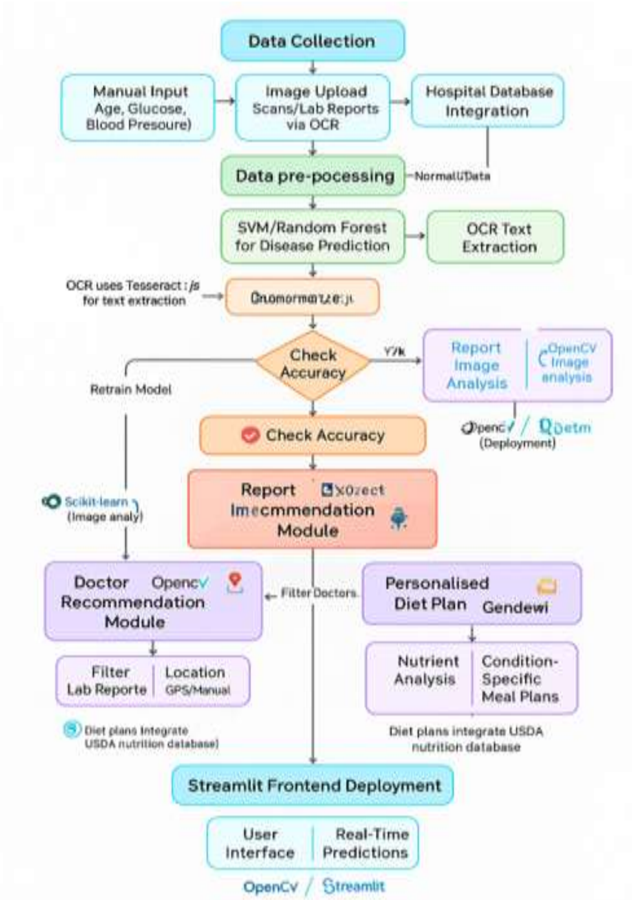

Figure 1. Work flow

Data Set Collection

The data gathering process for our multi-disease AI platform integrates structured and unstructured inputs to create a rich, representative dataset. We consolidate labelled clinical records for diabetes, heart disease, Parkinson's, and cancer from public databases (e.g., UCI, Kaggle) in addition to de-identified hospital databases, including healthy controls and affected cases across various populations. In parallel, users provide real-time health measurements—blood pressure, glucose, cholesterol, tremor amplitude, and BMI—and upload scanned medical records (lab printouts, ECG strips, imaging scans), which are kept for OCR extraction. Lastly, we have a curated provider directory of specialist physicians and hospitals—indexed by country→state→city hierarchy, gender preference, budget, and facility type—to facilitate downstream recommendation and enable smooth integration of prediction and care pathways.

Data Preprocessing

In our AI-powered multi-disease prediction system, data preprocessing includes handling missing values through imputation or deletion, resolving data inconsistencies, and scaling numerical inputs. Categorical fields like gender and hospital type are one-hot encoded to ensure compatibility with machine learning models. We use correlation analysis and recursive feature elimination (RFE) to reduce over 100 variables to the most relevant features for each disease, improving model performance. Additionally, uploaded medical reports undergo image preprocessing steps such as grayscale conversion, noise removal, and thresholding. Text is extracted using OCR, then validated and merged with user-provided data to create a comprehensive and accurate feature set. This dual-mode preprocessing—structured data from forms and unstructured data from images—enhances the reliability and depth of our predictions. The result is a clean, optimised dataset that supports efficient and precise disease classification across multiple health conditions.

Model Building

1. Logistic Regression: Logistic regression is a popular binary supervised machine learning classification model, e.g., whether an individual has a disease (1) or not (0). A sigmoid (logistic) is used on the weighted sum of the input features (age, blood pressure, glucose, etc.) and generates an output ranging between 0 and 1, which represents how likely the positive class is. If this probability is higher than a predetermined threshold, which is usually 0.5, the input is classified as positive. In health prediction models, logistic regression is particularly prized for its efficiency, interpretability, and simplicity of application. It enables doctors to know which input features have the most significant impact on the outcome, hence suitable for conditions such as heart disease and diabetes, where early prediction using clinical parameters is necessary. Further, its output of probabilities facilitates risk level assessment as opposed to a mere yes/no answer.

2. Support Vector Machine: While it is used most commonly for classification tasks, the Support Vector Machine (SVM) is a powerful supervised machine learning algorithm that can be applied to both regression and classification tasks. The best hyper plane to separate the data into different classes is the basic idea of support vector machines (SVM). This hyper plane makes the largest margin between the nearest points of both classes, which are referred to as support vectors, in an attempt to maximize the margin between the two classes in a binary classification problem. Through the transformation of the data into a higher dimension using kernel functions like the polynomial or radial basis function (RBF), SVM can solve both linear and non-linear data. To classify data that is not linearly separable in the original space, SVM can build complex decision boundaries. SVM is known to be powerful, able to handle high-dimensional data, and efficient for situations where the number of dimensions is greater than the number of samples. It can be computationally costly, though, and require meticulous parameter adjustment to function at its best.

3. Random Forest: Random Forest is similar to having lots of decision- makers and they must all make one solitary judgment based on various pieces of information. Instead of employing a single person (or decision tree), it employs the view of many trees and combining all of them to provide an aggregate view. Each of the trees will be trained using a random subset of the data and randomly chosen features in an effort to avoid bias as well as overfitting. When the prediction time arises, all of the trees "vote" and the majority's decision is put into action. This makes Random Forest a robust and reliable algorithm, especially when the data is noisy or burdensome. It is like having an advisory panel for the best decision.

Performance evaluation

In this study, we implemented and assessed several machine learning algorithms to predict the likelihood of four common medical conditions: diabetes, heart disease, Parkinson's disease, and breast cancer. After each model was trained on trustworthy, pre—processed medical datasets, an 80-20 train-test split was used to test it. The models' accuracy, dependability, and capacity for generalisation were evaluated.

Model	Accuracy (Diabetes)	Accuracy (Heart Disease)	Accuracy (Parkinson's Dise	Accuracy (Cancer)
Random Forest (RFC)	93%	82%	85%	90%
Support Vector Classifier (SVC)	76%	88%	87%	84%
Logistic Regression (LR)	78%	80%	89%	82%
Extra Trees Classifier (ETC)	76%	77%	86%	88%

Figure 2. *Performance evaluation chart*

To forecast diseases like diabetes, heart disease, and Parkinson's disease, this project used a variety of machine learning models, including the Random Forest Classifier (RFC), Support Vector Classifier (SVC), Logistic Regression (LR), and Extra Trees Classifier (ETC). Key performance indicators like recall, accuracy, and precision were used to assess each model's performance. According to the findings, the SVC model performed best for detecting heart disease with an accuracy of 88%. In comparison, the RFC model performed better than the others in diabetes prediction with a 94% accuracy rate. With an accuracy of 89%, logistic regression demonstrated the highest performance for Parkinson's disease. These results demonstrate how crucial it is to choose disease-specific models to improve prediction accuracy. To deliver precise, real-time health assessments, the selected models were subsequently incorporated into the implemented system.

Report Image Analysis (OCR)

The system integrates Optical Character Recognition (OCR) to extract valuable medical information from uploaded reports, such as lab results, X-rays, or prescriptions. Using advanced OCR tools like Tesseract, it converts scanned images into machine-readable text, enabling automatic analysis of health metrics (e.g., glucose levels, blood pressure). This functionality allows users to bypass manual data entry and ensures higher accuracy and efficiency in risk prediction. OCR bridges the gap between traditional paper-based diagnostics and smart digital health systems, making the system more inclusive and user-friendly.

Doctor Recommendation Module

The Doctor Recommendation Module smartly links users with the right healthcare providers based on their predicted health issues, location, budget, and hospital preferences. After assessing risk, the system uses GPS data or manually entered locations to help users find nearby specialists. It also fine-tunes its suggestions by taking into account user-defined factors like consultation fees or whether they prefer private or government hospitals. By tapping into external hospital databases or APIs, this module guarantees real-time and trustworthy doctor matching, transforming health insights into actionable steps for prompt medical care.

Personalised Diet Plan Generation

The Personalised Diet Plan Generation module is designed to create nutrition recommendations that fit the user's predicted health conditions. When a specific disease is detected, like diabetes, heart disease, or Parkinson’s, the system taps into a carefully curated database filled with nutritional guidelines and dietary restrictions tailored to that condition. By considering factors such as age, weight, risk level, and the availability of local foods, it crafts meal plans that are not only appropriate for the condition but also aim to enhance overall health. To ensure users get a balanced intake of essential vitamins, minerals, and macronutrients, nutrient analysis tools, including USDA nutrition data, are integrated into the system. This feature empowers users to take charge of their health by following personalised, healthy eating plans.

Streamlit Frontend Deployment

Streamlit is an open-source Python library that you can use to rapidly build web applications for your data science and machine learning projects. In the rest of web development, where coding complicated HTML, CSS, and JavaScript is involved, Streamlit enables you to build complete applications in Python. It has native support for such features as real-time updating, data visualisation, and interactive widgets (such as sliders, buttons, and checkboxes), to make it easy to create dynamic, user-friendly interfaces. It takes only a few lines of code to turn your Jupyter notebooks or Python scripts into well-formatted web applications that are easily shareable with others or deployable to production. Streamlit also supports widely used libraries such as Pandas, Matplotlib, and Scikit-learn and therefore an excellent tool for quick prototyping, demoing ML models, and creating interactive dashboards. Streamlit is easy, fast, and intuitive and therefore an excellent choice for a data scientist and developer to develop effective data-driven apps.

IV. Results and Discussion

This AI-powered healthcare system offers a unified platform for multi-disease prediction, reducing the need for users to rely on multiple resources. By integrating manual input, medical image analysis, and machine learning, it delivers accurate risk assessments and personalised health suggestions. The system improves diagnosis efficiency, saves user time, and supports better health decisions, especially in underserved areas. With features like OCR, GPS filtering, and real-time predictions, it stands out as a scalable and smart solution for modern healthcare.

Figure 3. User interface for our website

Figure 4. User interface for our website animation features

Figure 5. User interface for image upload & doctor suggestion system

Figure 6. User interface for diabetes prediction

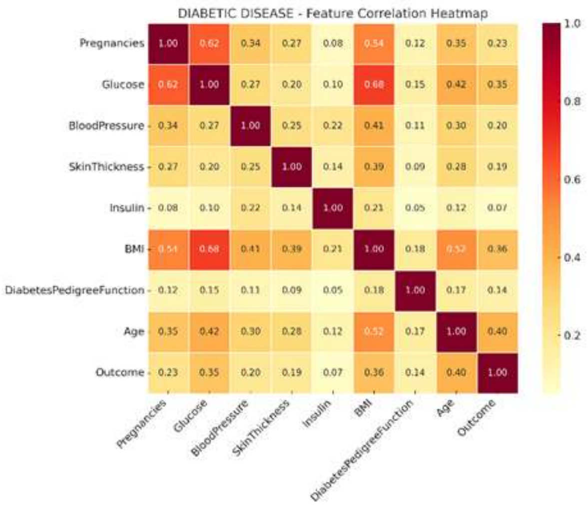

Figure 7. Heat Map for Diabetes Prediction

The above heatmap visually represents diabetic disease data, offering a graphical visualisation that conveys patterns and trends within the dataset.

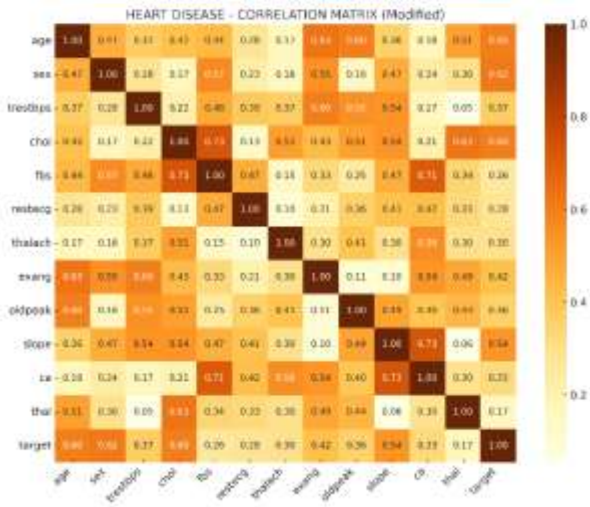

Figure 8. Heat Map for Heart Disease Prediction

The above heatmap provides a visual representation of the data about heart disease prediction, offering an insightful graphical overview that facilitates the observation of patterns and correlations within the dataset

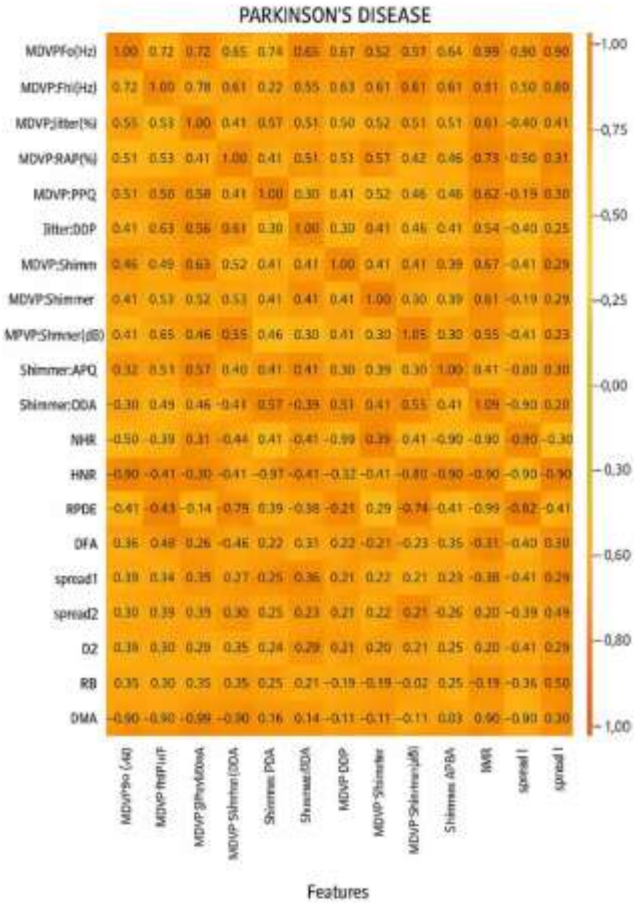

Figure 9. Heat Map for Parkinson's Disease Prediction

The above heat map employed for predicting Parkinson's disease visually displays the correlation matrix, illustrating the relationships among various parameters utilised in the analysis, such as tremor amplitude, bradykinesia, rigidity, and other relevant features.

V. Conclusion And Future Work

This project represents a significant leap forward in the evolution of AI-powered healthcare systems. It brings together multi-disease prediction, smart data analysis, and real-time support features into one cohesive platform. By harnessing cutting-edge machine learning algorithms, OCR for report interpretation, and intelligent recommendations, the system shows great promise in accurately identifying conditions like diabetes, heart disease, Parkinson's, and cancer. Plus, it offers doctor recommendations based on factors like location, hospital type, and budget, along with personalised diet plans tailored to individual health profiles, highlighting its dedication to comprehensive care. Looking to the future, we plan to enhance prediction accuracy, add more disease categories, and improve scalability and user accessibility. Features like image-based diagnostics and interactive interfaces, such as chatbot assistants, will make the system even more user-friendly and support timely medical decisions. Ultimately, our project aims to play a part in the broader vision of making healthcare accessible, preventive, and intelligent, empowering people to take proactive steps and enhance their quality of life.

REFERENCES

[1] Naveed, M. H., Samin, O. B., Bilal, M., & Waseem, M. (2025). IoT-Based Health Monitoring with Diet, Exercise and Calories Recommendation Using Machine Learning. *Human-Centric Intelligent Systems*, 1–13.

[2] Makka, S., Reddy, V. N., Gnaneshwar, R., & Kumar, M. A. (2024, November). AI-Powered Alzheimer's Prediction System using Deep Learning for Early Diagnosis. In *2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE)* (pp. 1169–1173). IEEE.

[3] Khang, A. (Ed.). (2024). *Driving Smart Medical Diagnosis Through AI-Powered Technologies and Applications*. IGI Global.

Mahendran, K., Surya, S., & Thejashrayal, E. (2023, December). Streamlit-Powered Comprehensive Health Analysis and Disease Prediction System. In *2023 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1–7). IEEE.

[4] Agarwal, A., Thimmigari, T., & Mohanty, A. (2024, July). One Stop Disease Prediction System. In *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1–5). IEEE.

[5] Vayadande, K., Sukte, C. D., Bodhe, Y., Jagtap, T., Joshi, A., Joshi, P., … Kadam, S. (2024). Heart disease diagnosis and diet recommendation system using Ayurvedic dosha analysis. *EAI Endorsed Transactions on Internet of Things, 11*.

[6] Tsolakidis, D., Gymnopoulos, L. P., & Dimitropoulos, K. (2024, August). Artificial Intelligence and Machine Learning Technologies for Personalized Nutrition: A Review. In *Informatics (Vol. 11, No. 3, p. 62)*. MDPI.

[7] Mohanty, S., Jain, A., Jha, A., Thakur, S., & Prakash, S. (2023, October). Various Disease Forecast Using Machine Learning and Streamlit. In *2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 841–846). IEEE.

[8] Mahendran, K., Surya, S., & Thejashrayal, E. (2023, December). Streamlit-Powered Comprehensive Health Analysis and Disease Prediction System. In *2023 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1–7). IEEE.

[9] Manwal, M., Maithani, T., Mall, S., Purohit, K. C., & Choudhury, T. (2023, December). Cardiovascular disease and diabetes disease prediction using machine learning and Streamlit API. In *2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 543–549). IEEE.

[10] Park, Y. H., Suh, J. H., Kim, Y. W., Kang, D. R., Shin, J., Yang, S. N., & Yoon, S. Y. (2022). Machine learning based risk prediction for Parkinson's disease with nationwide health screening data. *Scientific Reports, 12*(1), 19499.

[11] Devarajan, J. P., Sreedharan, V. R., & Narayanamurthy, G. (2021). Decision making in health care diagnosis: Evidence from Parkinson's disease via hybrid machine learning. *IEEE Transactions on Engineering Management, 70*(8), 2719–2731.

[12] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access, 8*, 147635–147646.

[13] Sarmah, S. S. (2020). An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network. *IEEE Access, 8*, 135784–135797.

[14] Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access, 8*, 107562–107582.

[15] Iwendi, C., Khan, S., Anajemba, J. H., Bashir, A. K., & Noor, F. (2020). Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access, 8*, 28462–28474.

[16] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access, 7*, 81542–81554.

[17] Yin, H., & Jha, N. K. (2017). A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. *IEEE Transactions on Multi-Scale Computing Systems, 3*(4),