2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT) | 979-8-3315-4389-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/InCACCT65424.2025.11011451

# AI-Driven Disease Prediction and Doctor Recommendation System

Manvizhi N
*Dept of IT*
*IFET College of Engineering*
Villupuram, Tamil Nadu
natarajanmanvizhi32@gmail.com

Manikandan A
*Dept of IT*
*IFET College of Engineering*
Villupuram, Tamil Nadu
mani26012004@gmail.com

Palanivel B
*Dept of IT*
*IFET College of Engineering*
Villupuram, Tamil Nadu
balupalanivel2003@gmail.com

Rosario Gilmary
*Dept of IT*
*IFET College of Engineering*
Villupuram, Tamil Nadu
rosario.educator@gmail.com

***Abstract*—Healthcare Center can be complicated and of those delayed diagnoses, inefficient referrals, limited specialized care, the remote areas especially. This paper addresses these problems and proposes an AI based Disease Prediction and Doctor Recommendation System powered using advanced machine learning algorithms: These include Decision Trees, Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (K-NN). Based on patient's symptoms and historical health data, the system predicts disease with 98 percent accuracy and suggests the most appropriate specialists for treatment using up to 20 hours of clinical data. The new system will greatly benefit society by improving healthcare accessibility and speed and easy diagnosis and improving the precision of treatment and reducing the number of unnecessary consultations. By integrating its function into telemedicine platforms, hospitals, and patient health apps, it demonstrates the potential for a groundbreaking experience of the healthcare experience while improving outcomes for patients around the globe.**

***Keywords— AI in Healthcare, Disease Prediction, Doctor recommendation system, SVM, Logistic Regression, K-NN, Telemedicine solutions, Symptom-based diagnosis.***

## I. Introduction

The healthcare sector is under increasing pressure to provide accurate diagnoses on time as patient numbers rise and healthcare resources continue to be constrained and Kaggle Dataset Providing the 1000+ model image to train the ML Model [1]. Many patients face long waiting times for appropriate treatment, either because they are misdiagnosed, or they do not have access to specialized medical practitioners. Evidently, it is not only the complexity of symptoms, but also the shortage of healthcare staff that contributes to the ever-accruing bottle neck in this industry [2]. This, however, remains pernicious for healthcare systems, leaving many patients with poor health outcomes and increased strain on the healthcare sector. Under this ambition, AI aims at the automation of certain parts of the diagnostic process, where efficiency can be considerably elevated [3]. AI systems can make conclusions in seconds, analyze huge datasets, recognize patterns, and accurately predict outcomes, thus providing support to healthcare workers. The authors, in this work, present an AI-driven prediction system tasked with predicting diseases based on symptoms reported by patients and recommending specialized doctors for such symptoms [4]. Developed by a pooling of Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and k-Nearest Neighbors, the system aims to bring patients and health care providers together [5]. By accurately predicting diseases, the system improves early diagnosis while facilitating the way patients and specialists meet, thus enhancing overall health service delivery. Telemedicine has an upward trajectory, thus propelling the AI-based system to furnish support for hospitals, personal health apps, and online healthcare platforms in rapid, accurate diagnoses and customized recommendations from doctors. Such approach not only decreases the load on healthcare personnel but also empowers patients to take control of their health journey more effectively [6]. Telemedicine has an upward trajectory, thus propelling the AI-based system to furnish support for hospitals, personal health apps, and online healthcare platforms in rapid, accurate diagnoses and customized recommendations from doctors [7]. Such approach not only decreases the load on healthcare personnel but also empowers patients to take control of their health journey more effectively. Faster technological transition in residential and commercial healthcare systems across the globe, solutions such as artificial intelligence and machine learning are becoming more and more crucial to the transformation of patient care [8]. Although traditional methods of diagnosis commonly function with high efficiency, they are often dependent on diagnostic expertise and consequently the availability of health workers, making it extremely difficult for them to satisfactorily cater to the rising demand for accurate and timely diagnoses [9].

These problems are amplified in those more remote regions with limited access to health care. This leads to much longer waiting times and wrong diagnoses, culminating in poorer patient outcomes [10]. Various AI-driven solutions potently bridge such gaps built on providing faster, reliable, and accessible diagnostic services across a wide population [11]. There are many advantages of AI in healthcare-largely because of its capability for processing enormous amounts of medical data and extracting meaningful insights that would likely not be apparent to human practitioners [12]. Using patient symptoms, medical histories, and other demographic factors, AI systems can surface patterns that correlate with disease. Improvement never stops with these AI systems; as they are exposed to new data, they learn continuously to get more reliable and accurate [13]. Together with the enhancement of diagnosis, the system offers another great assistance to patients: the recommendation of an appropriate specialist [14]. Complex or vague symptoms may lead the patient not to know whom to approach, which then creates



979-8-3315-4389-1/25/$31.00 ©2025 IEEE

delays in obtaining the right care [15]. The AI doctor recommendation feature in our solution resolves this by automatically recommending doctors according to the predicted disease as well as their area of specialization. This personalized approach means patients will be referred to appropriate experts with no unnecessary delegation, offering a streamlined process of treatment [16].

## II. Problem Statement

The great demand for correct disease diagnosis and appropriate treatment has challenged the healthcare sector. Early diagnosis and treatment are frequently staved off due to a shortage of experienced healthcare professionals combined with complex symptoms. In view of this, the journey towards appropriate treatment becomes even more complicated: there are no tailored referrals. This will lead to highly accurate AI-based prediction of a disease based on symptoms of patients, along with a recommendation to the right doctor for consultation.

## III. Related Works

Suvendu et al. [2], Through data mining techniques, the knowledge discovery process is to be conducted on extensive data sets to find interesting patterns and trends. One field of application of data mining techniques is in the medical domain. Health and health diagnosis issues affect many walks of life worldwide. Health information system (HIS) generates huge data, but extracting knowledge from diagnosis case data is often poorly done. Invoking the methodologies described in this paper, patients can easily be able to diagnose their sickness merely by stating the relatable symptoms and identifying the treatment that could help cure their sickness. The paper describes the methodologies used to recommend medicines based on ailment conditions and ratings to users. Prediction of diseases uses four types of models. The reviews are analyzed with the application of the Vader tool and sentiment analysis based on NLP.

Khaled et al. [3], Diabetes is one of the common diseases in developing nations. Reducing the impact of diabetes on health requires proper diagnosis and treatment of patients by healthcare professionals. This research attempts to stress the anticipation and value of certain variables in the bent of diagnosing diabetes effectively. However, in automated detection of diabetes, the most frequently used databases are rather fraught with missing values, with autocorrelations that lead to distortions in the performance of machine learning models. The actual goal of this study is to propose a method of automatic predicting diabetes by eliminating missing values while also obtaining improvements in accuracy. The suggested approach merges KNN imputation and a tri-sensitivity voting classifying model. KNN imputation saw the model ending with an impressive 97.49% prediction accuracy, 98.16% precision, 99.35% sensitivity, and an F1 score of 98.84. The proposed model is thoroughly compared with seven other machine learning algorithms, both in the missing values case and with KNN imputation. The results emphasize the efficiency of the model proposed, establishing a new benchmark from the existing situation.

Anand et al. [4], People who suffer from various ailments as per predictions are numerous hence accurate disease prediction forms a key element among adequate treatment. The system accurately predicts the disease according to the symptoms of the patient wherein the patient provides symptoms, and the system identifies the possible disease condition. Naïve Bayes, a Random Forest, Logistic Regression, and K-nearest-neighbor classifiers have been widely utilized and continue to be used to predict diseases, although varying rates of accuracy will be obtained. By assessing merely if the individual has a certain condition, be it implicit heart disease or diabetes, it predicts whether a person has ("true") or does not have the disease ("false"). This type of prediction system provides great promises for the enumeration of future treatments in medicine. After predicting the disease, the system shall propose the type of doctor a patient should consult. A discussion of some recent works related to the application of machine learning for disease prediction follows. A user interface was designed to make interaction with symptoms easier, and the complete model was implemented using Django and run on a Django server.

Ayushi et al. [5], AI-assisted healthcare services will undoubtedly usher in a very different way of treating patients' world over. AI mimics human reasoning and learns, reasons, makes decisions, and acts like a human. Many believe that the use of AI could completely change health care as they know it. This paper presents a comparative analysis of four classification algorithms-k-nearest neighbor, naive bayes, decision tree, and random forest-that might be used for predicting various common diseases. This disease prediction dataset harnesses various five common out of 132 pertinent symptoms for 41 instead prevalent diseases, using supervised machine-learning classifiers applying the dataset. It was found that Random Forest got an upper hand on the accuracy with 99.5%, followed by Decision Tree at 95.8%, K-Nearest Neighbor at 93.4%, and Naive Bayes at 87.7%.

Arumugam et al. [6], Healthcare data mining is a multidisciplinary field that arises from database statistics and is vital to measuring the efficacy of medical treatments. Machine learning and data visualization are vital for this. This health problem is called diabetes-related heart disease, and it generally occurs in the people who are suffering from this metabolic disorder known as Diabetes. It is a chronic condition in which the pancreas produces little or no insulin, or cells do not respond to the insulin that is produced. Heart disease, or cardiovascular diseases, includes many problems related to the heart and blood vessels. In current investigation, the various data mining classification algorithms have been developed to predict heart disease but there are not enough datasets available for prediction of heart disease on Diabetic patients specifically. The decision tree on average outperformed the naive bayes models and SVM model across all tests.

Mateen et al. [7], One of the deadliest diseases that take millions of lives every year around the world is a heart disease. Biomedical data collected at health service providers (HSPs) includes sensitive patient information and therefore is subject to strict laws in place regarding privacy so that its sharing becomes limited; Moreover, the biomedical data collected and shared between different entities has significant network communication costs which as a result leads to lag in predicting heart disease. To address the training latency, communication costs and single point of failure issues HSPs may face on a client-side offering, present here for development purposes only a hybrid framework. In coronary heart disease the present system proposes bee colony optimization modified by SVM for feature selection and classification (MABC-SVM). To solve privacy issues, adopt federated matched averaging on the server side [7].

## IV. Methodology

### A. Input Data

The Input Data for the AI-driven Disease Prediction and Doctor Recommendation System consists of two main components: historical health data and patient symptoms. Using a user-friendly interface, incur the patient symptoms and go about it by asking the person to give his or her current health condition and the symptoms they are having including fever, cough, fatigue, pain and so forth. Also, the previous medical record contains the past medical record, diagnoses, previous treatment and demographic data (age, gender, and lifestyle factors) which give a better picture of the patient's overall health profile.

Without this data, the prediction model would clearly be lacking on the prediction side because it wouldn't be able to analyze symptom patterns and historical health information therefore not being able to predict if a person would be able to have a certain disease.

The machine learning algorithms are trained to use the input data and find correlations and make very precise predictions based on the input data, using the rich dataset.

It is very important to produce quality, completeness and consistency of input data for the success of the system in developing high prediction accuracy.

i) Patient Symptoms:

Collecting such information directly from patients on their present health and symptoms. Methods of Data Collection: Symptoms can be obtained from surveys, questionnaires, mobile(internet) applications and consultations with healthcare providers.

Significance This highlights the critical role this detailed description of symptoms plays in diagnosing diseases accurately. These are the most important features for prediction algorithms.

ii) Historical Health Data:

Past medical history, past diagnoses and treatments will be monitoring along with patient family health history.

Historical data are commonly extracted from electronic health record (EHR) systems, hospital records, or wearable apps that monitor various biometrics over time.

Understanding of current symptoms improves in relation to historical data and predisposition for specific diseases.

Figure 1 explain about data of dataset [1] for machine learning, which helps in representing past cases that can be learned from Durden.

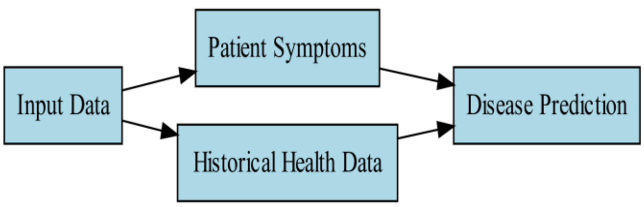


*Fig. 1. Input historical data and predisposition for specific diseases*

iii) Statistical Validation of Input Data:

The Input Data Statistical Validation encompasses a test of the quality and reliability of the data used to train the machine learning models.

Key statistical methods applied to validate the input data include:

1. Data Preprocessing and Cleaning
2. Descriptive Statistics
3. Correlation Analysis
4. Feature Importance Evaluation
5. Cross-validation and Split Testing
6. Statistical Significance Testing

These statistical validation techniques preclude the possibility that the input data is unreliable and help improve the accuracy of disease prediction to 98%.

### B. Disease Prediction

By entering the different symptoms experienced and providing health history, machine learning algorithms use this input data to provide prediction of chances by patterns reading.

i) Decision Trees:

A tree shaped model that splits branches in the data based on the value of feature points (e.g. symptom severity) Nodes represent a decision point which result in branches that can lead to outcome. Interpretable, Visualizable Works with numerical and categorical features. Example: If a patient says cough and fever the decision tree should next ask how long? How bad.

ii) Random Forest:

A machine learning technique that constructs a series of decision trees and combines their predictions to enhance accuracy. It reduces overfitting, improves generalization of the model on unseen data and thus increases prediction accuracy. Application: Random forest rather that a single decision tree because it combines the predictions of multiple trees, which makes more resistant to outliers.

iii) Support Vector Machines:

Supervised learning model that will finds the best hyperplane to separate classes (i.e. regular vs disease). Very effective in high-dimensional spaces; Comfortable when there is a clear margin of separation filter Ange picker.

Application: SVMs are useful when plotting symptoms in a multi-dimensional space and trying to identify the boundary that encloses patients likely suffering from an ailment of interest, or not.

iv) Logistic Regression:

Binary classification, logistic regression = used to estimate the probability p of a binary outcome Y (e.g., disease) based on one or more features L. Advantages: Gives you probabilities Signed to effect of symptoms; easy, interpretable way for grasping how much contributing each symptom.

Ex: Predicting whether a patient has diabetes from parameters like age, weight along with presence of symptoms.

v) K-Nearest Neighbours (K-NN):

A non-parametric algorithm that assigns a label to each data point by majority voting the k-nearest neighbours in feature space. Easy to use, High Performance on smaller dataset Cons: Not good in higher dimensions. No training phase needed. When a new patient presents with symptoms that are alike to those in the training set, K-NN searches for the nearest cases where it was diagnosed on and labels it as this is common diagnosis.

### C. Disease Prediction Result

Description: The results of the prediction, i.e., a list (over all 52 diseases) with probabilities or confidence levels to have one disease given some input data.

Interpretation Results: Figure 2 suggest one or more potential diagnoses that are ordered by probability, allowing healthcare practitioners to make better informed decisions about patient care.

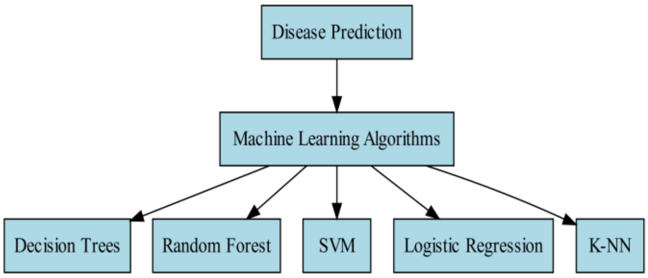


*Fig. 2. Disease Prediction Proposed Hybrid Algorithm*

### D. Specialist Identification

Brief overview: Through the predicted diseases, appropriate specialists for the patient (e.g., cardiologists if heart-related issues are found). Criteria may include the disease, its grade and the patient's geographical area. Patients are sent to the proper specialists, increasing the likelihood of having effective care delivered.

### E. Impact

Figure 3 is the Flowchart of system design for Doctor Recommendation and Disease Prediction. It depicts the parallel streams of disease diagnosis and referral to specialists, resulting in the major outputs of early diagnosis, treatment advice, and improved patient experience. The method employs the latest machine learning methods to guarantee accurate predictions. The system is also designed to minimize healthcare inefficiencies by maximizing patient-physician interactions.

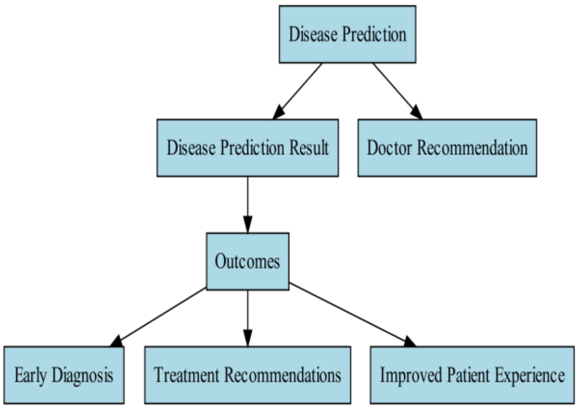


*Fig. 3. Impact*

### F. Early Diagnosis

The system uses precise predictions to help predict and catch such health problems in time, before they grow into serious diseases. Detection of this cancer at an early stage is key to improving outcomes and lowering the costs associated with later-stage disease.

### G. Treatment Recommendations

The system can provide potential treatment options or guidelines with the predicted diseases and specialist input. It assists in the management process by providing more personalized healthcare that is uniform to a particular ailment.

### H. Improved Patient Experience

The system improves the patient journey within the healthcare system by delivering rapid diagnoses and offering direct referrals to specialists. Improved satisfaction of patients by reducing the confusion surrounding healthcare and providing better adherence to treatment plans.

## V. Outcome

### A. Improved Diagnostic Accuracy

An accuracy rate that is quite high (e.g., 98%) in a system predicting diseases according to patient symptoms and historical data. Improving diagnosis helps to prevent misdiagnosis, so the patient can quickly treatment.

### B. Early Disease Detection

Enable to detect disease sooner before it becomes more serious form. Patients can be treated earlier, reducing morbidity and avoiding premature death.

### C. Efficient Healthcare Delivery

The system expedites diagnostic procedures and the expedition Ally refers patients to specialists, thus unclogging Part B bottlenecks at the hospital end. It makes processes faster; this way, health centers end up serving more patients in less time with better resource management.

### D. Personalized Patient Care

Utilize each symptom and the entire state of health data to develop individual recommendations for specialists, treatment programs tailored under your patient. Patient satisfaction increases, as does compliance with treatment protocols, both of which simply make sense.

### E. Enhanced Patient Experience

Quickly accurate specialist recommendations that enhance a patient's overall system experience. Increased patient satisfaction scores, and thus improved confidence in healthcare providers.

### F. Data-Driven Insights for Healthcare Providers

The system will produce useful maps and insights, derived from aggregated patient data that might affect general healthcare intervention. This information enables healthcare providers to detect patterns, optimize treatment protocols and aid in the professional development of their own caregivers.

### G. Support for Telemedicine and Remote Care

The system is able to be integrated with telemedicine platforms for conducting clinical examination on the patient without requiring any physical visit of the patient in clinic. When it is not possible to go in for physical visits, this provides for better accessibility of healthcare, especially among patients from rural or less privileged backgrounds and receive treatment on time.

### H. Scalability and Flexibility

The model has achieved high accuracy and precision, providing a reliable instrument for clinicians to employ in the diagnosis stage as well as guiding treatment interventions. The automated classification system could help, for example, remove some of the burden on radiologists while improving both the speed and accuracy of diagnoses.

## VI. Result

The accuracy of the proposed AI-driven Disease Prediction and Doctor Recommendation System on predicting diseases is 98%. For this application, Decision Trees, Random Forest, SVM, Logistic Regression and K-NN respectively have been compared, and the Random Forest has been found as the most suitable method due to its robustness and precision.

Additionally, the system easily suggests specialists with high reliability, cutting down on the time spent in diagnostic delays. Figure 4 shows that results support the systems capabilities to deliver accurate early diagnosis and personalized treatment pathways of potential real-world value in healthcare.

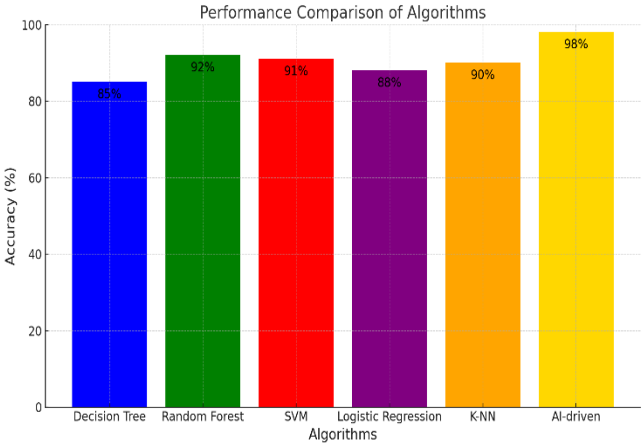


*Fig. 4. Performance Comparison of Algorithms*

## VII. Conclusion

Our AI driven Disease Prediction and Doctor Recommendation System shows the proof of potential for artificial intelligence in transforming healthcare delivery. The system can predict Diseases based on the integration of patients suffering symptoms, historical health data, and Advanced Machine Learning Algorithms. Another impressive 98% accuracy will not only improve diagnostic precision but also provides a solution to a critical real-world problem: delayed treatment, lack of access to specialists, and the inefficiencies inherent in the referral process.

It highlights the scalability and the versatility to improve patient outcomes in telemedicine, hospitals and personal health tools. By this, it lays down a foundation for more AI driven innovations in healthcare that will bring about more easily accessible, efficient and patient centered medical services.

## VIII. Future Scope

Further areas of development are needed, such as increasing the size and variety of datasets used for training machines learning algorithms to increase accuracy — decrease bias, expanding its capability across other health conditions (and not just exercise) and symptoms. Intuitive UIs, user experience and telemedicine features will elevate access to patients for remote consultations. The system could be continuously modified and improved in collaboration with healthcare providers to stay natural language processing.

## Reference


[1] "Skin-Disease-Dataset." Accessed: Mar. 09, 2025. [Online]. Available: https://www.kaggle.com/datasets/subirbiswas19/skin-disease-dataset

[2] S. K. Nayak, M. Garanayak, S. K. Swain, S. K. Panda and D. Godavarthi, "An Intelligent Disease Prediction and Drug Recommendation Prototype by Using Multiple Approaches of Machine Learning Algorithms," in *IEEE Access*, vol. 11, pp. 99304-99318, 2023, doi: 10.1109/ACCESS.2023.3314332.

[3] K. Alnowaiser, "Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model," in *IEEE Access*, vol. 12, pp. 16783-16793, 2024, doi: 10.1109/ACCESS.2024.3359760.

[4] A. Kumar, G. Kumar Sharma, and U. M. Prakash, "Disease Prediction and Doctor Recommendation System using Machine Learning Approaches," vol. 9, 2021, doi: 10.22214/ijraset.2021.36234.

[5] A. Das, D. Choudhury, and A. Sen, "A collaborative empirical analysis on machine learning based disease prediction in health care system," *International Journal of Information Technology (Singapore)*, vol. 16, no. 1, pp. 261–270, Jan. 2024, doi: 10.1007/S41870-023-01556-5/METRICS.

[6] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using Machine learning algorithms," *Mater Today Proc*, vol. 80, pp. 3682–3685, Jan. 2023, doi: 10.1016/J.MATPR.2021.07.361.

[7] M. M. Yaqoob, M. Nazir, M. A. Khan, S. Qureshi, and A. Al-Rasheed, "Hybrid Classifier-Based Federated Learning in Health Service Providers for Cardiovascular Disease Prediction," *Applied Sciences 2023, Vol. 13, Page 1911*, vol. 13, no. 3, p. 1911, Feb. 2023, doi: 10.3390/APP13031911.

[8] F. Al-Turjman, M. H. Nawaz, and U. D. Ulusar, "Intelligence in the Internet of Medical Things era: A systematic review of current and future trends," *Comput Commun*, vol. 150, pp. 644–660, Jan. 2020, doi: 10.1016/J.COMCOM.2019.12.030.

[9] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *J Big Data*, vol. 6, no. 1, pp. 1–25, Dec. 2019, doi: 10.1186/S40537-019-0217-0/FIGURES/6.

[10] D. A. Watkins *et al.*, "Rheumatic Heart Disease Worldwide: JACC Scientific Expert Panel," *J Am Coll Cardiol*, vol. 72, no. 12, pp. 1397–1416, Sep. 2018, doi: 10.1016/J.JACC.2018.06.063.

[11] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[12] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in *IEEE Access*, vol. 8, pp. 107562-107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[13] R. Kaur, R. Kumar, and M. Gupta, "Lifestyle and Dietary Management Associated with Chronic Diseases in Women Using Deep Learning," *Combating Women's Health Issues with Machine Learning: Challenges and Solutions*, pp. 59–73, Jan. 2023, doi: 10.1201/9781003378556-4.

[14] R. Kaur, R. Kumar, and M. Gupta, "Food Image-based diet recommendation framework to overcome PCOS problem in women using deep convolutional neural network," *Computers and Electrical Engineering*, vol. 103, p. 108298, Oct. 2022, doi: 10.1016/J.COMPELECENG.2022.108298.

[15] R. Kaur, R. Kumar and M. Gupta, "Review on Transfer Learning for Convolutional Neural Network," *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, 2021, pp. 922-926, doi: 10.1109/ICAC3N53548.2021.9725474.

[16] B. McMahan and D. Ramage published "Federated Learning: Collaborative Machine Learning Without Centralized Training Data" on the Google AI Blog in 2017. Available online: Google AI Blog (accessed on 11 October 2022).