Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P. 21779-21787 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Integrated Disease Forecasting: Leveraging Deep Learning and Machine Learning for Multi-Disease Prediction

Dr.Anitha Pradhan[1*], J.S.V.Gopala Krishna[2], B.Prasanna Kumar[3], G.Tabita[4], V.V.R.Lsastry[5] , Kattupalli Sudhakar[6]

[1*]Assistant Professor, Department of Computer Science & Engineering KLEF, Vaddeswaram, AP, India. anitha.pradhan15@gmail.com
[2]Associate Professor, Department of Computer Science & Engineering, SIR C.R.R. College of Engineering, Eluru, AP, India. Jsvgk4321@gmail.com
[3]Associate Professor, Department of Computer Science & Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, AP, India. prasannabpk@gmail.com
[4]Assistant Professor, Department of Electrical and Electronics Engineering Lakireddy Bali Reddy College Of Engineering Mylavaram, Andhra Pradesh, India gandham820@gmail.com
[5]Asst Professor Cse Dept, Vignan's Institute Of Information Technology (Autonomous), Duvvada, Visakhapatnam.sastryvvrl@gmail.com
[6]Senior Assistant Professor, Department of AI &DS, Lakireddy Bali Reddy College of Engineering, Mylavaram, AP, India.
sudhamtech@gmail.com

How to cite this article: Anitha Pradhan, J.S.V.Gopala Krishna ,B.Prasanna Kumar,G.Tabita, V.V.R.Lsastry,Kattupalli Sudhakar (2024). Integrated Disease Forecasting: Leveraging Deep Learning and Machine Learning for Multi-Disease Prediction. *Library Progress International,* 44(3), 21779-21787

Abstract—The paper presents how can we predict multiple disease prediction by utilizing machine learning(ML) and deep learning(DL) techniques by analyzing different datasets consisting of various medical parameters our approach integrates ML algorithms such as Logistic regression,KNN,Random forest,support vector machine for initial classification and then sequentially used DL methodology ANN(Artificial Neural Networks) through this extensive experimentation and validation of real world medical datasets, our proposed framework gives the promising outcomes in achieving a remarkable performance of predicting various diseases such as breast cancer, lung cancer, diabetes, monkeypox, parkinson's. thus offering valuable insights to get early diagnosis. which helps the healthcare professionals to detect and diagnose the disease at preventive stages to improve the patient care. The interface we used for predicting diseases is Streamlit.

Keywords: Machine Learning, Deep Learning, ANN, KNN, SVM, Random Forest, Logistic Regression, Gaussian naive bayes and Streamlit.

I. INTRODUCTION

In this modern health care,early detection and diagnosis of diseases are important for intensive treatment with the advantage of emerging technologies such as machine learning (Ml) and deep learning(DL) . There has been a need for using such new technologies in predicting health problems before they become complex ones. these techniques works on huge amount of real world medical data to find out weather the disease would occur to someone by based on the various medical parameters.most of the techniques figure out only one disease at a time.although it may not be at earlier stage.which may not be helpful because people often have multiple diseases.therefore our aim is to predict multiple diseases at initial stages by based on the medical parameters using these emerging technologies.by combining strengths of ML and DL we focused to create a strong predictive model which would be capable of analyzing the complex data and providing thevaluable information to assist the health care professionals as well as the people who are suffering can get cured easily.

Integrating the strengths of machine learning and deep learning we built a framework which predicts multiple diseases like heart, parkinson's, lung cancer, breastcancer, diabetes simultaneously. this predictive model offers the valuable information to get quick diagnosys for patients and help the health care professionals.here,in this

predictive model initially the ML algorithms logistic regression,KNN,random forest,support vector machine were employed to make initial predictions and subsequently DL approach ANN(Artificial Neural Network) to refine and enhance the model to produce the predictions.the ultimate goal of our research is to offer valuable insights for early diagnosis,which is crucial for improving patient care .by predicting and preventing diseases at early stages.the contributions of this paper exists in development of comprehensive predictive model which integrates both ML and DL for multi-disease prediction .we focused to overcome the limitations of traditional methodologies and achieve the best performance in disease prediction tasks Beyond the technical aspects, our framework holds significant implications for healthcare systems and society as a entirely.. By automating the disease prediction process and providing valuable insights for early diagnosis, our framework can empower healthcare professionals across diverse settings, including resource-constrained environments and underserved communities.Another important extension of our research involves the ethical and societal implications of deploying
ML and DL models in healthcare settings. As with any technology-driven innovation, careful consideration must be given to issues such as data privacy, transparency, and algorithmic bias. Our framework emphasizes the importance of responsible AI practices, including robust data governance, model interpretability, and bias mitigation strategies, to ensure the equitable and ethical use of predictive analytics in healthcare.The process outlined in our framework for predicting multiple diseases involves several key steps, each leveraging different machine learning (ML) and deep learning (DL) techniques. The first step is to gather diverse medical datasets containing various parameters related to the diseases of interest. The dataset we used for the prediction of diseases is taken from the kaggle. These datasets may include information such as medical parameters and test reports. Once collected, the data undergo preprocessing steps to clean and prepare it for analysis. This may involve handling missing values, normalizing data, and encoding categorical variables.then the relevant features or variables are selected from the dataset based on their potential importance in predicting the target diseases. Feature engineering techniques may also be applied to create new informative features from existing ones, enhancing the predictive power of the model.next Initial Classification with various ML algorithms such as Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM) are employed for initial classification. These algorithms analyze the dataset and learn patterns to classify instances into different disease categories based on the selected features. Sequentially utilized DL Methodology (ANN) After the initial classification by ML algorithms, the framework incorporates DL methodology, specifically Artificial Neural Networks (ANN), for further refinement. ANN is particularly adept at learning complex patterns and representations from high-dimensional data. By sequentially integrating DL techniques into the process, the framework aims to improve the accuracy and robustness of disease prediction.Throughout the process, the performance of the predictive models is evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. Additionally, the models are validated using real-world medical datasets to ensure their effectiveness in predicting various diseases accurately.

II. RELATED WORK

Anuradha D.Gunasinghe [1] This paper explores using machine learning to predict and diagnose lung diseases early, aiding doctors in saving lives by analyzing patient data including age, gender, and chest X-rays to determine the presence of conditions like asthma, COPD, tuberculosis, pneumothorax, and lung cancer.

Anup Lal Yadav [2] This system utilizes artificial intelligence to analyze heart disease data accurately, employing different algorithms like decision trees and logistic regression, aiming to predict various heart conditions effectively.

Muhammad Usman Ghani [3] Using data mining techniques, this research identifies key factors like age, BMI, and glucose levels as biomarkers for breast cancer prediction, with artificial neural networks achieving the highest accuracy of 80%. The study aims to aid early diagnosis of breast cancer, benefiting medical practitioners and patients alike.

Min Chen [4] This paper explores using machine learning to predict chronic disease outbreaks in communities, focusing on central China. By combining structured and unstructured hospital data with a new convolutional neural network algorithm, the prediction accuracy reaches 94.8%, surpassing existing methods and addressing the challenge of incomplete medical data.

Lu Men [5] This study introduces a deep learning method for predicting multiple diseases based on patients' clinical visit records, outperforming traditional methods. By incorporating time-aware and attention-based mechanisms, it enhances accuracy, aiding physicians in diagnoses and improving healthcare service quality.

K. Arumugam [6] This research focuses on using data mining techniques to predict heart disease in diabetic individuals, highlighting the effectiveness of decision tree models over other algorithms. By fine-tuning the decision tree model, it aims to improve accuracy in forecasting the likelihood of heart disease, addressing a gap in predictive capabilities for this specific population.

Dhiraj Dahiwade [7] This study focuses on predicting diseases based on symptoms using data mining techniques like K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN). Achieving an accuracy of 84.5%, CNN outperforms KNN, offering faster processing and lower memory requirements, providing risk assessments for general diseases based on individual habits and checkup information.

Mauro F. Pinto [8] This framework uses machine learning to predict the progression of Multiple Sclerosis, focusing on the development of secondary progressive courses and severity of disability. It aims to provide early insights into the disease's evolution, aiding in treatment planning and management.

Laxmi Deepthi Gopisetti [9] This study develops a system to predict multiple diseases using various machine learning algorithms and datasets, aiming to improve early detection and diagnosis, potentially saving lives. By comparing classification algorithms like K-Nearest Neighbor and Decision Tree, the research seeks to identify the most accurate method for disease forecasting via a user-friendly web application.

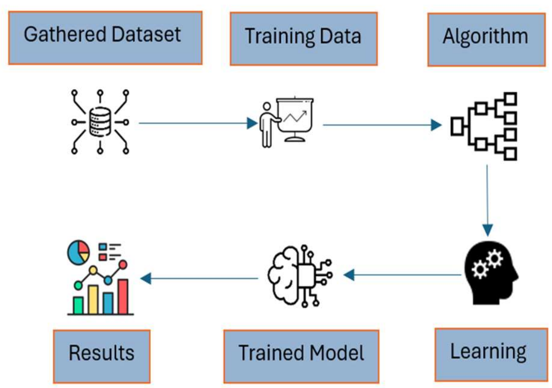
Keniya, Rinkal [10] We made an Excel file from a free dataset, listing symptoms for 230 diseases. There were over 1000 different symptoms in total. We used the symptoms, age, and gender of a person as inputs for different computer programs that learn patterns, called machine learning algorithms.

Alanazi, Rayan [11] We split the datasets into two parts: one for training and one for testing, and we clean up the collected data. Then, we use two machine learning methods, CNN and KNN, to teach the training data. After repeating this process several times until we reach our goal, we prepare the resulting model for testing its accuracy.

Harshit Jindal [12] In this study, we used three different methods—K nearest neighbors (KNN), Logistic Regression, and Random Forest Classifiers. The main aim of refining this project is to make sure it does its best in predicting the likelihood of heart disease using just one set of data.

III. PROPOSED WORK

Fig.1 shows how we use different computer programs to guess if someone has different diseases. These programs, like Naïve Bayes, RandomForest, Logistic Regression, KNN and ANN, help bridge the gap between patients and doctors. We test how good each program is at guessing by comparing their results. We combine lots of information to make the guesses as accurate as possible. There's also a website where people can type in their symptoms, and the program will guess what disease they might have.



A. WorkFlow of Proposed Work

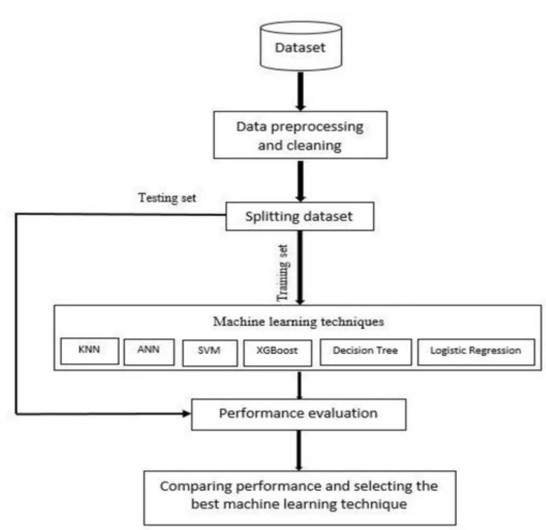


Fig. 1. WorkFlow Design Step 1: Get data from different places like Kaggle for diseases like heart disease, lung cancer, diabetes, breast cancer and parkinson's.
Step 2: Clean up the data by turning words like "male" or "female" into numbers like 0 or 1.
Step 3: Use different algorithms like KNN, Gaussian NB, SVM, Logistic Regression, Random Forest and ANN to make models for each disease.
Step 4: Teach each model with part of the data.
Step 5: Check how good each model is by checking the accuracy.
Step 6: Pick the best model based on its accuracy.

B. Machine Learning Models

1) Gaussian Naïve Bayes

Gaussian Naive Bayes is a basic way to make classifiers. It works by looking at the different features (like symptoms of a disease) and figuring out the probability that those features are related to a certain class (like having the disease or not). It assumes that each feature doesn't depend on any other feature when making this prediction.

Likelihood — Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability — Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

2) K Nearest Neighbor

The K-nearest Neighbors (KNN) algorithm is a tool used in both classification and regression tasks. It works by grouping similar data points together. When a new data point comes in, it's assigned to the category that matches the closest existing data points. The algorithm looks at how similar the new data point is to the ones it already knows about. Since it assumes that similar points are often located near each other, it's often used to group things into categories.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

3) Support Vector Machine

Support Vector Machine (SVM) is a method used for sorting things into groups or predicting values. It's commonly used for putting things into groups in real-world situations. In SVM, each piece of data is shown as a point in space with many dimensions, where each dimension represents a different feature. The goal is to find a line or plane that separates the different groups of data points. This line or plane is called a hyper-plane.

4) Random Forest

Random Forest (RF) works like this: it makes many decision trees using different parts of the data. Each tree gives its own prediction, and then all those predictions are put together to make a final prediction. It's like asking a bunch of people for their opinion, then taking a vote to decide on the answer. This method is called bagging, where we create different groups of data to work with. Each decision tree looks at one group and makes its own prediction. Then, all the predictions are combined, and the most common one is chosen as the final answer. This combining step is called aggregation.

$$Gini = 1 - \sum_{j} p_j^2$$

5) Logistic Regression

Logistic Regression is a way of sorting things into different groups. It works by figuring out the chance of something being in a certain group. We use a function called the Sigmoid Function to calculate this chance. The result will be a number between 0 and 1. If the chance is below a certain point called the threshold, it's put in one group, and if it's above the threshold, it's put in the other group.

$$ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$

$$=> P = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

C. Deep Learning Models

1) Artificial Neural Network (ANN)

An Artificial Neural Network, or ANN, is like a simplified version of how our brains work. It's made up of layers of "neurons" (just like brain cells). These neurons are organized into three main parts: input layer, hidden layers, and output layer.When we train an ANN, we adjust the connections between neurons to make the network better at its task, like predicting outcomes.

ANNs are used for lots of things, like speech recognition, and even predicting trends in data. They're like a simplified version of how our brains process information to solve problems.

$$z = f(b + x \cdot w) = f\left(b + \sum_{i=1}^{n} x_i w_i\right)$$

$$x \in d_{1\times n},\ w \in d_{n\times 1},\ b \in d_{1\times 1},\ z \in d_{1\times 1}$$

D. Streamlit

You can make cool machine learning web apps quickly and easily using a free tool called Streamlit. It's made for people who work with machine learning, so it's simple for them to use. With Streamlit, you don't need to spend weeks learning how to build a web app. Even if you don't know anything about web development, you can make your own online app right away. It's perfect if you're into data science and want to share your models fast and without writing a lot of code.

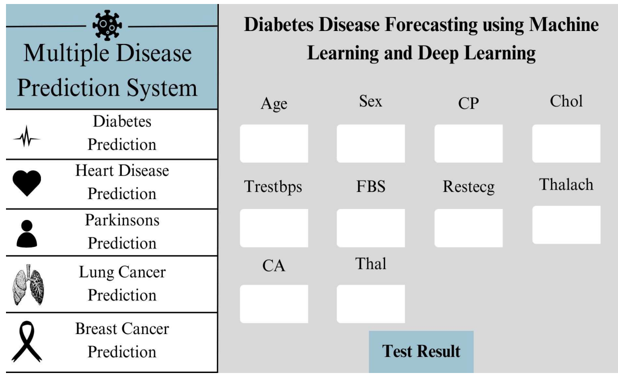
Streamlit has several key features:

- You don't need to know HTML, CSS, or JavaScript.
- You can quickly build impressive machine learning or data science apps in just hours or even minutes, instead of days or months.
- It supports many Python libraries like Pandas, Matplotlib, Seaborn, Plotly, Keras, PyTorch, and SymPy (latex).
- You can create fantastic online applications with less code.
- It simplifies and speeds up computation pipelines with data caching.

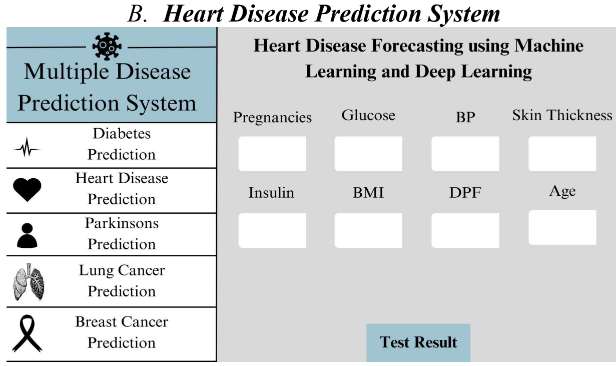
IV. USER INTERFACE

A user-friendly screen is created where people can choose a particular disease and type in their symptoms for prediction.

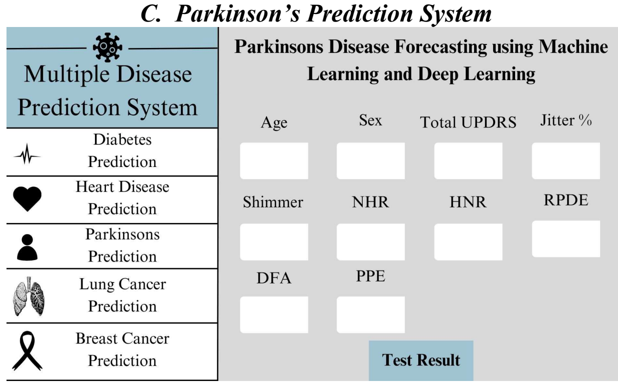
A. Diabetes Prediction System



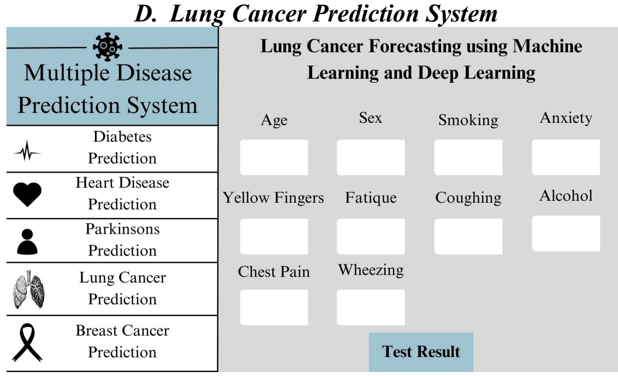
B. ***Heart Disease Prediction System***



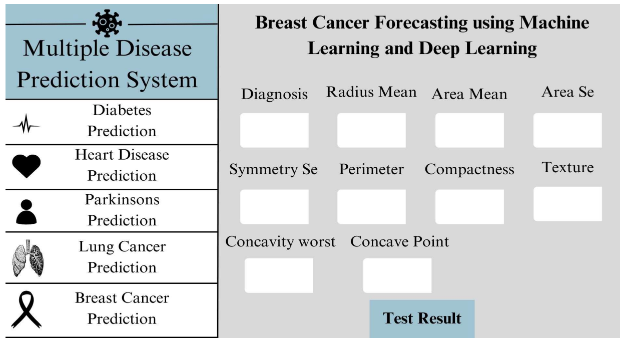
C. Parkinson's Prediction System



D. Lung Cancer Prediction System



E. Breast Cancer Prediction System



V. RESULTS

Following are the accuracy tables for each algorithm for various diseases:

A. Gaussian Naïve Bayes

Table 1: Results obtained using Gaussian Naive Bayes

Disease	Accuracy
Diabetes	84.92
Heart	88.76
Parkinson's	79.83
Lung Cancer	82.50
Breast Cancer	78.6

B. K Nearest Neighbor

TABLE 2: Results obtained using KNN

Disease	Accuracy
Diabetes	79.7
Heart	86.79
Parkinson's	79.86
Lung Cancer	80.37
Breast Cancer	79.81

C. Support Vector Machine

TABLE 3: Results obtained using SVM

Disease	Accuracy
Diabetes	86.80
Heart	80.79
Parkinson's	84.61
Lung Cancer	79.37
Breast Cancer	84.22

D. Random Forest

TABLE 4:* Results *obtained using Random Forest

Disease	Accuracy
Diabetes	89.36
Heart	88.98
Parkinson's	97.58
Lung Cancer	95.64
Breast Cancer	94.75

E. Logistic Regression

TABLE 5: Results obtained using Logistic Regression

Disease	Accuracy
Diabetes	90.78
Heart	83.90
Parkinson's	82.05
Lung Cancer	90.32
Breast Cancer	60.02

F. Artificial Neural Network (ANN)

TABLE 6: Results obtained using ANN

Disease	Accuracy
Diabetes	92.81
Heart	89.12
Parkinson's	90.00
Lung Cancer	94.73
Breast Cancer	93.88

VI. CONCLUSION

In this study, We investigated the Artificial Neural Network (ANN) given the most accurate prediction in the heart and diabetes diseases. Similarly, The Random Forest model (RF) given the promising results of Parkinson’s, Lung Cancer and Breast cancer. Therefore, We conclude that the Random Forest model is best for predicting cancer type diseases.

VII. REFERENCES

[1] A. D. Gunasinghe, A. C. Aponso, and H. Thirimanna, "Early Prediction of Lung Diseases," in 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033668.

[2] A. L. Yadav, K. Soni, and S. Khare, "Heart Diseases Prediction using Machine Learning," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10306469.

[3] M. U. Ghani, T. M. Alam, and F. H. Jaskani, "Comparison of Classification Models for Early Prediction of Breast Cancer," in 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 2019, pp. 1-6, doi: 10.1109/ICIC48496.2019.8966691.

[4] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in IEEE Access, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[5] Lu Men, Noyan Ilk, Xinlin Tang, and Yuan Liu, "Multi-disease prediction using LSTM recurrent neural networks," Expert Systems with Applications, vol. 177, pp. 114905, 2021, ISSN 0957-4174, doi: 10.1016/j.eswa.2021.114905.

[6] K. Arumugam, Mohd Naved, Priyanka P. Shinde, Orlando Leiva-Chauca, Antonio Huaman-Osorio, and Tatiana Gonzales-Yanac, "Multiple disease prediction using Machine learning algorithms," Materials Today: Proceedings, vol. 80, pt. 3, pp. 3682-3685, 2023, ISSN 2214-7853.

[7] D. Dahiwade, G. Patle, and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.

[8] M. F. Pinto, H. Oliveira, S. Batista, et al., "Prediction of disease progression and outcomes in multiple sclerosis with machine learning," Sci Rep, vol. 10, article 21038, 2020.[Online].

[9] L. D. Gopisetti, S. K. L. Kummera, S. R. Pattamsetti, S. Kuna, N. Parsi, and H. P. Kodali, "Multiple Disease Prediction System using Machine Learning and Streamlit," in 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 923-931, doi: 10.1109/ICSSIT55814.2023.10060903.

[10] R. Keniya, A. Khakharia, V. Shah, V. Gada, R. Manjalkar, T. Thaker, M. Warang, and N. Mehendale, "Disease Prediction From Various Symptoms Using Machine Learning," July 27, 2020.

[11] Alanazi, Rayan. "Identification and prediction of chronic diseases using ma chine learning approach." Journal of Healthcare Engineering 2022 (2022).

[12] H. Jindal et al., "Heart disease prediction using machine learning algorithms," in IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, IOP Publishing, 2021.

[13] D. E. Salhi, A. Tari, and M. Kechadi, "Using machine learning for heart disease prediction," in International Conference on Computing Systems and Applications, Springer, Cham, 2020.

[14] Z. Wang et al., "Machine learning-based prediction system for chronic kidney disease using associative classification technique," International Journal of Engineering & Technology, vol. 7, no. 4.36, pp. 1161, 2018.

[15] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," Procedia Computer Science, vol. 165, pp. 292-299, 2019.

[16] A. Kunjir, H. Sawant, and N. F. Shaikh, "Data mining and visualization for prediction of multiple diseases in healthcare," in 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, Andhra Pradesh, India, 2017, pp. 329-334, doi: 10.1109/ICBDACI.2017.8070858.

[17] C.-D. Chang, C.-C. Wang, and B. C. Jiang, "Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors," Expert Systems with Applications, vol. 38, no. 5, pp. 5507-5513, 2011, ISSN 0957-4174.