Disease Prediction System using Machine Learning

A Project Report

Submitted by

Vaibhav Raheja (C056)
Viraj Shah (C064)
Mayank Shetty (C075)
Purav Patel (C098)

Under the Guidance of

Prof. Manisha Tiwari

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY (Integrated)

COMPUTER ENGINEERING

At

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT AND ENGINEERING

October 2021

Specimen B

DECLARATION

I, Vaibhav Raheja Roll No. C056 B. Tech Integrated (Computer Engineering), IX semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs, or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism, and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student:

Name: Vaibhav Raheja

Roll No. C056

Place: Mumbai

Date: 29-10-2021

Specimen B

DECLARATION

I, Viraj Shah Roll No. C064 B. Tech Integrated (Computer Engineering), IX semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs, or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism, and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student:

Name: Viraj Shah

Roll No. C064

Place: Mumbai

Date: 29-10-2021

Specimen B

DECLARATION

I, Mayank Shetty Roll No. C075 B. Tech Integrated (Computer Engineering), IX semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs, or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism, and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student:

Name: Mayank Shetty

Roll No. C075

Place: Mumbai

Date: 29-10-2021

Specimen B

DECLARATION

I, Purav Patel Roll No. C098 B. Tech Integrated (Computer Engineering), IX semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs, or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism, and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student:

Name: Purav Patel

Roll No. C098

Place: Mumbai

Date: 29-10-2021

Specimen C

CERTIFICATE

This is to certify that the project entitled "Disease Prediction System using Machine Learning" is the bonafide work carried out by Vaibhav Raheja of B.Tech Integrated (Computer Engineering), MPSTME (NMIMS), Mumbai, during the IXth semester of the academic year 2021 in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering, Integrated as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

_____

<Name of the Mentor>

Internal Mentor

_____ _____

Examiner 1 Examiner 2

_____

Director

Specimen C

CERTIFICATE

This is to certify that the project entitled "Disease Prediction System using Machine Learning" is the bonafide work carried out by Viraj Shah of B.Tech Integrated (Computer Engineering), MPSTME (NMIMS), Mumbai, during the IXth semester of the academic year 2021 in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering, Integrated as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

_____

<Name of the Mentor>

Internal Mentor

_____ _____

Examiner 1 Examiner 2

_____

Director

Specimen C

CERTIFICATE

This is to certify that the project entitled "Disease Prediction System using Machine Learning" is the bonafide work carried out by Mayank Shetty of B.Tech Integrated (Computer Engineering), MPSTME (NMIMS), Mumbai, during the IXth semester of the academic year 2021 in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering, Integrated as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

_____

<Name of the Mentor>

Internal Mentor

_____ _____

Examiner 1 Examiner 2

_____

Director

Specimen C

CERTIFICATE

This is to certify that the project entitled "Disease Prediction System using Machine Learning" is the bonafide work carried out by Purav Patel of B.Tech Integrated (Computer Engineering), MPSTME (NMIMS), Mumbai, during the IXth semester of the academic year 2021 in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering, Integrated as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

_____

<Name of the Mentor>

Internal Mentor

_____ _____

Examiner 1 Examiner 2

_____

Director

Chapter – 1

1. Introduction

Machine Learning in Disease Prediction/Diagnosis

Machine Learning is a branch of Artificial Intelligence (AI), where the main objective is to give the computer the ability to learn from a provided set of data. The structure of the data is understood, after which the data is fit into models. These models can be successfully utilized by people for any given application where machine learning is required.

Despite being a field within computer science, it radically differs from the traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers for calculations or problem-solving, whereas machine learning algorithms allow for computers to train on data inputs and use statistical analysis to output values that fall within a specific range.

This approach in machine learning facilitates computers in building models from sample data, in order to automate decision-making processes based on data inputs.

Our project will be harnessing the potential of machine learning, in which a model will be trained in identifying various diseases included in our scope, where the output will be boolean values.

An Overview of Expert Systems in Medicine

Expert systems in medicine are defined as systems with the ability to capture and store expert knowledge, facts, and reasoning techniques to assist doctors in diagnosing a patient's condition.

These systems attempt to mimic a doctor's expertise by applying several computational methods to help in decision support and problem solving, by coming up with reasoned conclusions for a patient's illness or condition.

Our project will incorporate the core elements of an expert system by supporting medical experts with their claims, precursing their diagnosis of a chronic disease in their patients using trained machine learning model.

Chapter – 2

2.1. Review of Literature

Papers on machine learning for Chronic Kidney Disease Detection

Sr No.	Paper Name	Authors	Field of Research	Algorithm	Dataset	Paper Summary	Reference
1	Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease	- Rahul Gupta - Nidhi Koli - Niharika Mahor - N. Tejashri	Chronic kidney disease, Prediction, Machine Learning, Decision Tree, Random Forest and Logistic Regression.	- Decision Tree - Random Forest - Logistic Regression	UCI Chronic Kidney Disease Dataset	- The paper evaluates the performance of Decision Tree, Random Forest, and Logistic Regression on the preprocessed and filtered Chronic Kidney Disease Dataset. - Accuracy = Decision Tree (98.48%), Random Forest (94.16%), Logistic Regression (99.24%). - The authors conclude the paper by stating that Logistic Regression provides highest accuracy and recall, while Decision Tree provides highest precision.	International Conference for Emerging Technology (INCET)
2	Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease	- F. M. Javed Mehedi Shamrat - Pronab Ghosh - Mahbubul Hasan Sadek - Md. Aslam Kazi - Shahana Shultana	Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Accuracy.	- Decision Tree - Random Forest - Logistic Regression - K-Nearest Neighbors (KNN)	UCI Chronic Kidney Disease Dataset	- The paper compares the accurate prediction rates of Chronic Kidney Disease using Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN) over the presented dataset. - Accuracy = Decision Tree (97.91%), Random Forest (100%), Logistic Regression (100%), KNN (95.82%). - Random Forest takes the most time to predict, and has the best rating in Receiver Operating Characteristic (ROC) curve.	International Conference for Innovation in Technology (INOCON)
3	A Machine Learning Methodology for Diagnosing Chronic Kidney Disease	- Jiongming Qin - Lin Chen - Yuhua Liu - Chuanjun Liu - Changhao Feng - Bin Chen	Chronic kidney disease, machine learning, KNN imputation, integrated model.	- Random Forest - Logistic Regression - K-Nearest Neighbors (KNN) Imputation - Integrated Model (Random Forest + Logistic Regression)	UCI Chronic Kidney Disease Dataset	- The paper proposes an integrated model, which combines Logistic Regression and Random Forest, to accurately diagnose Chronic Kidney Disease. - Accuracy = Logistic Regression (98.95%), Random Forest (99.75%), Integrated Model (99.83%). - The authors conclude by stating that the integrated model has the highest accuracy and can be further perfected by the increase of size and quality of the data, as it currently cannot diagnose the severity of CKD because only two	Institute of Electrical and Electronics Engineers (IEEE)

						categories of data samples exist in the dataset (ckd and no ckd).	

Papers on machine learning for Diabetes Detection

Sr No.	Paper Name	Author	Field of Research	Algorithm	Dataset	Paper Summary	Reference
1	Diabetes Disease Prediction Using Data Mining	-Deeraj Shetty, -Kishor Rit, -Sohail Shaikh, -Nikita Patil	Diabetes, Prediction, Naïve Bayes, KNN.	- Naive Bayes Algorithm - K-Nearest Neighbors	Pima Indians Diabetes Database	-The Paper recommends using a larger dataset for better prediction -The larger dataset has proved to increase the accuracy of the algorithm, therefore working on some more attributes to better diagnose diabetes.	International Conference on Innovations in Information, Embedded and Communication Systems
2	Diabetes Prediction using Machine Learning Algorithms	-Aishwarya Mujumdara -Dr. Vaidehi Vb	Diabetes Mellitus, Big Data Analytics, Healthcare Machine Learning,	- Logistic Regression - LDA - Random Forest - Extra Trees Classifier	Self created Dataset based on Pima Indians Diabetes Database	-In this study, various machine learning algorithms are applied on the dataset and the classification has been done using various algorithms of which Logistic Regression gives highest accuracy of 96%. - Application of pipeline gave Ada Boost classifier as best model with accuracy of 98.8%.	International Conference On Recent Trends In Advanced Computing
3	Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm	-Samrat Kumar Dey -Ashraf Hossain -Md. Mahbubur Rahman	Diabetes, SVM, ANN, Naïve Bayes, Min Max Scaling	- ANN - Support Vector Machine - Naive Bayes Algorithm - KNN	Pima Indians Diabetes Database	- From different machine learning algorithms Artificial Neural Network (ANN) provide us highest accuracy with Min Max Scaling Method on Indian Pima Dataset.	International Conference of Computer and Information Technology

4	A comparison of machine learning algorithms for diabetes prediction	-Jobeda Jamal Khanam, -Simon Y. Foo	Machine learning, Data Mining, K-fold Cross Validation,	- Decision Tree - K-Nearest Neighbors - Random Forest - Naive Bayes - A/B - Linear Regression - Support Vector Machine	Pima Indians Diabetes Database	- The paper evaluates the performance of various algorithms and measures to get a high accuracy model for Diabetes prediction. All models give an average accuracy of 70%. -LR and SVM have 77-78% accuracy in both K-fold and train-test split -Among all the proposed models, the NN with two hidden layers is considered the most efficient and promising for analyzing diabetes with an accuracy rate of approximately 86%	The Korean Institute of Communications and Information Sciences (KICS).

Papers on machine learning for Heart Disease Detection

Sr No.	Paper Name	Author	Field of Research	Algorithm	Dataset	Paper Summary	Reference
1	Implementation of Machine Learning Model to Predict Heart Failure Disease	- Fahd Saleh Alotaibi	Machine learning model, heart failure diagnosis, KNN method.	- Decision Tree - Naïve Bayes - Random Forest - Support Vector Machine - Logistic Regression	UCI Heart Disease Dataset	- The paper aims to improve the heart failure (HF) prediction using the UCI heart disease dataset which is available on the internet. Multiple machine learning approaches were used to predict HF. - Accuracy = Decision Tree (93.19%), Random Forest (89.14%), Logistic Regression (87.36%), SVM (92.30%) & Naïve Bayes (87.27%). - In comparison this study showed significant improvement and higher accuracy than previous work.	International Journal of Advanced Computer Science and Applications (IJACSA)
2	Heart Disease Prediction using Hybrid machine Learning Model	- Dr. M. Kavitha - G. Gnaneswar1 - R. Dinesh - Y. Rohith Sai - R. Sai Suraj	Cleveland Heart Disease Database, Hybrid algorithm, Machine learning	- Decision Tree - Random Forest - Hybrid (Decision Tree + Random Forest)	Cleveland Heart Disease Dataset	- The interface is designed to get the user's input parameter to predict heart disease, for which the authors used a hybrid model of Decision Tree and Random Forest. - Accuracy = Decision Tree (79%), Random Forest	Institute of Electrical and Electronics Engineers (IEEE)

						(81%) & Hybrid Model (88.7%).	
3	Heart Disease Diagnosis using Extreme Learning Based Neural Networks	- Muhammad Fathurachman - Umi Kalsum - Noviyanti Safitri - Chandra Prasetyo Utomo	Heart Disease, Extreme Learning Machine, Medical Diagnosis.	- Artificial Neural Networks - Support Vector Machine - Decision Tree - Extreme Learning Machine	- Cleveland Heart Disease Dataset - Hungarian Institute of Cardiology Dataset - University Hospital Zurich Dataset - Medical Center Long Beach Dataset (All available on UCI Machine Learning Repository)	- The authors divided the experiments into five parts. For each part, we used a different training and testing dataset which had been divided using K-Fold Cross Validation. Then we set the hidden node of our ELM model for predicting heart disease - Accuracy = Decision Tree (75%), ELM (80%), SVM (68%), BP ANN (77%) - From the experimental results and analysis, the authors concluded that the performance of the ELM algorithm tends to be better when compared with SVM, DT and BP ANN. With an average of 83% accuracy, 88% sensitivity and 82% specificity, the application of ELM algorithm can be an alternative solution to help clinicians predict heart disease.	International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)

Papers on machine learning for Pneumonia:

Sr No.	Paper Name	Author	Field of Research	Algorithm	Dataset	Page Summary	Reference
1	Pneumonia Detection Using CNN based Feature Extraction	-Dimpy Varshni -Rahul Nijhawan -Kartik Thakral -Ankush Mittal -Lucky Agarwal	DensetNet, Deep Convolutional Neural Networks, SVM, Transfer Learning, Random Forest, Naive Bayes, K-nearest	- CNN - SVM	-The dataset used is ChestX-ray14 released by Wang et al. (2017) also publicly available on the Kaggle	- The authors have used a customized model which is a combination of CNN based feature extraction and supervised classifier algorithm. - The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. The C parameter trades off correct classification of training examples	Institute of Electrical and Electronics Engineers (IEEE)

			neighbors, Feature extraction.			against maximization of the decision function's margin. The Area Under the ROC curve (AUC) is an aggregated metric that evaluates how well a logistic regression model classifies positive and negative outcomes at all possible cutoffs. It can range from 0.5 to 1, and the larger it is the better. - Results = SVM (rbf kernel), C = 3.5, gamma = 2e-05 & AUC = 0.7904	
2	Pneumonia Detection using CNN with Implementation in Python	-Muhammad Ardi	Computer science undergraduate student of Universitas Gadjah Mada	- CNN	-Chest X-Ray Images (Pneumonia)	- This paper successfully detected pneumonia using CNN - The model is able to predict pneumonia caused by bacteria pretty well since 232 out of 242 samples are classified correctly -Which gives a accuracy of 95.86%	Pneumonia Detection using CNN with Implementation in Python
3	Detection of Pneumonia using ML & DL in Python	-A Sharma, -M Negi, -A Goyal, -R Jain, -P Nagrath	Neural network, confusion matrix, keras, recall, hyper parameters	- CNN	-Chest X-Ray Images (Pneumonia)	-Detection of diseases with the assistance of computers from various Machine and Deep learning techniques are very beneficial in such places where there is shortage of people who are skilled in techniques like radiology -Validation Accuracy 0.8410 -Validation Loss 0.8395 -Accuracy 0.9806 -Loss 0.0742	IOP Conference Series: Materials Science and Engineering

Machine learning algorithms have not only proven to be an essential tool in the field of mathematics and engineering, but also in the field of medicine as well.

The authors in paper [1] survey various machine learning algorithms and evaluate their accuracy on the Pima Indian Diabetes Database. Among the various models that are surveyed in the paper, Logical Regression and Support Vector Machine provided an accuracy of 78% in both K-fold and test-train split. All models give an average accuracy of 70%. Among all the proposed models, the neural network with two hidden layers is considered the most efficient and promising for analysing diabetes with an accuracy rate of approximately 86%. In paper [3], the authors have used a larger dataset based on the same parameters as the PIMA Indian Diabetes dataset which boosted the accuracy of the Logistic Regression classifier to 96%, alongside the pre-processing of the dataset.

The authors in paper [5] design an interface to get the user's input parameters and predict whether the user is healthy or has a probability of getting a heart disease. Decision Tree and Random Forest are used to make the prediction. With Decision tree the authors achieved an accuracy of 79% and with Random Forest they the accuracy achieved was 81%. Along with these two algorithms the authors

experimented with a Hybrid model which was a combination of Decision Tree and Random Forest. The Hybrid model was the most efficient and had the highest accuracy between the three algorithms with 88.7%.

In paper [7], the authors establish different machine learning algorithms to diagnose chronic kidney disease using the UCI Chronic Kidney Disease dataset using their proposed model. The performance of several high-accuracy algorithms was analyzed, after which only two were shortlisted for an integrated model, namely Logistic Regression and Random Forest. The accuracy of both these algorithms were 98.95% and 99.75% respectively, while the accuracy of the integrated model was 99.83%, which was higher than the individual accuracies of the algorithms used in the integrated model.

The authors of paper [11] utilized a pneumonia dataset from Kaggle, which consisted of chest X-ray images of several patients. The pre-processing of the dataset consisted of resizing the images and creating a function which stores the pre-processed images. The images of the chest X-ray were converted to an array of numbers, after which the authors used Convolutional Neural Network (CNN) to train the model to detect pneumonia, which ultimately provided an accuracy of 98.06%.

Chapter – 3

3.1. Proposed System

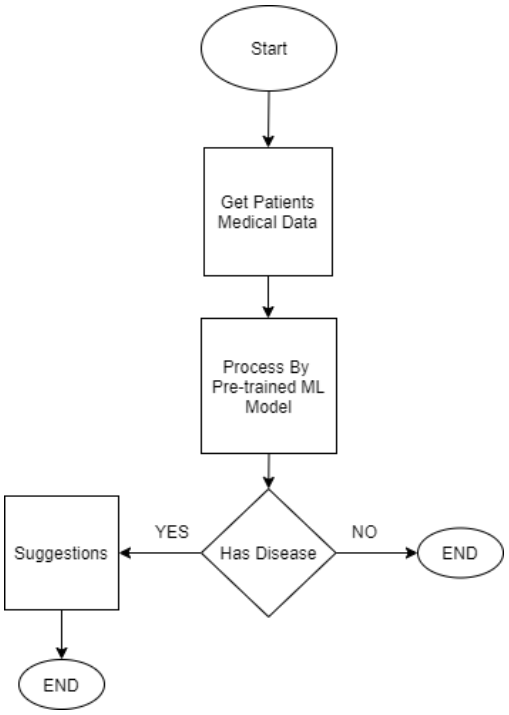

Fig 3.1: Flow Chart for disease detection model

3.1.1. Get Patients Medical Data

This state deals with collection of the input parameters for the pre-trained model. Depending on the chronic disease selected, a specific set of biological parameters of the patients are collected and fed to the pre-trained machine learning model.

3.1.2. Process By Pre-Trained Machine Learning Model

This stage contains a pre-trained model for each chronic disease considered in our project. Appropriate features will be selected for each chronic disease, which will help in further classification of a patient having or not having the chronic disease.

3.1.3. Suggestions

This stage is the output of the pre-trained model for the disease. In the event of a positive result, appropriate suggestions will be provided to the doctors, as there is a high probability of the patient having a chronic disease. If the result is negative, the patient has a low probability of the patient having a chronic disease.

3.2 Architecture Diagram

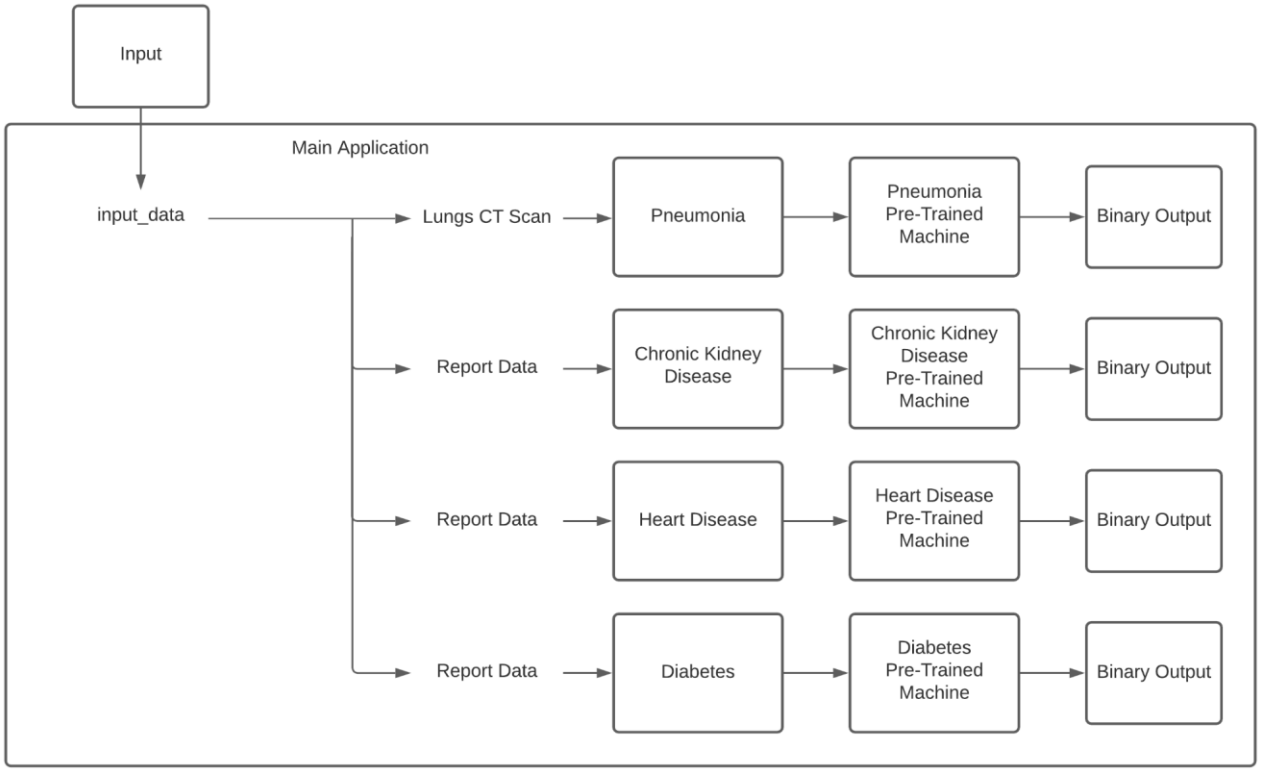

Fig 3.2: Architecture Diagram for Expert System

Disease Detection System

The user will first decide which disease machine they want to use and upon selecting the disease, the data is input into Disease Detection Systems' Pre-Trained Machine model. The machine will then give a binary output, stating whether the patient has the selected disease, or they are healthy.

The Pre-Trained Machines will be trained on a pre-processed dataset. The dataset will go through multiple processing steps before it will be fed to the machine learning models to train them. The data will be checked for any noisy and missing data. K Nearest Neighbor (KNN) will be used for imputing missing values, outlier detection methods will be used to estimate any noise in the data and rapid miner will be used to remove noise in the dataset. The dataset will also be checked for discrepancies, data transformation, discretization and binning techniques will be used.

3.3 UML Diagram

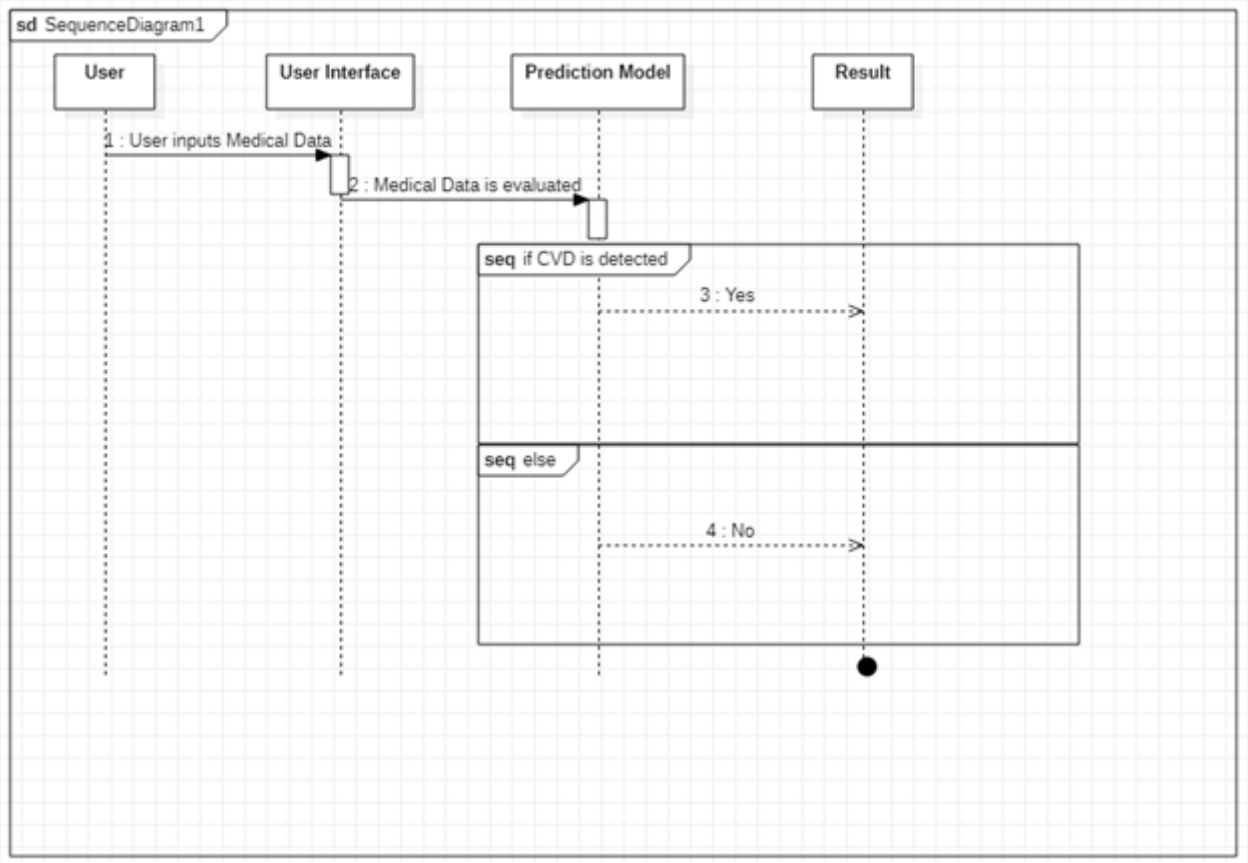

Fig 3.3: Sequence Diagram for CVD Detection

A Sequence Diagram depicting the sequential events of cardiovascular disease (CVD) detection system has been illustrated in Figure 3.3.

The patient's medical data consisting of 13 parameters collected from various tests are entered into the user interface by the medical expert, where the model predicts the probability of the patient having a CVD. The model will rely on values from the data set to predict the probability of a CVD, based on the set of data entered by the user.

Chapter – 4

Implementation

We have implemented a Logistic Regression model for Heart Disease Detection, using the Cleveland Clinic Foundation Heart Disease Dataset, in which the classifier model received an accuracy of 88.5%. The model has been implemented with a basic graphical user interface using Gradio, in which the patient's medical data can be entered using radio buttons, sliders, and text boxes. The classifier model predicts the probability of the patient having a CVD, and the UI conveniently shows the same on the right-hand side. Figure 4.1 shows a sample screenshot of the deployed model's UI.

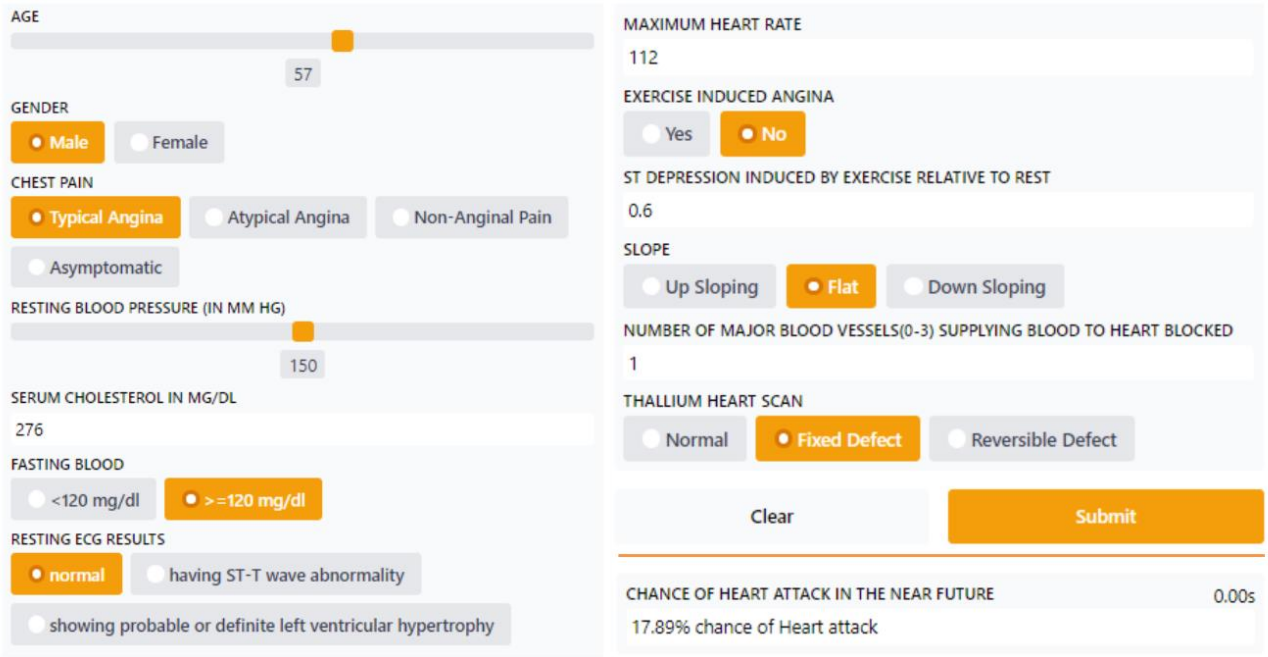

Fig 4.1: UI for CVD detection

Chapter – 5

Plan for Next Semester

Expert systems in medicine are defined as systems with the ability to capture and store expert knowledge, facts, and reasoning techniques to assist doctors in diagnosing a patient's condition. During this semester we have analyzed the problem statement and formed an approach to solve this problem. A few algorithms were shortlisted from the referenced research papers, which would be considered while implementing the chronic disease detection system. We were able to perform the initial implementation for heart disease detection using machine learning during this semester.

For the next semester, we plan to start with the implementation of classifier models for the remaining chronic diseases of the proposed disease detection expert system. Once we have the models ready, we will look into hyperparameter tuning for each of the models. Achieving a high accuracy for each classified model is our goal.

Once the individual models are ready, we will start working on the GUI for our expert system. After working with the GUI, we will be working on collaborating all of the machine learning models into one unified expert system, which will help us accomplish our goal for the project.

After collaborating the models together, there will be a testing phase to check the accuracy of each model again as well as check the system requirements of our deployed model.

Chapter – 6

References

[1] J. J. Khanam, S. Y. Foo, “A comparison of machine learning algorithms for diabetes prediction”, *ICT Express (Feb. 2021)*, 2021, doi: 10.1016/J.ICTE.2021.02.004.

[2] F. G. Woldemichael and S. Menaria, "Prediction of Diabetes Using Data Mining Techniques," *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2018, pp. 414-418, doi: 10.1109/ICOEI.2018.8553959.

[3] A. Mujumdar, V. Vaidehi, “Diabetes Prediction using Machine Learning Algorithms”, *Procedia Computer Science, Volume 165*, 2019, pp. 292-299, doi: 10.1016/J.PROCS.2020.01.047.

[4] S. K. Dey, A. Hossain and M. M. Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm," *2018 21st International Conference of Computer and Information Technology (ICCIT)*, 2018, pp. 1-5, doi: 10.1109/ICCITECHN.2018.8631968.

[5] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj, "Heart Disease Prediction using Hybrid Machine Learning Model," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

[6] F. Alotaibi, “Implementation of Machine Learning Model to Predict Heart Failure Disease”, *International Journal of Advanced Computer Science and Applications (IJACSA), 10(6)*, 2019, doi: 10.14569/IJACSA.2019.0100637.

[7] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," in *IEEE Access*, vol. 8, pp. 20991-21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[8] R. Gupta, N. Koli, N. Mahor and N. Tejashri, "Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease," *2020 International Conference for Emerging Technology (INCET)*, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154147.

[9] F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298026.

[10] T. Ozturk, M. Talo, E. Yildirim, U. Baloglu, O. Yildirim, U. Acharya, “Automated detection of COVID-19 cases using deep neural networks with X-ray images.”, *Computers in biology and medicine vol. 121 (2020)*, 2020, doi: 10.1016/j.compbiomed.2020.103792.

[11] A. Sharma, M. Negi, A. Goyal, R. Jain and P. Nagrath, “Detection of Pneumonia using ML & DL in Python.”, *IOP Conference Series: Materials Science and Engineering* 1022, 2021, doi: 10.1088/1757-899X/1022/1/012066.