**FYP - Final Script**

**Viraj Shah**

**Slide 1 - Title Slide**

Good morning, everyone! Today Vaibhav, Purav, Mayank and I will be presenting our project topic, that is Disease Detection System using Machine Learning.

**Slide 2 - Roadmap**

Our group has created this presentation according to the roadmap, and as you can see we will be starting off with the introduction, problem statement, and the literature review, followed by the details of the proposed system, after which we will be concluding the presentation with the plan of action for the rest of the project, along with the references used in our project.

**Slide 3 - Introduction**

* Chronic diseases are diseases that last for a period of 1 year or more, and are mostly caused by environmental, genetic, and physiological factors.
* These diseases have become a part of everyday life for most people in the present day, impeding them from physically exerting themselves to the fullest.
* In the last 20 years, millions of people have died from the untimely diagnosis of chronic diseases. Several of these chronic diseases were mentioned by the World Health Organization (WHO) in the list of top ten leading causes of death on a global scale, in the year of 2019.
* According to the statistics provided by the WHO, cardiovascular disease was the leading cause of death globally, claiming an estimated 17.9 million lives. Respiratory disease, diabetes, and kidney disease were also a part of the list, following cardiovascular disease in the 4th, 9th, and 10th place respectively.
* Early diagnosis of chronic diseases may save multiple lives, ultimately reducing the mortality rate of these diseases in the long run.
* Now, Mayank will be continuing with the Literature Review.

---------------------------------------------------------------------------------------------------------------------------

**Mayank Shetty**

**Slide 4 - Literature Review**

Good morning! I'm Mayank Shetty, and I'll be talking about our Literature Review. So, after reviewing all the references thoroughly, our group prepared a literature review document in a tabular format. The reference papers were collected for detection of the following chronic diseases - Chronic Kidney Disease, Diabetes, Heart Disease and Pneumonia. For each paper, we have selectively analysed the field of research, the algorithms implemented, and the dataset used for detecting the specified disease. Using this information, we have successfully analysed the performance and efficiency of the algorithms used for detecting the specified disease.

**Slide 5 - Proposed System**

Here’s a simplified representation of our proposed system - We begin by collecting the medical data of patients, which will be processed by a pre-trained machine learning model. If a patient is found to have any disease that our machine is successfully able to detect, the medical expert using our system will be provided with suggestions for diagnosing that particular disease. The following datasets will be used for detecting the corresponding chronic diseases.

**Slide 7 [HALF] - Proposed Algorithm #1 (Chronic Kidney Disease)**

For detection of Chronic Kidney Disease, we have selected Logistic Regression and Random Forest as potential algorithms. Logistic Regression creates a regression model that distinguishes between several samples and provides outputs in probabilistic values, which lie between 0 and 1 . As for Random Forest, it is a supervised machine learning algorithm which aggregates several decisions to make a single decision. Both models provide high-accuracy and high-performance, which is absolutely perfect for our system. Both these models can also be combined to form an Integrated Model, which has been implemented in one of our references, where it offered a higher chronic kidney disease detection accuracy than both the models individually. Now, Viraj will go forward with the next half of the slide.

**Viraj Shah**

**Slide 7 [HALF] - Proposed Algorithm #1 (Diabetes)**

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. KNN receives the highest accuracy in detection of diabetes on PIMA dataset. Hence it can be used for a highly accurate ML model to detect diabetes in patients as well as allow an opportunity for improving the accuracy of the model.

Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or False, etc. but instead it gives the probabilistic values which lie between 0 and 1. LR receives high accuracy in detection of diabetes on PIMA dataset. Hence it can be used for a highly accurate ML model to detect diabetes in patients.

---------------------------------------------------------------------------------------------------------------------------

**Slide 10 - Implementation**

*\*SCRIPT NOT NEEDED\**

We have implemented a random forest classifier on the UCI heart disease. After hyper parameter tuning we have received an accuracy of XX %. We intend to further explore other algorithms to improve accuracy.

**Slide 11 - Implementation Schedule**

*\*SCRIPT NOT NEEDED\**

**Vaibhav Raheja**

**Slide 8 [HALF] - Proposed Algorithm #2 (Pneumonia)**

For detection of Pneumonia ,

* We have selected CNN and SVM as potential algorithms. The core concept of CNN is to use convolution of image and filters to generate invariant features which are passed onto the next layer.
* As for SVM, it is a supervised machine learning model that uses classification algorithms for two-group classification problems. Both models provide high-accuracy and high-performance, which is perfect for our system.
* Both these models can also be combined to form an Integrated Model, which has been implemented in one of our references, where it offered a higher pneumonia detection accuracy than both the models individually.

**Slide 9 - Architecture Diagram**

x

---------------------------------------------------------------------------------------------------------------------------

**Slide 12 - Plan for Next Semester**

1. Maam for we plan to start with the implementation of classifier models for the remaining chronic diseases of the proposed system
2. Once we have the models ready,we will start working on the GUI after working with the GUI, we will be working on collaborating all of the machine learning models into one unified system
3. After collaborating the models together, there will be a testing phase to check the accuracy of each model and check the system requirements of our deployed model.

**Slide 13 - Plan of Action**

x

**Slide 14 - References**

These are our referenced implementation papers, of which we've performed a Literature Review. We believe that the implementation strategy and the results of these references will contribute to the overall success of our project.