Nayan: *Take care of your body. It's the only place you have to live*

Sohan: *Your body is your most priceless possession. Take care of it*

Nayan: Good morning, to our respected teacher. It’s our pleasure to being here and thank you all for being here today. My name is Mehadi Hasan, reg. no: 46341, exam roll:180696 along side with my friend.

Sohan: Sohanur Rahman, reg. no: 47939, exam roll:180703 and we were supervised by our respected teacher Professor Dr. Israt Jahan mam. We would like to welcome you to this presentation on “**A Comparative Study of Machine Learning for Predicting Multiple Diseases**”.

2

Nayan: This is the out line of our today’s presentation. First of all we will give you an overview of the project, then comes motivation, then objectives and then problem definition. After problem definition we will talk about existing system, then show some related research, and then our proposed system and the result we have got from our system. And finally the future scope of our project and conclusion.

Sohan:

Key challenges in multiple disease prediction include:

1. Data Complexity: Integrating and analyzing diverse data types, including structured and unstructured data, from various sources such as electronic health records, medical imaging, genetic profiles, and wearable devices.
2. Disease Interactions: Capturing the complex interactions and dependencies between different diseases, considering that the presence of one disease may affect the occurrence or progression of others.
3. Imbalanced Data: Handling imbalanced datasets where the occurrence of certain diseases may be rare, leading to skewed distributions that can affect the performance of machine learning models.
4. Feature Selection: Identifying the most relevant and informative features from a wide range of potential predictors, including demographic factors, medical history, clinical measurements, and genetic markers.
5. Model Interpretability: Ensuring the interpretability and explainability of the machine learning models used for prediction to gain trust and acceptance from healthcare professionals and patients.
6. Scalability and Real-time Prediction: Developing models that can handle large-scale datasets efficiently and provide real-time predictions to support timely interventions and decision-making.

DATA preprocessing:

1. Data Cleaning: This involves handling missing values, dealing with outliers, and resolving inconsistencies in the data. Missing values can be imputed or removed based on the specific context of the data. Outliers, which are extreme values that deviate significantly from the normal distribution, can be detected and handled appropriately.
2. Data Integration: In real-world scenarios, data may be collected from multiple sources, leading to data integration challenges. This task involves combining data from different sources and resolving inconsistencies in attribute names, data formats, and values.
3. Data Transformation: Data often needs to be transformed to meet the assumptions and requirements of specific algorithms. Common transformations include normalization, which scales the data to a standard range, and encoding categorical variables into numerical representations suitable for modeling.
4. Data Reduction: In cases where the dataset is large or contains redundant information, data reduction techniques can be applied. Dimensionality reduction methods, such as Principal Component Analysis (PCA) or feature selection algorithms, can be used to reduce the number of variables while preserving important information.

Nayan: 2,6,7,8,9,10,12,13, 19

Sohan: 3,4,5,11,14, 15,20