Final Project – Deliverable I Report

Course: CSCI 4750/5750 – Machine Learning
Deliverable: Final Project – Deliverable I
Student Name: Navyasree Chenchu and Shri Gouri Pinjarla
Major: Bioinformatics and computational Biology
Group No: 12
Date: April 13, 2025

Preferred Method for Communication: We would like to go with Discord , Mail and WhatsApp.

Prior Collaboration Experience & Concerns: We both already worked together in Bioinformatics-02 in same Project.

Project Selection: We would like to choose the Option 1 - ML application in healthcare

Introduction :

After reading about Warfarin and after some online research, we discovered that it is an anticoagulant, which essentially means that it thins the blood. It is an oral medication whose primary purpose is to prevent and treat blood clots in the body.

Clots can be extremely harmful because they prevent blood flow to vital sections of the body. For example, if a clot prevents blood from reaching the heart, it can result in a heart attack; if it prevents blood from reaching the brain, it can cause a stroke. Warfarin reduces this risk by thinning the blood and making it more difficult for clots to form.

One interesting fact we discovered is that Warfarin dosage varies by individual. It varies based on the individual's condition and frequent blood test results. Doctors frequently check something called the INR (International Normalized Ratio) to see how well the blood clots, and modify the Warfarin dosage accordingly.

This makes Warfarin dosing an ideal use case for machine learning because so many patient characteristics, such as age, weight, genetics, and medical history, influence how much Warfarin should be taken. Building an ML model can help estimate the appropriate dose more accurately, perhaps making treatments safer and more tailored.

ML End-to-End workflow to address this problem :

1. Understanding the problem:

First, we should understand the problem we trying to solve. In this case, the goal is to predict the right daily Warfarin dosage for a patient based on their medical data. This is a regression test because we are predicting a number (dosage) instead of a category.

2. Collecting Data
Next, we'll need data to train the model. We discovered that there exist public databases, like the IWPC dataset, that provide information such:
Age
Weight
Height
Race
Gender
Other medicines
Genetic information (such as VKORC1 and CYP2C9 genes)
Warfarin Dosage

3. Cleaning the data.
This part is about preparing the data so that the model could benefit from it. We plan to:
Fill in or remove the missing values.
Convert text or category fields to numbers (such as encoding race or gender).
Scale numerical data (like weight and height) to make sure that everything is in the same range.

4. Splitting the data
After cleaning, we'll split the dataset into two parts.
Training set (maybe 80%)—this is what the model learns from.
Test set (20%) – this is how we check if the model performs well.

5. Choosing a Model
Since this is a regression problem, we'll try out a few machine learning models like:
Linear Regression (basic yet useful as a starting point).
Random Forest Regressor (Keeps Complex Patterns)
Gradient Boosting (often gives good outcomes)
We'll compare them to decide which one best fits our data.

6. Training the Model
We give the training data into our selected model(s), who attempts to learn patterns between the input features and the right dosage for them. This process may take a little time, depending on the model.

7. Evaluate the Model
We'll then run our model against the test set to verify how well it predicts fresh,

previously unseen data. We will utilize measures such as:
Mean absolute error (MAE)
Root Mean Squared Error(RMSE)
R^2 Score (a measurement of how effectively a model explains dose variation)

8. Improving the model
If it's not accurate enough, we could:
Try different options.
Tune the model parameters (hyperparameter tuning).
Try a new model entirely.

9. If we choose to Deployment
If we have time, we'd like to build a simple web interface with tools like Streamlit or Flask. This would allow a doctor or user to enter patient data and get the predicted dose immediately.

Reference :

https://pmc.ncbi.nlm.nih.gov/articles/PMC11487003/

https://my.clevelandclinic.org/health/treatments/16182-warfarin-a-blood-thinning-drug-what-you-need-to-know-

Zoom Link :

Share Link

https://slu.zoom.us/rec/share/R36b0RycstOTyAuMoxLndCmD6o5roVQV7FakvpA-Xoe1Bjd2QeOtjROA3kILLKS3.BfsNuQCCTBIRUe_W

Passcode: 1Av6%J@b