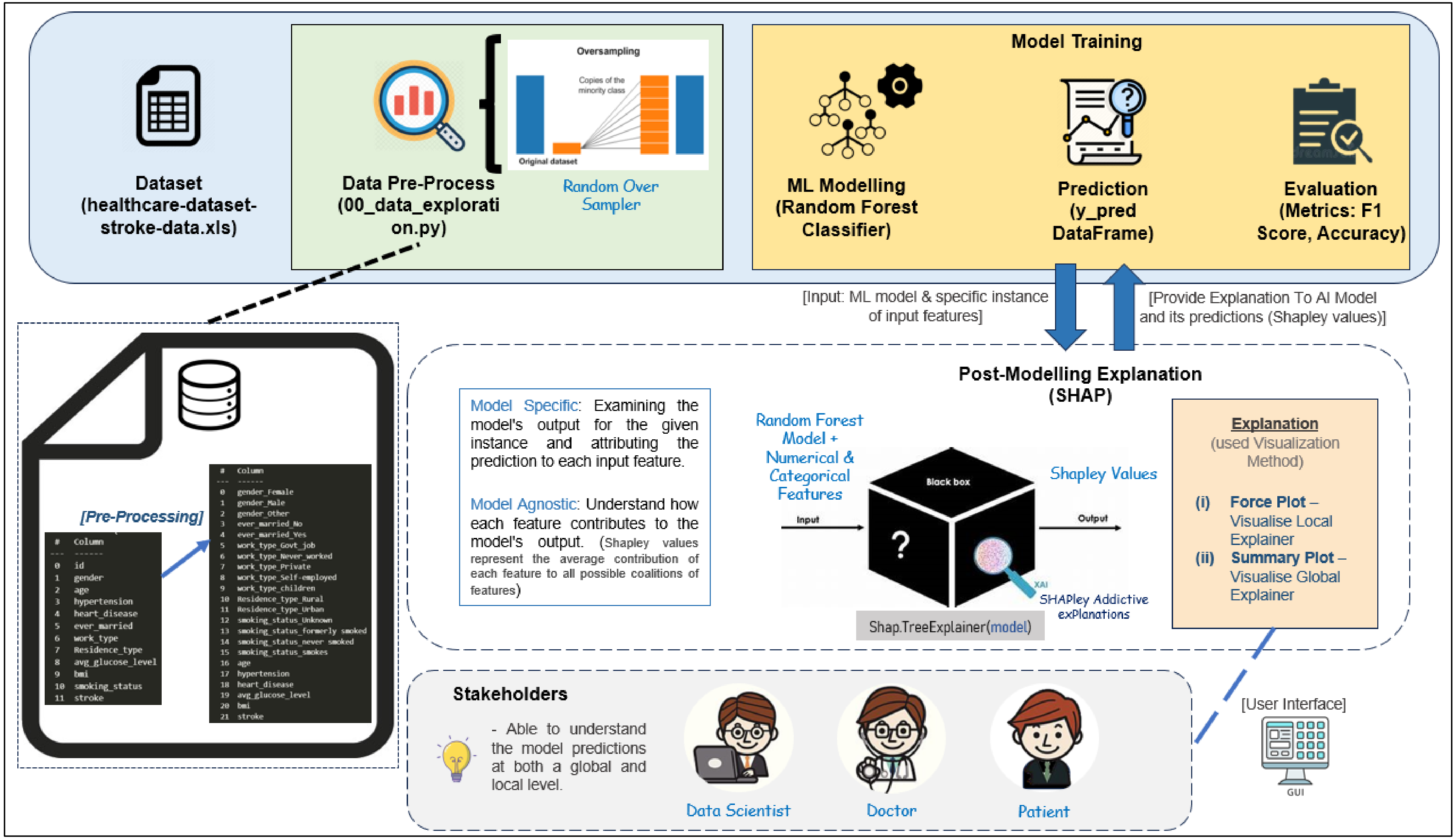
Process Flow :

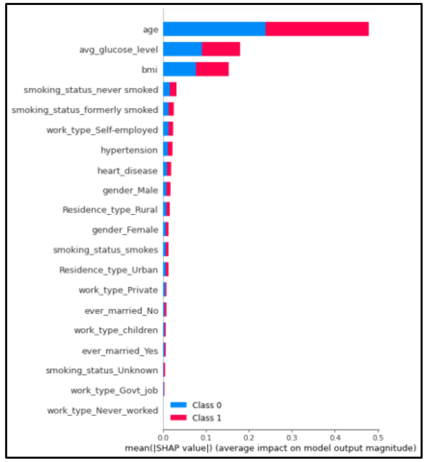


Outcome :

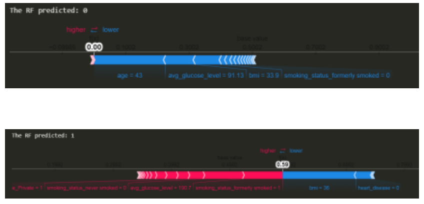
```
✓ rf = RandomForestClassifier()

F1 Score 0.5312473277101878
Accuracy 0.9422700587084148
```

The visualised global explainer using summary plot is as below:

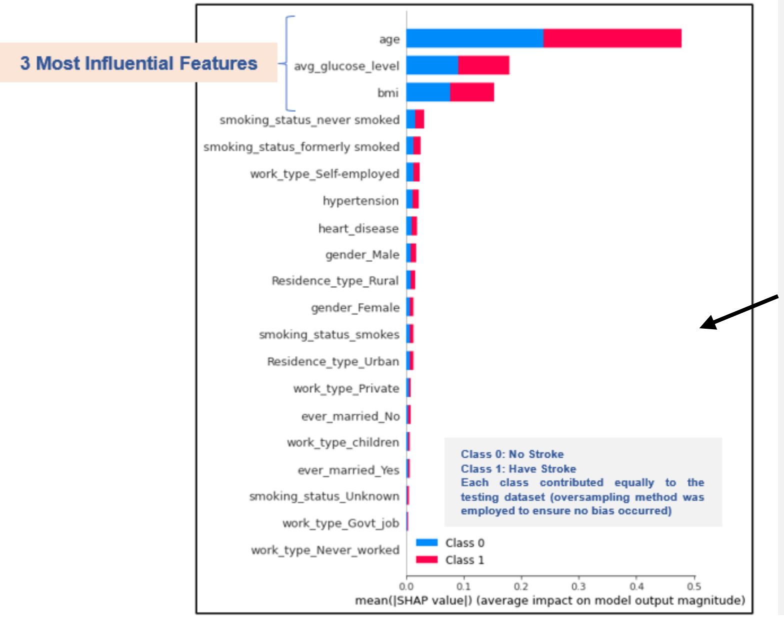


The visualised local explainer using force plot is as below:



From the process flow diagram above, we can get visualized explanations through the SHAP method for the random forest classifier model, utilizing the prepared dataset for stroke prediction. Our model is seemed to have an averagely good F1 score and high accuracy in traditional AI modeling. However, the incorporation of Explainable AI (XAI) in modeling is crucial as users not only require information on the accuracy of the model and its predictions but also need to understand why the model makes these predictions. Mere statement about high accuracy is not enough because most of the time users will have trouble to comprehend the decision and doubt the decision-making process. Relying on blind trust without explanations is not effective in gaining user confidence in the model's decisions. At this moment, XAI model can help to provide transparency to the model and its decisions.

Explanation of the Plotted Graphs:

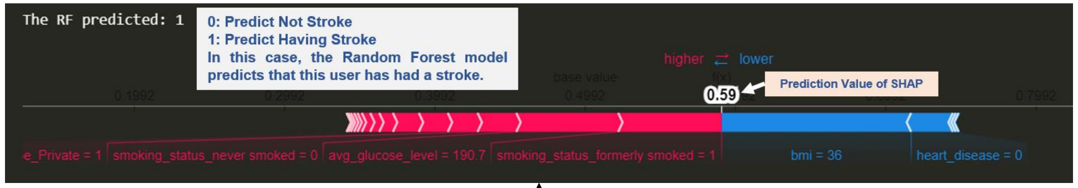


As SHAP provides insights into feature importance and how each feature contributes to the model's predictions, a model that is predisposed towards the majority class may fail to discern the intricacies of the minority class. Therefore, RandomOverSampler was employed to address class imbalance in our dataset. This is the reason why the number of predictions for class 0 (not having a stroke) and class 1 (having a stroke) is the same, due to having a balanced dataset.

Summary Plot gives an overview of feature importance across the dataset. The length of the bar indicates how much influence the feature has on the prediction. A longer length indicates a higher influence of that feature.

The features are ordered by how much they influenced the model's prediction. The x-axis represents the average of the absolute SHAP value of each feature. For this example, "Age" is the most important feature, followed by "Average Glucose Level" and "BMI".

By comparing these insights with a general understanding of the problem, users can trust that the model is intuitive and making the right decisions.



Force Plot Visualizes the impact of each feature on a specific prediction. All of the feature values, including smoking_status_never smoked, avg_glucose_level, bmi, and more, contribute to the prediction value of 0.59.

SHAP plots the top most influential features for the sample under study. Features in red indicate a positive influence as they drag the prediction value closer to 1 (increase the output raw value), while features in blue have the opposite effect (decrease the model output value).

Three Primary Stakeholders in XAI:

Data Scientist

Role in XAI:
- The person in charge of developing machine learning models, training them, and ensure models work just right. Generally, they are concerned about how accurate, efficient, and well the model works.

- Use XAI tools to learn how models makes such decisions decide. This helps them fix bugs, make models better, and find the best settings by finding problems like biases, overfitting, or poor performance on certain inputs.

What SHAP can provide answers to for data scientist:
- Interpretability: How each feature contributes to the model's predictions?

- Features Importance: Which features have the most impact on the model's predictions for feature selection and model improvement?

Doctor

Role in XAI:
- Domain experts or professionals in a specific field, who work with the AI system and make decisions based on its outputs.

- Use XAI tools to gain insights into model predictions, helping themselves validate results and make informed decisions based on the model's outputs.

What SHAP can provide answers to for doctor:
- Explanation for Predictions: Why a model predicted a particular patient to be at higher risk for a certain condition?

- Clinical Interpretability: Which patient features contribute most to a prediction?

Patient

Role in XAI:
- The individual whose provides feedback on the development of AI systems to help creating more patient-centric models, as well as becomes new data which is valuable for refining and improving AI models.

- Use XAI tools to gain insights into why a particular treatment option is recommended, and also to know the factors considered by AI models, ensuring that decisions align with their personal health goals and values.

What SHAP can provide for patient:
- Trust and Transparency: More willing to accept and follow recommendations if patients understand the rationale behind them.

- Personalised Insight: Telling them how their individual characteristics impact predictions.

Two sample scenarios are presented to convey the messages of the global explanation in the summary plot and the local explanation in the force plot to the three stakeholders.

Scenario 1:

Doctor, I've analyzed the stroke prediction model using SHAP, and I wanted to share the insights from the summary plot with you.

Sure, go ahead. What does the summary plot tell us?

The summary plot ranks the features by their average impact on the model's predictions. The most influential feature, according to the model, is “age," followed by “average glucose Level" and “bmi”.

Interesting. How does this information help us?

Well, it means that, according to the model, a person's age is the most significant factor in predicting stroke risk. Higher age tends to increase the model's prediction of stroke. Similarly, elevated levels of glucose and BMI also contribute positively to the prediction.

That aligns with our general understanding. Older age and health indicators like glucose levels and BMI are known risk factors for strokes.

Exactly. Now, the length of each bar in the plot represents the magnitude of the impact. Longer bars mean higher influence.

So, if we see a long bar for a certain feature, we know that changes in that feature strongly affect the prediction.

Page 1

Precisely. This insight helps us interpret the model's decisions and understand which factors it considers most important.

That's valuable. It gives us a clearer picture of how the model views different features in relation to stroke prediction.

And this is just the global explanation. If we need more granular details, we can delve into individual predictions using local explanations like force plots.

Excellent. This information will certainly enhance our understanding and confidence in the model's predictions.

I'm glad you find it helpful. It's crucial for us to not only have accurate predictions but also to understand the rationale behind them.

Agreed. This approach adds a layer of interpretability that's essential, especially in healthcare decision-making. Thanks for your thorough analysis.

You're welcome, Doctor

Page 2

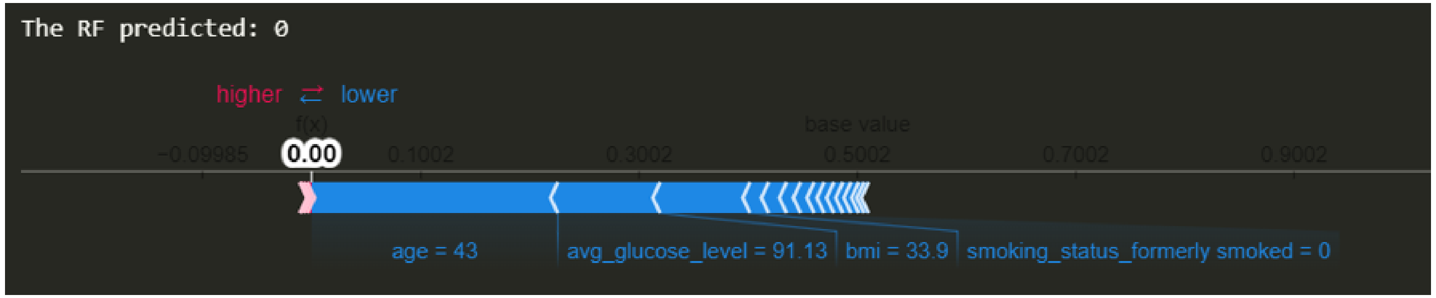
Scenario 2:

Patient 1 (ID 2813):

```
explainer = shap.TreeExplainer(rf)
# Calculate shapley values for test data
start_index = 1
end_index = 2
shap_values = explainer.shap_values(X_test[start_index:end_index])
X_test[start_index:end_index]
```

✓ 0.0s

	gender_Female	gender_Male	gender_Other	ever_married_No	ever_married_Yes
2813	0	1	0	0	1

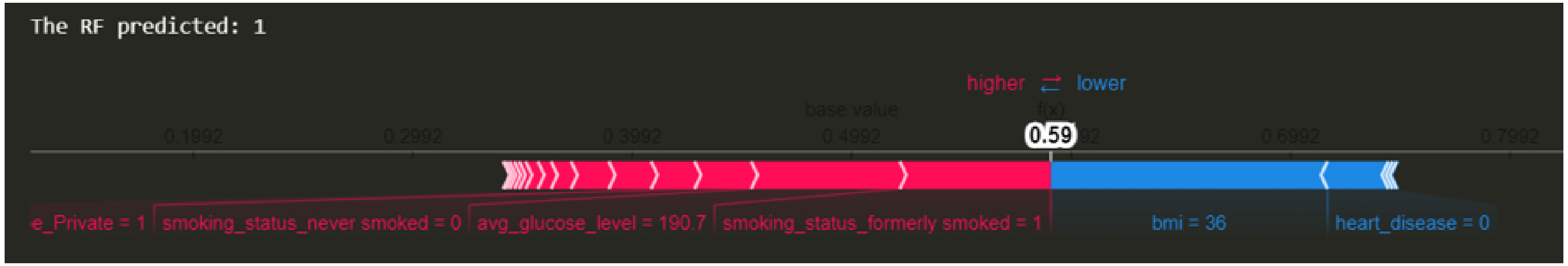


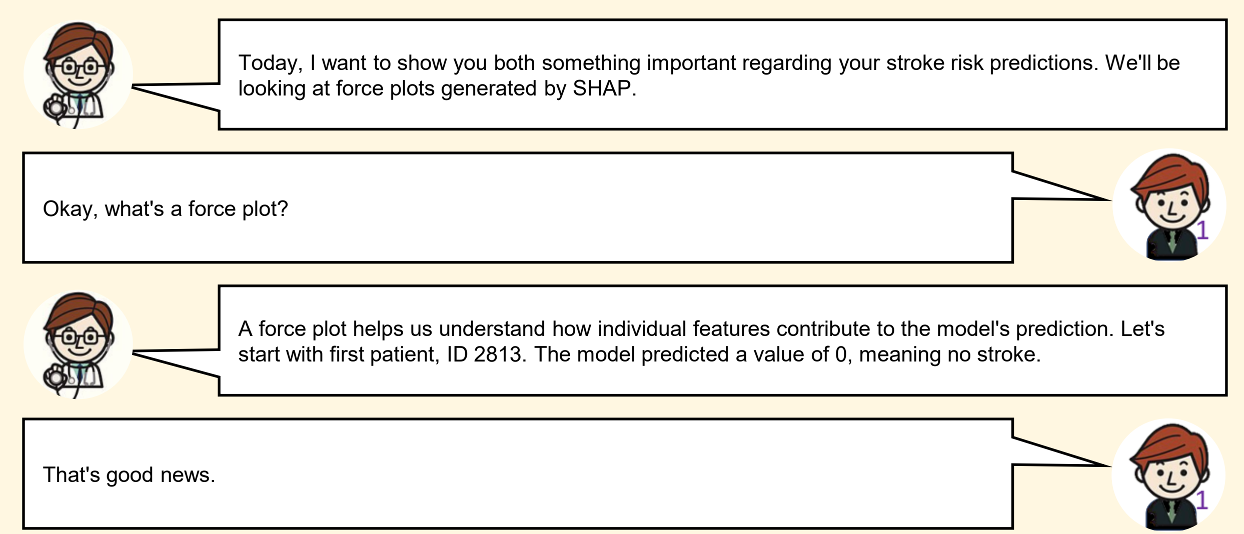
Patient 2 (ID 293):

```
explainer = shap.TreeExplainer(rf)
# Calculate shapley values for test data
start_index = 9
end_index = 10
shap_values = explainer.shap_values(X_test[start_index:end_index])
X_test[start_index:end_index]
```

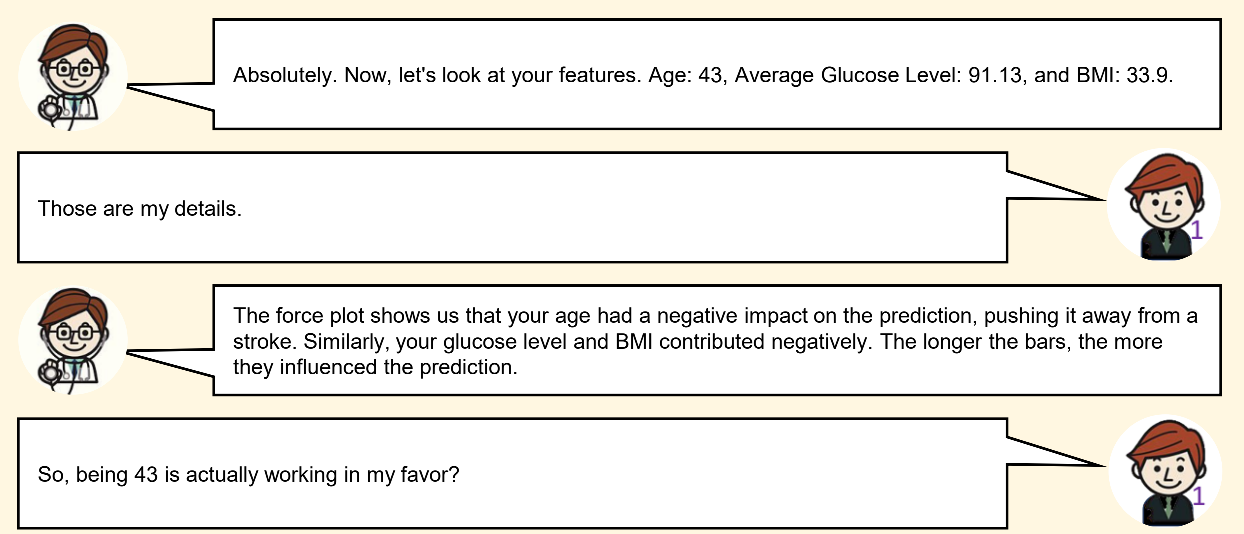
✓ 0.0s

	gender_Female	gender_Male	gender_Other	ever_married_No	ever_married_Yes
293	0	1	0	0	1





Page 1



Page 2

Scenario 2:

Yes, according to the model. Now, let's move to second patient, ID 293. The model predicted a value of 0.59, indicating a higher stroke risk.

That's concerning.

Don't worry; let's understand why. Your most influential features, in order, are BMI: 36, smoking history, and an average glucose level of 190.7.

That's higher than normal.

Indeed. The force plot illustrates that these factors contributed positively to the prediction. In particular, having a BMI of 36 had a significant impact, pushing the prediction towards a higher risk.

So, my weight is a big factor?

According to the model, yes. It's crucial for us to understand these details. Now, you both can see how individual features affect your predictions.

It is insightful. It helps me see where I might need to make some changes.

That's the idea. Remember, these are model predictions, and we use them alongside our medical knowledge. It's a tool to assist us in understanding your health risks better.

Page 3

Yes, it makes the prediction more understandable. I appreciate the clarity.

I'm glad you find it helpful. Understanding these details enables you both to make informed decisions about your health.

Page 4

Discussion:

Do these two visualized explanations clearly explain the prediction of Random Forest model?

- From my point of view, the lack of detailed explanation regarding how the value of a particular feature influences the SHAP prediction value makes the presentation unconvincing and not transparent enough, given only these two types of plotted graphs. While the absolute SHAP value provides information on how much a single feature affects the prediction, it does not explicitly give the answer like, High values of the Latitude variable have a high negative contribution to the prediction, while low values have a high positive contribution’. This crucial information is missing from the given summary plot.