

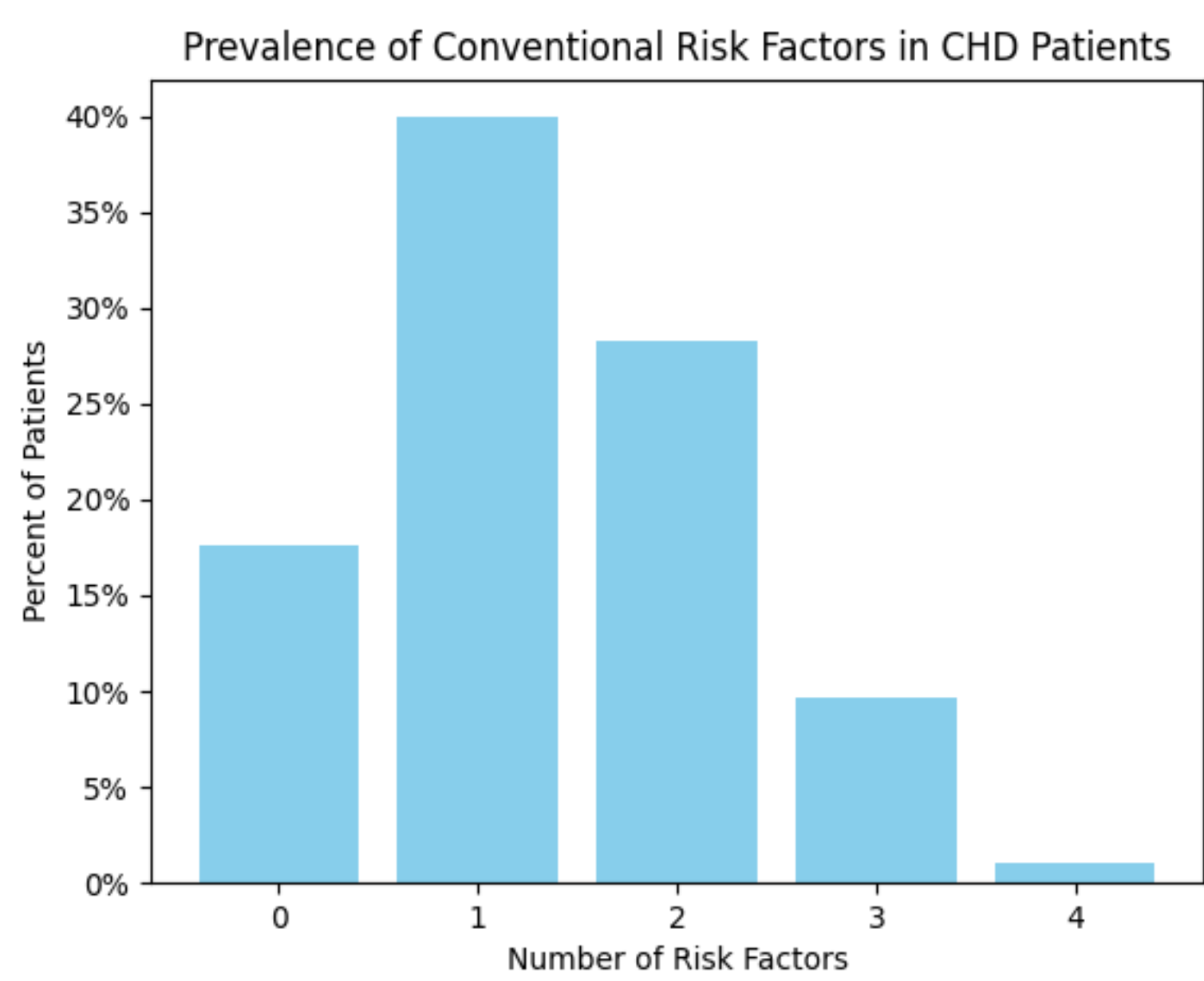
# Analyzing the Correlation Between Heart Disease & Various Risk Factors

Elena Batarseh, Krish Patel, Sean Wilweber

## Introduce problem

- Heart diseases, which are heart conditions such as diseased vessels, structural problems, and blood clots, are one of the leading causes of death worldwide. Early detection and prevention is therefore important.
- The task we chose is to determine the most impactful factors that lead to heart disease.

## Background



Data from [1]

- 4 conventional risk factors: cigarette smoking, diabetes, hyperlipidemia, and hypertension
- Study found that 82% of CHD patients were affected by at least one of the risk factors

## Problem Motivation

Finding the biggest causes of heart disease puts us a step ahead in many ways:

- Prevention focus: Knowing the primary causes allows for targeted public health initiatives to educate people on reducing their risk of developing heart disease
- Policy changes: Research on heart disease causes can direct policy decisions to promote healthier lifestyles, regulate products like tobacco, and more
- Early detection and treatment: By understanding the major factors that lead to heart disease, healthcare professionals would be able to identify high risk individuals and guide them towards lifestyle changes or medication to prevent heart disease

Limitations:

- Lack of control group
- Survival bias
- Patient self-reporting
- Hidden risk factors
- Lacking in prediction

## Approach

- We used a Random Forest model because it’s a powerful machine learning method that combines multiple decision trees to make more accurate predictions, especially if our data does not follow a linear trend. One of its strengths is that it reduces the risk of overfitting by averaging the results of all the trees. This makes it a reliable option for handling complex data sets.
- One of the best parts about Random Forests is that it shows which variables are the most important in determining the outcome. For example, our model ranked predictors like age and cholesterol at the top.

## Approach details

- For our Random Forest, we used 100 trees with the default settings for other parameters like tree depth. This allowed us to capture important patterns in the data without overcomplicating the model.
- Our dataset contained roughly 4000 individual rows, and we split the dataset into 70% for training the data and 30% for the actual testing. This ensured that the model was trained on a substantial portion of the data but still tested on unseen data to validate its performance.

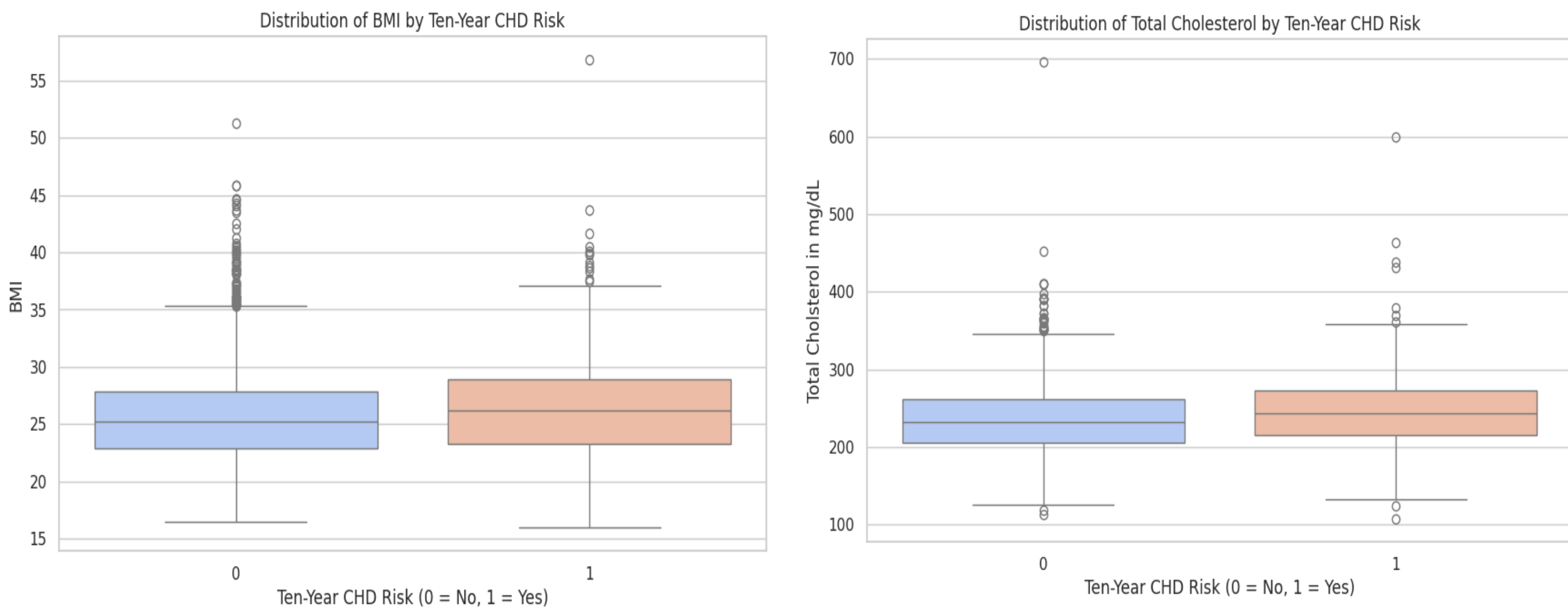
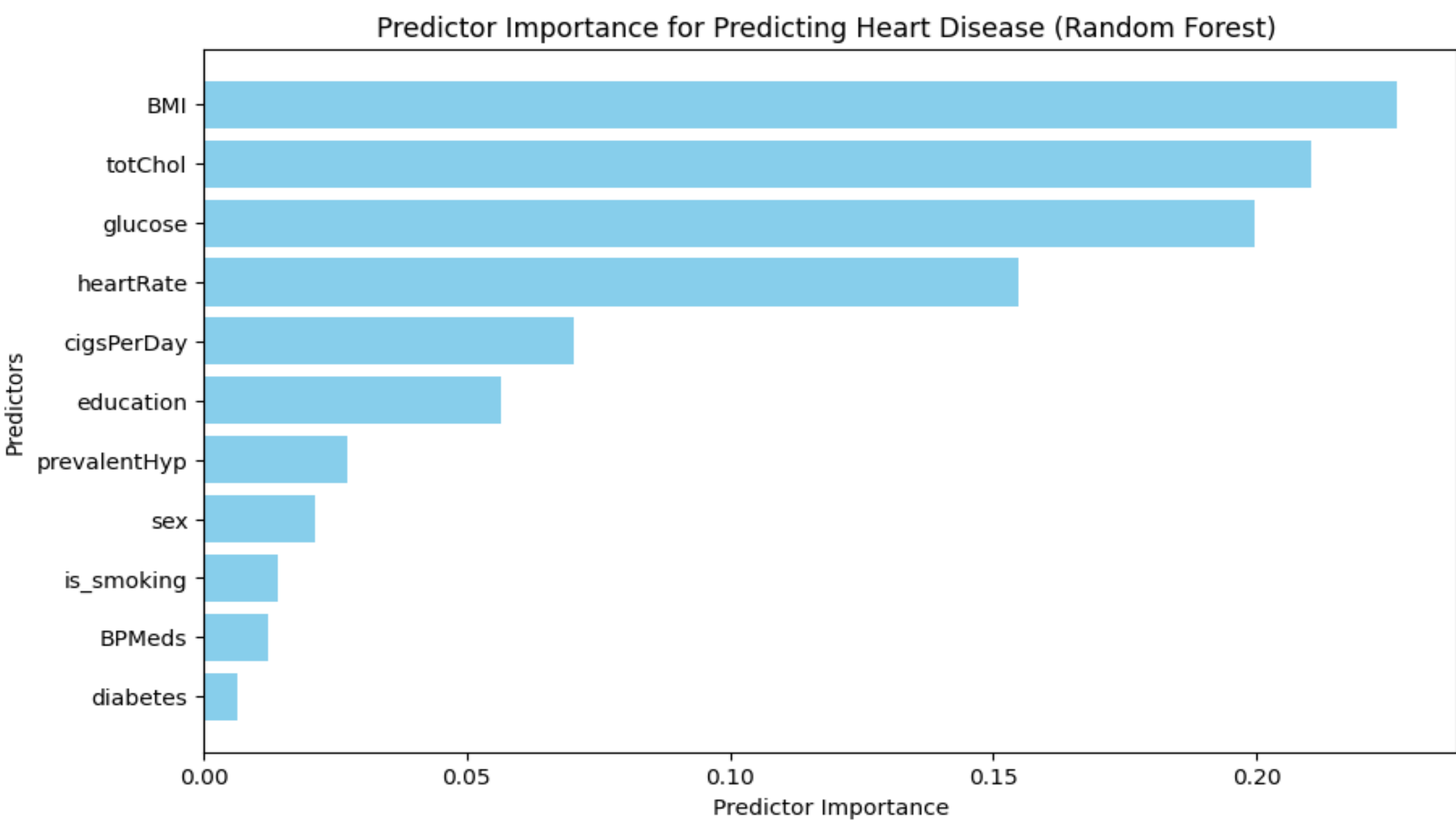
## Evaluation

Models	Accuracy	Classification error	Precision	F-measure	Sensitivity	Specificity
Naive Bayes	75.8	24.2	90.5	84.5	79.8	60.0
Generalized Linear Model	85.1	14.9	88.8	91.6	94.9	20.0
Logistic Regression	82.9	17.1	89.6	90.2	91.1	25.0
Deep Learning	87.4	12.6	90.7	92.6	95	33.3
Decision Tree	85	15.0	86	91.8	98.8	0.0
Random Forest	86.1	13.9	87.1	92.4	98.8	10.0
Gradient Boosted Trees	78.3	21.7	94.1	86.8	80.7	60.0
Support Vector Machine	86.1	13.9	86.1	92.5	100	0.0
VOTE	87.41	12.59	90.2	84.4	-	-
HRFLM (proposed)	88.4	11.6	90.1	90	92.8	82.6

Figure from [2]

- Using random forest, we can handle complex relationships in the data with a high accuracy.
- To improve accuracy further, larger datasets could be used.

## Results



## Conclusions

- In conclusion, we demonstrated that machine learning algorithms such as random forests can predict heart disease well, with BMI and cholesterol being the most important factors.
- BMI (Body Mass Index) is directly related to your health in general, which is why it makes sense that a higher BMI directly correlates to heart disease.
- High cholesterol blocks your arteries leading to reduced blood flow to the heart, explaining why it is also a good predictor of heart disease.
- These insights can help doctors make more informed decisions and improve patient care specifically with heart disease.

## References

[1] Khot UN, Khot MB, Bajzer CT, et al. (2003). "Prevalence of Conventional Risk Factors in Patients With Coronary Heart Disease". *JAMA*; 290(7):898-904  
[2] S. Mohan, C. Thirumalai and G. Srivastava. (2019). "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques". *IEEE*; 10.1109/ACCESS.2019.2923707