



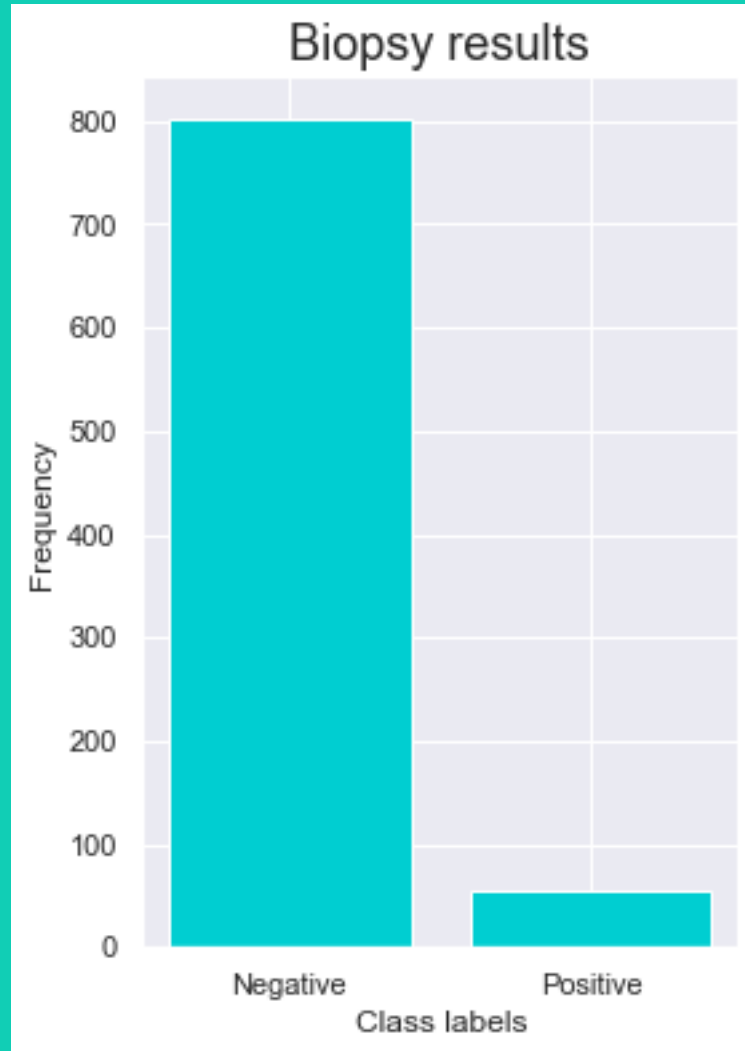
# Diagnosing Cervical cancer using Machine Learning

Team Members: Aayushi Tripathi,  
Mansi Palekar, Aditya Doomra  
and Hetansh Mehta



# Cervical cancer is a problem that affects millions of people globally

- The only procedure that can confirm this disease is a biopsy
- This requires medical expertise and sound infrastructure that may not be available in rural and developing areas
- Machine learning models can be used in those areas to help screen potential patients
- We used data of 858 patients collected in a Hospital in Venezuela to train machine learning models to detect cancer without a biopsy



# Dealing with Class Imbalance

- The response variable that we want to predict is Biopsy
- This had huge class imbalance with 13.6 times more negative entries than positive
- Changing metrics to judge model performance to precision, recall, F1 score
- Model training using under sampling, oversampling and class weights

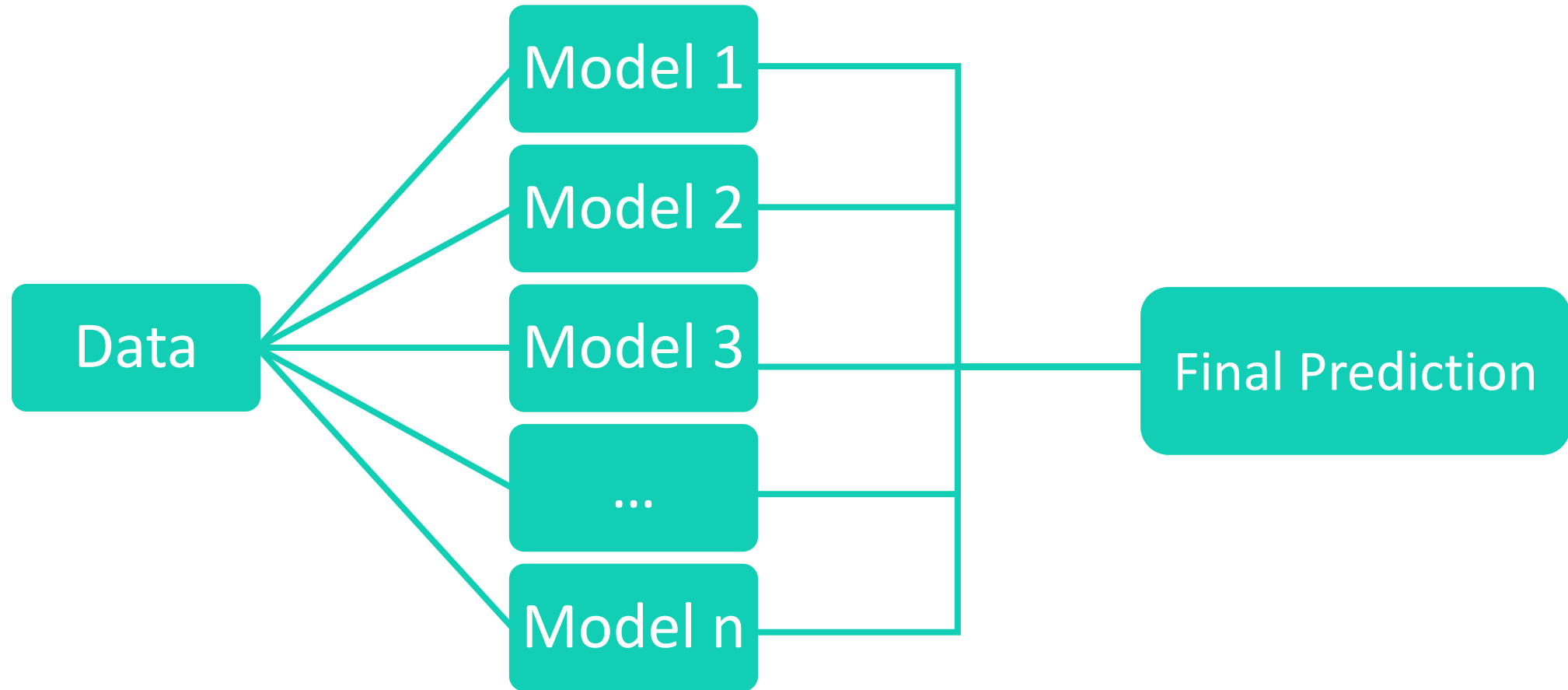
# We trained 4 models to predict the biopsy results of patients



- Logistic Regression
- Decision Tree with Bagging
- Random Forest Algorithm
- XGBoost

# Ensemble models

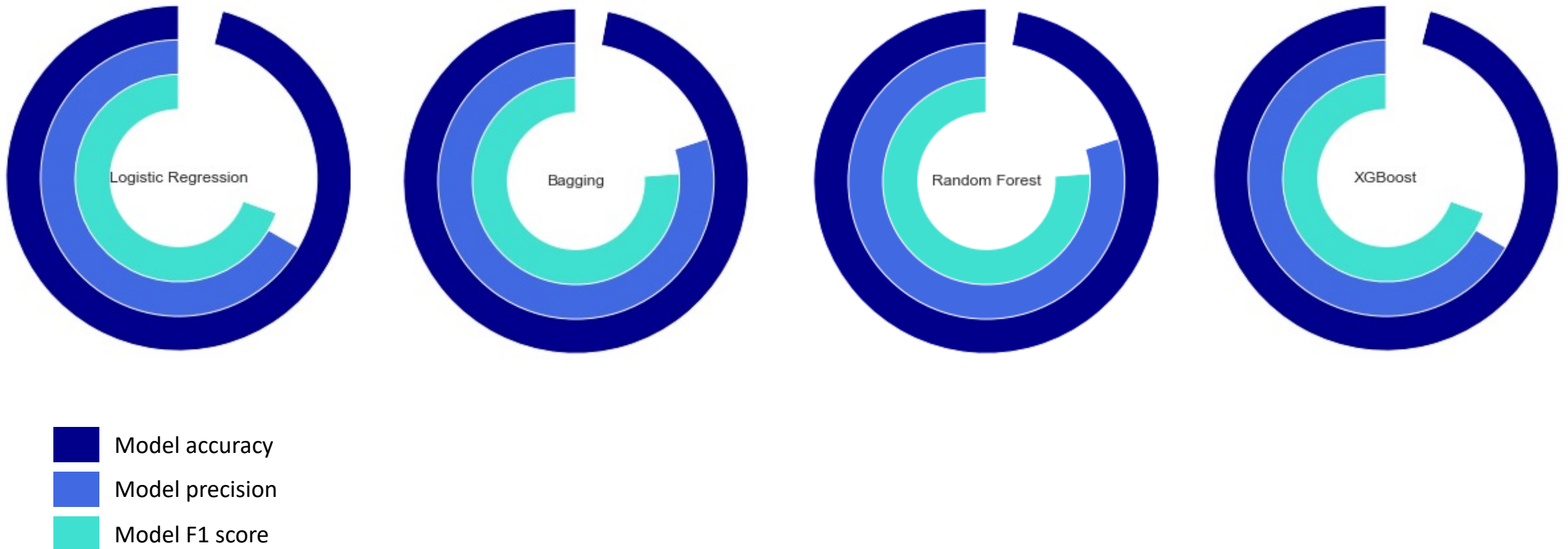
Bagging – Boosting – Random Forest Algorithm

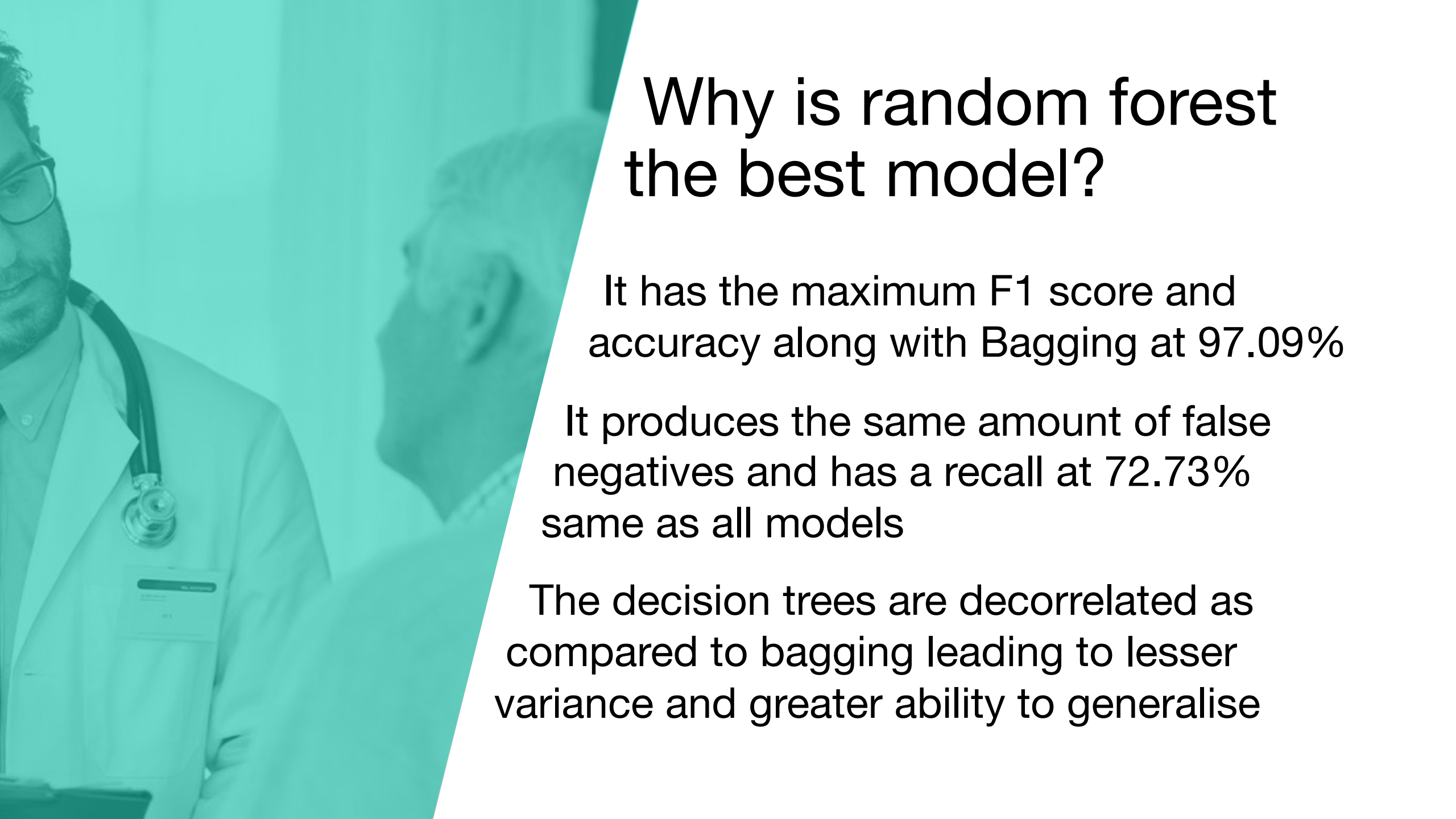




# Results

After training, tuning hyperparameters of models, tackling class imbalance using various methods, following are the best performance of all models





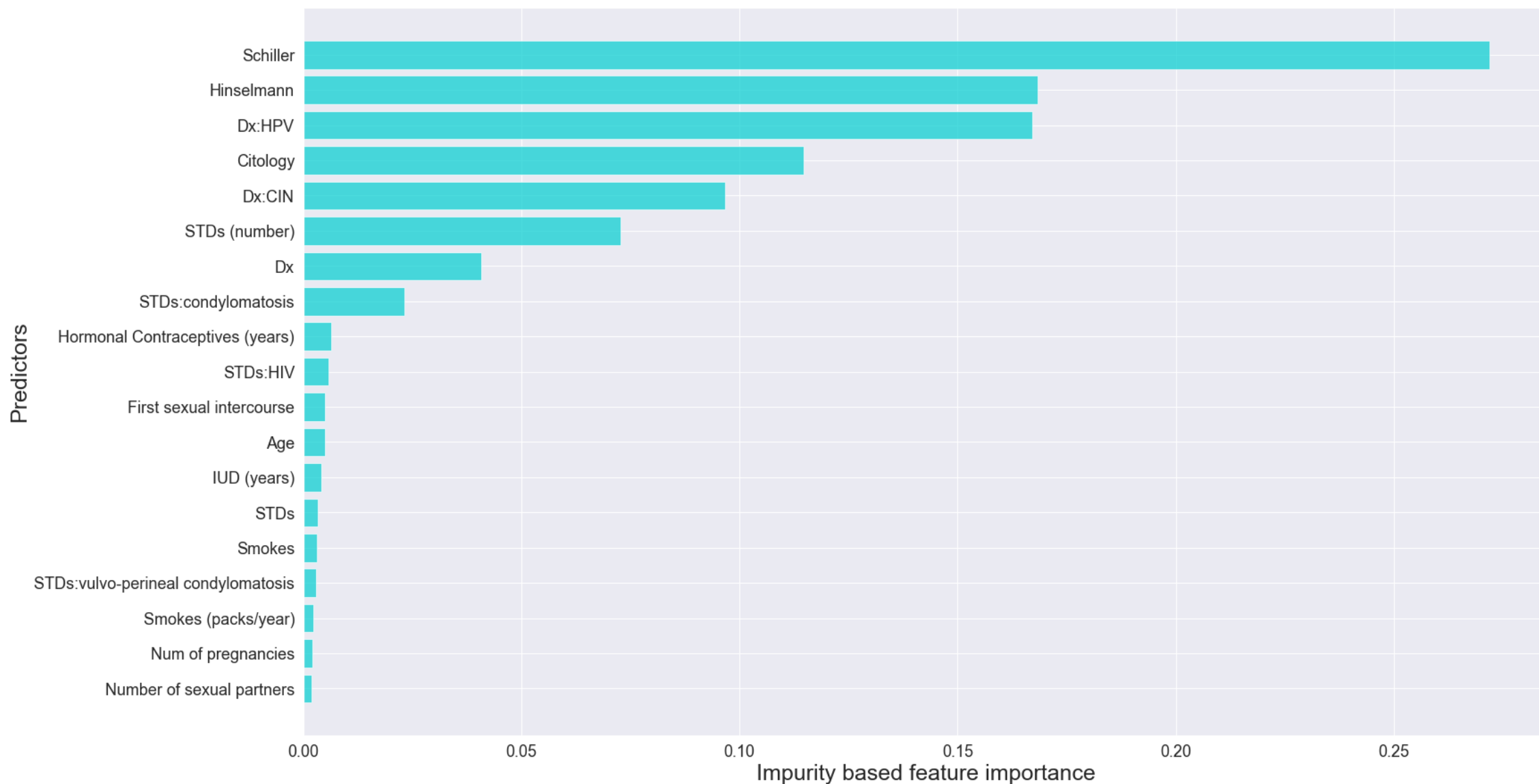
# Why is random forest the best model?

It has the maximum F1 score and accuracy along with Bagging at 97.09%

It produces the same amount of false negatives and has a recall at 72.73% same as all models

The decision trees are decorrelated as compared to bagging leading to lesser variance and greater ability to generalise

# Random forest model feature importance





# Model Performance

97.1%	80.0%	72.7%	76.2%
Accuracy	Precision	Recall	F1 score

- This performance was achieved by training on the data of just 686 patients
- More data will lead to better and more robust models that can be deployed in rural and developing areas
- According to WHO “When diagnosed, cervical cancer is one of the most successfully treatable forms of cancer, as long as it is detected early and managed effectively.”
- A full scale model can be deployed in collaboration with hospitals to get more data to train the model and help screen patients as early as possible to help in better treatment