Predicting Left Ventricular Diastolic Volume

Andrew Bergman

27 August 2019

Agenda

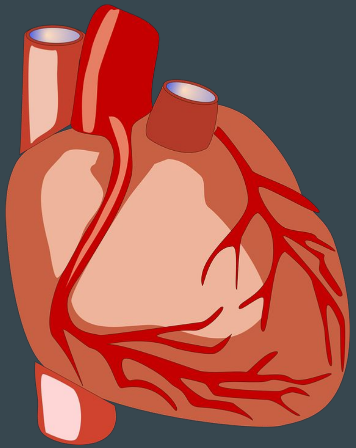
- ❖ Problem Statement
- ❖ Background Information
- ❖ Methods & Models
- ❖ Model Evaluation
- ❖ Conclusions
- ❖ Next Steps

Problem Statement

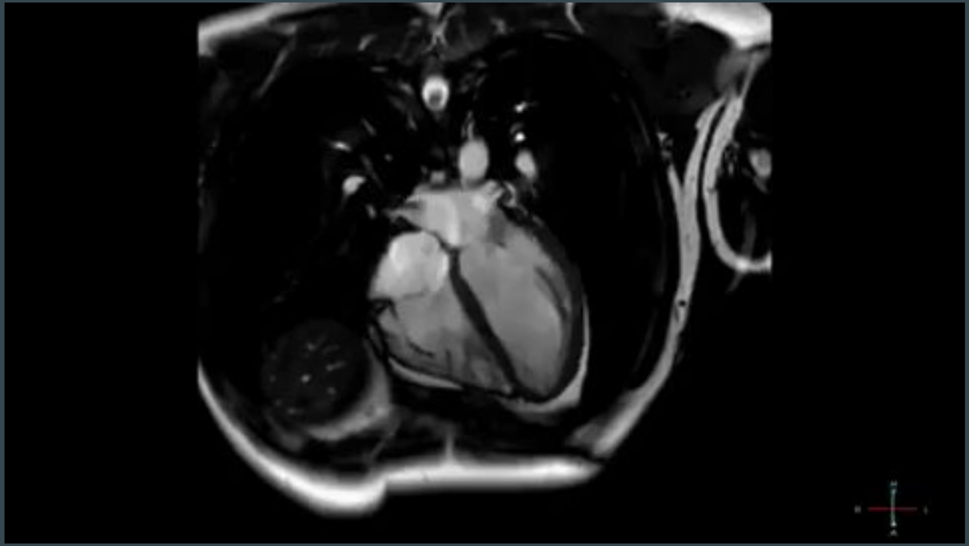
Left ventricular dilation is a serious condition because the muscular wall of the ventricle thins and becomes less able to pump blood. This is a major contributor to left-sided congestive heart failure & indication of heart function. Being able to predict the left ventricular volume would give cardiologists an idea of who needs therapy the most and help increase efficiency in the healthcare system.

Background

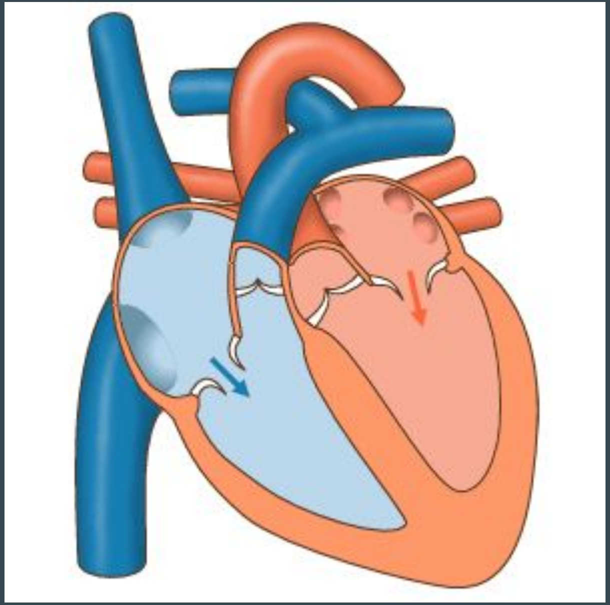
- The left ventricle pumps oxygenated blood into systemic circulation
- If the left ventricle cannot pump efficiently, blood will back up into the lungs causing further cardiac and pulmonary problems
- Diagnosis of dilation is typically done by cardiac MRI
  - Cardiac MRIs are expensive and unpleasant



Background

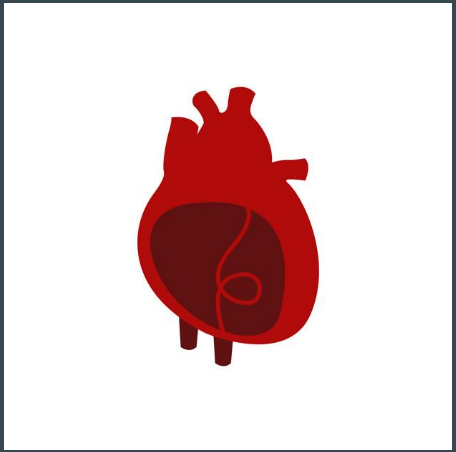


MRI source



Preprocessing

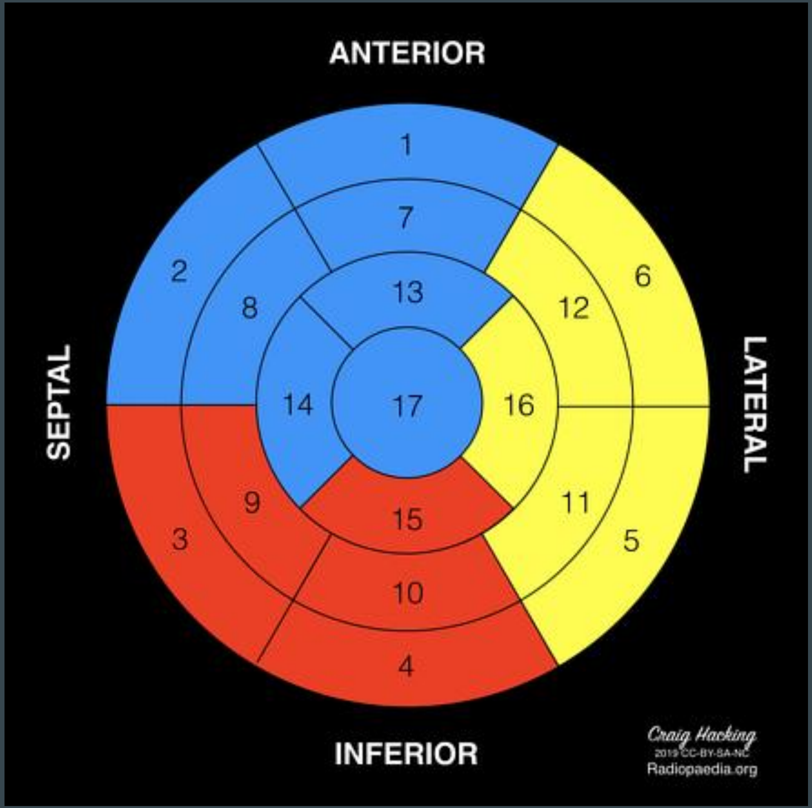
- Minimal cleaning was needed
- Most of the cleaning was making ordinal columns were numeric
- Some values did not make sense and had to be removed
- Missing data was imputed via a KNN method from the fancyimpute library



Feature Engineering

- We were able to identify two primary groups of data
- They are segmental features for HE and ischemia
- The left ventricle can be divided into 17 segments
- We intended to create interaction columns based off of the segmentation
- We ended up having to create “summary” columns instead
- There were only 3 summary columns: basal, mid, and apex

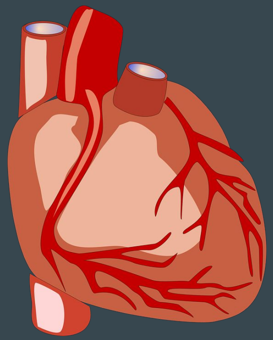
Preprocessing



Source

Models

- We began with a simple linear model
- Branched into regularized linear models: Ridge, LASSO, ElasticNet
- We used two non-linear models: random forest and XGBoost
- The models were run on a set containing the original and engineered features
- Models were evaluated on four metrics
- Model performance was determined by metric scores and residual plots

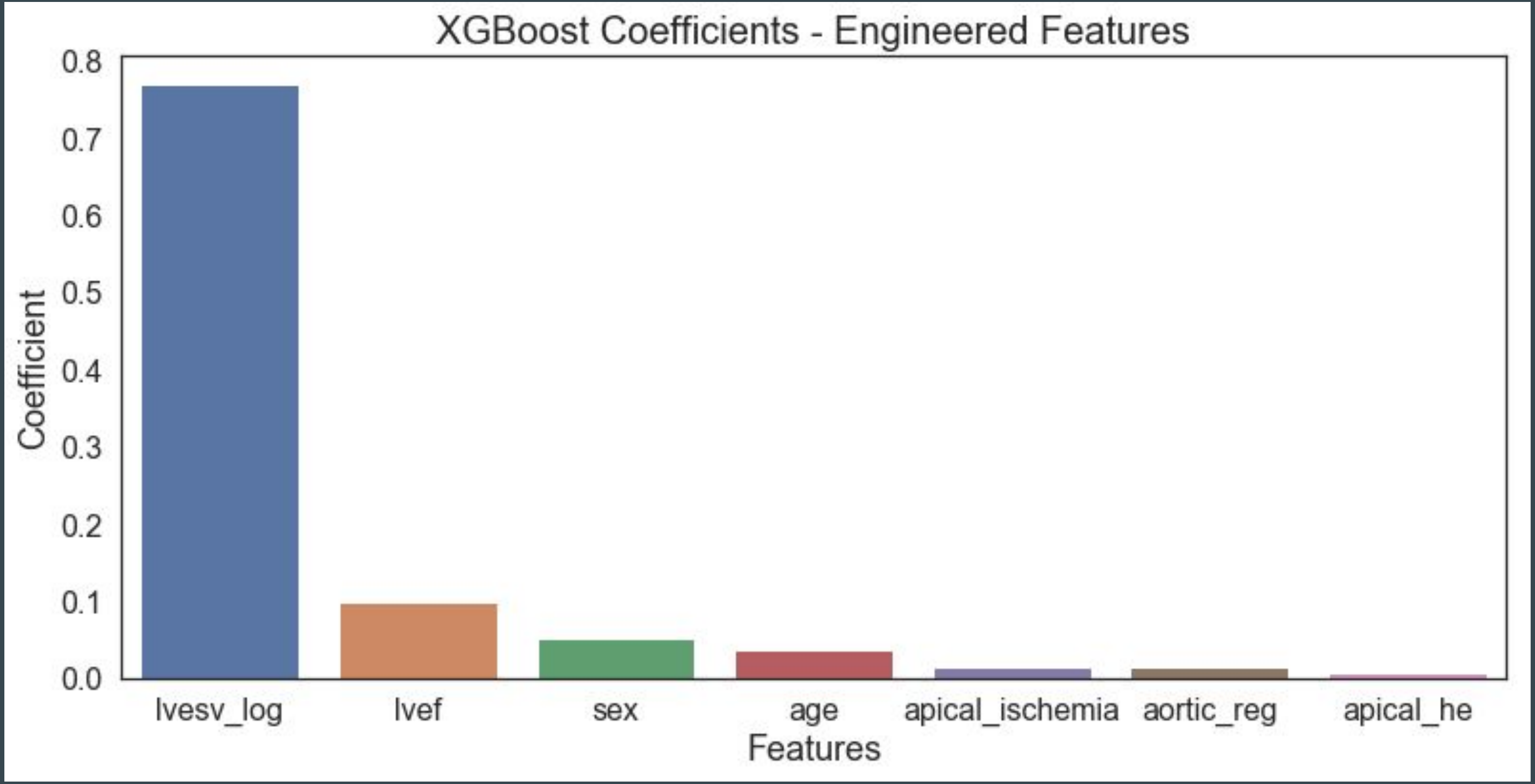


Conclusions

- We chose 2 best models
  - Random forest regression for performance, but is a black box
  - XGBoost regression for interpretability

Metric	Random Forest	XGBoost
RMSE	5.429	6.787
MAE	2.770	3.961
R^2	0.9926	0.9884
Adjusted R^2	0.9926	0.9884

Conclusions



Next Steps

- Model on information from patient charts
- Continue optimizing the random forest and XGBoost models
- Implement PDPbox to improve the interpretability of the random forest
- Continue feature engineering

Weill Cornell
Medicine

Definitions Of Some Terms

- Ventricle → two lower chambers that pump blood out of the heart
- End Diastolic Volume (EDV) → maximum volume of blood when filling
- End Systolic Volume (ESV) → volume at the end of contraction
- Ejection Fraction (EF) → how much blood is pumped during contraction
- Regurgitation → blood flowing back through a valve
- Hyper-Enhancement (_he) → scarification/infarction
- Ischemia → restriction of blood flow