Predicting Death And Cardiovascular Disease

Jonathan Garnier & Philip Terrill

INRODUCTION

Obstructive Sleep Apnea (OSA) is an increasingly common sleep disorder among the modern population, occurring when upper airway muscles relax during sleep, causing breathing to stop and start. The condition is linked to many other severe health complications including Hypertension.

AIM

The purpose of this study, is to determine which characteristics of OSA are most associated with hypertension. In doing so, the results will not only enhance the current understanding of OSA and hypertension, but may also give insight as to how to treat the conditions more effectively.
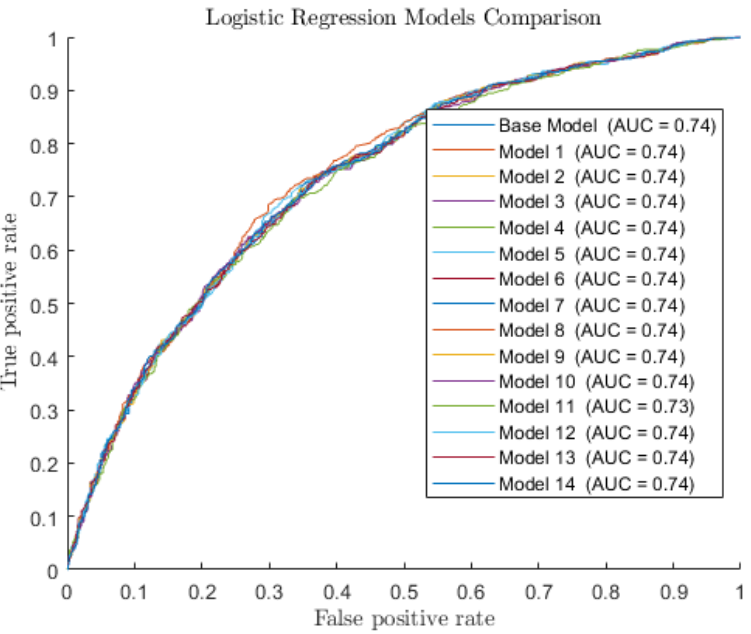
METHODS

Experiment 1
- Selecting features from dataset
- Refining features using feature selection algorithms
- Comparing classifiers

Experiment 2
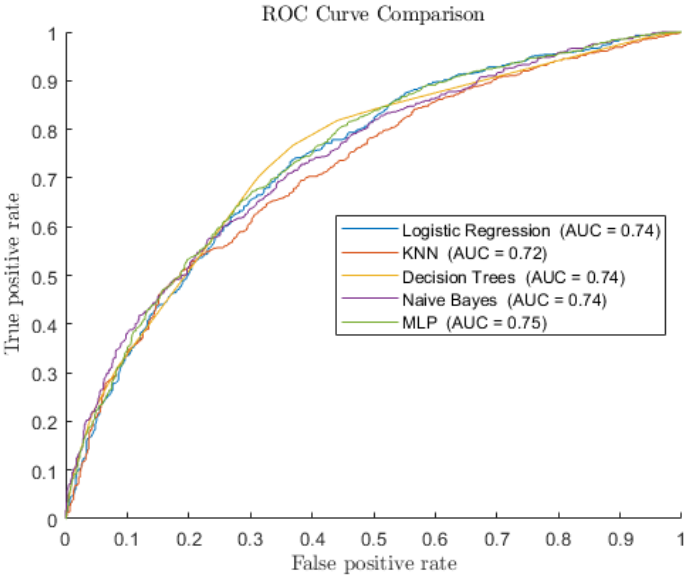- Further analysing best classifier
- Comparing performance using different subsets of OSA related data
- Interpreting individual models of interest

Algorithm	Training Accuracy	Testing Accuracy	AUC Score
Naive Bayes	65.70%	62.96%	0.74
K-Nearest Neighbours	66.57%	64.99%	0.72
Logistic Regression	67.04%	65.94%	0.74
Decision Trees	63.30%	62.94%	0.73
Multi-Layer Perceptron	68.04%	67.84%	0.75

Comparison of performance metrics of chosen classifiers. Multi-Layer Perceptron performs the highest followed by Logistic Regression based off test accuracy.

Comparison of ROC curves for chosen classifiers. No discernible difference that suggests one is a much higher performing classifier than another.

Comparison of ROC curves for Logistic Regression models with different subsets of data. Again, no discernible increase or decrease in performance when testing OSA related variables.

	Estimate	SE	t-Stat	p-value	Odds Ratio
Age	0.036	0.0064	5.6329	0.0001	1.0366
Smoking Status	-0.5192	0.2093	-2.4803	0.0131	0.595
Sleep Efficiency	-0.0226	0.0111	-2.0376	0.0416	0.9776
Diabetic Status	0.4857	0.2273	2.1367	0.0326	1.6252
FEV	-0.6856	0.1003	-6.8326	0.0002	0.5038
Neck Circumference	0.096	0.0242	3.9705	0.0001	1.1008

Table of coefficients table for base model using features selected directly out of feature selection algorithms. Only variables with p-values < 0.05 are shown to highlight which features were statistically significant.

RESULTS

Experiment 1
- MLP performed highest as expected
- Logistic Regression was chosen to further analyse due to its performance and interpretability

Experiment 2
- No significant increase or decrease in performance from testing other models
- No significant revelations from analysing p-values and odds ratios from tables of coefficients

CONCLUSIONS

Summary
- No outstanding unknown OSA predictors of hypertension
- Hypertension is made of many confounding factors – methods of prediction in this thesis are limited

Future Work
- Take statistical driven approach rather than literature driven
- Neural Network Backpropagation
- Split data by gender
- Regression rather than binary classification
- Imputate missing data rather than remove

INNOVATION EXPO