Diagnosing Asthma and Chronic Obstructive Pulmonary Disease with machine learning

THESIS PROJECT by DIMITRIS SPATHIS

ADVISOR: PANAYIOTIS VLAMOS

BIHELAB

IONIAN UNIVERSITY

Why

Asthma

chronic disease of the airways

symptoms appear after exposure to stimuli

300 million patients diagnosed yearly

250,000 deaths

COPD

chronic bronchitis and emphysema

main factor of the progressive airway obstruction is smoking

330 million patients diagnosed yearly

3 million deaths

Workflow

Data collection → Machine Learning → Mobile app

Doctor Visits

132 unique patients who visited a private doctor in Thessaloniki during 2014-2015. Each patient record describes 22 different features: demographic profile, medical and special lung measurements, habits and associated symptoms.

Statistics and Learning

Descriptive statistics of the sample. Machine learning with algorithms such as NNs, SVM and DTs. Feature significance analysis.

Building a mobile app

Extracting the machine learning model into a mobile-web app ready to use.

Tools used

WEKA	JAVA
SPSS	STATISTICS
SKLEARN	PYTHON
BACKBONE.JS JQUERY.MOBILE UNDERSCORE.JS	JAVASCRIPT

Variables 1/3

#	Attributes	Value	Comments
1	Age Group	numerical	Age groups are: 0-10 years old :1 11-20 years old: 2 21-30 years old: 3 31-40 years old: 4 41-50 years old: 5 51-60 years old: 6 61-70 years old: 7 71-80 years old: 8 81-90 years old: 9
2	Sex	M / F	
3	Oxygen Saturation	numerical	The fraction of oxygen-saturated hemoglobin relative to total hemoglobin in the blood. Normal blood oxygen levels in humans are considered 95-100 percent.
4	Pulse	numerical	
5	Cough	yes, no	
6	Breath Shortness	yes, no	
7	FEV1	numerical	Forced Expiratory Volume. FEV1 is the volume of air that can forcibly be blown out in one second, after full inspiration
8	FVC	numerical	Forced vital capacity (FVC) is the volume of air that can forcibly be blown out after full inspiration,

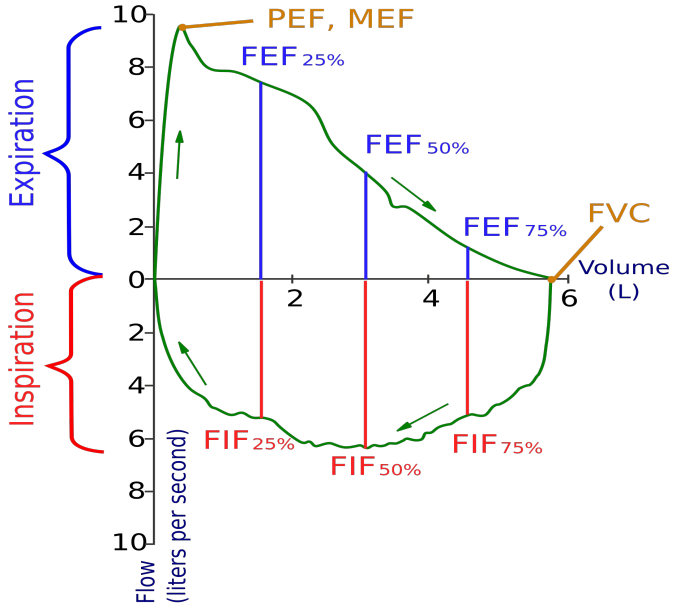
Variables
2/3

9	TIFF	numerical	Tiffeneau-Pinelli index = FEV1 / FVC
10	PEF	numerical	Peak expiratory flow (PEF) is the maximal flow (or speed) achieved during the maximally forced expiration initiated at full inspiration
11	MEF25	numerical	Maximal Expiratory Flow at 25% of the FVC curve.
12	MEF50	numerical	
13	MEF75	numerical	
14	MEF25-75	numerical	
15	Inhaler	yes, no	
16	Smoker	yes, no	

Variables
3/3

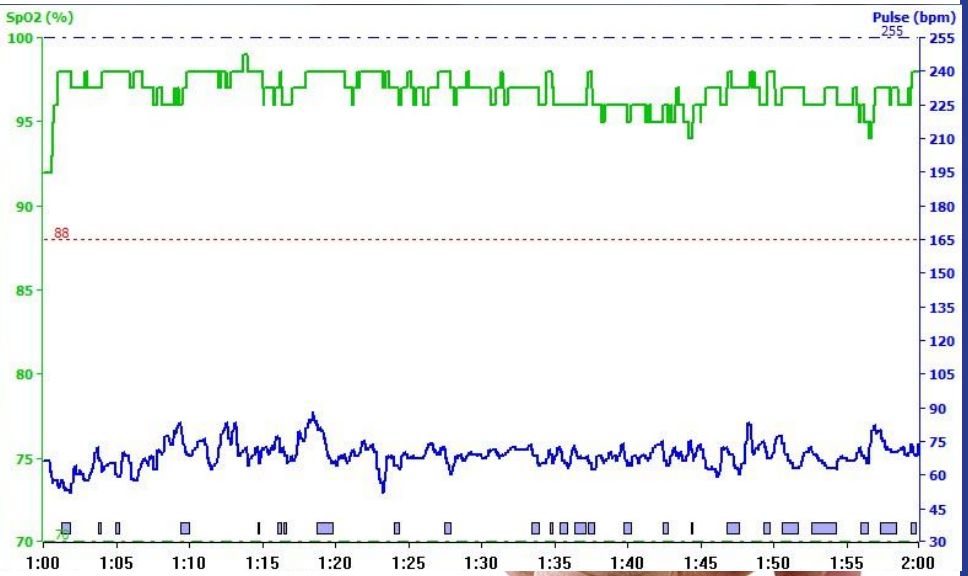
17	Wheeze	yes, no	
18	Spit	yes, no	
19	Chest Pain	yes, no	
20	Patient Name	char	
21	Asthma	yes, no	Dependent variable
22	COPD	yes, no	Dependent variable

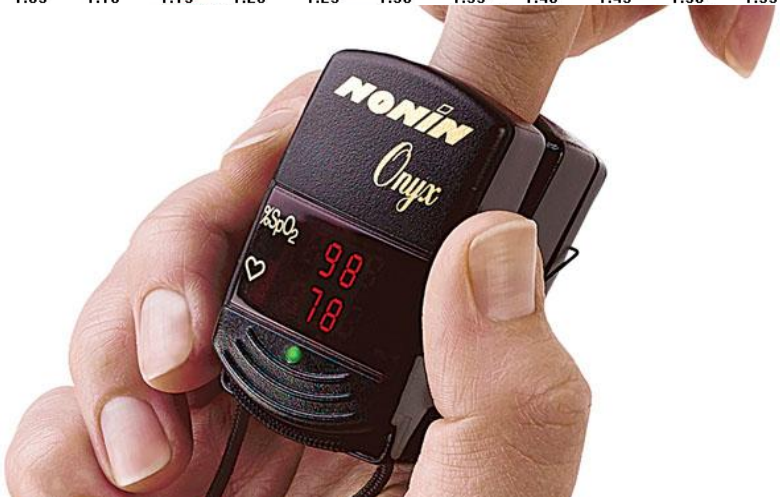
FEV1 = Forced Expiratory Volume, FVC = Forced Vital Capacity, TIFF (Tiffeneau-Pinelli index) = FEV1 / FVC, PEF = Peak Expiratory Flow, MEF = Maximal Expiratory Flow at 25%/50%/75% of the FVC curve. Patient names initials only, due to privacy reasons. (N=132)



Devices

Spirometer

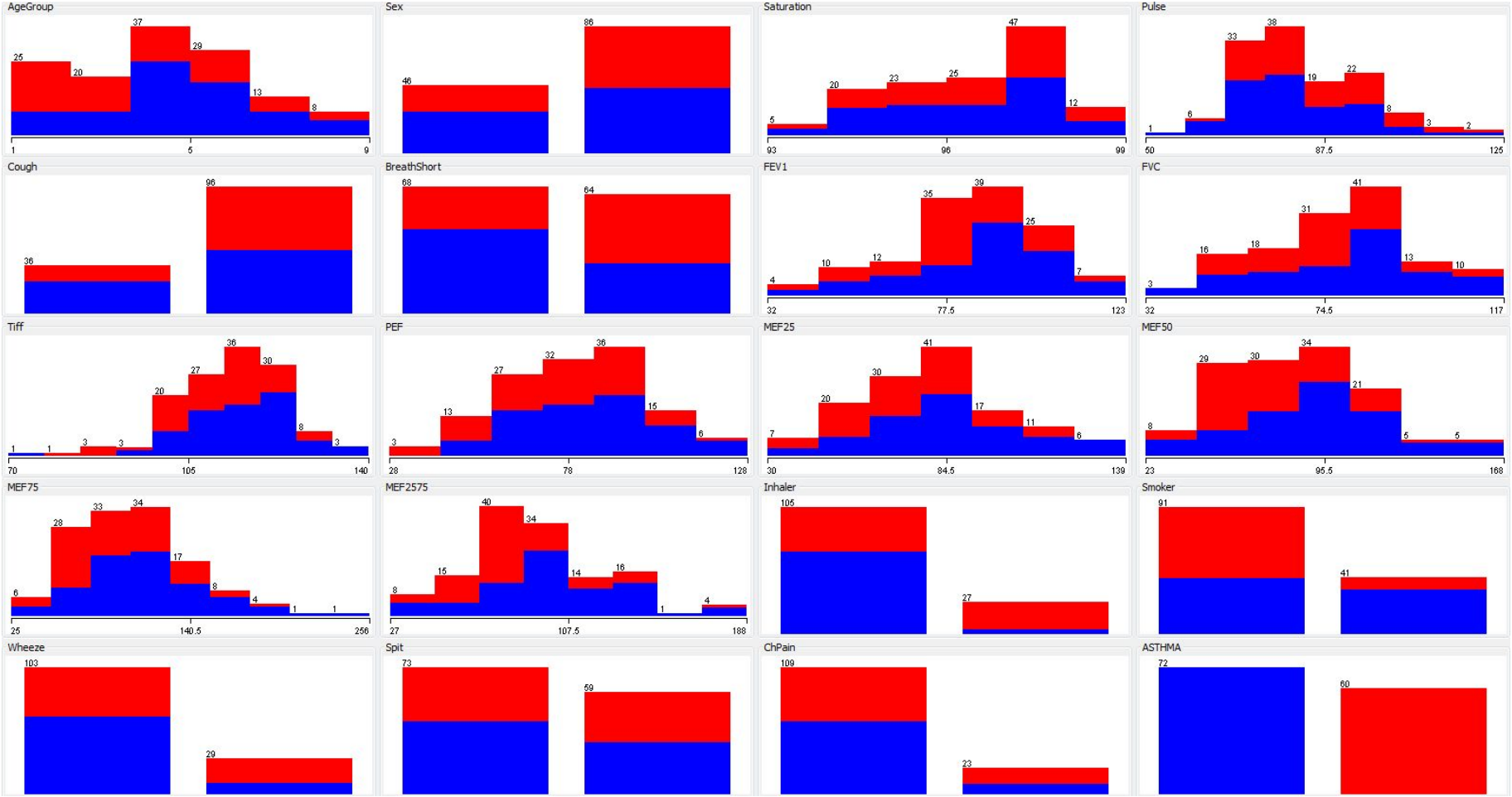




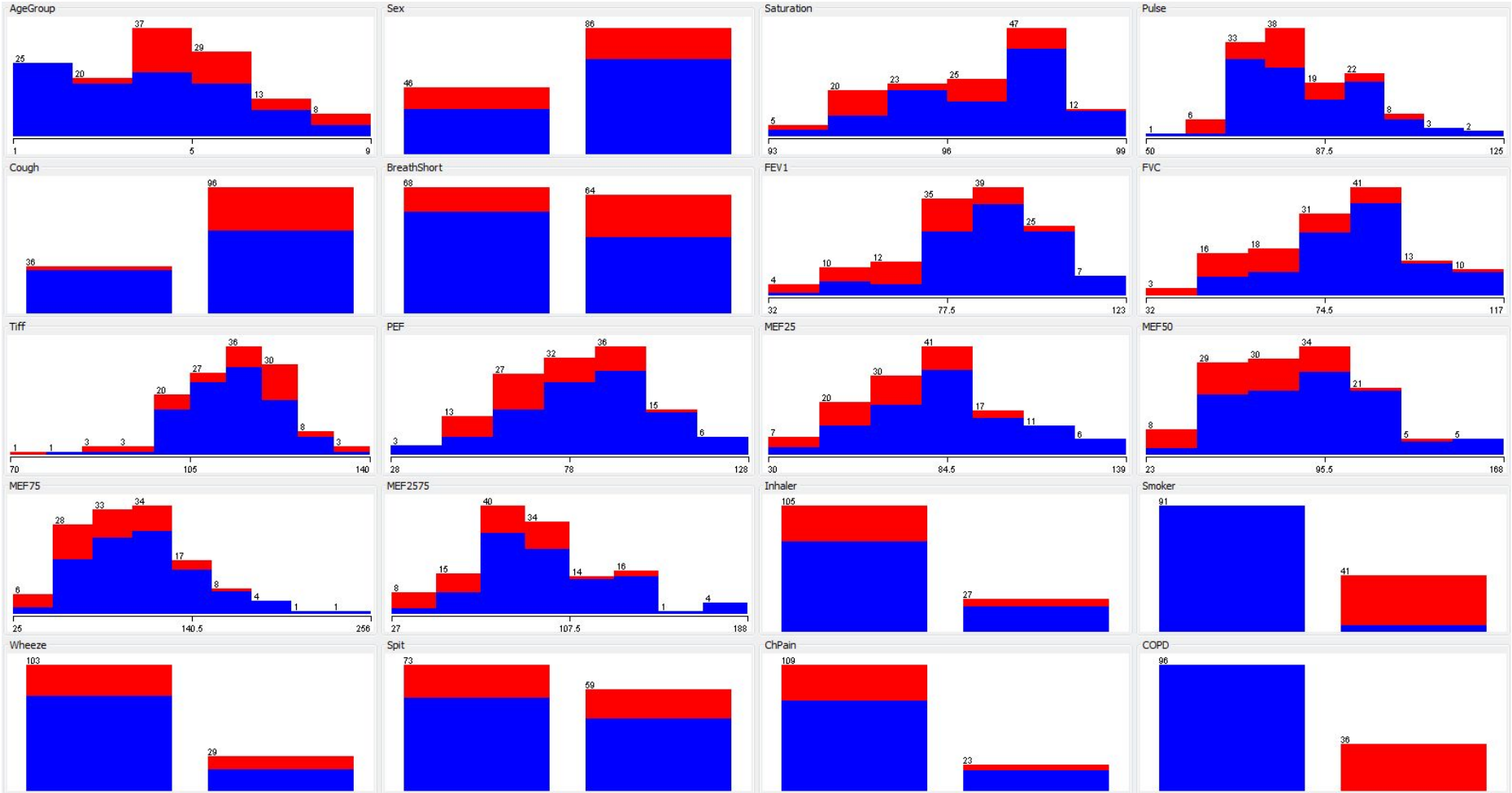
Devices

Oxymeter

Statistics



Crosstab visualization: asthma



Crosstab visualization: COPD

Machine Learning

Training Accuracy

Algorithms	Asthma %	COPD %
Naive Bayes	75.2	86.1
Logistic Regression	79.6	89.3
Logistic Regression (logitboost)	83.3	96.4
Neural Networks	81.1	94.7
Stochastic Gradient Descent	79.5	93.9
SVM	75.8	96.7
K-nearest neighbor	74.2	89.4
C4.5 Decision Tree	78.1	93.9
Random Forest	84.8	94.7

Logistic Regression is estimated with ridge estimator. Artificial Neural Networks follow the Multilayer Perceptron technique. SVM is conducted with sequential minimum optimization. C4.5 Decision Tree is implemented in J48.

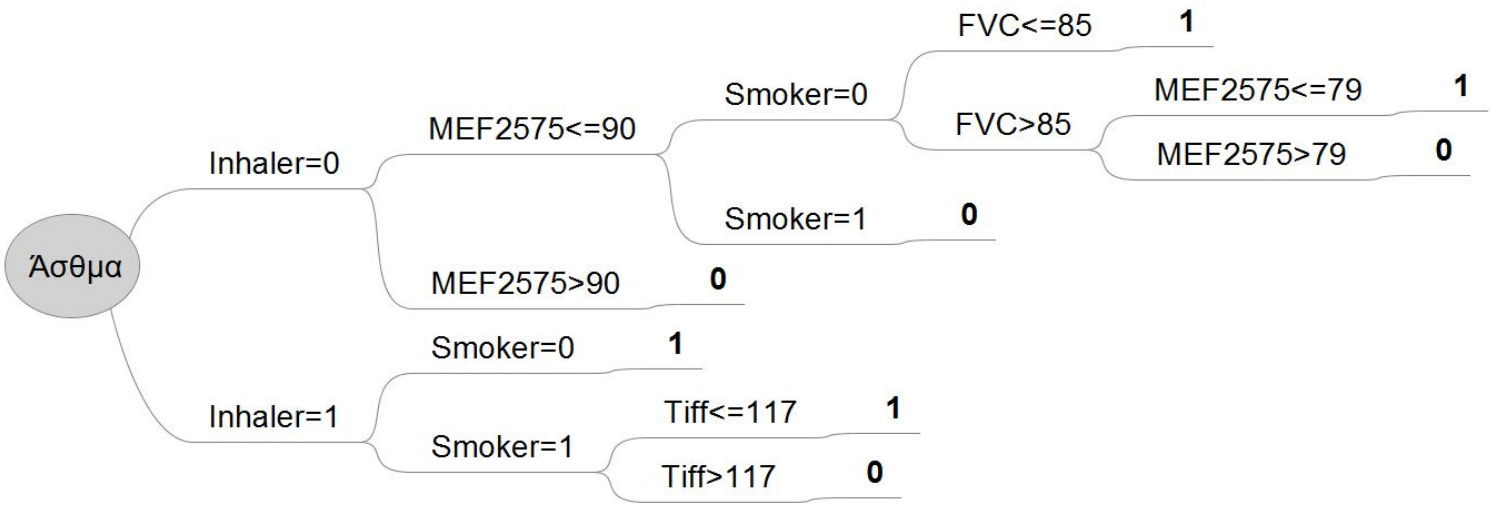
	COPD		Asthma	
Rank	Infogain	Correlation	Infogain	Correlation
1	Smoker	Smoker	Inhaler	Inhaler
2	FEV1	FEV1	MEF2575	Smoker
3	Age	FVC	Age	MEF2575
4	FVC	MEF50	Smoker	MEF50
5	Cough	Age	Breath Shortness	Age
6	Breath Shortness	MEF25	Wheeze	Wheeze
7	MEF50	MEF2575	Cough	MEF25
8	Wheeze	PEF	Chest Pain	Breath Shortness
9	MEF25	MEF75	Sex	Tiff
10	Sex	Cough	Spit	MEF75

Feature Selection

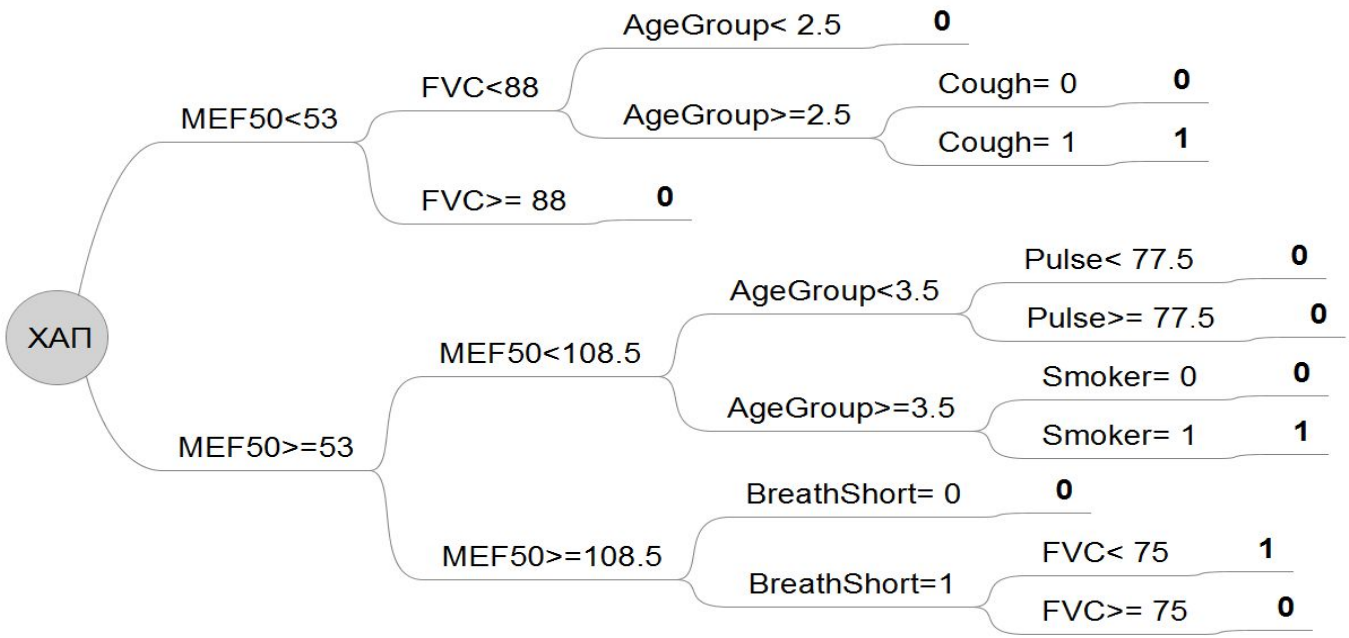
Infogain vs Correlations

InfoGain (Class, Attribute) = H (Class) - H (Class | Attribute)

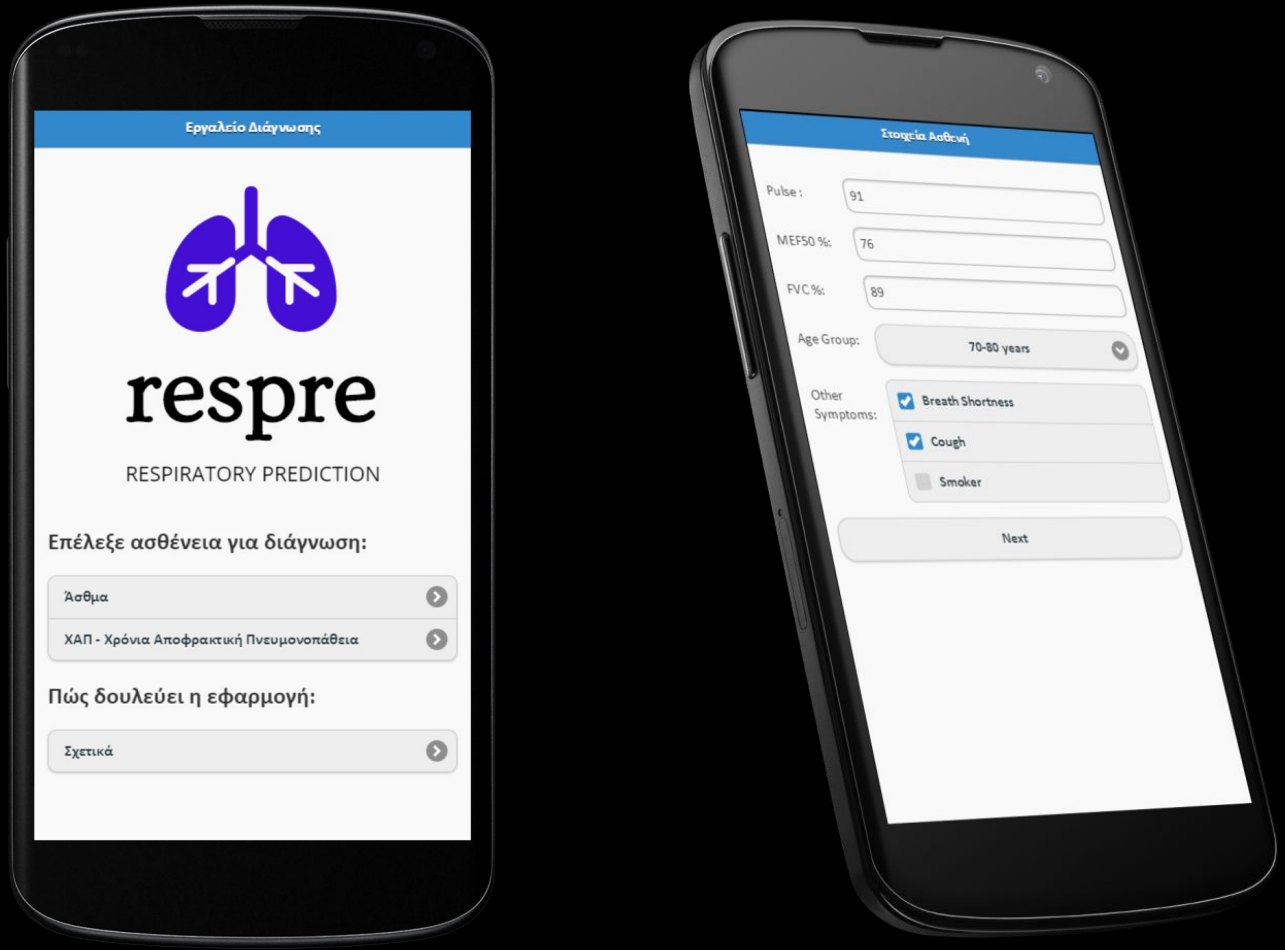
Asthma Decision Tree

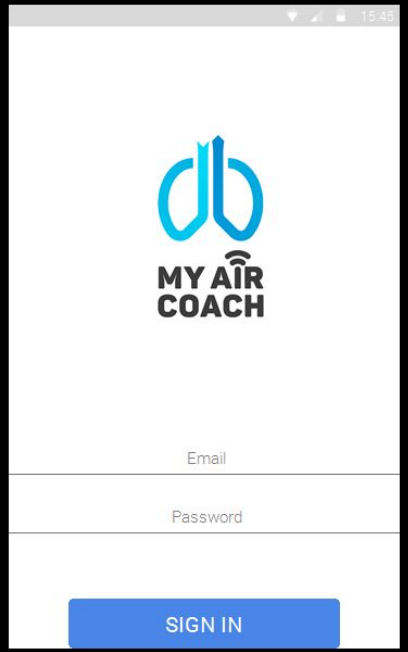


COPD Decision Tree

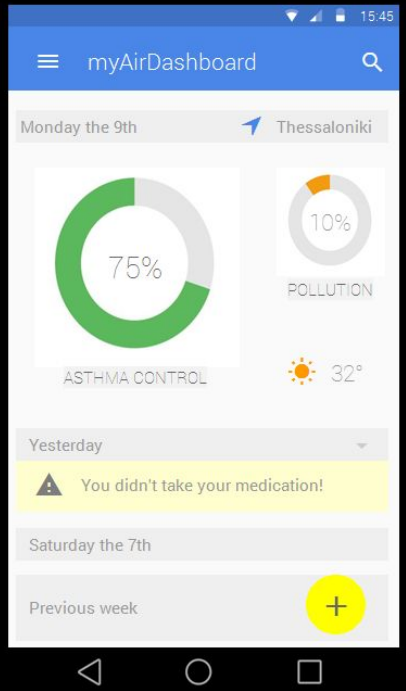


Extracting the training model into a mobile app

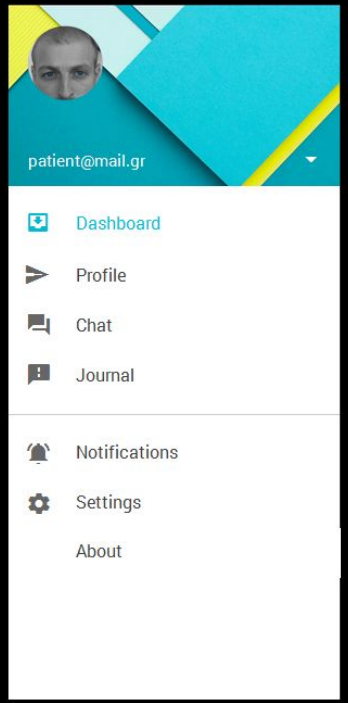

Εργαλείο Διάγνωσης
respre
RESPIRATORY PREDICTION
Επέλεξε ασθένεια για διάγνωση:
Άσθμα
ΧΑΠ - Χρόνια Αποφρακτική Πνευμονοπάθεια
Πώς δουλεύει η εφαρμογή:
Σχετικά
Στοιχεία Ασθενή
Pulse :
91
MEF50 %:
76
FVC %:
89
Age Group:
70-80 years
Other Symptoms:
Breath Shortness
Cough
Smoker
Next



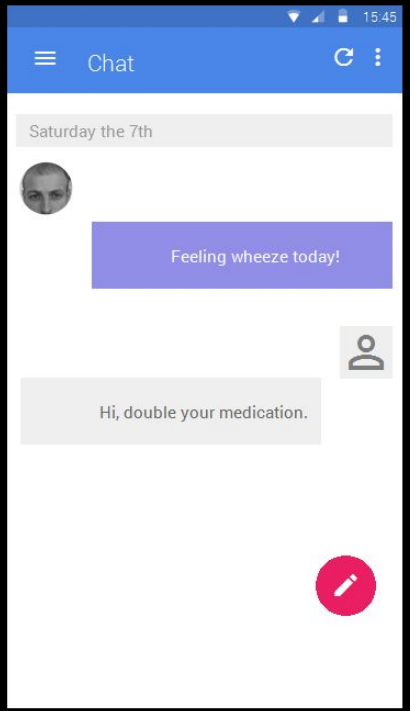
LOGIN



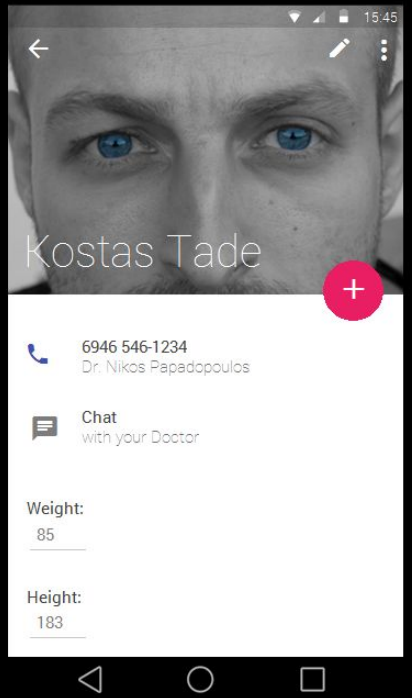
DASHBOARD



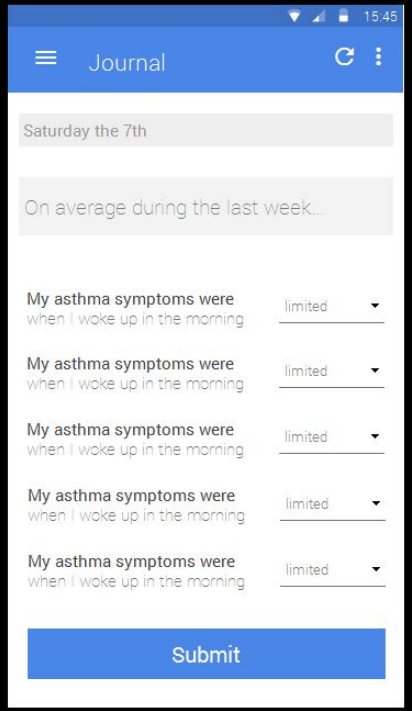
NAVIGATION



DOCTOR CHAT



PROFILE



JOURNAL

Code and thesis available

bihelab.di.ionio.gr/?page_id=209