# lime for time

# Outline

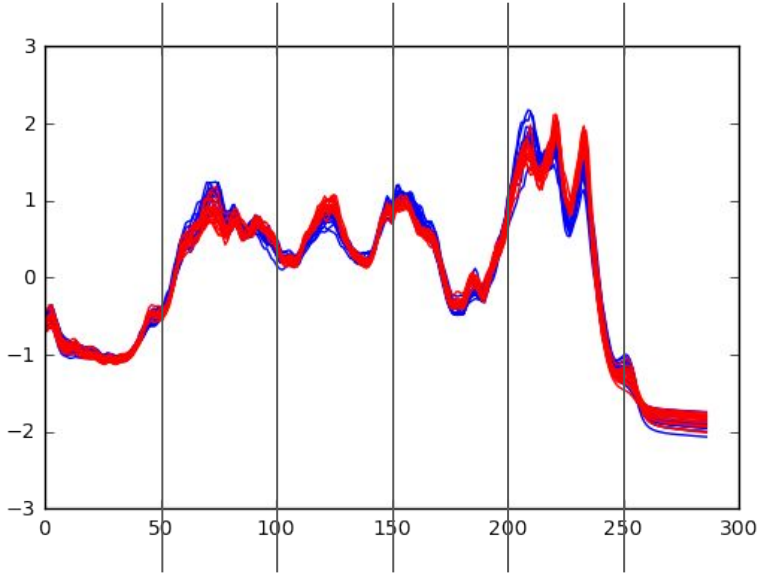
- Task
- Idea & Hypothesis
- Implementation
- Result
- Verification
- Learnings

# Task

- **Equivalent to super-pixels?**
- How to perturb a time series?
- What is the equivalent to grey super-pixels?

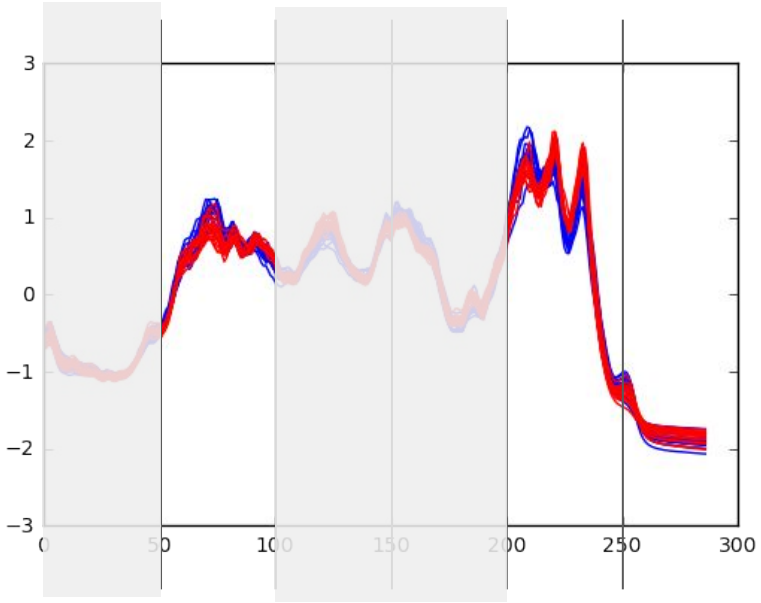
# Idea

- **Equivalent to super-pixels?**
- How to perturb a time series?
- What is the equivalent to grey super-pixels?



# Idea

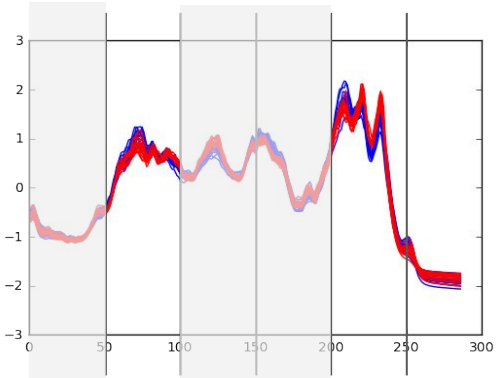
- Equivalent to super-pixels?
- **How to perturb a time series?**
- What is the equivalent to grey super-pixels?



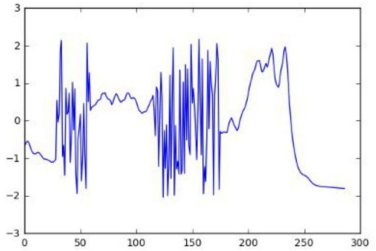

[ 0 1 0 0 1 1 ]

# Idea

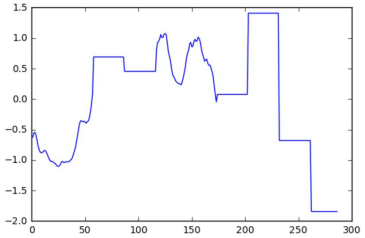
- Equivalent to super-pixels?
- How to perturb a time series?
- **What is the equivalent to grey super-pixels?**



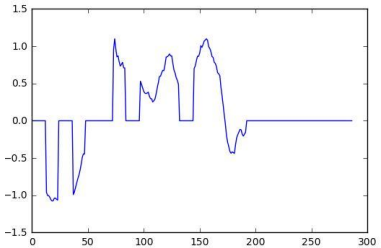
Noise



Mean over the time period

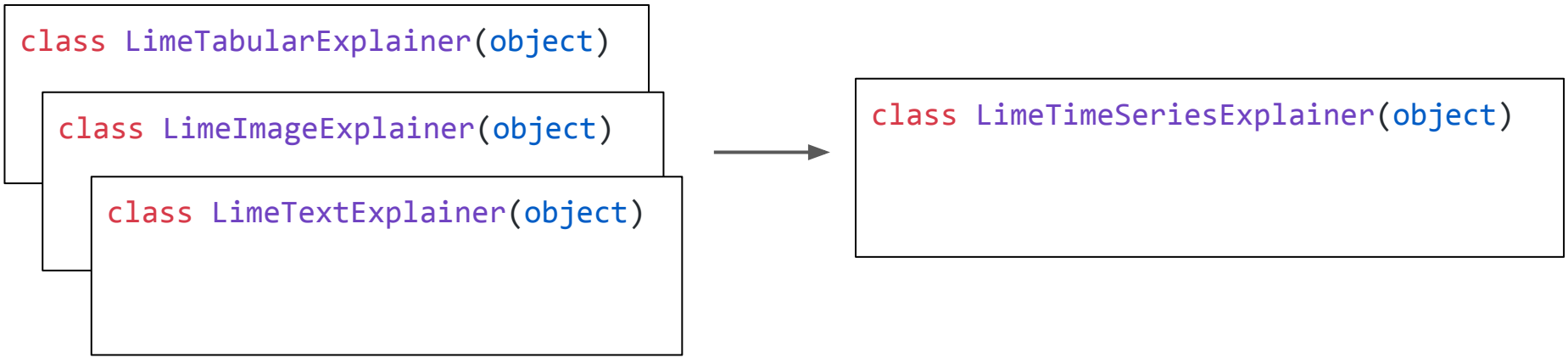


Entire mean



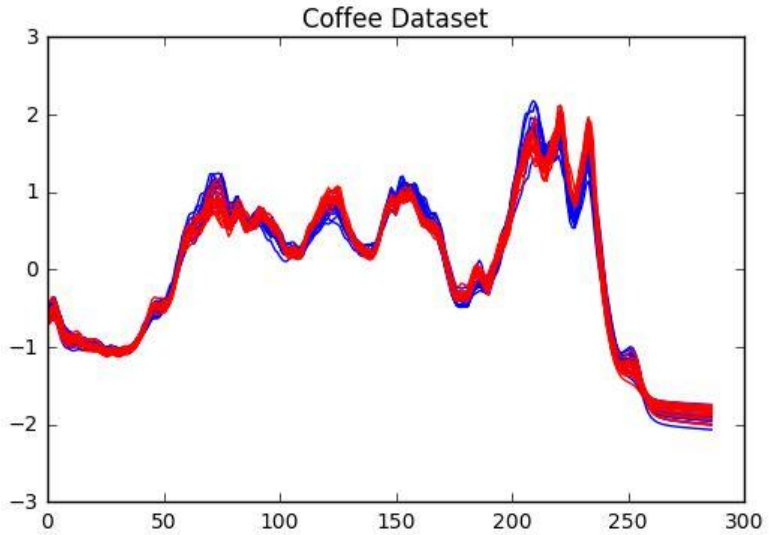
# Implementation

Implement an Explainer for Timeseries:



# Result on Coffee Dataset

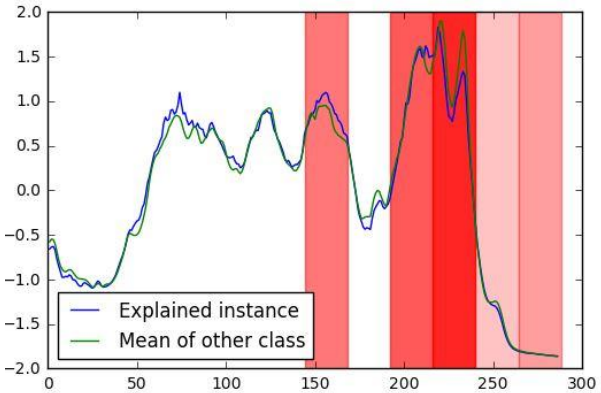
- 2 classes problem
- **“Food spectrographs** are used in chemometrics to classify food types, a task that has obvious applications in food safety and quality assurance. The coffee data set is a two class problem to **distinguish between Robusta and Arabica coffee beans**.”



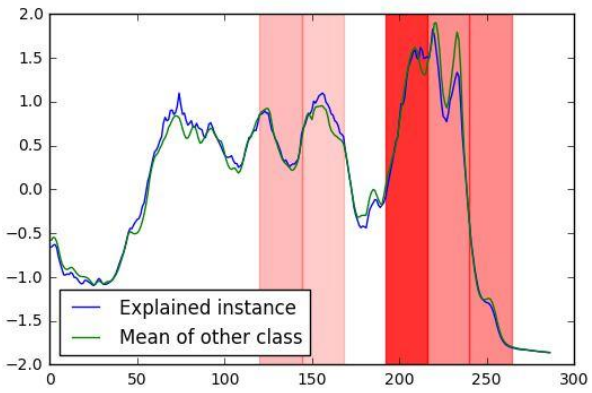
# Result on Coffee Dataset

## slices=12, top=5

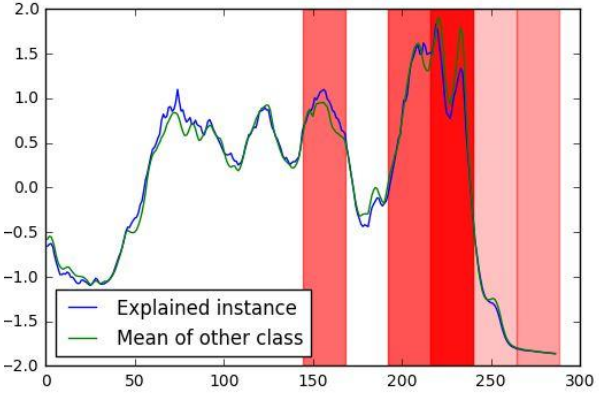
Noise




Mean over the time period



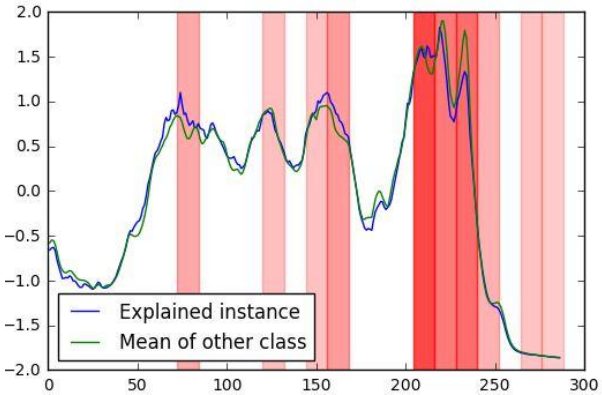

Entire mean



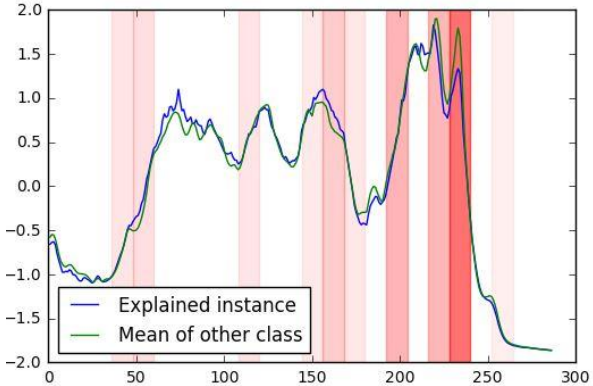
# Result on Coffee Dataset

## slices=24, top=10

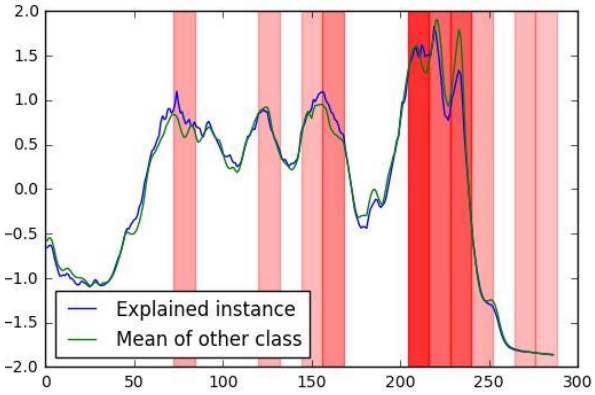
Noise




Mean over the time period



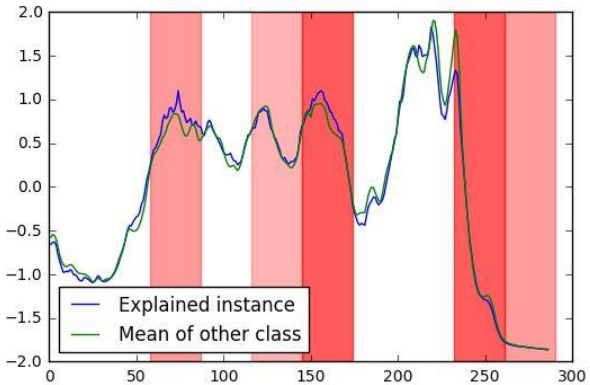

Entire mean



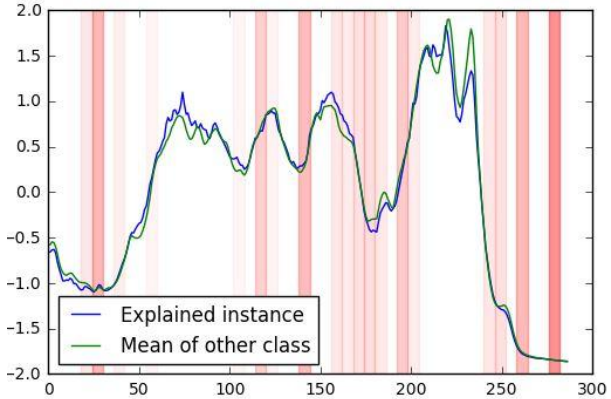
# Result on Coffee Dataset

## slices={10,50}, top={5,20}

Noise



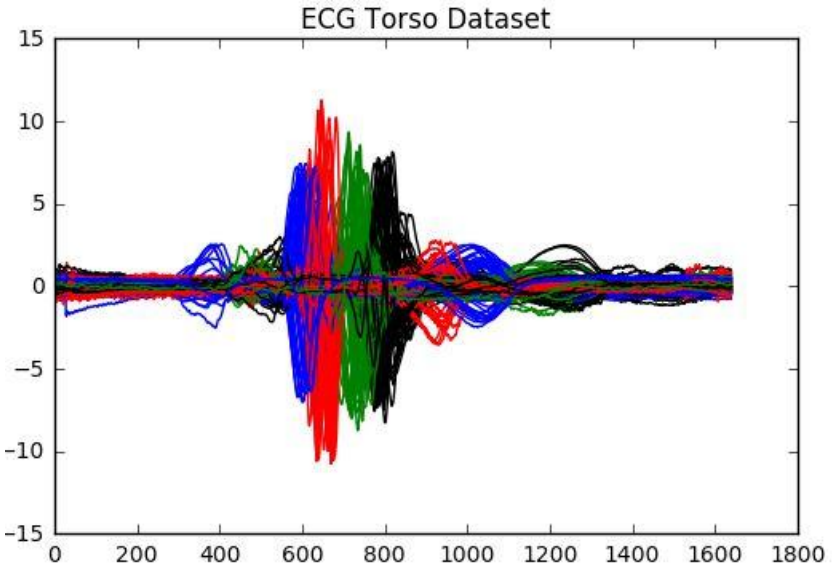

Noise




Determining the right amount of slices and top slices to be explained by LIME.

# Result on ECG Torso Dataset

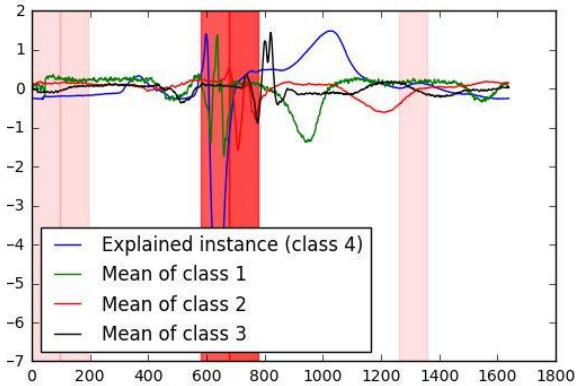
- 4 classes problem
- “This data is derived from one of the **Computers in Cardiology challenges**, an annual competition that runs with the conference series of the same name and is hosted on physionet. Data is taken from ECG data for multiple torso-surface sites. There are 4 classes (**4 different people**).”



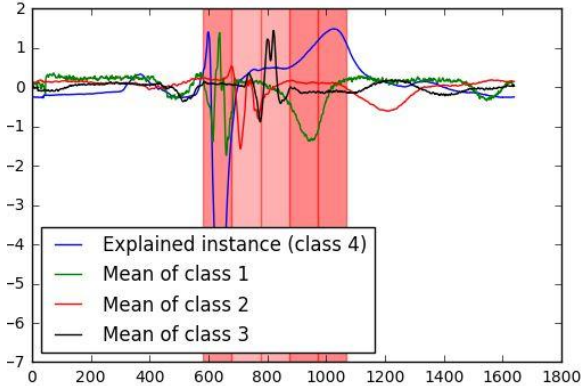
# Result on ECG Torso Dataset

## slices=17, top=5

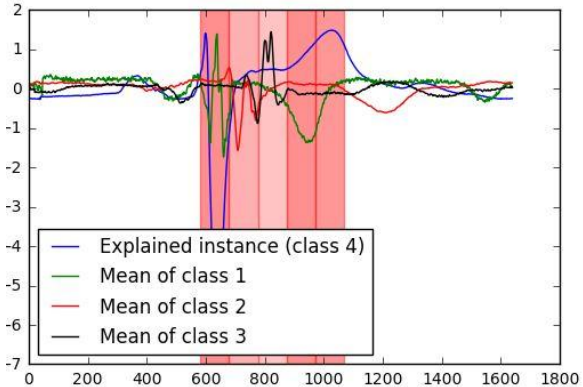
Noise




Mean over the time period




Entire mean



# Verification

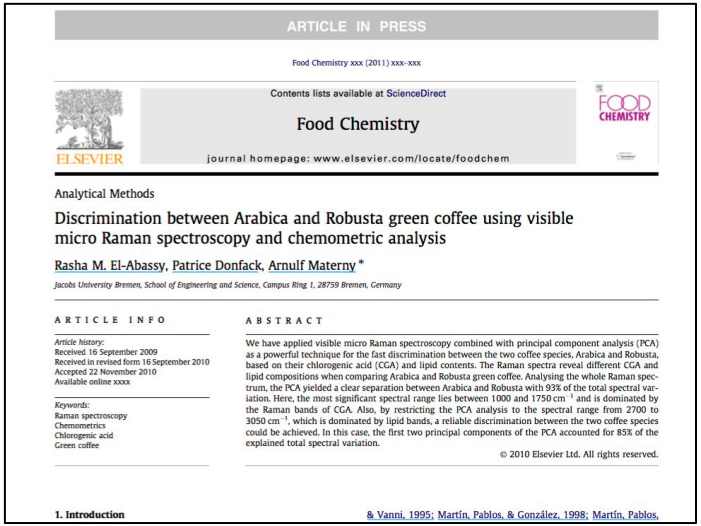
**Does our Lime explainer provide solid explanations?**

Verify Lime results with:

- Expert knowledge
- With classification models

# Verification
## Coffee Dataset

- **“Food spectrographs** are used in chemometrics to classify food types, a task that has obvious applications in food safety and quality assurance. The coffee data set is a **two class problem to distinguish between Robusta and Arabica coffee beans**.”



ARTICLE IN PRESS

Food Chemistry xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis

Rasha M. El-Abassy, Patrice Donfack, Arnulf Materny *

Jacobs University Bremen, School of Engineering and Science, Campus Ring 1, 28759 Bremen, Germany

ARTICLE INFO

*Article history:*
Received 16 September 2009
Received in revised form 16 September 2010
Accepted 22 November 2010
Available online xxxx

*Keywords:*
Raman spectroscopy
Chemometrics
Chlorogenic acid
Green coffee

ABSTRACT

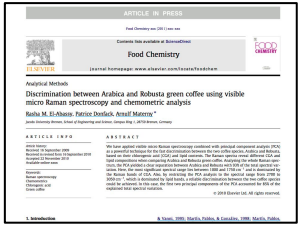
We have applied visible micro Raman spectroscopy combined with principal component analysis (PCA) as a powerful technique for the fast discrimination between the two coffee species, Arabica and Robusta, based on their chlorogenic acid (CGA) and lipid contents. The Raman spectra reveal different CGA and lipid compositions when comparing Arabica and Robusta green coffee. Analysing the whole Raman spectrum, the PCA yielded a clear separation between Arabica and Robusta with 93% of the total spectral variation. Here, the most significant spectral range lies between 1000 and 1750 $cm^{-1}$ and is dominated by the Raman bands of CGA. Also, by restricting the PCA analysis to the spectral range from 2700 to 3050 $cm^{-1}$, which is dominated by lipid bands, a reliable discrimination between the two coffee species could be achieved. In this case, the first two principal components of the PCA accounted for 85% of the explained total spectral variation.

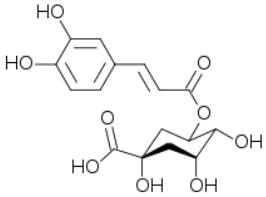


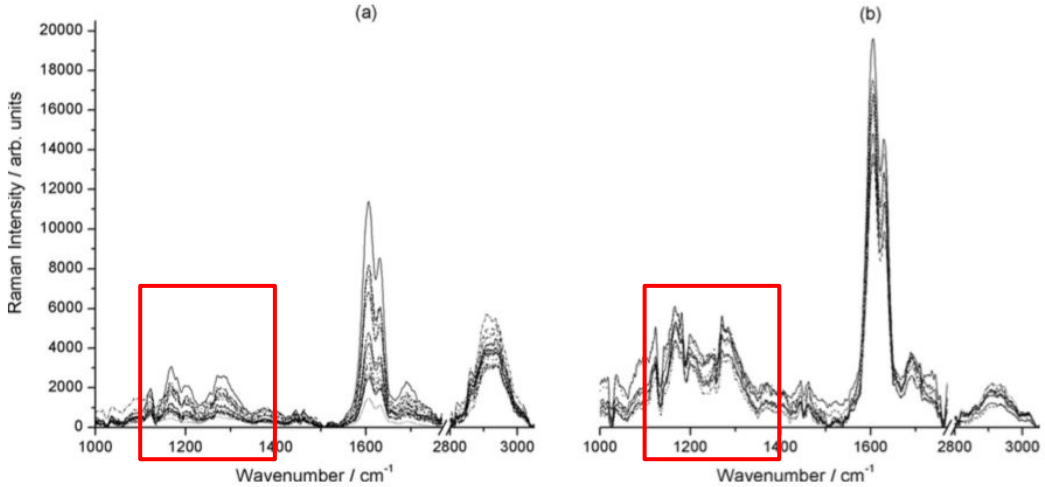
1. Introduction

& Vanni, 1995; Martín, Pablos, & González, 1998; Martín, Pablos,

# Verification
## Coffee Dataset



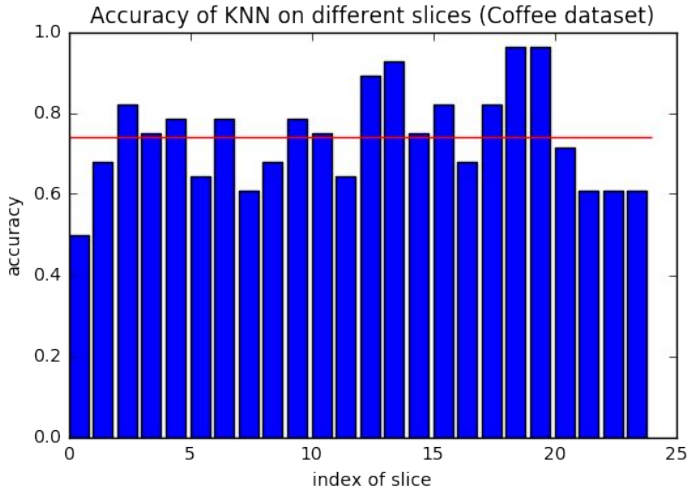



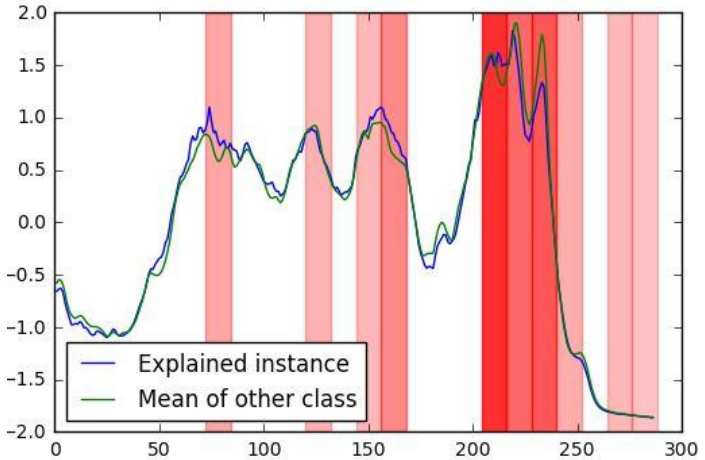




"[…] **the Robusta** green coffee samples with the weakest CGA peaks still **show stronger intensities than the Arabica** green coffee sample with the most intense CGA peaks."

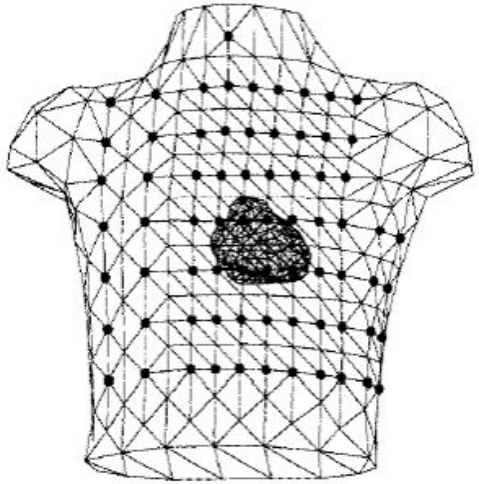
(a) Arabica, (b) Robusta

# Verification
## Coffee Dataset



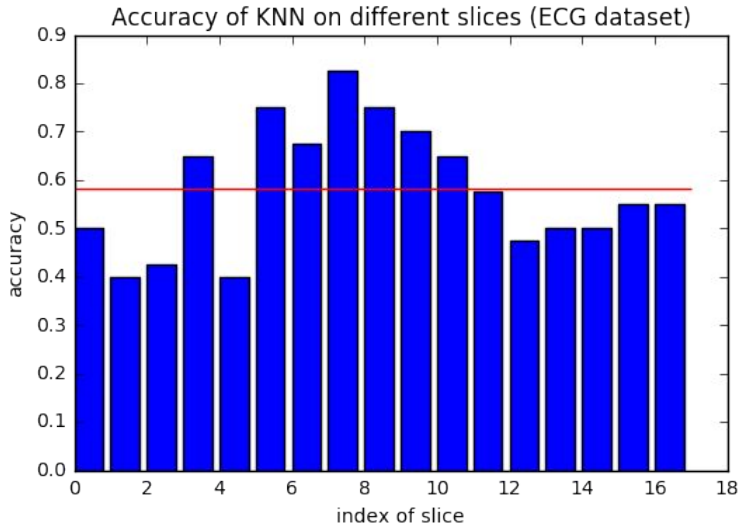



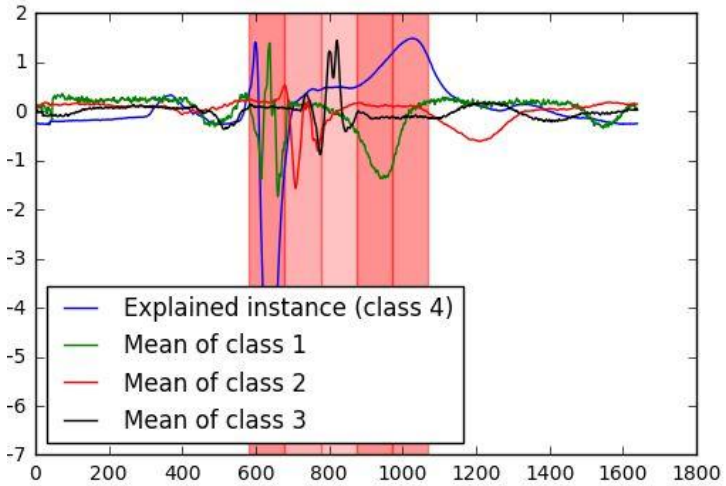
# Verification
## ECG Torso Dataset



- “This data is derived **from one of the Computers in Cardiology challenges**, an annual competition that runs with the conference series of the same name and is hosted on physionet. Data is taken from ECG data for multiple torso-surface sites. There are 4 classes (**4 different people**).”

# Verification
## ECG Torso Dataset



# Learnings

- Quality of explanation is highly dependent on number of slices and samples
- Verifying one's results is important
- Try it out

# lime for time

Thank you for your attention!

Any questions?