CLASSIFICATION OF FIBERS USING IR-SPECTRA

Karl Kaupmees, Oliver Meikar, Edgar Sepp

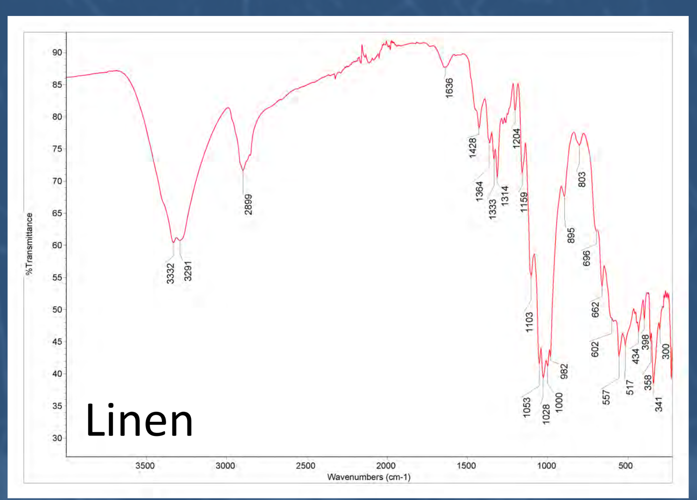
INTRODUCTION

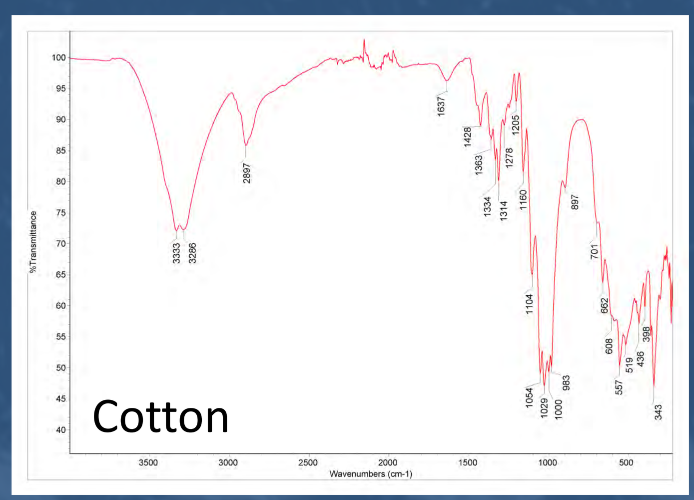
Identification of textile fibers is important in industry (quality control), forensic science (identification of fibers on crime scene), but also in conservation and archaeology (identification of historical textile fibers).

Common methods for fiber identification are microscopic observation, burning test and various solubility tests**. Infrared spectroscopy (IR)** has many advantages for fiber identification, because it offers highly characteristic information, is easy, fast, non-destructive and relatively inexpensive.

However the analysis of IR spectra is tedious and requires a trained scientist and some peak-analyzer software. The main difficulties are spectral inhomogeneities of repeated measurements and the intrinsic similarity between the spectra of some fibers.

Hence, a better method to analyze the IR specta of textiles is needed.





http://lisa.chem.ut.ee/IR_spectra/textile-fibres/

OBJECTIVES

To meet the increasing demand from the academic and industry, we aimed **build a classifier that can identify fibers by their IR spectra**. The initial aim is to be able to detect pure fibers at 80% probability.

For successful classification, we aimed to:

- reduce the complexity of the dataset
- engineer additional features from the dataset
- generate data normalization tools

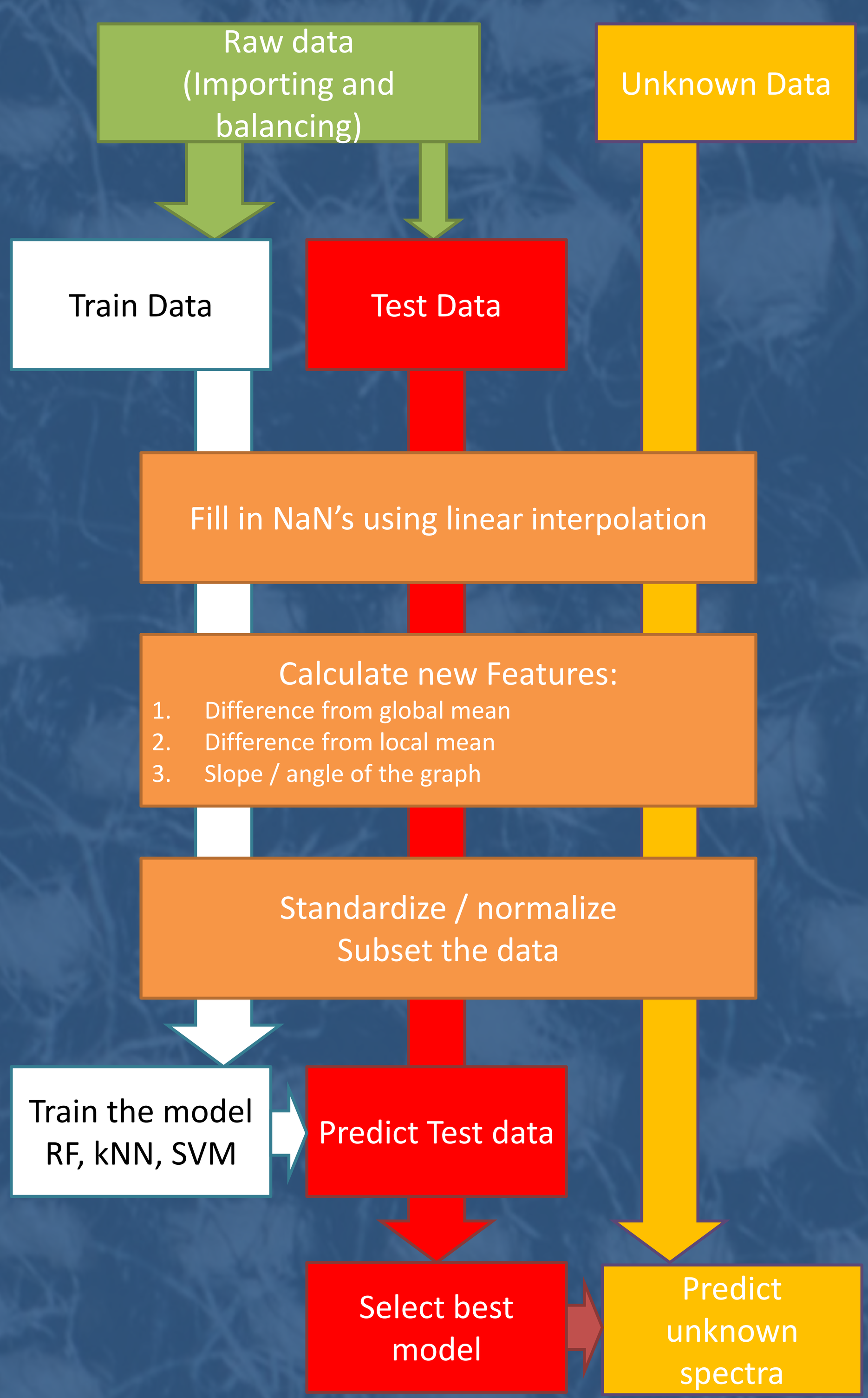
DATASET AND METHODOLOGY

Our dataset contains of 438 IR spectra of 12 pure textile fibers both, natural and synthetic:

```
Cat  count Name
1.11 52    silk
3.5  44    polyacrylic
1.5  44    cotton
1.6  40    linen
2.1  35    viscose
3.1  33    polyester
1.1  30    wool
1.9  24    jute
3.11 22    elastane
3.2  18    polyamide
2.3  16    acetate
3.12 15    polyetylene
```

Each spectra contains 1700 measured datapoints.

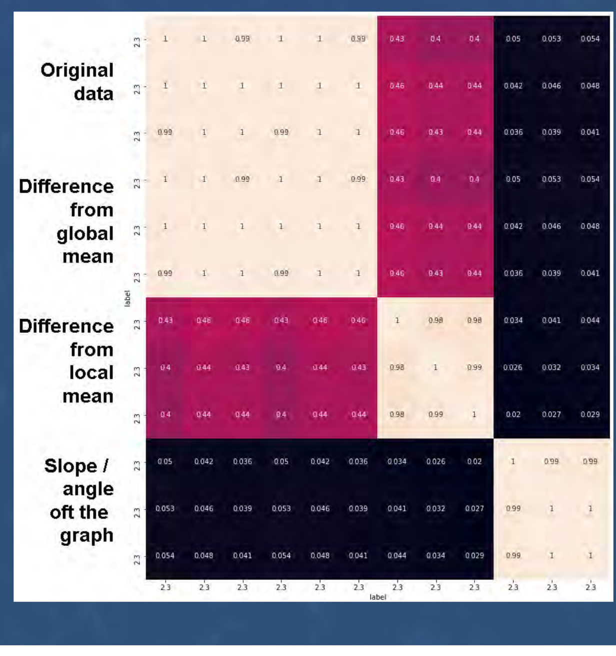
Train data – 30 spectra from each class
Test data – the remaining spectra from classes with more than 30 spectra

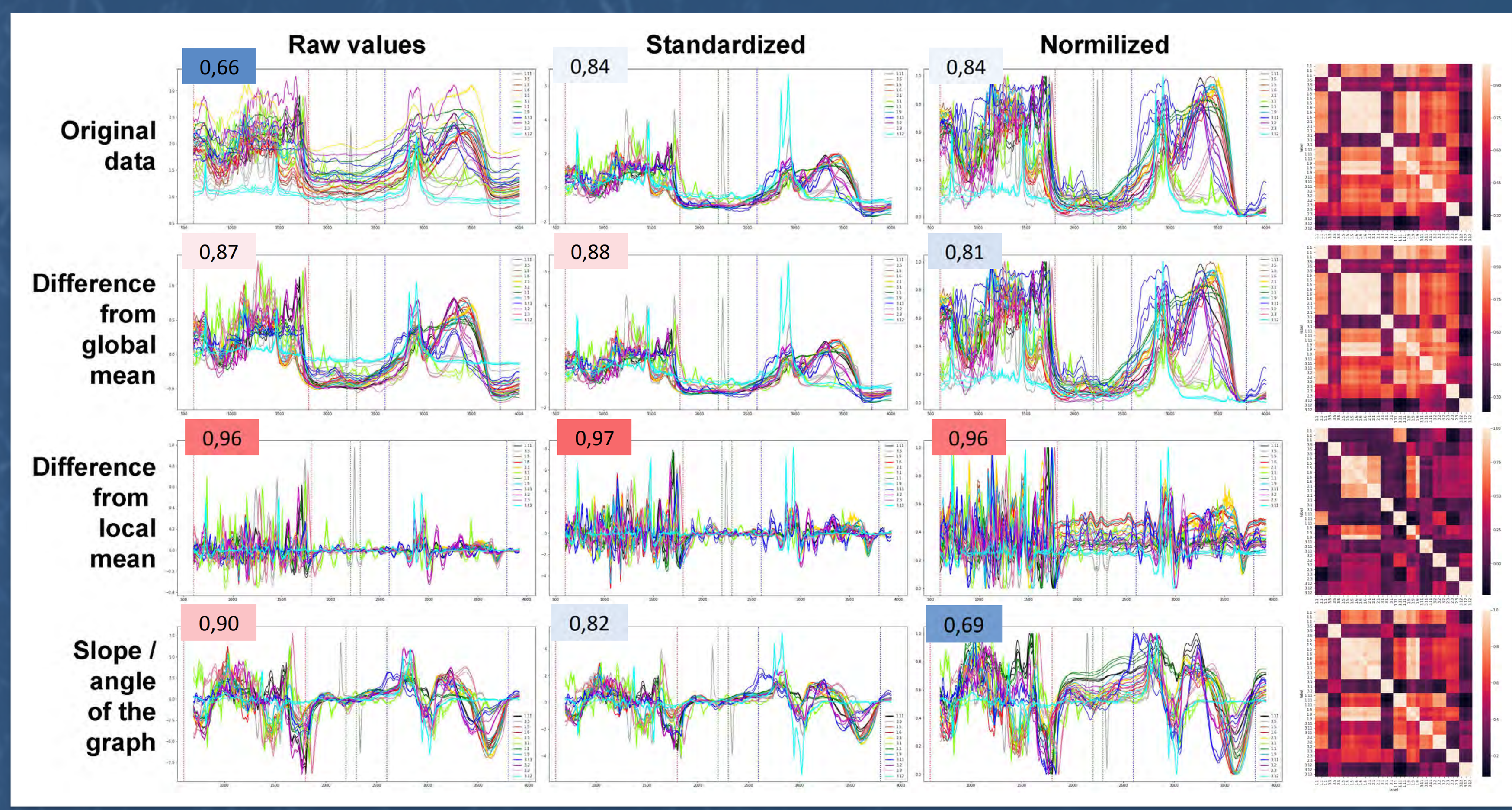


RESULTS

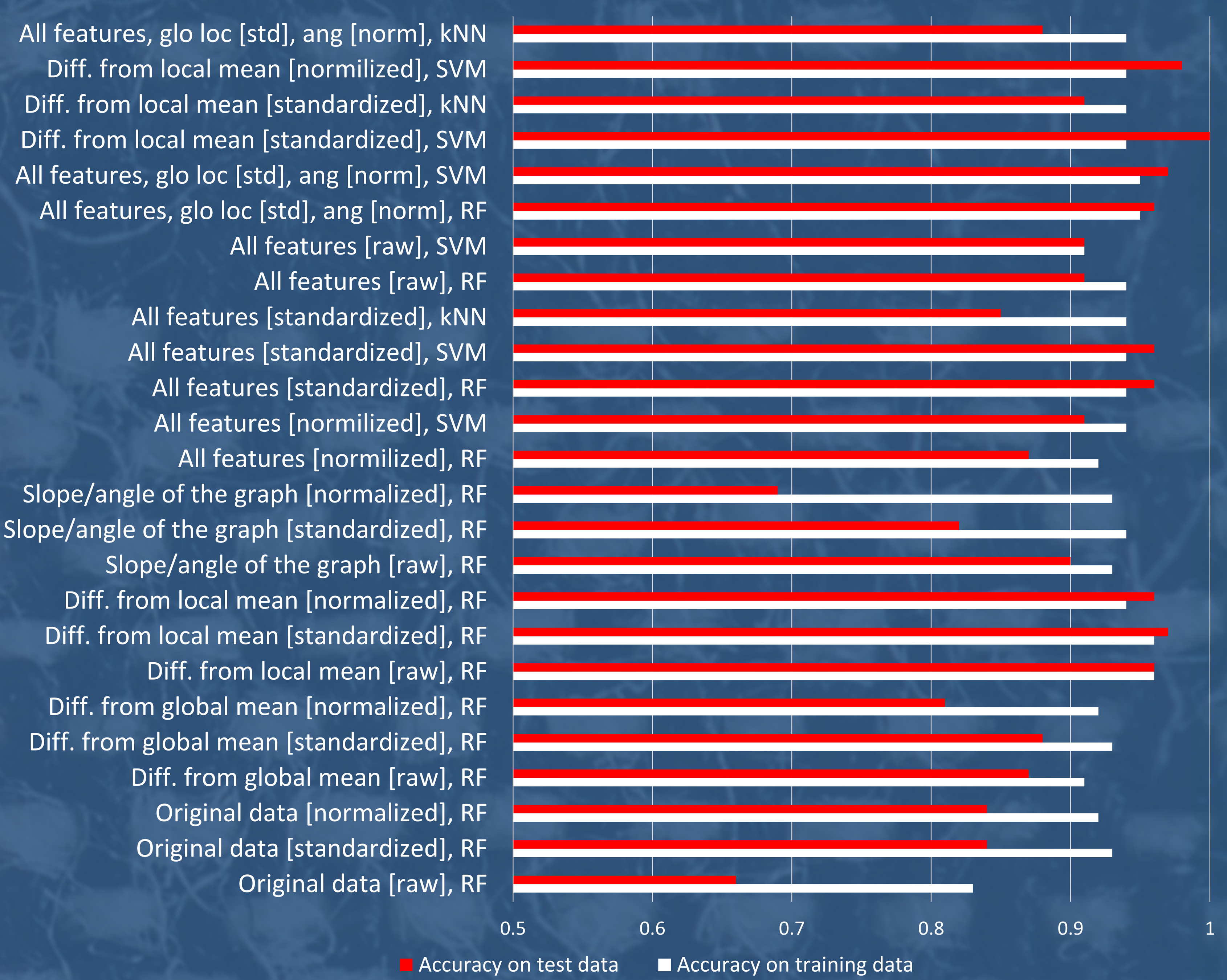
- **Feature Engineering**

Right: Correlations between original data and features.
Down: 3 randomly sampled spectra from all classes and correlations between them. Big numbers show random forest (RF) classification accuracy on test data when using only this feature.





- **Models**



As the amount of data is small the best performing models can be consider more-or-less equal in performance.

Predicted

Actual	**1.5**	**3.1**	**1.6**	**3.5**	**2.1**	**1.11**
1.5	14	0	0	0	0	0
3.1	0	3	0	0	0	0
1.6	3	0	7	0	0	0
3.5	0	0	0	14	0	0
2.1	0	0	0	0	5	0
1.11	0	0	0	0	0	22

Left: Confusion matrix of model All features [normalized], RF for predicting test data. The same model was used the classify textiles (old scarfs) from restorers of Kanuti Gild. It got 3 out of 4 correct and made a mistake by predicting cotton instead of viscose with probabilities of 0.44 vs 0.33, respectively.

CONCLUSIONS

- Quite small amount of data - all best models have similar accuracies
- Feature Engineering helped to improve the classification.
- Most useful feature seems to be difference from local mean.
- Feature engineering could be developed further and additional filtering of data applied to concentrate on areas with more information and thus providing better separation.
- Separating different spectra works very well in general, accuracy above 0.9, but there are still difficulties with more similar fibers: linen, cotton, viscose.

ACKNOW-LEDGEMENTS

We would like to thank researchers Signe Vahur and Pilleriin Peets from the Cultural Heritage research group, University of Tartu, for providing the data as well as domain knowledge throughout the project.