



# Master Seminar WS 18/19

## Flash Präsentationen

Patrick Schäfer

URL: <https://hu.berlin/landnutzung>

# Agenda

- Today:
  - Flash Presentations  $\pm 5$  minutes
    - Thema vorstellen,
    - Querverbindungen erkennen und
    - Beschäftigung mit dem Thema sicherstellen :)
  - Some Dataset Information

# Topics and Groups

Topic	Assigned to (groups of 2-3)
<b>(non-time series) based-Classifiers</b>	Alexej
<b>Whole-Series-based Classifiers</b>	Sedir und Darko
<b>Univariate Shapelet-based Classifiers</b>	Martin und Tim
<b>Univariate Dictionary-based Classifiers</b>	Erik und David
<b>Multivariate Dictionary-based Classifiers</b>	Arik und Markus
<b>Deep Learning Classifiers</b>	Melina und Michael
<b>Ensembles of Core Classifiers</b>	-----

Topic	Approach
<b>(non-time series)</b>	Features/Preprocessing: NDVI, <b>Statistische Werte</b> (Avg, Stddev, Mean, TSFRESH), Sampling, Dimensionality Reduction: PCA, <b>Autoencoder</b> Classifier: Random Forests
<b>Whole-Series</b>	Features/Preprocessing: Imputation (Lineare Interpolation), z-normalization Dimensionality Reduction: NDVI Classifier: Dynamic Time Warping, 1-Nearest Neighbors
<b>Univariate Shapelet</b>	Features/Preprocessing: Shapelets Discovery (Fast, Ultra-Fast, Learning), Imputation (Interpolation), NDVI, normalization, <b>sampling</b> Dimensionality Reduction: NDVI Classifier: SVM, Random Forests, Decision Tree
<b>Univariate Dictionary</b>	Features/Preprocessing: Bag-of-Pattern (SFA), Imputation Dimensionality Reduction: NDVI <b>Feature Selection: Chi-Squared-Test</b> Classifier: BOSS, WEASEL ( <b>Bigrams</b> )
<b>Multivariate Dictionary</b>	Features/Preprocessing: Bag-of-Pattern (SAX), Range Normalization, <b>Time Synchronization, TF-IDF, Bigrams</b> , Imputation, <b>non-overlapping windows (Quartale)</b> Dimensionality Reduction: Moving Averages (PAA) <b>Feature Selection: Chi-Squared-Test</b> Classifier: Logistic Regression, Random Forests, KNeighbors
<b>Deep Learning</b>	Features/Preprocessing: <b>None</b> , Imputation, <b>Offset of missing values</b> Classifier: Convolutional Neural Networks

# Train dataset

- A **massive** land cover pixel time series (TS) dataset
  - 46 geometrically and radio-metrically corrected images taken by FORMOSAT-2
  - Train data: 6 mio pixels TS, 2,4GB
  - Test data (hold-back): 2.6 mio pixels TS, 1,0GB
- In total 3x46 values per pixel time series
  - 46 time stamps between 06.2 and 29.11.2006
  - 3 surface reflectances: Near-Infra-Red, Red, Green
- Contains **missing values** ,?’
- Overall, 24 land cover classes, labelled by experts
- **Note: This data is provided for the class only and it has to be deleted once the seminar is over**

# Class Labels

prairie temporaire is mapped to #0

ble is mapped to #1

pre is mapped to #2

feuillus is mapped to #3

tournesol is mapped to #4

mais ensilage is mapped to #5

jachere is mapped to #6

bati dense is mapped to #7

bati diffus is mapped to #8

friche is mapped to #9

resineux is mapped to #10

sorgho is mapped to #11

pois is mapped to #12

orge is mapped to #13

bati indu is mapped to #14

soja is mapped to #15

eau is mapped to #16

eucalyptus is mapped to #17

colza is mapped to #18

lac is mapped to #19

peupliers is mapped to #20

mais is mapped to #21

graviere is mapped to #22

surface minerale is mapped to #23

# 1-NN ED on a 90/10 Train/Test Split

precision	recall	f1-score	support	
0	0.90	0.90	0.90	49011
1	0.98	0.98	0.98	99923
2	0.89	0.89	0.89	39030
3	0.89	0.92	0.91	39278
4	0.97	0.96	0.96	65796
5	0.95	0.95	0.95	2973
6	0.89	0.90	0.89	83546
7	0.75	0.75	0.75	27832
8	0.79	0.76	0.77	46246
9	0.81	0.83	0.82	33598
10	0.84	0.86	0.85	4259
11	0.97	0.97	0.97	6852
12	0.96	0.95	0.96	3326
13	0.96	0.96	0.96	15144
14	0.72	0.58	0.64	4919
15	0.98	0.98	0.98	7965
16	0.92	0.90	0.91	4444
17	0.71	0.78	0.74	636
18	0.98	0.98	0.98	13646
19	0.98	0.98	0.98	1182
20	0.34	0.38	0.36	34
21	0.99	0.99	0.99	53877
22	0.98	0.98	0.98	3819
23	0.79	0.87	0.83	1768
<b>avg / total</b>	<b>0.91</b>	<b>0.91</b>	<b>0.91</b>	<b>609104</b>

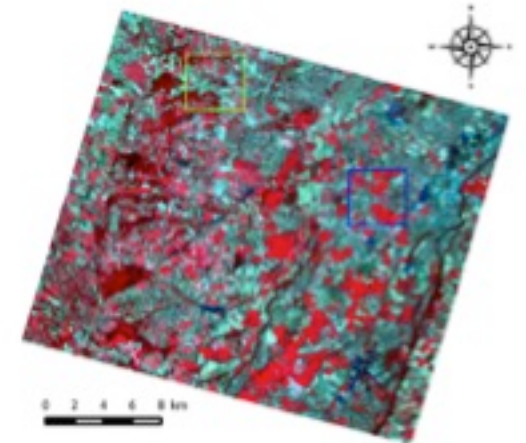
# Random Forest on a 90/10 Train/Test Split

precision	recall	f1-score	support	
0	0.87	0.79	0.83	49011
1	0.97	0.95	0.96	99923
2	0.92	0.72	0.81	39030
3	0.83	0.93	0.88	39278
4	0.95	0.94	0.94	65796
5	0.98	0.92	0.95	2973
6	0.78	0.85	0.81	83546
7	0.70	0.67	0.68	27832
8	0.65	0.74	0.69	46246
9	0.72	0.80	0.75	33598
10	0.92	0.79	0.85	4259
11	0.99	0.91	0.95	6852
12	0.98	0.88	0.93	3326
13	0.98	0.89	0.93	15144
14	0.78	0.42	0.55	4919
15	0.99	0.94	0.96	7965
16	0.91	0.90	0.90	4444
17	0.93	0.56	0.70	636
18	0.98	0.95	0.96	13646
19	0.97	0.98	0.98	1182
20	1.00	0.21	0.34	34
21	0.97	0.98	0.98	53877
22	0.98	0.98	0.98	3819
23	0.82	0.78	0.80	1768
<b>avg / total</b>	<b>0.87</b>	<b>0.86</b>	<b>0.86</b>	<b>609104</b>



# Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series

- The train data might come from this area?
  - The study area is located at the South West of France, near Toulouse city (110E, 4327N).
  - It is 24 km × 24 km area where about 60 % of the soil correspond to arable surfaces.
  - The area has a temperate continental climate with hot and dry summer – average temperature about 22.4 C and rainfall about 38 mm per month.
  - The Figure displays a satellite image of the area



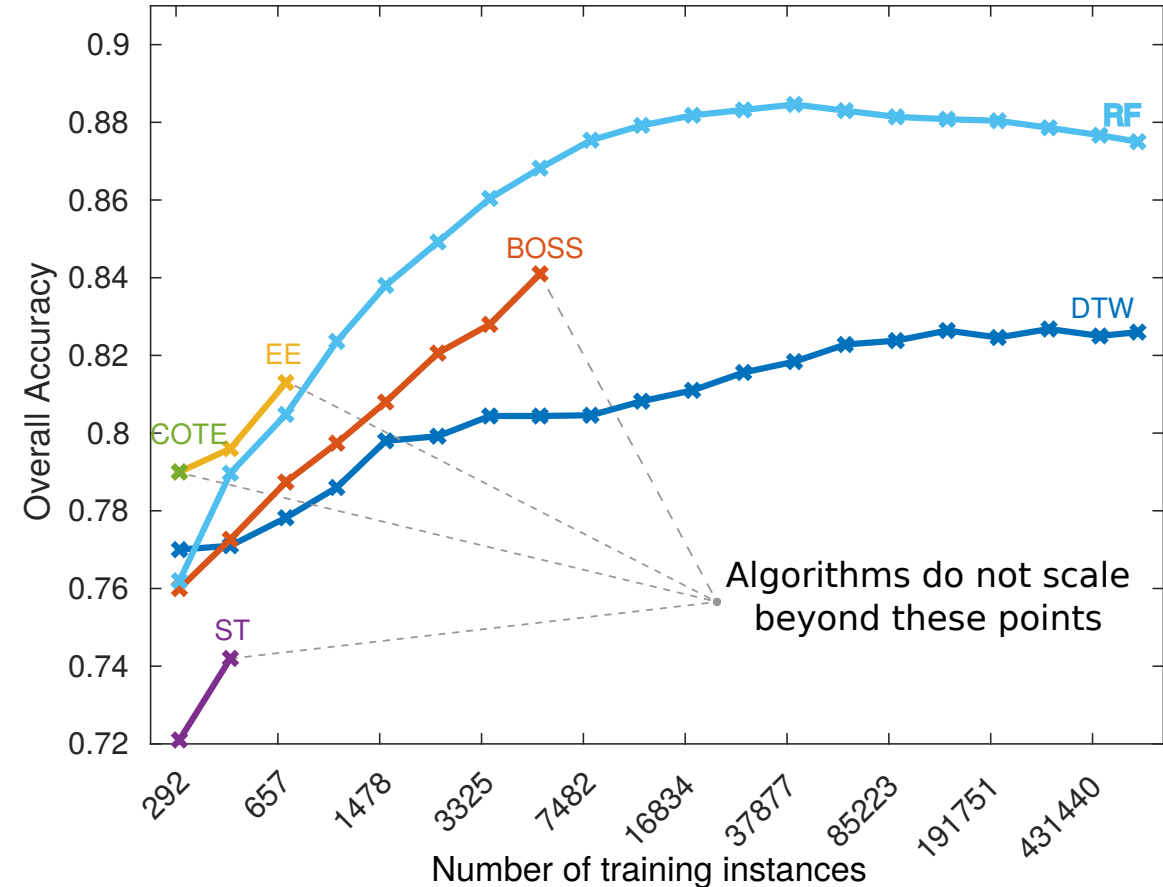
<https://arxiv.org/pdf/1811.10166.pdf>

# Data Preparation

- *Temporal sampling:* [...] We use here a temporal **linear interpolation** for imputing invalid pixel values. [...]
- *Feature extraction:* [...] The contribution of the spectral features is analyzed by adding the three computed spectral indexes (**NDVI**, NDWI and IB) to the spectral bands [...]
- *Feature normalization:*
  - [...] In remote sensing, the input time series are generally standardized by subtracting the **mean and divided by the standard deviation for each feature** where each time stamp is considered as a separate feature [...]
  - [...] **z-normalization** [...] **leads to a loss of the significance of the magnitude** that it is recognized as crucial for vegetation mapping, *e.g.* the corn will have higher NDVI values than other summer crops.

# Accuracy...

- Competitors
  - DTW (warping window size is fixed at 25%)
  - Elastic Ensemble
  - BOSS
  - Shapelet Transform
  - COTE
- Only **NDVI** features
- Only **1000** test samples
- Limit at **24 hours single core** runtime
- Using (inefficient) codes from [www.timeseriesclassification.com](http://www.timeseriesclassification.com)



# Next steps...

- **Please send me your flash presentation slides!**
- Competition (January)
  - You will be given an unlabeled test set
  - We will use an automated evaluation web platform (Kaggle)
  - Submissions will be possible in January
  - Small price for best average accuracy among all groups
- Blockseminar (**1.2.19 15-18 Uhr, RUD 25 4.410**)
  - Before 31.1.19: meet me to discuss slides
  - Present your topic (30-40min) at the Blockseminar
- Seminar Thesis before 31.3.2019!
  - write seminar thesis (~20 pages)