Science Day

Chadi Barakat

2020-04-30

UNIVERSITY OF ICELAND
FACULTY OF INDUSTRIAL ENGINEERING,
MECHANICAL ENGINEERING AND COMPUTER SCIENCE

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

Joint Research Center for Computational Biomedicine | RWTH AACHEN UNIVERSITY

Current State of Affairs - Briefly

- Continued working on ARDS dataset until recently.

- Given the COVID-19 situation:
  - Currently shifted towards a broad spectrum of influenza patients.
  - Attempts to extract relevant features from the dataset.
  - The approaches being tested will be applied to COVID-19 data.

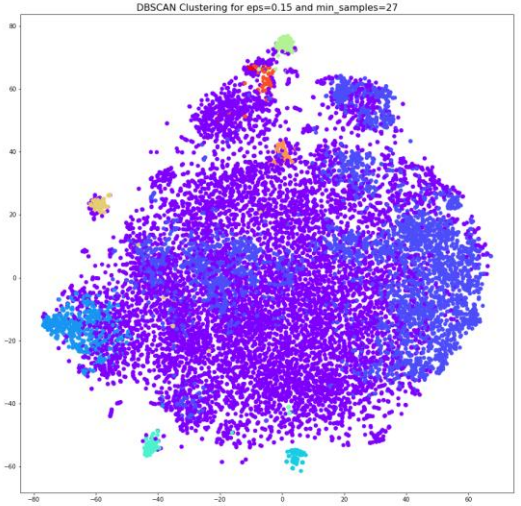
Dataset Description

- First dataset comprises baseline values of clinical data.
  - 17906 patients.
  - A separate file lists comorbidities of these patients.
  - Patients belong to 5 different cohorts.

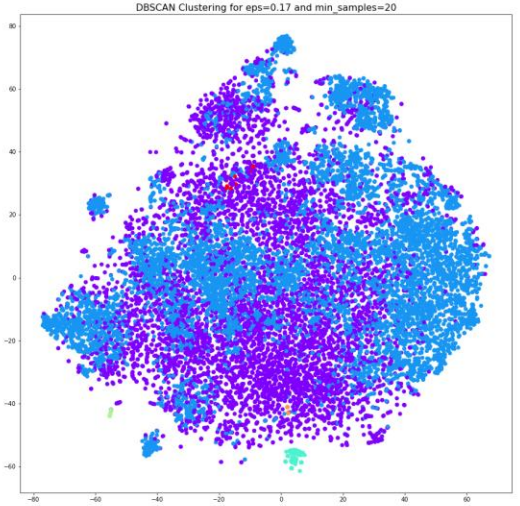
- Second dataset lists min, mean, max, median, and std values for several features of 16794 patients 8 hours after admission.

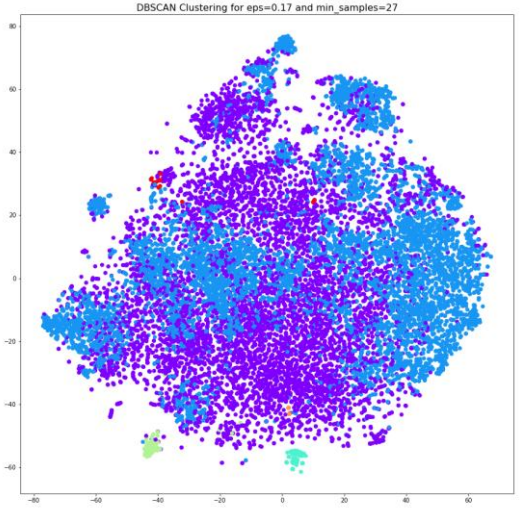
Approach

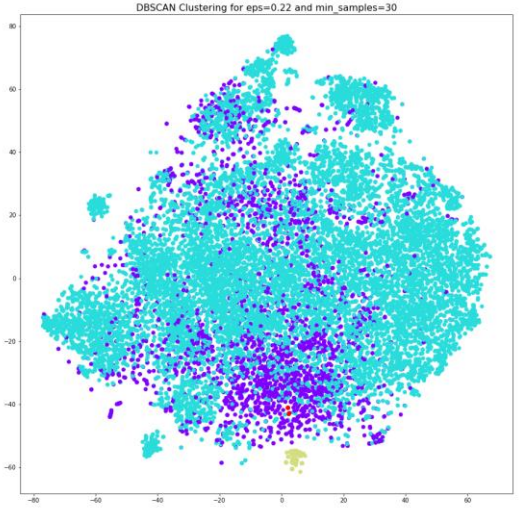
- Reduce number of features by removing components with more than 20% missing values:
  - Baseline dataset: 35 -> 24 features.
  - 8-hours dataset: 305 -> 70 features.
- Scale and apply PCA:
  - Baseline dataset: 24 -> 17 features.
  - 8-hours dataset: 70 -> 20 features.
- Find the best parameter combination for DBSCAN clustering.

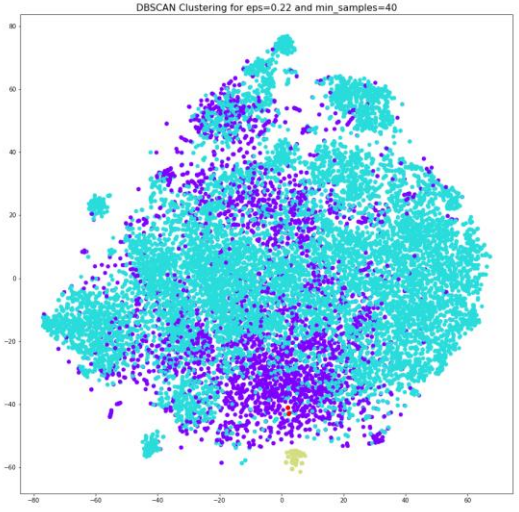


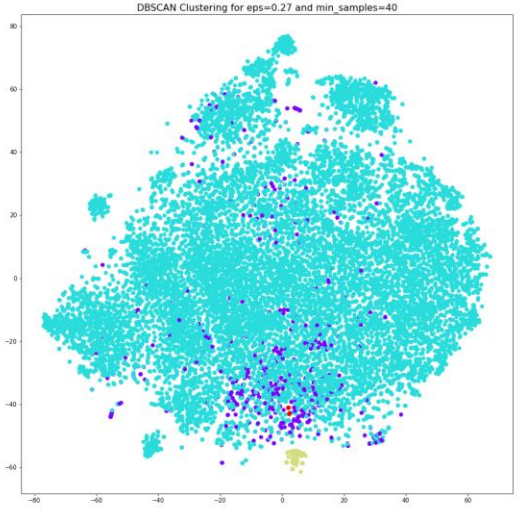
Parameter selection in DBSCAN







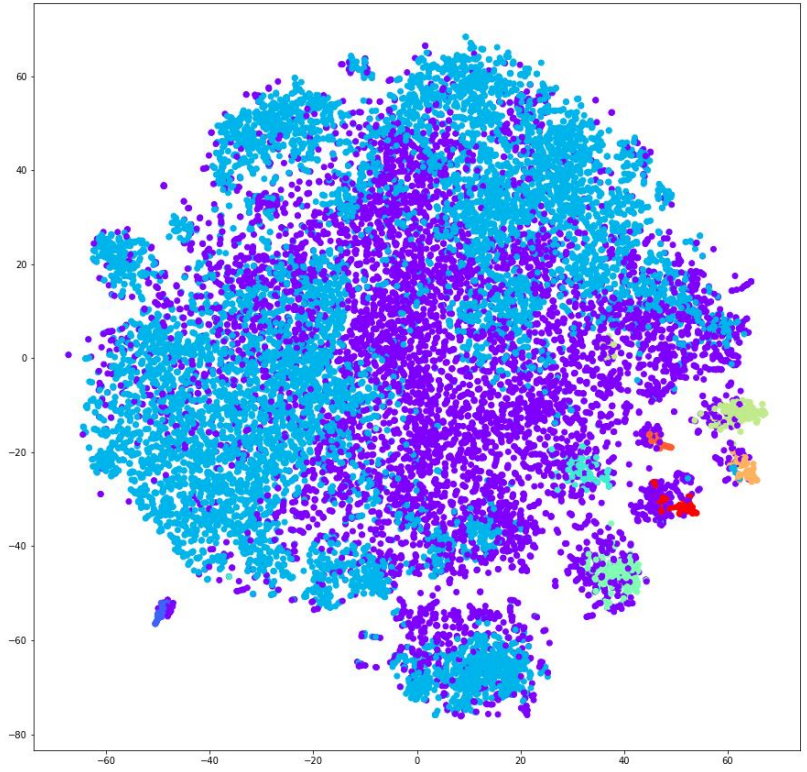


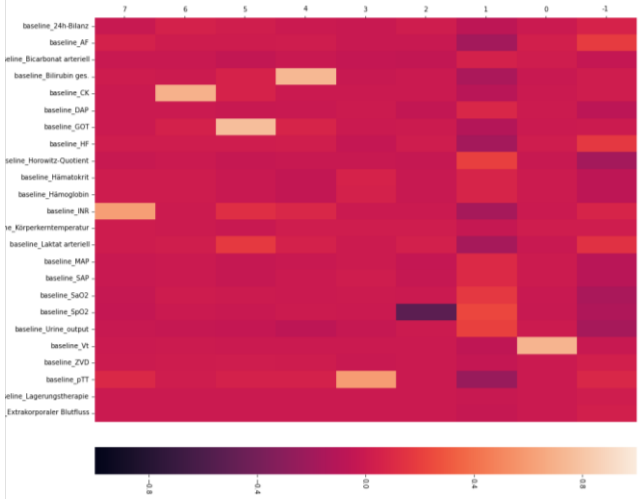


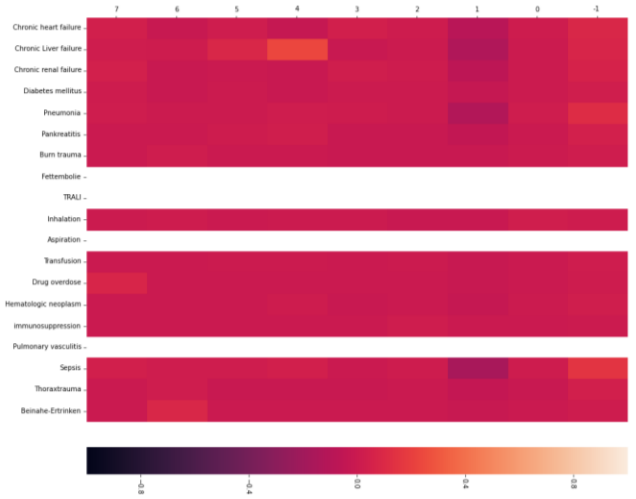
Clustering and Correlation Analysis

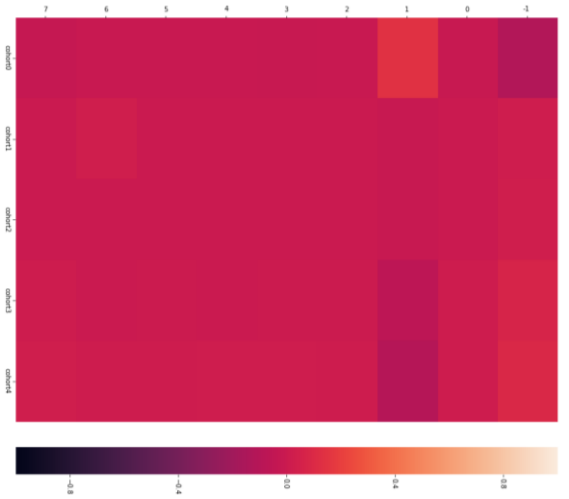
- Selected one potential combination of parameters that produces an acceptable number of clusters, without too many points being considered as noise.

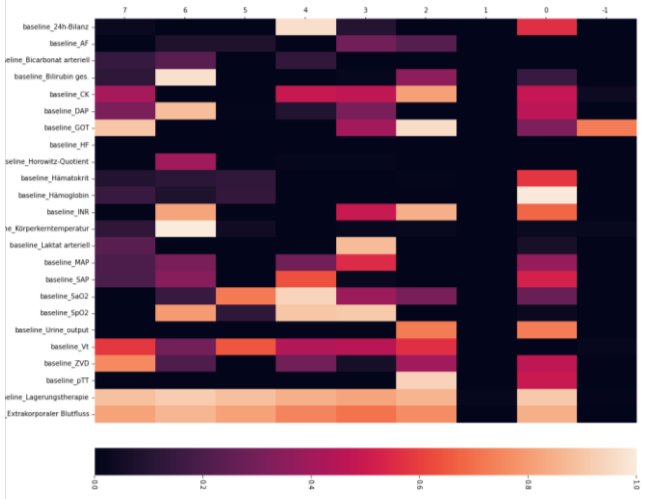
- Analysed correlation between clusters and the features, the cohorts, and the comorbidities.

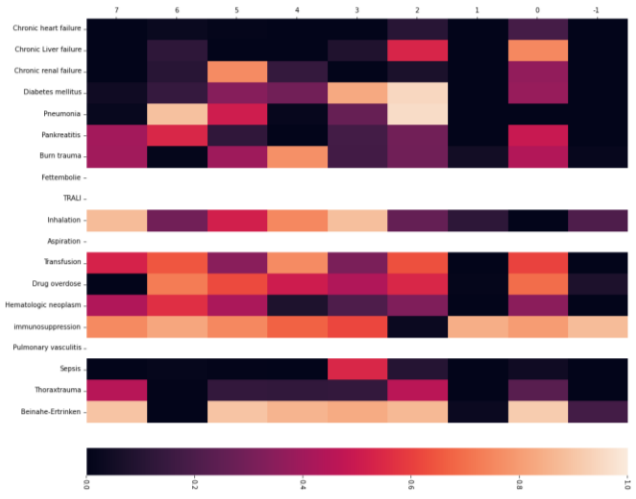


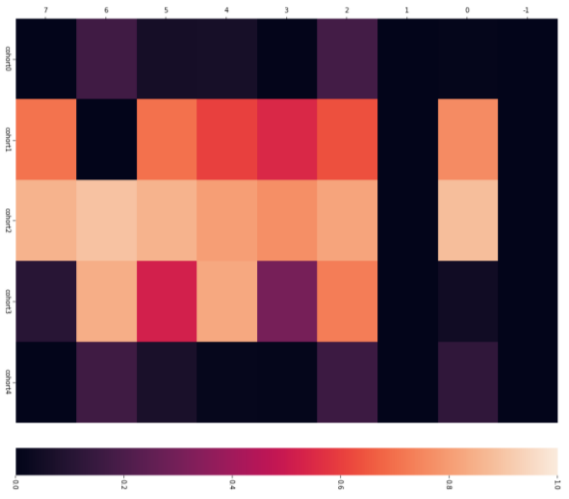

7 6 5 4 3 2 1 0 -1
baseline_24h-Bilanz
baseline_AF
eline_Bicarbonat arteriell
baseline_Bilirubin ges
baseline_CK
baseline_DAP
baseline_GOT
baseline_HF
seline_Horowitz-Quotient
baseline_Hämatokrit
baseline_Hämoglobin
baseline_INR
e_Körperkerntemperatur
baseline_Laktat arteriell
baseline_MAP
baseline_SAP
baseline_SaO2
baseline_SpO2
baseline_Urine_output
baseline_Vt
baseline_ZVD
baseline_pTT
eline_Lagerungstherapie
_Extrakorporaler Blutfluss
-0.8 -0.4 -0.0 -0.4 -0.8


7 6 5 4 3 2 1 0 -1
Chronic heart failure
Chronic Liver failure
Chronic renal failure
Diabetes mellitus
Pneumonia
Pankreatitis
Burn trauma
Fettembolie
TRALI
Inhalation
Aspiration
Transfusion
Drug overdose
Hematologic neoplasm
immunosuppression
Pulmonary vasculitis
Sepsis
Thoraxtrauma
Beinahe-Ertrinken
-0.8 -0.4 -0.0 -0.4 -0.8


7 6 5 4 3 2 1 0 -1
cluster0
cluster1
cluster2
cluster3
cluster4
-0.8 -0.4 -0.0 -0.4 -0.8


7 6 5 4 3 2 1 0 -1
baseline_24h-Bilanz
baseline_AF
eline_Bicarbonat arteriell
baseline_Bilirubin ges
baseline_CK
baseline_DAP
baseline_GOT
baseline_HF
seline_Horowitz-Quotient
baseline_Hämatokrit
baseline_Hämoglobin
baseline_INR
e_Körperkerntemperatur
baseline_Laktat arteriell
baseline_MAP
baseline_SAP
baseline_SaO2
baseline_SpO2
baseline_Urine_output
baseline_Vt
baseline_ZVD
baseline_pTT
eline_Lagerungstherapie
_Extrakorporaler Blutfluss
0.0 0.2 0.4 0.6 0.8 1.0


7 6 5 4 3 2 1 0 -1
Chronic heart failure
Chronic Liver failure
Chronic renal failure
Diabetes mellitus
Pneumonia
Pankreatitis
Burn trauma
Fettembolie
TRALI
Inhalation
Aspiration
Transfusion
Drug overdose
Hematologic neoplasm
immunosuppression
Pulmonary vasculitis
Sepsis
Thoraxtrauma
Beinahe-Ertrinken
0.0 0.2 0.4 0.6 0.8 1.0


7 6 5 4 3 2 1 0 -1
cluster0
cluster1
cluster2
cluster3
cluster4
0.0 0.2 0.4 0.6 0.8 1.0

Next Steps

- Select the best parameters for clustering the 8-hours data and attempt correlation analysis.
- Compare clustering and correlation results with other methods.
- Suggestions from colleagues.