



Science Day

Chadi Barakat

2020-07-15



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



JÜLICH
SUPERCOMPUTING
CENTRE



RWTHAACHEN
UNIVERSITY



Current State of Affairs - Briefly

- Summary of previous discussion:
 - Attempting to use DBSCAN clustering and correlation analysis to uncover and highlight potentially relevant features in the patient data and their relationship with the diagnosis.
 - Potentially can implement other clustering algorithms and implement more adequate correlation analysis for the results.



Clustering Approaches

- A lot of work went into defining parameters for DBSCAN, which proved ineffective as most resulting clusters were dominated by noise points.
- Other clustering methods produced less noisy clusters, but there was limited correlation between these clusters and the diagnoses.
- Relevant features were confirmed from work done by Konstantin.



Attempting to Fine-tune Pre-Processing

- Scaling-only approach vs. Scaling+Normalising approach
- StandardScaler() vs. RobustScaler()

```
[318]: stand_norm.max()

[318]: baseline_24h-Bilanz          0.371474
baseline_AF spontan              0.325837
baseline_AF                      0.346078
baseline_Bicarbonat arteriell    0.291480
baseline_Bilirubin ges.         0.609811
baseline_CK                      0.753878
baseline_DAP                     0.290594
baseline_GOT                     0.620258
baseline_HF                      0.282479
baseline_Horowitz-Quotient (ohne Temp-Korrektur) 0.339453
baseline_Hämatokrit             0.263132
baseline_Hämoglobin             0.316049
baseline_INR                    0.628135
baseline_Körperkerntemperatur   0.240822
baseline_Laktat arteriell       0.340680
baseline_MAP                     0.271720
baseline_P EI                   0.374236
baseline_PEEP                   0.553483
baseline_SAP                     0.303057
baseline_SaO2                   0.216755
baseline_SpO2                   0.215368
baseline_Thrombozyten           0.377077
baseline_Urine_output           0.373356
baseline_Vt                     0.783373
baseline_ZVD                    0.285643
baseline_individuelles Tidalvolumen pro kg idealem Körpergewicht 0.332894
baseline_pTT                    0.326224
baseline_Lagerungstherapie      1.000000
baseline_ECMO                   0.210785
dtype: float64
```

```
[319]: robust_norm.max()

[319]: baseline_24h-Bilanz          0.131081
baseline_AF spontan              0.154806
baseline_AF                      0.141584
baseline_Bicarbonat arteriell    0.121204
baseline_Bilirubin ges.         0.283344
baseline_CK                      0.577615
baseline_DAP                     0.117916
baseline_GOT                     0.873541
baseline_HF                      0.117236
baseline_Horowitz-Quotient (ohne Temp-Korrektur) 0.126657
baseline_Hämatokrit             0.112633
baseline_Hämoglobin             0.120527
baseline_INR                    0.297738
baseline_Körperkerntemperatur   0.116101
baseline_Laktat arteriell       0.152312
baseline_MAP                     0.115549
baseline_P EI                   0.134399
baseline_PEEP                   0.158613
baseline_SAP                     0.118926
baseline_SaO2                   0.106125
baseline_SpO2                   0.105660
baseline_Thrombozyten           0.133722
baseline_Urine_output           0.128724
baseline_Vt                     1.000000
baseline_ZVD                    0.122134
baseline_individuelles Tidalvolumen pro kg idealem Körpergewicht 0.122674
baseline_pTT                    0.151410
baseline_Lagerungstherapie      0.107773
baseline_ECMO                   0.105660
dtype: float64
```

```
[321]: icca_patients.baseline_Lagerungstherapie.describe()

[321]: count    17908.000000
mean         0.000223
std          0.014944
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max          1.000000
Name: baseline_Lagerungstherapie, dtype: float64
```

```
[328]: len(icca_patients[icca_patients.baseline_Lagerungstherapie.eq(0)])
```

```
[328]: 17904
```



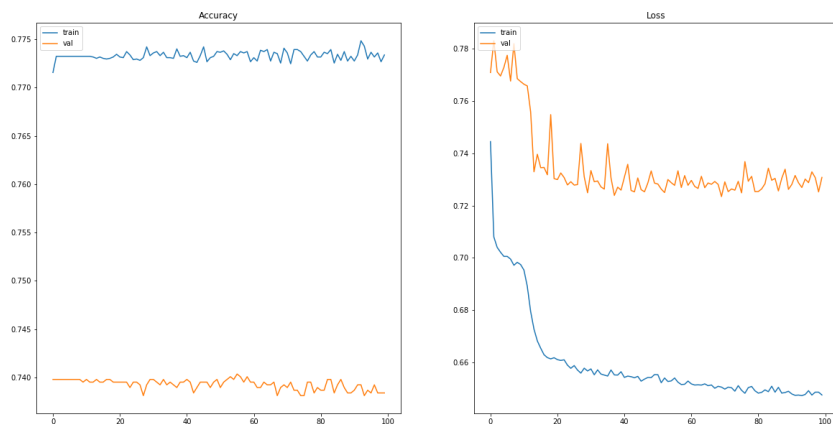
Using Neural Networks

- Two main datasets:
 - ICCA baseline values for 17888 patients with 27 features.
 - ICCA mechanically ventilated patients (13461 patients) with 234 features.
- 3- or 4-layer CNN with either Softmax or Sigmoid activation function at the output layer.
- Optimising for cohorts (0-5) using categorical crossentropy

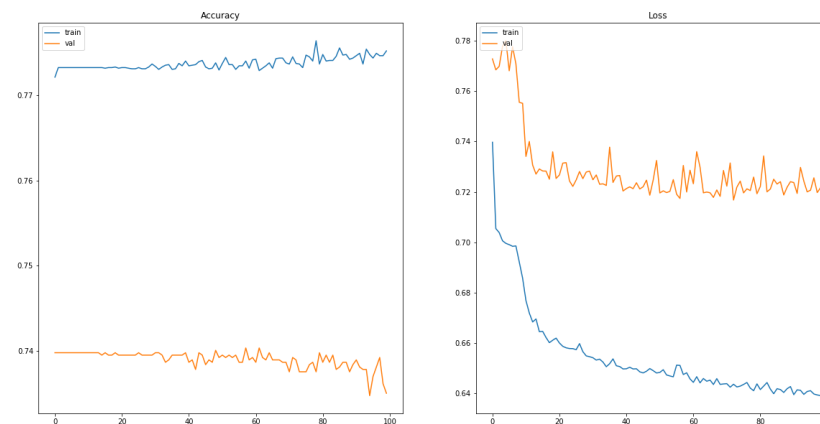


ICCA Baseline Patients

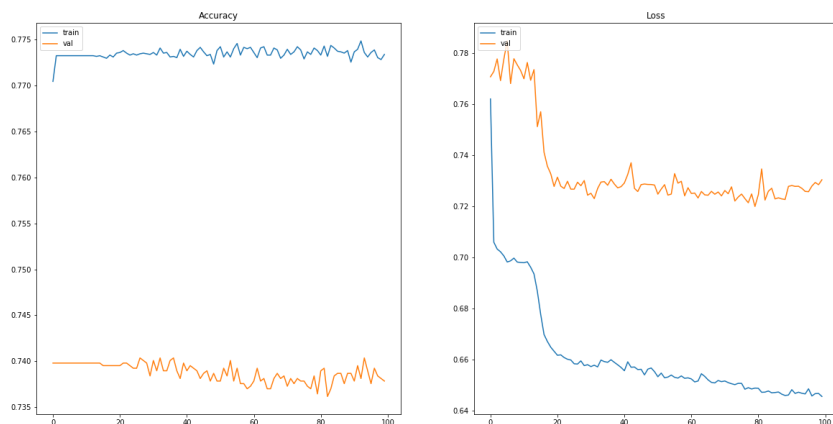
ICCA patients, 3xCONV1D, softmax



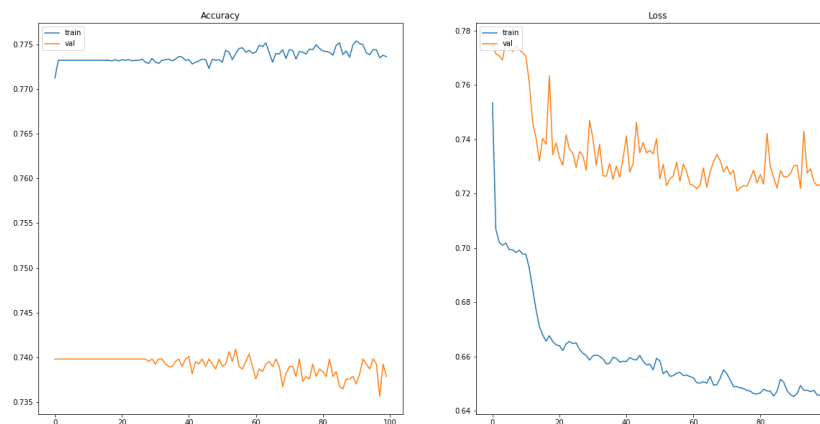
ICCA patients, 4xCONV1D, softmax



ICCA patients, 3xCONV1D, sigmoid



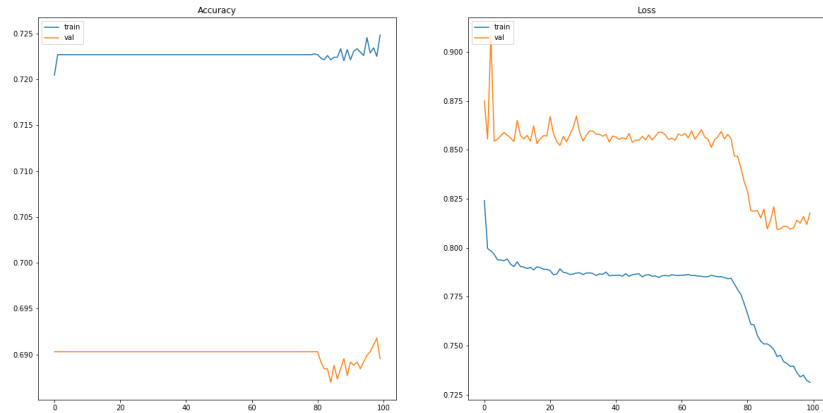
ICCA patients, 4xCONV1D, sigmoid



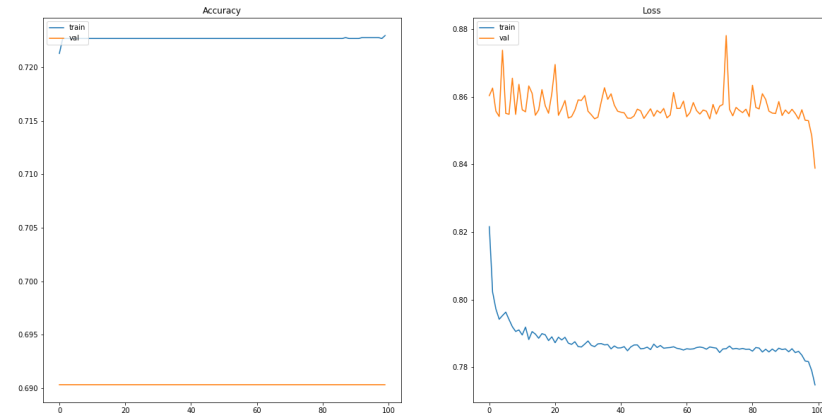


Mechanically-Ventilated Patients

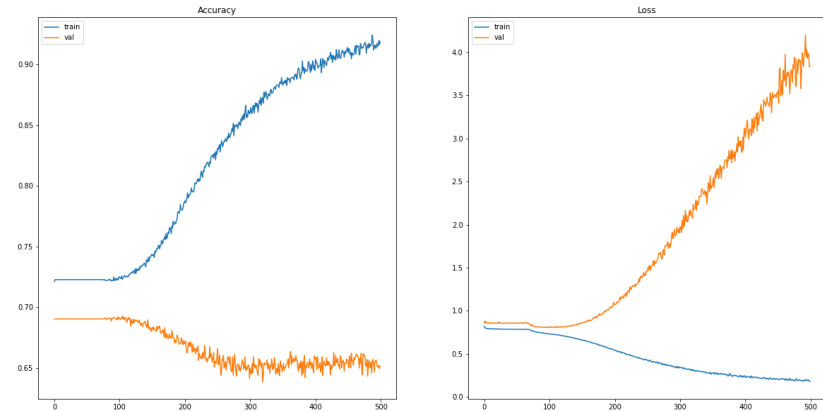
MV patients, 3xCONV1D, softmax



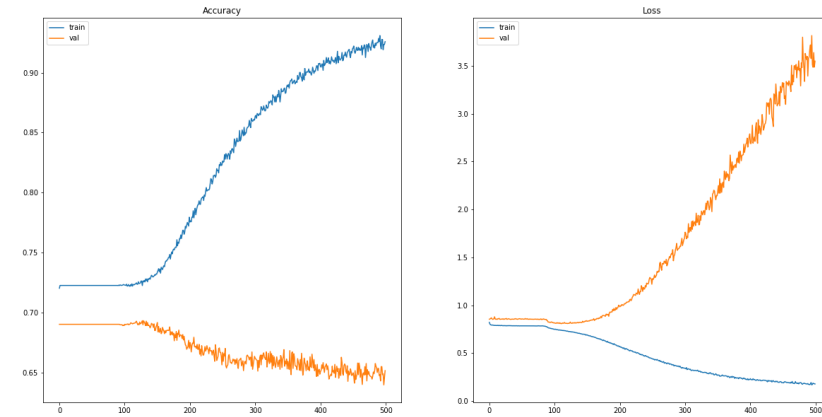
MV patients, 3xCONV1D, sigmoid



MV patients, 3xCONV1D, softmax, 500 epochs



MV patients, 3xCONV1D, sigmoid, 500 epochs





Conclusions

- Cohorts are not balanced with the majority (~76% of patients) belonging to the same cohort.
 - Next step: use a more balanced training set.
- Parameters may not be as indicative of original condition as they are of potential direction of disease.
 - Next (current) step: attempt a mortality prediction (in parallel with Moein).
- Currently only working with ICCA data until positive results or change of approach.
 - Next step: attempt analysis on COVID data.