CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN
BEUTH HOCHSCHULE
FÜR TECHNIK
BERLIN
University of Applied Sciences

Prediction of Permanent Kidney Graft Loss via Deep Ensemble Learning and NLP using a biased Dataset of Electronic Health Records

Master Thesis Defense - Rouven Reuter - 26.7.2021

Supervisor / Surveyor:
Prof. Dr. Alexander Löser
Prof. Dr. Petra Sauer
Prof. Dr. Henrik Tramberend
Dipl.-Ing. Jens-Michalis Papaioannou
Dr. Marcel Naik

Content

Introduction

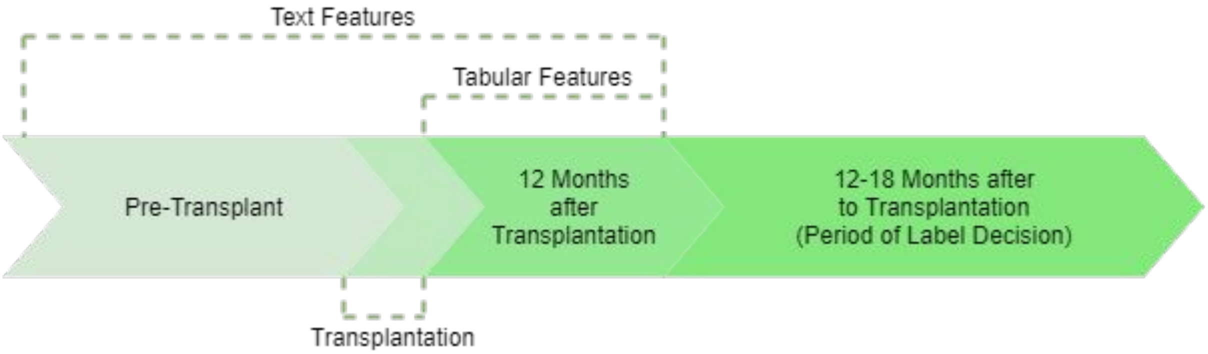
- Acute Kidney Injury (AKI) is the loss of kidney function within a short period of time. [1]
- This work focuses on the last stage of an AKI graft loss
- Patients that underwent a kidney transplant are susceptible to AKI’s [2]
- Deep Learning can help identifying the patients that are likely to develop an AKI.
- A database consisting of patients medical records that underwent transplantation was provided by the Charite of Berlin.
- The database features tabular data as well as unstructured Text.
- As baseline a random forest model was used [3]

Hypothesis

1. That medical text holds information complementary to the tabular data of the patient and will therefore, by combining the two data types, improve the performance of the model.
2. A MLP model using a balancing technique will outperform a RF model on the same imbalanced dataset.
3. Further pre-training BERT with AKI risk-factors will improve graft loss prediction on clinical notes.

Problem Definition

Problem Definition



Problem Definition - Challenges

- As baseline a Random Forest model using the tabular data was chosen.
- The dataset is strongly biased towards the success class.
- The text corpus is unstructured and drafted by multiple people.
- The data needs to be cleaned and prepared for machine learning.
- High dimensional input vector in comparison to the amount of examples.

Data Description

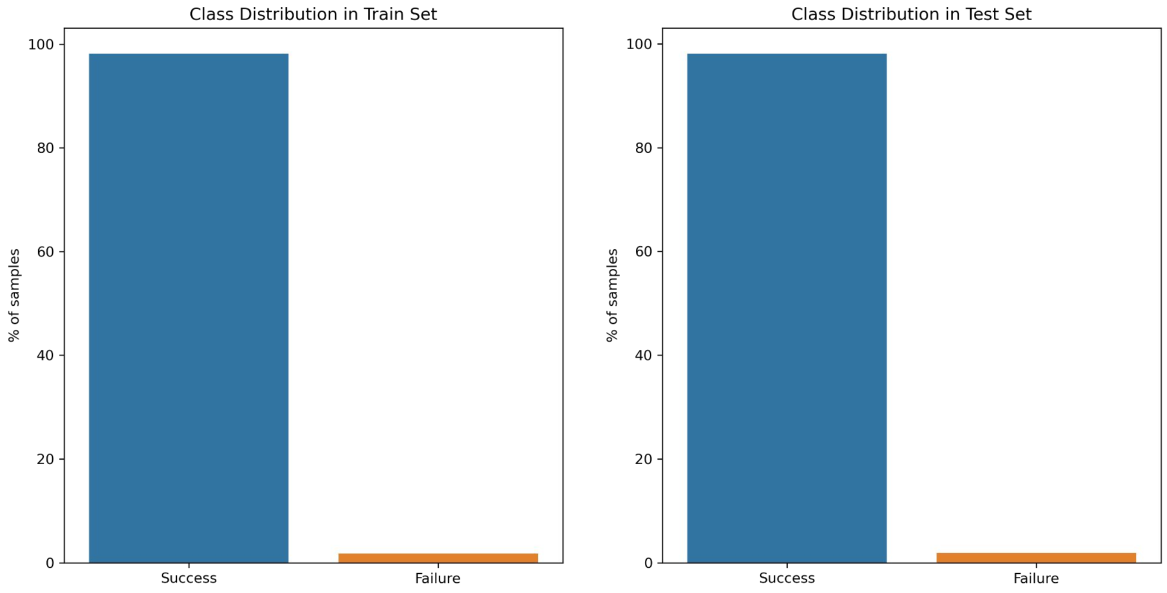
Data Description - Cohort Selection

- 19.642 Patients kidney transplant patients were found in the database of which 1263 were suited for the project.

  Cohort criteria

- Only patients with one or more kidney transplants, but no other kind of organ transplant.
- Discard all patients younger than 18
- Transplants before the year 2000 were not included
- The transplant must have been conducted at the location of Charité Mitte or Virchow
- Patients who had a failing transplant within 12 months after transplantation were not considered

Data Description - Data balance



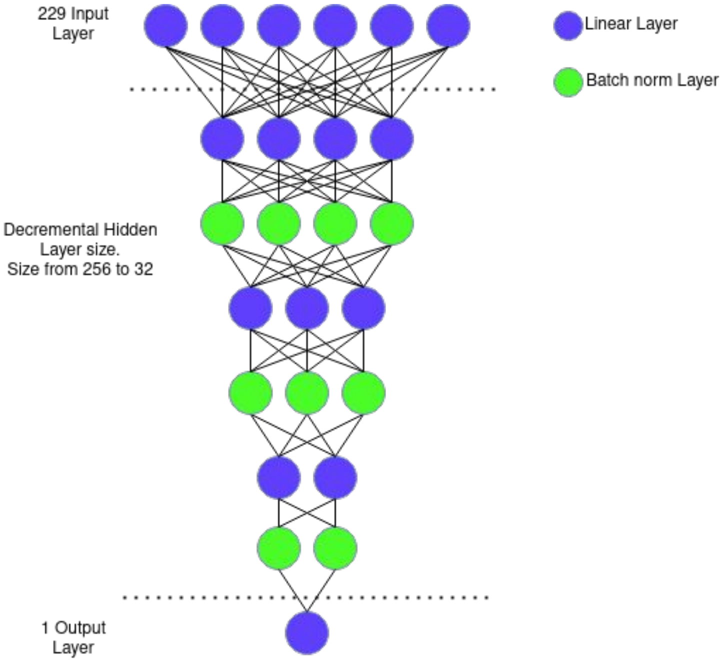
Methodology

Methodology

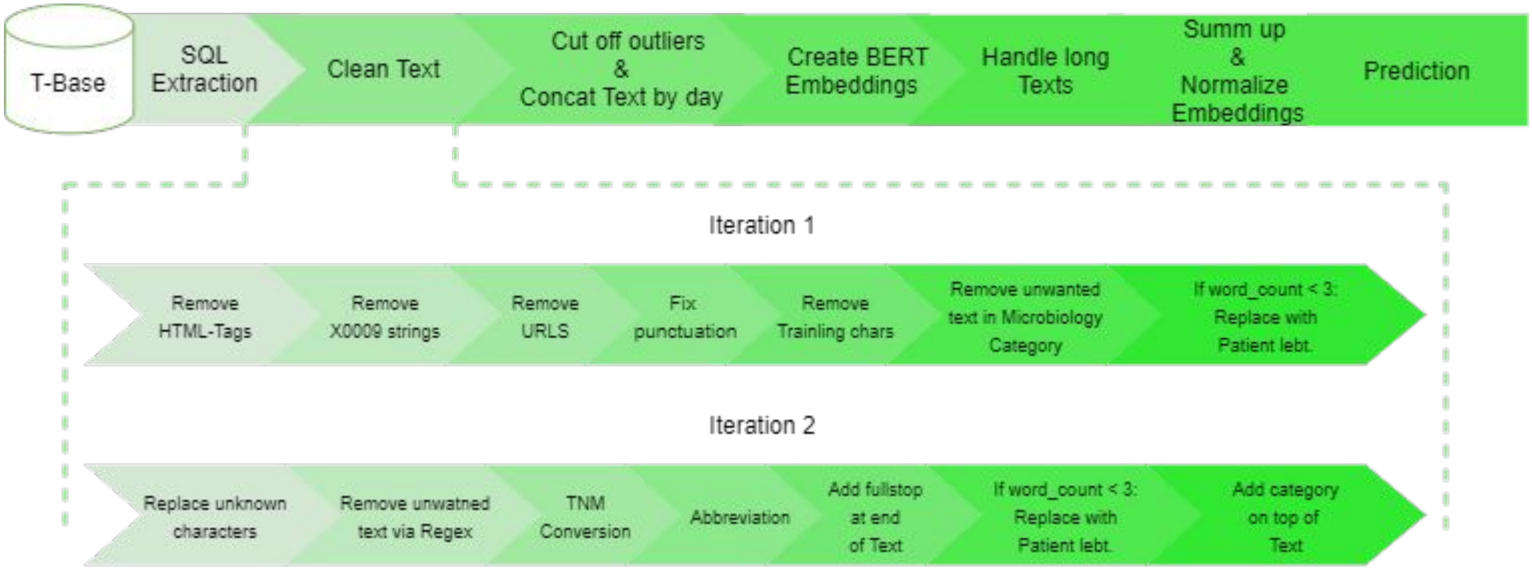
- To find the best suiting balancing technique we conducted experiments in which the only changing parameter was the balancing technique.
- To find the best combination of BERT model and balancing technique, we conducted the same kind of experiment
- All parameters of the final models were found using a hyper parameter optimisation
- Important to note is that the text data was produced in two iterations
- Produced three models one for tabular data, one for text and a combination.

Methodology - Tabular Model

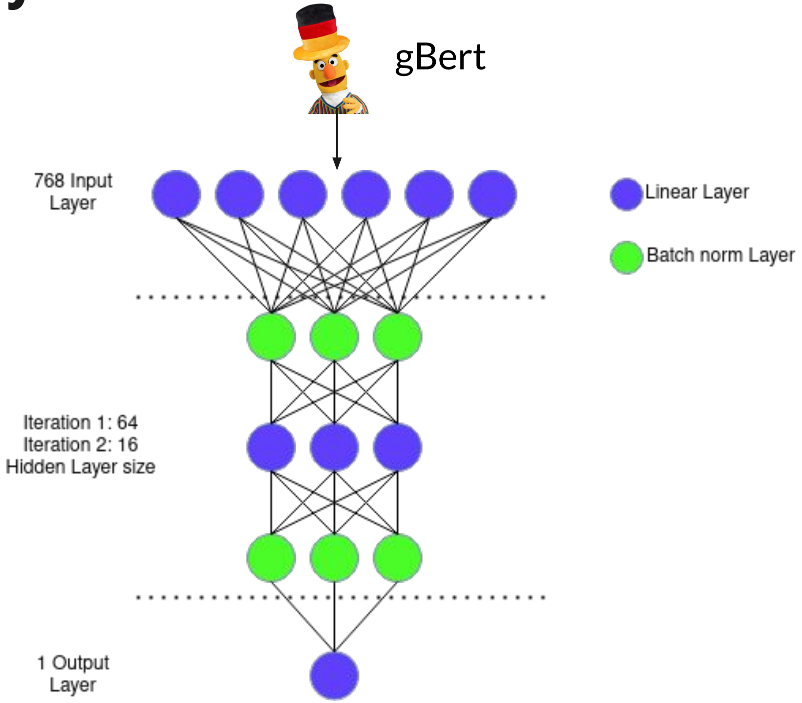
Methodology - Tabular Model



Methodology - Text Model



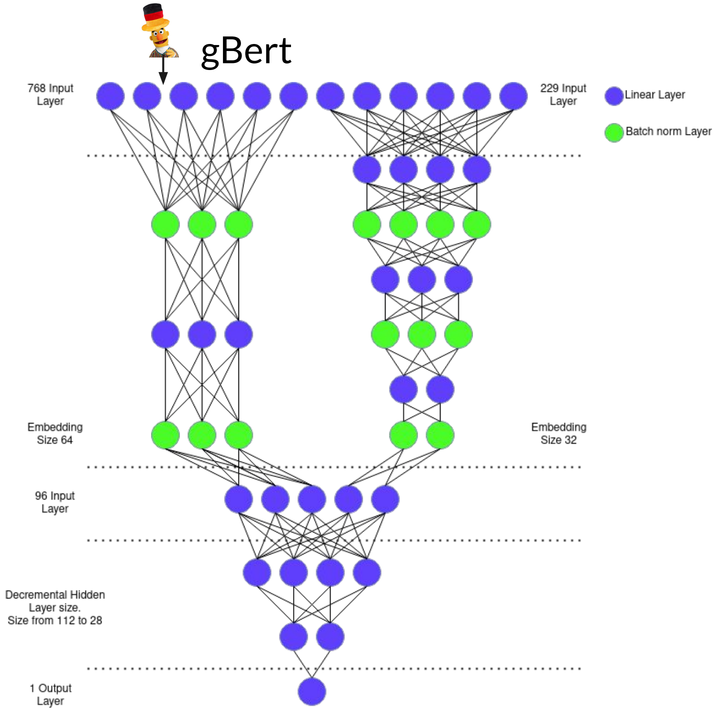
Methodology - Text Model



Weighted Random Sampler + Class weighting

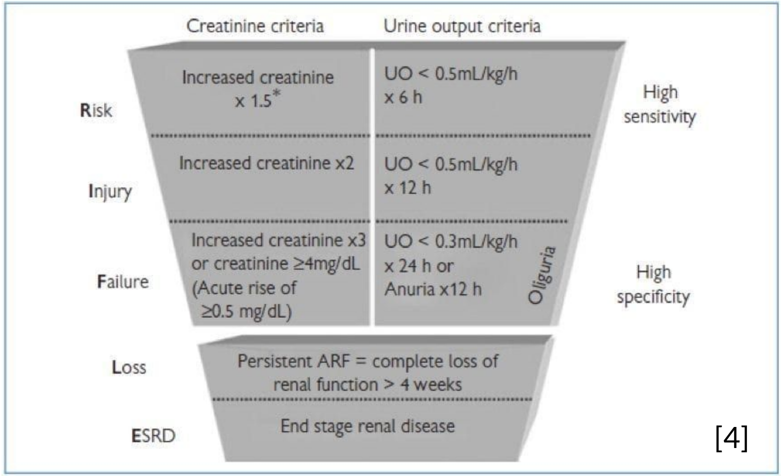
Methodology - Ensemble Model

Weighted Random Sampler + Class weighting



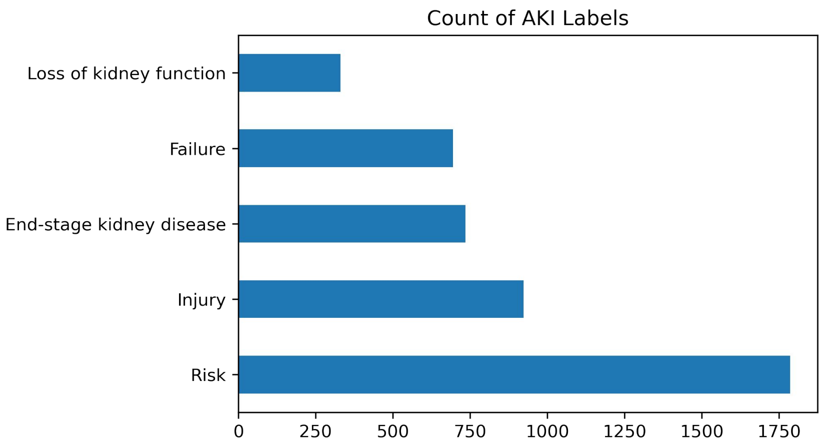
Methodology - BERT Pretraining

- Text found in the database were found almost daily.
- However, our label chosen for the main task would be suited to label daily examples
- To tackle this problem, we choose to create new labels from the AKI RIFLE risk score.
- The labels will still be related to the main task and further will reflect the AKI risk in a text more granular.



[4]

Methodology - BERT Pretraining



Results

Results - Main Task

Metrics	RF Baseline [26]	MLP Decremental Tabular Simple	MLP Homogeneous Text WRS+CW*	Ensemble WRS+CW
Recall	0.55	0.61	**0.75**	0.69
Precision	0.52	0.54	0.52	**0.57**
Accuracy	**0.94**	0.93	0.65	**0.94**
F1-Score	0.53	0.55	0.43	**0.59**
F2-Score	0.54	0.6	**0.69**	0.66
AUC	0.55	0.61	**0.75**	0.69
MCC	0.06	0.13	0.14	**0.22**

CW - Class Weighting   WRS - Weighted Random Sampler
MLP - Mulit-Layer perceptron

* Iteration one without a pretrained BERT model

Results - BERT Pretraining

	Metrics	Text Iteration 1	Text Iteration 2	Text Iteration 2 + Fine-Tune
	Recall	**0.75**	0.62	0.63
	Precision	0.52	0.56	**0.58**
	Accuracy	0.65	0.95	**0.96**
MLP	F1-Score	0.43	0.57	**0.60**
(Homogeneous)	F2-Score	**0.69**	0.61	0.62
	AUC	**0.75**	0.62	0.63
	MCC	0.14	0.17	**0.21**

Evaluation of Text Model Results. All models used the same balancing technique WRS+CW. However, different text features have been used and pre-training was applied.

Conclusion

Conclusion - Hypothesis 1

That medical text holds information complementary to the tabular data of the patient and will therefore, by combining the two data types, improve the performance of the model.

	MLP Decremental Tabular Simple	MLP Homogeneous Text WRS+CW	Ensemble WRS+CW
F1-Score	0.55	0.43	0.59

Conclusion - Hypothesis 2

A MLP model using a balancing technique will outperform a RF model on the same imbalanced dataset.

	RF Baseline [26]	MLP Decremental Tabular Simple
F1-Score	0.53	0.55

Conclusion- Hypothesis 3

Pre-training BERT with AKI risk-factors will improve graft loss prediction on clinical notes.

	Metrics	Text Iteration 1	Text Iteration 2	Text Iteration 2 + Fine-Tune
	Recall	**0.75**	0.62	0.63
	Precision	0.52	0.56	**0.58**
	Accuracy	0.65	0.95	**0.96**
MLP (Homogeneous)	F1-Score	0.43	0.57	**0.60**
	F2-Score	**0.69**	0.61	0.62
	AUC	**0.75**	0.62	0.63
	MCC	0.14	0.17	**0.21**

Future outlook

- Upgrade ensemble model with pre-trained BERT
- Fixing the tabular data pipeline
- Improve pre-training approach by using AKAIN risk factors
- Use produced labels as input for the tabular model
- Use LSTM to give the model a sense of time
- Investigate drop of recall in text iterations

Sources

[1] - Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury

[2] - Influence of delayed graft function and acute rejection on outcomes after kidney transplantation from donors after cardiac death

[3] - Artificial Neural Network for the prediction of short-term graft failure of transplanted kidneys

[4] - Acute Kidney Injury - Chris Nickson

Questions?