Human Age Prediction Based on Real and Simulated RR Intervals using Temporal Convolutional Neural Networks and Gaussian Processes

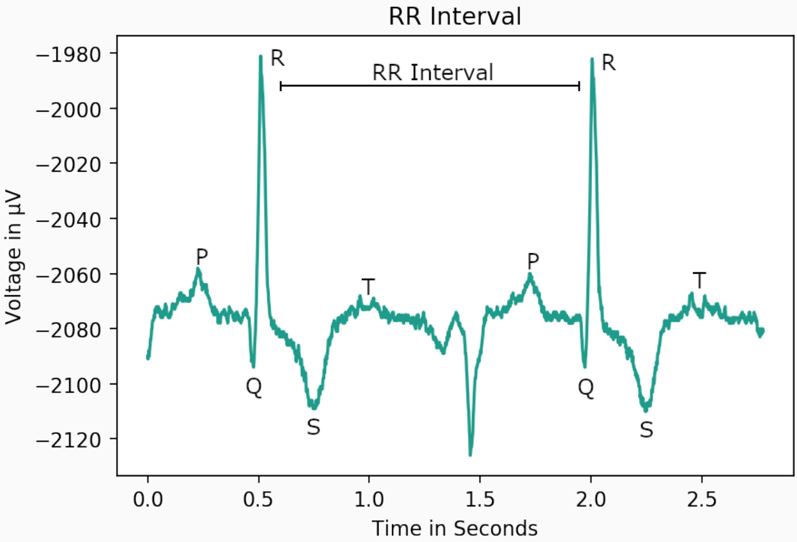
Maximilian Pfundstein (maxpf364)
June 2, 2020
Linköpings University

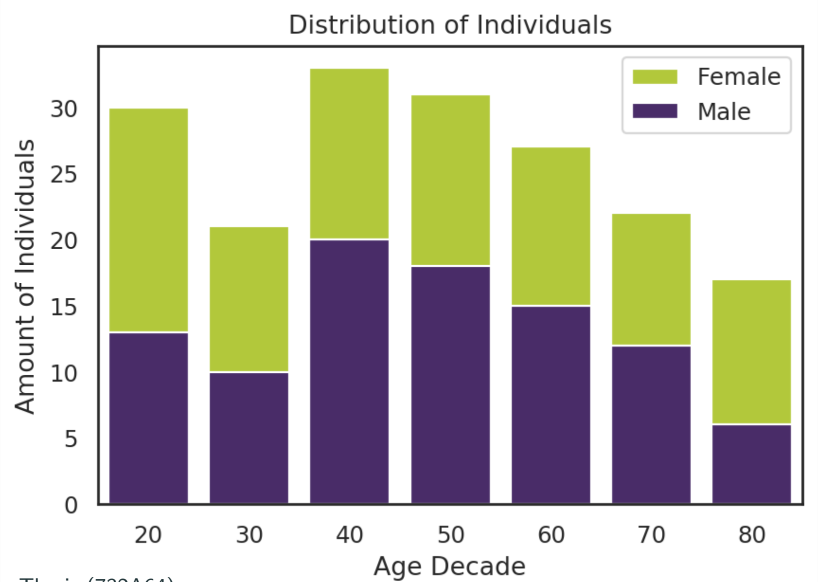
Examiner: Oleg Sysoev
Supervisor: Krzysztof Bartoszek

Table of contents

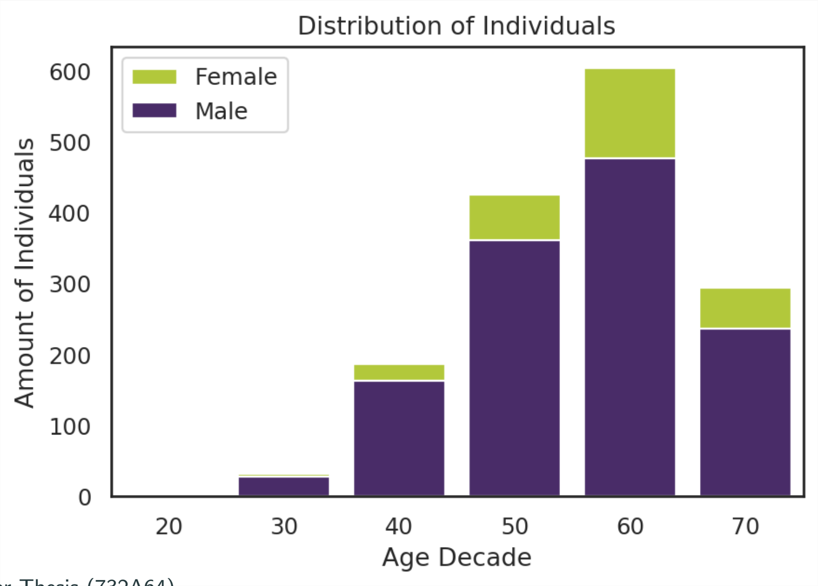
Introduction

Introduction — Age Prediction




Distribution of Individuals
Female
Male
Amount of Individuals
0
5
10
15
20
25
30
20
30
40
50
60
70
80
Age Decade

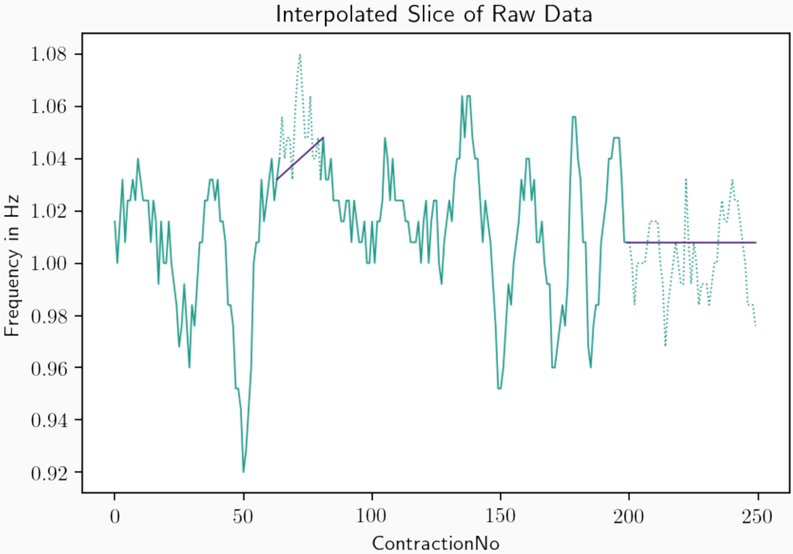
Introduction — Data Sets — PhysioNet



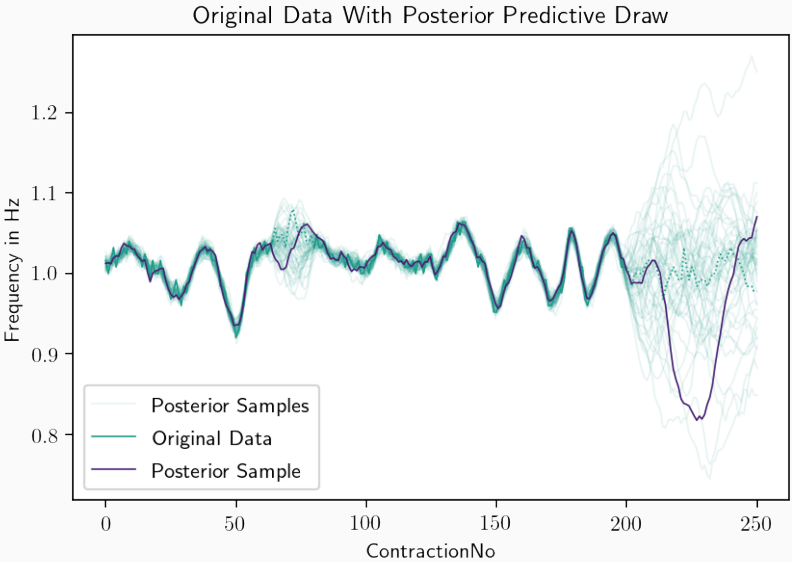
- Can the (cardiovascular) age of a person be predicted using recorded RR intervals and if yes, to which extend?
- Can RR intervals be simulated to generate training data, especially for deep learning models? How well does this work and does it improve the accuracy of the models?

Data Cleaning

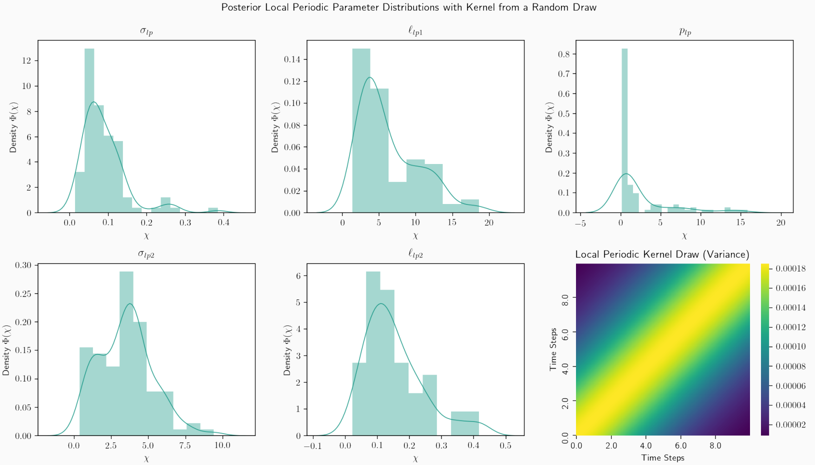
Data Cleaning — Linear Spline Interpolation



Data Cleaning — GP — Posterior Predictive Distribution



Data Cleaning — GP — Kernel Posterior Distributions



Methods and Models

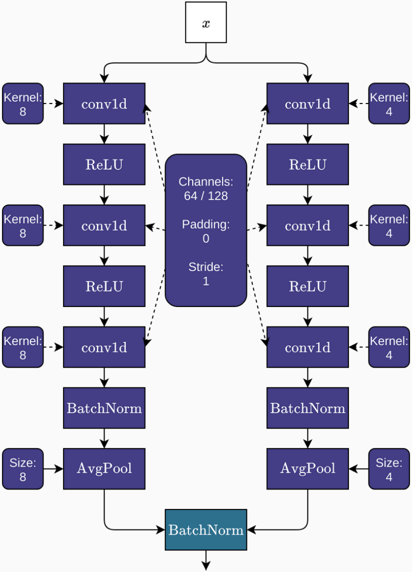
Methods and Models — Methods

- *complete* and *constant*
- *classification* and *regression*
- Feature-Based:
  - 33 features as used in a previous paper (`hrvanalysis`)
  - 3-fold cross validation with grid hyper-parameter search
  - weighted loss
- Deep-Learning:
  - Dropout and early stopping for deep-learning
  - weighted loss, under-sampling and over-sampling

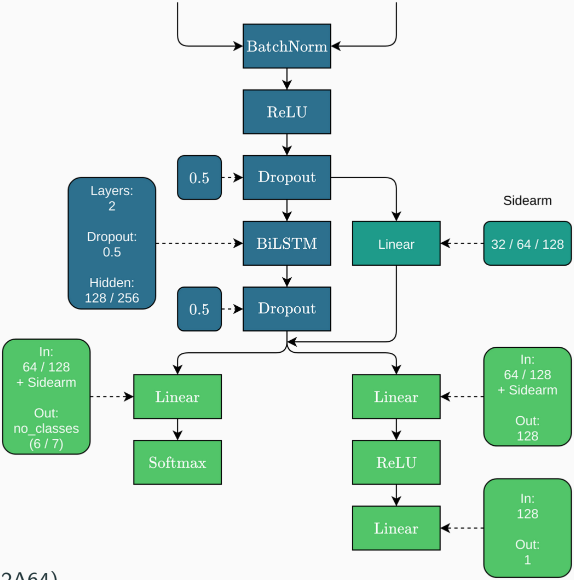
Methods and Models — Models

- Naive Bayes
- Support Vector Machines
- Random Forest
- XGBoost
- DeepSleep-based

Methods and Models — DeepSleep



Methods and Models — DeepSleep



Results

Results — Gdańsk

	Model	Method	Accuracy
1.	Random Forest	complete regression	37.83%
2.	Random Forest	constant classification	32.43%
3.	DeepSleep-based	unweighted constant regression	32.43%

Table 1: Results for the Gdańsk dataset. Baseline is 18.23%.

Results — Physionet

	Model	Method	Accuracy
1.	DeepSleep-based	unweighted constant classification	42.58%

Table 2: Results for the Physionet dataset. Baseline is 39.14%.

Results — Simulated

	Model	Method	Accuracy
1.	SVM	constant regression	31.16%
2.	DeepSleep-based	oversampled constant regression	29.51%
3.	DeepSleep-based	unweighted constant regression	27.66%

Table 3: Results for the Simulated dataset. Baseline is 18.23%.

Discussion

Discussion — Results

- Models perform better on the Gdańsk dataset
- Simulation and impurity handling did not improve the results
- The error distributions for all methods seem to be equal
- Results could be used for a very low number of classes
- Regression helps for deep learning, but not for feature-based models
- Precise age prediction based on RR intervals is **not** possible

Discussion — Results

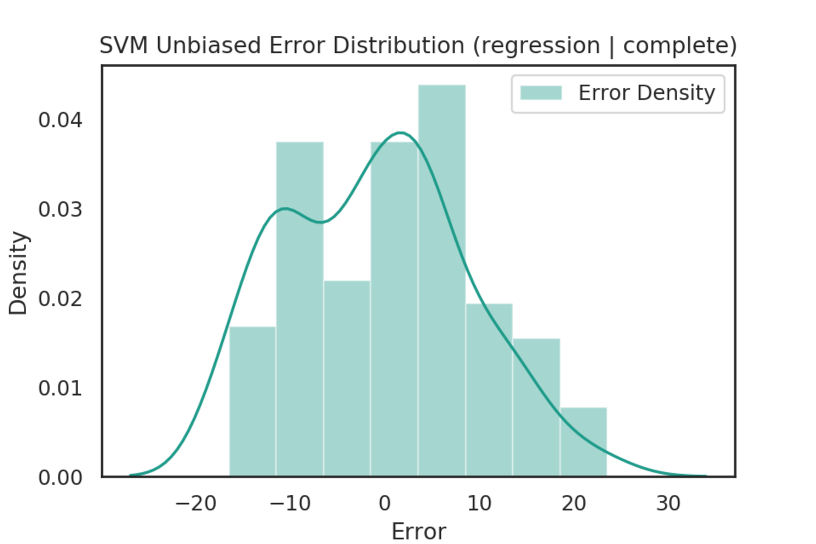
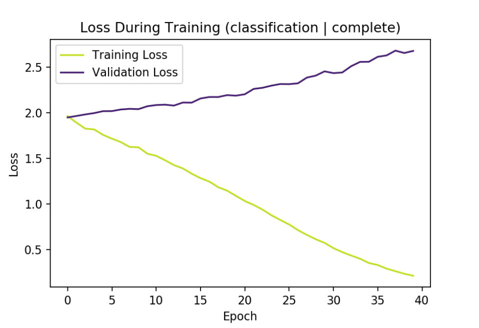


Figure 1: Physionet.

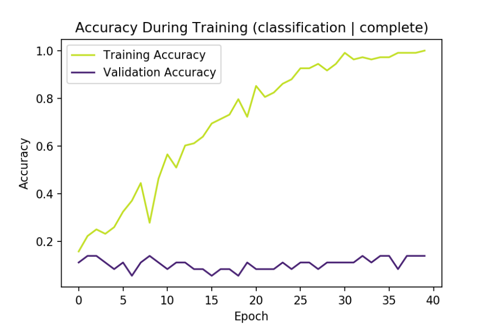
Discussion — Findings

- Tree-based models overfitted while still generalising best
- Weighting and oversampling gives worse results
- Deep learning scales better than most feature-based models
- Loss and accuracy during training for classification and regression

Discussion — Findings



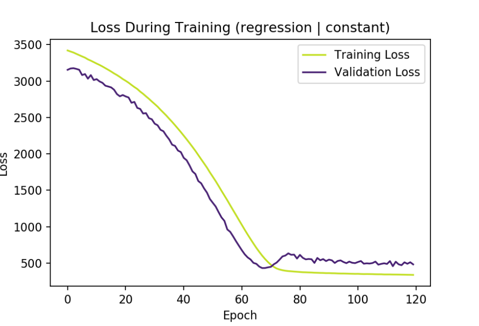
(a) Loss



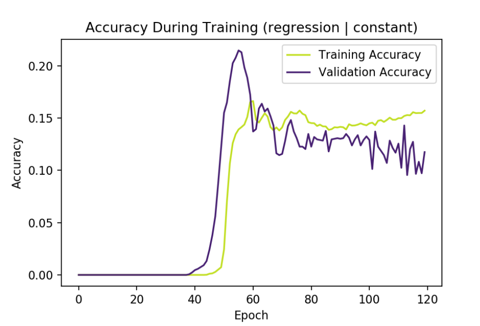
(b) Accuracy

Figure 2: **Training** and **validation** loss and accuracy for DeepSleep performing standard complete classification.

Discussion — Findings



(a) Loss



(b) Accuracy

Figure 3: **Training** and **validation** loss and accuracy for DeepSleep performing oversampled constant regression.

Discussion — Validity

- Comparison with other papers
- [Poddar et al., 2015] report accuracies of $\sim 70\%$ for 3 classes
- [Makowiec and Wdowczyk, 2019] report accuracies over 90% for 7 classes
    - No test dataset
    - Their SVM chose $C = 1.0$ and $\gamma = 0.2$
    - Probably overfitted to the training data

Conclusions & Outlook

Conclusions

- Cardiovascular age prediction based on RR intervals is hard
- Linear Spline Interpolation works sufficiently well
    - Maybe due to the fact that the information we seek is not embedded in the signal

Outlook

- Analyse simulation on a signal where information is obviously embedded
- Use the raw signal no embed for information
    - e.g. P, Q, S and T denotations

Appendix

References i

- Makowiec, D. and Wdowczyk, J. (2019).
  Patterns of heart rate dynamics in healthy aging population: Insights from machine learning methods.
  Entropy.
- Poddar, M., Kumar, V., and Sharma, Y. (2015).
  Heart rate variability: Analysis and classification of healthy subjects for different age groups.
  Conference: INDIACom-2015, At New Delhi (INDIA).