Can Supervised Learning be used to Classify Cardiac Rhythms?

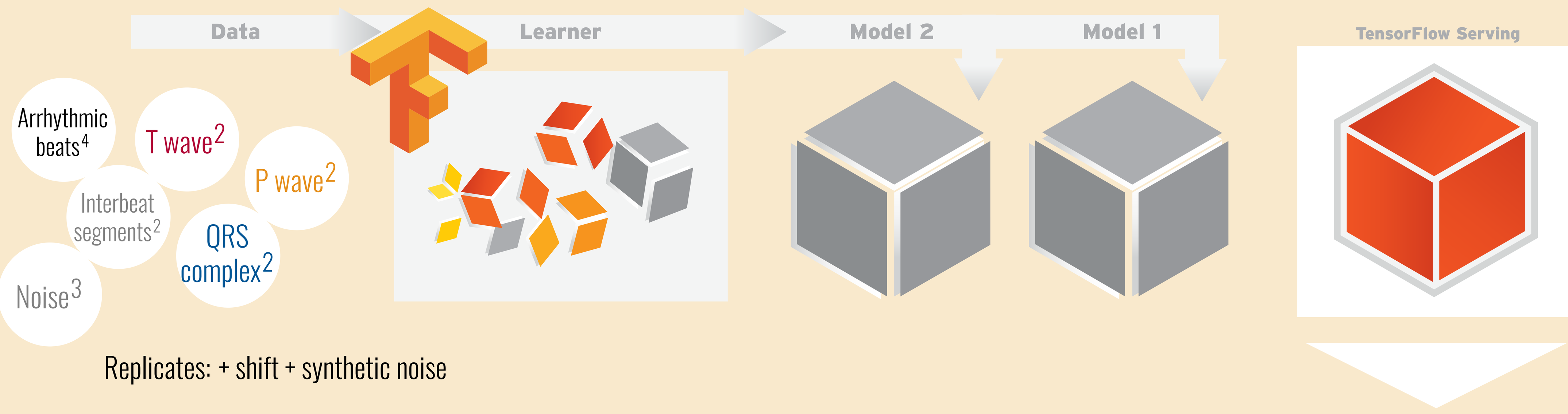
Marcus Vollmer[1], Philipp Sodmann[1], Leonard Caanitz[1], Neetika Nath[1,2], Lars Kaderali[1]

[1] Institute of Bioinformatics, [2] Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Germany

1 Deep learning / Annotation of an ECG

- A convolutional neural network was trained to annotate peaks of an ECG
- Labeled ECG segments from PhysioNet[1] data served as input and data was augmented

Raw ECG
Challenge records

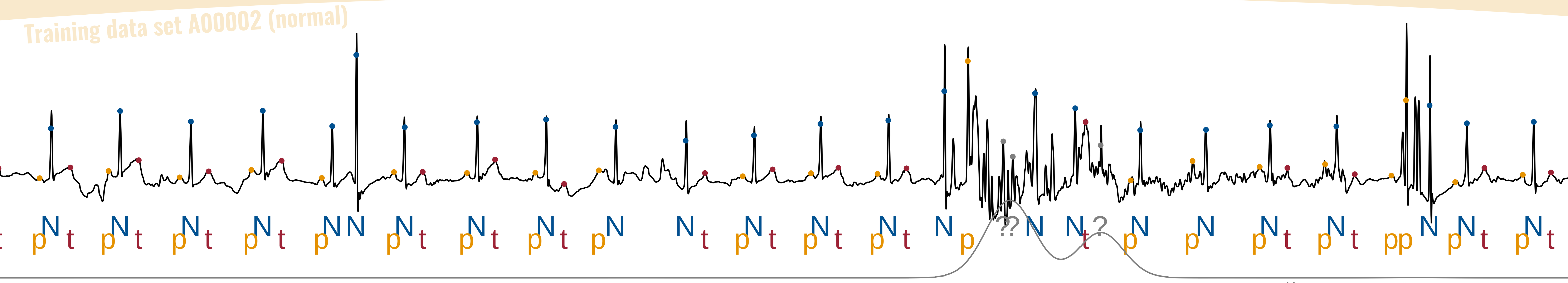


Replicates: + shift + synthetic noise

Annotation performance for QTdb

True positive rate and positive predictivity by comparing the resulting annotations with the reference annotations given for 82 records of the QT database[2]. A strict tolerance level of 25 ms and 50 ms was set for the time difference between both annotations to count as successful (true positive).

		N	p	t
Reference counts		86892	78665	88013
Model counts		86020	75126	79047
True positive rate (sensitivity)	25 ms	0.981	0.886	0.732
	50 ms	0.987	0.910	0.807
Positive predictive value	25 ms	0.991	0.928	0.815
	50 ms	0.997	0.953	0.898

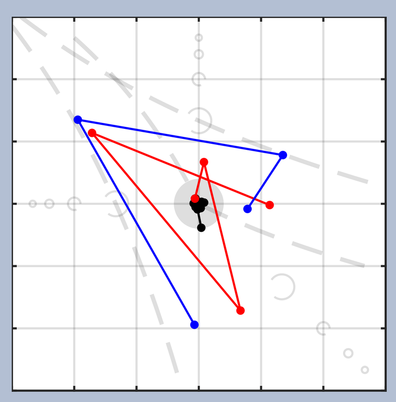


2 Feature Extraction

- 174 basic features were extracted from an annotated record
- Interval data: absolute values, percentiles, and interquartile ranges for *RR*, *RT*, and *PR* intervals
- Extra beats: absolute counts and percentage of extrasystoles with and without compensatory pause, doublets, triplets
- Complexity (entropy) of *RR* intervals: standard deviation of the shortened relative *RR* intervals, from which we removed detected extrasystoles
- Entropy on higher grades: considering a lag when computing relative *RR* intervals
- Normalization: adjusting interval data by heart rate (estimated by the 25% trimmed mean of *RR* intervals) or using relative intervals, defined as successive differences divided by their mean[5]
- Shape information: basic cluster characteristics like the silhouette score and distance information derived from k-Means and hierarchical clustering (average linkage, euclidean metric) on the basis of the cross correlation for each pair of heart beat waveforms

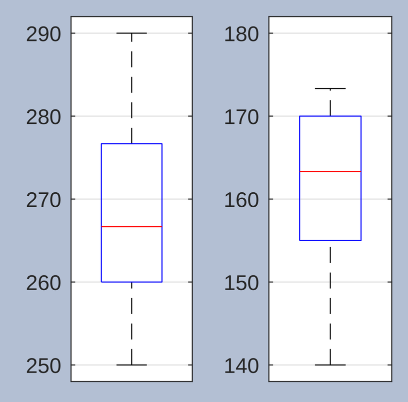
Identification of extra beats

Relative *RR* intervals and classification rules based on relations of successive intervals[6]:



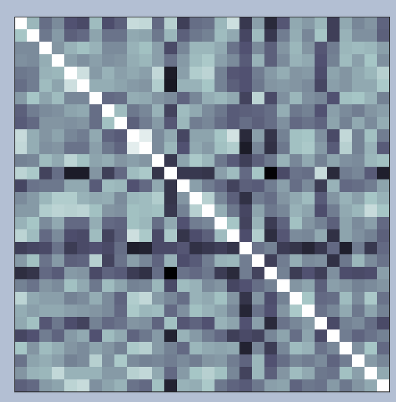
Interval data

Range of *RT* (left) and *PR* intervals (right) in milliseconds:



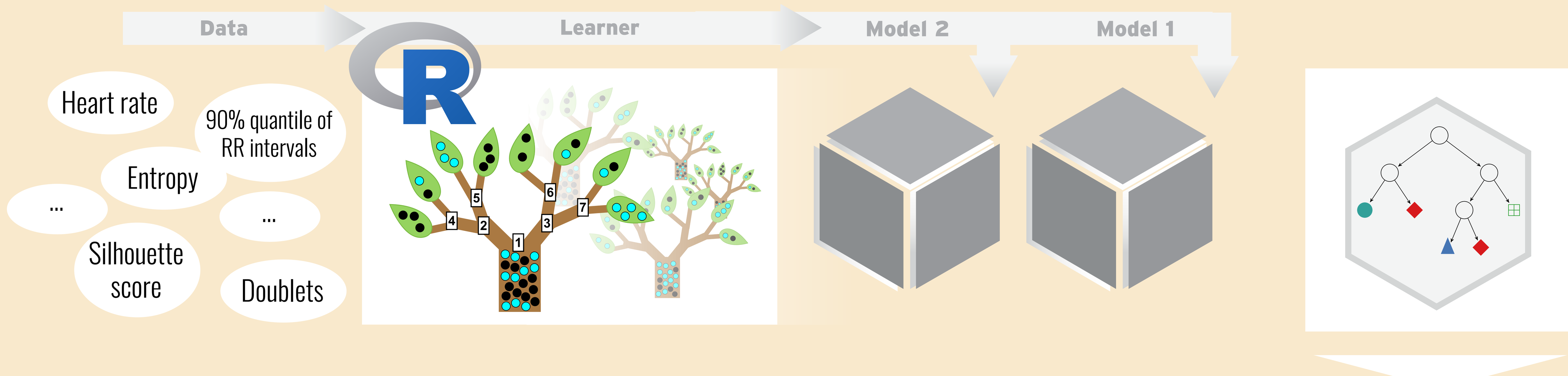
Shape information

Cross-correlation matrix of heart beats to identify classes of beats, e.g. multifocal PVCs:



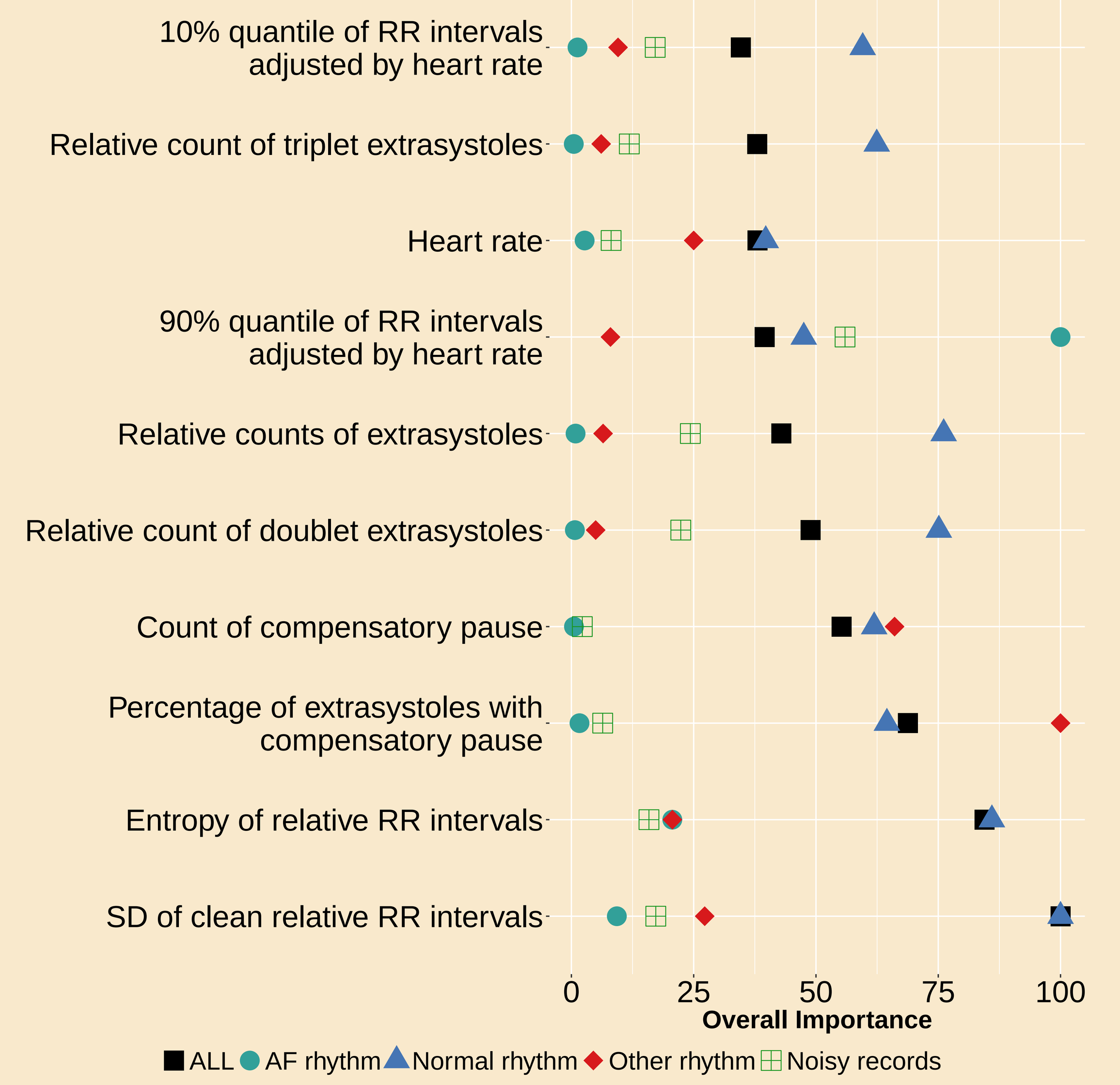
3 Supervised learning / Classification of an ECG

- Random Forest was used to build a regression tree for classifying the rhythm of an ECG
- The training dataset of the PhysioNet/CinC Challenge 2017[7] with 8,528 single lead recordings served as input data



Variable Importance

Features in ascending order by its overall importance and differentiated according to rhythm classes.



Classification performance

The table reports the overall and partial F_1 scores for the training and hidden test set. The noisy records are the hardest of all to predict, the partial F_1 was 0.46 is the result of false negative classifications. Of 284 noisy labeled records 99 were falsely classified as *A* and 94 were classified as *O*. This affects the partial scores of *A* and *O*.

	Overall	N	A	O	~
Training dataset	0.83	0.97	0.94	0.93	0.46
Test phase	NA	0.90	0.80	0.67	NA
Complete hidden set					

[1] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[2] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG," in *Computers in Cardiology 1997*, pp. 673–676, 1997.

[3] M. Vollmer, "Noise Resistance of Several Top-Scored Heart Beat Detectors," in *Computing in Cardiology*, vol. 44, 2017 in press.

[4] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[5] M. Vollmer, "A Robust, Simple and Reliable Measure of Heart Rate Variability using Relative RR Intervals," in *Computing in Cardiology*, vol. 42, pp. 609–612, 2015.

[6] M. Vollmer, "Arrhythmia Classification in Long-Term Data Using Relative RR Intervals," in *Computing in Cardiology*, vol. 44, 2017 in press.

[7] G. Clifford, C. Liu, B. Moody, I. Silva, Q. Li, A. Johnson, and R. Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017," in *Computing in Cardiology*, vol. 44, 2017 in press.

[8] P. T. Baker, S. Caudill, K. A. Hodge, D. Talukder, C. Capano, and N. J. Cornish, "Multivariate classification with random forests for gravitational wave searches of black hole binary coalescence," *Physical Review D*, vol. 91, no. 6, p. 062004, 2015.

Marcus_Vollmer
+49 (0)3834 86-5444
marcus.vollmer@uni-greifswald.de
MarcusVollmer.github.io

Supported by the DZHK (German Centre for Cardiovascular Research).

DZHK DEUTSCHES ZENTRUM FÜR HERZ-KREISLAUF-FORSCHUNG E.V.

09/2017