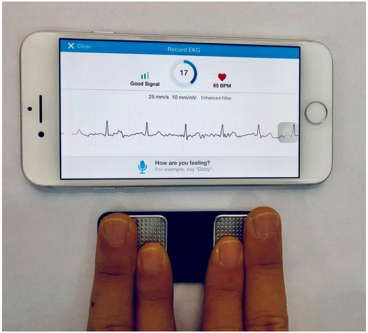
Convolutional Reccurent Neural Networks for Electrocardiogram Classification

Martin Zihlmann, Dmytro Perekrestenko*, Michael Tschannen

Dept. IT & EE, ETH Zurich, Switzerland

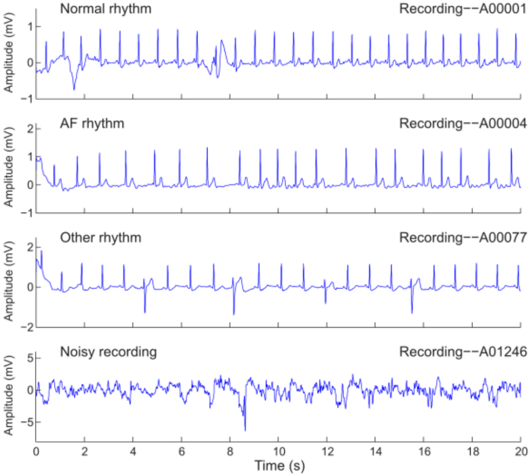
September 25, 2017

Problem



- Atrial fibrillation (AF) occurs in 1-2% of population
- Associated with significant mortality and morbidity
- Existing AF detection methods are not precise and require special multi-lead equipment
- It is proposed to use a single-lead ECG and leverage data processing techniques to be on par with multi-lead detectors
- Challenge - limited data + strict inference requirements

Data



Type	#	%
Normal	5050	59.2
AF	738	8.7
Other rhythm	2456	28.8
Noisy record	284	3.3
Total	8528	100.0

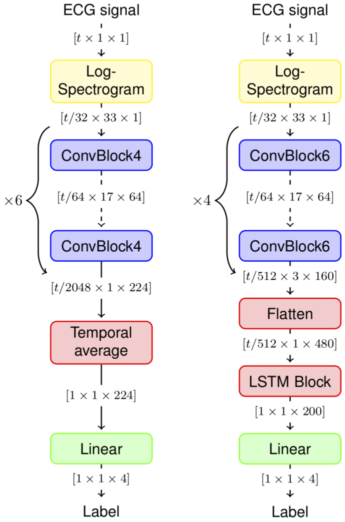
Table 2.1.: Rate of Occurence

Network architectures

- 24-layer CNN
- CRNN that combines a 24-layer CNN with a 3-layer LSTM network.

Both architectures consist of four parts:

1. data preprocessing with log-spectrogram
2. stack of convolutional layers for feature extraction
3. aggregation of features across time by averaging and LSTM block in case of CNN and CRNN, respectively
4. linear layer with Softmax classifier



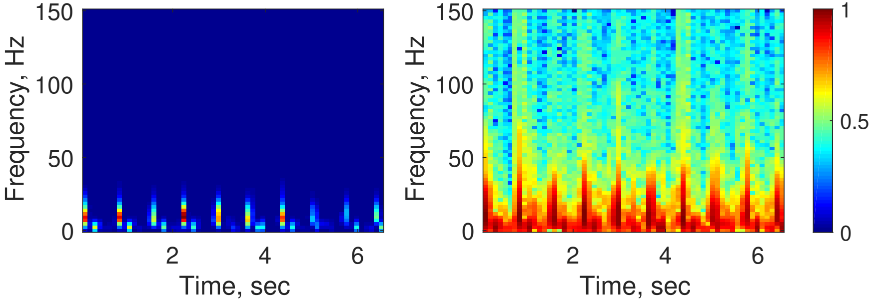
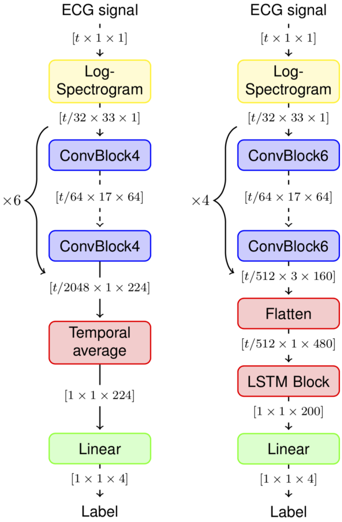


Figure: Normalized spectrogram (left) and normalized logarithmic spectrogram (right) of an example ECG signal.

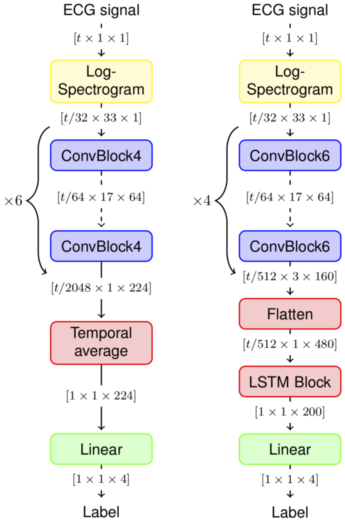
Convolutional layers

- 5×5 filters, ReLU activation
- grouped in blocks of 4 and 6 layers for CNN and CRNN, respectively
- number of channels is constant inside blocks, last layer in block applies max-pooling over 2×2 windows and increases number of channels
- number of channels after first block is 64 and increased by 32 by each subsequent block



Feature aggregation across time

- ConvBlocks produce variable length outputs, need to be aggregated across time to be fed into classifier.
- CNN - aggregation via simple averaging. Temporal smoothing of features, not suited to classify episodic phenomena occurring only during a short time span.
- CRNN - aggregation via 3 layer LSTM with 200 neurons in each layer. Aggregation of features in a highly non-linear manner across time, potentially better preservation of episodic phenomena.



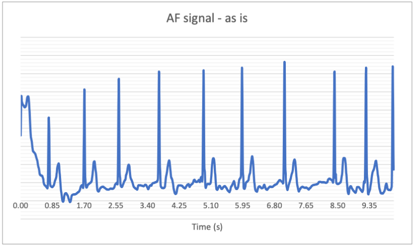
Data augmentation

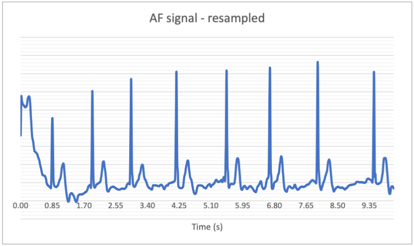
We employ two data augmentation techniques, namely *dropout bursts* and *random resampling*.

- Dropout bursts - selecting time instants uniformly at random and setting the ECG signal values in a 50ms vicinity of those time instants to 0. Dropout burst hence model short periods of weak signal due to, e.g., bad contact of ECG leads.

- Random resampling - assuming a heart rate of 80bpm, uniformly resamples the ECG signals such that the heart rate of the resampled signal would be uniformly distributed on the interval $[60, 120]$bpm. Random resampling thus emulates a broader range of heart rates.

Random resampling





Results (estimated using 5-fold cross validation)

Arch.	metric	N	A	O	~	overall
CNN	acc.	88.1	83.6	66.9	77.1	81.2
	F_1	87.8	79.0	70.1	65.3	79.0
CRNN	acc.	89.9	77.8	69.4	71.5	82.3
	F_1	88.8	76.4	72.6	64.5	79.2

Table: Accuracies (acc.) and F_1 scores (in %) for the proposed network architectures.

Arch.	metric	N	A	O	~	overall
CNN	acc.	90.5	64.2	68.0	54.9	80.5
	F_1	88.3	69.9	69.1	59.6	75.8
CRNN	acc.	90.2	69.1	63.0	51.1	79.2
	F_1	87.4	69.9	66.5	54.9	74.6

Table: Accuracies (acc.) and F_1 scores (in %) for the proposed network architectures with *data augmentation deactivated*.

Challenge submission

Ensemble of 5 CRNN networks.

- Partitioning into 5 equally sized subsets, 4 for training and 1 for validation/early stopping.
- Each network in the ensemble uses a different validation subset. Therefore, all available data was used in training.
- The size of submission is 273.9mb, the ensemble on average consumed 58.1% of computational quota available.

Discussion and Conclusion

- Results show that aggregation of features with LSTM is more effective than averaging. However CRNN has more parameters and thereby higher model capacity.
- The data augmentation proved to be effective, thus indicating that it captures certain real world phenomena.
- Tried to pretrain the CNN and the convolutional layers of the CRNN on the PTB Diagnostic ECG Database [Bousseljot et al., 1995]. Did not lead to improvements.
- Possible improvement - extending the data augmentation scheme, e.g., by taking the actual heart rate into account for random resampling (instead of assuming 80bpm).

Related work

- [Cakır et al., 2017] - CRNN for polyphonic sound detection. Input signal is mapped to the sequence of labels.
- [Bashivan et al., 2016] - CRNN for mental state classification from electroencephalogram (EEG) data.
- [Lipton et al., 2016] - LSTM networks are used for multilabel classification of diagnoses in electronic health recordings.
- [Rajpurkar et al., 2017] - deep CNN architecture for arrhythmia detection in ECGs, but unlike in the classification problem considered here, maps the ECG signal to a sequence of rhythm classes.

References I

Bashivan, P., Rish, I., Yeasin, M., and Codella, N. (2016).
Learning representations from EEG with deep recurrent-convolutional neural networks.
In *Proc. Int. Conf. on Learn. Representations (ICLR)*.

Bousseljot, R., Kreiseler, D., and Schnabel, A. (1995).
Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet.
Biomedizinische Technik/Biomedical Engineering, 40(s1):317–318.

Cakır, E., Parascandolo, G., Heittola, T., Huttunen, H., and Virtanen, T. (2017).
Convolutional recurrent neural networks for polyphonic sound event detection.
IEEE/ACM Trans. on Audio, Speech, and Language Processing, 25(6):1291–1303.

References II

Lipton, Z. C., Kale, D. C., Elkan, C., and Wetzell, R. (2016).
Learning to diagnose with LSTM recurrent neural networks.
In *Proc. Int. Conf. on Learn. Representations (ICLR)*.

Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., and Ng, A. Y. (2017).
Cardiologist-level arrhythmia detection with convolutional neural networks.
arXiv preprint arXiv:1707.01836.