Contextual and Temporal Distribution of False Positives in a Deep Learning Based Atrial Fibrillation Detection Algorithm: An Investigation

Devender Kumar
PhD Student @ DTU Health Tech and CACHET

"When you find someone that makes your heart skip a beat, stop the search and take the risk." Beware! it could be arrhythmia in disguise.

Overview

- ❖ What are arrhythmias/AF?
- ❖ How are they diagnosed?
- ❖ Issues in ambulatory AF detection algorithms
- ❖ Our hypothesis on context-aware AF detection
- ❖ Experiment to test our hypothesis
- ❖ Implications for future AF detection algorithm design

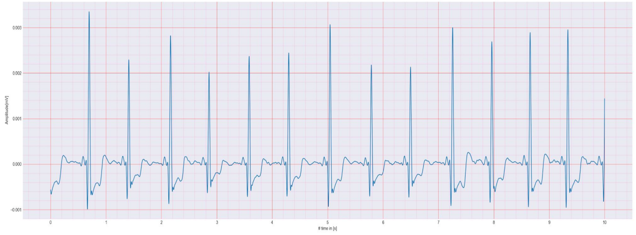
Arrhythmias

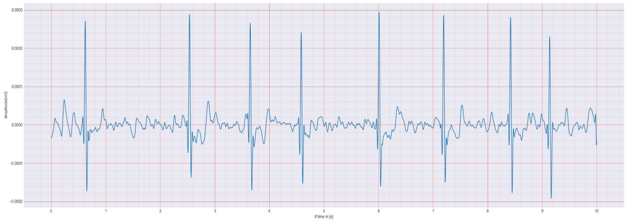
- What are arrhythmias?
  --Abnormality of the heart's rhythm
- What is Atrial fibrillation?
  -- The most prevalent of all arrhythmia

  -- 12.1 million people in the US will have AFib in 2030 [1]

  -- Costs ~2 % of the health budget in EU countries

Normal Sinus Rhythm





Atrial fibrillation

How are they diagnosed?

- ❖ In hospital ECG
- ❖ Ambulatory ECG under free-living conditions

Challenges in ambulatory Arrhythmia diagnosis

- Algorithms work well on public datasets, which are relatively clean and collected under clinical supervision.

- Large number of false positives (FP) on ambulatory ECG data collected under free-living conditions

- Even a small FPR in longitudinal screening could lead to overdiagnosis and patient anxiety

Our hypothesis

- Under free-living conditions, there could be specific user contexts resulting in more FP in an algorithm

- **Impact:** Algorithm's sensitivity and specificity can be fine tuned for false positives prone contexts, and the FPR can be reduced

How do we test the hypothesis?

- ❖ What are the relevant user contexts that need to be collected?

- ❖ Need contextualized ECG dataset to test the hypothesis

- ❖ Existing public arrhythmia datasets are either collected under in-hospital settings or do not have user's context information during ambulatory ECG recording.

mCardia: Collecting longitudinal contextualized ambulatory ECG

- ❖ mCardia system

  --uses a single channel ECG Holter

  --ensures patient engagement & participation

  during the longitudinal screening process

- ❖ Type of context collected

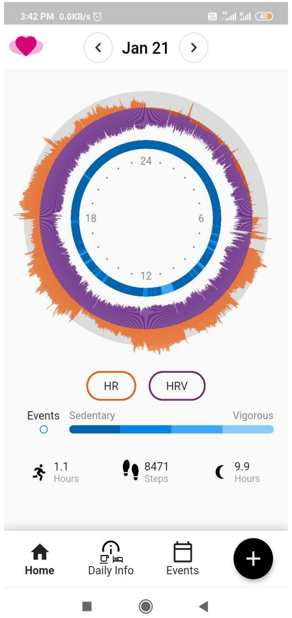
  --Activity

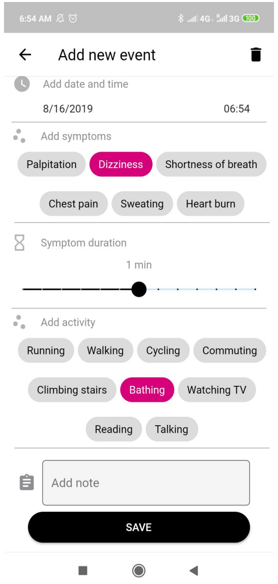
  --Body Position

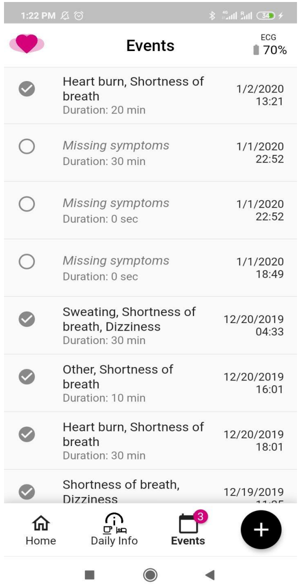
  --Movement acceleration

  --Unusual symptoms experience during the screening

  --Food intake, sleep, and stress levels



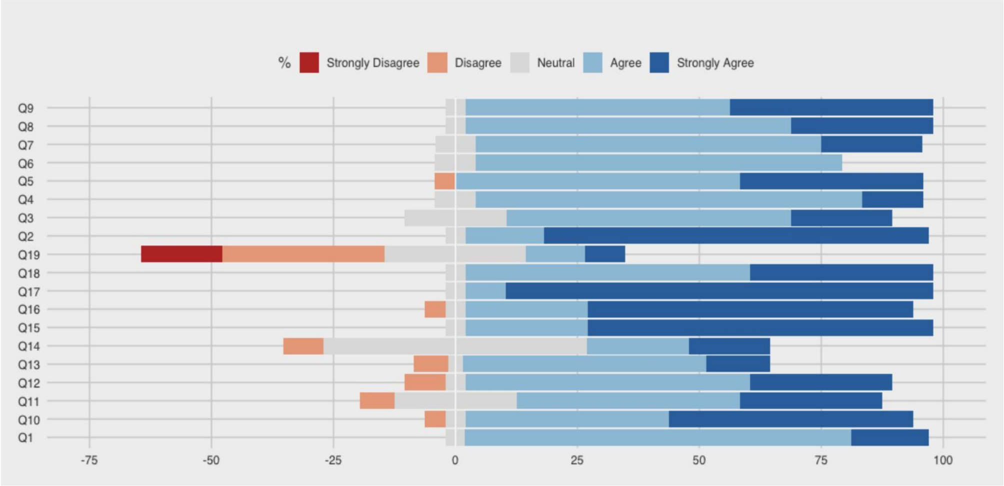




Clinical Feasibility and Usability Study

❖ Recruited 30+ patients suspected of AF and collected minimum two weeks long contextualized ECG data from each participant.

❖ Perceived usefulness and usability of mCardia

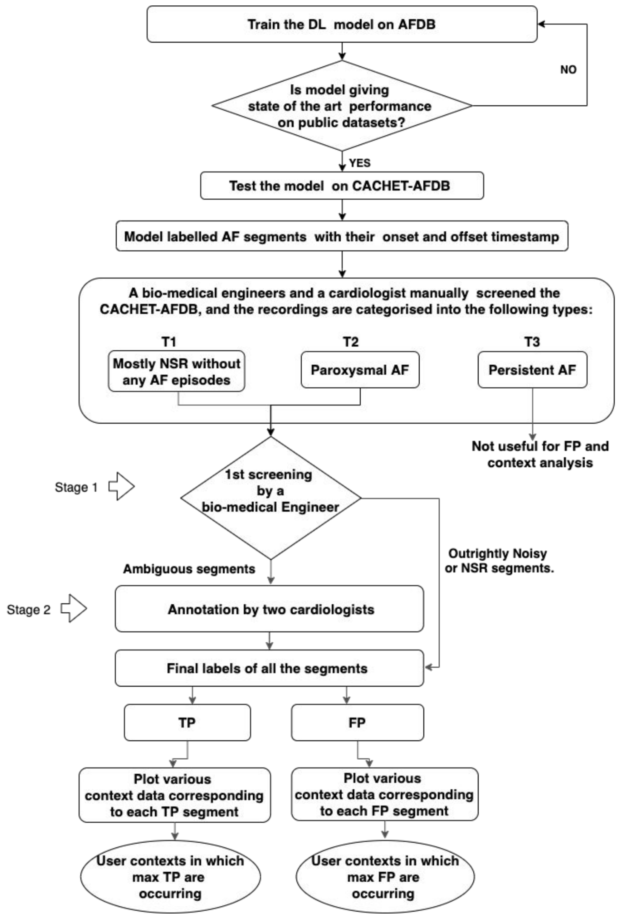


CACHET Unified Method for Assessment of Clinical Feasibility scores

CACHET-AFDB: Contextualized ECG Dataset

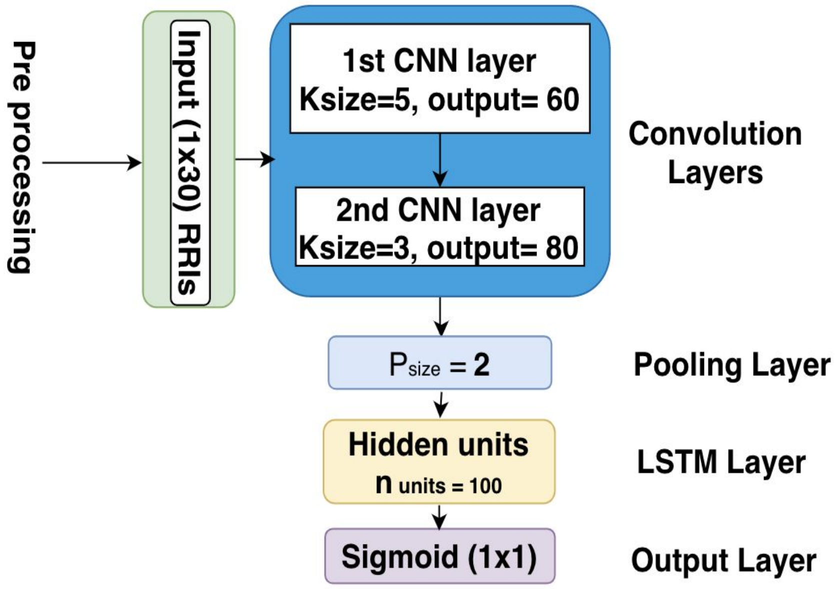
❖ The mCardia feasibility study resulted in over 215 days long contextualized ambulatory ECG data collected under free-living conditions.

Experimental process for analyzing user-context and false positives in a deep learning-based AF detection algorithm



Deep Learning-based AF detection model

- Trained on RRI features
- Combination of 1D CNN and RNN (LSTM)
  -- Two convolutions and a pooling layer with a kernel size of 3 and 5
- Binary classifier (AF or Non-AF class)



Results

- State-of-the-art performance on public datasets

Algorithm	C	AFDB				MITDB				NSRDB			
		Se	Sp	Acc	FPR	Se	Sp	Acc	FPR	Se	Sp	Acc	FPR
[48]	1	97.4	97.2	97.3									
[33]	2	94.28	94.91	94.59									
[18]	2	98.22	98.11	98.18									
[49]	2	99.93	97.03	96.59									
[10]	2	98.17	96.29	97.10	3.71	98.96	86.04	87.40	13.96		95.01		4.99
This work	1	96.06	98.29	97.04	1.7	96.87	86.94	87.98	13.06		94.44		5.56

Performance on data from free-living condition

- As expected, the number of AF increased on the CACHET-AFDB

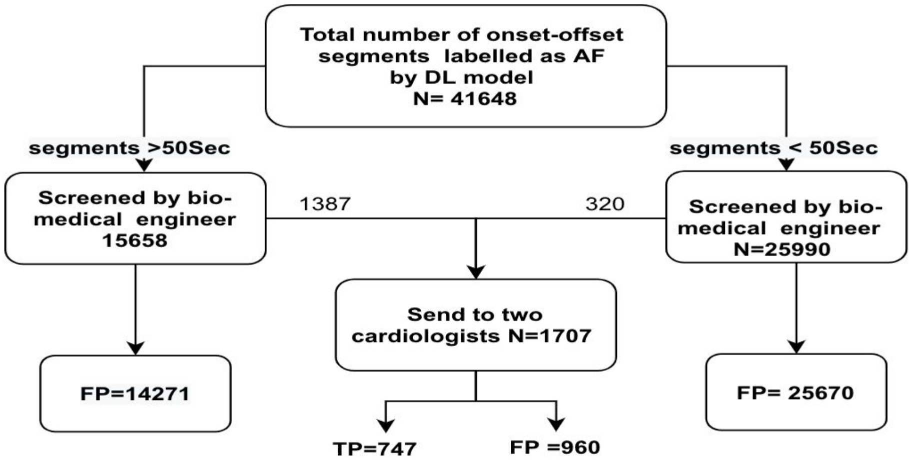
STATISTICS OF DL MODEL'S PERFORMANCE ON CACHET-AFDB. AF-DL: NUMBER OF SEGMENTS DETECTED AS AF BY DL MODEL

Subject	AF-DL	Days	Avg/day	Seg$\leq$50s	Seg$\geq$400s
S1	1991	12	166	1348	68
S2	2826	5	565	1857	25
S3	3419	16	214	1705	317
S4	3712	10	371	1711	107
S5	2308	11	209	1381	166
S6	3198	12	266	1877	65
S7	1702	12	141	1374	32
S8	3058	8	382	2132	153
S9	2415	12	201	1646	55
S10	4290	16	268	2835	71
S11	1236	19	65	883	37
S12	2470	12	205	1707	116
S13	1453	14	103	764	283
S14	787	5	157	614	8
S15	2075	4	518	1152	46
S16	2742	8	342	1569	97
S17	1966	7	280	1435	10

Manual annotation process

Ground truth after manual annotatior

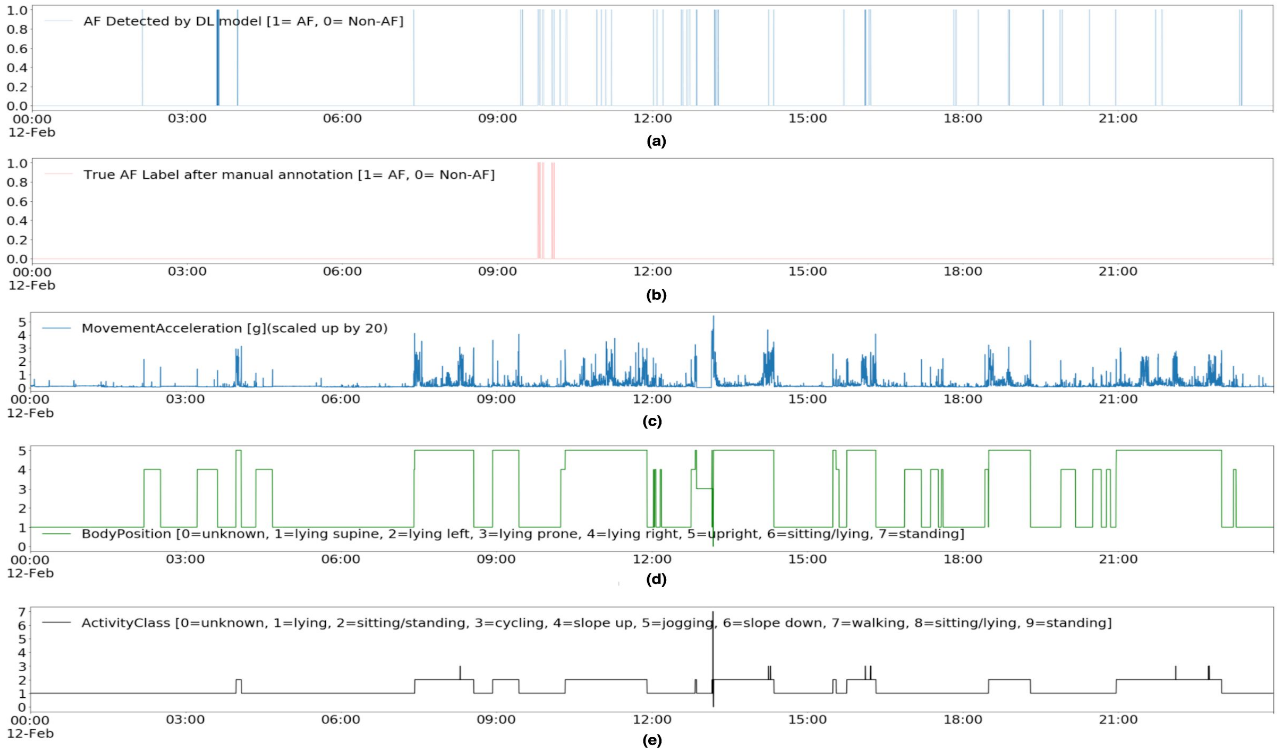
- 62% of total AF segments detected by the model were <= 50 seconds, and 99.9% of them were FP



Relation Between User Contexts and False Positives

The short FP segments of length <=50 sec were mostly associated with:

- Activity change
- Change in body position
- Movements acceleration



True Positives and User-Context

- In three paroxysmal AF subjects, the true positives were concentrated in the morning and late evening hours.

- Palpitations and shortness of breath were the most commonly reported unusual symptoms, and they too were concentrated in the morning and late evening hours.

- The relation between TP and sleep/stress level was non-conclusive from our data.

- More FP in female subjects data as compared to male

Implications of findings for future model design

- Context-aware heuristics around three user context for dynamic adjustments of sensitivity and specificity

- Include more data specific from these three context changes in the training dataset

- Along with ECG, use user context info as features in the multi-model DL models.

What Next?

- Building a new model, "DeepAware," that takes context-aware heuristic and P-wave into account and improves FPR in ambulatory ECG

Thank you!

Comments/Questions/Suggestions?

References

[1]. https://www.cdc.gov/heartdisease/atrial_fibrillation.htm