Not Every Peak Is a Beat:

R-Peak Detection Errors in Single-Lead Holter ECG

Nika Bukolovaitė[1], Zigmantas Kęstutis Juškevičius[1,3], Žygimantas Abramikas[1,2], Viktor Medvedev[1], Jolita Bernatavičienė[1]

[1] Institute of Data Science and Digital Technologies, Faculty of Mathematics and Informatics, Vilnius University, Akademijos g. 4, Vilnius, 04812, Lithuania [3] UAB Zive, Vilnius, Lithuania
[2] Clinic of Cardiovascular Diseases, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Vilnius, 03101, Lithuania nika.bukolovaite@mif.vu.lt

Institute of Data Science and Digital Technologies

1 Problem

RR intervals, beat labels and rhythm analysis are all built on R-peaks. Each beat gets a label: N is a normal beat, S and V are early beats that start in the atria or in the ventricles, and U is an unclassified beat. One extra or missed R-peak distorts the neighbouring RR intervals and can create false S, V or U beats. While cleaning and annotating single-lead Holter ECG, we kept finding R-peak errors, even in clean signal (Fig. 1). Benchmarks rank detectors on reference databases [1, 2, 5], but which beat shapes make them fail is still an open question [5].

(a) R-peak on the T wave (U) **(b) Extra marks (U), beats as V** **(c) Extra low marks** **(d) Unusual beats unmarked** **(e) Negative QRS: missed or U** **(f) Deep S as V, after T as U**

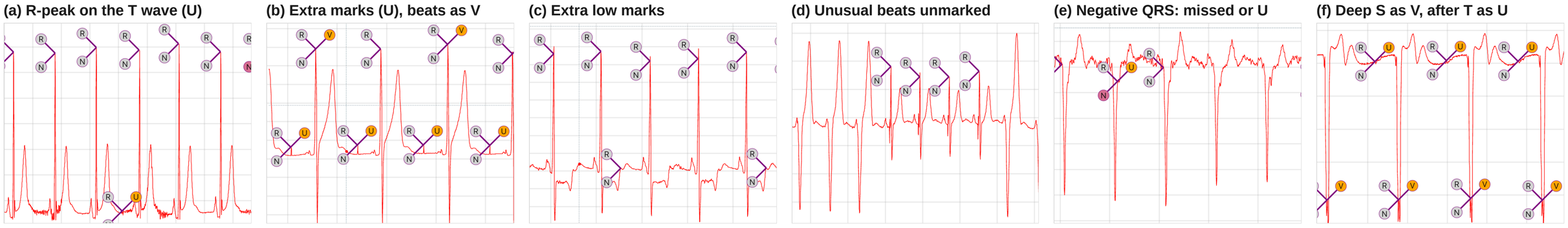

Fig. 1. Good and bad R-peak marks of the baseline method. Each flag joins a mark to its R label and its beat label. N is grey, S, V and U are orange, and the selected beat is magenta.

2 Data and Approach

- **16 Holter recordings with R-peak errors:** single lead, 200 Hz, 3.2–10.7 min each, about 8,400 beats in total. 12 are clean, 3 have EMG noise, 1 has a possible pacemaker.
- **Methods:** the rule-based baseline method (its R-peak marks are stored with the recordings), NeuroKit on 0.5–10 Hz, NeuroKit, Pan–Tompkins and Hamilton (NeuroKit2 [4]), XQRS (WFDB) and Elgendi with the settings of [5] on 5–20 Hz, and Christov (NeuroKit2 [4]) on 0.5–40 Hz. All were run on the same 16 recordings.
- **Fix:** the signal is inverted if most QRS complexes are negative, and each mark is moved to the largest deflection within 50 ms.
- **Vote:** there is no verified reference yet, so each method is compared with the vote of NeuroKit, Pan–Tompkins, Hamilton and XQRS without the fix. A beat counts when at least 3 of the 4 mark it within 150 ms. Agreement is not accuracy.
- **Beat-shape groups:** beats are cut from 200 ms before to 400 ms after each mark and grouped by shape (correlation at least 0.8).

3 Identified Problems and What Helps

Table 1. Problems found in the 16 recordings and what helps. A missed beat is a beat of the vote (Section 2) with no mark within 150 ms, and an extra mark is a mark with no such beat. XQRS with fix agreed best with the vote (Table 3).

Problem	What helps	Baseline method	XQRS with fix	Solved?
R-peak on the T wave (Fig. 1a, 2a)	A narrow 5–20 Hz band	10 extra marks	None	**Yes**
Extra marks between beats (Fig. 1b, c, f)	Another detector	67, 71 and 211 extra	1, 0 and 1 extra	**Yes**
Missed beats (Fig. 1d)	Another detector	7 missed	None	**Yes**
QRS pointing down (Fig. 1e, 2b)	Squared or absolute signal	324 of 623 and 425 of 628 missed	0 and 1 missed	Beats: **yes**. Mark on r or S: **open**
Inverted T wave, changed QRS	Probably the QRS direction and a wider band	11 missed, 6 extra	3 missed, 4 extra	**Not known**
Mark away from the QRS peak	The fix	Within 25 ms: Pan–Tompkins 25 % and 92 % with the fix, Hamilton 72 % and 88 %		Pan–Tompkins: **yes**. Hamilton: **partly**

What helps for each problem: github.com/NikaBukolovaite/R_peak_detection

(a) R-peak on the T wave: B adds a mark, X does not 1704502 **(b) Negative QRS: B misses beats, X finds them** 1640106 **(c) Where to mark? B and X mark S, NK marks r** 1716089



Fig. 2. What the detectors do in the same recordings: B baseline method X XQRS with fix NK NeuroKit

4 Results

Table 2. Errors of the baseline method compared with the vote within 150 ms, as a share of 8,439 vote beats (missed) or 8,099 marks (extra).

	Clean	Noisy (EMG)	Possible pacemaker	T-wave errors	Small-peak errors	Negative QRS	All
Recordings	12	3	1	5	4	4	16
Missed beats	0.3 % (20)	19.1 % (337)	67.7 % (425)	0.1 %	0.9 %	**36.0 %**	**9.3 %** (782)
Extra marks	6.4 % (413)	1.9 % (27)	1.0 % (2)	4.3 %	**15.6 %**	13.9 %	**5.5 %** (442)

The T-wave, small-peak and negative QRS groups overlap, because one recording can have several error types.

Table 3. All methods on the same 16 recordings, compared with the vote.

Method	F1	Missed	Extra	Marks within 25 ms of the QRS peak
Wavelet matched filter				
XQRS with fix	0.999	0.05 %	0.2 %	91 %
Slope energy and adaptive threshold				
Pan–Tompkins with fix	0.998	0.09 %	0.3 %	92 %
Hamilton with fix	0.986	0.5 %	0.8 %	88 %
Christov with fix	0.835	5.6 %	40 %	80 %
Gradient and most prominent peak				
NeuroKit	0.993	0.2 %	1.0 %	72 %
NeuroKit on 0.5–10 Hz	0.890	4.9 %	12.3 %	75 %
Blocks of interest (two moving averages)				
Elgendi with fix [5]	0.937	3.2 %	0.7 %	94 %
Rule-based baseline				
Baseline method	0.911	9.3 %	5.5 %	84 %

The vote comes from NeuroKit, Pan–Tompkins, Hamilton and XQRS without the fix, so their scores are not independent. Elgendi finds only 19 marks for 269 beats in 1716089, where a saturated 9 s segment raises its global threshold, and scores at least 0.97 in the other recordings. On other datasets, Elgendi was the best of 17 detectors [5].

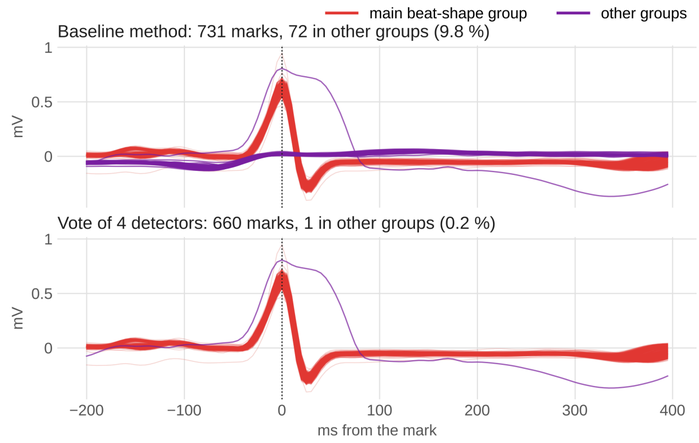

Fig. 3. R errors become false beat groups. Beats of 1652786 (Fig. 1c), one line per beat. The baseline method's 71 extra marks form a flat group, while the vote's only outlier is one wide beat. Over all recordings, 97 % of the baseline method's extra marks (427 of 442) fall outside the main group, as do 8.8 % of all its marks (vote 3.9 %, XQRS with fix 1.9 %). Marks switching between r and S split groups too: 28.6 % fall outside in 1645756 (AF), and 0.2 % with the fix.

5 Conclusions

- Most R-peak errors come from **beat morphology rather than noise**, since 12 of the 16 problem recordings are clean and 99 % of missed beats come from recordings with negative QRS.
- R-peak errors **carry over into beat analysis**, because extra or inconsistent marks form false beat-shape groups (Fig. 3).
- **Not only the detector but also the filter band matters**: NeuroKit scored F1 0.890 on 0.5–10 Hz but 0.993 on 5–20 Hz, where three detectors with the fix agree (F1 0.986–0.999).
- QRS polarity **decides where the mark is placed, not whether a beat is found**.

6 Discussion and Future Work

- The **mechanisms generalise** because they follow from how detectors work, but the **error rates do not**, as the recordings were chosen for their errors and agreement is not accuracy.
- The next step is to **agree where to place the mark (r or S) and which tolerance to use**, since with a 75 ms tolerance (IEC 60601-2-47, as in [5]) NeuroKit, which marks r, drops from 0.993 to 0.936, a drop that the 150 ms tolerance of EC57 [3] hides.
- Each fix should then be **tested on verified error-free recordings** (6 being prepared), so that it does not break recordings that already work.

Acknowledgement. This project has received funding from the Research Council of Lithuania (LMTLT), agreement No. S-ITP-25-9.

References. [1] Liu F. et al. (2018) J Healthc Eng 9050812. [2] Kristof F. et al. (2024) PLOS Digit Health 3: e0000538. [3] ANSI/AAMI EC57. [4] Makowski D. et al. (2021) Behav Res Methods 53: 1689–1696. [5] Wolf S. M. et al. (2026) Sci Rep 16:15748.