

CLUSTERING OF ECG CYCLES

Karolína Němečková

Bachelor Degree Programme (3), FEEC BUT

E-mail: xnemec77@stud.feec.vutbr.cz

Supervised by: Marina Ronzhina

E-mail: ronzhina@vut.cz

Abstract: The paper deals with application of cluster analysis to different ECG records in order to identify particular cardiac pathologies. The work is mainly focused on the detection of premature atrial and premature ventricular beats. Presented approach is based on the signal correlation and further beat type identification and beats clustering via specific ECG features and detection rules, including fuzzy expert rules. By evaluation the method on test data, we obtained Se 76,0 %, Sp 90,2 %, F1 43,8 %, Acc 89,5 %, and PPV 31,1 %. Pure F1 and PPV is due to high number of false positive detections mainly in noisy ECG or ECG with manifested atrial fibrillation.

Keywords: ECG correlation, extrasystols detection, cardiac beats clustering, fuzzy inference system

1 ÚVOD

Elektrokardiografické vyšetření (elektrokardiogram, EKG) je jedno z nejčastějších lékařských vyšetření, je rychlé, neinvazivní a má vysokou vypovídající schopnost o stavu pacienta. Právě z tohoto důvodu se neustále pracuje na zdokonalování systémů pro automatické vyhodnocení EKG křivky. Cílem této práce je detekce síňových a komorových extrasystol v 10 s záznamu EKG, což by mohlo přispět k časnému odhalení závažných postižení srdce, jako je například fibrilace komor, jejichž indikátorem mohou být právě extrasystoly. Většina dosavadních vyhodnocovacích systémů pracují na základech umělé inteligence a strojového učení, což sice vede k dosažení kvalitních výsledků detekce, ale za cenu složité či zcela nemožné interpretace rozhodovacího procesu a tím i výsledků. [2] Rozhodování systému představeného v této práci kopíruje postup zdravotnického personálu při vyhodnocování záznamu EKG. Celá práce se zakládá na hlavní charakteristice extrasystol. Předčasné síňové stahy (premature atrial contraction, PAC) se v EKG projevují změnou morfologie P vlny beze změn QRS komplexu či vlny T. Dále také odpovídající RR intervaly v okamžiku výskytu extrasystoly mají charakteristické vlastnosti, konkrétně krátký extrasystolický a dlouhý postextrasystolický RR interval. Komorové extrasystoly (premature ventricular contraction, PVC) se liší morfologií QRS komplexu, absencí P vlny a odlišnou délkou RR intervalů. Charakteristika vyplývá ze samotného vzniku a šíření elektrického impulzu, který je generován v komoře namísto sinusového uzlu, jak tomu bývá za normálních okolností. [1]

2 POUŽITÁ DATA

V této práci byla použita databáze patientských záznamů EKG z UBMI. Jednalo se o 6884 deseti-sekundových EKG záznamů snímaných z osmi svodů (II, III, V1-V6) se vzorkovací frekvencí 500 Hz. U těchto dat již byla vyfiltrována stejnosměrná složka a 50 Hz brum. Spolu s databází jsem měla možnost pracovat se souborem detekovaných pozic QRS komplexů a souborem anotací. Data byla pro objektivní hodnocení rozdělena na trénovací a testovací tak, aby v každé skupině byly patologie zastoupeny rovnoměrně.

3 PROGRAMOVÉ ŘEŠENÍ

Program a všechny jeho části byly navrženy a implementovány v prostředí MATLAB R2019b. Celý program můžeme rozdělit do několika dílčích částí: příprava signálu, korelační analýza, detekce PAC a detekce PVC. Tyto dílčí části budou dále postupně popisovány.

3.1 PŘÍPRAVA SIGNÁLU

V práci byla použita již filtrovaná data a pro zamezení zkreslení se aditivní filtrace neprováděla. Změna byla provedena pouze v rámci R kmitů, jejichž pozice byly zpřesněny. Toto zpřesnění bylo provedeno tak, že vrchol byl vždy zvolen jako maximum v úzkém okně kolem detekovaného vrcholu. Hlavním krokem zde bylo signál rozdělit na jednotlivé cykly, které pak budou následně korelovány mezi sebou. Rozdělení proběhlo podle dvou kritérií na delší (QRS a okolní oblast) a kratší (pouze QRS) cykly. Cykly byly následně použity pro korelační analýzu a výpočet tvarového faktoru (form factor, FF) podle vzorce (2). Přičemž FF vypočtený z delšího cyklu napomáhá odlišení od sebe normálních cyklů a PVC. Velikost okna byla odvozena zvlášť pro každý záznam s využitím tepové frekvence (dále jen TF). Jako průměrný tep bylo vybráno 80 bpm, kde se bere okno délky 100 ms (50 ms před R kmitem a 50 ms za R kmitem). Dále jsou délky kratšího okna vypočítávány pomocí kritérií z (1), kdy okno je zmenšováno pro nadprahovou TF, a naopak zvětšováno pro podprahovou TF:

$$okno = 50 + |80 \pm TF| * 0.1 \quad (1)$$

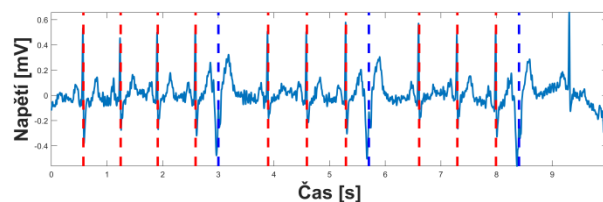
Pro výpočet delšího okna bylo potřeba vypočítat průměrný RR interval a okno bylo následně vybráno jako $\frac{1}{2}$ RR intervalu před i za R kmitem.

$$FF = \frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x} \quad (2)$$

3.2 KORELAČNÍ ANALÝZA

Korelační analýza je hlavní část programu, která provede prvotní třídění QRS komplexů v záznamu na základě velikosti korelačního koeficientu. Na začátku každého cyklu se zvolí vzor – první QRS komplex – se kterým budeme zbylé cykly korelovat. Pro každou dvojici (vzor a další komplex) se nejprve vypočítá pomocí vzájemné korelační funkce vzájemný posun, při kterém jsou si nejpodobnější. Následně se při tomto posunu vypočítá korelační koeficient a také FF z delších dvou cyklů. Dále porovnáme FF vzoru a korelovaného signálu a pokud je jejich rozdíl větší než směrodatná odchylka FF v rámci celého záznamu, provede se snížení korelačního koeficientu o hodnotu 0,15, což tento analyzovaný QRS komplex znevýhodní při shlukování. Hodnota 0,15 byla zvolena na základě optimalizačních postupů, kdy byl program opakovaně spouštěn, avšak vždy s jinou hodnotou a snížení koeficientu o 0,15 dosahovalo nejlepších výsledků.

Jakmile je vzor korelován se všemi zbylými komplexy, provede se samotné shlukování tak, že vzor představuje geometrický střed shluku a všechny komplexy, které mají korelační koeficient k tomuto vzoru vyšší než 0,85 jsou do tohoto shluku zařazeny. Hodnota 0,85 byla zvolena obdobně jako v případě snížení korelačního koeficientu. Nyní se vzor posouvá na další v pořadí nezařazený komplex a vše probíhá znovu. Celé shlukování je provedeno zvlášť pro svody II, V4 a V6, které jsou pro analýzu EKG nejčastěji volené. Finální svod se volí takový, který má ve výsledku nejmenší počet vzniklých shluků, protože se předpokládá, že vyšší procento šumu v analyzovaném EKG způsobuje snížení korelace, a tedy vznik více shluků, což je nežádoucí pro následnou analýzu cyklů. Možný výsledek shlukování můžete vidět na Obrázku 1.



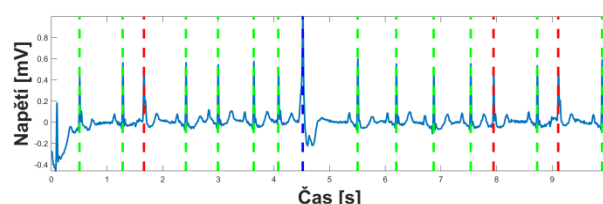
Obrázek 1: Ukázka správné funkce shlukování, kde jsou jednotlivé shluky barevně odlišeny. Dle dostupných anotací: cykly označené červenou tečkovanou čarou odpovídají normálním cyklům a modře jsou označeny PVC.

3.3 DETEKCE SÍŇOVÝCH EXTRASYSTOL

Při detekci PAC se rozpoznávají odlišnosti RR intervalů charakteristické pro daný typ extrasystoly. K této detekci jsou použity vektor RR intervalů konkrétního záznamu, konstanta označující nejpočetnější shluk/y (právě v tomto/těchto probíhá detekce PAC) a vektor vypovídající o výsledku prvotního shlukování (vektor čísel, přičemž každé číslo odpovídá jednotlivému shluku). Detekce probíhá v nejpočetnějších shlucích právě z toho důvodu, že v případě PAC by se QRS komplex neměl lišit od ostatních (kromě PVC) a korelace by tyto cykly měla zařadit do stejného shluku.

Obdobnými metodami, kterými byl zvolen úbytek korelačního koeficientu při odlišných FF byly zvoleny prahy, kdy je RR interval brán již za extrasystolický. Výchozími orientačními body během optimalizace byly střední hodnoty extrasystolických a postextrasystolický RR intervalů v rámci trénovacích dat. Opět je zde zohledněna TF, a proto byl zvolen práh zvláště pro bradykardii, tachykardii a fyziologickou TF. Je monitorován jak extrasystolický RR interval (bere se jako extrasystolický, pokud je jeho hodnota nižší než hodnota průměrného RR intervalu v záznamu zmenšená o hodnotu prahu) tak i postextrasystolický interval (k průměrnému RR intervalu práh přičítáme a RR interval musí být větší, aby byl považován za postextrasystolický).

Algoritmus ošetřuje i případy výskytu následných mnohočetných extrasystol, kde předpokládáme kaskádu krátkých RR intervalů ukončených neúplnou kompenzační pauzou. Pokud do algoritmu vstoupí cyklus, který má kratší RR interval (již brán za extrasystolický) před i za QRS komplexem, funkce iterativně pokračuje v kontrole RR intervalů následných cyklů. Sada těchto cyklů se vyhodnotí jako extrasystolická pouze v případě, pokud po posledním cyklu v řadě krátkých RR intervalů před i za QRS komplexem následuje cyklus s krátkým RR intervalem před komplexem a delší pauzou za komplexem QRS (interval brán již za postextrasystolický). Ukázku správné funkce detekce můžeme vidět na Obrázku 2.



Obrázek 2: Ukázka správné detekce PAC, červeně označen nově vzniklý shluk pro PAC.

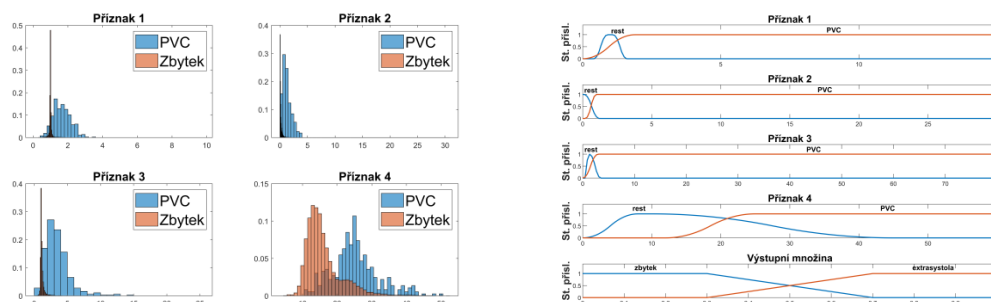
3.4 DETEKCE KOMOROVÝCH EXTRASYSTOL

Nyní již máme samostatný shluk pro PAC a nejpočetnější shluk/y, u kterých nepředpokládáme výskyt PVC kvůli jejich výrazně odlišné morfologii. Do další části programu vstupují zbylé shluky, v nichž proběhne detekce PVC na základě vybraných specifických EKG příznaků a vyhodnocení pomocí navrženého fuzzy expertního systému.

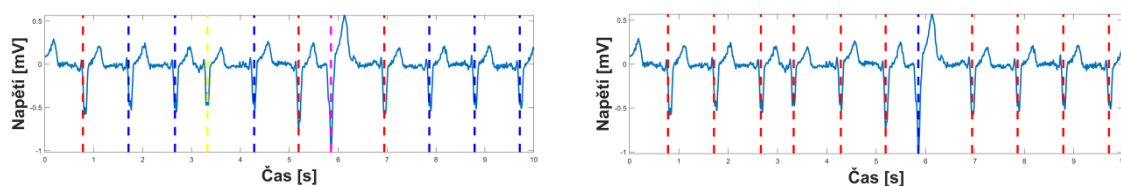
Z několika různých EKG příznaků byly v rámci experimentování vybrány 4 nejvhodnější příznaky. Jedná se o i) podíl RR intervalů před a po komplexu QRS, ii) rozdíl FF zkoumaného cyklu a reference, která byla vypočtena jako průměr z cyklů shluknutých v nejpočetnější skupině (pova-

žujeme za sinusové), iii) podíl plochy pod křivkou zkoumaného komplexu a reference a iv) počet vzorků mezi extrémy komplexu.

Trénovací data byla rozdělena do tří skupin: PAC, PVC a ostatní. V následujících krocích bylo pracováno pouze se skupinami PVC a ostatní. u obou těchto skupin byly vybrány příznaky a odděleně vykresleny histogramy. Na základě těchto výsledků byl navržen a implementován fuzzy interferenční rozhodovací systém. Postup můžete vidět na Obrázku 3. Rozhodovací pravidla byla nastavena tak, aby byl komplex vyhodnocen jako PVC pouze v případě, pokud alespoň tři ze čtyř příznaků nabývají hodnot typických pro PVC. Před implementací fuzzy systému byla detekce realizována obdobně jako v případě detekce PAC, kdy byly pomocí ROC analýzy nalezeny pevné prahové hodnoty pro jednotlivé příznaky. Implementace fuzzy systému značně přispěla k získání lepších výsledků detekce PVC. Ukázku správné funkce můžete vidět na Obrázku 4.



Obrázek 3: Ukázka návrhu fuzzy interferenčního systému. Podle vykreslených histogramů příznaků (vlevo) byly navrženy jednotlivé funkce příslušnosti proměnných (vpravo).



Obrázek 4: Výsledek prvotního shlukování (vlevo) a výsledná úprava detekce PVC pomocí FIS (vpravo), kde jsou barevně oddělené jednotlivé shluky. Dle dostupných anotací: na grafu vpravo cykly označené červenou tečkovanou čarou odpovídají normálnímu cyklům (raménková blokáda) a modře označeny cykly PVC.

4 ZÁVĚR

Navržený algoritmus dosahuje nejkvalitnějších výsledků u minimálně zašuměných signálů. V přítomnosti šumu procento falešně pozitivních (FP) detekcí roste. Vysoký počet FP detekcí PAC se vyskytuje také v případě atrální fibrilace, kde dochází k atrio-ventrikulární disociaci. Na trénovacích datech byly obdrženy výsledky: Se 77,8 %, Sp 89,7 %, F1 46,9 %, Acc 89,0 % a PPV 34,5 %. Nízké hodnoty F1 a PPV jsou způsobeny vysokým výskytem falešně pozitivních detekcí a proto by bylo vhodné algoritmus doplnit o aditivní filtraci a detekci atrální fibrilace.

REFERENCE

- [1] HAMPTON, John R. *EKG stručně, jasně, přehledně*. 7. Praha: Grada, 2013. ISBN 978-802-4742-465.
- [2] LI, Qichen, Chengyu LIU, Qiao LI, Supreeth P SHASHIKUMAR, Shamim NEMAT, Zichao SHEN a Gari D CLIFFORD. Ventricular ectopic beat detection using a wavelet transform and a convolutional neural network. *Physiological Measurement*. 2019, **40**(5). DOI: 10.1088/1361-6579/ab17f0. ISSN 1361-6579.