

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE

Brno, 2019

Richard Ředina



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

TŘÍDĚNÍ EKG CYKLŮ POMOCI SHLUKOVÉ ANALÝZY

CLUSTERING OF ECG CYCLES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Richard Ředina

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Marina Ronzhina, Ph.D.

BRNO 2019

Bakalářská práce

bakalářský studijní obor **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Student: Richard Ředina

ID: 195715

Ročník: 3

Akademický rok: 2018/19

NÁZEV TÉMATU:

Třídění EKG cyklů pomocí shlukové analýzy

POKyny PRO VYPRACOVÁNÍ:

1) Proveďte literární rešerši v oblasti elektrokardiologie, popište projevy různých patologií srdce v EKG. Dále se seznamte s metodami zpracování a analýzy EKG. Zaměřte se na metody používané pro třídění cyklů EKG. 2) Seznamte se s databází patientských EKG záznamů dostupnou na ÚBMI. 3) Realizujte vybranou metodu třídění cyklů v prostředí Matlab. 4) Otestujte navrženou metodu na záznamech od pacientů s různými patologiemi. Vyhodnoťte úspěšnost použité metody. 5) Na základě předchozího bodu proveďte optimalizaci metody a automatizujte nastavení parametrů metody. 6) Dosažené výsledky vhodně interpretujte. Diskutujte výhody a limity navržené třídící metody.

DOPORUČENÁ LITERATURA:

[1] CURTIS, Michael J., Jules C. HANCOX, András FARKAS, et al. The Lambeth Conventions (II): Guidelines for the study of animal and human ventricular and supraventricular arrhythmias. Pharmacology and Therapeutics, Amsterdam: Elsevier Inc., 2013. roč. 139, vyd. 2, s. 213 – 248. ISSN 0163-7258.

[2] SMISEK, Radovan, Jakub HEJC, Marina RONZHINA, et al. Multi-stage SVM approach for cardiac arrhythmias detection in short single-lead ECG recorded by a wearable device. Physiological Measurements, 2018.

[3] YEH, Yun-Chi, Che W. CHIOU, Hong-Jhih LIN. Analyzing ECG for cardiac arrhythmia using cluster analysis. Expert Systems with Applications, 2012. roč. 39, vyd. 1, s. 1000-1010.

Termín zadání: 4.2.2019

Termín odevzdání: 24.5.2019

Vedoucí práce: Ing. Marina Ronzhina, Ph.D.

Konzultant:

prof. Ing. Ivo Provazník, Ph.D.
předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Bakalářská práce pojednává o aplikaci shlukové analýzy na různé záznamy EKG s cílem vytvoření spolehlivého algoritmu na detekci odlišných QRS komplexů. Samotný algoritmus zahrnuje filtraci, upravení pozic R kmitů, vytvoření vzorového cyklu. Následuje porovnání na základě korelace a střední kvadratické odchylky, ze kterých se stanou data pro shlukování metodou k-means. Výsledný počet shluků je odvozen z průměrné hodnoty silulet pro různé počty shluků.

KLÍČOVÁ SLOVA

EKG, QRS komplex, srdeční arytmie, shluková analýza, k-means, PCA, korelace, siluety

ABSTRAKT

The bachelor thesis explores the application of cluster analysis on different ECGs in order to create a reliable algorithm for detecting different QRS complexes. Algorithm comprises filtration, R-wave positions adjustment, model cycle creation and comparison based on mean square error and correlation. Both, correlation and mean square error, become data for k-means clustering. The number of clusters is derived from silhouette values for different numbers of clusters.

KLÍČOVÁ SLOVA

ECG, QRS complex, cardiac arytmiias, cluster analysis, k-means, PCA, correlation, silhouette

ŘEDINA, Richard. *Třídění EKG cyklů pomocí shlukové analýzy*. Brno, Rok, 41 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce: Ing. Marina Ronzhina, Ph. D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Třídění EKG cyklů pomocí shlukové analýzy“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucí diplomové práce Ing. Marině Ronzhině, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno

.....

podpis autora

Obsah

Úvod	9
1 Morfologie EKG	10
1.1 Fyziologické EKG	10
1.1.1 Popis EKG	10
1.2 Patologické stavy	12
1.2.1 Fibrilace síní	12
1.2.2 Síňový flutter	12
1.2.3 Hypertrofie komor	13
1.2.4 Raménkové blokády	13
1.2.5 Ischemie myokardu	14
1.2.6 Komorové extrasystoly	15
1.2.7 Komorová tachykardie	15
1.2.8 Fibrilace komor	16
2 Shluková analýza	17
2.1 Hierarchické metody	17
2.2 Nehierarchické metody	17
2.2.1 Metoda k-means	18
2.3 Hodnocení shlukování	19
3 Databáze EKG dostupná na UBMI	20
4 Programové řešení	23
4.1 Filtrace	23
4.2 Optimalizace detekce R vln	26
4.3 Vytvoření vzoru	27
4.4 Metriky podobnosti	29
4.5 Analýza hlavních komponent	29
4.6 Shlukování	30
5 Statistické zhodnocení	36
6 Závěr	39
Literatura	40
Seznam symbolů, veličin a zkratk	41

Seznam obrázků

1.1	Fibrilace síní s charakteristickým rozvlněním izoelektrické linie [2] . . .	12
1.2	Flutter síní s dobře rozeznatelnými „zuby pily“ [2]	13
1.3	Kopletní blokáda levého Tawarova raménka - délka QRS je vyšší než 120 ms [1]	14
1.4	Komorové extrasystoly [2]	15
1.5	Komorová tachykardie [2]	15
1.6	Fibrilace komor [2]	16
2.1	Algoritmus metody k-means [5]	18
3.1	Signál č. 4563, svod II	20
3.2	Signál č. 4563, svod V4	21
3.3	Signál č. 4563, svod V6	21
3.4	Ukázka signálu obsahujícího pouze šum (signál č. 100, svod II)	22
4.1	Schéma algoritmu	23
4.2	Signál číslo 955 před a po filtraci	24
4.3	Signál číslo 955 po filtraci	24
4.4	Signál číslo 4834 před filtrací	25
4.5	Signál číslo 4834 po filtraci	25
4.6	Detekce R kmitu v signálu č. 1968 algoritmem zabudovaným do měřícího software (zeleně) a po optimalizaci navrženým algoritmem (červeně)	26
4.7	Signál č. 4563 s detekovanými S kmity	28
4.8	Vzor vytvořený kumulací cyklů signálu č. 4563	28
4.9	Signál č. 1002 s dvěma abnormálními QRS komplexy	31
4.10	Průměrné hodnoty siluet pro různé počty shluků	31
4.11	Výsledek shlukování	32
4.12	Signál č. 2607 s vyznačenými R kmity	32
4.13	Průměrné hodnoty siluet pro různé počty shluků	33
4.14	Výsledek shlukování	33
4.15	Signál č. 1568 s polohami R kmitů před korekcí	34
4.16	Signál č. 1568 po korekci poloh R kmitů	34
4.17	Průměrné hodnoty siluet pro různé počty shluků	35
4.18	Výsledek shlukování	35

Seznam tabulek

5.1	Kontingenční tabulka k výsledkům shlukování	36
5.2	Dopočítané parametry	37
5.3	Kontingenční tabulka po korekci	37
5.4	Dopočítané parametry po korekci	38

Úvod

Tématem bakalářské práce je aplikace shlukové analýzy na signály EKG. Pozornost je kladena především na nehierarchické metody, jako je například k-means. Cílem práce je navrhnout optimální algoritmus pro vyhodnocení elektrokardiogramu, který zaznamenanou periodickou srdeční aktivitu rozřadí do několika shluků, nebo ji ponechá v jednom společném. Rozhodování bude probíhat na základě různých parametrů podobnosti.

Součástí práce je také přehled patologických srdečních stavů a popis jejich průběhu do zachyceného signálu. Uvedené by mělo pomoci s orientací v signálech, rozlišení fyziologického od patologického stavu a v budoucnu možná i vést k automatické analýze spojené s diagnózou pacientova stavu.

1 Morfologie EKG

Tvar elektrokardiogramu je pro jeho hodnocení zcela zásadní, a to jak v klinice tak v přístrojovém zpracování. V následující kapitole je tedy popsán fyziologický průběh elektrického signálu srdeční periody a v kontrastu k němu popsány vybrané odchylky a jejich význam. Pakliže nebude řečeno jinak, bude se popisu držet záznamů z druhého srdečního svodu.

Zdrojem obsahu kapitoly je [1].

1.1 Fyziologické EKG

Křivka EKG u zdravého člověka sestává z několika základních částí. Tyto části jsou označovány jako vlny P, Q, R, S a T. Některé zdroje zmiňují i vlnu U, jejíž význam však není jasný.

Vlna P reprezentuje v záznamu depolarizaci síní. Ta se neděje samovolně, ale na popud přicházející ze SA (sinoatriálního) uzlíku, což je část vzrušivého myokardu, jež umí samovolně generovat akční napětí. Napětí se šíří přes síně ke komorám a je zakončeno na uzlíku AV (atrioventrikulárním) – jediném vodivém spojení mezi síněmi a komorami. Signál se šíří z atrioventrikulárního uzlíku Hissovým svazkem na Tawarova raménka, která přivádějí vzruch septem srdečním do oblasti srdečního hrotu a předávají jej Purkyňovým vláknům, jež se zakončují ve stěnách obou komor. Komplex kmitů Q, R a S pak odpovídá depolarizaci komor. Vlna T pak odpovídá repolarizaci komor, tedy uvedení membránových napětí na kardiomyocytech na hodnoty před depolarizací.

1.1.1 Popis EKG

Když už máme naměřená data od pacienta a jsou ve vhodném zobrazení, pak je i potřeba je umět správně interpretovat. Obvykle se při popisu postupuje následovně:

Rytmus

Rytmus je hodnocen podle místa, ze kterého vzruch vychází. Může být sinusový z SA uzlu, junkční z AV uzlu (dále dělený na horní, střední a dolní nodální) nebo komorový z Purkyňových vláken (idioventrikulární). Dále se hodnotí pravidelnost a frekvence.

AV převod

Při hodnocení AV převodu sledujeme prodlevu mezi vlnou P a QRS komplexem. Za fyziologických podmínek trvá tato izoelektrická perioda méně než 200 ms. Při prodloužení tohoto intervalu mluvíme o AV blokádách.

Sklon elektrické osy srdeční

Tato vlastnost je stanovována z končetinových svodů I, II a III. Jde o sumu výchylek napětí na EKG během QRS komplexu, promítnutou do frontální roviny. Jinými slovy se jedná o integrál napětí v čase od vlny Q po vlnu S. Vyneseme-li výslednou hodnotu do Einthovenova trojúhelníku a provedeme-li vektorový součet, získáme konečný vektor elektrické osy srdeční. Dle jeho sklonu ve stupních pak můžeme odhadnout přibližný náklon srdce v mediastinu. Elektrická osa srdeční směřuje nejčastěji doleva a dolů ve frontální rovině.

QRS komplex

Jak bylo uvedeno výše v kapitole 1.1, QRS komplex nám zobrazuje fázi depolarizace komor. Kontrakce srdečního svalu, která na depolarizaci navazuje, udržuje tok krve řečištěm. Celý komplex by měl trvat od 60 ms do 100 ms. V hrudních svodech můžeme sledovat progresi vlny R ve směru od V1 po V6.

ST úsek

Konec depolarizace komor před započítáním repolarizace se označuje jako ST úsek. Tato část EKG by měla být u zdravého srdce izoelektrická (nulová). Pokud tomu tak není, je to často známka patologie, jakou může být například ischemie myokardu.

Vlna T

Vlnou T se dostáváme k repolarizaci komor. Tato vlna by měla ve svodech I, II a V3 až V6 nabývat vždy hodnot se stejným znamenkem jako dominantní kmit QRS komplexu. V uvedených svodech je její negativní výchylka téměř vždy patologická.

QT interval

Poslední vlastností je délka QT intervalu. Délka jeho trvání je proměnlivá v závislosti na srdeční frekvenci. Jeho nepřiměřené prodloužení, nebo naopak zkrácení by mohlo být předzvěstí nebezpečných arytmií.

1.2 Patologické stavy

Patologií se rozumí téměř jakákoli odchylka od výše zmíněného úzu. Patologie mohou mít různý původ a také různé projevy v EKG signálu. V této části je kladen zřetel především na patologie, které se projeví změnou tvaru EKG křivky. Ty se dají v širším pohledu dělit na supraventrikulární a ventrikulární.

1.2.1 Fibrilace síní

Síňová fibrilace je způsobena neřízenou aktivitou atriálního myokardu, která může dosahovat frekvence větší než je 300 stahů za minutu. Jak je vidět na obrázku 1.1, v EKG signálu se to projevuje jako nepravidelné kolísání izoelektrické linie s absencí P vln. Ty totiž vznikají řízeným stahem síní.

V závislosti na činnosti komor se dále dělí na fibrilace s pomalou a rychlou odpovědí. V prvním případě je srdeční frekvence (nyní myšlená jako stahy komor) zpomalená, a to až pod 40 úderů za minutu. V případě druhém jsou QRS komplexy obvykle štíhlé a přicházejí s frekvencí, která může vzrůst i nad 95 stahů za minutu. Nebezpečí této arytmie spočívá v tom, že krev v síních neproudí a má tendenci vytvářet sraženiny, ty se pak mohou přesunout do periferie a způsobit ucpání tepny s následnou ischemií vyživované tkáně.



Obr. 1.1: Fibrilace síní s charakteristickým rozvlněním izoelektrické linie [2]

1.2.2 Síňový flutter

Stejně jako fibrilace síní je i síňový flutter supraventrikulární tachykardií. Na rozdíl od síňové fibrilace je flutter pravidelná aktivita síní a proto se na EKG ukazuje jako pravidelná salva P vln, která tvoří specifické „zuby pily“, viz 1.2. Frekvence síňového flutteru se pohybuje obvykle mezi 240 až 300 úderů za minutu. Vzhledem k tomu, že se jedná o tachykardii, je zde možný přesun aktivity na komory a způsobení kolapsu oběhu stejně jako je tomu u výše zmíněné fibrilace.



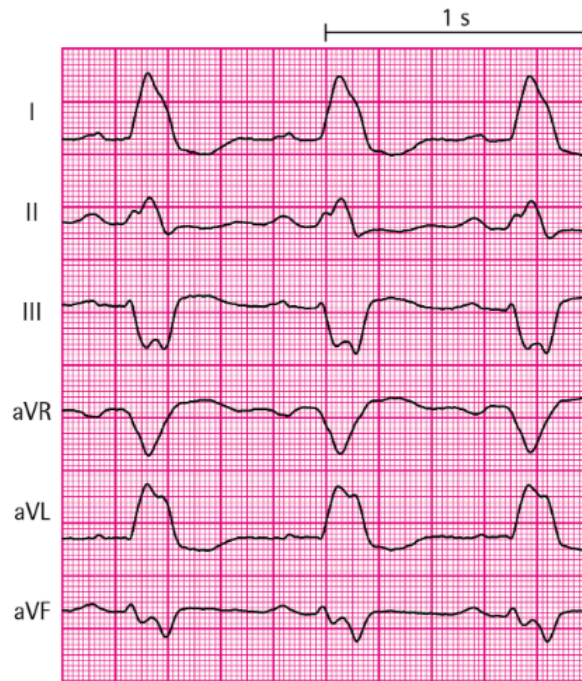
Obr. 1.2: Flutter síní s dobře rozeznatelnými „zuby pily“ [2]

1.2.3 Hypertrofie komor

Hypertrofie označuje zbytnění svaloviny komor, což bývá nejčastěji reakcí na chronickou objemovou nebo tlakovou zátěž komor. U hypertrofií se nejčastěji hodnotí výchylka dominantního kmitu QRS komplexu v různých svodech. Velikost výchylky koresponduje s právě depolarizovaným objemem buněk myokardu. Spolu s tím se také počítá tzv. Sokolowův index, který je vyjádřen pro každou komoru zvlášť. Pro pravou komoru je součet $R_{V_2}+S_{V_5}$ vyšší než 1,05 mV. U levé komory je Sokolowův index počítán trochu jinak, však obdobně: $S_{V_2}+R_{V_5}$ a jeho hodnota je u hypertrofie vyšší než 3,5 mV.

1.2.4 Raménkové blokády

Poruší-li se některé z Tawarových ramének, mluvíme o raménkových blokádách. RB (raménková blokáda) může být inkompletní, nebo kompletní podle toho, zda se vzruch šíří pouze pomaleji, nebo vůbec. Dále je rozeznávána pravá, levá přední nebo levá zadní podle provazce vodivého myokardu, který je zasažen. Kompletní RB se projevuje rozšířeným QRS komplexem a to nad 120 ms, viz 1.3. U inkompletních RB se délka QRS komplexu pohybuje v rozmezí 100-120 ms. Ve všech případech dochází k náklonu elektrické osy srdeční, což je důsledek pozmeněného šíření vzruchu srdcem. Jednotlivé podtypy mohou být spolu navzájem v kombinaci.



Obr. 1.3: Kopletní blokáda levého Tawarova raménka - délka QRS je vyšší než 120 ms [1]

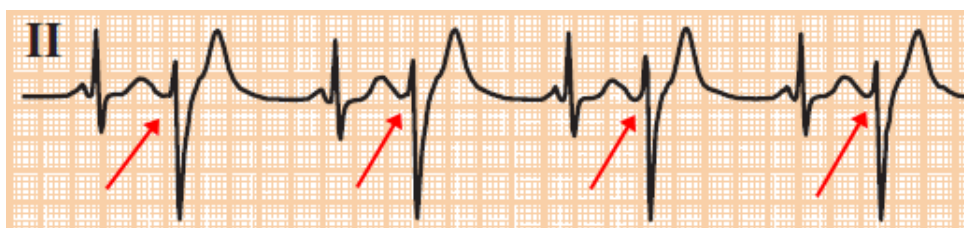
1.2.5 Ischemie myokardu

Patologickým uzávěrem věnčité tepny srdce dojde k nedostatečnému zásobení tkáně kyslíkem a živinami. Takováto tkáň se po krátké době stává ischemickou a začne odumírat. Odumřelá tkáň není elektricky aktivní ani kontraktlní. Z toho důvodu nepřispívá k výslednému elektrickému vektoru srdečnímu. Celá ischemie je dynamický proces a záleží v jakém momentu bylo EKG naměřeno. Hovoříme-li například o transmurální ischemii, můžeme průběh celé ischemie rozdělit do několika fází. V časném stádiu, které trvá několik prvních minut, můžeme zaznamenat hrotnaté a široké T vlny. V následujících šesti hodinách pozorujeme elevaci ST úseku v tzv. Pardeeho vlnu. Po více než šesti hodinách od počátku ischemie se začne zvětšovat záporná výchylka kmitu Q, hovoříme tedy o infarktovém Q. ST úsek se v této době normalizuje a T se stává zápornou. V chronickém stádiu perzistuje infarktové Q a vlna T se vrací do kladných výchylek.

1.2.6 Komorové extrasystoly

KES (Komorová extrasystola) je definovaná jako předčasný komorový stah způsobený vznikem vzruchu v ektopické lokalizaci v oblasti komor. KES mají rozšířený QRS komplex a změněnou morfologii, jako například na obrázku 1.4. Mohou být ojedinělé nebo vázané na sinusový stah, pak hovoříme o komorové bigemínii. Mohou se objevovat v párech nebo salvách případně mohou nasedat bezprostředně za QRS komplexem na vlnu T, což je popisováno jako fenomén R na T.

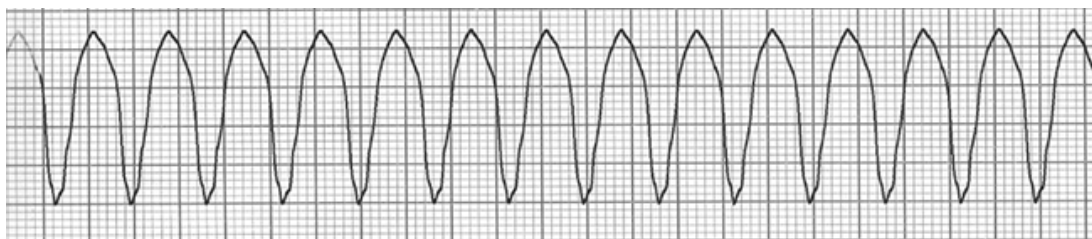
Tyto neřízené stahy se do jisté míry objevují i u zdravých lidí (asi jednou za pět minut). Jejich častější výskyt však zvyšuje pravděpodobnost, že udeří do tzv. vulnerabilní fáze (první polovina vlny T) a vyvolají komorovou fibrilaci.



Obr. 1.4: Komorové extrasystoly [2]

1.2.7 Komorová tachykardie

Komorová tachykardie patří mezi tachyarytmie, které se vyznačují zvýšenou tepovou frekvencí (nad 90 tepů za minutu). Jak už napovídá název, komorové tachykardie mají původ v ektopickém ložisku, které je umístěno v komorách. Další podmínkou, jež musí být splněna, aby se dalo hovořit o komorové tachykardii, je doba trvání, která by měla být vyšší než 30 sekund. Dale se dělí na tachykardie monoforní nebo polymorfní podle morfologie jednotlivých QRS komplexů v rámci záznamu jednoho pacienta. Všechny se však vyznačují rozšířenými QRS komplexy, které na sebe bezprostředně navazují. Komorovou tachykardii můžeme vidět na obrázku 1.5.

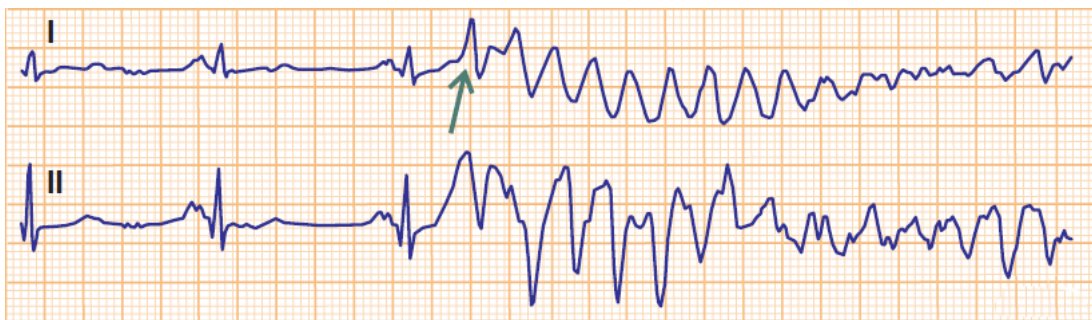


Obr. 1.5: Komorová tachykardie [2]

1.2.8 Fibrilace komor

Komorová fibrilace je život ohrožující stav, který je často příčinou náhlých úmrtí z plného zdraví. Je často způsobena akutní ischemií myokardu nebo setrvalou komorovou tachykardií. Z elektrofyziologického hlediska se jedná o chaotickou komorovou aktivitu způsobenou krouživými vzruchy v komorách. V tomto stavu je srdce hemodynamicky neúčinné, dochází proto k zástavě oběhu. Na EKG můžeme pozorovat absenci jakýchkoli komorových kmitů, jedná se pouze o náhodné chvění svaloviny (obrázek 1.6).

Pokud není osobě s touto arytmií poskytnuta první pomoc v podobě kardiopulmonální resuscitace s případným použitím defibrilátoru, je velice pravděpodobné, že zemře.



Obr. 1.6: Fibrilace komor [2]

2 Shluková analýza

Shluková analýza (anglicky *Cluster analysis*) je nástrojem pro vyhodnocování podobnosti mezi objekty na základě zvolené vzdálenostní metriky (např. Euklidovská, Manhattanská, atd.). Jejím úkolem je tedy najít vztahy mezi jednotlivými objekty a rozdělit je do skupin (shluků), které od sebe budou jasně oddělené. Cílem algoritmu mají být shluky, jejichž obsahem jsou objekty s největší podobností a zároveň docílit co největšího odlišení mezi shlucích. Shluková analýza (především hierarchické metody) se také využívá k hledání taxonomických struktur. Jako další příklady využití můžeme uvést segmentace obrazu, analýza sociálních sítí nebo trhu. [3]

V základu můžeme dělit shlukování podle přístupu k datům do dvou skupin, a to na nehierarchické a hierarchické.

2.1 Hierarchické metody

Algoritmy spadající do hierarchických metod na výstupu poskytují nejčastěji dendrogram, což je grafické znázornění vzdálenostních metrik mezi shlukovanými objekty.

Podskupinou těchto metod jsou metody aglomerativní, které vychází ze shluků tvořenými objekty jako takovými a jejich skládáním do větších skupin. Druhou takovou podskupinou jsou metody divizní, které naopak zařadí všechny objekty do jednoho shluku a ten pak v krocích dělí do shluků menších. Nevýhodou těchto metod je, že jakmile zařadí objekt do nějaké skupiny, již jeho přiřazení nemění (platí pro metody aglomerativní, u metod divizních platí, že jakmile je objekt oddělen od výchozího shluku již se jeho zařazení nemění).

2.2 Nehierarchické metody

Proti hierarchickým metodám stojí metody nehierarchické, které se snaží realizovat co možná nejlepší rozklad skupiny objektů do shluků. Ideálním řešením by bylo vyzkoušet všechny možné kombinace, to však není postupem optimálním. Proto nehierarchické metody pracují nejčastěji s již předem daným počtem shluků, jejichž složení se postupně optimalizuje. Algoritmus je ukončen jakmile se skladba shluků již po několik iterací nemění nebo se dosáhne předem nastaveného maximálního počtu kroků. Dále pak můžeme sledovat vývoj dříve definovaného kritéria, které klesne pod určitou hodnotu (příkladem může být třeba součet druhých mocnin vzdáleností mezi objekty ve shlucích. Jednou z nehierarchických metod je *k-means*, která bude rozebrána v následující podkapitole a je zároveň stěžejní pro tuto práci.

2.2.1 Metoda k-means

K-means (metoda k-středů) je založená na přiřazování objektů k definovaným centroidům, jež představují těžiště shluku. Poloha středů se průběžně mění tak, aby byla co možná nejmenší vzdálenost mezi nimi a objekty, které jsou k nim přidruženy. Máme-li tedy v prostoru definovaném třemi vlastnostmi N objektů a chceme je rozdělit do K shluků, zvolíme výchozí souřadnice každého centroidu (postupů je vícero, předpokládejme nyní, že jsou roztroušeny náhodně mezi objekty) a iteračně je měníme v závislosti na vzdálenostní metrice, kterou si zvolíme. Postup tedy spočívá ve výpočtu vzdálenosti mezi centroidem a objektem, přiřazením objektu k centroidu, ke kterému je nejbližší a přepočtení souřadnic centroidů podle objektů jež mu náleží. Nová poloha centroidu je obvykle průměrem souřadnic objektů shluku. Zmíněný postup je vyobrazen na obrázku 2.1. Celý tento proces se opakuje, dokud není splněna nějaká z podmínek, které jsou uvedeny výše. Nevýhodou této konkrétní metody je fakt, že musíme předem znát počet shluků, do kterých chceme objekty rozřadit. Druhým parazitním jevem může být závislost výsledku na původní poloze centroidů.[4]



Obr. 2.1: Algoritmus metody k-means [5]

Adaptací této metody je tzv. fuzzy c-means metoda, která nezařazuje jednotlivé objekty pevně do jednoho shluku, nýbrž každému objektu přiřazuje stupeň příslušnosti, a to pro každý shluk. Při rozdělování do dvou shluků může nastat situace, kdy objekt A patří do shluku prvního na 70 % a do druhého shluku na 30 %.

2.3 Hodnocení shlukování

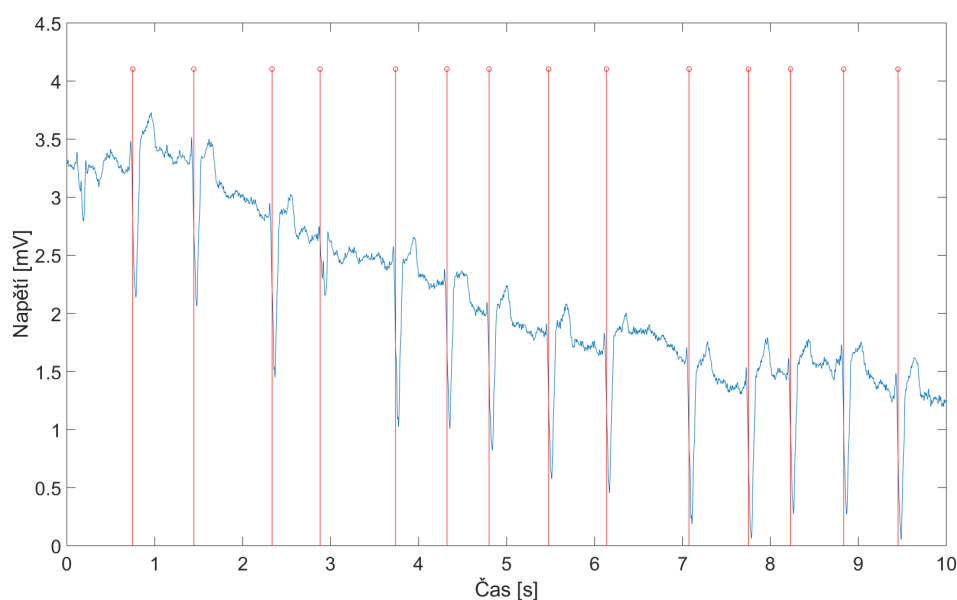
Po proběhnutém shlukování je potřeba vyhodnotit jeho úspěšnost. Asi nejjednodušším řešením tohoto problému je hodnocení pohledem, které je však zatíženo i největší subjektivitou. Zároveň taky záleží na zkušenostech vyhodnocujícího. Pro objektivní posouzení může posloužit například analýza siluet. Silueta vyjadřuje s jakou mírou je objekt zařazen do daného shluku. Vypočte se následujícím vztahem [5]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (2.1)$$

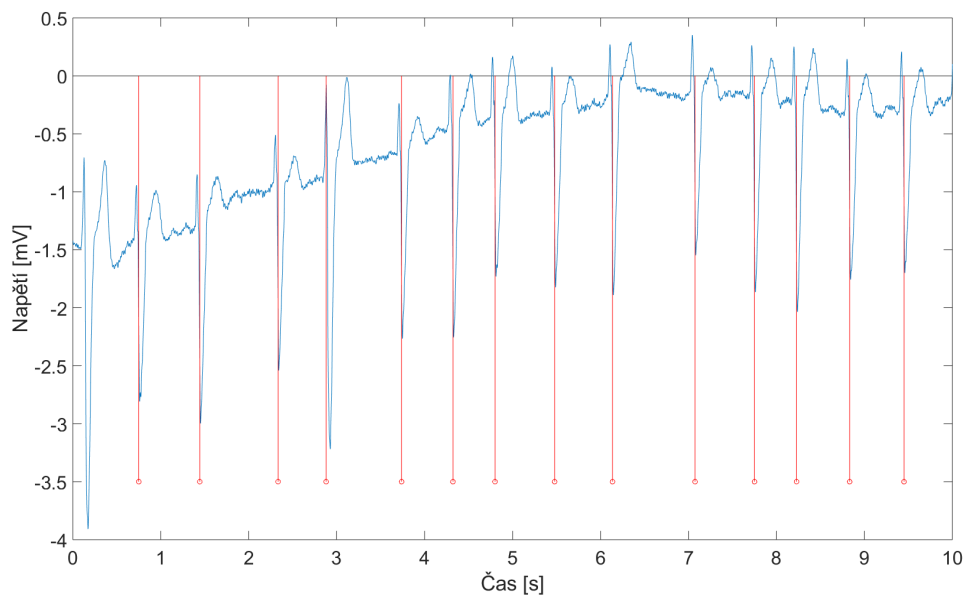
kde i představuje vybraný objekt, $b(i)$ průměrnou vzdálenost mezi objektem a všemi objekty z jiného shluku a $a(i)$ průměrnou vzdálenost mezi objektem a zbylými objekty v jeho vlastním shluku. Je počítána pro každý bod zvlášť a její průměrná hodnota v rámci jednoho shluku by měla být co nejvyšší při maximu 1. Čím nižší je její hodnota, tím je více pravděpodobné, že daný objekt je zařazen do příslušného shluku náhodou. Sledováním proměnlivosti průměrné hodnoty siluety pro různé počty shluků můžeme docílit optimálního počtu shluků pro danou množinu objektů.

3 Databáze EKG dostupná na UBMI

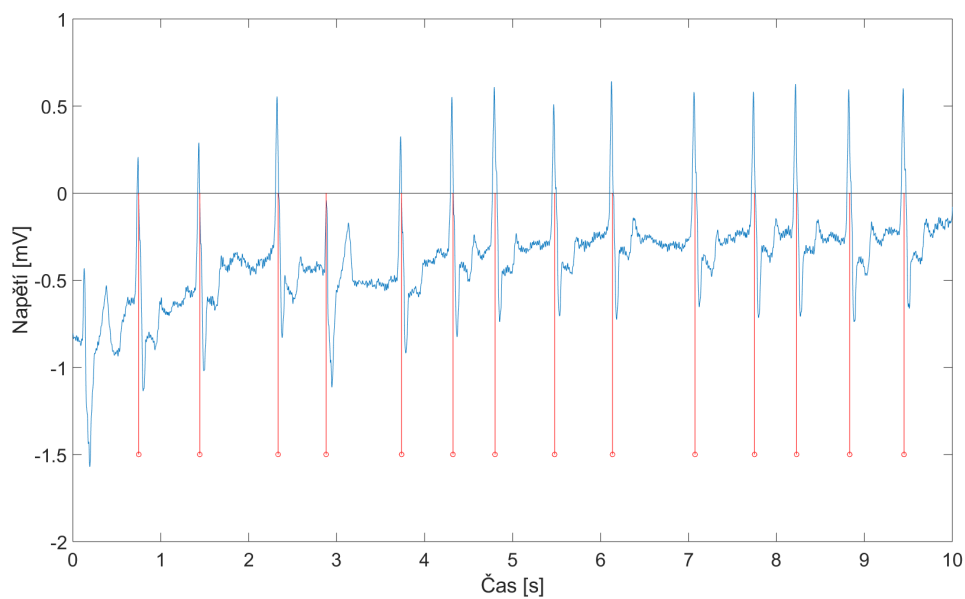
K mojí práci mi byl poskytnut soubor EKG záznamů naměřených firmou BTL. Jedná se celkem o 6884 desetisekundových úseků EKG snímaných osmi svody (konkrétně II, III, V1-6) se vzorkovací frekvencí 500 Hz. Dle přiloženého dokumentu je obsahem každého souboru raw záznam EKG a k němu přiložená automatická analýza firemním softwarem. Součástí analýzy bylo základní rozměření signálu s detekcí R vln. Tato detekce byla pravděpodobně uskutečňována pomocí jednoho vypočteného (syntetického) svodu. Tuto teorii podporuje fakt, že se pozice R vln neshodovala s pozicí největší výchylky v jednotlivých svodech. Krátká ilustrace zmíněného je na obrázcích 3.1, 3.2 a 3.3, kde jsou vyneseny svody II, V4 a V6. Indexy R vln se téměř vůbec neshodují s maximální výchylkou kmitu.



Obr. 3.1: Signál č. 4563, svod II

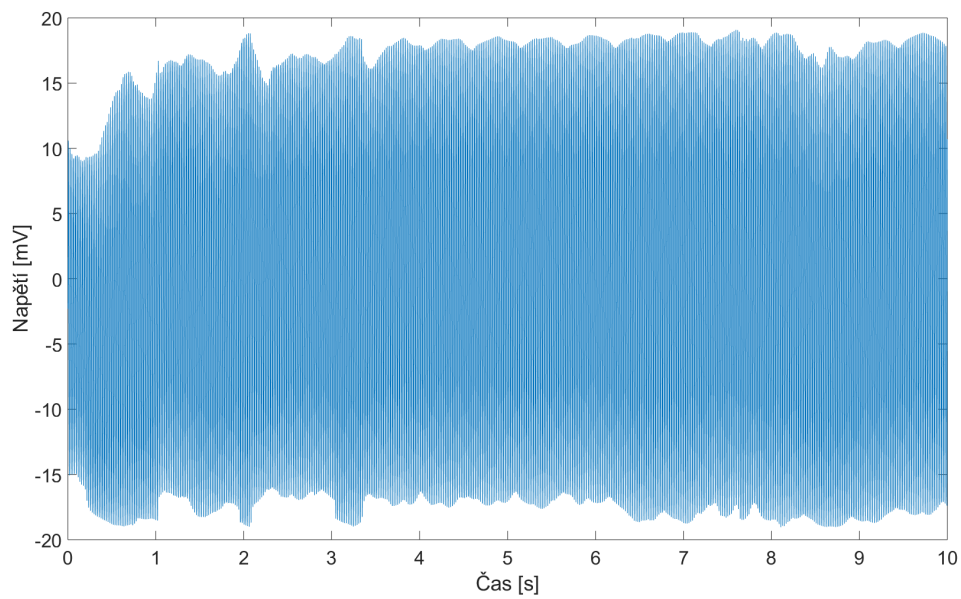


Obr. 3.2: Signál č. 4563, svod V4



Obr. 3.3: Signál č. 4563, svod V6

Součástí dat je také anotace. Ve většině případů se jedná o záznamy sinusového rytmu, dále se pak v záznamech objevují různé tachykardie, bradykardie, AV blokády různých stupňů, záznamy extrasystol a také atriálních fibrilací a flutterů. Objevují se však i záznamy, které obsahují jenom šum, jako například na obrázku 3.4.

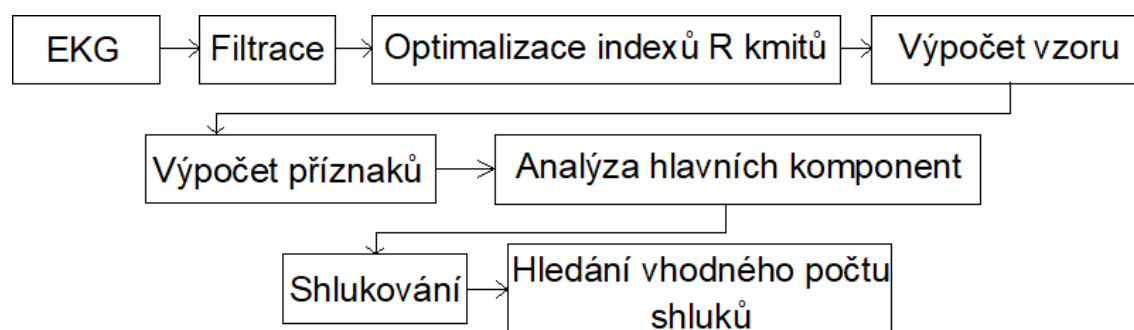


Obr. 3.4: Ukázka signálu obsahujícího pouze šum (signál č. 100, svod II)

4 Programové řešení

Jak již bylo řečeno v úvodu, cílem této práce je navržení algoritmu, jež bude schopen na základě shlukové analýzy rozdělit jednotlivé srdeční cykly (konkrétně QRS komplexy) do skupin za použití podobnostních metrik. Tato kapitola tedy obsahuje popis programového řešení (viz na obecném schématu 4.1) a také interpretaci dosažených výsledků. Každý záznam v databázi je jedinečný a proto se algoritmus musí každému zvlášť přizpůsobit. Algoritmus byl zpracován ve vývojovém prostředí Matlab.

Všechny kroky, které zde budou popsány se týkají svodu II. Právě tento svod je fyziologicky paralelní s elektrickou osou srdeční, viz kapitola 1.1.1. Díky tomu v něm můžeme pozorovat nejčastěji největší výchyly v signálu.



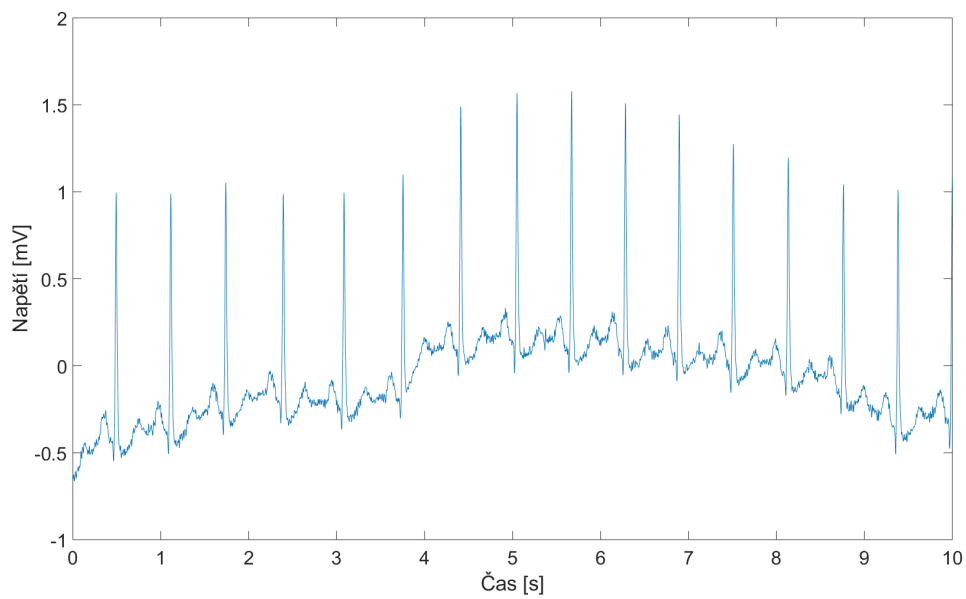
Obr. 4.1: Schéma algoritmu

4.1 Filtrace

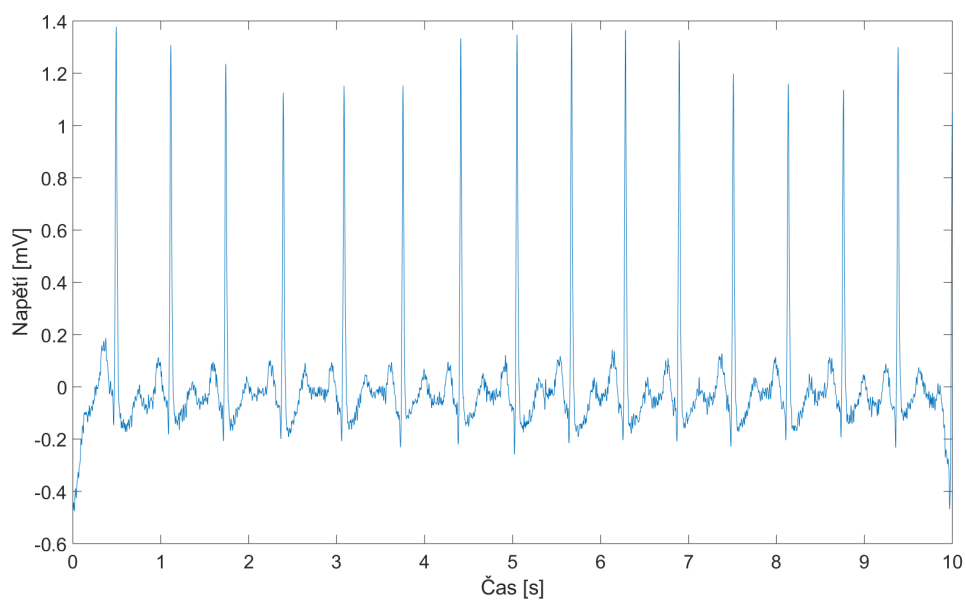
Velká část souborů obsahuje rušení v podobě kolísání izoelektrické linie. Z tohoto důvodu byla zařazena do algoritmu i filtrace. Jak je vidět na obrázcích 4.2 a 4.3, filtr typu horní propust s mezní frekvencí 1,2 Hz odstraní tato kolísání.

Na druhé straně frekvenčního spektra je zařazen filtr typu pásmová zádrž na frekvenci 50 Hz pro odstranění síťového rušení. Metodou filtrace je nulování spektrálních čar, která má tu výhodu, že je dosaženo ideální přenosové charakteristiky signálu na nefiltrovaných frekvencích. Nevýhodou této metody je její neaplikovatelnost v reálném čase, a to kvůli nutnosti provedení Fourierovy transformace, která vyžaduje celý záznam. Vzhledem k tomu, že algoritmus pracuje s již naměřenými záznamy, nebude tato nevýhoda omezením. [6]

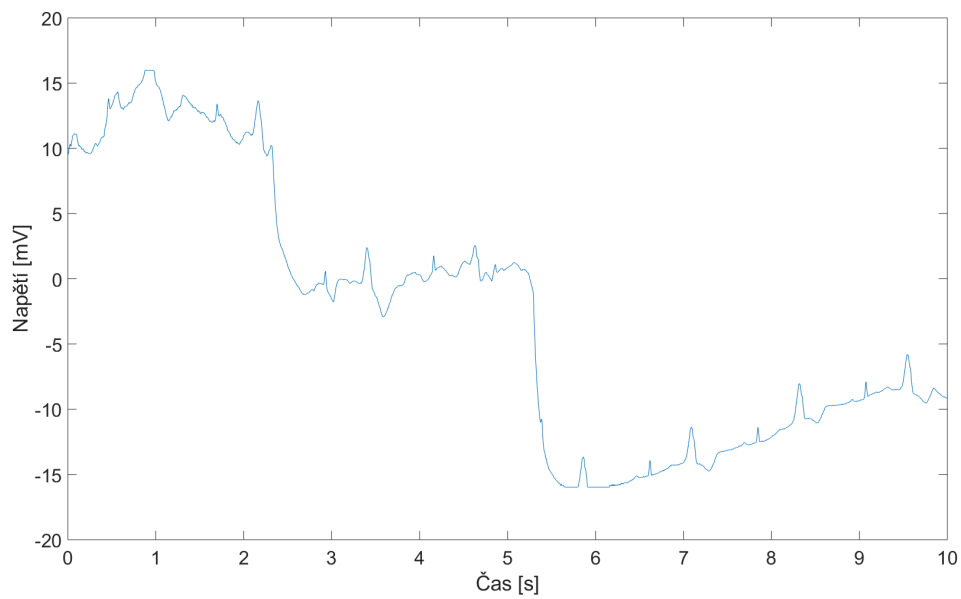
Někdy jsou záznamy rušené natolik, že filtrace jednoduše není účinná. Takovým případem je signál číslo 4834 na obrázcích 4.4 a 4.5. Signály, u nichž byla filtrace neúčinná a tento fakt byl odhalen, byly vyňaty ze závěrečného hodnocení.



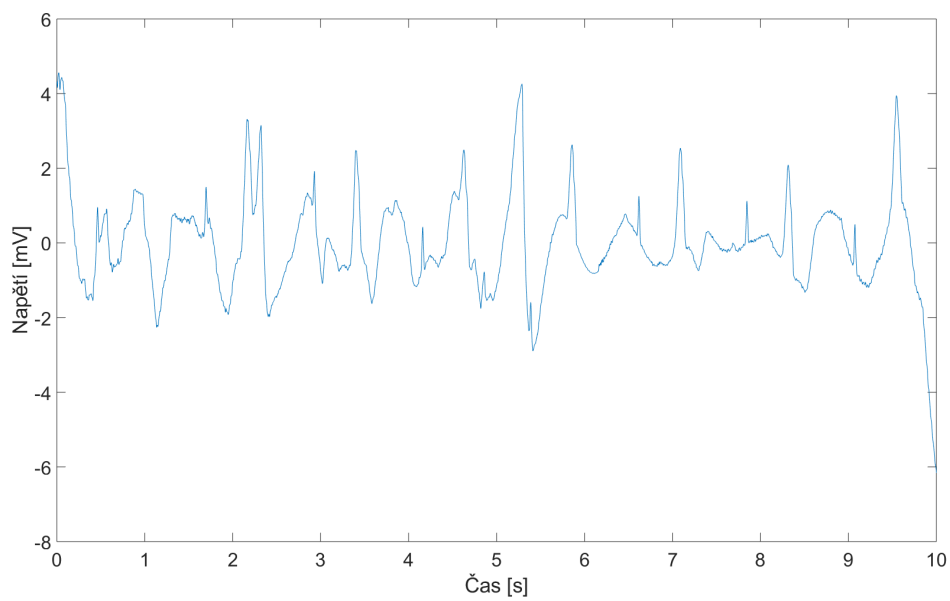
Obr. 4.2: Signál číslo 955 před a po filtraci



Obr. 4.3: Signál číslo 955 po filtraci



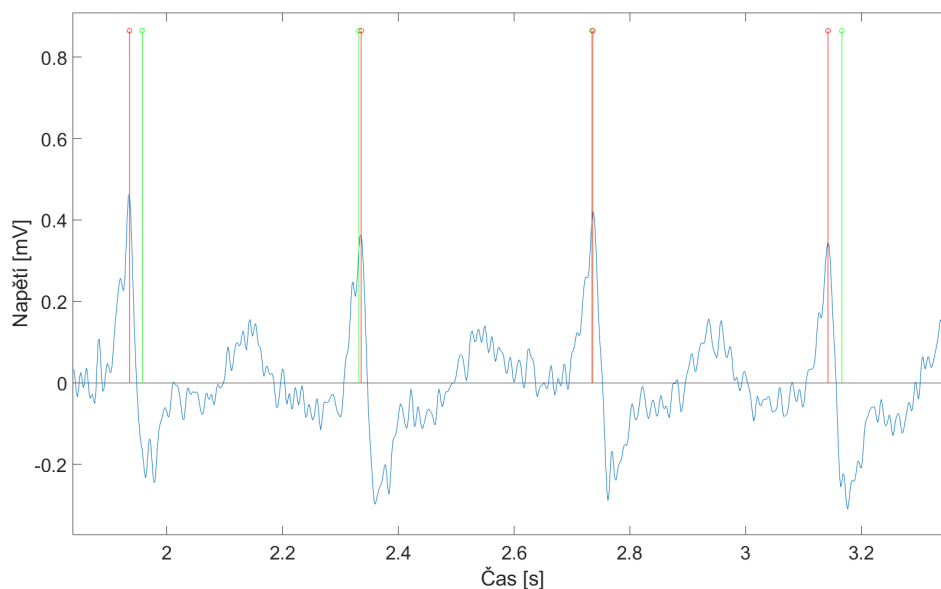
Obr. 4.4: Signál číslo 4834 před filtrací



Obr. 4.5: Signál číslo 4834 po filtraci

4.2 Optimalizace detekce R vln

Ve třetí kapitole je zmíněno, že v záznamech od firmy BTL jsou určeny pozice R kmitů, i to, že tato detekce nebyla příliš přesná, viz Obr. 4.6. Aby tento fakt následně ovlivňoval práci algoritmu co nejméně a v nejlepším případě ji neovlivňoval vůbec, je třeba detekci poupravit a indexy R kmitů přepsat. Tento krok lze provést výpočtem okolí detekovaného R kmitu a v něm najít maximální výchylku signálu a to buď kladnou nebo zápornou. Byly tedy rozměřeny vzdálenosti mezi jednotlivými R kmity, z nich byla vybrána ta nejmenší, která se podělila dvěma. Výsledek se zaokrouhlil dolů. Tím byl získán koeficient, který určuje okolí R kmitu na obě strany a je stejný pro celý záznam. V tomto úseku signálu bylo vybráno maximum a přepsána pozice R kmitu. U některých signálů se výška R kmitu a následného S kmitu (případně Q kmitu následovaného R kmitem) téměř shodovala s tím, že občas měl větší výchylku první a občas druhý. Protože tato nepřesnost by později zásadně ovlivnila vzdálenostní metriky, byl porovnán počet kladných a stejně tak záporných výchylek v analyzovaných úsecích kolem R kmitu. Pokud bylo více kladných kmitů, byly v úsecích s většími Q nebo S kmity nalezeny nejvyšší kladné výchylky et vice versa s vyšším počtem záporných výchylek. Přidáním tohoto kroku se zpřesní přepisování indexů R kmitů. Úseky signálu, jejichž R kmity jsou odlišné, ale správně zdetekované se tímto sice přepíše podle majoritní skupiny, ale to neovlivní výsledné vyhodnocení jejich odlišnosti.



Obr. 4.6: Detekce R kmitu v signálu č. 1968 algoritmem zabudovaným do měřícího software (zeleně) a po optimalizaci navrženým algoritmem (červeně)

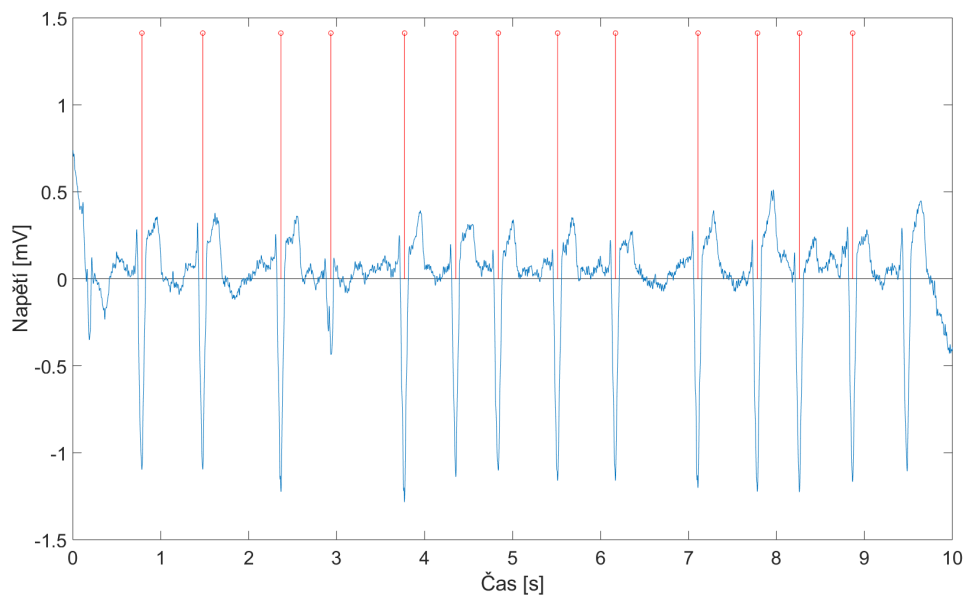
Je zapotřebí uvést, že v některých záznamech se R kmity vyskytují skoro na jeho začátku nebo konci, a tudíž by pro nedostatek hodnot před nimi nebo za nimi bylo získáno velmi malé okolí. Uvedené bylo ošetřeno podmínkou, že QRS komplex nebude uvažován, pakliže je jeho vzdálenost od kraje záznamu nižší než průměr a medián všech vzdáleností mezi R vlnami v záznamu. Tímto krokem se sice mohou ztratit některá data, ale na druhou stranu bude zajištěno, že velikost porovnávaných úseků QRS komplexů, které vstoupí do shlukování, nebude omezena nedostatkem hodnot z okrajů signálu.

Dalším faktem, který ovlivní závěrečné hodnocení funkčnosti algoritmu je, že někdy byly označeny za R kmit i jiné morfologické útvary v záznamu, a to nejčastěji vlna T. S touto variantou algoritmus neumí pracovat. Naštěstí se tato situace netýkala mnoha záznamů z databáze BTL. Z celkového počtu to bylo necelých 200 záznamů.

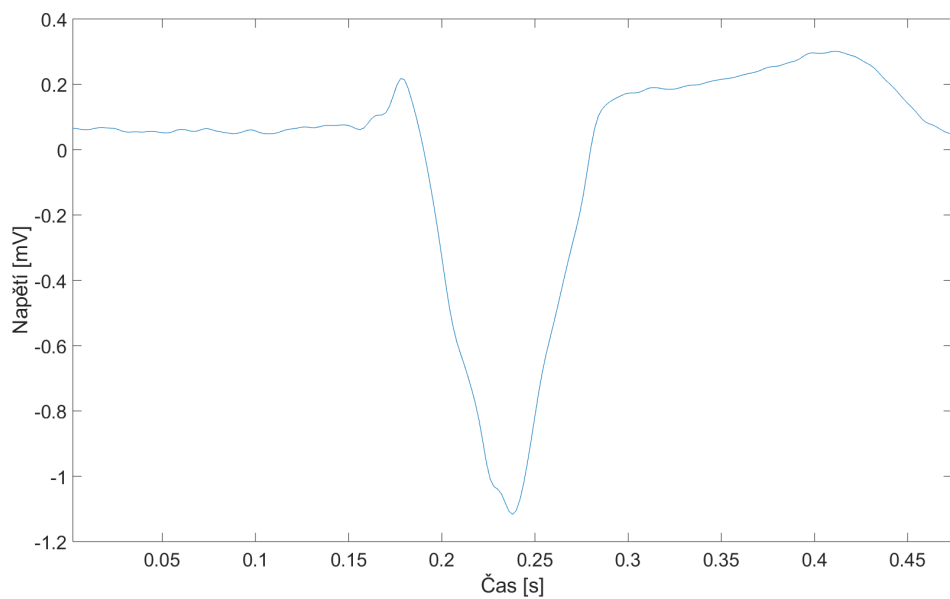
4.3 Vytvoření vzoru

Aby se srdeční cykly v rámci jednoho záznamu daly porovnávat, je třeba vytvořit umělý vzor, se kterým budou srovnávány. Ten je vytvořen kumulací jednotlivých cyklů v záznamu (obrázky 4.7 a 4.8). Jako jeden cyklus je uvažovaný detekovaný R kmit s jeho okolím, které bylo definováno v kapitole 4.2. Nejprve jsou zkumulovány cykly v základním signálu. Ten je následně zderivován a je vytvořen vzor i pro signál derivovaný. Nakonec jsou jednotlivé cykly převedeny na spektra, ze kterých se také udělá kumulací vzorové spektrum. Tímto krokem vznikly tři vzory pro tři podoby signálu. Základní signál ukazuje časovou posloupnost jednotlivých elektrických dějů v srdci. Derivovaná podoba pak ukazuje dynamiku v signálu. Spektra jednotlivých úseků ukazují frekvenční zastoupení, které se bude měnit s rozdílným tvarem křivky.

V závislosti na RR intervalu se bude měnit délka vytvářeného vzoru. Pakliže srdce bilo rychle, budou RR intervaly kratší a tudíž se zkrátí i vypočtené okolí. Při tachykardii je možné, že se zkoumaný úsek omezí pouze na QRS komplex a nebudou do ní zavzaty zbylé vlny. Zároveň je zde riziko špatně zdetekovaného R kmitu v původních datech. Pokud by byl detekován R kmit a hned za ním i T vlna, bylo by zkoumané okolí mnohem kratší. Výsledek bude tím pádem neprůkazný nebo i nesmyslný. Jak bylo již řečeno výše (kapitola 4.2), zmíněný problém se týká minoritního počtu záznamů.



Obr. 4.7: Signál č. 4563 s detekovanými S kmity



Obr. 4.8: Vzor vytvořený kumulací cyklů signálu č. 4563

4.4 Metriky podobnosti

Již vytvořený vzor bude následně porovnáván s jednotlivými cykly za účelem zjištění míry podobnosti. Celkem je vytvořeno šest metrik (příznaků), jimiž jsou Pearsonův korelační koeficient a střední kvadratická odchylka v základním a derivovaném signálu a ve spektru. Hodnoty korelačního koeficientu budou stoupat s vyšší podobností, zatímco kvadratická odchylka bude klesat. Všechny vypočtené hodnoty jsou následně převedeny do rozsahu od nuly do jedné, aby byly lépe srovnatelné. Nyní program může pracovat se šestirozměrným prostorem, ve kterém může proběhnout shlukování. Je však velice možné, že informace poskytnuté jednotlivými metrikami mohou být redundantní a ve výsledku i zavádějící. Pro výběr pouze relevantních z nich je do algoritmu zařazena PCA (analýza hlavních komponent).

4.5 Analýza hlavních komponent

Pakliže existuje popis obecného objektu, který je vytvořen z více úhlů pohledu, může se stát, že se některé informace v těchto náhledech překrývají. V takovém případě ne všechny pohledy přispívají k celkovému popisu a tudíž nejsou nezbytně nutné pro rozhodování. Řešení je dvojit. Lze provést selekci příznaků, kdy se vyberou pouze ty, které v sobě mají relevantní informace a nebo příznaky extrahovat.

Extrakce, kam se řadí i analýza hlavních komponent (PCA), je postup, kdy jsou z vypočtených příznaků vypočteny nové, které však obsahují majoritní část relevantní informace.

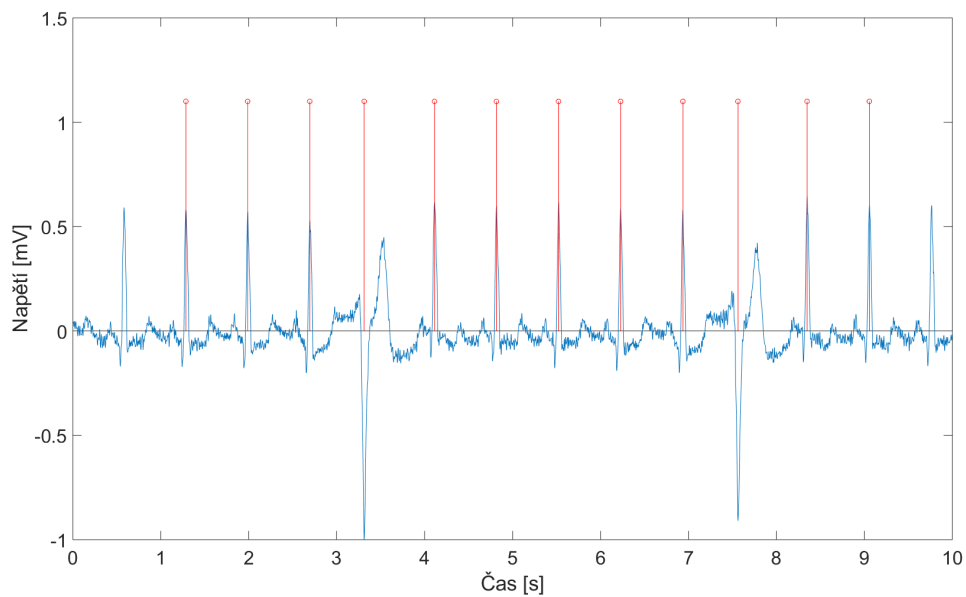
Výstupem z předchozího kroku je charakteristika každého srdečního cyklu pomocí šesti příznaků. Tyto hodnoty jsou umístěny do matice. Je zhotovena autokovarianční matice. Následně jsou vypočteny vlastní vektory a čísla pro každý příznak. Vlastní čísla ukazují na komponenty s největší variabilitou v datech, tudíž na ty komponenty, které v sobě obsahují nejvíce informace pro pozdější rozhodování. Pomocí vlastních čísel je vybráno tolik komponent, aby v sobě obsahovaly alespoň 95 % z původní informace (tj. celkové variability dat). Vybraná čísla k sobě mají vlastní vektor. Maticovým vynásobením vlastních vektorů a matice původních příznaků vznikne nová matice, ve které jsou jednotlivé objekty (QRS komplexy s přílehlým okolím) charakterizovány novými extrahovanými příznaky, jež jsou vzájemně minimálně korelované. Tím vznikl i nový prostor příznaků, ve kterém bude probíhat shlukování. Nevýhodou této metody je, že není vždy zaručen přínos nejvýznamnější komponenty pro rozlišení objektů. Na druhou stranu komponenty, které mohli být původně považovány za redundantní, mohly k rozlišení významně přispět. [7]

4.6 Shlukování

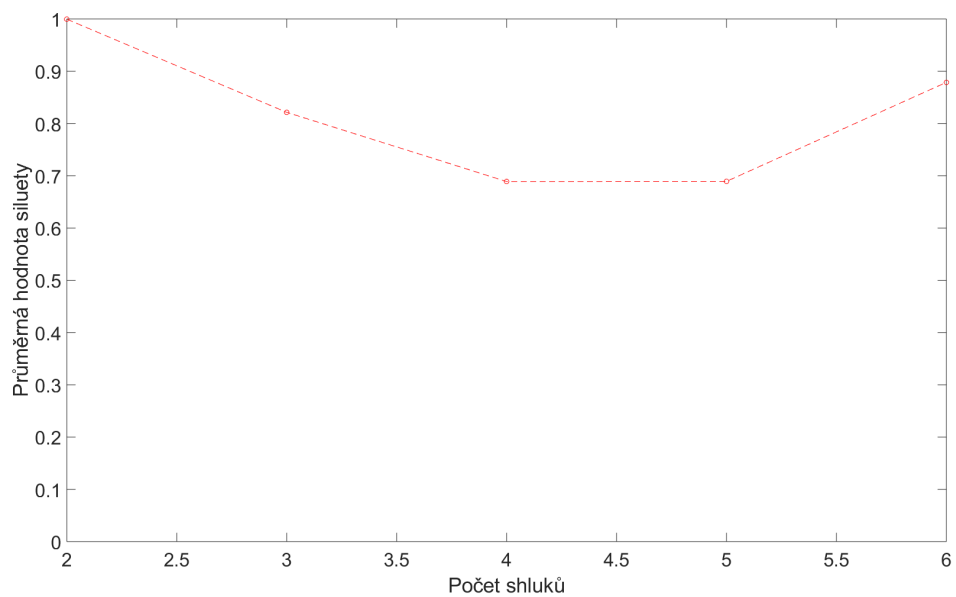
Nyní máme QRS komplexy reprezentovány v nových souřadnicích, které poskytla PCA a na kterém proběhne shlukování metodou K-means. Metoda však vyžaduje znát počet shluků, do kterých se budou data rozřazovat. Bude se tedy vycházet z premisy, že v záznamu se nevyskytuje každá morfologie QRS komplexu ojedinele a maximální počet shluků N bude nastaven na polovinu všech cyklů, které budou hodnoceny. Dále proběhne shlukování pro počet shluků dvě až N . Z každého výsledku shlukování budou spočteny siluety. Z průměrné hodnoty siluet se dále bude moci usoudit, jaký počet shluků je optimální. Limitem tohoto postupu je fakt, že vyhodnotí pouze varianty pro dvou a více shluků. To jestli jsou QRS komplexy dostatečně podobné pro zařazení do jednoho shluku, musí být vyvozeno až z toho, že pro ně není vhodný žádný z vyšších počtů shluků. Pokud shlukování proběhlo dobře blíží se hodnota průměrné siluety jedné. Pohybuje-li se kolem 0,5, nebylo shlukování úspěšné. S cílem dosáhnout co nejlepšího hodnocení u záznamů, kde se nevyskytují žádné odlišné srdeční cykly, byla hranice průměrné siluety postupným zkoušením nastavena na hodnotu 0,95. Tato hodnota značí, že shlukování proběhlo dobře. Byly však i případy, kdy silueta nevystoupala tak vysoko. Proto byla přidána druhá hranice, a to 0,86 s dodatečnou podmínkou, že siluety pro každý objekt zvláště nesmí být nižší než 0,825. Pakliže průměrná hodnota siluet pro nějaký počet shluků splní výše zmíněné podmínky, budou objekty rozřazeny do onoho počtu shluků.

Konečným výstupem algoritmu je okénko (*figure*), ve kterém se nachází informace v podobě grafu nebo grafů, na kterých jsou zobrazeny morfologie, které se vyskytují v záznamu a jejich četnost.

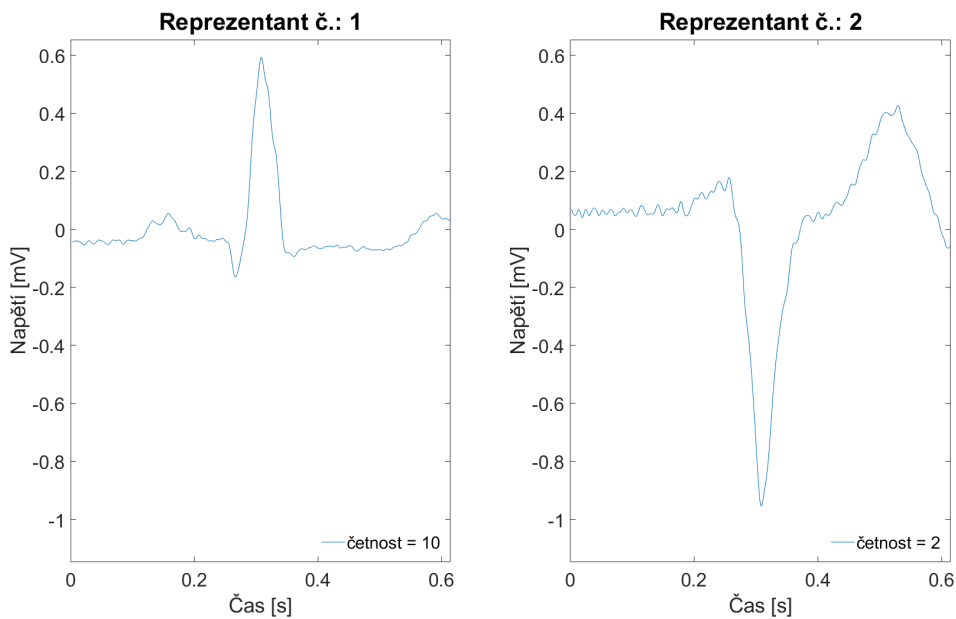
Názornou ukázkou může být záznam č. 1002 (obrázek 4.9), ve kterém je dle anotace přítomný jeden typ abnormálního QRS komplex, a to celkem dvakrát. Na obrázku 4.10 je zobrazen vývoj průměrné hodnoty silulet. Ta přesáhla hodnotu 0.99 při řazení do dvou shluků. Také je zde vidět zvýšení průměrné siluety v případě šesti shluků. To ukazuje na další variabilitu, která je mezi jednotlivými QRS komplexy. Tato variabilita je uvnitř větších shluků. Na dalším snímku (4.11) je ukázán výstup algoritmu. Jsou zde zobrazení reprezentanti dvou shluků, které vyšly jako ideální rozdělení z procesu shlukování. Reprezentanti sami jsou průměrem ze všech objektů nacházejících se v každém shluku.



Obr. 4.9: Signál č. 1002 s dvěma abnormálními QRS komplexy

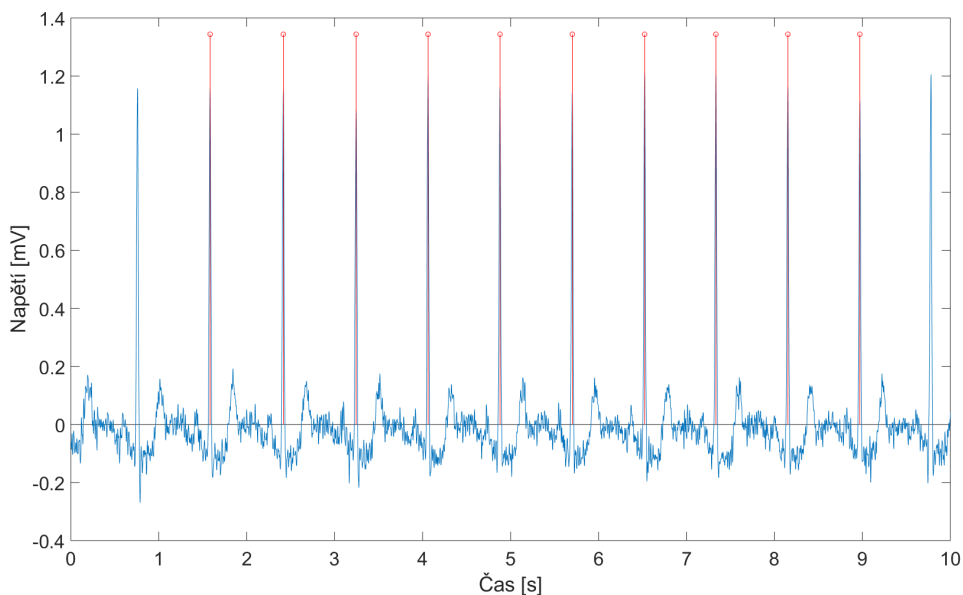


Obr. 4.10: Průměrné hodnoty siluety pro různé počty shluků

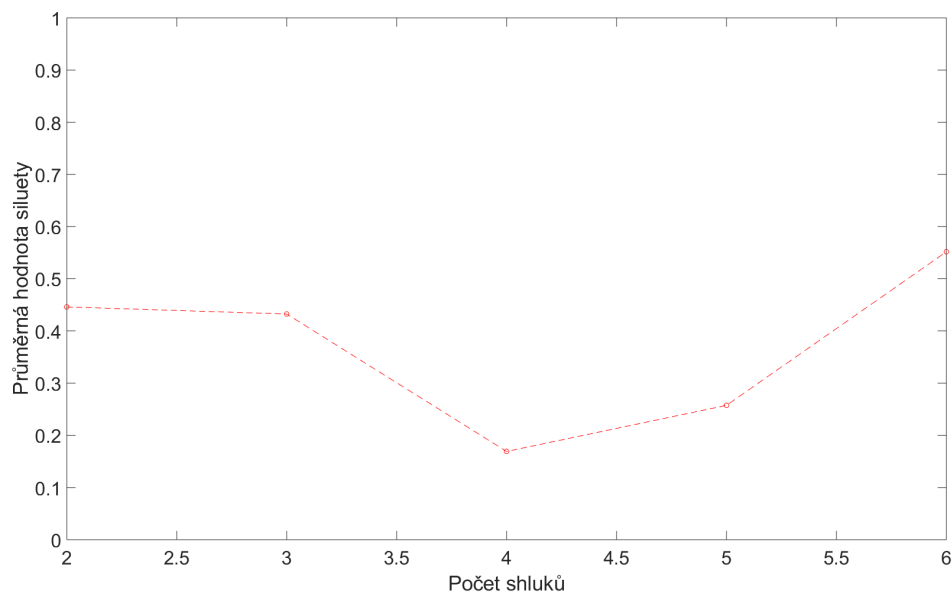


Obr. 4.11: Výsledek shlukování

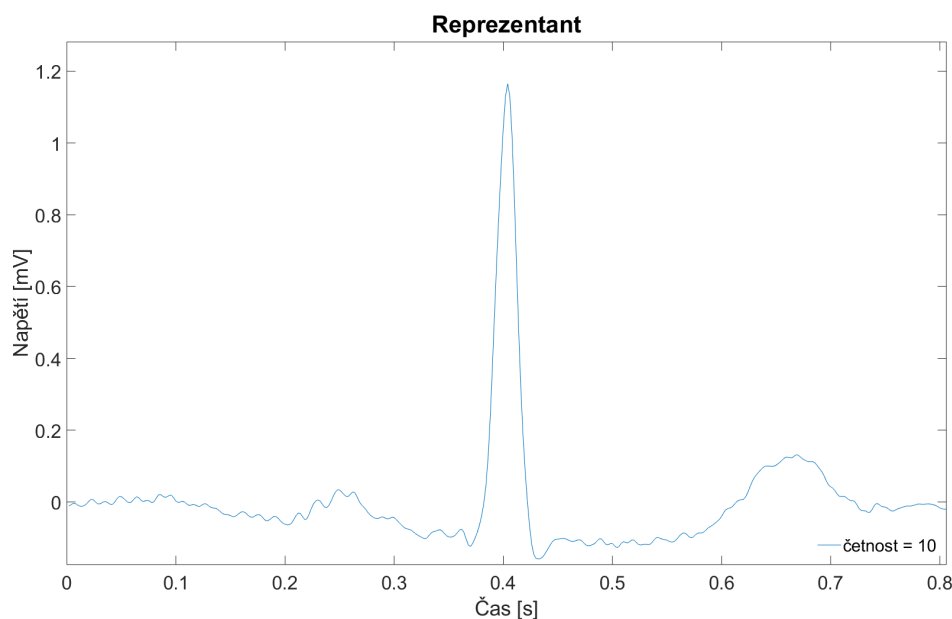
Pro porovnání je stejným algoritmem zpracován záznam sinusového rytmu bez abnormalit. Průměrná hodnota silulet nevystoupala pro žádný počet shluků nad hranici podmínek, což ukazuje, že žádný z vyšších počtů shluků není vhodný. Signál č. 2607, shlukování i průměrné hodnoty silulet jsou ukázány na obrázcích 4.12, 4.13 a 4.14.



Obr. 4.12: Signál č. 2607 s vyznačenými R kmity



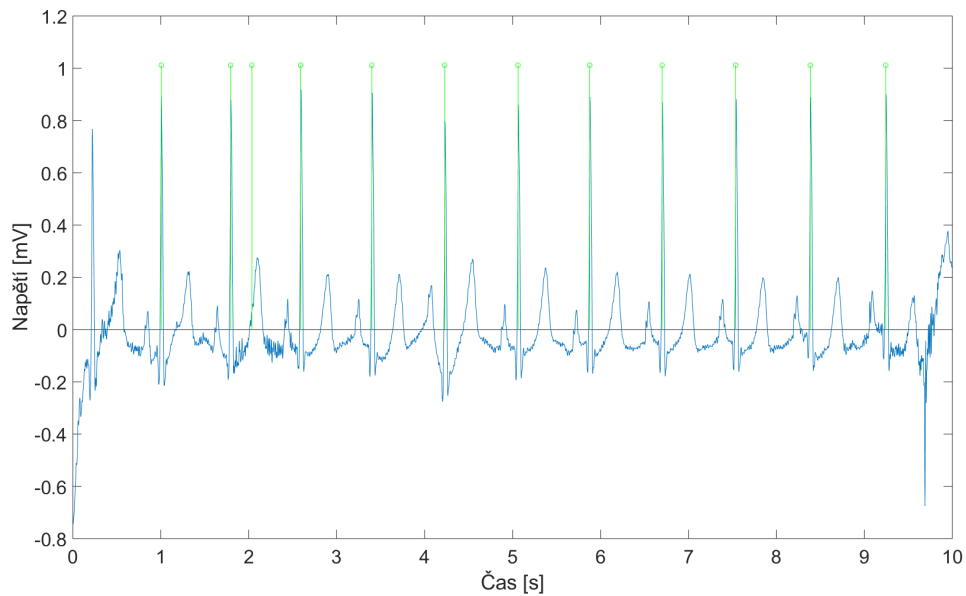
Obr. 4.13: Průměrné hodnoty siluet pro různé počty shluků



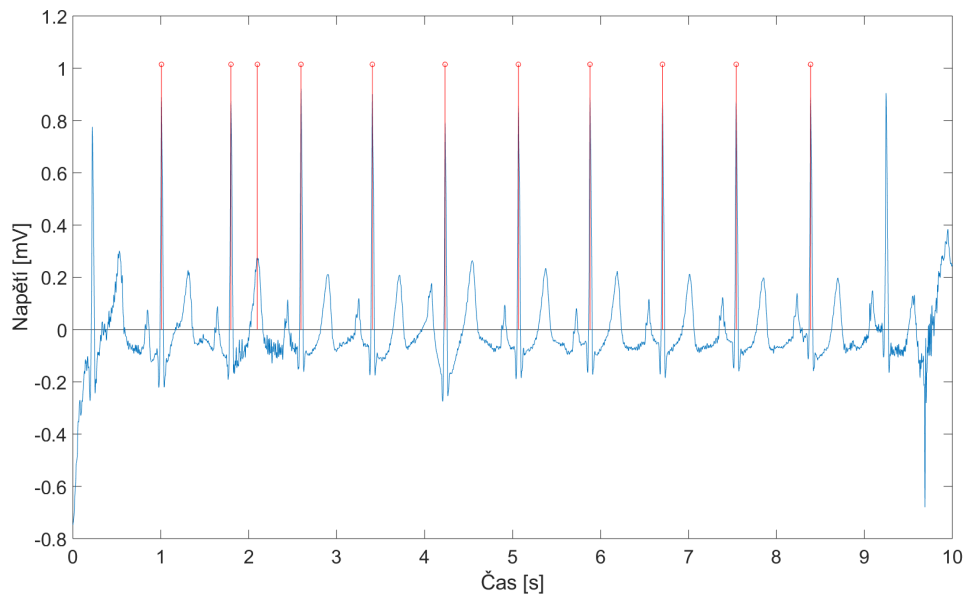
Obr. 4.14: Výsledek shlukování

Na snímku 4.15 je zobrazen záznam číslo 1568 spolu s detekovanými R kmity (zeleně). Je zřejmé, že jeden kmit je detekován špatně. Okolí, které bude určovat hodnocené úseky se tímto zkrátí na méně než polovinu délky RR intervalu v záznamu. Po korekci polohy R kmitu není vadná pozice smazána, ale přepsána do místa s nejvyšší výchylkou v okolí (v tomto případě vlna T), viz snímek 4.16. Podle vývoje

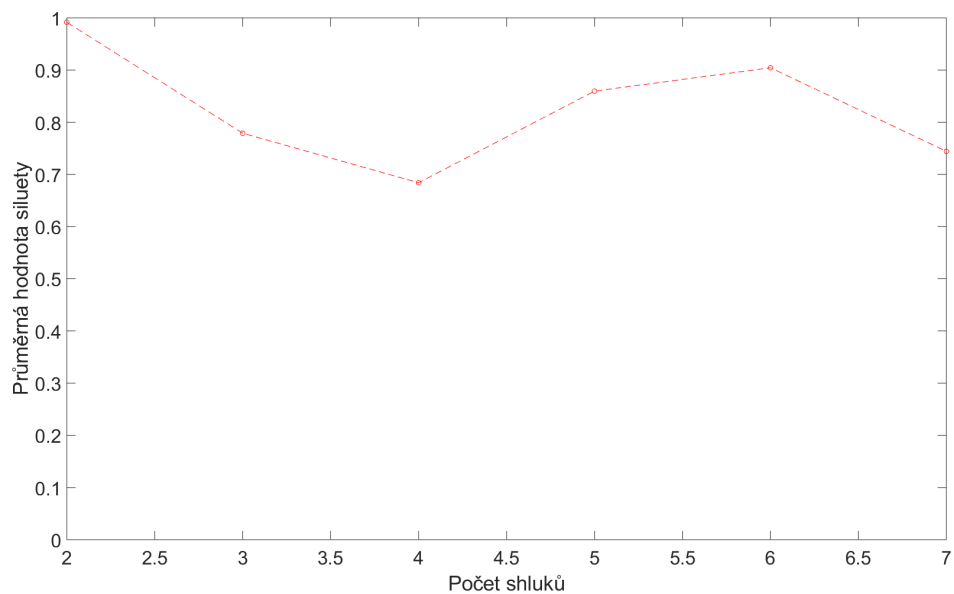
průměrné hodnoty siluet, který je zachycen na obrázku 4.17, je zřejmé, že v záznamu byla nalezena morfologická odlišnost v podobě T vlny. Výstup algoritmu (obrázek 4.18) je tedy správný, protože vlna T má nepochybně jinou morfologii oproti QRS komplexu. Zároveň je však výsledek nesmyslný.



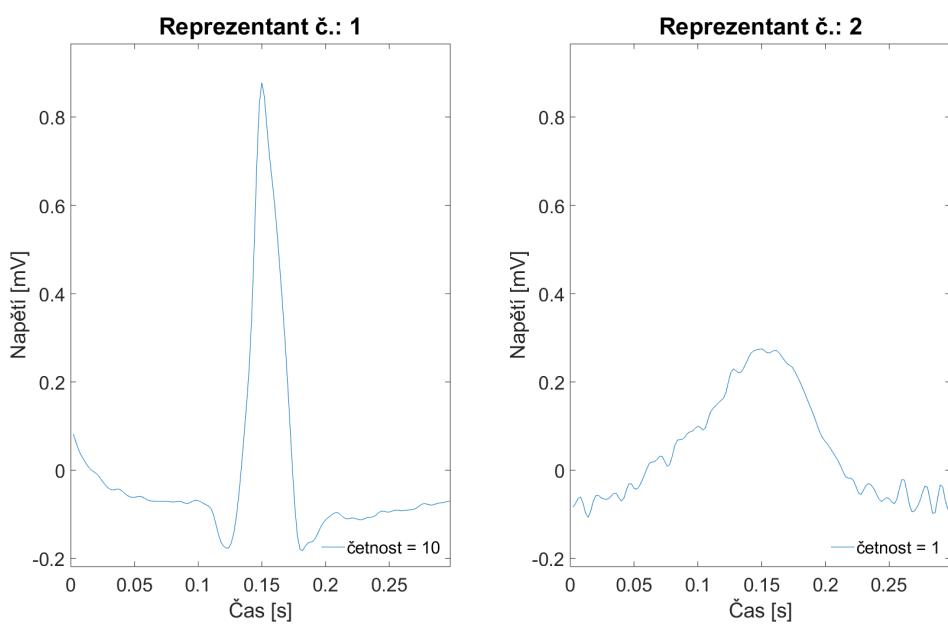
Obr. 4.15: Signál č. 1568 s polohami R kmitů před korekcí



Obr. 4.16: Signál č. 1568 po korekci poloh R kmitů



Obr. 4.17: Průměrné hodnoty siluet pro různé počty shluků



Obr. 4.18: Výsledek shlukování

5 Statistické zhodnocení

S nastavenými podmínkami pro zařazení srdečních cyklů do více shluků je možné přejít ke statistickému zhodnocení algoritmu. To bude probíhat na záznamech od firmy BTL a to konkrétně na svodu II. Z poskytnutých 6884 záznamů vstoupilo do analýzy pouze 6763, tedy jinými slovy 121 bylo ze statistiky vynecháno. Byly to záznamy, které byly špatně naměřené, nebyly k nim dodány indexy R kmitů nebo ani po provedeném předzpracování nebyly hodnotitelné. I přes vyřazení těchto nekvalitních záznamů by měla být statistika dostatečně vypovídající. Hodnocení algoritmu probíhalo srovnáním výsledku jak vizuálně se záznamem, tak s přiloženou anotací od firmy BTL.

V níže umístěné tabulce 5.1 je možné vidět výsledek hodnocení. V řádcích tabulky je zachyceno, zdali byl (+) nebo nebyl (-) v anotaci zapsán abnormální komorový komplex nebo jiná morfologická odchylka. Sloupce odpovídají výsledkům poskytnutých algoritmem, tedy jestli rozřadil srdeční cykly do více shluků (>1) či nikoli (1).

Tab. 5.1: Kontingenční tabulka k výsledkům shlukování

		Metoda		
		>1	1	Σ
Anotace	+	669	362	1031
	-	634	5098	5732
	Σ	1303	5460	6763

Z tabulky 5.1 je též patrné, že celkem 996 záznamů bylo nesprávně zařazeno a to buď jako falešně pozitivní nebo negativní. Chyby I. řádu, kdy se abnormality nevykytovaly, ale i přes to je algoritmus našel, byly často způsobeny nepřesnou detekcí R kmitů. Často se jednalo o špatné posouzení vlny T jako QRS komplexu. Naopak chyby II. řádu, kdy se abnormality vyskytovaly, ale nebyly detekovány, by se pravděpodobně daly eliminovat přidáním dalších hodnotících příznaků. Ty by mohly vnést další pohledy na srdeční cykly. Z tabulky 5.1 byly následně dopočítány parametry, a to konkrétně senzitivita (TPR), specificita (TNR), pozitivní prediktivní hodnota (PPN), negativní prediktivní hodnota (NPV) a správnost (ACC). Parametry byly vypočteny podle následujících rovnic [8]:

$$TPR = \frac{TP}{TP + FN}, \quad (5.1)$$

$$TNR = \frac{TN}{TN + FP}, \quad (5.2)$$

$$PPV = \frac{TP}{TP + FP}, \quad (5.3)$$

$$NPV = \frac{TN}{TN + FN}, \quad (5.4)$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \quad (5.5)$$

kde TP jsou správně určené záznamy s více morfologiemi, TN jsou správně určené záznamy s jednou morfologií, FP představují záznamy, které neobsahovali morfologicky odlišné QRS komplexy, ale byly tak vyhodnoceny a nakonec FN jsou záznamy s více morfologiemi, které nebyly nalezeny.

Tab. 5.2: Dupočítané parametry

TPR	64 %
TNR	88 %
PPV	51 %
NPV	93 %
ACC	85 %

Zanesené hodnoty v tabulce 5.2 ukazují, že algoritmus je poměrně specifický, bohužel už ne tolik senzitivní. Pozitivní prediktivní hodnota, která říká jaká je pravděpodobnost správného zařazení je nižší a hraničí s náhodným určením. Hranici náhody odpovídá 50 %. Naopak NPV ukazuje, že se v záznamu s vysokou pravděpodobností nevyskytuje více morfologií srdečních cyklů.

Jak je již výše nekolikrát zmíněno, algoritmus je citlivý na správnou detekci R kmitů. Jejich indexy se během algoritmu přepisují a součástí není ověření, zda-li konkrétní index je skutečně R kmit nebo jiný útvar v EKG (snahou bylo využít výsledky detekce QRS dodané se záznamy). Během hodnocení bylo těchto chybně zdetekovaných záznamů celkem 171. Za velmi optimistického předpokladu, že by byly zdetekované správně a shlukování proběhlo úspěšně, vypadaly by výsledky následovně.

Tab. 5.3: Kontingenční tabulka po korekci

		Metoda		
		>1	1	Σ
Anotace	+	840	362	1202
	-	463	5098	5561
	Σ	1303	5460	6763

Tab. 5.4: Dupočítané parametry po korekci

TPR	69 %
TNR	91 %
PPV	64 %
NPV	93 %
ACC	87 %

Po ruční korekci je v tabulce 5.4 vidět, že senzitivita dosahuje téměř k 70 % a specifita 91 %. Pozitivní prediktivní hodnota se také zvedla o 13 %.

6 Závěr

Výstupem této bakalářské práce je algoritmus využívající shlukovou analýzu k vyhledání morfologicky odlišných srdečních stahů. Pro dosažení cíleného výsledku jsou použity dva příznaky vypočtené z různých forem signálu – základního, derivovaného a spektra – jimiž jsou Pearsonův korelační koeficient a střední kvadratická odchylka.

Metoda je náchylná na nepřesně detekovaný R kmit, což byl jeden z hlavních problémů řešených v rámci předzpracování. Konečná statistika ukazuje, že pokud by na vstupu byl záznam s precizně určenými indexy R kmitů, byly by výsledky také lepší. Jako další omezení může být vnímána neschopnost vyhodnotit, zda-li jsou srdeční cykly dostatečně podobné na zařazení do jediného shluku. Tento fakt byl vyřešen vhodně zvolenou podmínkou, viz kapitola 4.6.

Naopak značnou výhodou algoritmu je, že mu stačí dobře určené indexy R kmitů, které se dnes zcela běžně detekují s vysokou přesností při automatickém rozměrování EKG signálu, a to i v jednoduchých přístrojích. Výsledná senzitivita metody dosahuje téměř 70 %, specifická 91 %. Je na místě zdůraznit, že byly použity pouze morfologické příznaky.

Algoritmus by mohl být nápomocen v ordinacích kardiologů při hodnocení dlouhých holterovských záznamů, které mohou a bývají snímány i 24 hodin. I přes neúplně dokonalé výsledky statistického hodnocení poskytne algoritmus dobrý odrazový můstek lékaři pro komplexnější vyhodnocení EKG záznamu. Nabízí se i uplatnění v telemedicině u osob s voperovaným přístrojem na snímání EKG. Po pravidelném přenášení dat do externího zařízení, které po aplikaci algoritmu odešle výsledky (v podobě nejčastěji se vyskytujících morfologií) do ordinace lékaře.

Literatura

- [1] HABERL, Ralph. *EKG do kapsy*. Praha: Grada Publishing 2012, ISBN 978-80-247-4192-5
- [2] BLAHÚT, Peter. *EKG a Arytmologia* <https://www.techmed.sk/ekg-a-arytmologia-kniha/>
- [3] RONZHINA, Marina. *Shluková analýza*. Presentation presented at: [Umělá inteligence v medicíně AUIN; 2018 Oct 23, Brno, Czech Republic]
- [4] DUDA, Richard O., Peter E. HART, David G. STORK. *Pattern Classification*. New York, United States: John Wiley and Sons Inc., 2000 ISBN: 0-471-05669-3
- [5] THEODORIDIS Sergios, KOUTROUMBAS, Konstantinos. *Pattern Recognition*. 2nd ed. San Diego (USA): Academic Press; 2003, ISBN: 0-12-685875-6
- [6] KOZUMPLÍK, Jiří. *Analýza biologických signálů* [online prezentace]. Brno: ÚBMI, VUT, [cit. 2. 12. 2013]
- [7] HILL, Martin, MELOUN, Milan, MILITKÝ, Jiří. *Počítačová analýza vícerozměrných dat*. Praha (CZ): Academia; 2005, ISBN 80-200-1335-0
- [8] Vincent LABATUT, Hocine CHERFI. *Accuracy Measures for the Comparison of Classifiers*. Al-Dahoud Ali. The 5th International Conference on Information Technology, May 2011, amman, Jordan. Al-Zaytoonah University of Jordan, pp.1,5, 2011. <hal-00611319>

Seznam symbolů, veličin a zkratk

SA	sinoatriální
AV	atrioventrikulární
RB	raménková blokáda
KES	Komorová extrasystola
PCA	Principal component analysis – Analýza hlavních komponent
TPR	True positive rate – Senzitivita
TNR	True negative rate – Specificita
PPN	Positive prediction value – Pozitivní prediktivní hodnota
NPV	Negative prediction value – Negativní prediktivní hodnota
ACC	Accuracy – Správnost