

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA ELEKTROTECHNICKÁ

KATEDRA TEORIE OBVODŮ



Diplomová práce

Diploma Thesis

**Stanovení prediktorů pooperační fibrilace síní u pacientů
po aortokoronárním bypassu**

**Determination of Predictive Factors of Postoperative Atrial
Fibrillation after Coronary Artery Bypass Grafting**

Autor: Bc. Kristýna Vieweghová
Studijní program: Lékařská elektronika a bioinformatika
Vedoucí práce: Ing. Petr Ježdík, Ph.D.
Konzultant: doc. MUDr. Pavel Osmančík, Ph.D.

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Vieweghová** Jméno: **Kristýna** Osobní číslo: **487226**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra kybernetiky**
Studijní program: **Lékařská elektronika a bioinformatika**
Specializace: **Lékařská technika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Stanovení prediktorů pooperační fibrilace síní u pacientů po aortokoronárním bypassu

Název diplomové práce anglicky:

Determination of Predictive Factors of Postoperative Atrial Fibrillation after Coronary Artery Bypass Grafting

Pokyny pro vypracování:

1. Seznamte se s danou problematikou a vypracujte krátkou rešerši.
2. Ohodnoťte záznamy předoperačních EKG pacientů pro následnou analýzu.
3. Nalezněte možné prediktory pooperační síňové fibrilace.
4. Pokuste se stanovit skóre ohodnocující pravděpodobnost vzniku pooperační síňové fibrilace.
5. Dosažené výsledky shrňte do zprávy.

Seznam doporučené literatury:

- [1] THOREN, E., et al. Prediction of postoperative atrial fibrillation in a large coronary artery bypass grafting cohort. *Interactive CardioVascular and Thoracic Surgery*. 2012
- [2] MATHEW, Joseph P. A Multicenter Risk Index for Atrial Fibrillation After Cardiac Surgery. *JAMA*. 2004
- [3] AMAR, D., et al. Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting. *Journal of the American College of Cardiology*. 2004
- [4] ZAKKAR, M., et al. Inflammation, oxidative stress and postoperative atrial fibrillation in cardiac surgery. *Pharmacology & Therapeutics*. 2015

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Petr Ježdík, Ph.D., katedra měření, katedra teorie obvodů, LVR

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **10.01.2020** Termín odevzdání diplomové práce: **05.01.2021**

Platnost zadání diplomové práce: **30.09.2021**

Ing. Petr Ježdík, Ph.D.
podpis vedoucí(ho) práce

doc. Ing. Tomáš Svoboda, Ph.D.
podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomantka bere na vědomí, že je povinna vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studentky

Abstrakt

Tato diplomová práce se zabývá statistickým zpracováním dat u pacientů po aortokoronárním bypassu (CABG). Cílem této práce je určení prediktorů pooperační fibrilace síní s pomocí multivariantních modelů. Teoretická část shrnuje informace o vedení elektrického vzruchu svalovinou myokardu, hodnocení záznamu tohoto procesu a parametrech získaných z analýzy variability srdeční frekvence. V metodické části je čtenář obeznámen se získanými daty a statistickými metodami použitými při zpracování této práce. Výsledkem práce jsou klasifikační modely SVM a GLM, implementované v programu MATLAB, obsahující parametry CRP, SD2, Triangular Index a vzorkovou entropie. Finálně je získán práh predikce pro každý model, který je schopen rozřadit pacienty do skupin dle možného vzniku pooperační fibrilace síní.

Klíčová slova

Fibrilace, generalizovaný lineární model, SVM, MATLAB, predikce, analýza variability srdeční frekvence

Abstract

This diploma thesis focuses on the statistical analysis of patient's data after coronary artery bypass grafting (CABG). The main aim is to find predictors of the post-operative atrial fibrillation using multivariate statistical models. Theoretical part summarizes information about heart conduction system, evaluation of EKG and analysis of heart rate variability. In the methodical part the main focus is on the gained data and used statistical methods. The results consist of the classification models SVM and GLM programmed in MATLAB with parameters CRP, SD2, Triangular Index and Sample Entropy. Prediction threshold, which is able to classify patients into groups according to the possible occurrence of postoperative atrial fibrillation is obtained for each model.

Key words

Fibrillation, generalized linear model, SVM, MATLAB, prediction, Heart Rate Variability analysis

Poděkování

Tímto bych chtěla poděkovat Ing. Petrovi Ježdíkovi, Ph.D., za odborné vedení a také za jeho trpělivost a vstřícnost při konzultacích. Dále bych chtěla poděkovat doc. MUDr. Pavlovi Osmančíkovi, Ph.D. a Ing. Janě Veselé, Ph.D. za podnětné rady a poskytnutí dat pro vypracování této práce. Závěrem bych ráda poděkovala své rodině a nejbližším za podporu během studia.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne:

Kristýna Vieweghová:.....

Obsah

| | |
|---|----|
| Použité zkratky | 8 |
| Úvod | 9 |
| 1 Teoretická část | 10 |
| 1.1 Převodní systém srdeční | 10 |
| 1.2 Elektrokardiografie | 11 |
| 1.2.1 EKG křivka | 12 |
| 1.3 Fibrilace síní | 14 |
| 1.4 Analýza variability srdeční frekvence | 15 |
| 1.4.1 Časové parametry | 17 |
| 1.4.2 Frekvenční parametry | 17 |
| 1.4.3 Nelineární parametry | 18 |
| 1.5 Aortokoronární bypass | 19 |
| 2 Metodická část | 21 |
| 2.1 Klinická data | 21 |
| 2.2 Univariantní analýza | 22 |
| 2.2.1 Explorační analýza | 23 |
| 2.2.2 Testování normality dat | 23 |
| 2.2.3 Parametrické testování rozdílu skupin | 24 |
| 2.2.4 Neparametrické testování rozdílu skupin | 24 |
| 2.3 Multivariantní analýza | 25 |
| 2.3.1 Generalizovaný lineární model (GLM) | 25 |
| 2.3.2 Support Vector Machine (SVM) | 26 |
| 2.4 Křížová validace | 29 |
| 2.5 ROC křivka a AUC | 30 |
| 2.6 Korekce vícenásobného testování | 31 |
| 2.7 Zaslepená studie | 31 |
| 2.8 Práh predikce | 32 |
| 3 Výsledky | 33 |
| 3.1 Výsledky explorační analýzy a testování normality | 33 |
| 3.2 Výsledky parametrických testů | 34 |
| 3.3 Výsledky neparametrických testů | 35 |
| 3.4 Výsledky modelu GLM | 35 |
| 3.5 Výsledky modelu SVM | 38 |
| 3.6 Výsledky zaslepené studie | 39 |
| 3.7 Problematika určení prahu predikce | 43 |
| 4 Diskuze | 45 |
| 5 Závěr | 47 |
| Literatura | 48 |

| | |
|----------------------|----|
| Seznam obrázků | 52 |
| Seznam tabulek..... | 53 |
| Příloha A..... | 54 |
| Příloha B..... | 55 |

Použité zkratky

| | | |
|---------|----------------------------------|------------------------------------|
| NoFis | No Fibrillation | Bez fibrilace |
| Fis | Fibrillation | Fibrilace |
| CABG | Coronary artery bypass grafting | Aortokoronární bypass |
| HRV | Heart Rate Variability | Variabilita srdeční frekvence |
| EKG | Electrocardiogram | Elektrokardiogram |
| AD test | Anderson-Darlington test | Andersonův-Darlingtonův test |
| GLM | Generalized Linear Model | Generalizovaný lineární model |
| SVM | Support Vectore Machine | Metoda podpůrných vektorů |
| TP | True Positive | Pravdivě pozitivní |
| TN | True Negative | Pravdivě negativní |
| FN | False Negative | Falešně negativní |
| FP | False Positive | Falešně pozitivní |
| FPR | False Positive Rate | Poměr falešně pozitivních |
| TPR | True Positive Rate | Poměr pravdivě pozitivních |
| CV | Cross-validation | Křížová validace |
| ROC | Receiver Operator Characteristic | Operační charakteristika přijímače |
| AUC | Area Under Curve | Plocha pod křivkou |

Úvod

Aortokoronární bypass je operací prováděnou u pacientů s diagnostikovanou ischemickou chorobou srdeční či v akutní formě po infarktu myokardu. Ischemická choroba srdeční je jednou z hlavních příčin úmrtí ve vyspělých zemích. Tato skutečnost má v posledních letech klesající tendence, které souvisí s vývojem nových postupů a zlepšováním dostupnosti a kvality zdravotnické péče

Nejčastější pooperační komplikací po provedení aortokoronárního bypassu je fibrilace síní. Prevalence této komplikace se pohybuje kolem 20-40 %. Její přítomnost prodlužuje dobu pobytu pacienta v nemocnici, roste pravděpodobnost dalších zdravotních komplikací a celkově navyšuje náklady na péči o pacienta. Cílem této práce je nalézt prediktory této komplikace s větším důrazem na multivariantní modely. Dále je zkoumán přínos parametrů pořízených při standardním EKG vyšetření před provedením operace. Získání prediktorů z těchto parametrů by bylo výhodné vzhledem k tomu, že vyšetření elektrokardiogramem je součástí běžného hodnocení stavu pacienta při přijetí do nemocnice. Při zjišťování hodnoty parametrů by tak nedocházelo ke zvýšené zátěži nemocničního personálu či navýšení souvisejících nákladů. Konečným požadavkem na tuto práci je nalezení vhodného prahu predikce modelů pro rozřazení nových pacientů do skupiny Fis a NoFis.

V úvodní kapitole práce je rozebrána problematika fyziologického vedení akčního potenciálu srdcem a jeho snímání s pomocí standardního 12svodového EKG. Tímto způsobem poskytnuté informace, jakožto nové parametry v této práci, jsou rozebrány v další podkapitole. Dále jsou zde rozebrány parametry týkající se variability srdeční frekvence a její analýzy. Nechybí ani popis samotné fibrilace síně a operace typu aortokoronárního bypassu.

Další část práce se zabývá získanými daty a statistickými metodami použitými v praktické části této práce. Je zde popsána univariantní analýza, multivariantní klasifikační modely GLM a SVM, způsob ohodnocení jejich výkonnosti a postup určení prahu predikce.

Na závěr je rozebrána problematika práce a její porovnání s dostupnou literaturou. Jsou zde zmíněny i limitace práce a možné návrhy jejich řešení. Speciální část je věnována porovnání výsledků této práce s dizertační prací Ing. Jany Veselé, PhD., která byla z části provedena na stejném vzorku dat. Vedlejším přínosem této práce má být verifikace těchto výsledků. Vzhledem k novým parametrům v této práci a vybráním zúženého vzorku pacientů bude toto ověření jen orientační.

1 Teoretická část

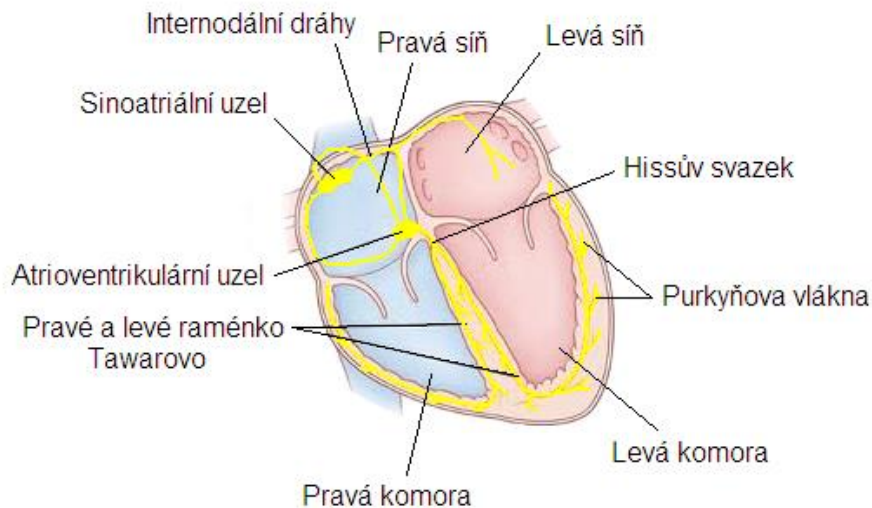
Pro zpracování této diplomové práce bylo nutné porozumět problematice vedení elektrického signálu po srdeční svalovině a obeznámit se s klinickými parametry, které popisují tento děj a jeho poruchami s důrazem na fibrilaci síní. Konec teoretické části je věnován obeznámení s tématem aortokoronárního bypassu.

1.1 Převodní systém srdeční

Tato kapitola popisuje převodní systém srdeční a funkce jeho jednotlivých částí. Srdeční svalovina se skládá z buněk kardiomyocytů, které se dle své funkce dělí na dva typy. Jedním je tzv. pracovní myokard, který má schopnost kontrakce. Druhým jsou buňky převodního systému srdečního často označované jako pacemakerové buňky. Tato specializovaná tkáň má schopnost bez vnějšího stimulu generovat vzruchy a rozvádět je k buňkám pracovního myokardu. Pracovní myokard tuto autonomní aktivitu přebírá jen za patologických stavů. [16]

Převodní systém srdeční se skládá ze sinoatriálního (primárního) uzlíku, internodálních spojů, atrioventrikulárního (sekundárního) uzlíku, Hissova svazku, Tawarových ramének a Purkyňových vláken. Jednotlivé části převodního systému srdečního je možné vidět na obrázku číslo 1.[17]

Sinoatriální uzel se nachází ve stěně pravé síně poblíž ústí horní duté žíly a pravého ouška. Říká se mu „přirozený pacemaker“, právě zde dochází ke vzniku spontánní elektrické aktivity a udává tak srdci sinusový rytmus (70-80 tepů za minutu). Vzniklé elektrické podráždění se přenáší s pomocí internodálních drah na pracovní myokard síní a na atrioventrikulární uzlík, který leží v bázi vnitřního cípu trojcípé chlopně. Někdy se mu říká také „sekundární pacemaker“, protože při poruše funkce sinoatriálního uzlu přebírá tvorbu vzruchů, které je ale schopen generovat pouze s frekvencí 40-60 tepů za minutu a vzniká tak zpomalení akce srdeční (bradykardie). Vzruch se dále šíří cestou Hissova svazku v mezikomorové přepážce, ten se dělí na Tawarova raménka (pravé a levé) a ta se rozpadají v Purkyňova vlákna, která jsou konečnou částí převodního systému. [16][17]



Obrázek 1: Převodní systém srdeční. Převzato z [18]

1.2 Elektrokardiografie

Elektrokardiografie je základní kardiologické vyšetření nutné pro správnou diagnostiku kardiovaskulárních chorob. Časová závislost rozdílů elektrických potenciálů je snímána z povrchu těla pacienta s pomocí elektrod. Základním způsobem rozmístění elektrod je tzv. standardní 12svodový systém, který se skládá ze čtyř končetinových a šesti hrudních elektrod. Končetinové elektrody jsou rozmístěny na všech 4 končetinách (R-pravá ruka, L-levá ruka, F-levá noha, N-pravá noha, referenční elektroda) a dohromady tvoří Einthovenův trojúhelník, ten je vyobrazen na obrázku číslo 2. Bipolární končetinové svody je možné definovat (rovnice 1 až 3) jako [25]

$$U_I = V_{LA} - V_{RA} \quad (1)$$

$$U_{II} = V_{LL} - V_{RA} \quad (2)$$

$$U_{III} = V_{LL} - V_{LA} \quad (3)$$

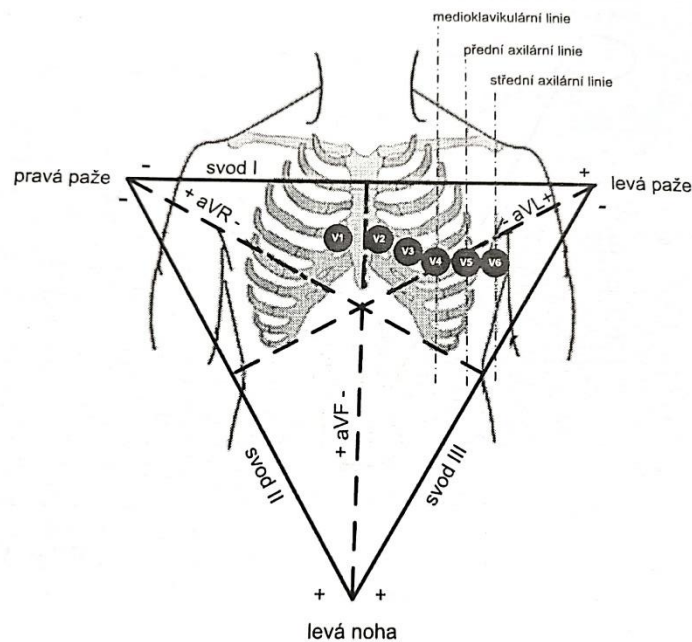
kde V_{RA} je snímáný potenciál pravé ruky, V_{LL} je snímáný potenciál levé nohy a V_{LA} snímáný potenciál levé ruky. Dále 12svodový systém doplňují Goldbergovy augmentované bipolární svody. Výpočet (viz rovnice 4 až 6) je proveden ze vztahu potenciálů ve vrcholech Einthovenova trojúhelníku k průměrnému potenciálu protilehlé strany trojúhelníku. [25]

$$U_{aVR} = V_{RA} - 0,5(V_{LA} + V_{LL}) \quad (4)$$

$$U_{aVL} = V_{LA} - 0,5(V_{RA} + V_{LL}) \quad (5)$$

$$U_{aVF} = V_{LL} - 0,5(V_{LA} + V_{RA}) \quad (6)$$

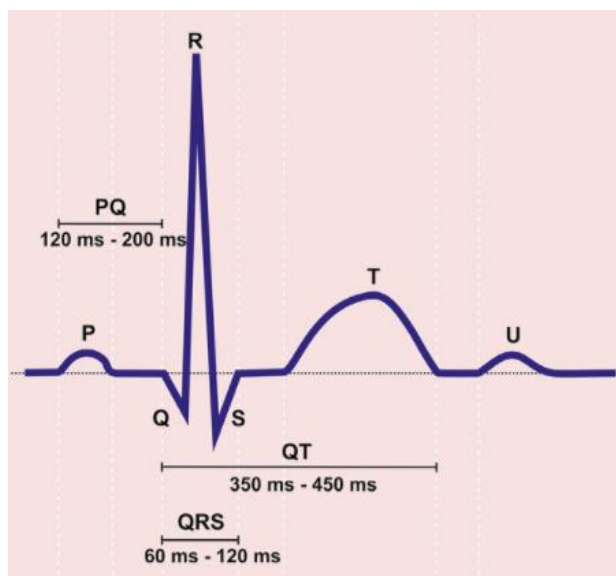
Poslední součástí systému jsou Wilsonovy unipolární hrudní svody V_1 - V_6 , které udávají napětí mezi Wilsonovou svorkou (aritmetický průměr tří končetinových potenciálů) a elektrodami položenými na hrudníku. Jejich rozmístění je vidět na obrázku číslo 2. [25]



Obrázek 2: Einthovenův trojúhelník a poloha elektrod hrudních unipolárních svodů. Převzato z [25].

1.2.1 EKG křivka

Záznam pořízený elektrokardiografem během srdečního cyklu se nazývá elektrokardiogram. Skládá se z jednotlivých PQRST(U) křivek neboli EKG křivek. Fyziologická EKG křivka je vyznačena na obrázku číslo 3. Ve většině případů je záznam pořízen s posunem papíru rychlostí $25 \text{ mm} \cdot \text{s}^{-1}$, což značí, že jeden čtvereček na papíře o straně 1 mm reprezentuje 40ms. V následné části budou probrány významné parametry, které se dají z EKG křivky vyčíst a byly v této práci použity. [1][19]



Obrázek 3: EKG křivka. Převzato z [1].

- **Tepová frekvence:** U zdravého jedince srdce pumpuje krev s frekvencí 60-90 tepů/min, u trénovaného jedince až kolem 50 tepů/min. S pomocí EKG pravitka se počítá ze tří po sobě jdoucích RR intervalů¹.
- **PQ interval:** Jedná se o období depolarizace síní, převod vzruchu ze síní ke komorám. Měří se od začátku P vlny až k začátku QRS komplexu. Standardně trvá 120-200 ms.
- **QRS komplex:** Během QRS komplexu probíhá depolarizace komor. Standardně trvá 60-120 ms. Široký QRS komplex může značit blokádu Tawarova raménka.
- **QT interval:** Popisuje depolarizace a repolarizace komor a na EKG trvá od začátku QRS komplexu až do začátku vlny T. Tepová frekvence významně ovlivňuje dobu trvání QT intervalu, se zvyšující se frekvencí se zkracuje. Standardně trvá 350-450 ms.
- **QTc:** Prodloužený normalizovaný QT interval se zavádí z důvodu vlivu tepové frekvence na QT interval. Upravenou hodnotu vypočteme dle Bazettovy formule (vzorce 7). Hodnota fyziologické doby trvání je udávána mezi 340-440 ms a ovlivňuje ji stáří a pohlaví pacienta.

$$QTc = \frac{QT}{\sqrt{RR}} \quad (7)$$

- **P vlna:** P vlna odpovídá depolarizaci síní a trvá přibližně 80 ms.

¹ RR interval je vzdálenost mezi dvěma po sobě jdoucími kmity R.

- **Elektrická osa srdeční QRS:** Jedná se o vektor vyznačující směr postupu vzruchu přes srdeční komory během stahu srdeční svaloviny (systoly). U zdravého jedince leží v rozmezí -30° až $+105^{\circ}$.
- **Elektrická osy srdeční P:** Vektor popisující postup depolarizace síní. Běžně se jako klinický parametr nepoužívá, ale vzhledem k tomu, že fibrilace síní se týkají právě vlny P, byl tento parametr do práce zahrnut. Fyziologické rozmezí 0° až 75° .
- **Přítomnost blokády Tawarova raménka:** Může se jednat o blokádu levého Tawarova raménka (LBBB – Left Bundle Branch Block) či pravého (RBBB – Right Bundle Branch Block). Obě se projevují QRS komplexem delším než 120 ms. Dochází také k výchytkám srdečních os. Levá se na EKG projevuje především ve svodu V6, kde QRS vypadá jako písmeno M a následuje negativní vlna T². Dále je často změna ve svodu V1, QRS je zde negativní ve tvaru písmene V. Blokáda pravého raménka se hledá především ve svodu V1, kde mívá QRS dva vrcholy vedle sebe připomínající písmeno M a objevuje se zde také negativní vlna T. [1][15]

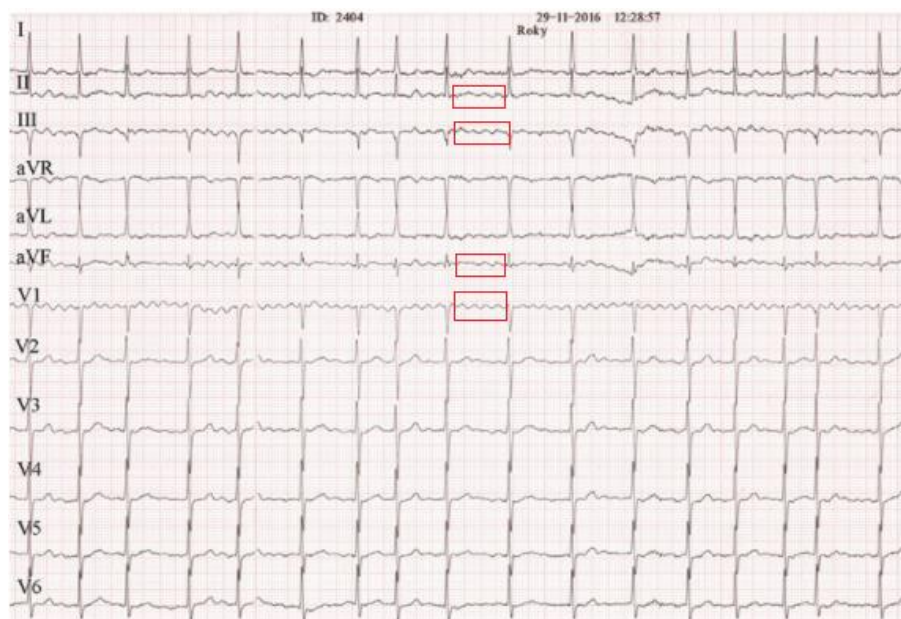
1.3 Fibrilace síní

Srdeční arytmie se vyznačují poruchami vedení nebo tvorbou elektrického vzruchu, jde tedy o narušení funkce převodního systému srdečního. Dle rychlosti rytmu je můžeme dělit na bradyarytmie ($<60/\text{min}$) a tachyarytmie ($>90/\text{min}$). Fibrilace síní se řadí mezi supraventrikulární tachyarytmii, což znamená, že vzniká nad komorami a její tepová frekvence přesahuje $100/\text{min}$. Fibrilace síní je popisována neorganizovanou elektrickou aktivitou v myokardu síní. [1][7]

Jako klinický projev se u fibrilací síní objevuje palpitace (bušení srdce vnímané pacientem), námahová dušnost, únava, závratě, fyzická nevykonnost, bolest na hrudi, poruchy spánku a výjimečně i presynkopy či synkopy (krátkodobá ztráta vědomí). Objevují se i případy, kdy je Fis (fibrilace síní) zcela asymptomatická. [2]

Při pohledu na záznam standardního 12svodového EKG je viditelná nepřítomnost P vln, místo kterých se objevují nesourodé fibrilační vlnky. Tento jev můžeme pozorovat na obrázku 4 například na červeně vyznačených místech. Přítomnost P vln se nejlépe identifikuje ze svodů II, III, aVF nebo V1. Dále z elektrokardiogramu rozpoznáme rozdílné RR intervaly mezi QRS komplexy dávající informaci o nepravidelné tepové frekvenci. [1]

² Fyziologická vlna T značí repolarizaci komor



Obrázek 4: Fibrilace síní. Chaotická vlnitá aktivita nahrazující P vlny ve svodech II, III, aVF a V1 zvýrazněna červenými obdélníky. Převzato z [1].

1.4 Analýza variability srdeční frekvence

Variabilita srdeční frekvence (HRV, Heart Rate Variability) je přirozená fluktuace mezi časovými intervaly jednotlivých srdečních akcí. HRV reflektuje regulační systémy v našem těle, jako například změny krevního tlaku, výměnu plynů apod. a pomáhá nám přizpůsobit se různým vnějším vlivům či psychologickým změnám. Větší srdeční variabilita není ale vždy dobrá a může naznačovat patologické změny a zdravotní problémy jako je například fibrilace síní. [20]

HRV můžeme analyzovat v časové, frekvenční či nelineární oblasti. Přehled parametrů, získaných programem Kubios HRV, užívaných v této práci je popsán v tabulce 1. V tabulce je uvedena příslušnost parametru do jednotlivých oblastí, jednotka, ve které je parametr zaznamenáván a stručný popis parametru. Uváděné postupy získání určitých parametrů se vztahují k softwaru Kubios HRV.

Tabulka 1: Parametry analýzy variability srdeční frekvence v časové, frekvenční a nelineární oblasti. Převzato a upraveno z [21][24].

| Časová oblast | | |
|---------------|----------|---|
| Parametr | Jednotky | Popis |
| Mean RR | [ms] | Průměrná hodnota RR intervalů |
| Mean HR | [1/min] | Průměrná hodnota tepové frekvence |
| RMSSD | [ms] | Druhá mocnina součtu čtverců rozdílu RR intervalů |

| | | |
|--------------|---------|---|
| NN50 | [počet] | Počet po sobě jdoucích RR intervalů, které se od sebe liší o více než 50 ms |
| pNN50 | [%] | Procentuální zastoupení NN50 v celkovém počtu RR intervalů |
| RR Tri Index | - | Integrál histogramu RR intervalů dělený výškou histogramu |
| TINN | [ms] | Trojúhelníková interpolace normálních RR intervalů |

Frekvenční oblast

| Parametr | Jednotky | Popis |
|------------------------------|--------------------|---|
| VLF, LF a HF maxima | [Hz] | Frekvence s maximální amplitudou v pásmu velmi nízkých (VLF, <0,04 Hz), nízkých (LF, 0,04-0,15 Hz) a vysokých (HF, 0,15-0,4 Hz) frekvencí |
| VLF, LF a HF absolutní výkon | [ms ²] | Absolutní výkon v pásmu velmi nízkých, nízkých a vysokých frekvencí |
| VLF, LF a HF relativní výkon | [%] | Relativní výkon v pásmu velmi nízkých, nízkých a vysokých frekvencí |
| LF a HF normalizovaný výkon | [n.u.] | Výkon v pásmu nízkých a vysokých frekvencí v normalizovaných jednotkách |
| Celkový výkon TP | [ms ²] | Celkový spektrální výkon |
| LF/HF | - | Poměr mezi výkonem v pásmu nízkých a vysokých frekvencí |
| EDR | [Hz] | Frekvence dýchání odvozená z EKG |

Nelineární oblast

| Parametr | Jednotky | Popis |
|---------------------------------|----------|--|
| SD1, SD2 | [ms] | Směrodatné odchylky získané z Poincarého mapy |
| ApEn | - | Přibližná entropie |
| SampEn | - | Vzorková entropie |
| D ₂ | - | Korelační rozměr |
| α ₁ , α ₂ | - | Krátkodobá a dlouhodobá fluktuace fluktuační analýzy přizpůsobující se trendu (DFA - Detrended Fluctuation Analysis) |
| Lmean | [beats] | Průměrná délka diagonální přímky v RP (Recurrence plot) |
| Lmax | [beats] | Maximální délka diagonální přímky v RP |
| REC | [%] | Recurrence rate (procento rekurentních bodů v RP) |
| DET | [%] | Determinismus (procento rekurenčních bodů, které formují diagonální přímku v RP) |
| ShanEn | - | Shannonova entropie pravděpodobnostního rozdělení délek diagonálních přímek |
| MSE | - | Víceúrovňová entropie (Multiscale Entropy) |
| Portův index (PI) | - | Vyhodnocení asymetrie bodů ve dvou regionech Poincareho mapy oddělených přímkou identity |

| | | |
|--------------------|---|--|
| Guzikův index (GI) | - | Vyhodnocení asymetrie na základě kumulativní vzdálenosti každého bodu od přímky identity |
| Slope Index (SI) | - | Kvalifikace asymetrie fázových úhlů |

1.4.1 Časové parametry

Nejjednodušším přístupem je výpočet parametrů v časové oblasti. Je zde zkoumána variabilita v intervalu mezi jednotlivými QRS komplexy, tedy v tzv. RR intervalech. Do výpočtu jsou začleněny pouze intervaly s normálním vedením vzruchu bez výskytu patologických změn. K parametrům časové oblasti, které jsou s jednotkami vypsány v tabulce 1, řadíme střední hodnotu RR intervalů, střední hodnotu tepové frekvence, druhou mocninu součtu čtverců rozdílu RR intervalů (RMSSD), počet po sobě jdoucích RR intervalů, s rozdílem více než 50 ms (NN50) a procentuální zastoupení NN50 mezi všemi RR intervaly (pNN50). [21]

Pod časovou oblast se řadí i parametry geometrické, které se získávají z histogramu RR intervalů. Mezi tyto parametry patří Triangular index (RR Tri Index), který je charakterizován integrálem histogramu RR, jehož výškou je dělený. Výška histogramu je závislá na zvolení šířce jednotlivých binů histogramu, doporučována je šířka 1/128 s. Další geometrickou metrikou je TINN, což je parametr, který popisuje šířku základny histogramu ohodnocenou přes trojúhelníkovou interpolaci. [24]

1.4.2 Frekvenční parametry

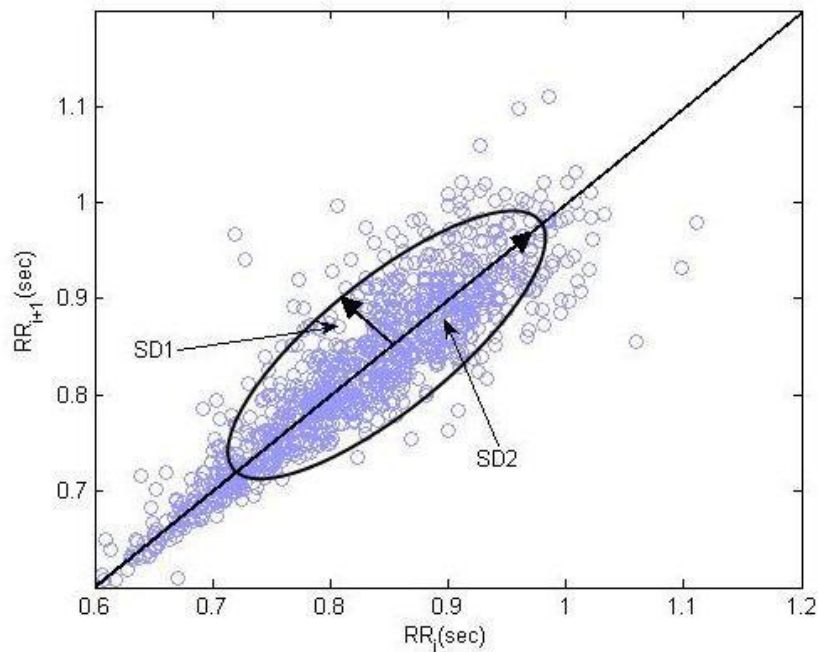
Hlavní podstatou frekvenční analýzy je, že oproti časové oblasti podává informaci o inervaci vegetativními nervy – sympatikem a parasympatikem. Mezi těmito nervy dochází k vzájemné regulaci. V případě vyšší aktivity sympatiku dochází ke zrychlení a zesílení činnosti srdce, což implikuje zeslabení činnosti parasympatiku. V situaci zvýšené aktivace parasympatiku srdeční akce zpomaluje a oslabuje. Předpokladem je, že aktivita sympatiku je propojena s nízkými frekvencemi 0,04-0,15 Hz a parasympatikus si spojuje především s vyššími frekvencemi v oblasti 0,15- 0,4 Hz. Nerovnováha mezi sympatikem a parasympatikem může přispět ke vzniku fibrilace síní. [1][21]

Pro získání frekvenčních parametrů je nutno vypočítat ze série RR intervalů frekvenční spektrum. Prvním krokem je převzorkování RR řady pomocí interpolace kubickou splinou (typ nelineární regrese), teprve poté je možné získat frekvenční spektrum. K výpočtu program Kubios HRV užívá dva způsoby – metodu rychlé Fourierovy transformace (FFT – Fast Fourier Transform) a metodu parametrických autoregresivních modelů (AR). [24]

1.4.3 Nelineární parametry

Vzhledem k složitosti celého řídicího systému srdce je vhodné pro popis použít i nelineární metody, aby bylo možné tento komplexní systém dostatečně charakterizovat. Parametry nelineární oblasti umožňují popisovat nepředvídatelnost a složitost časových řad.

Základem pro získání nelineárních parametrů je Poincareho mapa. Jedná se o přístup převzatý z nelineární dynamiky. Poincareho mapa je grafickým zobrazením korelace mezi po sobě jdoucími RR intervaly. RR_{i+1} je funkcí RR_i , což značí předpoklad vlivu probíhajícího RR intervalu na ten následující. Tvar mapy tvoří elipsa, která je orientovaná podle dat a LOI (line od identity). Šířka a délka elipsy je popisována standardními odchylkami $SD1$ a $SD2$. Poincareho mapa je vidět na obrázku číslo 5. [24][34]



Obrázek 5: Poincareho mapa. Převzato z [34].

$$SD1^2 = \frac{1}{2}SDSD^2 \quad (8)$$

$SD1$ je definována vzorcem 8, kde $SDSD$ značí směrodatnou odchylku rozdílů po sobě přicházejících RR intervalů. $SD1$ je standardní odchylkou bodů kolmých na LOI a popisuje krátkodobou variabilitu, oproti tomu $SD2$ měří dlouhodobou variabilitu a je definována jako:

$$SD2^2 = 2SDNN^2 - \frac{1}{2}SDSD^2 \quad (9)$$

kde SDNN je směrodatná odchylka RR intervalů popsána následující rovnicí 10, kde N je celkový počet po sobě následujících RR intervalů, mean RR je průměrná hodnota RR intervalů a RR_i je i-tý RR interval.[24]

$$SDNN = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (RR_i - \text{mean } RR)^2} \quad (10)$$

Z Poincareho mapy pochází i parametry Guzikův index (GI), Portův index (PI) a Slope index (SI), které jsou obsažené v práci.

Dalšími nelineárním parametrem je entropie, která je popisována jako míra neuspořádanosti systému. V této práci jsou zmiňovány čtyři druhy entropií – přibližná entropie, vzorková entropie, víceúrovňová entropie a Shannonova entropie.

První tři zmiňované druhy se týkají chaotického chování RR intervalů. Přibližná entropie (ApEn, Approximate Entropy) vyčísluje pravidelnost a složitost časových řad. Vyšší hodnoty poukazují na trend stoupající variability RR intervalů. Vzorková entropie (SampEn, Sample Entropy) je určitým vylepšením přibližné entropie, jelikož poskytuje lepší vyhodnocení časových řad. Nižší hodnoty jsou odrazem větší podobnosti časových řad. Dalším druhem je víceúrovňová entropie (MSE, Multiscale Entropy), kde je signál během analýzy rozdělen na bloky o určité velikosti. V každé úrovni signálu určené tímto blokem je hodnota entropie vypočtena zvlášť, z důvodu zkoumání změn patologií variability, která se nemusí odrážet stejně ve všech korelacích. Pro analýzu je rozhodující doba naměřeného signálu. [36]

Posledním druh entropie obsažený v této práci je Shannonova entropie (ShanEn, Shannon Entropy). Tento typ entropie ohodnocuje míru neuspořádanosti distribuce v tzv. rekurentním grafu, která je odvozen z recurrence plot analýzy. Další parametry ohodnocující tento graf jsou Lmean, Lmax, REC a DET jejichž charakteristiku najdeme v tabulce číslo 1. [21][24][35]

1.5 Aortokoronární bypass

Aortokoronární bypass (CABG, Coronary Artery Bypass Graft) je velmi častou operací prováděnou u pacientů trpících ischemickou chorobou srdeční, která se v tabulkách pohybuje mezi předními příčinami úmrtí na světě. Jedná se o pacienty se srdečními tepnami zúženými postupným ukládáním tuku do jejich stěn (procesem aterosklerózy), u kterých nemají dostatečný efekt léky. Variantou akutního neplánovaného zákroku je případ, kdy zhoršením ischemické choroby srdeční dojde k infarktu myokardu. Principem operace je přivést krev do oblastí, které trpí jejím nedostatečným zásobením v důsledku zúžení či úplného uzavření. Aortokoronární céva je přemostěna s pomocí žilního či tepenného štěpu. [15][26]

Nejčastější komplikací nastupující po této operaci je fibrilace síní. Výskyt této komplikace se dle studií pohybuje kolem 20-40 %, do tohoto čísla jsou ale i zahrnuty operace chlopní na srdci. Pooperační fibrilace síní je charakteristická krátkými, paraxysmálními (záchvatovými) stavy s největším výskytem kolem 2-4 dnu po výkonu a navrácením se během prvního týdne po operaci. Návrat k sinusovému rytmu je často spontánní bez ohledu na terapeutické strategie. Tato komplikace je spojena s prodloužením pobytu v nemocnici v řádu průměrně 5 dní a s rostoucími náklady okolo hospitalizace společně s dodatečnými vyšetřeními, léčbou a prodlouženou dobou strávenou v nemocnici. Pacienti s pooperační fibrilací síní mají zvýšené riziko pozdějších zdravotních komplikací, jako jsou například cévní mozkové příhody či tvorba jiných arytmií.[6] [29]

2 Metodická část

Kapitola metodická část je rozdělena na několik podkapitol. První část je zaměřena na klinická data a jejich popis. V další části jsou popsány použité statistické metody a modely k nalezení vhodných prediktorů nebo kombinace prediktorů, díky kterým by bylo možné predikovat, zda se u pacienta vyskytne pooperační fibrilace síní. V podkapitolách jsou zmíněny i určité postupy v programu MATLAB, použité pro sestavení těchto predikčních modelů.

2.1 Klinická data

Data pro diplomovou práci byla poskytnuta Fakultní nemocnicí Královské Vinohrady. Základní soubor dat byl pořízen v rámci retrospektivní studie APOFIS, do které byli začleněni pacienti podstupující kardiologický výkon v období března 2015 až září 2016. Podmínkou pro vstup do studie byla předchozí absence fibrilace síní. Další kritérium pro účast byla nepřítomnost implantovaného kardiostimulátoru či defibrilátoru, neurgentní provedení operace a jiný než sinusový rytmus při přijetí. Na pacientech zahrnutých ve studii byl proveden aortokoronární bypass z důvodu ischemické choroby srdeční, korekce vady srdečních chlopní či kombinace těchto výkonů. Operace se uskutečnila na Kardiologické klinice 3.lékařské fakulty Univerzity Karlovy a Fakultní nemocnice Královské Vinohrady. Pro záměry práce byli vybráni pouze pacienti s aortokoronárním bypassesem. [21]

Celkově studie APOFIS obsahuje 106 parametrů týkající se základní charakteristiky pacienta jako například pohlaví, věk či hmotnost, popis chlopenních vad, léky podané na příjmu, hodnoty získané laboratorně, typ operačního výkonu a další. Velká část parametrů byla získána z HRV analýzy skládající se z parametrů časové, frekvenční a nelineární oblasti. Pro výpočet parametrů byl u každého pacienta naměřen přibližně dvouhodinový záznam EKG systémem VLV vzniklým na Fakultě biomedicínského inženýrství ČVUT. Zpracovaný signál byl vyhodnocen s pomocí programu Kubios HRV verze 2.2. Z HRV analýzy bylo celkově vypočteno 70 parametrů. [21]

Nejdůležitějším parametrem této práce je kategoriální parametr Skupiny, který poskytuje informaci o přítomnosti vzniku pooperační fibrilaci síní. Pacienti se tedy dělí do dvou skupin Fis (s fibrilací) a Nofis (bez fibrilace).

Pro praktickou část byly nemocníci poskytnuty elektrokardiogramy účastníků studie naměřené při příjmu. Z těchto záznamů elektrické aktivity srdce bylo ručně tzn. pomocí EKG pravítka a automaticky (myšleno vyčtením z dodaných dokumentů ověřených pomocí EKG pravítka) rozměřeno celkově 9 nových parametrů obsahující základní charakteristiku EKG křivky. Jde o parametry PQ, QRS, QT, QTc, P, Osa QRS, osa P, frekvence a přítomnost blokády Tawarova raménka. Základní informace, k těmto devíti novým parametrům, v podobě středních hodnot, mediánů a směrodatných odchylek jsou zpracované v tabulce číslo 2. Z důvodu chybějícího či nekvalitního záznamu u některých účastníků byl celkový vzorek pacientů, přidáním těchto parametrů, zúžen na číslo 82. Tento vzorek obsahuje 67 mužů a 15 žen pohybujících se ve věkovém rozmezí 39 až 85 let. U 31 pacientů se po operaci objevila komplikace v podobě fibrilace síní.

Tabulka 2: Přehled parametrů získaných ze záznamu EKG a jejich středních hodnot (mean) se směrodatnou odchylkou (std) a mediánem. Hodnota 0 u blokády Tawarova raménka značí nepřítomnost blokády.

| Počet pacientů | NoFis | | Fis | |
|-----------------------------|------------|--------|------------|--------|
| | 51 | | 31 | |
| | Mean ± Std | Median | Mean ± Std | Median |
| PQ [ms] | 169 ± 26 | 164 | 180 ± 29 | 172 |
| QRS [ms] | 102 ± 15 | 100 | 99 ± 18 | 94 |
| QT [ms] | 411 ± 52 | 410 | 414 ± 32 | 418 |
| QTc [ms] | 424 ± 49 | 423 | 431 ± 27 | 431 |
| PQ [ms] | 124 ± 43 | 118 | 123 ± 17 | 120 |
| OsaQRS [°] | 7 ± 35 | 5 | 16 ± 36 | 15 |
| OsaP [°] | 33 ± 34 | 45 | 42 ± 21 | 47 |
| Frekvence [tep/min] | 64 ± 10 | 63 | 66 ± 11 | 62 |
| Blokáda Tawarova raménka | 0 | 0 | 0 | 0 |

Finálně byly modely testovány na zaslepených datech, viz kapitola 2.7. Tento vzorek dat obsahuje celkově 61 pacientů, s rozložením pohlaví na 47 mužů a 14 žen, z toho 17 osob s pooperační fibrilací síní.

2.2 Univariantní analýza

Na získaných datech byla provedena základní explorační analýza, při které byla zkoumána korelace jednotlivých parametrů, hledání odlehlých hodnot a testována normalita dat. Dále byly použity parametrické a neparametrické testy pro porovnání rozdílů skupin.

2.2.1 Explorační analýza

Data byla prozkoumána s pomocí krabicových diagramů (boxplot) vyobrazujících podstatné popisné informace jako medián, první a třetí kvartil, maximální a minimální hodnotu a odlehlá data (outlinery).

Dále byla vypočtena korelační matice obsahující Pearsonův korelační koeficient R pro všechny možné kombinace dat a posouzena jejich konekce. K výpočtu byla použita funkce *corrcoef*.

Normalita dat byla zkoumána na histogramech a byla navržena hypotéza, že určité parametry nepocházejí z normálního rozdělení. Hypotéza byla testovaná Andersonovým-Darlingtonovým testem, viz následující kapitola.

2.2.2 Testování normality dat

Normalita byla otestována pomocí Andersonova-Darlingova testu (AD test). Je testována nulová hypotéza, že data pocházejí z normálního rozdělení. Test měří vzdálenost mezi empirickou distribuční funkcí dat $F(x)$ a distribuční funkcí teoretického normálního rozdělení $F_T(x)$.

$$D = n \int_{-\infty}^{\infty} (F_T(x) - F(x))^2 w(x) dF(x) \quad (11)$$

kde n je velikost vzorku, hodnoty $\{x_1 < \dots < x_n\}$ jsou seřazené hodnoty vzorku a $w(x)$ je váha funkce, která je stanovena podle následující rovnice číslo 12.

$$w(x) = [F(x)(1 - F(x))]^{-1} \quad (12)$$

AD test dává větší váhu na okrajové části rozdělení, je tedy více citlivý na odlehlé hodnoty a má větší sílu v detekování odklonu od normálního rozdělení v těchto okrajových částech rozdělení. [9]

Statistika testu je definována vztahem 13:

$$AD = -n - \sum_{i=1}^n \frac{2i-1}{n} [\ln(F(x_i)) + \ln(1 - F(x_{n+1-i}))] \quad (13)$$

Pro ověření normality dat byla využita funkce *adtest* z knihovny Statistics and Machine Learning Toolbox programu MATLAB.

2.2.3 Parametrické testování rozdílů skupin

Data, která pochází z normálního rozdělení, byla dále testována nezávislým Studentovým T testem. Předpokladem pro Studentův T test jsou stejné rozptyly skupin a normální rozdělení. Jsou testovány střední hodnoty rozdělení, $H_0: \mu_1 = \mu_2$. K výsledku testování byla vypočtena i velikost účinku (Cohen's effect size) d . Velikost účinku větší jak 0,8 značí velký rozdíl skupin. Hodnota je vypočtena s pomocí středních hodnot μ_1 a μ_2 a sdružené směrodatné odchylky σ_p :

$$d = \frac{\mu_1 - \mu_2}{\sigma_p} \quad (14)$$

kde sdruženou směrodatnou odchylku σ_p pro rozdílně velké skupiny vypočteme podle vzorce 15, kde N_1 a N_2 označují velikost populací. [9]

$$\sigma_p = \sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2}} \quad (15)$$

K testování se užívala funkce `ttest2` z knihovny Statistics and Machine Learning Toolbox programu MATLAB.

2.2.4 Neparametrické testování rozdílů skupin

Na parametrech, u kterých byla přijata alternativní hypotéza, která nám říká, že data nepocházejí z normálního rozdělení byl dále proveden neparametrický test pro porovnání skupin. Pro spojitá data byl zvolen Wilcoxon-Mann-Whitney test k ověření nezávislosti populací skupiny Fis a NoFis. Tomuto testu se také říká Wilcoxon Sum Rank Test, každé hodnotě je přiřazeno určité pořadí s hodnotou (rank) a porovnávají se mediány populací. Pro každý parametr byla otestována hypotéza H_1 : Mezi pacienty skupiny Fis a NoFis je statisticky signifikantní rozdíl. [40]

K výsledkům testu Wilcoxon-Mann-Whitney byla vypočtena velikost účinku (effect size) r , jejíž výpočet udává rovnice 16. [39]

$$r = \frac{z}{\sqrt{N}} \quad (16)$$

kde N je počet pacientů a z značí z -skóre³. Hodnota r 0,1-0,3 naznačuje malý efekt, 0,3-0,5 střední a $>0,5$ velký efekt.

Na otestování zbylých kategoričkových dat byly použity kontingenční tabulky a Pearsonův chí-kvadrát test.

³ Pro získání z -scóre se provádí lineární z -transformace, která převádí znaky do standardizované škály. z -skóre = (původní hodnota – střední hodnota) / směř. odchylka hodnot[22]

K uvedeným testům byly využity funkce *ranksum* a *crosstab* z knihovny Statistics and Machine Learning Toolbox.

2.3 Multivariantní analýza

Pro multivariantní analýzu byly zvoleny dva přístupy. Prvním z nich byl GLM (Generalizovaný lineární model) a druhým SVM (Support Vector Machine). Oba modely jsou v závěrečné části práce porovnány.

2.3.1 Generalizovaný lineární model (GLM)

GLM je metoda, která umožňuje větší svobodu při specifikaci regresního modelu než klasická lineární regrese. Skládá se ze tří generalizovaných částí – lineárního prediktoru, link funkce a náhodné části, která je specifikovaná typem rozdělení. [12]

Jedním z hlavních rozdílů oproti klasickému lineárnímu modelu je, že zkoumaná závislá proměnná pochází z exponenciální rodiny. Mezi exponenciální rodinu patří nejen Gaussovo, ale i například exponenciální, inverzní Gaussovo, Poissonovo nebo binomické rozdělení. Porovnání zápisu distribucí modelu lineární regrese a GLM modelu je možno vidět níže,

$$\text{Lineární model: } Y_i \sim N(\mu, \sigma^2) \quad (17)$$

$$\text{GLM: } Y_i \sim \text{exponenciální rodina} \quad (18)$$

kde Y je závislá proměnná, N je normální rozdělení, μ je střední hodnota a σ^2 značí rozptyl. [12]

Model GLM má lineární prediktor s $k \geq 1$ nezávislých vysvětlujících proměnných, což je stejné jako u lineárního modelu, ale levá strana rovnice je jiná. Lineární prediktor totiž přímo nemodeluje střední hodnotu závislé proměnné. Nachází se zde link funkce, která transformuje střední hodnotu závislé proměnné do lineární predikční škály.

$$f(\mu) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (19)$$

Kde $f(\mu)$ je link funkce, x jsou vysvětlující nezávislé proměnné, α je intercept a β neznámé parametry.

Link funkce pro binomické rozdělení, která byla použita v práci:

$$\log \frac{\mu}{1-\mu} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (20)$$

GLM je ve skutečnosti speciální třídou nelineární regrese. Vzhledem k nelinearitě link funkce či daným rozdělením je většina GLM modelů ze své podstaty nelineární.[13]

V rámci práce byly v programu MATLAB využity funkce *stepwiseglm*, *fitglm* a *predict* z knihovny Statistics and Machine Learning Toolbox.

Funkce *stepwiseglm* každým krokem hledá a přidává či vylučuje proměnné na základě odchylky modelu. Například z modelu odstraňuje příliš korelované parametry. Výsledná kombinace parametrů byla trénována při CV (Cross-validation, křížové validaci).

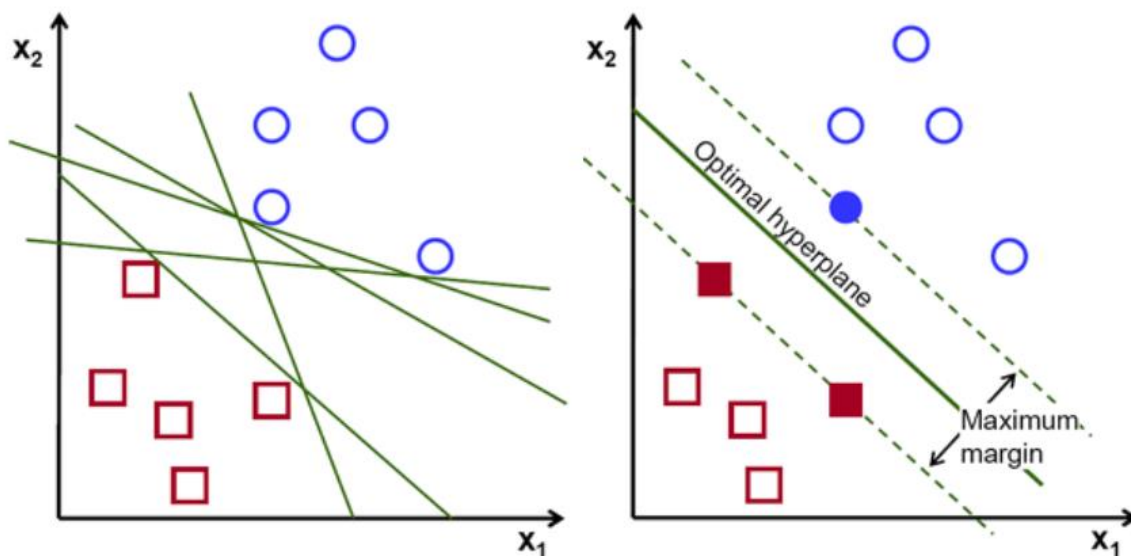
2.3.2 Support Vector Machine (SVM)

SVM model je typem „supervised learning“ algoritmu (učení s učitelem) používané buďto pro regresi nebo klasifikaci. V případě této práce byl tvořen klasifikační model. Hlavní myšlenkou SVM je na základě trénovacích dat nalézt optimální nadrovinu (hyperplane), která rozdělí data ve vícedimenzionálním prostoru a u testovacího datasetu se bude užívat pro klasifikaci. Pro zjednodušení bude problematika vysvětlena na příkladu dvoudimenzionálního problému, kde tvoří nadrovinu pouhá přímka.

Způsob klasifikace do skupin je u SVM odlišný oproti ostatním klasifikačním modelům., kde se většinou hledají nejčastější a největší rozdílnosti skupin, a na základě této charakteristiky jsou data rozdělena. SVM se oproti tomu zaměřuje na nejvíce podobné body mezi skupinami, tyto body jsou označovány jako support vectors (podpůrné vektory). Na základě těchto bodů je umístěno separační pásmo (margin), jehož šířku se snažíme maximalizovat. Rozhodovací hranice je umístěna do středu separačního pásma. Šířka se maximalizuje dle vzorce číslo 21, který popisuje maximalizaci nejmenší vzdálenosti bodu násobící predikci. Φ zde značí zobrazení parametrů, w je váhový vektor, t značí skupinu a b zaujetí.

$$\arg \max_{w,b} \frac{1}{\|w\|} \min_i [t_i(\phi(x)^T w + b)] \quad (21)$$

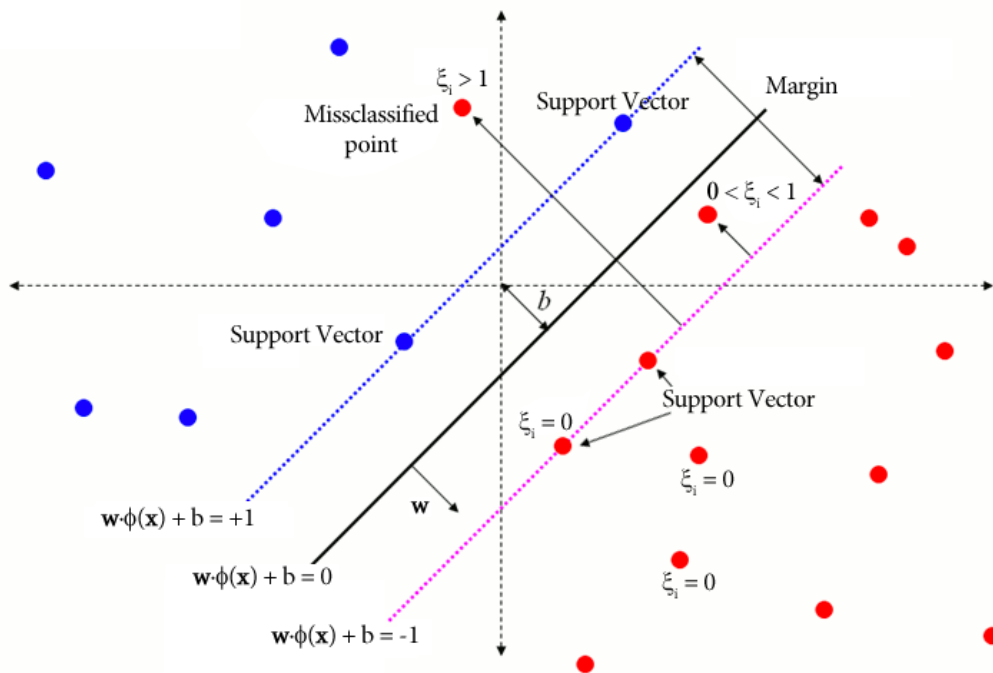
Support vectors leží přímo na okrajích pásma a pro model jsou esenciální, jejich odstranění by vedlo ke změně modelu. Pro predikci stačí algoritmu pamatovat si tyto body a zbytek vstupních dat může vypustit. Toto platí při předpokladu lineární separability dat, kdy je typ SVM označován jako Hard margin a počet podpůrných vektorů je roven $D+1$, kde D je počet dimenzí (parametrů) modelu. V případě dvoudimenzionálního prostoru, který je vyznačen na obrázku číslo 6, by nám tedy pro predikci stačily pouhé tři body. Support vectors jsou na obrázku znázorněny výplní bodu. [31][32]



Obrázek 6: Hledání ideální nadroviny (optimal hyperplane) ve dvoudimenzionálním prostoru. Šířka pásma je zde značena jako maximum margin a support vectors jsou body obsahující barevnou výplň. Převzato z [30].

V případě dat, která nejsou lineárně separabilní je nutno použít metodu Soft margin. Na obrázku 7 je vidět, že některé body se objevují v separačním pásmu nebo jsou dokonce na druhé straně nadroviny. Za všechny špatně klasifikované body se platí určitá penalta označována jako slack variable ξ_i . Při maximalizaci pásma je nutné tuto penalizaci započítat, a proto je původní funkce rozšířena o penalizační část, kterou shrnuje rovnice číslo 22, kde $C > 0$ je tzv. cost parametr, který rozhoduje o důležitosti špatně umístěných pozorování. Malé hodnoty C vedou k více benevolentním modelům, zatímco vysoká hodnota C by způsobila řešení blížíící se metodě Hard margin. [33]

$$\min_{w, \xi, b} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \right) \quad (22)$$



Obrázek 7: Soft margin SVM případ s lineárně neseparabilními daty. Převzato z [33]. (missclassified point – špatně zařazený bod)

V případě reálných dat jsou málokdy vztahy mezi parametry lineární a snadno separovatelné. K řešení tohoto problému SVM používá metodu nazývanou kernel trik, což je univerzální řešení nevztahující se pouze k SVM. Tato metoda popisuje mapování dat z jejich originálního prostoru do prostoru s vyšší dimenzí. K tomu jsou užívány tzv. kernelovy funkce, mezi které patří například lineární kernel, polynomiální kernel, gaussian kernel (RBF, radial basis function) a další. [33]

Vzhledem k tomu, že model SVM pracuje pouze s podmnožinou dat, vyniká svou nižší výpočetní náročností. Je vhodný pro použití u více dimenzí a vyznačuje svou efektivitou proti jiným modelům, obzvláště v případě vyššího počtu dimenzí než pozorování. Nevýhodou je, že kernel modely jsou náchylnější na přetrénování. [31]

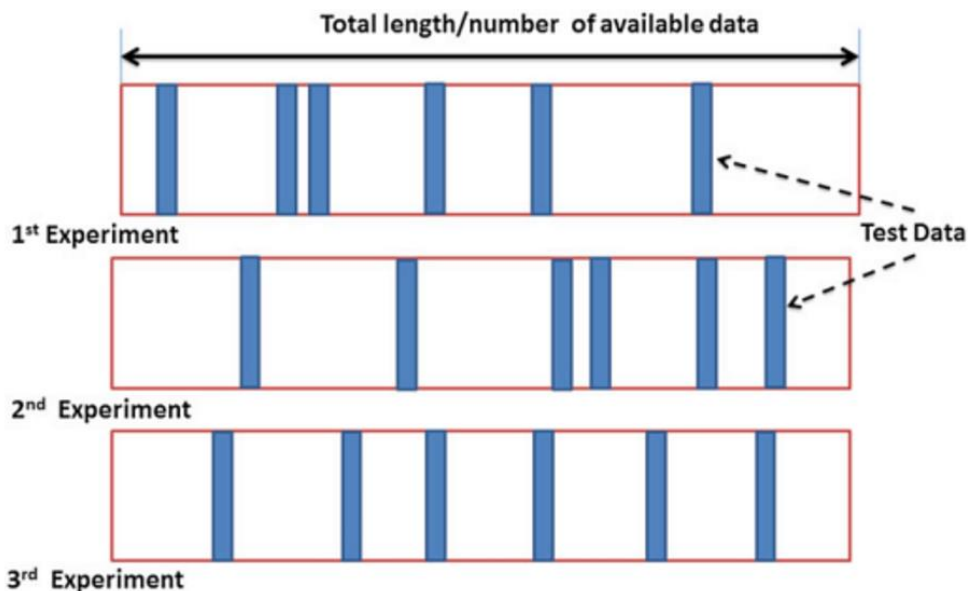
Pro tuto práci byla zvolena metoda Soft margin a lineární kernel, který je uváděn jako vhodný pro užití na datech s větším počtem parametrů než pozorování. [32]

Byly využity funkce *fitcsvm*, *fitSVMPosterior* a *predict* z knihovny Statistics and Machine Learning Toolbox.

2.4 Křížová validace

Důležitým aspektem statistického modelu je určení schopnosti predikce a odhadu její chyby. Z toho důvodu byla provedena křížová validace (CV). Tato metoda se používá v případě, kdy nemáme žádná nezávislá data na testování. Pomáhá nám vybrat ten nejvhodnější model a ujistit se, že model není přetřénovaný. Přetřénovanost modelu se vyznačuje tím, že model až příliš dobře pasuje na trénovací data, v případě testovací dat je ale nepoužitelný. Takový model má velký rozptyl, ale nízkou zaujatost. Data trénovacího setu si pamatuje, neučí se ale charakteristiku a vztahy bodů. [23]

Pro potřeby této diplomové práce byla vybrána metoda typu Monte Carlo. K tomuto typu křížové validace je nutno provést více iterací⁴ (minimálně 10), při každé iteraci je dataset náhodně rozdělen na trénovací a testovací část, jejich poměr je dopředu určen. Postup náhodného výběru dat metody Monte Carlo je vizualizován na obrázku číslo 8. V případě této práce trénovací dataset obsahoval při každé iteraci 70 % dat a iterací bylo 15.



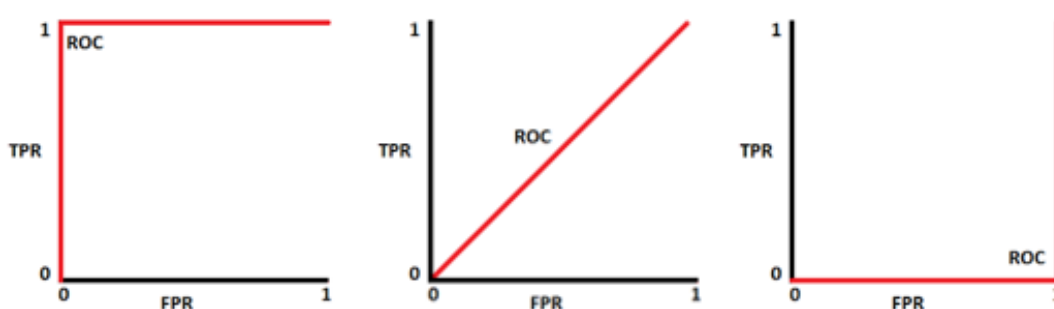
Obrázek 8: Postup náhodného výběru dat metody Monte Carlo při 3 iteracích. Převzato z [27]. Total length/number of available data (celkový vzorek pacientů), 1st-3rd Experiment (Označení počtu vzorově naznačených iterací)

Tato metoda je vhodná k verifikaci lineárních modelů a větších datasetů. Tento přístup udává v porovnání s jinými typy CV více pesimistické výsledky, což může být považováno za výhodu vzhledem k tomu, že testování na externích datech mimo CV generuje realističtější výsledky. Nevýhodou je, že se může stát, že některý pacient nebude nikdy vybrán do testovací podmnožiny. [8][27]

⁴ Opakující se průchod cyklem.

2.5 ROC křivka a AUC

ROC (Receiver Operator Characteristic) křivka je způsob vizualizace a výběru klasifikačních modelů na základě jejich výkonu. Jedná se o graf o dvou dimenzích, v kterém osa x popisuje FPR (False Positive Rate, 1-Specificita) a osa y TPR (True Positive Rate, Senzitivitu). Jednotlivé body tvořící ROC křivku jsou vypočítané z pravděpodobnostního skóre správné predikce a tvoří jednotlivé prahové hodnoty křivky. Pro každý práh existuje poměr pravdivých pozorování, které mají skóre větší nebo rovno tomuto prahu a stejně tak u poměru falešných pozorování. Pro dobrou predikci modelu chceme, aby se křivka nacházela v horním levém trojúhelníkovém prostoru, tedy aby byla plocha pod křivkou (AUC, Area Under Curve) co největší. [9][10]



Obrázek 9: Extrémní případy ROC křivek. Převzato z [28].

Extrémní případy ROC křivek jsou vizualizovány na obrázku číslo 9. Nalevo je vyobrazen ideální model s nejvyšší možnou senzitivitou při nejnižším poměru FPR (AUC = 1). Prostřední obrázek ukazuje diagonální přímku $y=x$ reprezentující model, který nepredikuje. Tento případ je popisován jako model hádající výsledky (AUC = 0,5). Nejvíce na pravo je zvláštní případ, kdy model predikuje negativní třídu jako pozitivní a naopak (AUC = -1).

Senzitivita je schopnost testu správně najít hledaný znak u testovaných subjektů neboli pravděpodobnost positivity, když je znak skutečně přítomen. Postup výpočtu můžeme vidět u rovnice 23, kde TP znamená počet správně pozitivních výsledků a FN falešně negativní výsledky. [11]

$$\text{Senzitivita} = \frac{TP}{TP+FN} \quad (23)$$

Specificita udává pravděpodobnost, se kterou test přesně vybere případy, u nichž hledaný znak nenastává. Její výpočet je ukázán na rovnici 24, kde TN charakterizuje správně negativní výsledky a FP falešně pozitivní. [11]

$$\text{Specificita} = \frac{TN}{TN+FP} \quad (24)$$

Z hodnot specificity a senzitivity modelu je možno vypočítat přesnost (accuracy, ACC) dle vzorce 25.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (25)$$

S pomocí ROC křivky byl vybrán rozhodovací práh pro zařazení do skupiny Fis/NoFis viz kapitola 2.8.

K realizaci ROC křivky v programu MATLAB byla použita funkce *perfcurve* z knihovny Statistics and Machine Learning Toolbox.

2.6 Korekce vícenásobného testování

Při současném testování většího množství hypotéz je nutné udělat korekci hladiny významnosti α , za účelem omezení falešné pozitivivity testů. V případě této práce byla zvolena korekce chyby prvního druhu typu Holm-Bonferroni. Tato metoda v porovnání s klasickou Bonferroniho korekcí nemá tak ostrá kritéria a disponuje větší statistickou silou testu. Vypočteme ji dle rovnice 26, kde α je hladina významnosti klasicky stanovena na hodnotu 0,05, n je počet testů. Pořadí testu je stanoveno seřazením výsledných p-hodnot od nejmenší po největší, přičemž nejnižší hodnota dostane hodnotu 1, druhá nejmenší 2 apod. [38]

$$HB = \frac{\alpha}{(n-\text{pořadí testu}+1)} \quad (26)$$

Tato korekce byla v práci použita u univariatního testování nezávislých skupin.

2.7 Zaslepená studie

V rámci zaslepené studie byla použita data od pacientů, kteří byli vyloučeni ze základního datasetu z důvodu chybějícího či nekvalitního záznamu EKG. Jednalo se o 61 pacientů s průměrným věkem $64,5 \pm 10,9$ let, mezi kterými bylo 17 pacientů postihnuto pooperační fibrilací síní.

Podstatou zaslepené studie je posouzení predikce modelu na neznámých datech. Je posouzena úspěšnost klasifikace modelů SVM a GLM na datech, které nebyly použity pro trénování a model je dopředu neviděl, jsou tedy náhodná.

2.8 Práh predikce

ROC křivka nám kromě hodnocení velikosti plochy pod křivkou dává informaci o velikosti TPR a FPR. Ideální model by měl TPR na hodnotě 1 a FPR 0, odhalil by tedy všechny pacienty s hrozcí fibrilací síní a zároveň ohodnotil správně i ty, kterým fibrilace pravděpodobně nehrozí. Ideálnímu modelu se v realitě jen blížíme, z pohledu klinického hodnocení je důležité nastavit si vhodný poměr pro danou problematiku mezi FPR a TPR.

Pro výpočet hodnot TPR a FPR se používá pravděpodobnost správné predikce určité skupiny. Toto skóre dává funkce *predict*, která spočítá pro každého pacienta z testované skupiny pravděpodobnost zařazení do skupiny Fis. Tyto hodnoty pravděpodobností tvoří u ROC křivky jednotlivé prahy predikce.

S pomocí funkce *perfcurve* je možno vypočítat optimální bod na ROC křivce, která najde přímku s ideálním sklonem a posunem z horního levého rohu ROC grafu směrem k dolnímu pravému a najde optimální bod v místě protnutí ROC křivky. Tento způsob byl použit u každé iterace CV a z výsledků byla vypočtena střední hodnota pro výpočet prahu predikce, který byl následně otestován v rámci zaslepené studie.

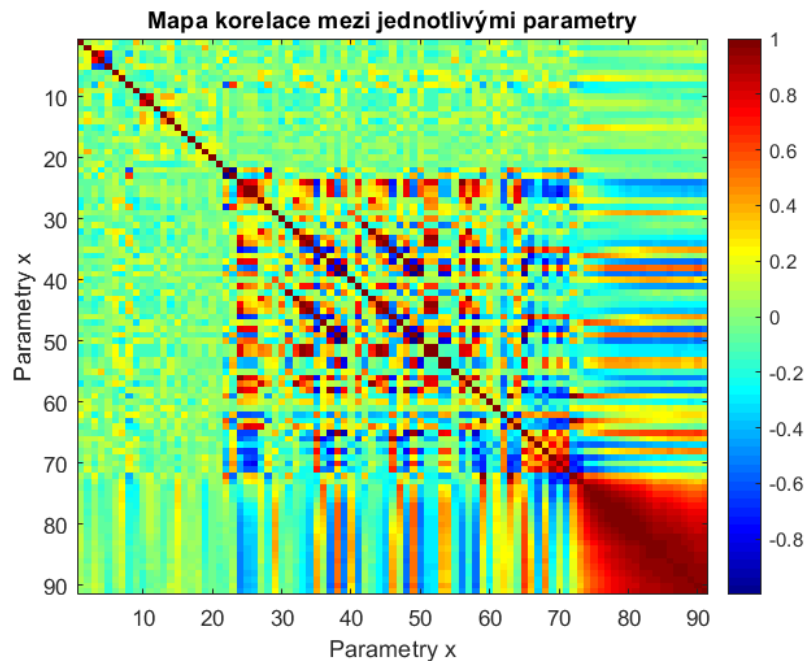
3 Výsledky

V této kapitole jsou popsány výsledky vyjmenovaných postupů v metodické části. Veškerá práce s daty byla prováděna v programu MATLAB. Práce byla vypracována na datech 82 pacientů, kteří byli popsáni pomocí 115 parametrů. Na tomto datasetu byla provedena univariantní a multivariantní statistika.

Finálně byly zvoleny 2 modely odlišného principu obsahující 4 prediktory, které generovaly nejlepší výsledky při predikování skupin NoFis a Fis. Schopnost predikce těchto modelů byla následně vyzkoušena na nezávislém vzorku 61 pacientů v rámci zaslepené studie.

3.1 Výsledky explorační analýzy a testování normality

V rámci explorační analýzy byla prozkoumána korelace mezi jednotlivými spojitými parametry, ta je vyobrazena na obrázku číslo 10. Sloupec hodnot na pravé straně obrázku udává hodnotu Pearsonova korelačního koeficientu R v rozsahu -1 až 1. Rudá barva značí hodnotu 1 udává velmi silnou závislost. Střední závislost je uváděna v rozmezí hodnot 0,4 až 0,8. Mezi parametry neexistuje žádný vztah v zelených oblastech mapy a záporné modré hodnoty implikují inverzní závislost. Rudá diagonála značí setkávání stejných parametrů. Parametry ležící na ose x a y jsou totožné, jejich seznam je uveden v příloze A. Významně korelovaná je zde oblast pravého dolního rohu, kde se nachází 20 parametrů MSE.

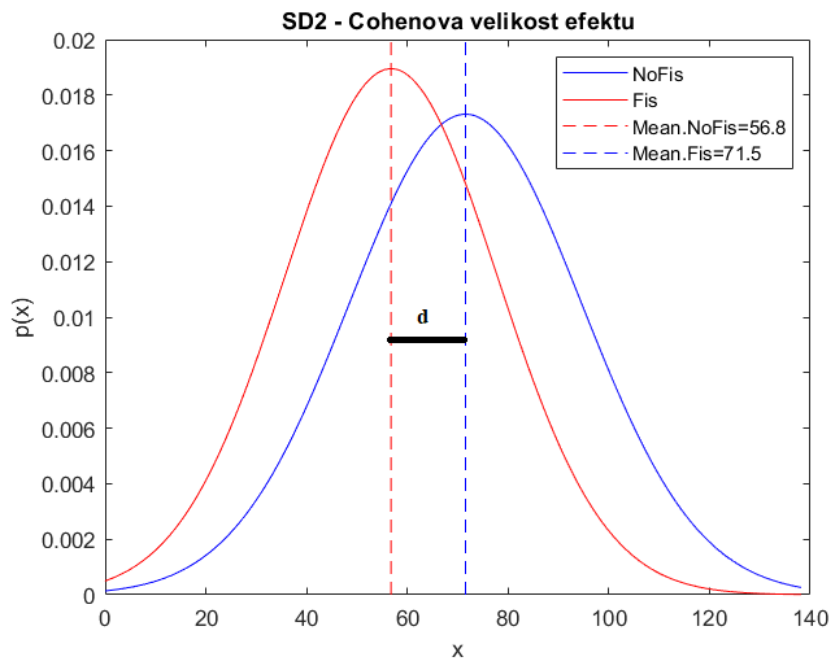


Obrázek 10: Korelační mapa mezi všemi parametry datasetu. V rozmezí hodnot x 1-8 najdeme parametry EKG, 9-21 obecnou charakteristiku pacienta, 22-28 časovou oblast HRV analýzy, 29-55 frekvenční oblast HRV analýzy a 56-91 nelineární oblast HRV analýzy.

Při testování normality dat metodou Anderson-Darling bylo zjištěno, že velká část dat nepochází z normálního rozdělení. Konkrétně byla data rozdělena na 24 parametrů s normálním rozdělením a 91 s odlišným rozdělením. Dataset obsahuje kromě spojitéch proměnných i několik kategoričkových proměnných, jako například cílovou proměnnou obsahující informaci o skupině Fis/NoFis.

3.2 Výsledky parametrických testů

Data s normálním rozdělením byla testována nezávislým Studentovým T testem s hypotézou, že mezi skupinami Fis a NoFis existuje statisticky signifikantní rozdíl. Tato alternativní hypotéza byla přijata pouze u parametru SD2 (vysvětlení parametru viz tabulka číslo 1) s $p=0,0047$, jehož výsledek byl ale následně zkorigován Holm-Bonferroniho korekcí a byl také zamítnut. Podstatným hodnotícím parametrem u parametrických testů je i Cohenova síla efektu a ta označila rozdíl skupin jako velký ($d=0,895$). Cohenova síla efektu parametru SD2 je vidět na obrázku 11.



Obrázek 11: Normální rozdělení skupin parametru SD2 s vyznačenými středními hodnotami a Cohenovou velikostí efektu. Mean (střední hodnota).

3.3 Výsledky neparametrických testů

Pro všechny spojité parametry s jiným, než normálním rozdělením, byla s pomocí Wilcoxon-Mann-Whitneyho testu otestována hypotéza, že mezi skupinami Fis a NoFis je statisticky signifikantní rozdíl. Výsledkem testu je tabulka 3, která popisuje výsledné hodnoty před Holm-Bonfferoniho korekcí. Po korekci chyby prvního druhu nebyl nalezen žádný signifikantní výsledek. Společně s p-hodnotou je zde vypsan i parametr síly efektu r pro neparametrické testy, který naznačuje ve všech případech malý efekt.

Tabulka 3: Tabulka výsledků neparametrického Wilcoxon-Mann-Whitneyho testu před Holm-Bonfferoniho korekcí

| Parametry | | | Parametry | | |
|-------------------|-------|--------|---------------------------------|-------|-------|
| Časová oblast | p | r | Frekvenční oblast | p | r |
| RMSSD | 0,036 | 0,231 | AbsPowVLFAR | 0,007 | 0,299 |
| NN50 | 0,035 | 0,233 | AbsPowLFAR | 0,022 | 0,252 |
| pNN50 | 0,031 | 0,233 | AbsPowHFAR | 0,027 | 0,245 |
| Nelineární oblast | p | r | PeakLFFFT | 0,041 | 0,226 |
| SD1 | 0,036 | 0,231 | AbsPowVLFFFT | 0,009 | 0,290 |
| SlopeIndex | 0,030 | -0,240 | AbsPowLFFFT | 0,019 | 0,259 |
| MSE14 | 0,049 | -0,218 | AbsPowHFFFT | 0,027 | 0,245 |
| MSE16 | 0,043 | -0,224 | TotPowAR | 0,012 | 0,278 |
| MSE17 | 0,047 | -0,220 | TotPowFFT | 0,012 | 0,279 |
| MSE18 | 0,037 | -0,230 | Obecná charakteristika pacienta | p | r |
| | | | CRP | 0,019 | 0,259 |

Kategoriální proměnné byly testovány s pomocí kontingenčních tabulek a Pearsonova chí-kvadrát testu nezávislosti skupin. Mezi kategoriálními proměnnými nebyla shledána žádná statisticky významná proměnná.

3.4 Výsledky modelu GLM

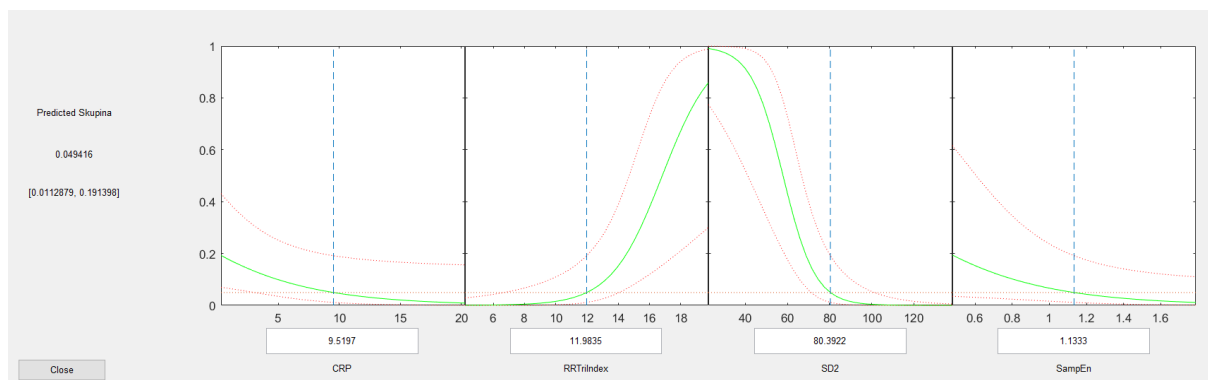
V této části práce byl hledán model, který by kombinací parametrů poskytl lepší výsledky než univariální statistika. Funkce `stepwiseglm` vytvořila model obsahující 4 parametry: SD2, CRP (C-reaktivní protein), `SampEn` a `RRTriIndex`. Rovnice 27 orientačně vyjadřuje jejich vztah viz níže.

$$P(\text{Skupina Fis}) \sim 1 + CRP + RRTriIndex + SampEn + SD2 \quad (27)$$

Z obrázku číslo 12 lze vyčíst, že parametry RRTriIndex a SD2 mají větší vliv na predikci než SampEn a CRP. Zároveň je vidět, že s rostoucím parametrem RRTriIndex stoupá pravděpodobnost určení do skupiny Fis, ostatní parametry mají opačný trend. To se dá vyčíst i z výstupní tabulky estimačních koeficientů modelu, kde je dle velikosti hodnoty T statistiky a hodnot p nejdůležitější parametr SD2.

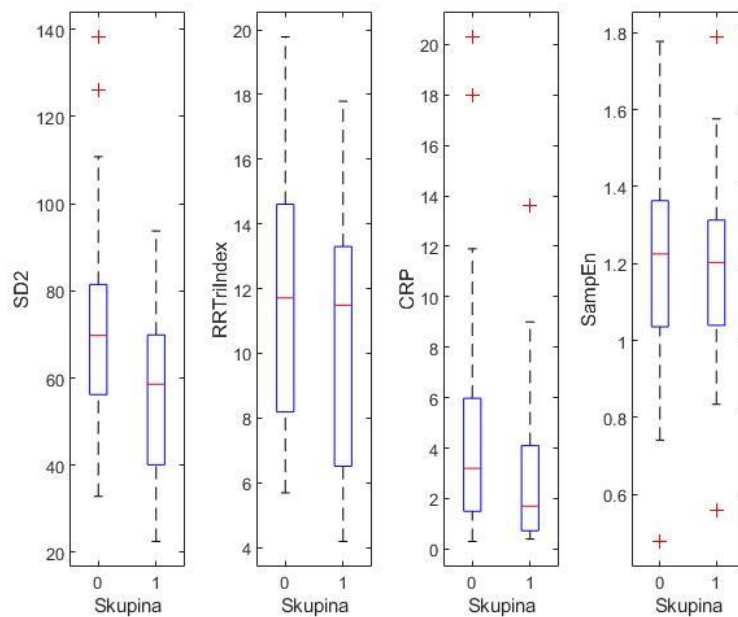
Tabulka 4: Estimační koeficienty modelu GLM (Estimate – sklon přímky, SE – standart error, bodové odhady směrodatných odchylek přímek, tStat – T-statistika, p – p-hodnota)

| | Estimate | SE | tStat | p |
|------------|-----------------|-----------|----------------|---------------|
| Intercept | 4,5304 | 1,7382 | 2,6064 | 0,0091 |
| CRP | -0,1652 | 0,0842 | -1,9626 | 0,0497 |
| RRTriIndex | 0,6102 | 0,2065 | 2,9548 | 0,0031 |
| SD2 | -0,1316 | 0,0370 | -3,5543 | 0,0004 |
| SampEn | -2,3341 | 1,1988 | -1,9470 | 0,0515 |



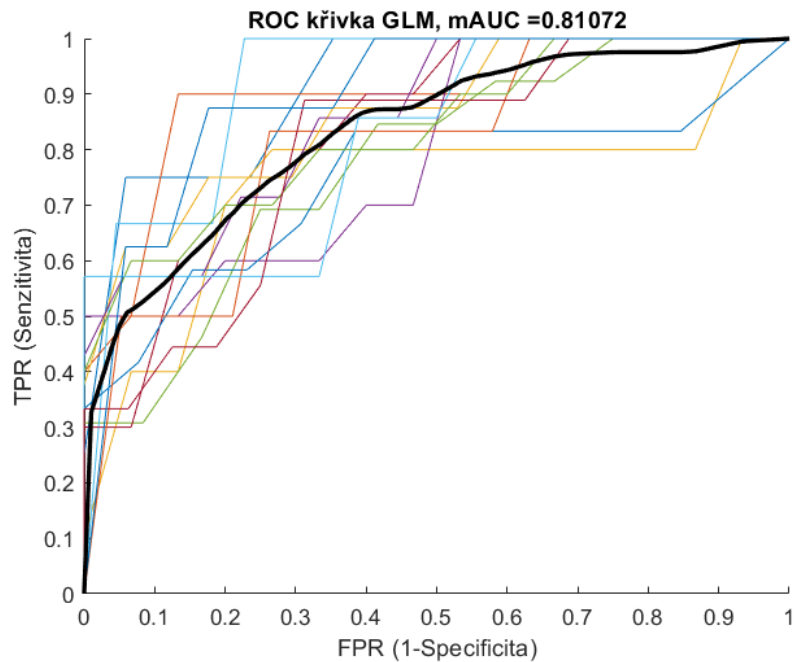
Obrázek 12: Výstupní křivky GLM modelu typu logistické regrese (funkce stepwiseglm) obsahující parametry CRP, RRTriIndex, SD2 a SampEn vztažené k predikované Skupině

Základní charakteristiku v podobě krabicových diagramů těchto parametrů je možné prohlédnout na obrázku 13, kde skupina 0 značí pacienty s NoFis a 1 pacienty s Fis.



Obrázek 13: Krabicové diagramy signifikantních parametrů používaných v obou výstupních modelech

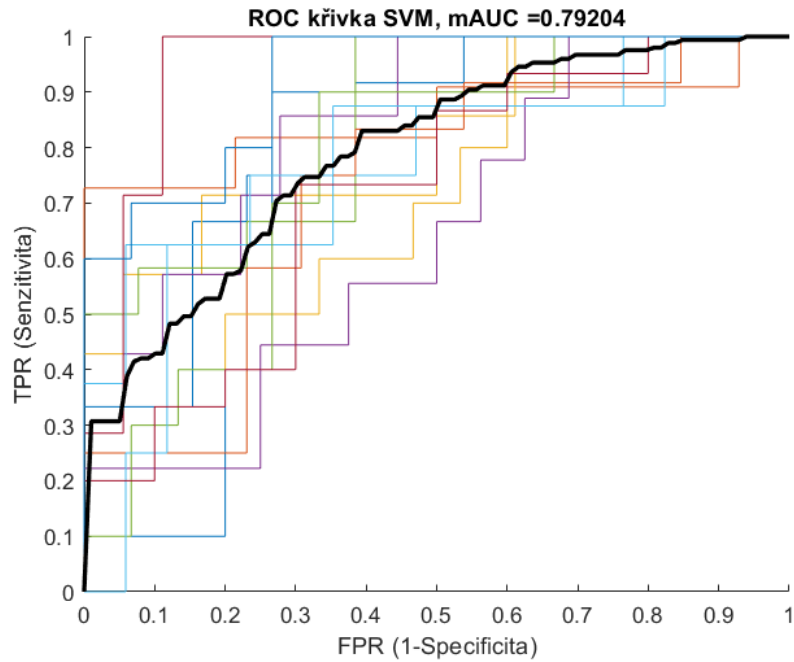
Získané parametry byly dále využity při trénování a testování modelů během křížové validace. Nejlepší výsledky byly získány u modelu GLM typu logistické regrese, kde byla, vzhledem k cílové proměnné disponující binomickým rozdělením, zvolena logit link funkce. Výsledky modelů byly porovnávány dle průměrné hodnoty AUC z 15 iterací získané z ROC křivky. ROC křivka modelu je vizualizována na obrázku 14. Na ose x je vynesena hodnota FPR a na ose y TPR. Na grafu je viditelný průběh každé ROC křivky z 15 iterací, černě je vyobrazená zprůměrovaná ROC křivka. Z výstupních hodnot ROC křivky byl vypočten práh predikce 0,5499.



Obrázek 14: Výsledná ROC křivka GLM modelu

3.5 Výsledky modelu SVM

Dle dosažených výsledků byl jako druhý nejlepší vyhodnocen model SVM. U tohoto modelu dosahuje zprůměrovaná hodnota AUC z 15 iterací hodnoty 0,79. Příslušná ROC křivka je zobrazena na obrázku číslo 15. Stejně jako u předchozího grafu je na ose x vynesena hodnota falešně pozitivního poměru výsledků a na ose y poměr správně pozitivních výsledků. Variability jednotlivých křivek je větší než v případě modelu GLM, což může značit méně stabilní model. Práh predikce byl ze zprůměrovaných výstupů 15 iterací ROC křivky určen na hodnotu 0,5291.

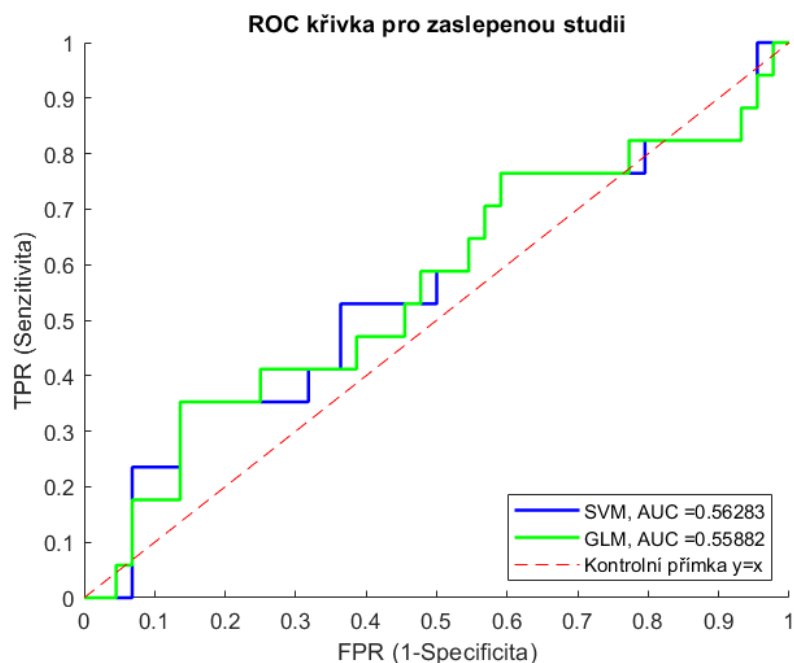


Obrázek 15: Výsledná ROC křivka SVM modelu

3.6 Výsledky zaslepené studie

Jednou ze zkoumaných problematik práce bylo zhodnotit užitečnost nových parametrů vyčtených z EKG pacientů. Vzhledem k tomu, že žádný z těchto parametrů nebyl začleněn do výstupních modelů práce, otevřela se možnost otestování modelů na nezávislých datech, která byla v prvním kroku práce vyřazena z důvodu chybějících elektrokardiogramů.

Bohužel ani jeden z modelů zde nedosahuje příliš příznivých výsledků, výkonnost predikce obou modelů je přibližně stejná, dle hodnoty AUC vychází model SVM o trochu lépe. Příslušná ROC křivka je zobrazena na obrázku číslo 16. Zeleně je zde vyznačena křivka GLM modelu s výsledným AUC = 0,559, modře ROC křivka modelu SVM s hodnotou AUC = 0,563. Pro přehlednost je na grafu červenou přerušovaně vyznačena kontrolní přímka s hodnotou AUC = 0,5.



Obrázek 16: Porovnání výkonu modelů SVM a GLM na nových datech

Přehled výsledků nastaveného prahu predikce najdeme v tabulce 5. Klasifikační práh, který byl určen během křížové validace na trénovacích dat, byl pro GLM stanoven na hodnotu pravděpodobnosti predikce skupiny Fis $P(\text{Fis}) = 0,5499$ a pro model SVM $P(\text{Fis}) = 0,5291$. Výsledky vycházejí velmi podobně, ačkoliv přesnost prahu GLM je o 1 % větší, což (vzhledem k lepšímu výsledku SVM dle AUC) značí větší stabilitu predikce během trénovací části křížové validace.

Tabulka 5: Úspěšnost nastaveného prahu predikce u modelu SVM a GLM

| | GLM – P(Fis) 0,5499 | SVM – P(Fis) 0,5291 |
|--------------------|----------------------------|----------------------------|
| TP | 6 | 3 |
| TN | 38 | 40 |
| FN | 11 | 14 |
| FP | 6 | 4 |
| Senzitivita | 0,3529 | 0,1765 |
| Specificita | 0,8636 | 0,9091 |
| FPR | 0,1364 | 0,0909 |
| ACC [%] | 72 | 71 |

Pro přiblížení situace je zde přiložena tabulka číslo 6 obsahující příklady hodnot důležitých parametrů vždy jednoho zástupce z pacientů charakterizující ukazatel TP, TN, FN a FP. Dle našich předpokladů by měl mít největší vliv parametr SD2, jež se s klesající hodnotou přibližuje skupině Fis.

Tabulka 6: Vybrané subjekty zastupující ukazatele TP, TN, FN a FP

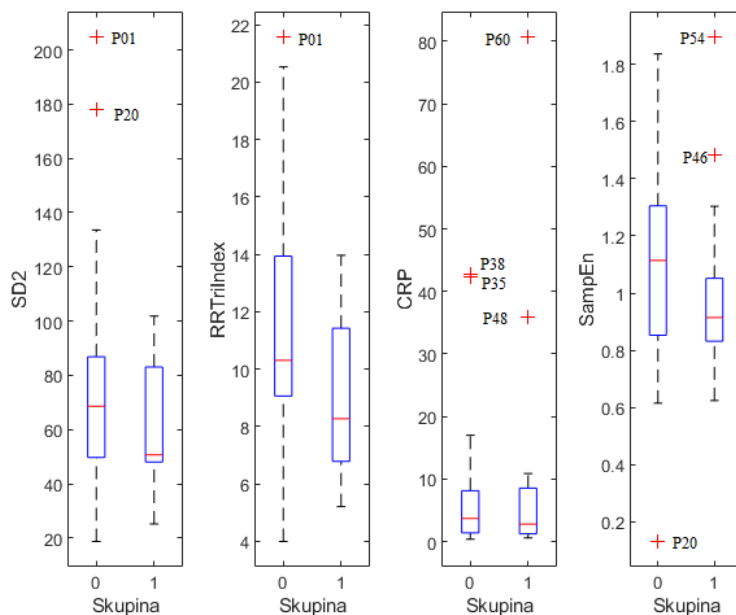
| | CRP | RRTriIndex | SD2 | SampEn |
|-----------------|------------|-------------------|------------|---------------|
| TN (P02) | 2,30 | 13,96 | 81,95 | 1,29 |
| FN (P49) | 1,70 | 13,78 | 83,16 | 1,16 |
| TP (P61) | 8,70 | 5,21 | 25,20 | 0,91 |
| FP (P03) | 1,20 | 9,87 | 47,08 | 1,47 |

Vzhledem k nepříliš dobrému výsledku zaslepené studie byla zvážena možnost, zda se vzorek nových dat příliš nevymyká datasetu trénovacímu či jestli nový dataset neobsahuje příliš odlehlých hodnot, které by mohly zkreslovat celkový výsledek. K prohlédnutí základních charakteristických hodnot jako je střední hodnota, standardní odchylka a medián slouží tabulka číslo 7. Velmi rozdílná střední hodnota a medián (zároveň i široký rozptyl) u parametru CRP naznačuje přítomnost velmi odlehlé hodnoty.

Tabulka 7: Porovnání důležitých parametrů základních dat a dat zaslepené studie. Mean (střední hodnota), std (směrodatná odchylka).

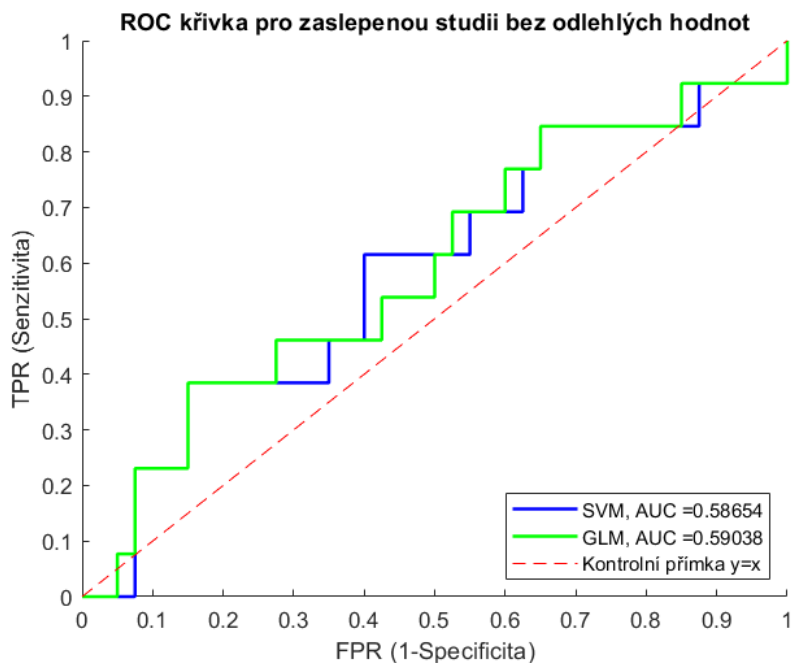
| Skupina | Základní data | | Zaslepená data | |
|-------------------|---------------|----------|----------------------|-------------|
| | NoFis | | NoFis | |
| | Mean ± Std | Median | Mean ± Std | Median |
| CRP | 4,59 ± 4,28 | 3,20 | 6,46 ± 8,93 | 3,70 |
| RRTriIndex | 11,41 ± 3,76 | 11,71 | 11,65 ± 4,55 | 10,31 |
| SD2 | 71,54 ± 23,05 | 69,81 | 75,24 ± 38,05 | 68,61 |
| SampEn | 1,21 ± 0,28 | 1,22 | 1,10 ± 0,33 | 1,11 |
| Skupina | Základní data | | Zaslepená data | |
| | Fis | | Fis | |
| | Mean ± Std | Median | Mean ± Std | Median |
| CRP | 2,84 ± 3,01 | 1,7 | 10,14 ± 20,02 | 2,8 |
| RRTriIndex | 10,57 ± 3,99 | 11,486 | 9,14 ± 2,92 | 8,271688 |
| SD2 | 56,77 ± 21,05 | 58,5955 | 60,35 ± 22,93 | 50,65221 |
| SampEn | 1,19 ± 0,25 | 1,202711 | 0,99 ± 0,32 | 0,9144205 |

Nalezení odlehlých dat bylo provedeno s pomocí krabicových diagramů, které jsou zobrazeny na obrázku 17. Celkově bylo detekováno 10 odlehlých hodnot, z toho 2 pacienti se zde objevili dvakrát. Všechny nalezené odlehlé hodnoty byly následně z datasetu odstraněny pro další testování.



Obrázek 17: Základní charakteristika důležitých parametrů nových dat. Hodnoty PXX označují číslo pacienta, který byl stanoven jako odlehlá hodnota a byl z testování vyřazen.

Na obrázku 18 je vidět, že klasifikace modelů se po odstranění odlehlých hodnot zlepšila. Model SVM má hodnotu $AUC = 0,587$ a GLM se dostal na hodnotu $AUC = 0,590$. Na základě provedené analýzy je možné říct, že model GLM byl více ovlivněn odlehlými hodnotami.



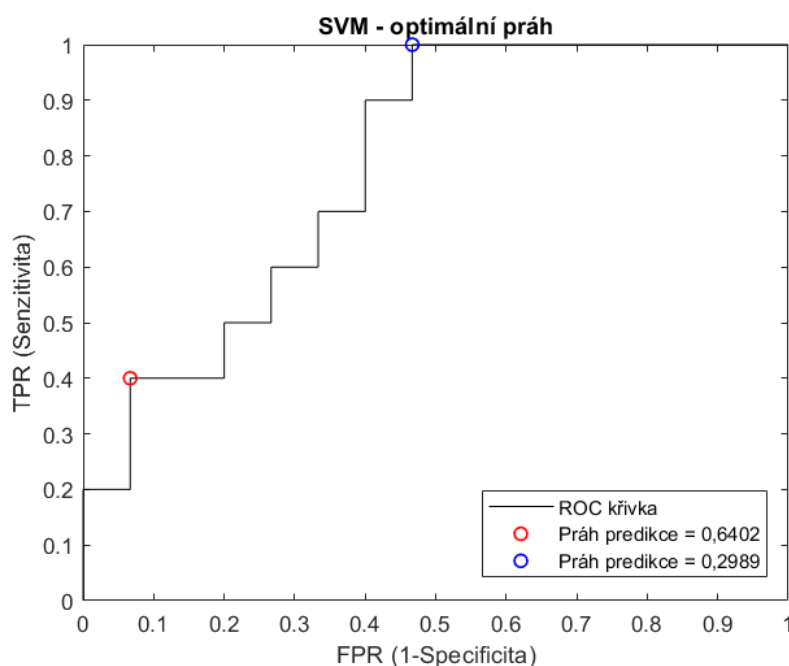
Obrázek 18: Porovnání modelů GLM a SVM na nových datech s odstraněnými odlehlými hodnotami

3.7 Problematika určení prahu predikce

V závěru výsledků práce bych poukázala na určitém příkladu problematiku určení vhodného prahu predikce. V tabulce 8 je možno vidět hodnoty správně pozitivních, správně negativních, falešně pozitivních a falešně negativních výsledků jedné predikce modelu SVM během CV. Z těchto výsledků je vypočtena Senzitivita, Specificita, FPR a přesnost (ACC). Je zde testováno celkově 25 subjektů. Práh $P(\text{Fis}) = 0,6402$ byl automaticky vypočten s pomocí ROC křivky a udává nám velmi nízkou falešnou pozitivitu na úkor senzitivity. Pokud bychom se ale zaměřili na nalezení prahu, který vybere správně všechny pacienty s hledaným znakem musíme práh snížit na hodnotu $P(\text{Fis}) = 0,2989$. Tento postup je tedy nutno zkorigovat dle specifických potřeb.

Tabulka 8: Porovnání výsledků 2 různých prahových hodnot u jedné náhodně vybrané iterace modelu SVM

| | P(Fis) = 0,6402 | P(Fis) = 0,2989 |
|--------------------|------------------------|------------------------|
| TP | 3 | 10 |
| TN | 14 | 8 |
| FP | 1 | 7 |
| FN | 7 | 0 |
| Senzitivita | 0,3 | 1 |
| Specificita | 0,9333 | 0,5333 |
| FPR | 0,0667 | 0,4667 |
| ACC [%] | 68 | 72 |



Obrázek 19: Porovnání dvou prahů jedné náhodně vybrané iterace SVM

Na obrázku číslo 19 jsou znázorněny oba prahy na ROC křivce. Červeně je vyznačen optimální práh predikce dle funkce *perfcurve* s hodnotou pravděpodobnosti správného určení skupiny Fis. Modrý bod obsahuje ideální práh dle vizuálního hodnocení s předpokladem, že hledáme model s co nejvyšší senzitivitou.

4 Diskuze

Zkoumanou skupinou této práce byli pacienti indikovaní ke kardiochirurgickému výkonu typu aortokoronární bypass. Jejich předoperační historie neobsahovala výskyt fibrilace síní, ale během zotavovacího období po operaci se u nich objevila komplikace v podobě této patologie srdečního rytmu. Tato komplikace působí negativně na zotavení pacienta a prodlužuje jeho dobu strávenou na lůžku v nemocnici. Tato práce se zaměřuje na zkoumání parametrů, které by mohly působit jako prediktory pro odhalení rizika možnosti vzniku pooperační fibrilace síní. Lepší porozumění těchto mechanismů indukující riziko vzniku by mohlo pomoci pro nalezení vhodného terapeutického přístupu k omezení této komplikace.

V návaznosti na předešlou dizertační práci doktorky Veselé, která také pracovala s daty studie APOFIS, bylo navrženo zaměřit se pouze na část pacientů s operací CABG s důrazem na multivariantní statistiku. V porovnání s výše avizovanou prací byly přidány nové parametry v podobě hodnocení elektrokardiogramů. V rámci multivariantní analýzy byly navrženy 2 modely nejlépe hodnocené dle ROC křivky. Tyto modely obsahovaly kombinaci parametrů SD2, Triangular Index, CRP (C-reaktivní protein) a vzorkovou entropii. První dva zmiňované parametry se objevily jako signifikantní dle univariantní analýzy i v práci doktorky Veselé. [21] Tyto modely byly v závěru práce otestovány na nezávislém vzorku, kde byla zhodnocena neuspokojivá kvalita predikce. Tento problém byl dále vyšetřován a po odstranění velmi odlehlých hodnot ze zaslepených dat lehce vylepšen. Tento problém může být způsoben příliš malým vzorkem dat a nízkým procentem pacientů ve skupině Fis. Pro další rozvoj této práce by bylo vhodné celkový vzorek rozšířit o nová pozorování. Pro lepší výsledky při křížové validaci, stabilnější predikci a lepší interpretovatelnosti výsledků bych jako kvalitnější volila model GLM. Ačkoliv je model SVM velmi flexibilní, vzhledem k jeho podstatě (modelování ve vícedimenzionálním prostoru a možností použití nelineárních kernelů) je mnohem složitější popsat jeho výsledky než u případu generalizovaného lineárního modelu, kde je možné sledovat vliv jednotlivých parametrů. Tento jev je možné vyčíst i z literatury, kde se většina prací zaměřuje na univariantní statistiku, multivariantní regresní modely, logistickou regresi či GLM.

Vzhledem k tomu, že tato pooperační komplikace je jednou z nejčastějších ve spektru kardiochirurgických operací, byl tento jev už vícekrát zkoumán. V literatuře je uváděno, že předoperační hladiny CRP nemají signifikantní souvislost s postoperační fibrilací síní, existují ale protichůdná data o jeho pooperačních hladinách a vlivu. Mé výsledky naznačují spojitost nižší hladiny CRP se skupinou Fis. V naprosté většině studií je uváděn jako důležitý faktor věk. Negativní výsledek v této oblasti může být způsoben velmi jednotnou skupinou pacientů, kdy věk(FIS) = 66 ± 7 a věk(NoFis) = 64 ± 9 let. Při porovnání parametrů zkoumaných v této práci s literaturou se objevují zmínky o mužském pohlaví zvyšující pravděpodobnost vzniku Fis, vyšší hypertenzi, délce P vlny >110 ms, přítomnosti operací chlopně (pacienti s operací chlopni byli z našeho datasetu vyloučeni za účelem soustředit se pouze na jeden druh operace) či prodělaného infarktu myokardu. [14][29][37]

Nové parametry EKG se v této práci ukázaly z pohledu univariantské i multivariantské analýzy jako nedůležité.

Poslední zkoumanou problematikou práce bylo určení vhodného prahu predikce pro rozřazení nových pacientů do skupin Fis a NoFis. Prahy obou modelů, určené během křížové validace, byly testované na vzorku zaslepených dat. Výsledné procento přesnosti se sice zdá uspokojivé, ale je potřeba zvážit, zda preferujeme ideální poměr FPR a TPR či upřednostňujeme co největší poměr pozitivních výsledků. Při preferenci vysoké senzitivity by se kritérium pro určení prahu muselo přepracovat.

5 Závěr

Cílem této diplomové práce bylo nalézt prediktory indikující vznik pooperačních fibrilací síní a vytvořit práh predikce pro klasifikaci nových pacientů do skupin dle přítomnosti hledaného znaku. Dalším úkolem bylo ohodnotit přínos nových parametrů z elektrokardiogramu a posoudit podobnost výsledků s dříve provedenou prací Ing. Jany Veselé Ph.D.

V teoretickém rozboru byla popsána problematika týkající se vzruchové činnosti srdce a z toho vyvozených parametrů pro tuto práci. Dále zde byla uvedena prováděná operace zkoumané skupiny pacientů a patologie srdce z pohledu fibrilace síní. V rámci metodické části byla popsána klinická data a jejich získání. Následně se tato kapitola věnovala všem statistickým postupům, na základě kterých byla provedena praktická část v programu MATLAB.

Výstupem této práce jsou klasifikační modely SVM a GLM obsahující signifikantní parametry SD2 (směrodatná odchylka získaná z Poincareho mapy), Triangular index, CRP (C-reaktivní protein) a vzorkovou entropii. Při křížové validaci dosahují oba modely kvalitních výsledků. Plocha pod křivkou (AUC) se u obou modelů blíží hodnotě 0,8. Funkce klasifikace obou modelů byla testována na nezávislých datech v rámci zaslepené studie. U těchto dat ale nebyla prokázána stejná kvalita predikce. Hodnota parametru plochy pod křivkou se v tomto případě blíží pouze hodnotě 0,6. Práh predikce je určen s důrazem na ideální poměr mezi TPR a FPR a jeho přesnost se pohybuje okolo 72 %.

Parametry hodnotící průběh EKG, o které byl doplněn dataset původních klinických dat nasbíraných v rámci studie APOFIS, nebyly k výsledkům práce přínosné, v univariantsní ani multivariantsní analýze nevykazovaly statisticky významné výsledky.

V závěru práce byla provedena diskuze. Obsahem diskuze byly dosažené výsledky praktické části, jejich úskalí a možnosti pro vylepšení dalšího přístupu k problematice. Zároveň zde byla práce posouzena v souvislosti s dizertační prací doktorky Veselé. S prací paní doktorky Veselé se výsledky shodují v parametru SD2 a Triangular index, které jí ale vyšly signifikantní při univariantsní analýze na rozdíl od výsledku z multivariantsní analýzy uvedené v této práci.

Jako vhodnější model byl zvolen přístup GLM díky své lepší interpretovatelnosti a stabilitě výsledků. Pro další postup je navrženo zvýšit vzorek zkoumaných pacientů díky čemuž by bylo možné dosáhnout přesnějších výsledků. Dalším navrhovaným postupem je zaměření na práh predikce, s co největší senzitivitou pro odhalení všech rizikových pacientů.

Literatura

- [1] BULAVA, Alan. *Kardiologie pro nelékařské zdravotnické obory*. Praha: Grada Publishing, 2017. ISBN 978-80-271-0468-0.
- [2] EISENBERGER, M., A. BULAVA a M. FIALA. *Základy srdeční elektrofyziologie a katéetrových ablací*. Praha: Grada, 2012. ISBN 978-80-247-3677-8.
- [3] THOREN, E., et al. *Prediction of postoperative atrial fibrillation in a large coronary artery bypass grafting cohort*. Interactive CardioVascular and Thoracic Surgery [online]. 2012, **14**(5), 588-593 [cit. 2021-01-05]. ISSN 1569-9293. Dostupné z: doi:10.1093/icvts/ivr162
- [4] MATHEW, Joseph P. *A Multicenter Risk Index for Atrial Fibrillation After Cardiac Surgery*. JAMA [online]. 2004, **291**(14) [cit. 2021-01-05]. ISSN 0098-7484. Dostupné z: doi:10.1001/jama.291.14.1720
- [5] AMAR, D., et al. *Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting*. Journal of the American College of Cardiology [online]. 2004, **44**(6), 1248-1253 [cit. 2021-01-05]. ISSN 07351097. Dostupné z: doi:10.1016/j.jacc.2004.05.078
- [6] ZAKKAR, M., et al. *Inflammation, oxidative stress and postoperative atrial fibrillation in cardiac surgery*. Pharmacology & Therapeutics. [online]. 2015, **154**, 13-20 [cit. 2021-01-05]. ISSN 01637258. Dostupné z: doi:10.1016/j.pharmthera.2015.06.009
- [7] ČOUPKOVÁ, Hana, et al. *Ošetřovatelství v chirurgii*. 2 vyd.. Praha: Grada Publishing, 2019. Sestra (Grada). ISBN 978-80-247-2900-8.
- [8] XU, Qing-Song a Yi-Zeng LIANG. *Monte Carlo cross validation*. Chemometrics and Intelligent Laboratory Systems. 2001, **56**(1), 1-11. ISSN 01697439. Dostupné z: doi:10.1016/S0169-7439(00)00122-2
- [9] Statistics and Machine Learning Toolbox. *MathWorks - Makers of MATLAB and Simulink - MATLAB & Simulink* [online]. © 1994-2021 [cit. 2021-01-05]. Dostupné z: https://uk.mathworks.com/help/stats/index.html?s_cid=doc_ftr
- [10] FAWCETT, Tom. *ROC graphs: Notes and practical considerations for researchers* [online]. Machine learning, 2004, **31**(1), 1-38 [cit. 2021-01-05]. Dostupné z: <http://binf.gmu.edu/mmasso/ROC101.pdf>

- [11] PARIKH, R., et al. *Understanding and using sensitivity, specificity and predictive values*. Indian Journal of Ophthalmology. 2008, **56**(1). ISSN 0301-4738. Dostupné z: doi:10.4103/0301-4738.37595
- [12] PEKÁR, S. a M. BRABEC. *Modern analysis of biological data: generalized linear models in R*. Brno: Masaryk University, 2016. ISBN 9788021080195.
- [13] DOBSON, A. J. a A. G. BARNETT. *An introduction to generalized linear models*. 4th ed., revised. Boca Raton: CRC Press, Taylor & Francis Group, 2018. Chapman & Hall/CRC texts in statistical science series. ISBN 9781138741515.
- [14] EL-ESSAWI, A., et al. *Predictors of postoperative atrial fibrillation persisting beyond hospital discharge after coronary artery bypass grafting*. Perfusion [online]. 2020, 1-7[cit. 2021-01-05]. ISSN 0267-6591. Dostupné z: doi:10.1177/0267659120978647
- [15] ŠTEFÁNEK, Jiří. *Medicína, nemoci, studium na I. LF UK* [online]. © 2011 [cit. 2021-01-05]. Dostupné z: <https://www.stefajir.cz/>
- [16] Srdce. *Funkce buněk a lidského těla • Multimediální skripta* [online]. [cit. 2021-01-05]. Dostupné z: <http://fbt.cz/skripta/x-srdce-a-obeh-krve/1-srdce/>
- [17] POSPÍŠILOVÁ, B., J. ŠRÁM a O. PROCHÁZKOVÁ. *Anatomie pro bakaláře II.: systém kardiovaskulární, systém nervový, smyslové orgány, soustava kožní, žlázy s vnitřní sekrecí*. 2. vyd. Liberec: Technická univerzita, 2015. ISBN 978-80-7494-153-5.
- [18] BERNACIKOVÁ, M.. *Fyziologie* [online]. BRNO: Masarykova univerzita, 2012 [cit. 2020-12-31]. ISBN 978-80-210-5841-5. Dostupné z: <https://www.fsps.muni.cz/emuni/data/reader/book-3/08.html>
- [19] SOVOVÁ, Eliška. *EKG pro sestry*. Praha: Grada, 2006. Sestra (Grada). ISBN 80-247-1542-2.
- [20] SHAFFER, Fred a J. P. GINSBERG. *An Overview of Heart Rate Variability Metrics and Norms*. *Frontiers in Public Health* [online]. 2017, 5 [cit. 2020-12-31]. ISSN 2296-2565. Dostupné z: doi:10.3389/fpubh.2017.00258
- [21] VESELÁ, Jana. *Využití spektrální analýzy pro léčbu paroxysmální či perzistující fibrilace síní*. Praha, 2020. Dizertační práce. Univerzita Karlova 3. lékařská fakulta.
- [22] Scio - Statistické pojmy. *Scio – Oficiální stránky*[online]. © 2020 [cit. 2021-01-05]. Dostupné z: https://www.scio.cz/o-vzdelavani/teorie-a-metodika-testu/statisticke-pojmy/#zskor_t-skor

- [23] KOEHRSEN, Will. Overfitting vs. Underfitting: A Complete Example. In: *Towards Data Science* [online]. 2018 [cit. 2021-01-05]. Dostupné z: <https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765>
- [24] TARVAINEN, M. P., et al. *Kubios HRV – Heart rate variability analysis software*. Computer Methods and Programs in Biomedicine. 2014, **113**(1), 210. DOI: 10.1016/j.cmpb.2013.07.024. ISSN 01692607. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0169260713002599>
- [25] ROZMAN, J. *Elektronické přístroje v lékařství*. Praha: Academia, 2006. Česká matice technická (Academia). ISBN 80-200-1308-3.
- [26] SAVIO, M. T. a M. HARIHARAN. *Impact of psychosocial intervention on prognosis of cardiac surgery patients*. Health Psychology Research [online]. 2020, **8**(3) [cit. 2021-01-05]. ISSN 2420-8124. Dostupné z: doi:10.4081/hpr.2020.8887
- [27] REMESAN, R. a J. MATHEW. *Hydrological Data Driven Modelling* [online]. Cham: Springer International Publishing, 2015 [cit. 2021-01-05]. ISBN 978-3-319-09234-8. Dostupné z: doi:10.1007/978-3-319-09235-5
- [28] NARKHEDE, S. Understanding AUC - ROC Curve. In Machine Learning, performance. In: *Towards Data Science* [online]. 2018 [cit. 2021-01-05]. Dostupné z: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
- [29] DOBREV, D., et al. *Postoperative atrial fibrillation: mechanisms, manifestations and management*. Nature Reviews Cardiology [online]. 2019, **16**(7), 417-436 [cit. 2021-01-02]. ISSN 1759-5002. Dostupné z: doi:10.1038/s41569-019-0166-5
- [30] GANDHI, R. Support Vector Machine — Introduction to Machine Learning Algorithms. In: *Towards Data Science* [online]. 2018 [cit. 2021-01-05]. Dostupné z: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [31] ZOLTAN, C. SVM and Kernel SVM. In: *Towards Data Science* [online]. 2018 [cit. 2021-01-05]. Dostupné z: <https://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200>
- [32] STRAKA, M. Kernel Methods, SVM [online]. In: *Machine Learning for Greenhorns / ÚFAL* [online]. Praha, 9. listopadu 2020 [cit. 2021-01-05]. Dostupné z: [Machine Learning for Greenhorns | ÚFAL \(cuni.cz\)](https://machinelearningforgreenhorns.cuni.cz/)

- [33] HALTUF, Michal. *Support Vector Machines for Credit Scoring*. Praha, 2014. Magisterská práce. Vysoká škola ekonomická v Praze.
- [34] KARMAKAR, C. K., et al. *Complex Correlation Measure: a novel descriptor for Poincaré plot*. *BioMedical Engineering OnLine* [online]. 2009, **8**(1) [cit. 2021-01-03]. ISSN 1475-925X. Dostupné z: doi:10.1186/1475-925X-8-17
- [35] ZEMÁNEK, Ladislav. *Analyza variability srdečního rytmu pomocí entropie*. Brno, 2013. Magisterská práce. Vysoké učení technické v Brně.
- [36] CHEN, C., et al. *Complexity change in cardiovascular disease*. *International journal of biological sciences*, 2017, **13**(10), 1320-1328 [cit. 2021-01-05]. ISSN 1449-2288. Dostupné z: doi:10.7150/ijbs.19462
- [37] ALMASSI, G. H., et al. Predictors and impact of postoperative atrial fibrillation on patients' outcomes: A report from the Randomized On Versus Off Bypass trial. *The Journal of Thoracic and Cardiovascular Surgery* [online]. 2012, **143**(1), 93-102 [cit. 2021-01-05]. ISSN 00225223. Dostupné z: doi:10.1016/j.jtcvs.2011.10.003
- [38] Holm-Bonferroni Method. *Statistics How To: Elementary Statistics for the rest of us!* [online]. © 2020 [cit. 2021-01-05]. Dostupné z: <https://www.statisticshowto.com/holm-bonferroni-method/>
- [39] FRITZ, C. O., P. E. MORRIS a J. J. RICHLER. *Effect size estimates: Current use, calculations, and interpretation*. *Journal of Experimental Psychology: General* [online]. 2012, **141**(1), 2-18 [cit. 2021-01-05]. ISSN 1939-2222. Dostupné z: doi:10.1037/a0024338
- [40] VIDAKOVIĆ, Branislav. *Statistic for Bioengineering Science: with MATLAB and WinBUGS*. New York: Springer, 2011. Springer texts in statistics. ISBN 978-1-4614-0393-7.

Seznam obrázků

| | |
|---|----|
| Obrázek 1: Převodní systém srdeční. Převzato z [18]..... | 11 |
| Obrázek 2: Einthovenův trojúhelník a poloha elektrod hrudních unipolárních svodů. Převzato z [25]. | 12 |
| Obrázek 3: EKG křivka. Převzato z [1]..... | 13 |
| Obrázek 4: Fibrilace síní. Chaotická vlnitá aktivita nahrazující P vlny ve svodech II, III, aVF a V1 zvýrazněna červenými obdélníky. Převzato z [1]..... | 15 |
| Obrázek 5: Poincareho mapa. Převzato z [34]. | 18 |
| Obrázek 6: Hledání ideální nadroviny (optimal hyperplane) ve dvoudimenzionálním prostoru. Šířka pásma je zde značena jako maximum margin a support vectors jsou body obsahující barevnou výplň. Převzato z [30]. | 27 |
| Obrázek 7: Soft margin SVM případ s lineárně neseperabilními daty. Převzato z [33]. | 28 |
| Obrázek 8: Postup náhodného výběru dat metody Monte Carlo při 3 iteracích. Převzato z [27]. | 29 |
| Obrázek 9: Extrémní případy ROC křivek. Převzato z [28]..... | 30 |
| Obrázek 10: Korelační mapa mezi všemi parametry datasetu. V rozmezí hodnot x 1-8 najdeme parametry EKG, 9-21 obecnou charakteristiku pacienta, 22-28 časovou oblast HRV analýzy, 29-55 frekvenční oblast HRV analýzy a 56-91 nelineární oblast HRV analýzy. | 33 |
| Obrázek 11: Normální rozdělení skupin parametru SD2 s vyznačenými středními hodnotami a Cohenovou velikostí efektu..... | 34 |
| Obrázek 12: Výstupní křivky GLM modelu typu logistické regrese (funkce stepwiseglm) obsahující parametry CRP, RRTriIndex, SD2 a SampEn vztažené ke Skupině..... | 36 |
| Obrázek 13: Krabicové diagramy signifikantních parametrů používaných v obou výstupních modelech | 37 |
| Obrázek 14: Výsledná ROC křivka GLM modelu..... | 38 |
| Obrázek 15: Výsledná ROC křivka SVM modelu..... | 39 |
| Obrázek 16: Porovnání výkonu modelů SVM a GLM na nových datech..... | 40 |
| Obrázek 17: Základní charakteristika důležitých parametrů nových dat. Hodnoty PXX označují číslo pacienta, který byl stanoven jako odlehlá hodnota a byl z testování vyřazen..... | 42 |
| Obrázek 18: Porovnání modelů GLM a SVM na nových datech s odstraněnými odlehlými hodnotami | 42 |
| Obrázek 19: Porovnání dvou prahů jedné náhodně vybrané iterace SVM..... | 43 |

Seznam tabulek

| | |
|---|----|
| Tabulka 1: Parametry analýzy variability srdeční frekvence v časové, frekvenční a nelineární oblasti. Převzato a upraveno z [21][24]. | 15 |
| Tabulka 2: Přehled parametrů získaných ze záznamu EKG a jejich středních hodnot (mean) se směrodatnou odchylkou (std) a mediánem. Hodnota 0 u blokády Tawarova raménka značí nepřítomnost blokády. | 22 |
| Tabulka 3: Tabulka výsledků neparametrického Wilcoxon-Mann-Whitneyho testu před Holm-Bonferroniho korekcí | 35 |
| Tabulka 4: Estimační koeficienty modelu GLM | 36 |
| Tabulka 5: Úspěšnost nastaveného prahu predikce u modelu SVM a GLM..... | 40 |
| Tabulka 6: Vybrání subjekti zastupující ukazatele TP, TN, FN a FP..... | 41 |
| Tabulka 7: Porovnání důležitých parametrů základních dat a dat zaslepené studie. Mean (střední hodnota), std (směrodatná odchylka). | 41 |
| Tabulka 8: Porovnání výsledků 2 různých prahových hodnot u jedné náhodně vybrané iterace modelu SVM | 43 |

Příloha A

Seznam parametrů v korelační mapě:

| EKG parametry | | Frekvenční parametry HRV analýzy | | Nelineární parametry HRV analýzy | |
|--|------------------|----------------------------------|--------------|----------------------------------|------------|
| 1 | PQ | 29 | PeakVLFAR | 56 | SD1 |
| 2 | QRS | 30 | PeakLFR | 57 | SD2 |
| 3 | QT | 31 | PeakHFAR | 58 | SD1_SD2 |
| 4 | QTc | 32 | AbsPowVLFAR | 59 | ApEn |
| 5 | P | 33 | AbsPowLFAR | 60 | PortaIndex |
| 6 | OsaQRS | 34 | AbsPowHFAR | 61 | GuzikIndex |
| 7 | OsaP | 35 | RelPowVLFAR | 62 | SlopeIndex |
| 8 | Frekvence | 36 | RelPowLFAR | 63 | SampEn |
| Obecná charakteristika pacienta | | 37 | RelPowHFAR | 64 | D2 |
| 9 | vek | 38 | NormPowLFAR | 65 | Alpha1 |
| 10 | hmotnost | 39 | NormPowHFAR | 66 | Alpha2 |
| 11 | BMI | 40 | PeakVLFFFT | 67 | Lmean |
| 12 | EjekcniFrakce | 41 | PeakLFFFT | 68 | Lmax |
| 13 | LVED | 42 | PeakHFFFT | 69 | REC |
| 14 | LS_LAD_ | 43 | AbsPowVLFFFT | 70 | DET |
| 15 | Hemoglobin | 44 | AbsPowLFFFT | 71 | ShannonEn |
| 16 | Leukocyty | 45 | AbsPowHFFFT | 72 | MSE1 |
| 17 | Na | 46 | RelPowVLFFFT | 73 | MSE2 |
| 18 | K | 47 | RelPowLFFFT | 74 | MSE3 |
| 19 | Kreatinin | 48 | RelPowHFFFT | 75 | MSE4 |
| 20 | Urea | 49 | NormPowLFFFT | 76 | MSE5 |
| 21 | CRP | 50 | NormPowHFFFT | 77 | MSE6 |
| Časová oblast HRV analýzy | | 51 | TotPowAR | 78 | MSE7 |
| 22 | MeanRR_ms_ | 52 | TotPowFFT | 79 | MSE8 |
| 23 | MeanHR_tepu_min_ | 53 | LF_HFAR | 80 | MSE9 |
| 24 | RMSSD | 54 | LF_HFFFT | 81 | MSE10 |
| 25 | NN50 | 55 | EDR | 82 | MSE11 |
| 26 | pNN50 | | | 83 | MSE12 |
| 27 | RRTriIndex | | | 84 | MSE13 |
| 28 | TINN | | | 85 | MSE14 |
| | | | | 86 | MSE15 |
| | | | | 87 | MSE16 |
| | | | | 88 | MSE17 |
| | | | | 89 | MSE18 |
| | | | | 90 | MSE19 |
| | | | | 91 | MSE20 |

Příloha B

Obsah CD:

- Data.csv – soubor obsahující kompletní data APOFIS (anonymní)
- Data_ekg_cast2CSV.csv – soubor obsahující základní dataset s EKG parametry
- ROC_mean.m – GLM, testování normality, univariantní statistika, korelace
- SVM_pokus.m – SVM model
- Porovnaní_zaslepena.m – porovnání modelů na zaslepených datech
- adjust_unique_pointss.m – funkce na průměrování ROC křivky
- train_test_indices.m – funkce na křížovou validaci typu Monte Carlo