



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

EKG BIOMARKERY PRO ODHAD RIZIKA PŘÍDATNÉ SÍŇOKOMOROVÉ DRÁHY Z POHLEDU NÁHLÉ SRDEČNÍ SMRTI.

ECG-BASED BIOMARKERS FOR ESTIMATION OF SUDDEN DEATH RISK IN PATIENTS WITH ACCESSORY
ATRIOVENTRICULAR PATHWAY.

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Lucia Tomková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jakub Hejč

BRNO 2023

Bakalářská práce

bakalářský studijní program **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Lucia Tomková

ID: 231157

Ročník: 3

Akademický rok: 2022/23

NÁZEV TÉMATU:

EKG biomarkery pro odhad rizika přidatné síňokomorové dráhy z pohledu náhlé srdeční smrti.

POKYNY PRO VYPRACOVÁNÍ:

- 1) Seznamte se s projevy a riziky přidatné síňokomorové dráhy u dětských pacientů. Dále se seznamte s metodami pro extrakci příznaků z EKG záznamu a sestavte literární rešerši z obou oblastí.
- 2) V dostupných datech z elektrofyziologického vyšetření stanovte vhodné úseky EKG záznamu a další klinické parametry vhodné pro analýzu rizika přidatné dráhy.
- 3) V prostředí Python implementujte vybrané algoritmy pro extrakci příznaků z EKG záznamu.
- 4) Proveďte statistickou analýzu vybraných biomarkerů a implementujte vhodný model pro predikci rizika přidatné dráhy.
- 5) Proveďte diskuzi dosažených výsledků.

DOPORUČENÁ LITERATURA:

- [1] Katoh T, Ohara T, Kim EM, Hayakawa H. Non-invasive diagnosis of concealed Wolff-Parkinson-White syndrome by detection of concealed anterograde pre-excitation. *Jpn Circ J.* 2001 May;65(5):367-70. doi: 10.1253/jcj.65.367.
- [2] Li HY, Chang SL, Chuang CH, Lin MC, Lin YJ, Lo LW, Hu YF, Chung FP, Chang YT, Chung CM, Chen SA, Lee PC. A Novel and Simple Algorithm Using Surface Electrocardiogram That Localizes Accessory Conduction Pathway in Wolff-Parkinson-White Syndrome in Pediatric Patients. *Acta Cardiol Sin.* 2019 Sep;35(5):493-500. doi: 10.6515/ACS.201909_35(5).20190312A.

Termín zadání: 6.2.2023

Termín odevzdání: 29.5.2023

Vedoucí práce: Ing. Jakub Hejč

doc. Ing. Jana Kolářová, Ph.D.
předseda rady studijního programu

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

Cieľom tejto práce je zoznámiť čitateľa s prejavom delta vlny na elektrokardiograme u detských pacientov s Wolf-Parkinson-White syndrómom, ktorí sú ohrození náhlou smrťou srdca. Následne sa v práci venujeme spracovaniu dát z elektrofyziologického vyšetrenia a tvorbe algoritmov na extrakciu biomarkerov z EKG signálu v programovacom jazyku Python. Práca obsahuje štatistickú analýzu biomarkerov, pričom sa sústreďuje na rozdiel v datasetoch u rizikových a nerizikových pacientov. Biomarkery, ktoré majú štatisticky významný rozdiel v datasetoch sú použité k tvorbe a výberu vhodného modelu predikcie rizika akcesórnej dráhy u pacientov.

KLÚČOVÉ SLOVÁ

WPW syndróm, preexcitačný syndróm, akcesórna dráha, EKG, delta vlna, extrakcia biomarkerov

ABSTRACT

The aim of this work is to acquaint the reader with the delta wave manifestation on the electrocardiogram in pediatric patients with Wolf-Parkinson-White syndrome who are at risk of sudden cardiac death. Subsequently, in the work we deal with the processing of data from the electrophysiological examination and the creation of algorithms for the extraction of biomarkers from the ECG signal in the Python programming language. The work includes statistical analysis of biomarkers, focusing on the difference in datasets for at-risk and non-at-risk patients. Biomarkers that have a statistically significant difference in the datasets are used to create and select a suitable model for predicting the risk of the accessory pathway in patients.

KEYWORDS

WPW syndrome, pre-excitation syndrome, accessory pathway, ECG, delta wave, extraction of biomarkers

TOMKOVÁ, Lucia. *EKG biomarkery pro odhad rizika přídatné síňokomorové dráhy z pohledu náhlé srdeční smrti.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2023, 48 s. Bakalářská práce. Vedúci práce: Ing. Jakub Hejč

Vyhlásenie autora o pôvodnosti diela

Meno a priezvisko autora: Lucia Tomková
VUT ID autora: 231157
Typ práce: Bakalárska práca
Akademický rok: 2022/23
Téma záverečnej práce: EKG biomarkery pro odhad rizika přídavné síňokomorové dráhy z pohledu náhlé srdeční smrti.

Vyhlasujem, že svoju záverečnú prácu som vypracovala samostatne pod vedením vedúcej/cého záverečnej práce, s využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autorka uvedenej záverečnej práce ďalej vyhlasujem, že v súvislosti s vytvorením tejto záverečnej práce som neporušila autorské práva tretích osôb, najmä som nezasiahla nedovoleným spôsobom do cudzích autorských práv osobnostných a/alebo majetkových a som si plne vedomá následkov porušenia ustanovenia § 11 a nasledujúcich autorského zákona Českej republiky č. 121/2000 Sb., o práve autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov (autorský zákon), v znení neskorších predpisov, vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákonníka Českej republiky č. 40/2009 Sb.

Brno

.....

podpis autorky*

*Autor podpisuje iba v tlačenej verzii.

POĎAKOVANIE

Rada by som poďakovala vedúcemu bakalárskej práce pánovi Ing. Jakubovi Hejčovi za odborné vedenie, konzultácie, trpezlivosť a podnetné návrhy k práci.

Obsah

Úvod	10
1 Klinická časť	11
1.1 Prevodový systém srdca	11
1.2 Akčné napätie bunky	11
1.3 Popis EKG	12
1.4 Patofyziológia WPW syndrómu	13
1.5 Príznaky WPW syndrómu	14
1.6 Arytmie spôsobené WPW syndrómom	14
2 Metódy	15
2.1 Pacienti	15
2.2 Meranie dát	17
2.3 Zaznamenávanie dát	18
2.4 Detekcia QRS komplexov	18
2.5 Korelácia signálov	19
2.6 Extrakcia biomarkerov	21
2.7 Agregácia extrahovaných biomarkerov	23
2.8 Štatistické testovanie významnosti biomarkerov	24
2.8.1 Test normálneho rozloženia	24
2.8.2 Test štatisticky významnej odchýlky	24
2.9 Korelácie biomarkerov	25
2.10 Model predikcie	26
2.10.1 Klasifikačný model	27
2.10.2 Hodnotenie kvality modelu	27
2.10.3 Model rozhodovacieho stromu	28
2.10.4 Model logistickej regresie	29
2.10.5 Model SVM	29
2.10.6 Metóda bagging	31
3 Výsledky	32
3.1 Normálne rozloženie dát	32
3.2 Štatisticky významná odchýlka	32
3.3 Porovnanie výsledkov modelov predikcie	33
4 Diskusia	35
Záver	37

Literatúra	38
Zoznam symbolov a skratiek	43
Zoznam príloh	44
A Výsledky	45
A.1 Porovnanie rozloženia dát biomarkerov pre rizikových a nerizikových pacientov	45
B Obsah elektronickej prílohy	48

Zoznam obrázkov

1.1	Fázy akčného napätia myokardu a efektívna refraktérna perióda	12
1.2	QRS komplex s delta vlnou	13
2.1	Vývojový diagram selekcie pacientov	16
2.2	Škatuľové grafy pre hodnoty ERP AP	17
2.3	Postup korelácie signálov	20
2.4	Ukážka výstupu po kumulácii signálov	21
2.5	Znázornenie jednotlivých biomarkerov	22
2.6	Znázornenie jednotlivých biomarkerov	22
2.7	Korelačná matica signifikantných biomarkerov	26
2.8	Schématické znázornenie lineárneho modelu SVM	30
2.9	Schématické znázornenie metódy bagging	31
3.1	Matice zámeny modelov	34

Úvod

Wolf-Parkinson-White syndróm (WPW syndróm) je vrodeným ochorením srdca, pri ktorom spojenie medzi predsieňami a komorami nezabezpečuje len atrioventrikulárny uzol, ale aj akcesórna dráha. Táto dráha spôsobuje predčasnú elektrickú aktivitu komôr s obrazom preexcitácie na elektrokardiograme (EKG). V najhorších prípadoch môže dôjsť k tvorbe arytmií, ktoré spôsobia náhlu smrť srdca. Charakteristickým príznakom tohto syndrómu je nález delta vlny na EKG. Vzhľadom na to, že mnoho pacientov je asymptomatických, ich preventívne sledovanie a prípadná liečba nie sú dostatočne skoro začaté. Frekvencia výskytu WPW syndrómu u ľudí je 0,1% - 0,3%. Znamená to, že každý prvý až tretí človek z 1000 ľudí trpí WPW syndrómom [1].

Hlavným cieľom tejto práce je spracovanie signálov od pacientov s WPW syndrómom a implementovanie vhodných algoritmov v prostredí Python, ktoré budú použité pre extrakciu príznakov z EKG záznamu. Následné predspracovanie dát potrebných k implementácii modelu predikcie rizika akcesórnej dráhy z pohľadu náhlejšej srdcovej smrti.

1 Klinická časť

1.1 Prevodový systém srdca

Prevodový systém srdca je tvorený špecializovanými myocytmi, ktoré majú svoje jedinečné anatomické, funkčné a molekulárne vlastnosti, vďaka ktorým naše srdce dokáže pracovať ako samostatná jednotka. Systém je tvorený dvoma uzlami, sinoatriálnym (SA uzol) a atrioventrikulárnym (AV uzol), ďalej je tvorený pravou a ľavou vetvou zväzkov, ktoré vedú elektrický impulz až ku konečným Purkyňovým vláknam [2].

Sinoatriálny uzol má tvar podobný "banánu"[2] a nachádza sa v subepikardiálnej oblasti srdca. Pôsobí ako primárny pacemaker, ktorý automaticky udáva rytmus. Riadi sa spontánnou diastolickou depolarizáciou, ktorá po dosiahnutí určitého prahu spustí ďalšie akčné napätie [2].

Vzniknutý impulz sa vďaka preferenčným dráham dostáva do atrioventrikulárneho uzla, ktorý sa nachádza na dne predsieňového septa. Jeho hlavnou úlohou je viesť akčné napätie z predsiení na komory, pričom musí tento impulz spomaliť, aby vzniklo oneskorenie medzi predsieňovou a komorovou systolou, pretože je dôležité, aby sa najprv dokončila depolarizácia predsiení a až tak bola zahájená kontrakcia komôr. Zároveň AV uzol chráni komory pred predsieňovou tachykardiou a v prípade zlyhania SA uzlu môže AV uzol prevziať úlohu pacemekera [2].

Hisov zväzok prestupuje medzi väzivový skelet do interventrikulárneho septa a tvorí jedinou možnú elektrickú dráhu pre impulz z AV uzla na komory. Myokard predsiení a komôr je elektricky izolovaný pomocou väzivového skeletu. Systém vedenia od AV uzlu až po komorový pracovný myokard zabezpečuje synchronizáciu aktivácie komorového myokardu tak, aby sa zabezpečila čo najefektívnejšia aktivita srdca ako pumpy [2].

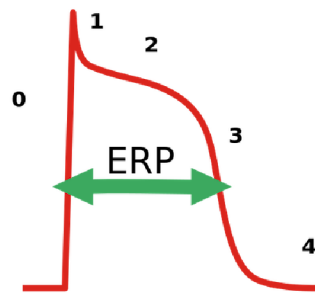
Poslednou časťou prevodového systému srdca sú Tawarové ramienka, na ktoré nadväzujú Purkyňove vlákna. Ich úlohou je previesť vzruch cez pravú a ľavú komoru až na pracovný myokard komôr [2].

1.2 Akčné napätie bunky

Akčné napätie bunky myokardu je možné merať pomocou intracelulárnej elektródy zavedenej priamo do srdca ku konkrétnej bunke. Akčné napätie je tvorené 5 základnými fázami, ktoré popisujú jeho vznik a priebeh v rámci jednej bunky. Fáza 0 popisuje nástup rýchlej depolarizácie, kedy sa aktivujú sodíkové kanály a zvýši sa vodivosť membrány. Fáza 1 nastáva vtedy, keď sa otvoria draslíkové kanály a nastane rýchly pokles napätia v bunke. Vo fáze 2 alebo inak nazývanej aj plató sa vyrovná

podiel vápnikových a draslíkových iónov a membránový potenciál je takmer konštantný. Fáza 3 zahrňuje repolarizáciu bunky, kedy sa uzatvoria vápnikové kanály, avšak draslíkové kanály ostanú otvorené až do dosiahnutia pokojového membránového potenciálu okolo -90 mV. Štvrtá fáza nastáva počas diastoly bunky, kedy sú uzatvorené sodíkové aj draslíkové kanály a akčné napätie má stabilnú hodnotu okolo -90 mV [3].

Efektívna refraktérna perióda (ERP) je doba po vyvolaní akčného napätia, počas ktorej nemôže dôjsť k vyvolaniu ďalšieho akčného napätia bez ohľadu na to, aký veľký impulz je vyvolaný, keďže membrána nie je dostatočne repolarizovaná a sodíkové kanály sú ešte inaktivované. Na obrázku 2.2 je znázornený jeden priebeh akčného napätia v kardiomyocytoch a názorne znázornená aj ERP [4].



Obr. 1.1: Fázy akčného potenciálu srdca a efektívna refraktérna perióda [8]

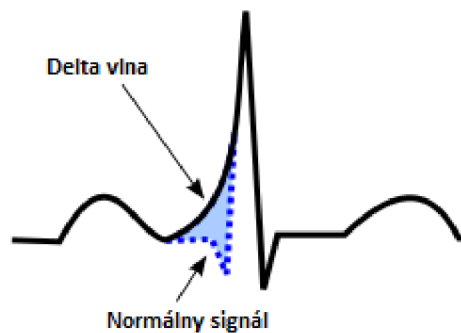
1.3 Popis EKG

Najčastejšie používaná vyšetrovacia metóda srdca 12-zvodový elektrokardiogram. V bunkách myokardu vzniká akčný potenciál, ktorý sa postupne depolarizáciou bunky po bunke šíri od sinoatriálneho uzlu až po Purkyňove vlákna. V oblastiach rozdielneho membránového napätia vznikajú malé prúdy, ktoré vytvárajú elektromagnetické pole. Vďaka vodivosti telesných tekutín môžeme zachytávať tieto zmeny potenciálu na elektródach na povrchu tela. Výsledkom tohto merania je EKG krivka, pomocou ktorej môžeme jednoducho a neinvazívne rozlíšiť niektoré typické ochorenia srdca.

EKG krivka sa skladá z jednotlivých QRS komplexov, ktoré nám zobrazujú priebeh srdca. QRS komplex sa skladá z 5 základných vln. Fyziologicky by sa pred každým QRS komplexom mala nachádzať P vlna, ktorá signalizuje depolarizáciu predsiení, pričom prvá časť signalizuje depolarizáciu pravej predsienie a druhá časť zasa ľavej predsienie. Nasleduje interval PQ (od začiatku vlny P až po začiatok vlny Q), ktorý predstavuje dobu, za akú sa signál dostane zo sinoatriálneho uzla až

na prevodový systém komôr. Nasleduje samotný komplex QRS, ktorý vo všeobecnosti opisuje depolarizáciu komôr a tvoria ho dve negatívne vlny Q a S a jedna pozitívna vlna R. Poslednou významnou vlnou na EKG je vlna T, ktorá reprezentuje repolarizáciu komôr. Predsiene sú samozrejme taktiež repolarizované, avšak táto aktivita nie je na EKG viditeľná, keďže ju prebýja depolarizácia komôr [5] [6].

Na EKG môžeme pozorovať isté patofyziologické zmeny EKG krivky, ktoré naznačujú rôzne choroby srdca. Jednou z nich je aj delta vlna, ako môžeme vidieť na obrázku 1.2 a ktorá je typickým príznakom pri WPW syndróme. Síce sa nám zobrazí klasický sínusový rytmus, avšak interval PQ, ktorý má bežne dĺžku 0,12 - 0,2 sekundy, je skrátený pod 0,12 sekundy, čo je spôsobené práve predčasnou depolarizáciou komôr. Práve delta vlna skracuje PQ interval a zároveň rozširuje spodnú časť QRS komplexu, a teda má QRS komplex dĺžku viac ako 0,1 sekundy. Druhá polovica QRS komplexu je normálna, keďže impulz, ktorý sa na komory dostal cez Kentov zväzok sa stretne s vlnou depolarizácie, ktorý prešiel cez AV uzol [7].



Obr. 1.2: QRS komplex s delta vlnou [9]

1.4 Patofyziológia WPW syndrómu

Wolf-Parkinson-White syndróm je vrodený preexcitečný syndróm s predčasnou depolarizáciou časti komôr cez akcesórnú dráhu (Kentov zväzok) v prevodovom systéme srdca. Ľudia s týmto syndrómom sa narodia s pridanou dráhou v srdci, ktorá umožní elektrickému signálu obísť AV uzol a depolarizuje tým komory skôr (preexcitácia), poprípade sa vzruch môže vrátiť pomocou dráhy naspäť z komory na predsieň. Keďže spojenie medzi predsieňou a komorou je pomocou dvoch zväzkov (AV uzol a Kentov zväzok), môže dochádzať k rôznym poruchám rytmu srdca. Elektrické impulzy, ktoré prejdú cez akcesórnú dráhu vedú k abnormálnej supraventrikulárnej tachykardii [7] [1].

1.5 Príznaky WPW syndrómu

Wolf-Parkinson-White syndróm je úkaz na EKG, ktorý je jedným z prejavov atri-oventrikulárnej recipročnej tachykardie (AVRT). Medzi najčastejšie príznaky AVRT patria synkopy (náhla a krátkodobá strata vedomia), palpácia (búšenie srdca), dyspnoe (pocit nedostatku vzduchu) a pocit chvenia. Hlavne u detí môže mať vyvolaná tachykardia fatálne následky, keďže srdcový rytmus dosiahne hodnoty nad 220 tepov za minútu a vedie k zástave srdca a náhlej smrti[10]. Podľa predchádzajúcich štúdií sa náhla smrť pri WPW syndróme vyskytuje aj bez predchádzajúcich príznakov, častokrát v prípade, kedy sú pacienti v klude, poprípade v spánku a zároveň zo štúdií vychádza, že častejšie sú ohrození muži ako ženy [11].

1.6 Arytmie spôsobené WPW syndrómom

Aksesórna dráha Kentovho zväzku pri WPW-syndróme je častou príčinou vzniku arytmií. U pacientov s týmto syndrómom sa s 95% pravdepodobnosťou vyskytuje ortodrómná AV reentry tachykardia (AVRT). Pri ortodrómej AVRT sa vzruch šíri anterogádne z predsiení na komory cez AV uzol, avšak retrogádne sa vracia na predsieň cez aksesórnu dráhu, čím vzniká krúživý vzruch. Jedná sa o pravidelnú tachykardiu s úzkym komplexom QRS a s P vlnou, ktorá sa nachádza až na počiatku ST úseku. Delta vlna v takomto prípade chýba a QRS komplex je štíhly [7] [12].

V opačnom prípade, keď antegrádne vedenie vzruchu je cez Kentov zväzok na komoru a ten sa nasledovne vracia retrogádne na predsieň cez AV uzol, hovoríme o antridromnej AV reentry tachykardii. EKG obsahuje typickú delta vlnu, skrátený PQ interval a rozšírený QRS komplex [7] [12].

Okrem tachykardií môže Wolf-Parkinson-White syndróm spôsobiť aj fibriláciu predsiení. Vzniká dvojité vedenie cez AV uzol a zároveň aj cez pridanú dráhu, čo sa prejavuje na EKG premenlivými RR intervalmi a výraznou delta vlnou. Ide o najnebezpečnejšiu formu arytmií, kedy hrozí prechod fibrilácie z predsiení aj na komory, čo môže spôsobiť náhlu smrť srdca [7].

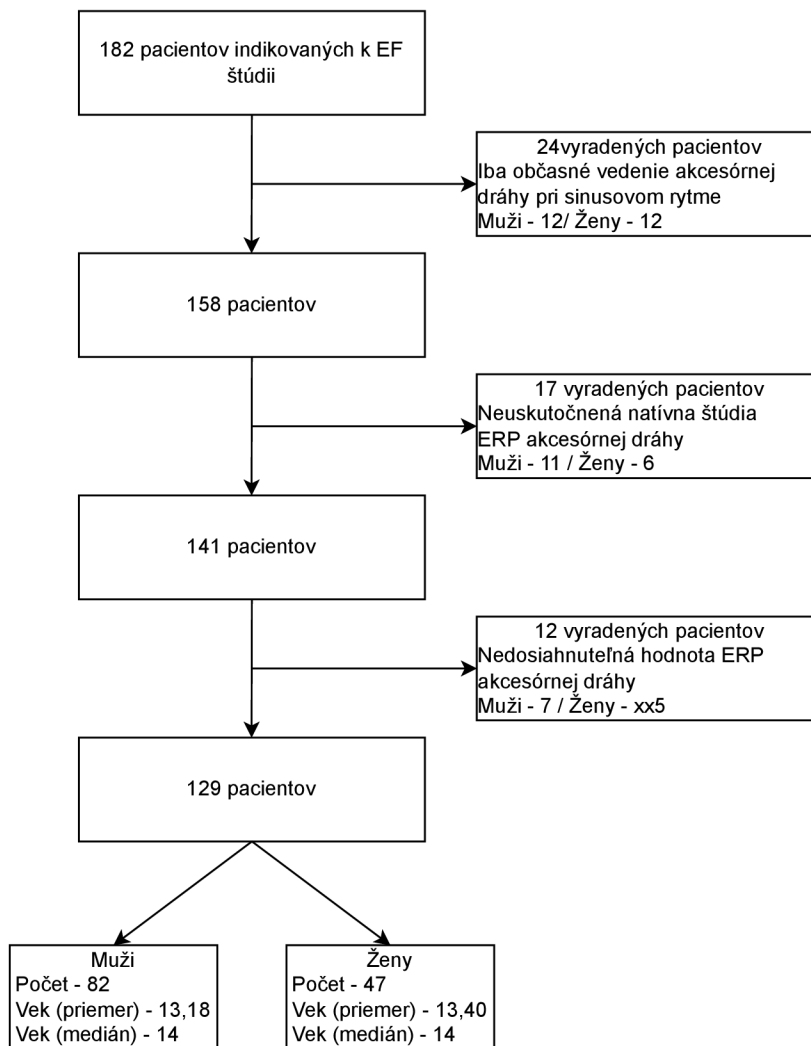
2 Metódy

2.1 Pacienti

Jedná sa o retrospektívnu štúdiu, v ktorej bolo zahrnutých 182 mladých pacientov vo veku od 5 do 19 rokov indikovaných k elektrofyziologickej štúdii (EF) pre obraz preexcitácie na 12-zvodovom EKG v Detskej nemocnici Fakultnej nemocnice Brno v rokoch 2012 až 2022. Počas EF štúdie bola pacientom meraná ERP akcesórnej dráhy v 4 rôznych situáciách. Anterográdna efektívna refrakterná perióda bola indikovaná pri doprednom vedení dráhy, kedy sa impulz šíril z predsieni na komoru. Retrográdna ERP bola indikovaná pri spätnom vedení AP, pričom sa impulz šíril presne opačným smerom ako pri anterográdnom vedení, a teda z komory na predsieň. U niektorých pacientov boli merané tieto hodnoty aj po podaní izoproterenolu. Izoprotenerol je beta-1 a beta-2 adrenergný receptorový agonista, ktorý po aplikácii zvýši kontraktilitu a frekvenciu srdca, a zároveň zvýši rýchlosť vedenia vzruchu v prevodovom systéme srdca [13].

Získané dáta boli pre každého pacienta preštudované a z dát EF vyšetrenia boli vybrané hodnoty nameraných ERP, ktoré boli následne uložené do tabuľky pre jednoduchší prehľad. V práci bolo určených niekoľko kritérií, ktoré pacienti museli spĺňať a podľa ktorých boli pacienti roztriedení do skupín. 24 pacientov (z toho 12 mužov a 12 žien) malo iba intermitentne (občasne) vedúcu prídavnú dráhu pri sínusovom rytme, a preto boli zo štúdie vyradení. Ďalších 17 pacientov (z toho 11 mužov a 6 žien) bolo zo štúdie vyradených, pretože pri nich nebola vykonaná natívna štúdia efektívnej refraktérnej periódy žiadnym z obvyklých postupov. Medzi obvyklé metódy sa zaraďuje vzostupná stimulácia predsiení, S1-S2 programovaná stimulácie predsiení a meranie najkratšieho R-R intervalu pri indukovanej fibrilácii predsiení. Posledná skupina 12 vyradených pacientov (z toho 7 mužov a 5 žien) nemali dosiahnuteľnú hodnotu ERP akcesórnej dráhy ani jedným z obvyklých spôsobov.

Z pôvodných 182 pacientov iba 129 spĺňalo všetky kritéria zaradenia do práce. 82 mužov s priemerným vekom 13,18 rokov a mediánom 14 rokov, čo predstavuje 63,6% všetkých pacientov. Zvyšných 36,4% pacientov tvorilo 47 žien s priemerným vekom 13,40 rokov a mediánom 14 rokov. Postup selekcie pacientov je zobrazený na obrázku 2.1.



Obr. 2.1: Vývojový diagram selekcie pacientov

Z tabuľky 2.1 je zrejmé, že najčastejšie bola nameraná anterográdna ERP AP, ktorá bola indikovaná u všetkých zaradených pacientov. Práve táto hodnota v práci slúži na klasifikáciu rizikových a nerizikových pacientov. Zároveň sú v tabuľke uvedené priemerné hodnoty ERP AP spolu s vypočítanou smerodajnou odchýlkou, ktorá bola vypočítaná podľa vzorca 2.1.

$$\delta = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (2.1)$$

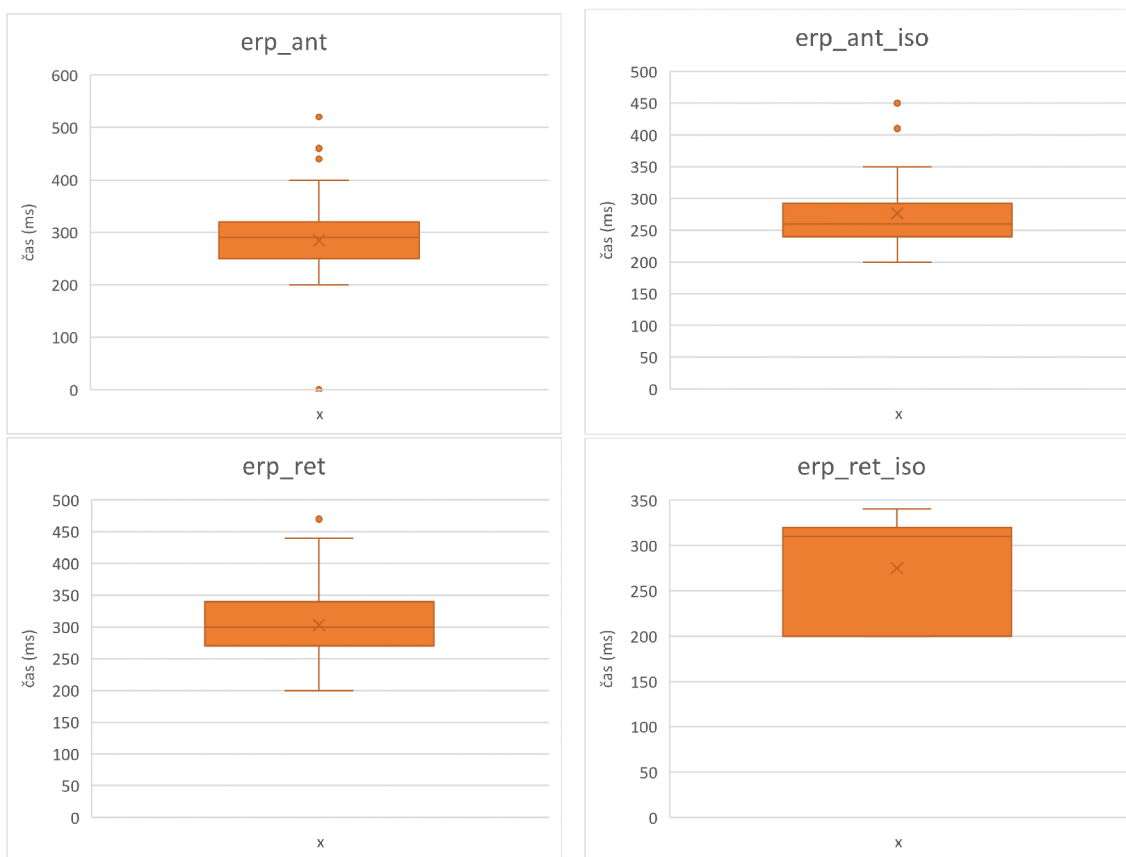
kde N je počet prvkov, x_i je hodnota na pozícii i vo vektore x a \bar{x} je priemerná hodnota vektoru x [14].

Škatulový graf (boxplot) zobrazuje rozdelenie hodnôt anterográdnej ERP akcesórnej dráhy pre všetkých pacientov zaradených do práce. Z grafov je zrejmé, že väčšina hodnôt sa pohybuje v okolí 300 ms. Pre ostatné typy ERP AP sú škatulové

	Ant ERP AP	Ant ERP AP s iso	Ret ERP AP	Ret ERP AP s iso
Počet	128	30	59	8
\bar{x}	286	277	303	275
δ	53.7	51.8	55.8	58.9

Tab. 2.1: Hodnoty ERP AP (ant - anterogádne, ret - retrogádne). \bar{x} - priemer a δ - smerodajná odchýlka

grafy vykreslené bez NaN hodnôt, pretože tie by značne ovplyvnili výsledný graf 2.2.



Obr. 2.2: Hodnoty anterogádnej (ant) a retrogádnej (ret) ERP akcesórnej dráhy pre všetkých pacientov zobrazené v škatuľovom grafe.

2.2 Meranie dát

Podstúpenie elektrofyziologického vyšetrenia je najdôležitejším krokom k určení rizika náhlej smrti a k vyhodnoteniu povahy biomarkerov, keďže neinvasívne štúdie

síce dokážu určiť polohu akcesórnej dráhy, ale zatiaľ pomocou nich nedokážeme predpovedať riziko náhlej smrti srdca [15].

Pri EF vyšetrení sa pacientovi zavedú elektródy do pravej časti srdca a ešte pred samotnou stimuláciou sa nameria natívne EKG. Začína sa so stimuláciou z predsiení na komory, aby sa zistilo, ako vyzerá vedenie signálu. Prvá stimulácia je uskutočnená metódou vzostupnej sériovej stimulácie predsiení, kedy sa začínajú stimulovať komory na základnom stimulačnom cykle, ktorý má frekvenciu vyššiu ako je frekvencia pacienta, pričom sa obvykle začína na hodnote okolo 100 tepov za minútu. Táto frekvencia sa postupne zvyšuje a sleduje sa odpoveď na komorách. Môžu nastať dva prípady, vďaka ktorým vieme určiť ERP na základe aktuálnej frekvencie stimulácie. V prípade, že vedenie prestane byť 1:1 (1 predsieň a 1 komora) znamená, že vypadla aj akcesórna dráha a aj AV uzol alebo môže nastať prípad, že sa zoštíhli QRS komplex na EKG, čo značí, že vypadla iba akcesórna dráha [15].

Ďalšou metódou, ktorá sa používa pri elektrofyziológii srdca je programovateľná stimulácia predsiení, ktorá sa nazýva aj ako protokol S1-S2. Funguje na princípe, že sa pacientovi stimuluje predsieň presne daným počtom stimulov pri napríklad 100 milisekundách a následne sa podá jeden extrastimul pri 400 milisekundách. Takto sa opakuje cyklus a postupne sa znižuje hodnota extrastimulu. Sleduje sa, kedy sa zastaví prevod z predsiení na komory (ERP AV uzlu alebo akcesórnej dráhy) alebo kedy zmizne delta vlna (ERP AP). Protokol S1-S2 sa znova opakuje, avšak s tým, že sa stimulujú komory a sleduje sa prevod signálu z komôr na predsieň. Takýmto spôsobom zisťujeme retrográdne ERP [15].

V prípade, že bol pacientovi podaný izoproterenol (iso), ktorý zvyšuje excitabilitu buniek srdca, opakuje sa celý postup znovu a merajú sa hodnoty ERP pri iso.

2.3 Zaznamenávanie dát

Na zaznamenávanie dát sa používa systém Workmate firmy Abbott, pričom sa zaznamenávajú hodnoty z 12-zvodového EKG a intrakardiálne záznamy z koronárneho sínusu a z Hisovho zväzku. Vzorkovacia frekvencia sa obvykle nastavuje na 2 kHz a rozlíšenie signálu je 78(μ V/lb). Tieto dáta sú poskytnuté z Dětské nemocnice FN Brno.

2.4 Detekcia QRS komplexov

Na spracovanie EKG signálu bol použitý program SignalPlant z Ústavu prístrojovej techniky Akadémie vied Českej republiky [16]. Tento program sa špecificky zameriava na biologické signály, avšak obsahuje nástroje užitočné pre akúkoľvek inú oblasť

spracovania signálu. Všetky EKG záznamy boli uložené v súbore typu .h5, ktorý je jedným z hierarchických dátových formátov, ktorý sa používa na veľké množstvo údajov vo forme viacrozmerných polí [17].

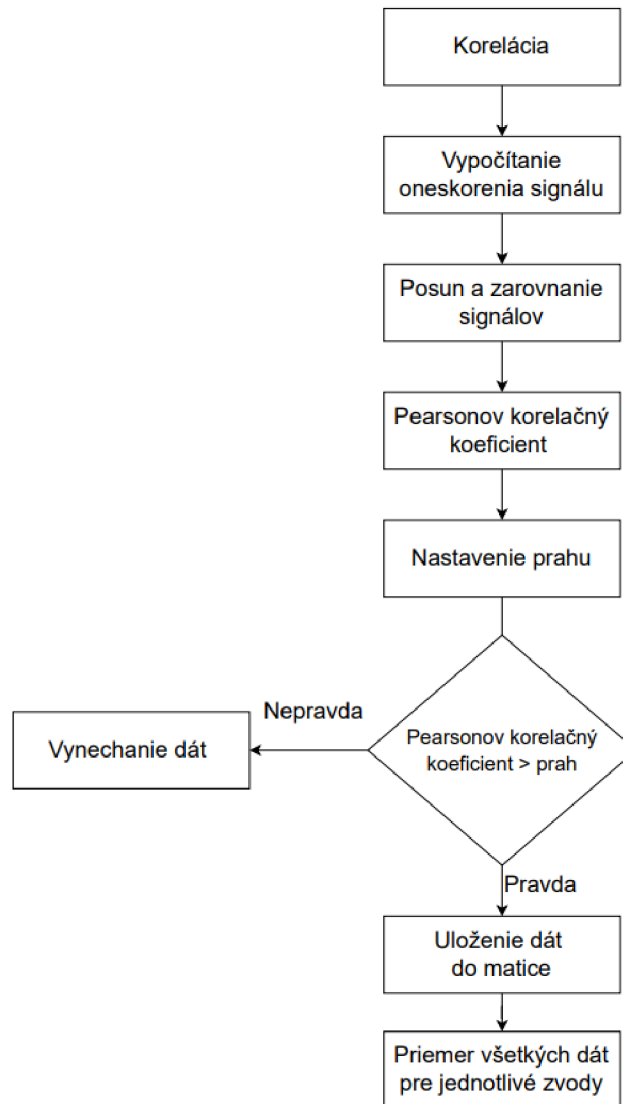
V rámci tejto bakalárskej práce bol najviac využitý softwarový modul QRS detektor, ktorý v celom signáli označil každý QRS komplex [18]. Model bol upravený pre potreby tejto bakalárskej práce rozšírením detekcie cez všetkých 12 zvodov. Následne boli tieto značky dávkovo zo všetkých signálov exportované do .sel súborov. Keďže program SignalPlant, označí iba stred QRS komplexu a na správnu lokalizáciu celého QRS komplexu a úseku P-Q je potrebné označiť väčšiu časť, boli preto všetky značky pomocou vytvoreného programu rozšírené ešte o 350 vzoriek dolava a 350 vzoriek doprava.

2.5 Korelácia signálov

Korelácia je štatistická metóda, ktorá sa používa na posúdenie linearity medzi dvoma spojitými premennými. Korelačný koeficient môže nadobúdať hodnoty od -1 do +1. Hodnoty -1 a +1 znamenajú úplnú koreláciu, zatiaľ čo nula znamená, že medzi premennými neexistuje žiaden lineárny vzťah [19]. V tejto práci bola korelácia signálov využívaná na porovnávanie jednotlivých QRS komplexov. Pri každom pacientovi bol zvolený jeden časový úsek s QRS komplexom, ktorý obsahoval delta vlnu a jeden časový úsek s QRS komplexom, ktorý už delta vlnu neobsahoval a bol zaznamenaný na konci elektrofyzikálneho vyšetrenia. Takýmto spôsobom boli určené vzory, voči ktorým prebiehala korelácia. Hlavným dôvodom, kvôli ktorému sme sa rozhodli takto postupovať, bolo znížiť výpočtovú náročnosť celého programu, aby nemusela prebiehať korelácia každého detekovaného úseku signálu s každým detekovaným úsekom.

Na obrázku 2.3 je znázornený vývojový diagram, na ktorom je graficky spracovaný postup tvorby korelácie, výpočtu korelačného koeficientu a vytvorenia priemeru vhodných úsekov signálu.

Pomocou knižnice SciPy[20] v prostredí Python sme využili funkciu `correlation_lags`, ktorá vypočíta posun indexov pre 1D krížovú koreláciu a vráti pole, ktoré obsahuje indexy posunu. Následne bol signál zarovnaný voči vzoru a vypočítaný Pearsonov korelačný koeficient zo vzorca 2.2. Ide o štatistický test, ktorý určuje mieru asociácie dvoch signálov. Pre výber dát veľmi podobných šablóne bolo potrebné nastaviť prah. V tejto práci bola nastavená hodnota prahu na 0,9. Všetky úseky QRS komplexov, ktoré spĺňali túto podmienku, boli uložené do slovníka, ktorého kľúče tvoria jednotlivé zvody. Dáta boli následne kumulované samostatne pre každý zvod každého pacienta, pričom tieto dáta boli uložené do trojrozmerného poľa typu `numpy` (700 vzoriek, 12 zvodov, korelované a kumulované dáta). Tie úseky, ktoré



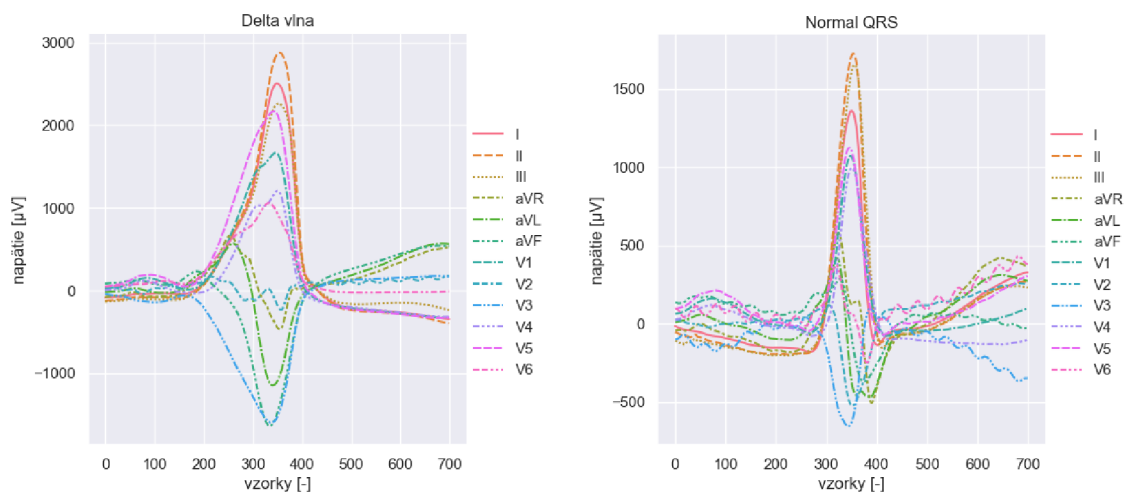
Obr. 2.3: Postup korelácie signálov

podmienku nespĺňali, boli vynechané. Kvôli výpočtovej a pamätovej náročnosti boli uložené už len dáta kumulovaného signálu.

$$r = \frac{\sum (x - m_x) \cdot (y - m_y)}{\sqrt{\sum (x - m_x)^2 \cdot \sum (y - m_y)^2}}, \quad (2.2)$$

kde m_x je priemerná hodnota vektoru x a m_y je priemerná hodnota vektoru y [21].

Grafy na obrázku 2.4 predstavujú vybrané .sel súbory u konkrétneho pacienta s kumulovanými hodnotami v jednotlivých zvodoch.



Obr. 2.4: Ukážka výstupu po kumulácii korelovaných signálov. Ľavý obrázok predstavuje signály s delta vlnou. Pravý obrázok predstavuje signály s normálnym QRS komplexom.

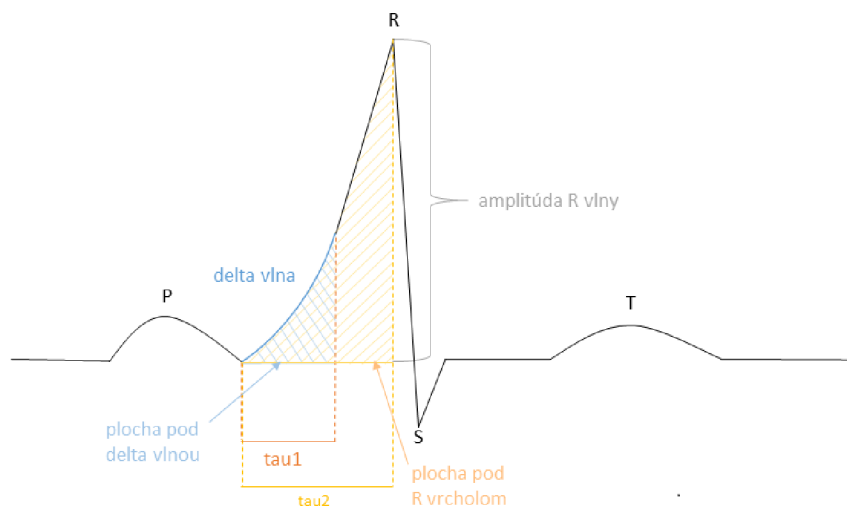
2.6 Extrakcia biomarkerov

Odhalenie biomarkerov na EKG je kľúčové k správne mu určení ochorenia a ako bolo spomenuté v kapitole 1.3 je najtypickejším príznakom pre Wolf-Parkinson-White syndróm výskyt pomalej delta vlny, a teda rozšírenie QRS komplexu nad 120 ms.

Získané korelované a priemerované dáta boli stále zašumené a obsahovali frekvencie, ktoré pre túto prácu nie sú dôležité. Keďže počas filtrovania môže dôjsť k posunutiu fáze, na náš signál sme aplikovali filtrovanie dopredu a odzadu (angl. forward-backward filtering). Ako filter bol použitý Butterworthov filter 4. radu s pásmovou priepustou od 0.5 Hz do 40 Hz, aby sme odstránili jednosmernú zložku signálu a zároveň odstránili striedavý interferenčný signál z napájania prístroja. [24].

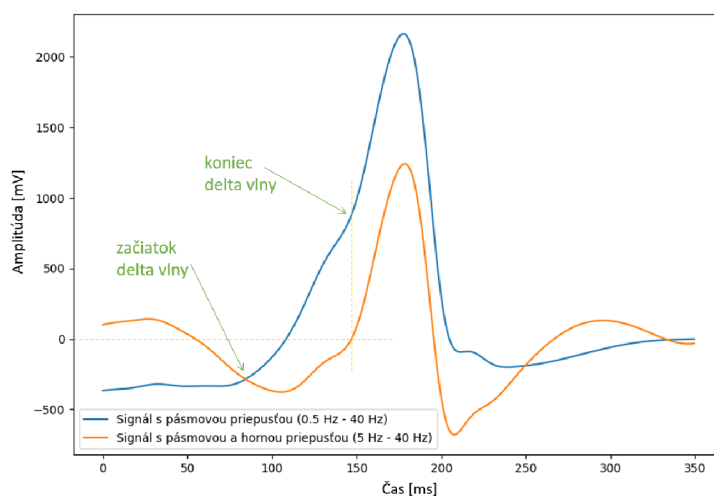
Pre určovanie biomarkerov zo signálu EKG a teda konkrétne z QRS komplexov je potrebné si v prvom rade určiť niekoľko hlavných bodov na QRS komplexe ako je začiatok delta vlny, koniec delta vlny a pozícia amplitúdy R peaku. Jednotlivé biomarkery sú znázornené na obrázku 2.5. Amplitúda R vrcholu (v práci skrátene ako $ampl_R$) bola určená nájdením maxima, a zároveň bola určená aj pozícia R vrcholu, ktorá bola použitá k ďalším výpočtom.

V ďalšom kroku bol na signál aplikovaný Butterworthov filter 4. rádu typu horná priepusť, ktorý filtroval hodnoty nižšie ako 5 Hz, čo je typická frekvencia delta vlny v EKG signáli. Získali sme tak signál, ktorý delta vlnu už neobsahoval a práve porovnanie týchto dvoch signálov sme využili k detekcii začiatku a konca delta vlny, pričom začiatok delta vlny sa určoval ako kríženie týchto signálov v mieste, ktoré sa nachádza pred miestom ukončenia delta vlny. Koniec delta vlny bol určovaný v bode,



Obr. 2.5: Znázornenie jednotlivých biomarkerov

kde signál s odfiltrovanou frekvenciou delta vlny prechádzal nulovou izolíniou, čo je znázornené aj na obrázku 2.6. Začiatok delta nebol pri všetkých signáloch pomocou programu nájdený, v takýchto prípadoch dodatočne ručne určený.



Obr. 2.6: Znázornenie jednotlivých biomarkerov

Medzi ďalšie príznaky, ktoré sme z našich signálov určovali, bola časová hodnota τ_1 , ktorá určovala čas od začiatku delta vlny až do maxima R vlny, čo sa zisťovalo odčítaním začiatku delta vlny od polohy amplitúdy R vlny. Určovali sme aj časovú hodnotu τ_2 , ktorá určovali dobu trvania delta vlny a bola určovaná ako odčítanie polohy začiatku delta vlny od bodu, kedy delta vlna skončila.

Dôležitým ukazovateľom delta vlny je jej rýchlosť v signáli, čo sme v našom prípade určovali pomocou derivácie, pričom sa vždy vzala maximálna hodnota de-

rivácie v úseku delta vlny (`der_d`). Okrem rýchlosti zmeny amplitúdy signálu môže byť dôležitým ukazovateľom EKG signálu aj plocha pod R vlnou (v našej práci ako `inte_R`), či plocha pod delta vlnou (`inte_d`) a ich vzájomný pomer (`inte_Rd`). Integrál R vlny bol počítaný pomocou funkcie `peak_widths` z knižnice `scipy.signal` [20], pričom sme si nastavili relatívnu výšku vrcholu na 0,85. Ide o výšku, pri ktorej sa meria šírka vlny ako percento jej významnosti, kde hodnota 1 znamená, že sa počíta výška najnižšieho bodu vlny. Následne bol signál prechádzaný for cyklom `p` v úseku nastavenej šírky R vrcholu a počítal sa integračný súčet ako absolútnu hodnotu jednotlivých vzoriek. Absolútna hodnota bola použitá z dôvodu, že občasne R vlna prekračovala nulovú izolíniu do záporných hodnôt. V prípade integrálu delta vlny bol výpočet jednoduchší a použitý iba integračný súčet absolútnych hodnôt v úseku od detekovaného začiatku delta vlny po jeho koniec. Posledným parametrom bol pomer týchto plôch.

Efektívna refraktérna perióda (ERP) akcesórnej dráhy rozdeľovala databázu pacientov na rizikových a nerizikových, pričom sme prah nastavili na 280 ms (typicky sa tento prah pohybuje od 280 ms do 300 ms). Všetci pacienti, ktorí mali hodnotu ERP väčšiu ako 280 milisekúnd boli považovaní za rizikových, zatiaľ čo pacienti s hodnotou ERP nižšou ako 280 ms boli považovaní za nerizikových. Práve tento parameter je kľúčový pre vytvorenie modelu predikcie rizika akcesórnej dráhy z pohľadu náhlej smrti srdca. Tieto parametre boli ku ostatným príznakom pridané až v poslednom krok pred tvorbou modelu.

2.7 Agregácia extrahovaných biomarkerov

Agregácia biomarkerov je proces, pri ktorom sa zhlukujú informácie z viacerých biomarkerov do jedného nového príznaku. Na agregáciu dát môžu byť použité rôzne metódy, pričom v tejto práci bolo zvolené určovania maxima, minima a priemeru hodnôt. Pre presnejšie určenie sme zvolili samostatne agregovať dáta pre končatinové zvody (I, II, III) a pre hrudné zvody (V1-V6). Pri tomto procese sa úplne vynechali augmentované zvody `aVL`, `aVF` a `aVR`, keďže ide o zvody, ktoré sú len dopočítané a nemerajú sa priamo.

Agregáciou pomocou maxima, minima a priemeru samostatne pre hrudné a končatinové zvody zo všetkých 7 biomarkerov sme získali 42 agregovaných biomarkerov, ktoré sme následne uložili do numpy poľa a pre lepšiu vizualizáciu sme ich previedli do .csv súboru. Príklad agregovaných dát môžeme vidieť v prílohe ??, pričom prvý stĺpec označuje pacienta (v prílohe sú pacienti označovaní poradovým číslo) a každý ďalší stĺpec je pomenovaný skrátenou formou príznaku. Vždy prvá časť predstavuje skratku príznaku, nasledujúce písmeno k značí, že ide o končatinové zvody, poprípade písmeno `h`, ktorá označuje hrudné zvody a poslednú časť tvorí typ agregácie,

pričom p značí priemer, \max značí maximálnu hodnotu a \min označujú minimálnu hodnotu (príklad: `ampl_R-k-p` značí priemer amplitúdy R vrcholu z hrudných zvodov).

2.8 Štatistické testovanie významnosti biomarkerov

2.8.1 Test normálneho rozloženia

Pred samotným testom na štatisticky významný rozdiel v biomarkeroch medzi rizikovými a nerizikovými pacientami bolo potrebné zistiť, či naše dáta majú normálne rozloženie alebo nie. Na tento účel sme využili D'Agostino a Pearsonov test, ktorý testuje nulovú hypotézu, že dáta jednotlivých biomarkerov majú normálne rozloženie [25]. D'Agostino a Pearsonov test počíta dva štatistické parametre šikmost a kurtózu. Šikmost určuje mieru asymetrie v distribúcii, zatiaľ čo kurtóza kvantifikuje, aká časť z celkového rozloženia je na konci. Ako prvé sa pomocou týchto dvoch parametrov vypočíta asymetria voči Gaussovskému rozloženiu a následne sa vypočíta p -hodnota zo súčtu nezrovnalostí medzi Gaussovským rozložením a rozložením našich hodnôt.

Keďže na výber správneho štatistického testu je potrebné, aby biomarker mal ideálne v oboch skupinách (ohrozených a aj neohrozených pacientov) rovnaké rozloženie dát, boli všetky príznaky roztriedené do troch skupín. Prvá skupina obsahovala tie biomarkery, ktoré majú v oboch prípadoch normálne rozloženie, druhá skupina tie, ktoré normálne rozloženie nemajú a posledná skupina obsahovala biomarkery, pri ktorých sa typy rozloženia nezhodovali v oboch kategóriách.

2.8.2 Test štatisticky významnej odchýlky

Pri analýze dát sa často stretávame s potrebou overiť, či existuje štatisticky významná odchýlka medzi dvoma skupinami alebo vzorkami. V prípade príznakov s normálnou distribúciou bol použitý T-test, zatiaľčo v prípade, že obe skupiny nemali normálne rozloženie alebo sa rozloženia nezhodovali bol použitý neparametrický Mann-Whitneyov U test. Studentov t-test je parametrický štatistický test, ktorý sa používa na porovnanie priemerov dvoch nezávislých skupín dát. Nulová hypotéza testuje, že neexistuje žiadny významný rozdiel medzi priemernými hodnotami oboch skupín. Implementáciou t-testu je funkcia `ttest_ind` v knižnici `scipy`. Pri použití t-testu predpokladáme nezávislosť pozorovaných skupín, čo znamená, že hodnoty v jednej skupine nemajú žiaden vzťah s hodnotami v druhej skupine, zároveň predpokladáme normálnu Gaussovskú distribúciu a homogenitu variácie dát.

Variácia dát nám vyjadruje kvantitatívnu mieru odchýlky dát od ich priemeru. Výstupom funkcie je vypočítaná t-štatistika a p-hodnota. P-hodnota (pravdepodobnostná hodnota) vyjadruje pravdepodobnosť dosiahnutia pozorovaného alebo ešte vzdialenejšieho výsledku, ak je nulová hypotéza pravdivá [28] [26].

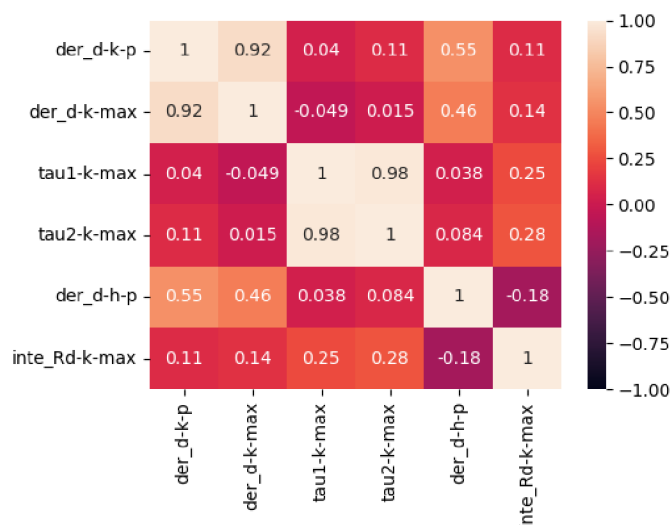
Ako druhý bol použitý Mann-Whitneyov U test, ktorý je analógiou studentovho t-testu, avšak nevyžaduje normálne gaussovské rozloženie. Porovnáva priemery medzi dvoma nezávislými vzorkami a testuje pritom nulovú hypotézu, že medzi priemermi dvoch vzoriek nie je štatisticky signifikantný rozdiel. Implementáciou tohto testu v jazyku Python je funkcia `mannwhitneyu` z knižnice `scipy`. Výstupmi funkcie sú u-štatistika a p-hodnota. Hodnota štatistiku U vyjadruje súčet hodnôt U pre jednu zo skupín. Vyššia hodnota U poukazuje na to, že hodnoty v druhej skupine sú pravdepodobne vyššie ako hodnoty v prvej skupine [27].

Oba testy boli hodnotené na základe výsledkov p-hodnoty. P-hodnota sa pohybuje v rozmedzí od 0 do 1, pričom hodnoty blízke 1 naznačujú vysokú pravdepodobnosť, že rozdiel medzi pozorovanými hodnotami je vysoko náhodný, zatiaľ čo hodnoty blízke 0 naznačujú, že existuje veľmi malá pravdepodobnosť, že by sa dosiahlo signifikantného rozdielu medzi skupinami, pokiaľ by vzorky pochádzali z totožného rozloženia. Typicky sa ako hraničná udáva hodnota 0.05, čo značí 5% hladinu významnosti. V prípade ak je p-hodnota menšia ako hladina významnosti 0.05, hypotézu zamietame, čo v prípade t-testu alebo u testu znamená, že medzi dvoma skupinami je štatisticky signifikantný rozdiel. Na druhej strane, ak je p-hodnota väčšia ako 0.05 znamená to, že nulovú hypotézu nezamietame a medzi dvoma skupinami nie je štatisticky významný rozdiel [28].

2.9 Korelácie biomarkerov

Multikolinearita sa vyskytuje v prípade, keď sú vstupné premenné modelu medzi sebou silne korelované, čo môže mať za následok, že model bude mať problém rozlíšiť vplyv jednej premennej od vplyvu iných premenných a bude pre neho problematické určiť, ktoré premenné sú skutočne dôležité pre predikciu. [30]

Získané agregované príznaky mohli byť medzi sebou lineárne závislé, čo bolo zisťované pomocou Pearsonovho korelačného koeficientu. V prípade, ak niektoré príznaky mali korelačný koeficient väčší ako 0.8, čo znamená, že 80% informácie mali rovnakých, bol z nich vybraný iba jeden a to na základe nami zvoleného poradia presnosti určenia jednotlivých biomarkerov. Na obrázku 2.7 je znázornená korelačná matica aj s hodnotami Pearsonovho korelačného koeficientu pre 6 biomarkerov, ktoré z predchádzajúceho kroku vyšli, že majú štatisticky významný rozdiel medzi hodnotami pre ohrozených a neohrozených pacientov.



Obr. 2.7: Korelačná matica signifikantných biomarkerov

V prípade maximálnej hodnoty derivácia delta vlny v končatinových zvodoch (der_d-k-max) a priemernej hodnoty derivácie delta vlny v končatinových zvodoch (der_d-k-p) je vidieť na obrázku 2.7, že Pearsonov korelačný koeficient dosahuje hodnotu 0.92, čo znamená, že medzi týmito dvoma biomarkermi je silná lineárna závislosť. Preto bol parameter der_d-k-max odstránený, keďže priemerná hodnota v sebe obsahuje obecnjšie informácie o derivácii delta vlny.

Zároveň z obrázku 2.7 vyplýva, že korelácia nastala aj medzi biomarkermi tau2-k-max (maximálna hodnota doby od začiatku delta vlny po R vrchol pre končatinové zvody) a tau1-k-max (maximálna hodnota doby trvania delta vlny pre končatinové zvody), keďže Pearsonov korelačný koeficient dosahoval hodnotu 0.98. Z týchto dvoch biomarkerov bol odstránený tau1-k-max, keďže parameter tau2-k-max v sebe zahŕňa aj oblasť celej delta vlny.

Zvyšné 4 biomarkery (der_k-p, tau2-k-max, der_d-h-p, inte_Rd-h-max) boli použité na tvorbu modelu predikcie rizika akcesórnej dráhy pre pacientov ohrozených náhlou smrťou srdca.

2.10 Model predikcie

Strojové učenie má v medicíne veľký potenciál a môže pomôcť zlepšiť kvalitu zdravotnej starostlivosti a predchádzať zhoršeniu stavu pacientov. Rôzne modely strojového učenia môžu byť využívané na predpovedanie rizika vzniku ochorení, pravdepodobnosti úmrtia, alebo aj na predikciu liečebných úspechov [32].

2.10.1 Klasifikačný model

Medzi základné nástroje strojového učenia patrí klasifikačný model, ktorý umožňuje rozdeliť dáta do jednotlivých kategórií na základe ich vlastností a charakteristík. Jeho hlavným cieľom je naučiť sa implicitné vzťahy medzi vstupnými údajmi a príslušnými triedami, čo umožňuje predikovať triedu pre nové, neznáme dáta. Pri klasifikácii sa predpokladá, že existuje určitý vzťah medzi vstupnými údajmi a ich príslušnými triedami, čo umožňuje vytváranie matematických modelov na odhadnutie týchto vzťahov.

Pri klasifikačných modeloch sa rozlišuje medzi dvoma hlavnými typmi učenia a to učenie s učiteľom (angl. supervised learning) a učenie bez učiteľa (angl. unsupervised learning). Tieto prístupy sa líšia v spôsobe tréningu a využitia dostupných údajov. Pri učení s učiteľom rozdelujeme dataset na dve skupiny a to na tréningové a testovacie dáta, pričom v prípade tréningových dát je ku vstupným atribútom priradená aj príslušná trieda. V našom prípade sú to dve triedy a to rizikovní a nerizikovní pacienti, kde každý pacient z tréningovej sady dát je zaradený do jednej z týchto dvoch tried. Testovacia množina slúži na overenie schopnosti modelu generalizovať naučené vzťahy na nové, neznáme dáta. V tejto práci boli testovacie a tréningové dáta rozdelené pomerom 30:70. Na naučený model sa aplikujú testovacie dáta a výsledné klasifikácie sa používajú na vyhodnotenie úspešnosti modelu. Porovnávajú sa klasifikácie z modelu s označenými triedami v testovacej množine.

V práci boli skúmané tri typy klasifikačných modelov pre určenie najlepšieho z nich a to konkrétne rozhodovací strom, model logistickej regresie a algoritmy podporných vektorov (SVM z angl. Support vector machine).

2.10.2 Hodnotenie kvality modelu

Na vyhodnotenie úspešnosti modelu sa využíva matica zámieny (angl. confusion matrix), ktorá predstavuje tabuľkovú metódu, ktorá sumarizuje správnosť predikcie modelu. Každý riadok predstavuje skutočnú triedu zaradenia prvku (pacienta) a každý stĺpec predstavuje predikovanú triedu. V matici zámieny, ktorú môžeme vidieť v tabuľke 2.2 rozoznávame štyri typy hodnôt: skutočne pozitívnu (TP z angl. True Positive), falošne pozitívnu (FP z angl. False Positive), skutočne negatívnu (TN z angl. True Negative) a falošne negatívnu (FN z angl. False Negative). Skutočne pozitívne sú tí pacienti, ktorí boli správne zaradení ako rizikovní, falošne pozitívni sú tí, ktorí boli nesprávne klasifikovaní ako pozitívni, ale mali byť klasifikovaní ako negatívni. Skutočne negatívni pacienti boli správne zaradení ako nerizikovní a poslední falošne negatívni mali byť klasifikovaní ako pozitívni, ale boli klasifikovaní ako negatívni.

	Skutočne pozitívni (1)	Skutočne negatívni (0)
Predikované pozitívni (1)	TP	FP
Predikované negatívni (0)	FN	TN

Tab. 2.2: Matica zámeny

Z konkrétnych hodnôt matice zámen je možné určiť niekoľko ďalších parametrov úspešnosti modelu akými sú presnosť (ACC z angl. accuracy), špecificita a senzitivita. Presnosť udáva pomer správne klasifikovaných pacientov ku celkovému počtu všetkých pacientov 2.3, špecificita vyjadruje percento skutočne negatívnych pacientov, ktorí boli správne klasifikovaní ako negatívni a je daná vzorcom 2.4, zatiaľ čo senzitivita, ktorá sa tiež nazýva aj citlivosť modelu, udáva percento skutočne pozitívnych pacientov, ktorí boli správne klasifikovaní ako pozitívni a je daná vzorcom 2.5.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

$$\text{Špecificita} = \frac{TN}{TN + FP} \quad (2.4)$$

$$\text{Senzitivita} = \frac{TP}{TP + FN} \quad (2.5)$$

2.10.3 Model rozhodovacieho stromu

Rozhodovací strom je v dnešnej dobe jedným z často používaných algoritmov strojového učenia, ktorí patrí do kategórie učenia s učiteľom. Je založený na princípe postupného rozdeľovania dát na základe príznakov a vytvárania istej hierarchie rozhodnutí. Reprezentovaný je štruktúrou stromu, v ktorej každý vnútorný uzol reprezentuje rozhodovaciu podmienku, ktorá oddeľuje dáta na základe hodnoty konkrétneho príznaku. Každá vetva predstavuje jednu možnosť rozhodnutia na základe konkrétneho príznaku, ktorý je spojený s daným vrcholom a každý list stromu predstavuje konečný výsledok rozhodovania. Rozhodovací strom je štruktúrovaný tak, aby minimalizoval chybu predikcie a zároveň bol čo najefektívnejší. Kvalita klasifikácie je hodnotená pomocou entropie, ktorá vyjadruje mieru neistoty v systéme. Entropia môže nadobudnúť hodnôt od 0 do $\log_2(n)$, pričom n udáva počet kategórií v množine. V prípade tejto práce je $n=2$. Ak sú všetky prvky v množine dát rovnakej triedy máme úplnú istotu a entropia nadobúda hodnoty 0. Ak sú všetky triedy v množine dát rovnomerne zastúpené entropia nadobúda maximálnu hodnotu

a znamená to, že je v bode maximálnej neistoty, čo vychádza z rovnice 2.6.

$$E(S) = - \sum_{i=1}^c p_i (\log_2 p_i), \quad (2.6)$$

kde S je množina dát, i je trieda a $p(i)$ je pravdepodobnosť výskytu triedy i v množine S . Pri tvorbe rozhodovacieho stromu je snaha minimalizovať hodnotu entropie, čo znamená, že je snaha dosiahnuť maximálnu homogenitu vrámci jednotlivých skupín. V programovacom jazyku Python bol model implementovaný z knižnice scikit-learn [35] pomocou funkcie `DecisionTreeClassifier()` [33] [34].

2.10.4 Model logistickej regresie

Druhým typom modelu použitý v práci bol model logistickej regresie, ktorý taktiež patrí do kategórie učenia s učiteľom. Princíp je založený na logistickej funkcii, ktorá transformuje lineárnu kombináciu vstupných príznakov pomocou sigmoidnej funkcie. Sigmoidná funkcia, ktorá má tvar ako písmeno 'S', zabezpečuje, že výstup je vždy v rozmedzí od 0 do 1, pričom hodnoty blízke 0 udávajú nízku pravdepodobnosť zaradenia k určitej triede. Na druhú stranu hodnoty blízke 1 vyjadrujú, že pravdepodobnosť príslušnosti k danej triede je veľmi vysoká. V prípade, že lineárna kombinácia vstupných príznakov je rovná nule, pravdepodobnosť príslušnosti k danej triede je 50%. Každý príznak je vážený koeficientom, ktorý udáva, aký význam má daný príznak pri predikcii. Podobne ako lineárna regresia, logistická regresia môže zahŕňať iba jednu alebo viac nezávislých premenných, hoci skúmanie viacerých premenných je vo všeobecnosti informatívnejšie, pretože odhaľuje jedinečný príspevok každej premennej po úprave pre ostatné, čo vychádza aj z rovnice 2.7.

$$\hat{Y}_i = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i}}, \quad (2.7)$$

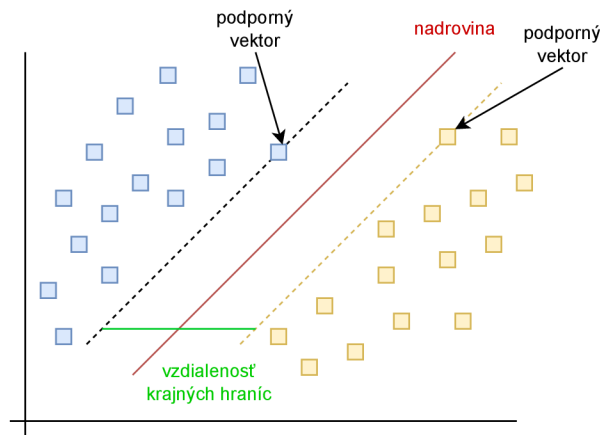
kde \hat{Y}_i vyjadruje aká je odhadovaná pravdepodobnosť, že výstupná hodnota nachádza v jednej binárnej skupine (i) oproti výskytu v druhej, X_i je nezávislá premenná a β_i je koeficient udávajúci sklon regresnej priamky, pričom β_0 je priesečník krivky s osou y [37].

Základné predpoklady, ktoré musia byť splnené pre logistickú regresiu, zahŕňajú nezávislosť chýb, absenciu multikolinearity a absenciu silne vplyvných extrémnych hodnôt. V práci bol model logistickej regresie implementovaných taktiež pomocou knižnice scikit-learn a funkcie `LogisticRegression()` [35] [36] [37].

2.10.5 Model SVM

Model podporných vektorov je model používaný pre regresné aj klasifikačné úlohy. Cieľom SVM je nájsť optimálnu hranicu separácie medzi jednotlivými triedami

v priestore príznačov. Nadrovina je definovaná polohou podporných vektorov, ktoré sú vybrané tak, aby boli najbližšie vzorky oboch tried maximálne vzdialené, čo je znázornené na obrázku 2.8.



Obr. 2.8: Schématické znázornenie lineárneho modelu SVM

SVM model môže byť buď lineárny alebo s takzvanou jadrovou metódou, ktorá umožňuje transformáciu nelineárne separovaných dát do vyššieho priestoru a tým vytvoriť nelineárnu separačnú hranicu. Existujú viaceré typy jadrových metód ako napríklad polynomiálne jadro, sigmoidálne jadro či radiálne bázové funkčné jadro (RBF z angl. Radial Basis Function), ktoré bolo použité aj v prípade tejto práce [38]. Funkcia jadra je matematická funkcia, ktorá prevádza nízkorozmerný vstupný priestor na priestor vyššej dimenzie. To sa dosiahne mapovaním údajov do nového priestoru funkcií. V tomto priestore budú údaje lineárne oddeliteľné. Jadro RBF funguje tak, že mapuje údaje do vysokodimenzionálneho priestoru nájdením bodov a štvorcov všetkých funkcií v množine údajov a následným vykonaním klasifikácie pomocou základnej myšlienky lineárneho SVM [39]. Pre projekciu dát do vyššieho dimenzionálneho priestoru používa jadro RBF takzvanú radiálnu bázickú funkciu, ktorú možno zapísať ako:

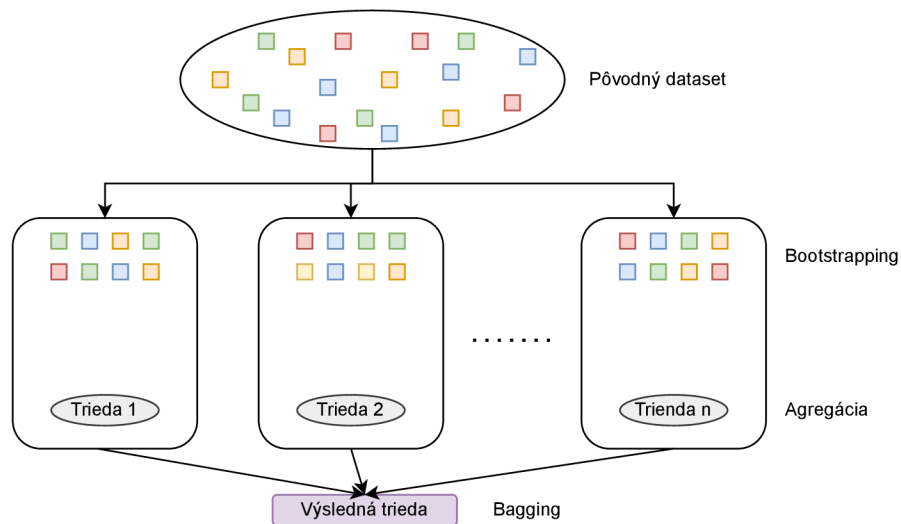
$$K(X_1, X_2) = \exp\left(-\frac{\|X_1 - X_2\|^2}{2\sigma^2}\right), \quad (2.8)$$

kde $\|X_1 - X_2\|^2$ vyjadruje štvorcovú euklidovskú vzdialenosť bodov a σ je parameter na vyladenie funkcie [39].

V prípade tejto práce bol implementovaný model pomocou funkcie `sklearn.svm.SVC` z knižnice `scikit-learn`[35]. Funkcia `SVC()` slúži na implementáciu SVM modelu pre úlohy klasifikácie.

2.10.6 Metóda bagging

Bagging je metóda súborového učenia (z angl. bootstrap aggregation), ktorá sa používa na zvýšenie presnosti klasifikačného modelu. Princípom metódy je kombinovať výstupy viacerých nezávislých modelov, ktoré boli trénované na rôznych podmnožinách tréningových dát. V prvom kroku sa vytvorí M tréningových podmnožín s opakovaním (bootstrapping). Tieto podmnožiny majú rovnakú veľkosť ako pôvodná tréningová množina. Pre každú podmnožinu sa vytvorí a učí samostatný model strojového učenia. Následne sa výsledky zo všetkých modelov spoja a určí sa finálny výsledok výberom triedy s najväčším zastúpením. Zjednodušený postup bagging metódy je znázornený na obrázku 2.9. Výsledný model je odolnejší voči preučeniu a šumu v dátach, keďže kombinuje viacero nezávislých modelov [40]. V práci bola použitá funkcia `BaggingClassifier()` z knižnice `scikit-learn` [35]. Vstupným parametrom boli vždy jednotlivé modely (vždy sa tvoril bagging len pre jeden druh modelu) a parameter `n_estimators`, ktorý určuje počet nezávislých modelov bol nastavený na hodnotu 20.



Obr. 2.9: Schématické znázornenie metódy bagging

3 Výsledky

3.1 Normálne rozloženie dát

V prípade testu normálneho rozloženia dát sa počítalo so 42 biomarkermi samostatne pre rizikových pacientov označených ako $ERP = 1$ a nerizikových pacientov označených $ERP = 0$. Porovnanie rozloženia dát bolo vykonané nie len pomocou D'Agostino a Pearsonoho testu, ale aj vizuálne pomocou škatulových grafov, ktoré sa nachádzajú v prílohe A.1. V prípade rizikových pacientov malo 24 príznakov normálne rozloženie, zatiaľ čo 18 nemalo normálne rozloženie. V prípade nerizikových pacientov bol pomer opačný čo znamená, že 23 biomarkerov nemalo normálnu distribúciu hodnôt a 19 biomarkerov malo normálne rozloženie. 15 biomarkerov malo v oboch skupinách normálne rozloženie, 14 biomarkerov nemalo normálne rozloženie v oboch skupinách a 13 biomarkerov malo v jednej zo skupín normálne rozloženie a v druhej zasa normálne rozloženie nemalo.

3.2 Štatisticky významná odchýlka

Pri testovaní štatisticky významnej odchýlky pomocou t-testu a Mann-Whitney U testu na dátach od pacientov s hladinou významnosti 0.05 mal iba jediný biomarker štatisticky významný rozdiel. Tau1-k-max je biomarker udávajúci maximálnu dobu trvania delta vlny pri snímaní z končatinových zvodov. Jeho p-hodnota bola 0.047. Keďže testovaním vyšiel ako jediný parameter, ktorý má štatisticky významnú odchýlku medzi rizikovými a nerizikovými pacientmi, rozhodli sme sa posunúť hranicu p-hodnoty na 0.1, čím bola minimalizovaná chyba II. typu, ktorá udáva chybu prijatia nulovej hypotézy, ak je v skutočnosti nesprávna [29]. Na druhej strane sa posunutím hranice k vyšším hodnotám zvyšuje pravdepodobnosť chyby I. typu, ktorá udáva chybné zamietnutie nulovej hypotézy, aj keď je v skutočnosti správna, čo vedie k zvýšeniu falošne pozitívnych výsledkov [29]. V prípade zvýšenia hranice hladiny významnosti sme získali celkovo 6 biomarkerov (der_d-k-max, der_d-k-p, tau2-k-max, tau1-k-max, der_d-h-p, inte_RD-h-max), ktoré mali hodnoty nižšie ako 0.1 a teda mali štatisticky významný rozdiel medzi kategóriou rizikových a nerizikových pacientov. Pri týchto 6 biomarkeroch bola testovaná aj ich vzájomná lineárna závislosť. V prípade der_d-k-max a der_d-k-p vyšla hodnota korelačného koeficientu 0.92, a preto bol príznak der_d-k-max odstránený. Rovnako aj v prípade tau2-k-max a tau1-k-max vyšiel Pearsonov korelačný koeficient vyšší ako stanovený prah 0.8, a preto bol príznak tau1-k-max odstránený.

3.3 Porovnanie výsledkov modelov predikcie

Pre získanie čo najpresnejšieho modelu pre predikciu rizika náhlej srdčnej smrti u pacientov boli porovnávané tri typy model strojového učenia a to konkrétne model rozhodovacieho stromu, model logistickej regresie a model podporných vektorov známy pod skratkou SVM. Všetky typy modelov boli porovnávané v jednoduchých implementáciách a zároveň v prípade, kedy na nich bola aplikovaná metóda bagging. Tabuľky 3.1 a 3.2 predstavujú porovnanie kvality jednotlivých modelov. Z výsledkov je zrejmé, že rozhodovací strom bol najmenej presný, čo mohlo byť spôsobené malým množstvom biomarkerov a zároveň celkovo menším datasetom. Zároveň pre každý model bola vytvorená matica zámeny, z ktorej bola vypočítaná presnosť, špecificita a senzitivita modelu. Matice zámeny pre jednotlivé modely a na nich implementovanú metódu bagging sú znázornené na obrázku 3.1.

	Rozhodovací strom	Logistická regresia	SVM
ACC	0.44	0.68	0.74
Špecificita	0.71	0.71	0.59
Senzitivita	0.18	0.65	0.88

Tab. 3.1: Hodnotenie kvality modelov na testovacej množine.

	Rozhodovací strom	Logistická regresia	SVM
ACC	0.59	0.68	0.68
Špecificita	0.76	0.59	0.53
Senzitivita	0.41	0.76	0.82

Tab. 3.2: Hodnotenie kvality modelov s aplikovanou metódou bagging na testovacej množine.

	Rozhodovací strom				Logistická regresia				SVM			
	Min	Max	\bar{x}	δ	Min	Max	\bar{x}	δ	Min	Max	\bar{x}	δ
Bagging	0.41	0.65	0.52	0.04	0.59	0.74	0.66	0.03	0.5	0.74	0.62	0.05
Bez bagging	0.47	0.59	0.52	0.03	0.68	0.68	0.68	0.00	0.74	0.74	0.74	0.00

Tab. 3.3: Vyhodnotenie presnosti modelov predikcie.

Z výsledkov presnosti modelov predikcie v tabuľke 3.3 je vidieť, že model rozhodovacieho stromu mal najmenšiu úspešnosť správneho zaradenia pacientov do tried, keďže priemerná hodnota v prípade klasického modelu aj v prípade použitia metódy bagging bola 0,52, čo znamená, že iba 52% všetkých testovacích pacientov

bolo zaradených správne. Najlepšie zo všetkých modelov vyšiel model podporných vektorov, ktorého úspešnosť bola vždy 74%. Z matíc zámery 3.1 je vidieť, že pre model SVM bolo 7 pacientov predikovaných ako rizikových, aj keď neboli ohrození, zatiaľ čo 2 pacienti boli predikovaní ako nerizikoví, aj keď mali patriť do skupiny rizikových. V prípade zaradenia pacientov do skupiny rizikových a nerizikových je síce nesprávne, ale lepšie, ak je pacient zaradený do skupiny ohrozených, aj keď mu nehrozí riziko náhlej smrti srdca, keďže pacient bude pod väčším dohľadom. Na druhej strane ak je pacient zaradený do skupiny neohrozených, avšak v skutočnosti je ohrozený, môže byť problém v nedostatočnej prevencii pacienta.

		Skutočne pozitívni (1)		Skutočne negatívni (0)	
Predikované pozitívni (1)	RS	3	5		
	LR	11	5		
Predikované negatívni (0)	RS	14	12		
	LR	6	12		

		Skutočne pozitívni (1)		Skutočne negatívni (0)	
Predikované pozitívni (1)	RS	7	4		
	LR	13	7		
Predikované negatívni (0)	RS	10	13		
	LR	4	10		

Obr. 3.1: Matice zámery pre jednotlivé modely (RS - rozhodovací strom, LR - logistická regresia), vpravo aj s aplikovanou metódou bagging.

Z matíc zámery je vidieť, že najpresnejšie určenie má model logistickej regresie, na ktorý bola aplikovaná metóda bagging, keďže najmenší počet pacientov bol zaradený do skupín falošne pozitívnych a falošne negatívnych.

4 Diskusia

Táto práca mala v úvode stanovené tri hlavné ciele. Prvým bolo objasniť z dostupnej literatúry prejavy a riziká prídavnej sinoatriálnej dráhy pri detských pacientoch, ktoré sú spojené s Wolf-Parkinson-White syndrómom. Z dostupných informácií je známe, že najčastejším prejavom týkajúcim sa akcesórnej SA dráhy je delta vlna a s ňou spojené rozšírenie QRS komplexu na EKG. Druhým cieľom tejto práce bolo spracovanie signálov EKG od pacientov po elektrofyzilogickom vyšetrení a posledným cieľom bola extrakcia vhodných príznakov typických pre WPW syndróm a následná implementácie modelu predikcie rizika akcesórnej dráhy spojenej s náhlou smrťou srdca.

Retrospektívne získané dáta od pacientov obsahovali aj dáta, ktoré neboli pre túto prácu vhodné, pretože nespĺňali niektorú z 3 podmienok opísaných v kapitole Pacienti 2.1. Pre dáta pacientov, ktorí spĺňali všetky podmienky, boli určené 4 základné hodnoty efektívnej refraktérnej periódy akcesórnej dráhy a ich výsledky môžeme vidieť v tabuľke 2.1. Z tabuľky je známe, že antegrádne vedenie ERP akcesórnej dráhy je najvhodnejším indikátorom predikcie WPW syndrómu a práve preto bolo v práci použité ako klasifikátor rizikových a nerizikových pacientov.

Meranie a zaznamenávanie dát prebiehalo počas elektrofyzilogického vyšetrenia vo FN Brno, ktorá poskytla dáta pre túto prácu. Počas vyšetrenia bolo pacientovi okrem iného snímané aj EKG. Práve dáta z 12-zvodového EKG boli pre túto prácu kľúčové. Dáta boli ako prvé spracované pomocou programu SignalPlant, v ktorom sme využili dávkové označovanie QRS komplexov. Pri dávkovom označovaní QRS komplexov nastal v prípade niekoľkých dát problém so správnym detekovaním QRS komplexov a program tieto súbory vynechal. Najčastejším dôvodom bol výskyt hodnôt na začiatku alebo na konci signálu, ktoré mali hodnoty rádovo aj 1000 krát väčšie, než sú normálne hodnoty EKG. Pokiaľ bol výskyt týchto nechcených hodnôt vo väčšine súborov alebo neboli správne vytvorené .sel súbory boli pacienti vyradení. Kvôli tejto chybe boli vyradení 4 pacienti.

Keďže bolo potrebné správne zaradiť každý jeden QRS komplex z pacientovho EKG záznamu bol pri každom pacientovi ručne vybraný jeden vzorový úsek QRS komplexu voči ktorému bola následne vykonaná korelácia pomocou korelačnej matice. V prípade 6 pacientov bolo veľmi problematické určiť vhodný úsek delta vlny, keďže ich dáta boli buď príliš nekonzistentné alebo obsahovali výraznú fibriláciu, kvôli ktorej nebolo možné odhaliť delta vlnu. Výber takýchto dát by zaneslo chybné hodnoty do množiny pre budúce tréningovanie a testovanie modelov. Vo viacerých prípadoch sa stalo, že niektorý zo zvodov chýbal alebo signál obsahoval výrazný šum, čo taktiež vnieslo chybu do datasetu. Výber vzorového úseku bol realizovaný hlavne kvôli zníženiu výpočtovej náročnosti celého programu, keďže korelácia každého QRS

komplexu ku každému v prípade aj niekoľko stoviek sekundových záznamov by bola výpočtovo a pamäťovo náročná.

V prípade detekovania biomarkerov na QRS komplexe bol najväčší problém detekovať začiatok delta vlny. V prípade 130 signálov (približne 1 zvod u 1 pacienta) nebol tento bod vôbec určený a bolo potrebné ručne zadať jeho hodnotu. Takéto hodnoty boli zapísané do samostatného .csv súboru, s ktorými následne program počítal. Keďže určenie začiatku delta vlny nebolo v každom prípade správne, zanáša to chybu do celkového výpočtu biomarkerov. V prípade, kedy delta vlna nebola v signáli zreteľne viditeľná, bol problém správne určiť aj jej koniec, čo ovplyvnilo hodnoty biomarkeru tau1.

Pri agregácii bol každý jeden pacient na základe hodnoty ERP zaradený ako rizikový alebo ako nerizikový. V prípade, že ERP hodnota bola väčšia ako 280 ms, bol tento pacient zaradený do skupiny ohrozených, zatiaľ čo ak hodnota ERP bola menšia ako 280 ms, bol pacient zaradený do skupiny nerizikových. Takýmto spôsobom nám vznikli dva samostatné datatsety, ktoré boli následne podrobené štatistickej analýze, pričom hodnoty ERP väčšie ako 280 boli označované 1 a hodnoty menšie ako 280 ms zase 0. V prípade 8 pacientov boli chybné vypočítané niektoré príznaky, čo znamenalo, že pacienti boli vylúčení.

Keďže výber správneho štatistického testu si vyžaduje zistenie rozloženia dát v jednotlivých príznakoch bol použitý test normálneho rozloženia, ktorý testuje nulovú hypotézu, že dáta jednotlivých biomarkerov majú normálne rozloženie. Bol na to použitý D'Agostino a Pearsonov test. Počítala sa miera asymetrie voči Gaussovskému rozloženiu. Jednotlivé biomarkery boli roztriedené do troch kategórií. Prvú kategóriu tvorili tie príznaky, ktoré mali v prípade $ERP = 1$ a aj $ERP = 0$ normálne rozloženie, druhú kategóriu tvorili tie príznaky, ktoré v oboch skupinách nemali normálne rozloženie a tretiu skupinu tvorili tie biomarkre, ktoré mali rozdielne rozdelenie pre hodnoty $ERP = 1$ a $ERP = 0$. Pre grafické porovnanie rozloženia dát biomarkerov pre dve skupiny ERP boli vytvorené krabicové grafy A.1.

Pri tréovaní modelov bola významná veľkosť datasetu tréovacej a testovacej množiny, keďže celý dataset obsahoval iba 112 pacientov so 4 biomarkermi a pri rozdeľovaní na tréovacie a testovacie dáta bol rozdelený v pomere 30% pre testovacie a 70% na tréovanie modelu, čo znamená, že z celkových 112 pacientov bolo 34 použitých na testovanie a 78 na tréovanie.

Záver

Odhalenie rizika akcesórnej dráhy z EKG záznamu by mohlo pomôcť ku včasnej a jednoduchšej predikcii a prevencii pred náhlou smrťou srdca u detských pacientov s Wolf-Parkinson-White syndrómom. V súčasnej dobe je potrebné na odhalenie rizika podstúpiť elektrofyziologické vyšetrenie, počas ktorého doktor z viacerých parametrov vyhodnocuje ohrozenie pacienta náhlou smrťou srdca. Táto práca bola zameraná na spracovanie 12-zvodového EKG záznamu z retrospektívnej štúdie v rozmedzí od roku 2012 do roku 2022. V štúdií bolo zahrnutých 182 pacientov vo veku od 5 do 19 rokov, pričom nie všetci boli vhodní pre túto prácu. Extrahované biomarkre z 12-zvodového EKG boli použité pre model predikcie rizika.

Pri testovaní štatistickej významnosti rozdielu biomarkerov medzi rizikovými a nerizikovými pacientmi vyšiel v prípade nastavenia prahu p-hodnoty na 0.05 iba jeden štatisticky významný biomarker a to tau1-k-max, ktorý udáva maximálnu dobu trvania delta vlny pri snímaní z končatinových zvodov. Keďže predikovať riziko len na základe jedného parametru by nebolo vhodné, bol prah p-hodnoty posunutý na hodnotu 0.1, pri ktorej vyšli 4 signifikantné biomarkre (der_k-p, tau2-k-max, der_d-h-p, inte_Rd-h-max), ktoré boli použité aj ako vstupné dáta modelov predikcie.

V práci boli otestované 3 typy modelov strojového učenia a to konkrétne model rozhodovacieho stromu, model logistickej regrezie a model SVM. Na každý z modelov bola aplikovaná aj metóda bagging. Zo získaných hodnôt zapísaných v matici zámeny môžeme konštatovať, že najvhodnejší na predikciu rizika akcesórnej dráhy u pacientov s náhlou smrťou srdca je model podporných vektorov (SVM), ktorý má úspešnosť modelu 74% a zo všetkých 34 testovacích pacientov, zaradil nesprávne 9, pričom 7 bolo zaradených ako falošne pozitívnych a 2 boli zaradení ako falošne negatívnych. Najmenej sa osvedčil model rozhodovacieho stromu.

Literatúra

- [1] CHHABRA Lovely, GOYAL Amandeep, BENHAM Michael D. Wolff Parkinson White Syndrome [online]. StatPearls Publishing, 2022 [cit. 2022-10-16]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK554437/>
- [2] DOBRZYNSKI, Halina, Robert H. ANDERSON a Andrew ATKINSON. Structure, function and clinical relevance of the cardiac conduction system, including the atrioventricular ring and outflow tract tissues. *Pharmacology & Therapeutics* [online]. 2013, 139(2), 260-288 [cit. 2022-12-22]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0163725813000892> doi:<https://doi.org/10.1016/j.pharmthera.2013.04.010>
- [3] SHIH, Hue-Teh. Anatomy of the action potential in the heart. *Texas Heart Institute Journal* [online]. 1994, 21(1), 30-41 [cit. 2022-12-22]. PMID: 7514060; PMCID: PMC325129. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC325129/>
- [4] FISHER, John D. a Andrew KRUMERMAN. Clinical Electrophysiology Techniques. In: *Electrophysiological Disorders of the Heart* [online]. Elsevier, 2012, 2012, s. 267-280 [cit. 2023-05-19]. ISBN 9781437702859. Dostupné z: doi: 10.1016/B978-1-4377-0285-9.00020-X
- [5] HAMPTON, John R. EKG stručně, jasně, přehledně. Překlad 7. vydání. Praha: Grada, 2013. ISBN 978-80-247-4246-5. [cit. 2022-11-20]
- [6] FONTANA, Josef, Josef KROUPA, Patrik MAŘA, Peter IVÁK a kolektiv. Srdce. Funkce buněk a lidského těla: Multimediální skripta [online]. Praha: 3. lékařská fakulta Univerzity Karlovy, 2016 [cit. 2022-11-20]. Dostupné z: <http://fb.lt.cz/skripta/x-srdce-a-obeh-krve/1-srdce/>
- [7] HABERL, Ralph. EKG do kapsy. 4. vydání. Praha: Grada, 2012. ISBN 978-80-247-4192-5. [cit. 2022-10-16]
- [8] What is the Effective Refractory Period?. In: Study.com [online]. 2022 [cit. 2022-11-25]. Dostupné z: <https://study.com/academy/lesson/effective-refractory-period-ventricular-atrial.html>
- [9] Wolff-Parkinson-White (WPW) syndrome Mnemonic. In: MED Explosion [online]. 04/2020n. 1. [cit. 2022-11-25]. Dostupné z: <https://medexplosion.com/wolff-parkinson-white-wpw-syndrome-mnemonic/>

- [10] PAGE, Richard L., José A. JOGLAR, Mary A. CALDWELL, et al. 2015 ACC/AHA/HRS Guideline for the Management of Adult Patients With Supraventricular Tachycardia: Executive Summary. *Circulation* [online]. 2016, 133(14) [cit. 2022-12-28]. ISSN 0009-7322. Dostupné z: doi:10.1161/CIR.0000000000000310
- [11] ETHERIDGE, Susan P., Carolina A. ESCUDERO a Andrew D. BLAUFox. Life-Threatening Event Risk in Children With Wolff-Parkinson-White Syndrome: A Multicenter International Study. *JACC: Clinical Electrophysiology* [online]. 2018, 4(4), 433-444 [cit. 2023-05-19]. ISSN 2405-500X. Dostupné z: <https://doi.org/10.1016/j.jacep.2017.10.009>.
- [12] BUTTNER Robert a BURNS Ed. Pre-excitation syndromes [online]. LITFL - Life in the FastLane, 2022 [cit. 2022-10-16]. Dostupné z: <https://litfl.com/pre-excitation-syndromes-ecg-library/>.
- [13] SZYMANSKI, Michael W. a Davinder P. SINGH. Isoproterenol. StatPearls Publishing [online]. Treasure Island (FL), 2022 Jan [cit. 2022-11-17]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK526042/>
- [14] Směrodatná odchylka. Matematika polopatě [online]. [cit. 2022-12-30]. Dostupné z: <https://www.matweb.cz/smerodatna-odchylka/>
- [15] BREMBILLA-PERROT, Beatrice. Electrophysiological evaluation of Wolff-Parkinson-White Syndrome. *Indian Pacing Electrophysiol J.* [online]. 2002, 2(4), 143–152 [cit. 2022-11-25]. PMID: 16951730; PMCID: PMC1557416. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1557416/>
- [16] F. Plesinger, J. Jurco, J. Halamek, and P. Jurak, “SignalPlant: an open signal processing software platform,” *Physiol. Meas.*, vol. 37, no. 7, pp. N38–N48, 2016.[cit. 2022-11-25] Dostupné z: doi:10.1088/0967-3334/37/7/N38
- [17] Introduction to HDF5. HDF Support Portal [online]. Champaign, USA, 2019 [cit. 2022-11-25]. Dostupné z: <https://portal.hdfgroup.org/display/HDF5/Introduction+to+HDF5#IntroductiontoHDF5-introprog>
- [18] F. Plesinger, J. Jurco, P. Jurak, J. Halamek, “Robust Multichannel QRS Detection,” in *Computing in Cardiology Conference (CinC)*, 2014, [cit. 2022-11-25], vol. 41, pp. 557–560.
- [19] ALTMAN, Douglas G. *Practical Statistics for Medical Research*. [cit. 2022-12-28] USA: CRC press, 1991. ISBN 0-412-27630-5.

- [20] SciPy [online]. [cit. 2022-12-30]. Dostupné z: <https://docs.scipy.org/doc/scipy/>
- [21] Scipy.stats.pearsonr. SciPy [online]. [cit. 2022-12-30]. Dostupné z: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html>
- [22] LANGNER, Paul H. a David B. GESELOWITZ. First Derivative of the Electrocardiogram. *Circulation Research* [online]. 1962, 10(2), 220-226 [cit. 2022-12-30]. ISSN 0009-7330. Dostupné z: doi:10.1161/01.RES.10.2.220
- [23] FARIHA, M. A. Z., R. IKEURA, S. HAYAKAWA a S. TSUTSUMI. Analysis of Pan-Tompkins Algorithm Performance with Noisy ECG Signals. *Journal of Physics: Conference Series* [online]. 2020, 1532(1) [cit. 2022-12-30]. ISSN 1742-6588. Dostupné z: doi:10.1088/1742-6596/1532/1/012022
- [24] JAGTAP, Sonal K. a M. D. UPLANE. The impact of digital filtering to ECG analysis: Butterworth filter application. *International Conference on Communication, Information & Computing Technology (ICCICT)* [online]. IEEE, 2012, 2012, 1-6 [cit. 2023-05-25]. ISBN 978-1-4577-2078-9. Dostupné z: doi:10.1109/ICCICT.2012.6398145
- [25] D'AGOSTINO, Ralph a E. S. PEARSON. Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$. *Biometrika* [online]. Oxford University Press, 1973, 60(3), 613 - 622 [cit. 2023-05-26]. Dostupné z: <https://www.jstor.org/stable/2335012?origin=crossref>
- [26] KIM, Tae Kyun. T test as a parametric statistic. *Korean Journal of Anesthesiology* [online]. 2015, 68(6), 540–546 [cit. 2023-05-23]. ISSN 2005-6419. Dostupné z: doi:10.4097/kjae.2015.68.6.540
- [27] HART, A. Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ* [online]. 2001, 323(7309), 391-393 [cit. 2023-05-20]. ISSN 09598138. Dostupné z: doi:10.1136/bmj.323.7309.391
- [28] DAHIRU, T. P-Value, a true test of statistical significance? a cautionary note. *Annals of Ibadan Postgraduate Medicine* [online]. 2011, 6(1), 21-26 [cit. 2023-05-20]. ISSN 1597-1627. Dostupné z: doi:10.4314/aipm.v6i1.64038
- [29] BANERJEE, Amitav, UB CHITNIS, SL JADHAV, JS BHAWALKAR a S CHAUDHURY. Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal* [online]. 2009, 18(2), 127–131 [cit. 2023-05-20]. ISSN 0972-6748. Dostupné z: doi:10.4103/0972-6748.62274

- [30] HUBERTY, Carl J. Multiple Correlation Versus Multiple Regression. *Educational and Psychological Measurement* [online]. 2003, 63(2), 271-278 [cit. 2023-05-23]. ISSN 0013-1644. Dostupné z: [doi:10.1177/0013164402250990](https://doi.org/10.1177/0013164402250990)
- [31] BARATLOO, Alireza, Mostafa HOSSEINI, Ahmed NEGIDA a Gehad El Ashal EL ASHAL. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg (Tehran)* [online]. 2015, 3(2), 48-49 [cit. 2023-05-23]. PMID: PMC4614595. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4614595/>
- [32] SIDEY-GIBBONS, Jenni A. M. a Chris J. SIDEY-GIBBONS. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology* [online]. 2019, 19(1) [cit. 2023-05-23]. ISSN 1471-2288. Dostupné z: [doi:10.1186/s12874-019-0681-4](https://doi.org/10.1186/s12874-019-0681-4)
- [33] LI, Mujin, Honghui XU a Yong DENG. Evidential Decision Tree Based on Belief Entropy. *Entropy* [online]. 2019, 21(9) [cit. 2023-05-23]. ISSN 1099-4300. Dostupné z: [doi:10.3390/e21090897](https://doi.org/10.3390/e21090897)
- [34] GUO, Sheng, Qiang HU, Chun NG a C.T. LIU. More than entropy in high-entropy alloys: Forming solid solutions or amorphous phase. *Intermetallics*, [online]. 2013, (41), 96-103 [cit. 2023-05-23]. ISSN 0966-9795. Dostupné z: <https://doi.org/10.1016/j.intermet.2013.05.002>.
- [35] Scikit-learn: Machine Learning in Python, Pedregosa et al., *JMLR* 12, pp. 2825-2830, 2011. [cit. 2023-05-23]. Dostupné z: <https://scikit-learn.org/stable/index.html#>
- [36] PREGIBON, Daryl. Logistic Regression Diagnostics. *The Annals of Statistics* [online]. 1981, 9(4) [cit. 2023-05-23]. ISSN 0090-5364. Dostupné z: [doi:10.1214/aos/1176345513](https://doi.org/10.1214/aos/1176345513)
- [37] STOLTZFUS, Jill C. Logistic Regression: A Brief Primer. *Academic Emergency Medicine* [online]. 2011, 18(10), 1099-1104 [cit. 2023-05-27]. ISSN 10696563. Dostupné z: [doi:10.1111/j.1553-2712.2011.01185.x](https://doi.org/10.1111/j.1553-2712.2011.01185.x)
- [38] HUANG, Shujun, Nianguang CIA, Pedro Penzuti PACHECO, Yang WANG a Wayne XU. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics & Proteomics* [online]. 2018, 15(1), 41-51 [cit. 2023-05-23]. ISSN 17906245. Dostupné z: [doi:10.21873/cgp.20063](https://doi.org/10.21873/cgp.20063)
- [39] SIDHARTH. The RBF kernel in SVM: A Complete Guide. *PyCodeMates* [online]. 2022 [cit. 2023-05-27]. Dostupné z:

https://www.pycodemates.com/2022/10/the-rbf-kernel-in-svm-complete-guide.html?utm_content=cmp-true

- [40] KADIYALA, Akhil a Ashok KUMAR. Applications of python to evaluate the performance of bagging methods. Environmental Progress & Sustainable energy [online]. 2018, 37(5), 1555-1559 [cit. 2023-05-23]. ISSN 1944-7442. Dostupné z: doi:10.1002/ep.13018

Zoznam symbolov a skratiek

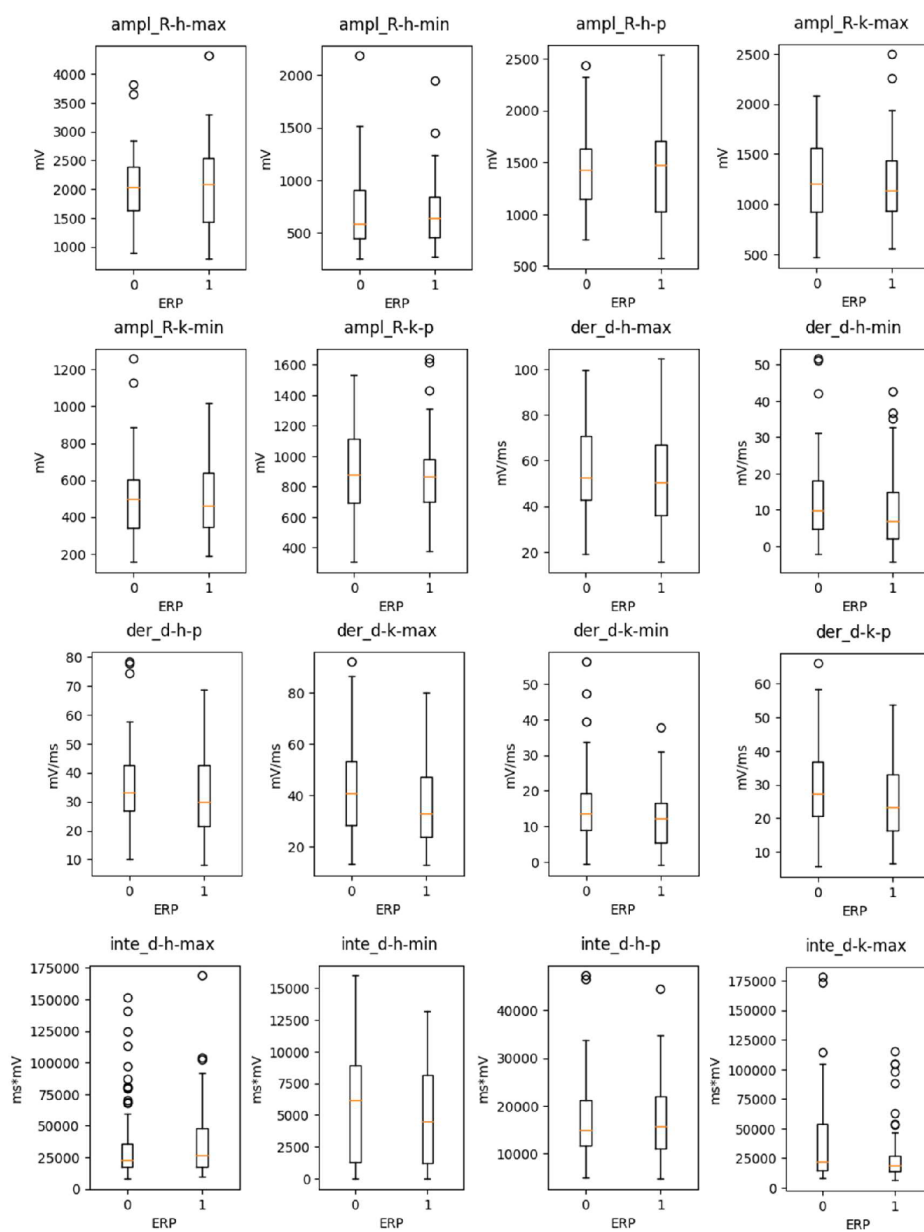
WPW	Wolf-Parkinson-White
EKG	Elektrokardiogram
SA	Sinoatriálny
AV	Atrioventrikulárny
mV	Milivolt
ERP	Efektívna refraktérna perióda
AVRT	Atrioventrikulárna recipročná tachykardia
EF	Elektrofyzologická
AP	Accessory pathway - prídavná dráha
iso	Izoproterenol
kHz	Kilohertz
μV	Mikrovolt
lsb	Least significant bit - menej významný bit
FN	Fakultní nemocnice
ms	Milisekunda
ant	Anterográdne
ret	Retrográdne
SVM	Support vector machine - Model podporných vektorov

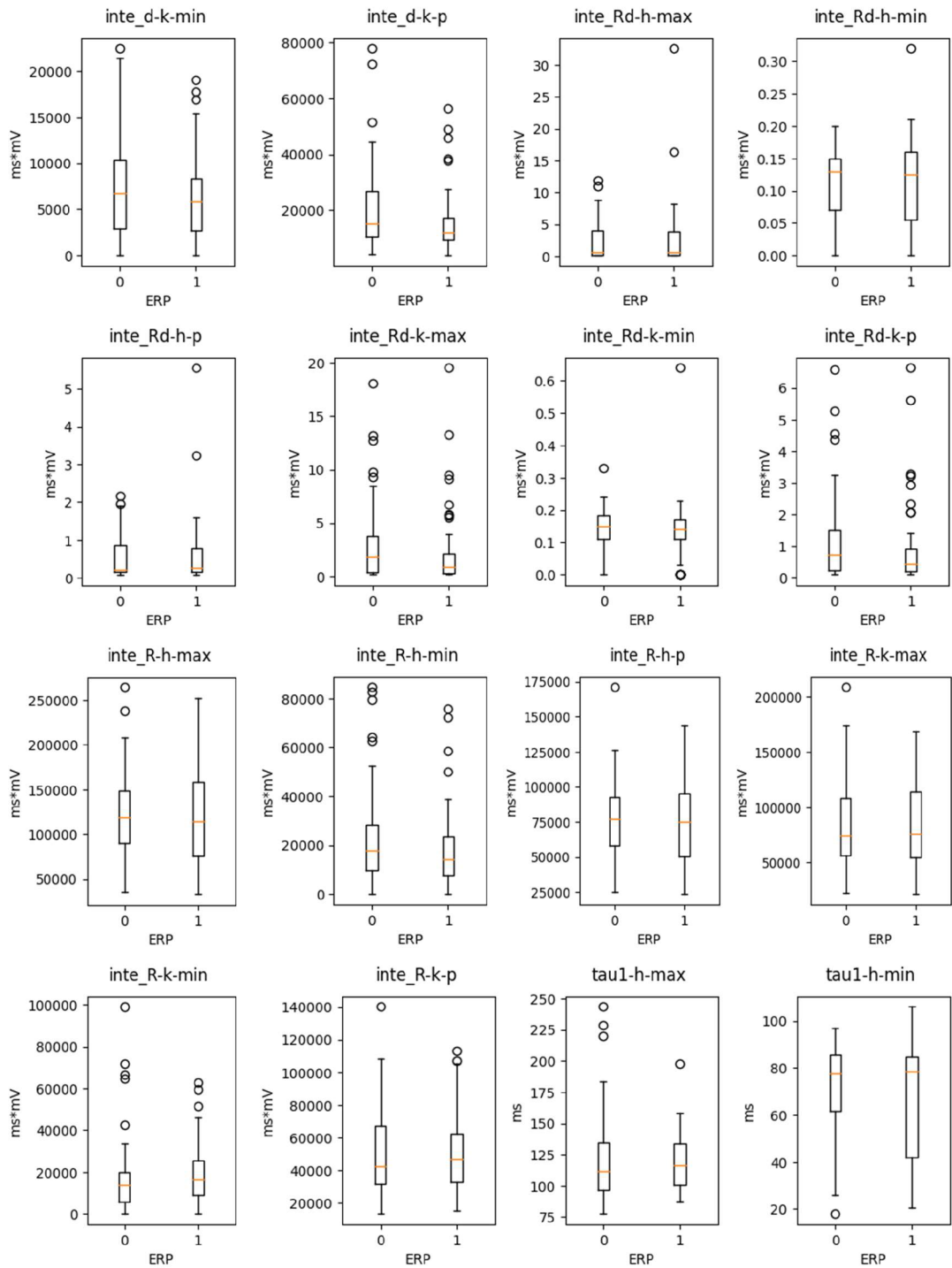
Zoznam príloh

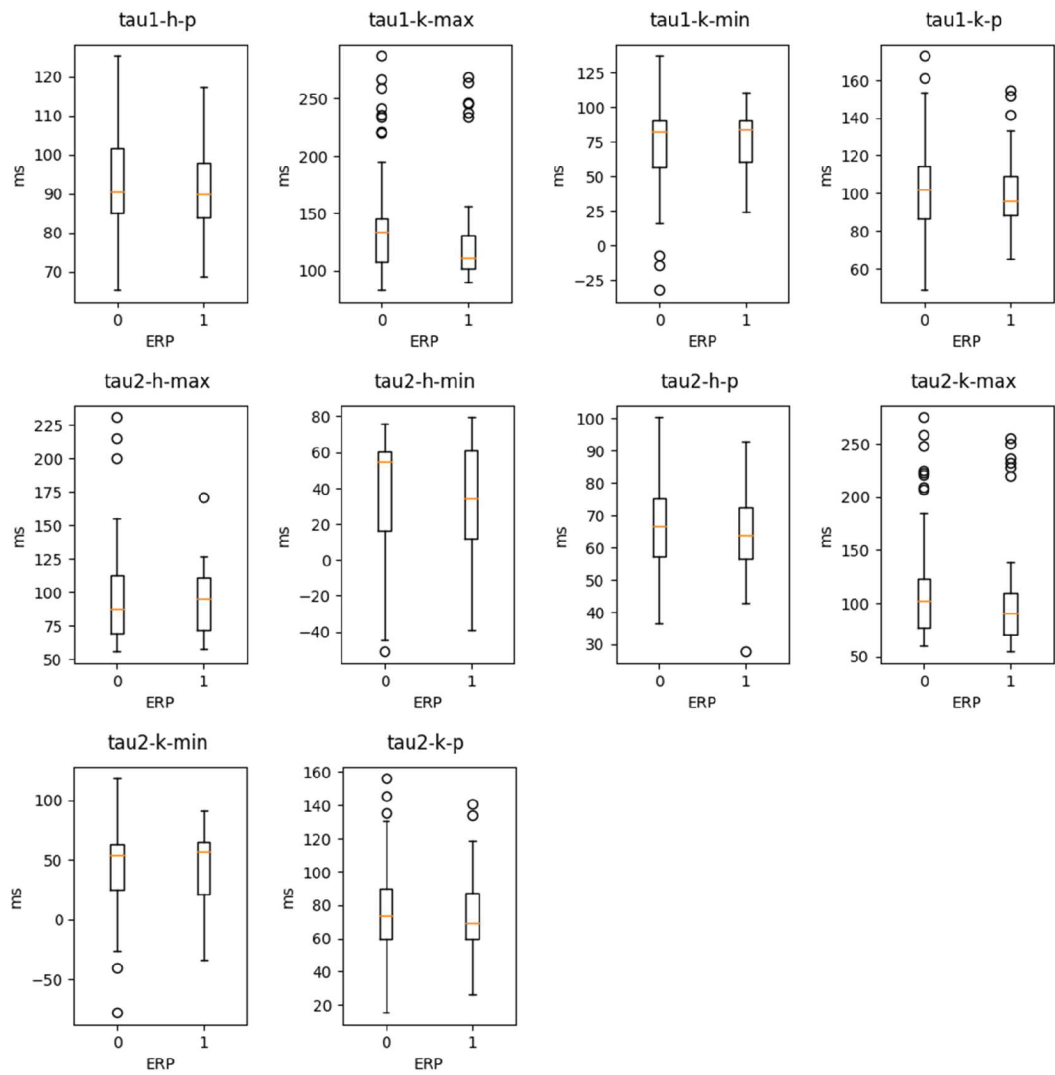
A	Výsledky	45
A.1	Porovnanie rozloženia dát biomarkerov pre rizikových a nerizikových pacientov	45
B	Obsah elektronickej prílohy	48

A Výsledky

A.1 Porovnanie rozloženia dát biomarkerov pre rizikových a nerizikových pacientov







B Obsah elektronickej prílohy

```
/.....koreňový adresár priloženého archívu
├── Programy.....programy v jazyku Python
│   ├── Extrakcia_priznakov .....spracovanie biomarkerov
│   │   ├── agregacia.py
│   │   ├── extrakcia.py
│   │   └── statisticke_testovanie_biomarkerov.py
│   ├── Korelacia_signalov .....korelácia signálov s delta vlnou
│   │   ├── main.py
│   │   ├── correlation_matrix.py
│   │   └── sel_parser.py
│   └── Modely_predikcie .....tvorba modelov predikcie
│       └── modely_predikcie.py
├── Modely .....natrénované modely (posledné 2 čísla v názve určujú presnosť)
│   ├── rozhodovaci_strom_56.sav
│   ├── logisticka_regresia_68.sav
│   ├── svm_74.sav
│   ├── bagging_rozhodovaci_strom_59.sav
│   ├── bagging_logisticka_regresia_68.sav
│   └── bagging_svm_68.sav
├── agregovanie_data.csv .....ukážka výstupu agregácie biomarkerov
└── model_data.csv .....dataset biomarkerov použitý k tvorbe modelov
```