



**ČESKÉ VYSOKÉ  
UČENÍ TECHNICKÉ  
V PRAZE**

**F3**

**Fakulta elektrotechnická  
Katedra teorie obvodů**

**Diplomová práce**

# **Klasifikace parasomních epizod z 3-osých aktigrafických záznamů**

**Mariana Castillo**

**Květen 2021**

**Studijní program: Lékařská elektronika a bioinformatika**

**Vedoucí práce: Ing. Eduard Bakštein, Ph.D.**



## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Castillo** Jméno: **Mariana** Osobní číslo: **465456**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra teorie obvodů**  
Studijní program: **Lékařská elektronika a bioinformatika**  
Specializace: **Lékařská technika**

## II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

**Klasifikace parasomních epizod z 3-osých aktigrafických záznamů**

Název diplomové práce anglicky:

**Classification of Parasomnia Episodes Using 3-axial Wrist Accelerometry**

Pokyny pro vypracování:

- Detekce parasomních epizod u pacientů s poruchou spánku probíhá manuálně a je problematická a zdoluhavá. Výskyt konkrétních epizod přitom může výrazně usnadnit diagnostiku pacienta a automatická detekce epizod by tak významně přispěla rychlejší a snadnější diagnostice. Cílem projektu je navrhnout a implementovat metodu automatické klasifikace parasomních epizod ze záznamu 3D aktigrafie (záznam pohybové aktivity zápěstí ve 30 nebo 100Hz a třech osách).
1. Nastudujte problematiku předzpracování a klasifikace aktigrafických signálů.
  2. Zkompletujte a předzpracujte zadaný dataset, obsahující anotované aktigrafické a polysomnografické záznamy pacientů s parasomnií.
  3. Proveďte explorační analýzu a identifikujte vhodné příznaky pro rozlišení a) oblastí klidu a aktivity, a b) parasomních a fyziologických epizod. Příznaky vytvořte z dat agregovaných po epochách i hrubých třiosých aktigrafických dat.
  4. Vytvořte klasifikátor pro rozlišení parasomních a fyziologických epizod a evaluujte jeho vlastnosti na zadaném datasetu, vč. realistického zhodnocení reálného nasazení vytvořeného algoritmu.

Seznam doporučené literatury:

- [1] Sadeh, A., & Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews*, 6(2), 113–124. <https://doi.org/10.1053/smr.2001.0182>
- [2] Yuda, E., Yoshida, Y., Sasanabe, R., Tanaka, H., Shiomi, T., & Hayano, J. (2017). Sleep Stage Classification by a Combination of Actigraphic and Heart Rate Signals. *Journal of Low Power Electronics and Applications*, 7(4), 28. <https://doi.org/10.3390/jlpea7040028>
- [3] Náhlík, M. (2019): Detekce cyklu spánků-bdění a spánkových stádií z aktigrafického záznamu, ČVUT v Praze, diplomová práce
- [4] Fois, C., Wright, M. A. S., Sechi, G. Pietro, Walker, M. C., & Eriksson, S. H. (2015). The utility of polysomnography for the diagnosis of NREM parasomnias: an observational study over 4 years of clinical practice. *Journal of Neurology*, 262(2), 385–393. <https://doi.org/10.1007/s00415-014-7578-2>

Jméno a pracoviště vedoucí(ho) diplomové práce:

**Ing. Eduard Bakštejn, Ph.D., Analýza a interpretace biomedicínských dat FEL**

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

Datum zadání diplomové práce: **05.01.2021** Termín odevzdání diplomové práce: **21.05.2021**

Platnost zadání diplomové práce: **30.09.2022**

Ing. Eduard Bakštejn, Ph.D.  
podpis vedoucí(ho) práce

doc. Ing. Radoslav Bortel, 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

## Poděkování / Prohlášení

Ráda bych poděkovala svému vedoucímu diplomové práce Ing. Eduardovi Bakšteinovi, Ph.D., za trpělivost, cenné rady a dobré vedení a MUDr. Jitce Bušové, Ph.D., za spolupráci a odborné rady.

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 21. 5. 2021

.....  
Mariana Castillo

## Abstrakt /

Poruchy spánku jsou utrpením mnoha lidí, přestože je spánek jedním z nejdůležitějších faktorů pro udržení dobrého zdraví. NREM parasomnie je porucha, jejíž diagnostika je typicky založena na polysomnografii, což je velmi komplexní a drahá metoda. Parasomnie se navíc nemusí během vyšetření projevit. Alternativní metodou pro vyhodnocení a klasifikaci parasomnie může být 3D aktigrafie, která by přinesla finančně přístupnější možnosti a také možnost dlouhodobého nahrávání.

V této práci byly explorační analýzou identifikovány příznaky z aktigrafických a polysomnografických dat a následně klasifikovány pomocí různých modelů. Data byla statisticky porovnána s expertními anotacemi parasomní a fyziologické aktivity. Anotace znázorňovaly čas počátku aktivity, kdy nejdříve byly k dispozici s přesností na minutu a později doplněny kvůli přesnosti na vteřinu.

V první části práce byly zpracovány aktigrafické záznamy s využitím minutových anotací. Na vytvořených příznacích byla data klasifikována pomocí SVM (*Support vector machine*) modelu. Přesnost těchto dat se pohybovala okolo 67 % se směrodatnou odchylkou 6 %. Senzitivita tohoto modelu se však pohybovala jen okolo 10 % se směrodatnou odchylkou 16 %.

Proto byly následně doplněny anotace s přesností na vteřinu a do analýzy byly zahrnuty i vybrané polysomnografické signály (EMG, EKG). Příznaky byly vypočítány po dvouvteřinových oknech. Analýza příznaků v čase prokázala, že nejužitečnější informaci o rozlišení fyziologické a parasomní aktivity obsahuje prvních několik vteřin od počátku aktivity. Klasifikací pomocí *Random forest* modelu bylo zjištěno, že nejvyšší schopnost rozlišení nese biologický signál EMG a těsně pod

ním aktigrafický signál. Celková přesnost klasifikace modelu byla 83 % a se senzitivitou 50 %.

Kromě klasifikace událostí získaných z anotací byla implementována detekce v celonočním záznamu. Byly vypočteny příznaky stejné jako při zpracování vteřinových anotací a klasifikovány na stejném modelu (RF). Celková senzitivita se průměrně pohybovala okolo 42 % (směrodatná odchylka 46 %) a pozitivní predikce 5 % (směrodatná odchylka 6 %).

Pro další studie byl vytvořen datový set z aktigrafických a polysomnografických dat obsahující události okolo anotací.

**Klíčová slova:** parasomní epizody; polysomnografie; aktigrafie; detekce; extrakce příznaků; klasifikace.

## Abstract /

Sleep disorders represent a burden for many people, although sleep is one of the most important factors for a good health. NREM parasomnia is a disorder that is very complex and expensive to diagnose and may not exhibit at all during the diagnostic polysomnography. An alternative method for the evaluation and classification of parasomnia can be 3D actigraphy, which could bring more financially accessible possibilities, as well as the possibility of long-term recording.

In this work, the characteristics of actigraphic and polysomnographic data were identified by exploratory analysis and subsequently classified using various models. The data were statistically evaluated with annotations of parasomnic and physiological activity. The annotations showed the start time of the activity, when they were first available with an accuracy of one minute and later supplemented for an accuracy to one second.

In the first part of the work, actigraphic records were processed using minute annotations. Using a set of actigraphy-based features, the data were classified using the SVM (*Support vector machine*) model. The accuracy of this method was around 70 % with a standard deviation of 6 %. However, the sensitivity of this model was around 9 % with a standard deviation of 11 %.

Therefore, annotation with a precision of one second was subsequently added and selected polysomnographic signals (EMG and ECG) were included in the analysis. Features were calculated after two-second windows. Analysing the features over time appeared that the most useful information on the distinction between physiological and parasomnic activity contained

the first few seconds from the onset of activity. By classification using the *Random forest* model, it was found that the highest resolution is carried by biological signal EMG and just below it by the actigraphic signal. The overall accuracy of the model classification was 83 % and the sensitivity 50 %.

In addition to the classification of events obtained from annotations, parasomnia detection was implemented in the overnight record. The same features as in the processing of the seconds annotations were calculated and classified on the same model (RF). The overall sensitivity was on average around 42 % (standard deviation 46 %) and the positive predictive value 5 % (standard deviation 6 %).

For further studies, a data set was created from actigraphic and polysomnographic records containing events around annotations.

**Keywords:** parasomnia episodes; polysomnography; actigraphy; detection; features extraction; classification.

**Title translation:** Classification of Parasomnia Episodes Using 3-axial Wrist Accelerometry

# Obsah /

<b>1 Úvod</b> .....	1
1.1 Parasomnie .....	1
1.2 Polysomnografie .....	2
1.3 3D Aktigrafie .....	3
1.4 Přehled současného stavu .....	5
1.4.1 Obecné využití aktigrafie ..	5
1.4.2 Podobné přístupy .....	5
1.5 Cíle práce .....	6
<b>2 Metody</b> .....	7
2.1 Naměřená data .....	7
2.1.1 Aktigrafické záznamy .....	7
2.1.2 Polysomnografické zá- znamy .....	8
2.2 Klasifikátory .....	8
2.2.1 SVM klasifikátor .....	8
2.2.2 Random forest klasifi- kátor .....	10
2.3 Návrh vlastního řešení .....	12
2.3.1 Předzpracování dat s minutovými anotacemi ..	12
2.3.2 Definice příznaků pro předzpracování minu- tových anotací .....	15
2.3.3 Předzpracování dat s vteřinovými anotace- mi .....	17
2.3.4 Definice příznaků pro předzpracování vteři- nových anotací .....	17
2.3.5 Detekce parasomnie v celonočním záznamu ...	20
2.4 Statistické vyhodnocení .....	22
2.4.1 AUC .....	23
<b>3 Výsledky</b> .....	24
3.1 Předzpracování dat .....	24
3.1.1 Předzpracování dat s minutovými anotace- mi .....	24
3.1.2 Předzpracování dat s vteřinovými anotace- mi .....	25
3.2 Klasifikace dat s minutovými anotacemi .....	28
3.3 Klasifikace dat s vteřinovými anotacemi .....	29
3.4 Detekce parasomnie v celo- nočním záznamu .....	30
<b>4 Diskuze</b> .....	31
4.1 Předzpracování dat .....	31
4.2 Klasifikace předzpracova- ných dat .....	32
4.3 Detekce parasomnie v celo- nočním záznamu .....	33
<b>5 Závěr</b> .....	34
<b>Literatura</b> .....	35
<b>A Zkratky a symboly</b> .....	39
A.1 Zkratky .....	39
A.2 Symboly .....	39
<b>B Soubory</b> .....	40
B.1 Přílohy diplomové práce .....	40
B.2 Soubory zdrojového kódu .....	40
B.3 Soubory vytvořeného dato- vého setu .....	41



## Tabulky / Obrázky

<b>3.1.</b> Statistické vyhodnocení SVM klasifikátoru trénovacích dat. ....	28
<b>3.2.</b> Statistické vyhodnocení SVM klasifikátoru testovacích dat. ....	29
<b>3.3.</b> Statistické vyhodnocení RF klasifikátoru trénovacích dat. ..	29
<b>3.4.</b> Statistické vyhodnocení RF klasifikátoru testovacích dat. ..	30
<b>3.5.</b> Statistické vyhodnocení <i>Random forest</i> klasifikátoru při detekci v celonočním záznamu. ....	30
<b>1.1.</b> Monitorace pacienta pomocí PSG metody. ....	3
<b>1.2.</b> Náhled PSG záznamu. ....	3
<b>1.3.</b> Náhled hypnogramu. ....	3
<b>1.4.</b> Aktigraf. ....	4
<b>1.5.</b> Mikrostruktura kapacitního akcelerometru ....	4
<b>2.1.</b> Hrubá data. ....	8
<b>2.2.</b> SVM. ....	9
<b>2.3.</b> SVM: mapování dat do prostoru o vyšší dimenzi. ....	10
<b>2.4.</b> Random forest algoritmus. ....	11
<b>2.5.</b> Filtrace hrubých dat. ....	12
<b>2.6.</b> Agregace filtrovaných dat. ....	13
<b>2.7.</b> Vypočtené příznaky agregovaného signálu. ....	13
<b>2.8.</b> Histogram všech anotovaných epizod. ....	14
<b>2.9.</b> Detekce aktivních segmentů. ..	15
<b>2.10.</b> Schéma předzpracování signálu s minutovými anotacemi. .	16
<b>2.12.</b> Schéma předzpracování signálu s vteřinovými anotacemi. .	17
<b>2.11.</b> Událost z ACG a PSG dat. ....	18
<b>2.13.</b> Měření výkonové spektrální hustoty. ....	19
<b>2.14.</b> Zpracování EKG signálu. ....	20
<b>2.15.</b> AUC hodnoty v závislosti na kvantilu čtyřvteřinového segmentu. ....	21
<b>2.16.</b> Schéma předzpracování signálu pro detekci parasomní aktivity. ....	22
<b>2.17.</b> ROC křivka. ....	23
<b>3.1.</b> Histogram příznaků. ....	25
<b>3.2.</b> ROC křivka příznaků. ....	25
<b>3.3.</b> AUC hodnoty příznaků maximální úhel a maximální rozdíl dvou os aktigrafického signálu. ....	26
<b>3.4.</b> AUC hodnoty výkonové spektrální hustoty ACG signálu. ....	26
<b>3.5.</b> AUC hodnoty výkonové spektrální hustoty EMG signálu brady. ....	27

<b>3.6.</b>	AUC hodnoty výkonové spektrální hustoty EMG signálu končetin. ....	27
<b>3.7.</b>	AUC hodnoty příznaku maximální rozdíl RR intervalu. ...	28

# Kapitola 1

## Úvod

Kvalitní spánek je jedním z důležitých faktorů pro dobré zdraví. Poruchy spánku mohou výrazně ovlivnit odpočinek mozku a tím mimo jiné i emocionální stav člověka. Poruchy spánku lze popsat různými parametry, jako je např. latence spánku, počet a trvání nočních probuzení, celková doba spánku, stav rychlého pohybu očí (REM) a bez rychlého pohybu očí (non-REM). Všechny tyto indikátory lze měřit sledováním různých fyziologických parametrů během spánku, jako je např. dýchací frekvence, srdeční frekvence, elektrická aktivita mozku (EEG), fyzická aktivita apod.

Jednou z poruch spánku je parasomnie, která snižuje kvalitu spánku. Má závažné dopady na psychický stav jedince, který může trpět nespavostí až náměsíčností. Její diagnostika je časově a finančně náročný proces. Tato práce se zabývá možnostmi automatického hodnocení aktigrafického a polysomnografického záznamu. Cílem této práce je vyvinout klasifikátor pro automatické rozeznání parasomních a fyziologických epizod na základě aktigrafických dat. Před klasifikací je potřeba data předzpracovat pomocí explorační analýzy a identifikovat tak příznaky pro rozlišení jednotlivých aktivit.

V úvodu diplomové práce jsem se zaměřila na přiblížení problematiky týkající se poruchy parasomnie. Následující kapitoly objasňují, jak je porucha charakterizovaná, jak se v dnešní době diagnostikuje a její porovnání se zvolenou alternativní metodou - 3D Aktigrafie. Na závěr je zařazen přehled současného stavu v oblasti 3D Aktigrafie.

## 1.1 Parasomnie

Parasomnie je porucha spánku charakterizovaná jako akutně probíhající abnormální epizodické nepředvíratelné jevy ve spánku. Projevuje se ve smíšeném stavu spánku a bdělosti, kdy jedinec je schopen vykonávat řadu automatických pohybů a jednání, ale není schopen je plným vědomím korigovat a řídit. Parasomnie může být způsobena poruchou koordinace biologických a psychologických procesů v organismu, resp. v centrálním nervovém systému, vznikajících v průběhu spánku. Často mohou být provokovány stresem. [1]

Obecně jsou tyto epizody zcela benigní povahy, méně často mohou být komplexní a dramatické, ev. spojené s nebezpečným chováním [2]. Jsou charakterizovány abnormálním chováním nebo fyziologickými projevy (např. chůze). Mezi klinické příznaky spadá probuzení se zmateností, somnambulismus (náměsíčnost) a noční děsy projevující se epizodou děsu s výkřikem či pláčem [3]. Pohyby jsou nekoordinované, řeč je nesrozumitelná. Pacient během epizody nereaguje na okolní svět a při snaze o probuzení může být agresivní [1].

Parasomnie se vyskytují převážně v dětském věku, ale postihují rovněž 2-4 % dospělých [4].

Tato porucha spánku (tak jako i všechny ostatní) jistým způsobem ovlivňuje život postiženého v jeho pracovních i mimopracovních činnostech a má negativní vliv na celkový výkon jedince. Cílená statistika úrazů, dopravních nehod, pracovní neschopnosti,

kteřá vzniká důsledkem zvýšené nespavosti není známá ani sledována. Některé z charakteristik parasomnie však mluví samy za sebe, jako např. somnambulismus (náměsíčnost) či noční děsy mohou znamenat jisté nebezpečí, jiné (skřípání zuby ve spánku - bruxismus) mohou při dlouhodobém trvání poškodit některé části organismu. Z výčtu těchto nepříjemností je tedy zřejmé, že je diagnostika a léčba poruchy vhodná. [5]

## 1.2 Polysomnografie

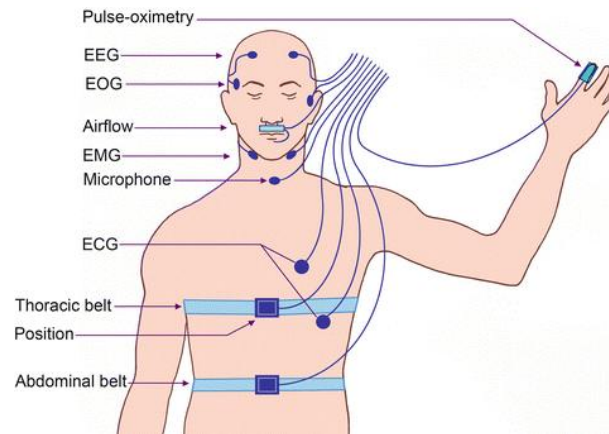
Zlatým standardem diagnostiky parasomnie je v dnešní době využívána tzv. video-polysomnografie, která prokazuje probouzečící reakce v elektroencefalogramu (EEG), či probuzení z hlubokého NREM spánku. Provedení polysomnografie (PSG) slouží též k vyloučení výskytu specifických poruch spánku typu obstrukční spánková apnoe (OSA) či periodické pohyby končetinami ve spánku (periodic limb movements in sleep - PLMS). V případě nejednoznačného klinického obrazu se provádí současně video-EEG monitorace k vyloučení noční epilepsie. Epizody bývají stereotypní, mohou se vyskytovat i několikrát za noc a nebývají vázány na první spánkový cyklus. [6]

Podstatou PSG je souběžné nahrávání více biologických signálů najednou (viz Obr. 1.1). Jde například o tyto: elektroencefalogram (EEG), elektromyogram brady a končetin (EMG), elektrokardiografie (EKG), elektrookulogram (EOG), hrudní dýchání, saturace krve kyslíkem ( $SpO_2$ ) a proud vzduchu před nosem a ústy. Příklad záznamu PSG lze vidět na Obr. 1.2. Následuje vyhodnocení hypnogramu, které je potřeba manuálně označit (viz Obr. 1.3). Hypnogram je grafický záznam spánkových stádií (bdění, REM, NREM 1-3), která jsou specifická EEG aktivitou.

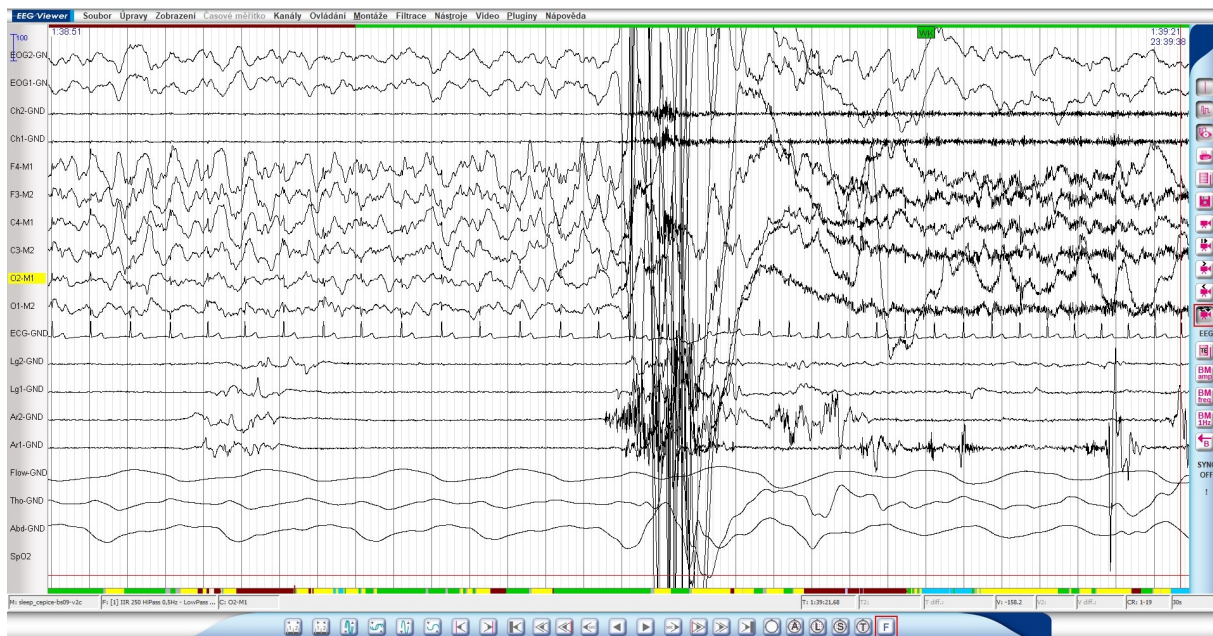
Toto standardní hodnocení spánku se provádí profesionálně ve spánkové laboratoři. Proto má použití PSG své limity, jako je laboratoř s potřebným vybavením a vyškoleným personálem. Toto vybavení je finančně nákladnější. Kromě toho může být chování spánku člověka ovlivněno novým prostředím. Po nahrání jedné noci je potřeba manuální zhodnocení záznamu expertem. Právě tato omezení motivují k hledání alternativních metod dlouhodobého hodnocení spánku doma s automatickým vyhodnocením.

Existuje studie [7], která zjišťuje užitečnost polysomnografie pro diagnostiku NREM parasomnie. Výstupem je, že PSG poskytuje užitečné diagnostické informace celkem u 65,5 % pacientů s podezřením na parasomnii. Je zde i potvrzeno, že videomonitorace má vysokou hodnotu při tomto vyšetření, a to až u 76,5 % studovaných pacientů.

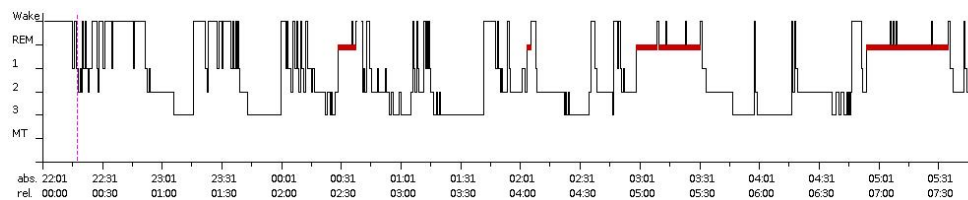
Pacienti, kteří jsou vyšetřováni pro podezření na parasomnii, často podstoupí záznam více než jednu noc, aby se zvýšila pravděpodobnost zachycení událostí. Toto je časově i finančně velmi náročné a také to vede ke zvýšení nepohodlí pro pacienty. Výhoda delších nahrávek nebyla nikdy dostatečně prozkoumána. V testované skupině pacientů ve studii [7] bylo zjištěno, že druhá noc zřídka poskytla jakékoli další informace, protože podíl pozitivních testů u lidí, kteří podstoupili jednu noc záznamu (70 %), byl podobný jako u lidí se dvěma nocemi PSG (64 %). [7]



**Obrázek 1.1.** Monitorace pacienta pomocí PSG metody. Převzato z [8].



**Obrázek 1.2.** Náhled polysomnografického záznamu. Převzato od MUDr. Jitky Buškové, Ph.D.



**Obrázek 1.3.** Náhled vyhodnoceného hypnogramu. Převzato od MUDr. Jitky Buškové, Ph.D.

## 1.3 3D Aktigrafie

Triaxiální aktigrafie se v dnešní době používá ke studii spánku-probuzení [9]. Výhodou aktigrafie oproti tradiční polysomnografii (PSG) je, že může pohodlně zaznamenávat

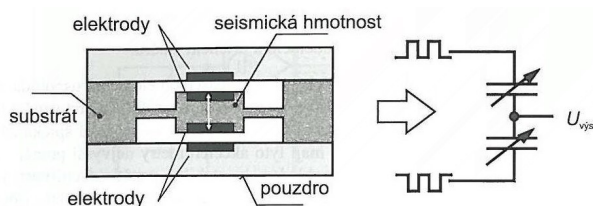
signál v nepřetržitém režimu 24 hodin denně po několik dní, týdnů nebo i déle. V budoucnu pak může jít o nákladově efektivní metodu pro hodnocení konkrétních poruch spánku. Nevýhodou ale zůstává to, že není tak účinná při detekci poruch, které nezahrnují pohyb končetin, jelikož zaznamenává pouze zlomek dat, který obsahuje PSG. Přesto má tato metoda velký potenciál sledovat kvalitu spánku. Je třeba poznamenat, že tento systém nevyžaduje sběr obrazových dat, vstupy ani analýzy od vyškolených odborníků.

Aktigraf (viz Obr. 1.4) je malé, lehké a neinvazivní zařízení umístěné na zápěstí využívané pro měření záznamu pohybu ve třech ortogonálních rovinách. Zařízení lze umístit i na jiné části těla, ale dle studie o umístění aktigrafu na těle [10] je nedominantní zápěstí ideálním místem pro analýzu. Shromážděná data jsou stažena do počítače pro zobrazení a analýzu aktivity/nečinnosti. Mezi výhody nositelných senzorů patří, že se subjekt může chovat přirozenějším a spontánnějším způsobem, není potřeba speciální místnosti a vyhodnocení experta.



**Obrázek 1.4.** Aktigraf. Převzato z [11].

Hlavní součástí aktigrafu je triaxiální akcelerometr, který poskytuje měření ve třech navzájem kolmých směrech (vertikální, anteroposteriorní a mediolaterální) v reálném čase až do 100 Hz [12]. Akcelerometr pracuje na principu, ve kterém je hmotnostní element (často označován jako seismická hmota) buzen v systému hmota-pružina-tlumení. Kapacitní akcelerometr (viz Obr. 1.5) měří tuto výchylku seismické hmoty kapacitně. Snímací prvek se skládá ze dvou elektrod, přičemž jedna z nich je pohyblivá v závislosti na zrychlení. Při přiblížení elektrod se kapacita zvětší a naopak. Vyhodnocuje se rozdíl hodnot kapacit. [13]



**Obrázek 1.5.** Mikrostruktura kapacitního akcelerometru. Převzato z [14].

Parametry kapacitních akcelerometrů mají frekvenční rozsah obvykle mezi 0-6 kHz [14]. Pro hodnocení lidské pohybové aktivity se však obvykle používá aktivita ve frekvenčním rozsahu do několika jednotek Hertz.

I přesto, že aktigrafie není tak přesná jako metoda PSG, studie se obecně shodují na tom, že aktigrafie je díky své schopnosti nepřetržitě zaznamenávat po dlouhou dobu spolehlivější než spánkové dotazníky, které jsou další rozšířenou metodou ve spánkové diagnostice [15] (viz také Kapitola 1.4.2).

Pro subjektivní hodnocení kvality spánku se používá také tzv. Pittsburský index kvality spánku, který obsahuje 19 otázek s vlastním hodnocením. Skóre každé složky se pohybuje od 0 do 3. [16] Součet dílčích skóre pak poskytuje globální skóre poruch spánku. Vyšší hodnoty označují závažnější poruchy spánku. Zpětné hodnocení kvality, které vyplňuje pacient, spoléhá se na to, že si pacient pamatuje předchozí noc. Toto hodnocení je však pro jednotlivce subjektivní, což vede k méně spolehlivému výsledku. [9, 17].

Jelikož si jedinci často ani nejsou vědomi, že se v noci několikrát probudili, sebehodnocení spánku není pro případ parasomnie nejlepší variantou. Velkou výhodou, která aktigrafie oproti PSG poskytuje, je finanční dostupnost. Použitím aktigrafie by mohly klesnout náklady až na 5 % oproti stávajícímu postupu PSG.

## 1.4 Přehled současného stavu

### 1.4.1 Obecné využití aktigrafie

Díky pokroku informační a komunikační technologie se zvyšuje využití technologických řešení v medicíně. Na trhu se zlepšuje dostupnost zařízení, jako jsou např. nositelná čidla (např. akcelerometry) a ceny senzorů se díky širokému rozšíření chytrých telefonů velmi snižují.

Aktigrafie je v poslední době hojně se rozvíjející metoda, jak zaznamenávat aktivitu člověka. Velmi často se využívá na sledování každodenní fyzické aktivity člověka jako je například chůze, běh, sportovní aktivita apod. Pomocí inteligentních nositelných zařízení lze také sledovat aspekty spánku a analýzu jeho kvality. Přesto se dnes velmi zřídka používají v klinických studiích. Poslední studie prokazují, že výstupy inteligentních zařízení, co se týče spánku, jen slabě korelují s výstupem PSG [18].

Tato alternativní metoda se rozvíjí různými směry. Některé studie se zabývají zpracováním fyzické aktivity denního života, jiné zase specifickými aktivitami tělesného postižení. Detekce aktivity během spánku je též jedna z možností.

Aktigrafie je metoda vhodná pro detekci spánků-bdění (nehybnost účastníka znamená začátek spánkového období) [19–20], ale dle studie [18, 21] není nejlepší metodou pro detekci spánkových stádií.

Analýzou pohybu je možné se zaměřit na specifické tělesné postižení a kvantitativně vyhodnotit specifické pohyby. Ve studii [22] byl zkoumán pohyb horních končetin u postižení mozkové obrny. V závěru práce byl popsán významný rozdíl aktivit horních končetin u malých dětí. Velkého potenciálu by tato metoda mohla dosáhnout při analýze nočního pohybu u Parkinsonovy choroby [23]. Další studie [24] využívají metodu k odhadu aktivity energetického výdeje (energy expenditure - EE). Dle výsledků je odhad EE pomocí aktigrafie lepší než metoda nepřímé kalorimetrie.

Některé studie se snaží využít aktigrafii zároveň s dalším fyziologickým parametrem jako je např. variabilita srdečního rytmu (heart rate variability - HRV) [25]. Tato studie též porovnávala rozdíl mezi aktigrafii na zápěstí a na hrudníku. Dle výstupu (např. specifická a přesnost) bylo zjištěno, že aktigrafické záznamy jsou užitečnějším signálem než jen samotné HRV. Kombinací aktigrafu na zápěstí a HRV parametru se výsledky pro analýzu spánku významně nelišily. [25]

### 1.4.2 Podobné přístupy

I přes existenci velkého počtu aktigrafických studií zabývajících se aktivitou člověka, pouze minimální část z nich se věnuje aktivitě během spánku. Pro detekci spánků-

bdění se však využívá nejméně 40 let [26]. Novější studie naznačují, že je velmi vhodná pro zkoumání variability spánku [27, 25].

Aktigrafie během spánku lze dělit na detekci spánek-bdění nebo měření konkrétní spánkové aktivity. Využit se dají tzv. chytré náramky, které používají pohybový senzor. Ze získaných dat je určeno několik parametrů, např. délka usínání a spánku, detekce a počet nočních probuzení spolu s přehledem jednotlivých stádií spánku.

Z několika studií vyplývá, že detekce spánek-bdění založená na aktivitě aktigrafu se velmi shoduje s detekcí založenou na tradičním hodnocení z EEG signálu. [19–20, 26]. Ve studii [28] dospěli k závěru, že normální subjekty vykazují více než 90 % korelaci při srovnání aktigrafie s metodou PSG.

Existují studie na měření dalších parametrů (např. nástup spánku, počet probuzení apod.). V porovnání se spánkovými dotazníky lze ze studie [15] říci, že výsledky obou metod detekce spánek-bdění silně korelují. Podobné parametry byly použity při analýze spánku u pacientů se spánkovou apnoe. Parametry jako celková doba spánku, účinnost spánku, latence nástupu spánku apod. odhadnuté z aktigrafického záznamu poskytují dobré výsledky pro budoucí práce [29].

Studie [30] pracovala s daty ze zařízení AppleWatch, ze kterého byla získána srdeční frekvence, zrychlení a tzv. „hodiny proxy“, představující vstup z bdění do režimu spánku a snaží se aproximovat měnící se chod cirkadiánních hodin. Výsledky byly porovnány s metodou PSG. Ke klasifikaci fází spánku (bdění-spánek nebo bdění-REM-NREM) využili několik modelů: logistická regrese, algoritmus k-nejbližších sousedů, Random forest a neuronové sítě. Nejpresnější určení spánkových stádií měl klasifikátor neuronových sítí při zahrnutí všech tří příznaků. Rozdíly oproti ostatním klasifikátorům však byly malé. Přesnost klasifikace spánek-bdění nejlepšího klasifikátoru byla 90 %.

Mimo aktigrafická data lze k vyhodnocení využít i další parametry. Ve studii [31] byly použity srdeční signály (HRV parametr) z tříminutových epoch. Klasifikace fází spánku (bdění, REM, NREM) byla určena s přesností okolo 75 % v kombinaci aktigrafických a srdečních signálů.

Z výše jmenovaných studií se jeví, že má aktigrafie velký potenciál při měření spánkové aktivity. Nevýhodou aktigrafie je, pokud není zaznamenán pohyb těla, ale přesto se děje jistá specifická mozková aktivita měřená PSG. V tomto stavu nelze pomocí aktigrafu zaznamenat aktivitu a aktigrafický záznam je klasifikován jako spánek. Tato výjimka nám předem definuje méně přesný odhad, o kterém je dobré vědět. [32]

Studie o aktigrafii a jejím využití u pacientů trpících parasomnií nebyly nalezeny.

## 1.5 Cíle práce

Cílem této práce je nastudovat problematiku předzpracování a klasifikace aktigrafických signálů. Je potřeba zkompletovat a předzpracovat zadaný dataset, který obsahuje anotované aktigrafické a polysomnografické záznamy pacientů s parasomnií.

Dále je třeba provést explorační analýzu a identifikovat vhodné příznaky pro rozlišení a) oblastí klidu a aktivity, a b) parasomních a fyziologických epizod. Příznaky budou vytvořeny z dat agregovaných po epochách i hrubých tříosých aktigrafických dat.

Posledním požadavkem je vytvořit klasifikátor pro rozlišení parasomních a fyziologických epizod a evaluovat jeho vlastnosti na zadaném datasetu, vč. realistického zhodnocení reálného nasazení vytvořeného algoritmu.



# Kapitola 2

## Metody

Cílem této kapitoly je přiblížit metodiku analýzy a zpracování 3D aktigrafického a polysomnografického signálu, která je následně použita v úloze klasifikace parasomních epizod. Důležitou částí práce je předzpracování signálu, u čehož je hlavním úkolem explorace dat. Dále jsou nastíněny použité algoritmy a jejich výhody a nevýhody. Následuje popis statistického vyhodnocení.

### 2.1 Naměřená data

Mezi data, která byla naměřena ve spánkové laboratoři Národního ústavu duševního zdraví (NÚDZ), patří tříosé aktigrafické a polysomnografické záznamy. Tyto záznamy pochází z vyšetření od pacientů s diagnostikovanou parasomnií.

Data byla naměřena na celkem 32 pacientech, kdy někteří byly nahrávány dvě noci. Celkový počet záznamů je 35. Nahrávání podstoupilo celkem 20 žen a 12 mužů v průměrném věku 31 let (rozmezí 19 až 44 let).

Každý záznam byl somnologem anotován parasomnickou a fyziologickou aktivitou během spánku. V záznamu byly spánkovým expertem (MUDr. Jitkou Buškovou, Ph.D.) identifikovány počátky epizod fyziologických a parasomních probuzení, tzv. anotace. Anotace byla založena na videu, hypnogramu a vyhodnocení fyziologických signálů, jako je např. EEG, EKG, EMG apod. Dále byl definován hypnogram (tj. posloupnost a časování spánkových fází).

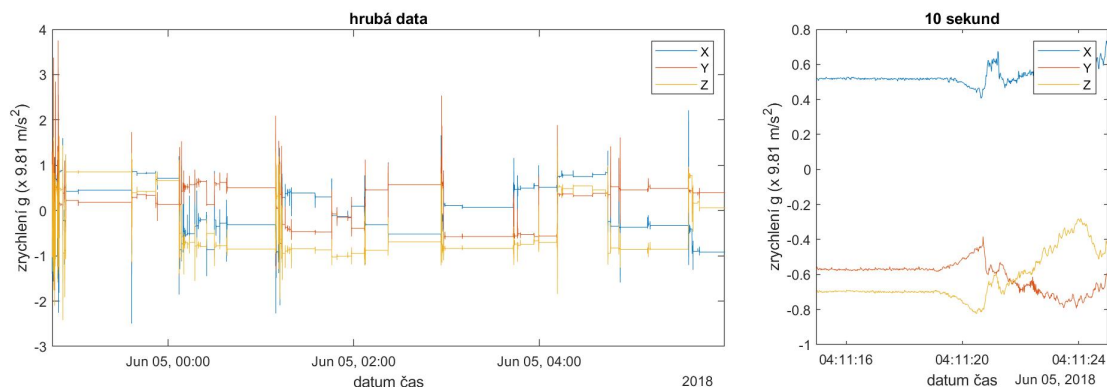
Tyto detekce v záznamu definovaly počátek aktivity v minutových intervalech. Na základě průběžných výsledků této práce byly pak anotace začátků epizod MUDr. Jitkou Buškovou, Ph.D. zpřesněny na formát HH:MM:SS. V první části této práce však anotace ve vteřinách nebyly k dispozici. Byly doplněny až během práce na této studii. Proto je studie rozdělena do dvou různých přístupů podle toho, jak přesné anotace byly k dispozici. Při prvotní analýze musela být explorace provedena velmi jednoduchým způsobem, tedy zpracováním dat v jednotkách minut okolo anotace a vytvořením jednoduchých agregátů. Po získání přesnějších detekcí bylo možné více zpřesňovat analýzu, a tím pádem lépe identifikovat příznaky rozlišující různé aktivity během spánku.

Celkový počet anotací obsahoval 49 parasomních a 101 fyziologických aktivit. Každý záznam obsahoval v průměru 1 parasomní a 3 fyziologické anotace. Jeden záznam obsahoval maximálně 6 parasomních a 7 fyziologických anotací.

Studii schválila etická komise NÚDZ. Data jsou zpracována v prostředí MATLAB®.

#### 2.1.1 Aktigrafické záznamy

Reálné aktigrafické záznamy se naměřily ve spánkové laboratoři přístrojem Actigraph GT3X s vzorkovací frekvencí 100 a 30 Hz souběžně s video-PSG. Výstupem aktigrafického snímání byla hrubá data, která se nijak nefiltrovala ani neagregovala do epoch (viz Kapitola 2.3.1). Příklad zobrazení hrubých dat získaných z aktigrafu lze vidět na Obr. 2.1.



**Obrázek 2.1.** Hrubá data ze záznamu jednoho pacienta a přiblížení v okně o délce 10 s.

### 2.1.2 Polysomnografické záznamy

Polysomnografická data se také naměřila ve spánkové laboratoři NÚDZ. Hrubá data byla vzorkována 1000 Hz a po anotaci somnologem podvzorkována na 250 Hz. Data se nahrála u stejných pacientů zároveň s aktigrafickými. Polysomnografické záznamy obsahovaly tyto biologické signály: EEG (elektrická aktivita mozku), EMG (elektrická aktivita svalů) brady a končetin, EKG (elektrická aktivita srdce) měřené mezi hrudními body V5-V6, hrudní a břišní dýchání, proud vzduchu před nosem a ústy a saturace krve kyslíkem. Náhled na polysomnografický záznam lze vidět na Obr. 1.2.

## 2.2 Klasifikátory

Výsledky předzpracovaných dat je potřeba nějakým způsobem klasifikovat, aby bylo možné zjistit, jak dobře lze parasomní a fyziologickou aktivitu od sebe rozlišit. Klasifikace příznaků, které vznikly na základě explorační analýzy (viz Kapitola 2.3), byla provedena pomocí SVM (Support Vector Machine) a *Random forest* klasifikátoru. Byl vytvořen model na trénovacích datech (70 % celkového datového setu), který se následně ověřil na testovací sadě (zbylých 30 % datového setu). Oba zmíněné klasifikátory jsou v následujících kapitolách popsány a porovnány.

### 2.2.1 SVM klasifikátor

SVM klasifikátor (*Support vector machine*) je populární metoda strojového učení pro klasifikaci, regresi a další úlohy učení s učitelem. Typické použití metod učení s učitelem zahrnuje tyto dva kroky: nejprve trénink datové sady k získání modelu, a poté použití modelu k predikci výstupů na testovací datové sadě. Pro ohodnocení stability a stálosti modelu se SVM klasifikátor aplikoval 100krát na aktigrafických datech, která byla pokaždé náhodně rozdělena do trénovací a testovací sady.

SVM je založen na myšlence nalezení optimální nadroviny, která nejlépe odděluje dvě třídy. Body nejbližší k nadrovině se nazývají *support vectors* a vzdálenost vektorů od nadroviny se nazývá *margins*. Výhodou tohoto klasifikátoru je efektivnost ve vysokodimenzionálních prostorech, k rozhodování používá podmnožinu tréninkových dat (z nich *support vectors*) a navíc je též všestranný (lze definovat, pro jaký typ dat je klasifikováno).

Jeden ze základních principů SVM klasifikátoru je nelineární transformace dat z prostoru původních dat do nového prostoru, ve kterém lze třídy od sebe lineárně oddělit

- tzv. Kernelova metoda. Cílem je najít rovinu, která má maximální okraj, tj. maximální vzdálenost mezi datovými body obou tříd. S vyšší vzdáleností budou datové body klasifikovány s větší jistotou.

Nadrovina je v  $p$ -dimenzionálním prostoru popsána následujícím vztahem:

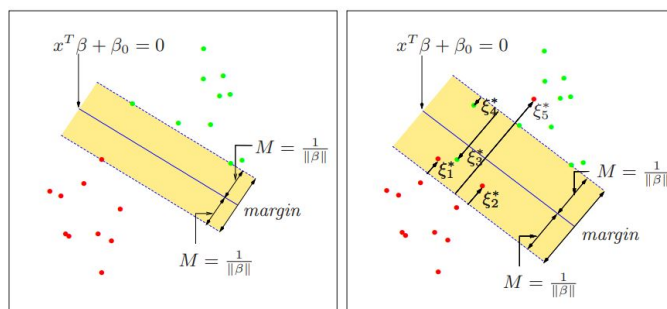
$$\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = 0, \quad (1)$$

kde  $\beta$  jsou parametry nadroviny. Tato rovnice pak rozděluje prostor na dvě části:

$$\text{pozorování} = \begin{cases} \text{třída 1} & \text{pokud } \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p > 0 \\ \text{třída 2} & \text{pokud } \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p < 0 \end{cases} \quad (2)$$

Rozhodnutí optimálního nalezení nadroviny spočívá ve výpočtu vzdálenosti od každého pozorování k oddělovací nadrovině. Cílem je nalézt takovou nadrovinu, která má okraj nejvzdálenější od pozorování. Toto se nazývá hyperplocha maximální marže (*Maximal margin hyperplane*). Maximalizací této hyperplochy nalézáme optimální parametry  $\beta$ . Body, které jsou ve stejné vzdálenosti od nadroviny, se nazývají podpůrné vektory (*Support vectors*). Tyto vektory „podporují“ maximální rovinu okraje, z čehož vyplývá, že pokud se pohybují, pohybuje se také nadrovina. U ostatních bodů pohyb nemá vliv na oddělovací nadrovinu.

Při oddělení pozorování dle nadroviny nelze vždy dokonale oddělit data. V těchto případech lze dovolit určitému počtu pozorování porušit pravidla, aby mohla ležet na nesprávné straně hranic okrajů. Pak je potřeba definovat „uvolněná“ (*slack*) pozorování  $\epsilon$ .  $\epsilon$  vyjadřuje jak velkou „chybu“ dovolíme SVM modelu udělat. Hledáme-li nadrovinu, která téměř oddělí třídy, použijeme tzv. *soft margin*. Příklad bodů, které je možné rozdělit nadrovinou, lze vidět na Obr. 2.2.



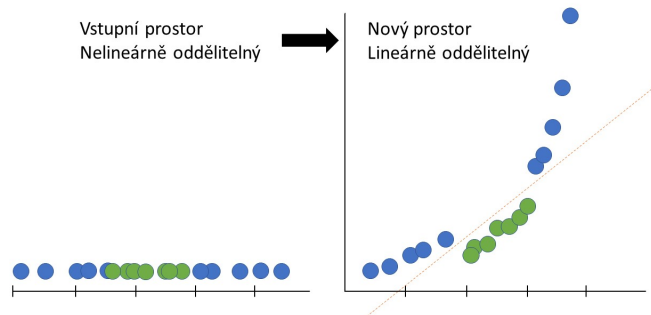
**Obrázek 2.2.** SVM klasifikátor ve 2D prostoru. Levý obrázek znázorňuje oddělení dvou tříd. Rozhodovací hranicí je plná čára, zatímco přerušované ohraničují maximální hranici šířky. Pravý obrázek zobrazuje neoddělitelný případ (překrývají se). Marže je maximalizována. Převzato z [33].

Výsledná optimalizace SVM modelu lze popsat následujícími vztahy [34]:

$$(\beta_0, \beta_1, \dots, \beta_p) = \operatorname{argmax} M, \quad (3)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M, \quad (4)$$

kde  $i = 1, \dots, N$ .



**Obrázek 2.3.** Transformace dat do prostoru o vyšší dimenzi, aby byla lineárně oddělitelná. V levé části lze vidět data, která nelze lineárně oddělit. Proto jsou mapována do vyšší dimenze, ve které je možné je oddělit (vpravo).

Nalezení optimálního prostoru, ve kterém lze třídy nejlépe oddělit, je vizuálně znázorněno na Obr. 2.3.

Tento popsaný klasifikátor nalézá lineární hranice ve vstupním prostoru prvků. Postup je však možné flexibilizovat rozšířením prostoru funkcí pomocí základních rozšíření, jako jsou např. polynomy nebo splajny. Obecně hranice dosahují ve zvětšeném prostoru lepšího oddělení tréninkové třídy a vznikají tak nelineární hranice. [33]

Takovou funkcí je například *Radial Basis (rbf) Kernel*, která nalézá vektorové klasifikátory v nekonečných dimenzích. Principem je porovnávání pozorování, kdy nejbližší mají velký vliv na to, jak bude klasifikováno nové pozorování, a pozorování, která jsou dále, mají relativně malý vliv na klasifikaci. Tato funkce vypočítává euklidovskou vzdálenost pozorování testovacího vzorku vzhledem k trénovacímu. Jinými slovy, pouze blízká testovací pozorování k tréninkovým mohou ovlivnit predikci třídy testovacích dat. Hlavní Radial Kernel parametr, který určuje velikost vlivu každého pozorování v tréninkové sadě na klasifikaci nových pozorování, je:

$$\kappa(x_1, x_2) = e^{-\gamma(x_1 - x_2)^2}, \quad (5)$$

kde  $x_1$  a  $x_2$  jsou dvě odlišná pozorování. Míra vlivu jednoho pozorování na druhé je tedy funkcí druhé mocniny vzdálenosti. Parametr  $\gamma$  (gamma), který je určen křížovou validací, mění měřítko druhé mocniny, a tím mění vliv dvou bodů. Čím větší hodnota  $\gamma$  je, tím zohledňuje více pozorování a bližší pozorování se více ovlivňují. Tato konstanta odpovídá za plynulost rozhodovací hranice a rozptyl modelu.

SVM klasifikátor lze pro optimalizaci výsledků ladit. K tomu je využívána *Optimize Hyperparameters* hodnota, která vyhledává takové hodnoty parametrů minimalizující chybu v křížové validaci. Touto optimalizací se opakovaním snaží model nalézt nejlepší nadrovinu.

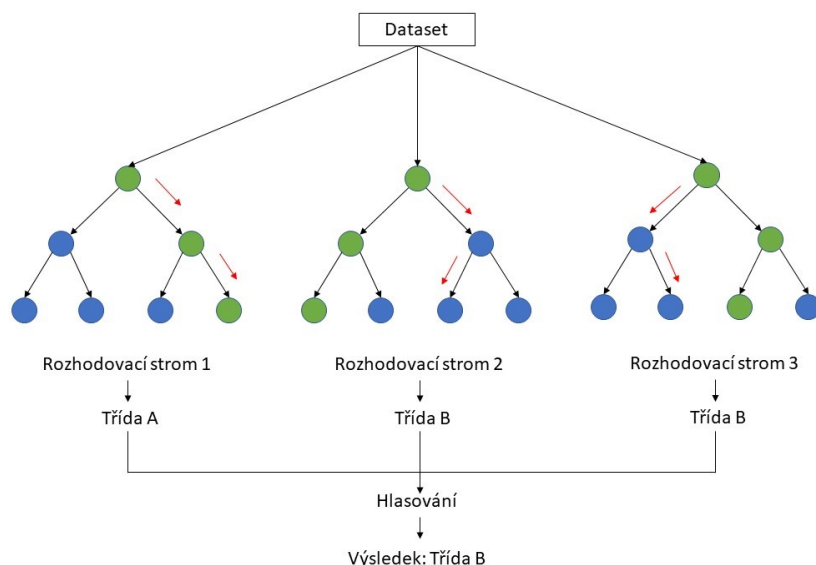
Použití SVM modelu se snižuje výpočetní náročnost, jelikož se model zaměřuje pouze na *support vectors* ležící na okraji funkce a není citlivý na změnu jediného pozorování. Na druhou stranu je potřeba výpočet ve vysokodimenzionálním prostoru.

## 2.2.2 Random forest klasifikátor

Random forest (náhodný les) je metoda strojového učení pro klasifikaci a regresi. Tento model využívá ve svém rozhodovacím procesu velké množství rozhodovacích stromů s použitím náhodně vybrané podmnožiny tréninkových souborů. Tato metoda představuje modifikaci metody *bagging* [35]. Random forest model využívá kombinace

rozhodovacích stromů. Každý z těchto vytvořených stromů závisí na hodnotách náhodného vektoru vzorkovaného pro všechny stromy se stejnou distribucí [36]. Každý strom pak dělá nezávislá rozhodnutí a konečný výsledek je určen na základě hlasování (viz Obr. 2.4). [33]

Výhodou algoritmu je, že není zkreslený, protože každý strom z daného velkého počtu stromů je trénován na podmnožině dat. Na základě síly „davů“ se algoritmus rozhoduje, do jaké třídy data spadají. Model je též velmi stabilní, jelikož není novým pozorováním příliš ovlivněn. Toto pozorování snadno ovlivní jeden strom, nikoliv všechny. Nevýhodou je výpočetní složitost kvůli velkému množství rozhodovacích stromů. Trénink tohoto velkého množství navíc zabírá více času než jiné srovnatelné algoritmy.



**Obrázek 2.4.** Ukázka Random forest modelu. Na obrázku lze vidět vytvořené tři rozhodovací stromy z trénovacích dat. Červené šipky pak znázorňují již klasifikaci dat na testovací sadě. Na základě hlasování lze vidět výslednou třídu.

Pro vytvoření Random forest klasifikátoru jsou potřeba tyto následující kroky: *bootstrapping* a vytvoření rozhodovacího stromu s ohledem na náhodnou podmnožinu. Zopakováním předchozích dvou kroků lze vytvořit další rozhodovací stromy. Bootstrapping je obecně statistická technika pro hodnocení statistické metody. V *Random forest* modelu se využívá k selekci dat na základě náhodného výběru pozorování a parametrů včetně opakování.

K vytvoření *bootstrap* datasetu jsou potřeba anotovaná vstupní data, která obsahují parametry jednotlivých pozorování a výslednou třídu. Z těchto dat jsou náhodně vybrána pozorování, přičemž nemusí být vybrána všechna a některá se mohou opakovat (typicky  $\frac{1}{3}$  vstupních dat není v novém datasetu). Rozhodovací strom se následně vytváří z tohoto vytvořeného trénovacího datasetu, ale v každém kroku je použita pouze náhodná podmnožina proměnných (parametrů). Z vybraných parametrů se vytváří rozhodovací větve. Postupně je vytvořen rozhodovací strom s ohledem na náhodnou podmnožinu proměnných v každém kroku. Ideálně je pak potřeba vytvořit stovky rozhodovacích stromů. Rozmanitost dělá Random forest efektivnějším než individuální rozhodovací stromy.

Po vytvoření Random forest z trénovacích dat jej lze využít na klasifikaci testovacích. Nové pozorování je klasifikováno každým rozhodovacím stromem. Na základě výsledků

hlasování všech stromů lze pozorování zařadit do třídy s více hlasy (viz Obr. 2.4). Vytvoření *bootstrap* datasetu a použití stromů k rozhodování se nazývá *bagging*.

## 2.3 Návrh vlastního řešení

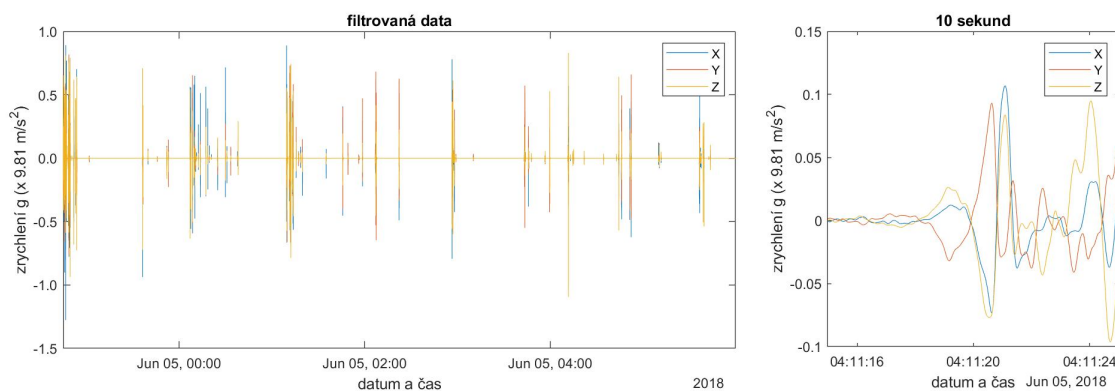
V této kapitole je přiblížen postup, který byl použit pro předzpracování dat, identifikaci příznaků pro rozlišení různých aktivit, klasifikaci vybraných příznaků a detekci v celonočním záznamu. Postup práce lze rozdělit v závislosti na přesnosti časových anotací, ve které první část byla více testovací a sloužila spíše ke zjištění či specifikaci dat, zda je možné aktivity od sebe nějakým způsobem oddělit. Nad rámec práce byla provedena detekce parasomnie v celonočním záznamu, která vycházela ze zjištěných poznatků předchozího předzpracování a klasifikace.

### 2.3.1 Předzpracování dat s minutovými anotacemi

Vzhledem k tomu, že zkoumání fyzické aktivity při parasomních epizodách je zcela novou úlohou, prvním krokem bylo provedení explorační analýzy naměřených dat. Vyhledávání nových informací bylo založeno na anotovaných časových detekcích od expertky MUDR. Jitky Buškové, Ph.D., díky jejíž pomoci bylo možné explorační analyzovat rozdíl fyziologické a parasomní aktivity. Cílem předzpracování bylo získat užitečné příznaky, které lze využít při klasifikaci dat.

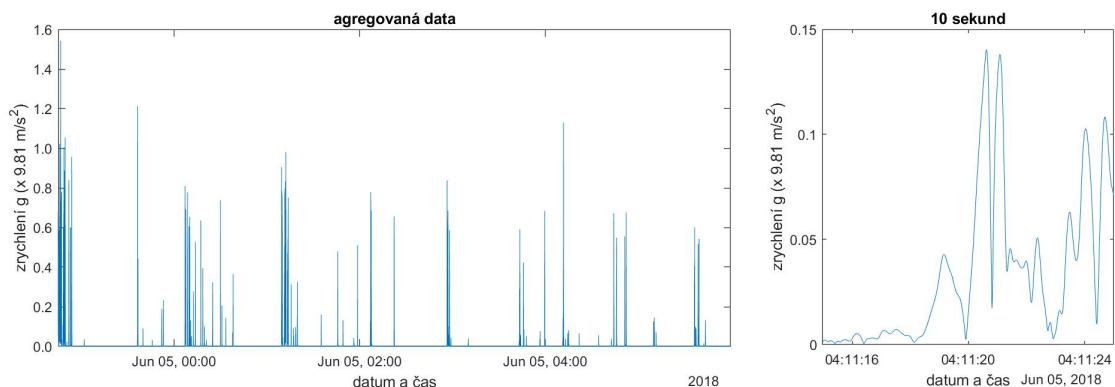
Klasický přístroj na měření aktigrafických dat pro úsporu paměti a snížení chyby měření velmi často agreguje data již při měření a výstupem jsou už jistým způsobem zpracovaná data. Pro úsporu dat se provádí tzv. *agregace* dat v epochách o délce typicky 30 s či 60 s. Po uvedené době se data nahrávají do mezipaměti a z celé epochy se uloží pouze jeden nebo více argumentů (např. kroky, „counts“, maximum apod.). [37]

Abyste nedošlo ke ztrátě informace zpracovaných dat, byla v této práci použita původní hrubá data naměřených zrychlení ve třech osách. Tato data převzatá z přístroje s vzorkovací frekvencí 100 a 30 Hz byla nejdříve filtrována pásmovou propustí pro odstranění gravitačního zrychlení a pomalých pohybů (horní propust) a pro odstranění vibrací (dolní propust). Na základě několika studií [38–40] byl vybrán frekvenční interval 0.25–2.5 Hz (viz Obr. 2.5).



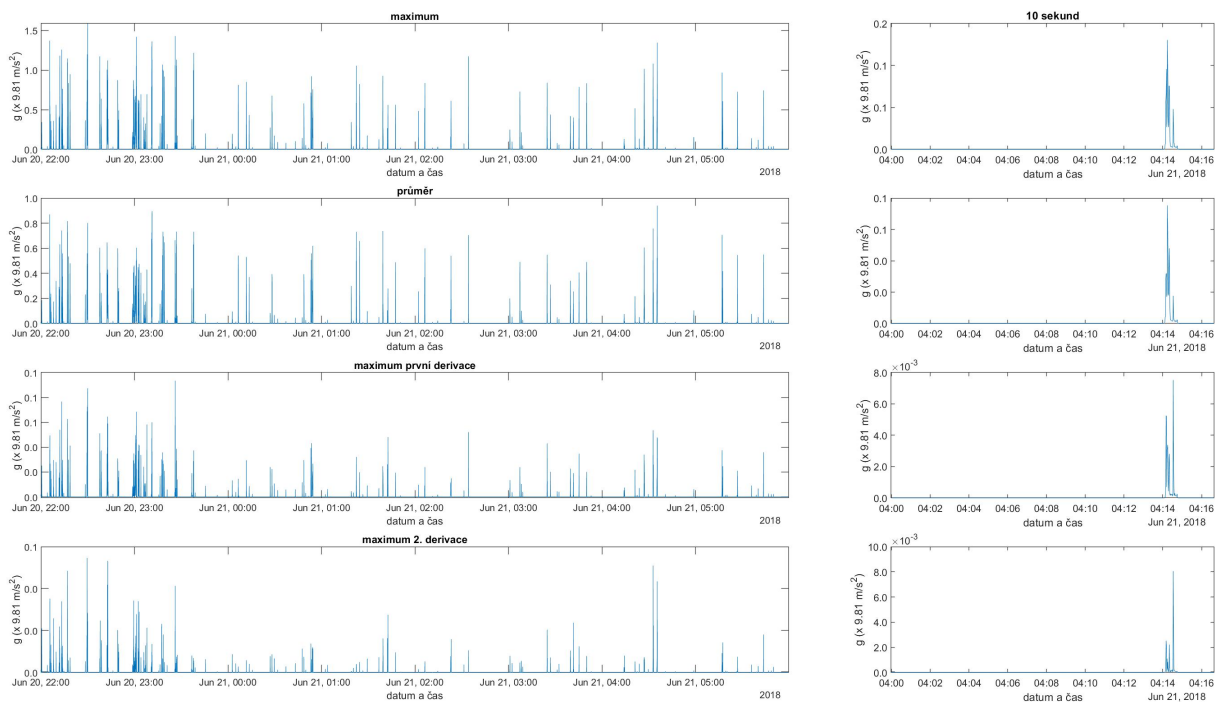
**Obrázek 2.5.** Filtrace hrubých dat pásmovou propustí jednoho pacienta v rozmezí 0.25–2.5 Hz a přiblížení v okně o délce 10 s.

Filtrací byly získány signály ve stejné izolínii, aby se následně sloučily do jednoho, čímž však došlo ke ztrátě 3D informace a detailu. Sloučení bylo provedeno Euklidovskou metodou [41] (viz Obr. 2.6).



**Obrázek 2.6.** Agregace dat Euklidovskou metodou jednoho pacienta a přiblížení v okně o délce 10 s.

Pro možnost detekovat aktivní část s menší výpočetní náročností byly vypočítány příznaky každé 1 vteřiny signálu. Vybraly se tyto základní parametry: maximum, průměr, maximum první derivace a maximum druhé derivace (viz Obr. 2.7).



**Obrázek 2.7.** Příznaky agregovaného signálu jednoho pacienta: maximum, průměr, maximum první derivace, maximum druhé derivace.

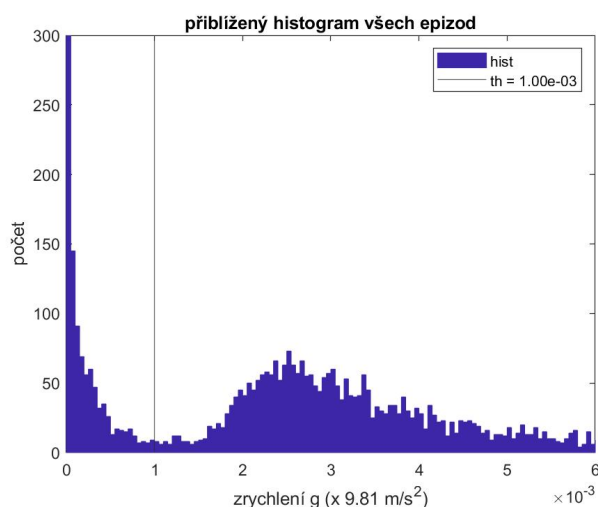
Na Obr. 2.7 lze vidět vypočítané příznaky agregovaného signálu. Aktivita během spánku je možné opticky rozeznat, jelikož se velikost zrychlení výrazně zvýší oproti klidové fázi. Pro další zpracování byl tedy vybrán parametr maximum, jelikož nese informaci maximální aktivity, při které lze metodou prahování snadno rozlišit aktivitu a je i jedním ze standardních příznaků pro detekci spánku-bdění.

Pro exploraci dat se muselo vzhledem k absenci přesných anotací začátků epizod vytvořit oblasti parasomnických a fyziologických aktivit, ze kterých by bylo možné vypočítat určité příznaky. Tyto oblasti se určily z epizod okolo anotovaných detekcí.

Ze známých časových detekcí (hodina a minuta) počátku události parasomnických i fyziologických aktivit byly vytvořeny epizody okolo tohoto času o délce 5 minut: 2 minuty před a 3 minuty po detekci, jelikož by aktivita neměla být delší. Později však vyšlo najevo, že je tento čas příliš dlouhý a může zanášet nepřesnosti do výsledků. Proto byly následně detekce anotovány s přesností na vteřinu a epizody se zkrátily na 90 vteřin (30 s před a 60 s po detekci).

Ze získaného signálu bylo potřeba identifikovat příznaky parasomní aktivity. Aby se využil pouze signál, který vykazuje dostatečnou aktivitu pacienta, byla v celé 90vteřinové epizodě prahováním určena tzv. aktivní oblast.

Pro optimální prahování vytvořených epizod se musela nalézt ideální hodnota prahu. Proto se vytvořil histogram všech epizod, ze kterého již bylo možné aktivitu vyčíst (viz. Obr. 2.8). Na histogramu lze vidět, že se epizody skládají ze dvou rozdělání, kdy každé ohraničuje aktivní a neaktivní oblast.



**Obrázek 2.8.** Histogram všech anotovaných epizod: přibližná zajímavá oblast a zobrazený vybraný práh.

Po získání prahové hodnoty  $th = 0.001$  už se dalo z vytvořené epizody prahováním získat oblasti aktivity. Ve dvouvteřinovém okně (tzv. epoše) bylo vypočteno maximum epizody a porovnáno s globální prahovou hodnotou. Při překročení prahu se epocha označila jako aktivní. Klasifikace aktivity dvouvteřinové epochy probíhala dle následujícího vztahu:

$$epocha = \begin{cases} aktivní & \text{pro } M_i > th, \\ neaktivní & \text{pro } M_i \leq th, \end{cases} \quad (6)$$

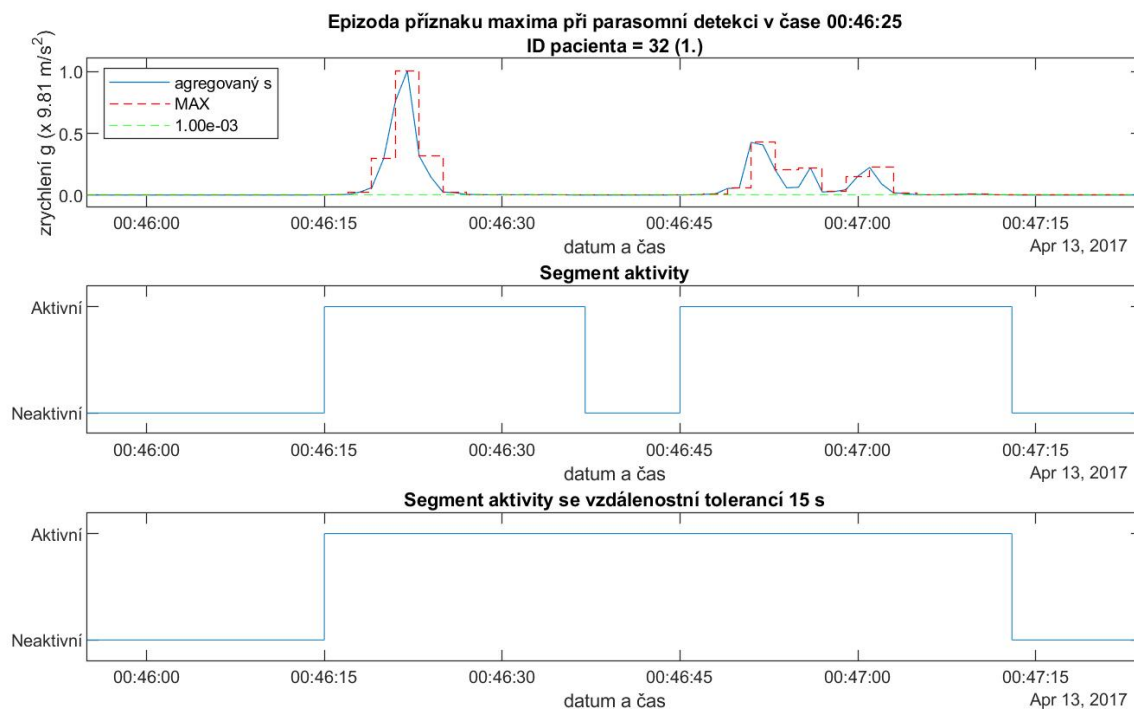
kde  $M_i$  je vypočítaný příznak maximum ve dvou vteřinách a  $th = 0.001$  je globální prahová hodnota vyčtená z histogramu všech epizod.

Pro sjednocení celé aktivity a odstranění krátkých neaktivních segmentů byla použita distanční tolerance **15 s**, která sjednocovala všechny segmenty bližší než zvolená vzdálenost. Sjednocení segmentů probíhalo na základě tohoto vztahu:

$$sousední\ segmenty = \begin{cases} sjednoceny & \text{pokud } t_{inactive} < t_{tolerance}, \\ nesjednoceny & \text{pokud } t_{inactive} \geq t_{tolerance}, \end{cases} \quad (7)$$

kde  $t_{inactive}$  je čas neaktivity mezi aktivními segmenty a  $t_{tolerance}$  je nastavených 15 s.





**Obrázek 2.9.** Detekce aktivních segmentů: Na horním obrázku lze vidět jednu epizodu parasomní detekce, dvouvteřinové epochy maxima a globální prahovou hodnotu (zeleně). Uprostřed lze vidět aktivní segment, který je určen na základě prahu. Poslední obrázek označuje segmenty aktivity s určenou distanční tolerancí.

Ukázka epizody s dvouvteřinovými epochami a výslednou klasifikací aktivních segmentů lze vidět na Obr. 2.9.

Díky zjištění aktivních segmentů už bylo možné z těchto oblastí vypočítat příznaky:

- maximum
- minimum
- průměr
- maximum první derivace
- procento aktivity
- procento neaktivity
- počet aktivních oblastí v jednom segmentu

Postupem času byly přidány příznaky popisující několik prvních vteřin segmentu:

- dynamika pohybu prvních 10 s segmentu (maximum, minimum, průměr).

Pro vizuální porovnání rozdílů aktivit se zobrazily histogramy (viz Obr. 3.1 a ROC křivky příznaků (viz Obr. 3.2).

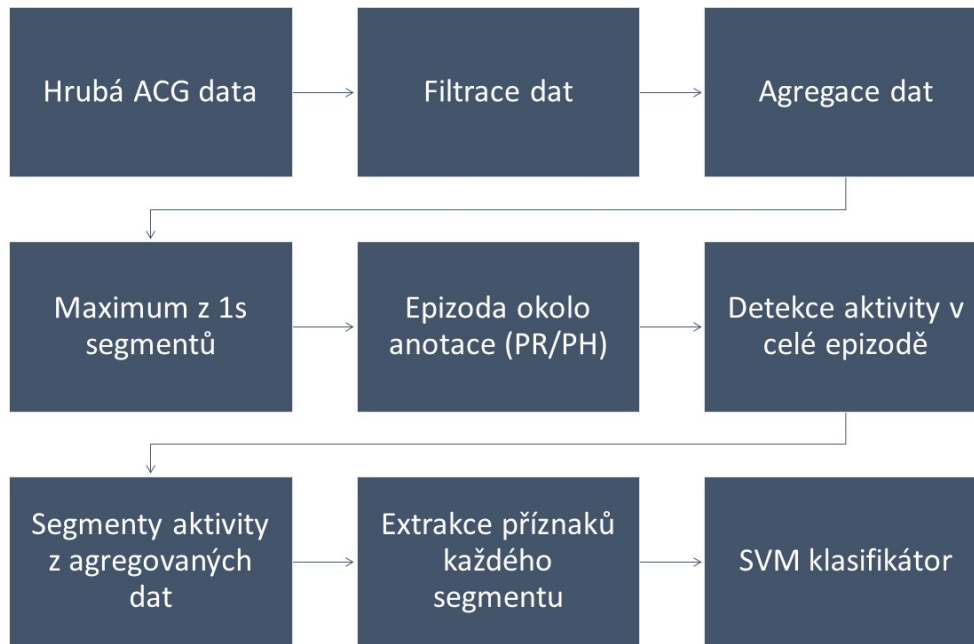
Proces celého předzpracování od získání hrubých dat po extrakci příznaků je znázorněn na Obr. 2.10.

### ■ 2.3.2 Definice příznaků pro předzpracování minutových anotací

Dle následujících vztahů byly vypočteny příznaky aktivních segmentů:

**Maximum** epizody lze vyjádřit vztahem

$$Max = \max(x_i), \quad (8)$$



**Obrázek 2.10.** Schéma předzpracování agregovaného aktigrafického signálu: Dle jednotlivých bloků lze vidět, jaké kroky byly provedeny k závěrečné extrakci příznaků pro SVM klasifikaci.

kde  $x_i$  jsou zrychlení aktigrafického signálu.

Opačný parametr **minimum** epizody lze vyjádřit obdobným vztahem

$$Min = \min(x_i), \quad (9)$$

**Odhad střední hodnoty** je možné vypočítat tímto způsobem:

$$\mu = \frac{1}{N} \cdot \sum_{i=1}^N x_i, \quad (10)$$

kde  $N$  je celkový počet vzorků a  $x$  je vzorek v bodě  $i$ ,

Příznak **procento aktivity či neaktivity** je vypočten tímto způsobem:

$$active = \frac{t_{active}}{t_{segment}} \cdot 100, \quad (11)$$

kde  $t_{active}$  je délka aktivity v celé délce aktivity segmentu  $t_{segment}$ .

Výpočet **dynamiky signálu v prvních několika vteřinách** vycházel z difference okna signálu a jejím zprůměrováním. Vypočítat lze takto:

$$delta(n) = |x_n - x_{n+1}|, \quad (12)$$

kde  $delta(n)$  je vypočítaná difference,  $x$  jsou vzorky signálu. Získaná difference byla následně zprůměrována v předem nastaveném okně (100 ms):

$$avg\_delta(m) = \frac{1}{N} \cdot \sum_{n=k}^{k+N-1} delta(n), \quad (13)$$

kde  $N = 0.1 \cdot fs$  je počet vzorků v daném okně a  $k = N \cdot 0 + 1, N \cdot 1 + 1, N \cdot 2 + 1, \dots, n - N$  je posloupnost průměrovaného okna. Z výsledného parametru  $avg\_delta$  bylo dále možné vypočítat příznaky minimum, maximum a průměr.

### 2.3.3 Předzpracování dat s vteřinovými anotacemi

Jelikož byla analýza dat s minutovými anotacemi velmi nepřesná a nešlo sledovat okolí počátku aktivity s vteřinovou přesností, zpřesnily se anotace od MUDr. Jitky Buškové, Ph.D. na vteřinu. Proto bylo možné dále více sledovat dynamiku počátku aktivity, která je dle dodatečných informací somnologů i odborné literatury pro rozlišení parasomních a fyziologických epizod charakteristická.

V rámci této studie byla také k dispozici nahraná polysomnografická data pacientů s diagnostikovanou parasomnickou poruchou. Data se využila ke klasifikaci zároveň s aktigrafickými. Z mnoha signálů se vybralo EMG (brady a končetin), které by mohlo potenciálně obsahovat další informaci o pohybu pacienta. Dalším vybraným signálem bylo EKG (měřeno mezi V5-V6), které by mohlo poskytnout výraznou odlišnost při parasomní a fyziologické aktivitě. Z EKG signálu se vypočetl RR interval, ze kterého byl vyčten jeden z příznaků.

Pro předzpracování dat se opět využily časové anotace od expertky. Kromě hrubých signálů byl z aktigrafických a EMG dat vypočítán frekvenční parametr výkonová spektrální hustota (power spectral density - PSD) (viz níže).

Další úprava se týkala EKG signálu, ze kterého se vypočítal RR interval (viz Kapitola 2.3.4).

Aby bylo možné extrahovat příznaky, vytvořily se události okolo anotace 4 s před a 10 s po detekci. Příklad jedné události lze vidět na Obr. 2.11, kde je možné sledovat dynamický vývoj signálů v závislosti na čase. Proto se příznaky extrahovaly ve dvou-vteřinových úsecích a pozorovány v čase.

Proces předzpracování od hrubých dat po identifikaci příznaků je znázorněn na Obr. 2.12.

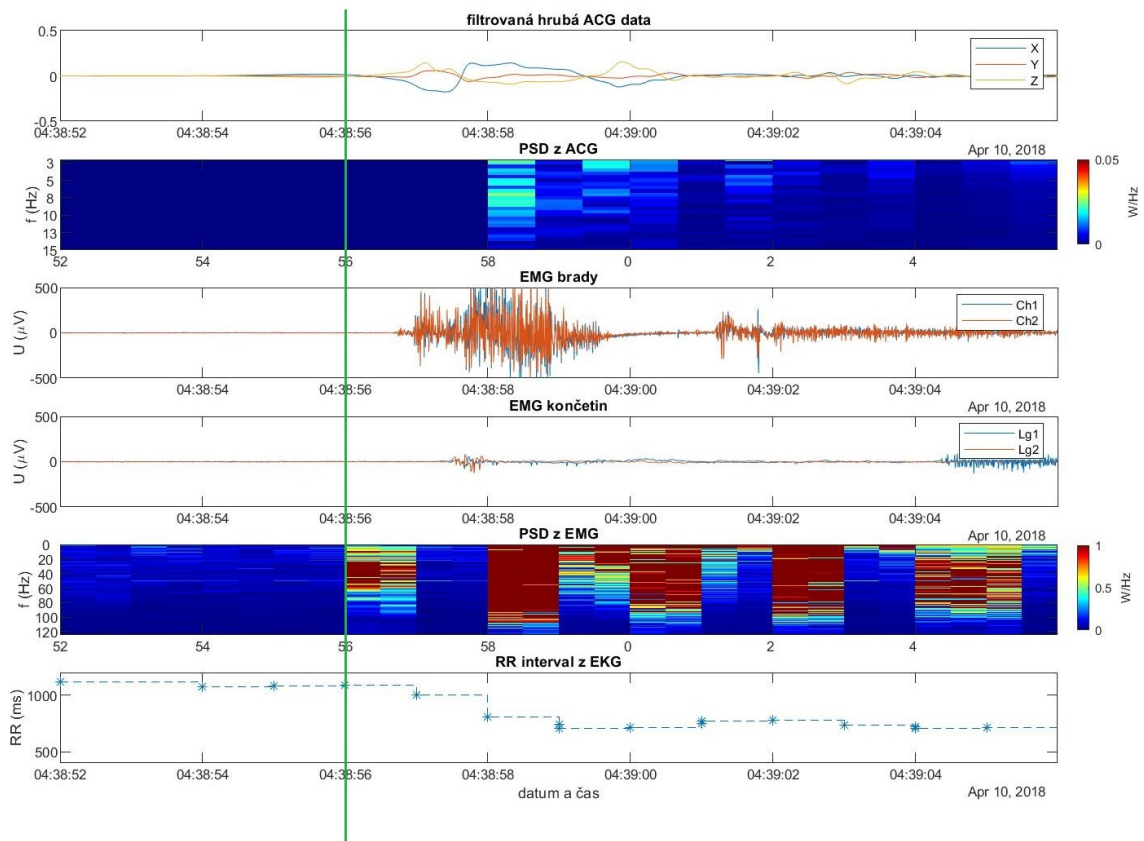


**Obrázek 2.12.** Schéma předzpracování aktigrafického a polysomnografických signálů: Dle jednotlivých bloků lze vidět, jaké kroky byly provedeny k závěrečné extrakci příznaků pro Random forest klasifikaci.

### 2.3.4 Definice příznaků pro předzpracování vteřinových anotací

Poté, co byly vytvořeny události o délce 14 s, zaměřila jsem se v této práci na zpracování příznaků v čase. Na Obr. 2.11 lze vidět, jak se signály mění. Proto byl vybrán úsek 2 vteřin, ve kterém se v každém tomto okně vypočítaly příznaky.

Z biologických signálů byly extrahovány následující příznaky:



**Obrázek 2.11.** Zobrazení události z ACG a PSG dat: Událost obsahuje signály aktigrafu, výkonové spektrum aktigrafického signálu, EMG brady a končetin a výkonové spektrum těchto signálů a RR interval z EKG signálu. Zelená čára znázorňuje počátek detekované aktivity (detekováno expertkou).

#### ■ ACG

- maximální úhel aktigrafického záznamu
- maximální rozdíl směrů dvou os (osy X-Y, X-Z, Y-Z)
- PSD ve frekvenčním spektru 0.5-10 Hz (příznaky po 0,5 Hz)

#### ■ EMG

- PSD ve frekvenčním spektru 0-100 Hz (příznaky po 0.5 Hz do 10 Hz, poté v logaritmickém měřítku)

#### ■ ECG

- změna RR intervalu

Příznak **Maximální úhel aktigrafického záznamu** vychází ze vztahu pro výpočet celkového směru:

$$\theta_x = \cos^{-1} \left( \frac{x}{\sqrt{x^2 + y^2 + z^2}} \right), \quad (14)$$

kde  $x, y, z$  jsou signály změřeného zrychlení v příslušné ose. Maximální úhel je pak maximum dvouvteřinového okna ze všech tří úhlů.

Příznak **Maximální rozdíl směrů 2 os** vychází ze stejného vztahu pro výpočet celkového směru. Místo následného výpočtu maxima byly však jednotlivé velikosti úhlů normalizovány metodou min-max. Tím se získaly 3 signály o velikosti 0 až 1. Dále se vypočítaly rozdíly mezi každými dvěma znormovanými velikostmi úhlů. Výslednými příznaky byly maximální rozdíly znormovaných úhlů X-Y, X-Z, Y-Z v každém dvouvtěřinovém okně.

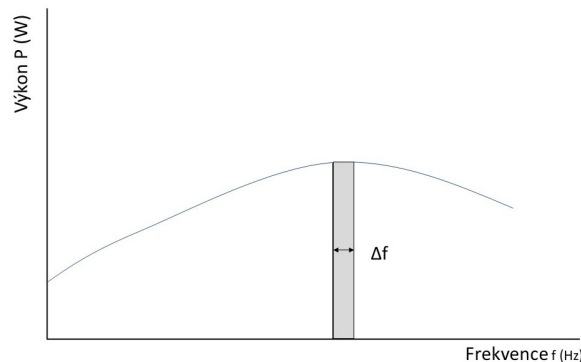
Dalším potřebným parametrem pro výpočet příznaků je výkonová spektrální hustota, která vyjadřuje výkon obsažený v intervalu spektra a je vyjádřena W/Hz. Vypočítat lze pomocí rychlé Fourierovy transformace (fast Fourier transform - FFT). Tato metoda se nazývá Bartlettova metoda periodogramů a zahrnuje tyto kroky: Datová posloupnost  $x(n)$  o délce  $n$  je rozdělena na  $I$  nepřekrývající se segmenty o délce  $N$ . Pro každý segment je vypočítán periodogram:

$$PSD(k)^i = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi kn} \quad (15)$$

Periodogramy pro  $I$  segmentů jsou průměrovány:

$$PSD(k)^I = \frac{1}{I} \sum_{i=0}^{I-1} PSD(k)^i \quad (16)$$

Měření výkonové spektrální hustoty segmentu lze vidět na Obr. 2.13



**Obrázek 2.13.** Měření výkonové spektrální hustoty.

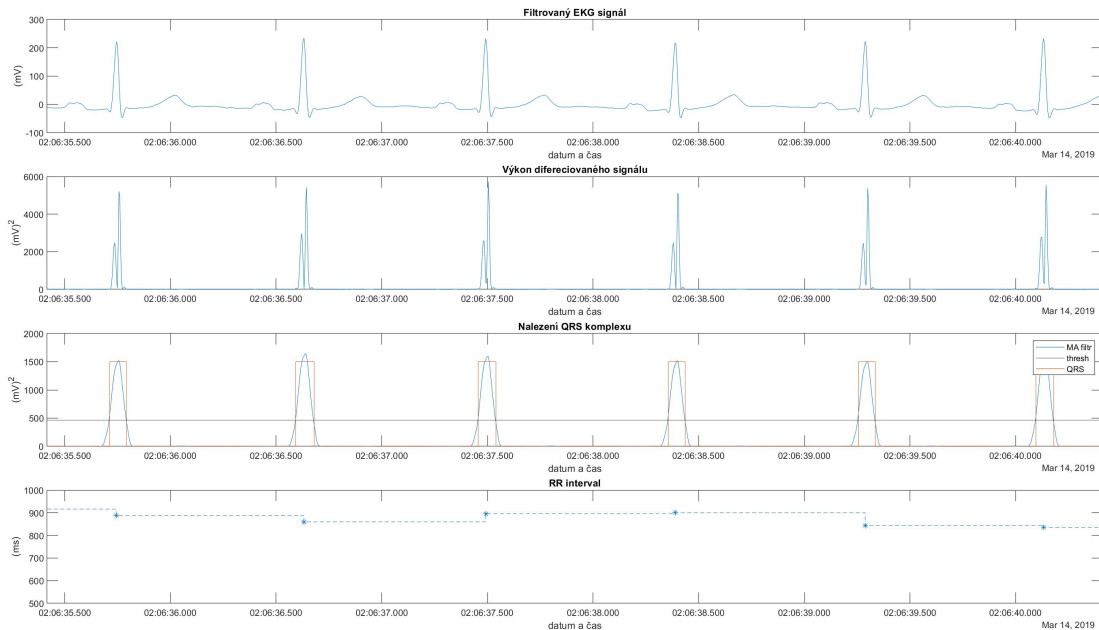
Z aktigrafického signálu bylo vybráno **spektrum PSD ACG 2.5-10 Hz** a z **EMG PSD 0-100 Hz** (od 10 Hz byla šířka spektrálních binů zvolena v logaritmickém měřítku: 13, 16, 21, 28, 36, 46, 60, 77, 100 Hz). Aktigrafické PSD obsahovalo pro každý dvouvtěřinový úsek frekvenci všech tří os. EMG PSD takto obsahovalo celkem 4 hodnoty, jelikož obsahuje 2 signály brady a 2 signály končetin. Toho si lze všimnout na Obr. 2.11, kde v druhém obrázku (PSD z ACG) obsahuje každé dvouvtěřinové okno 3 sloupce, v pátém obrázku (PSD z EMG) pro změnu 4 sloupce.

Poslední vypočítaný příznak **Rozdíl RR** vycházel z RR intervalu. Nejdříve bylo potřeba detekovat R-špičky QRS komplexu. Původní EKG signál se filtroval pásmovou propustí 0.5-40 Hz. Byla vytažena diference signálu, umocněna a výsledný signál se vyhladil filtrem klouzavého průměru (moving average - MA), který se běžně používá k vyhlazení pole signálu. Takto vyhlazená data se prahovala 90% kvantilem. Pomocí získaných pozic byla nalezena lokální maxima QRS komplexu - R-špičky. Postupné zpracování lze vidět na Obr. 2.14.

Po získání RR intervalu se následně vypočítal příznak. Minimum RR intervalu bylo odečítáno od průměru předchozího RR intervalu dle vztahu:

$$RR_{dif} = |\min(RR_N) - \text{mean}(RR_{N-1})| \quad (17)$$

kde  $RR_N$  je RR interval aktuálního dvouvtěrinového segmentu a  $RR_{N-1}$  předchozího dvouvtěrinového segmentu.



**Obrázek 2.14.** Zpracování EKG signálu: Na prvním obrázku lze vidět filtrovaný hrubý signál, na druhém vypočítaný výkon diference signálu. Třetí obrázek znázorňuje detekci QRS komplexu z vyhlazeného signálu MA filtrem. Poslední obrázek ukazuje již vypočítaný RR interval.

### 2.3.5 Detekce parasomnie v celonočním záznamu

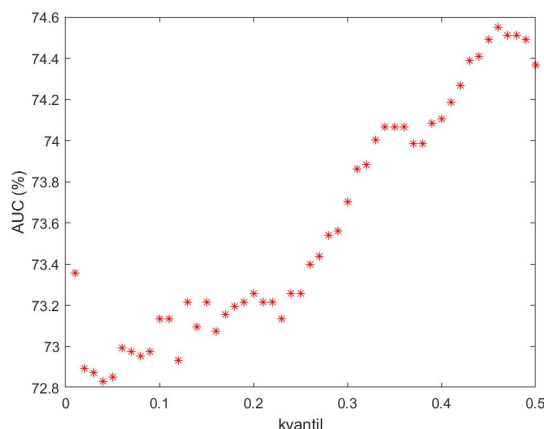
Předchozí klasifikace na základě anotovaných dat sloužila pro rozpoznání příznaků, které by mohly odlišit fyziologickou aktivitu od parasomní. Nedávala nám však žádnou informaci o tom, kde kromě anotovaných epizod by se parasomnie v celonočním záznamu mohla vyskytovat. Detekce v celonočním záznamu naopak nevychází z anotací, ale detekuje aktivitu parasomnie z celého záznamu na základě získaných poznatků. Výsledky jsou pak statisticky porovnávány s anotacemi od expertky.

V první fázi byly detekovány časy aktivity pacientů v celonočním záznamu. Aktivita byla detekována z aktigrafického záznamu pacienta a k jejímu zjištění bylo zrychlení porovnáváno s prahovou hodnotou. Tento parametr se zjišťoval dle následujícího algoritmu.

Původním záměrem bylo porovnání celkového zrychlení 2 s před a 2 s po anotaci, které by přineslo informaci o tom, jak přesně detekovat aktivitu v celém záznamu. Použitím ROC křivky však přineslo rozlišení aktivit AUC hodnoty okolo 60 %, což neznaznačuje přílišnou rozdílnost stavů. Proto se použila jiná varianta, která porovnávala různé aktivity po čase detekce.

Ze známých anotací byly vytvořeny aktigrafické události 4 vteřiny po detekci. Tyto události se rozdělily na parasomní a fyziologické. Zrychlení ve třech osách bylo zpraco-

váno do agregovaného signálu (pomocí Euklidovské metody viz Kapitola 2.3.1). Z tohoto výsledného signálu trvajících 4 vteřiny (tj. 400 nebo 120 hodnot agregovaného zrychlení dle vzorkovací frekvence) byly vypočítány kvantily 0.01 až 0.5, které byly vstupními parametry pro výpočet ROC křivky. Hodnoty se rozdělovaly do dvou tříd dle toho, jestli šlo o parasomní či fyziologickou aktivitu. Následně se vypočítala AUC hodnota z ROC křivky a optimální pracovní bod (viz vztah (22)), k jehož výpočtu byla využita funkce ze Statistického Toolboxu [42]. Přehled AUC hodnot je vidět na Obr. 2.15. Ze zmiňovaného obrázku lze vyčíst, že se AUC hodnoty v závislosti na kvantilech výrazně neliší (73-75 %). Proto byly všechny hodnoty nalezených optimálních bodů přes všechny kvantily zprůměrovány a tím se získala globální prahová hodnota  $0.047 g \times 9.81 m/s^2$  pro detekci parasomní aktivity.



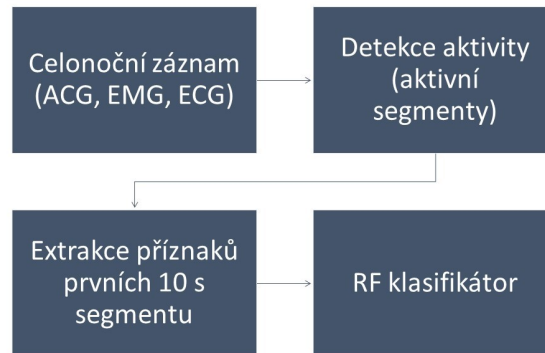
**Obrázek 2.15.** AUC hodnoty v závislosti na kvantilu čtyřvteřinového segmentu: rozlišovací schopnost parasomní a fyziologické aktivity.

Pomocí globálního prahu bylo tedy možné získat z celonočního záznamu pouze oblasti zájmu, ve kterých pacient vykazoval aktivitu. Agregovaný aktigrafický signál s vyššími hodnotami než prahová hodnota se označilo jako aktivní. Data se porovnávala s globálním prahem ve dvouvteřinovém okně přes celonoční záznam. Tento rozhodovací postup byl obdobný jako ve vztahu (6). Aby se nemusel každý segment vyhodnocovat zvlášť, použil se slučovací algoritmus. Aktivní segmenty, které mezi sebou měly menší „neaktivní“ oblast než stanovená hranice 15 vteřin, byly označeny jako jeden. Tento postup byl stejný jako ve vztahu (7). Ze získaných segmentů aktivity šlo dále pracovat pouze s úseky, ve kterých aktigrafická data vykazovala nějakou aktivitu.

Z předchozích poznatků bylo zjištěno, že nejvyšší informaci o rozlišení fyziologické a parasomní aktivity nese několik prvních vteřin události (viz Kapitola 3.3, Tab. 3.4). Proto se z celého jednoho segmentu aktivity použilo prvních deset vteřin, ze kterých se po dvouvteřinovém intervalu vypočetly příznaky z aktigrafických, EMG a EKG dat (stejně jako v Kapitole 2.3.4). Příznaky byly klasifikovány pomocí *Random forest* modelu, který již byl natrénován na datech z předchozího zpracování (viz Kapitola 2.3.3). Pokud se alespoň jeden dvouvteřinový úsek z prvních 10 vteřin klasifikoval jako pozitivní třída, byl celý aktivní segment označen jako pozitivní (parasomní aktivita).

Klasifikované úseky se statisticky porovnávaly s anotacemi. Přesnost detekce zahrnovala toleranci 30 vteřin před a 60 vteřin po anotaci, stejně jako tomu bylo při vytváření událostí v Kapitole 2.3.1. To znamená, že pokud reálná detekce od predikovaného času parasomní aktivity byla o 30 s dříve nebo o 60 s později, vyhodnotila se tato predikce jako správná. Jestliže do tohoto intervalu nespádala žádná reálná detekce, označila se jako falešně pozitivní. Statistické vyhodnocení je zobrazeno v Tab. 3.5.

Proces detekce parasomní aktivity je znázorněn na Obr. 2.16.



**Obrázek 2.16.** Schéma předzpracování dat pro detekci parasomní aktivity: Dle jednotlivých bloků lze vidět, jaké kroky byly provedeny k závěrečné klasifikaci pomocí Random forest modelu.

## 2.4 Statistické vyhodnocení

Aby se daly výsledky navržených klasifikátorů porovnat, bylo potřeba výsledky statisticky vyhodnotit. K tomu se využilo metody Monte Carlo. Rozdělení dat do trénovací a testovací sady a následná klasifikace se 100krát zopakovala, čímž byl ověřen vliv konkrétního rozdělení dat do trénovací a testovací množiny na výsledky klasifikace.

Jedna z metod srovnání původní expertní anotace a výsledků klasifikace je výpočet správných a falešných detekcí. V reálných signálech jsou pozice časových detekcí známy, proto lze vytvořit konfuzní matici obsahující následující prvky: správné detekce True Positive (TP), které byly klasifikovány algoritmem, falešné detekce False Positive (FP), které jsou nesprávně označeny algoritmem a reálné detekce False Negative (FN), které nebyly algoritmem označeny. Hodnoty True Negative (TN) lze započítávat pouze v případě klasifikace z předem vytipovaných epizod. Při detekci v celonočním záznamu však tato hodnota nelze určit, jelikož by mohl být každý vzorek signálu bez detekce určen jako TN. Parasomní aktivita byla označena jako pozitivní třída, negativní označovala fyziologickou aktivitu. Dále lze vypočítat i jiné parametry a to konkrétně senzitivitu, pozitivní predikci a přesnost.

*Senzitivita* (true positive rate TPR) vyjadřuje pravděpodobnost správné detekce příkladů z pozitivní třídy a vypočítá se následujícím vztahem [43]:

$$TPR = \frac{TP}{TP + FN} \quad (18)$$

*Pozitivní predikce* (positive predictive value PPV) je pravděpodobnost, že je pozitivně detekované pozorování opravdu z pozitivní třídy, lze ji vyjádřit takto [43]:

$$PPV = \frac{TP}{TP + FP} \quad (19)$$

*Přesnost* (accuracy ACC) představuje poměr mezi správně predikovanými pozorováními a všemi pozorováními v datové sadě [44]:

$$ACC = \frac{TP + TN}{P + N} \quad (20)$$



Aby bylo možné zjistit přesnost výsledků, vypočítaly se výše zmíněné parametry z více opakování a z toho následně hodnoty průměru a směrodatné odchylky.

Definice směrodatné odchylky je následující:

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \mu)^2}, \quad (21)$$

kde  $n$  je celkový počet dat a  $\mu$  odhad střední hodnoty (viz rovnice (10)).

### 2.4.1 AUC

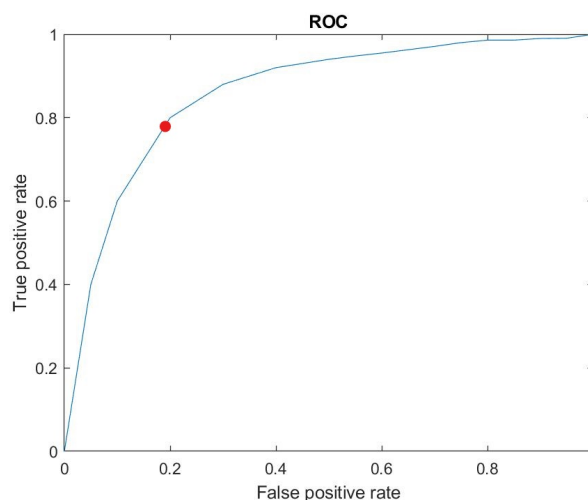
K dalšímu zhodnocení výsledků lze využít ROC (Receiver operating characteristics) křivky (viz Obr. 2.17). Ta se hodí pro vizualizaci a výběr vhodných klasifikátorů na základě jejich výkonu. Zobrazuje kompromis mezi senzitivitou (TPR) a mírou falešného poplachu (FPR) klasifikátorů. Pro porovnání výsledků ROC křivek lze vypočítat plochu pod ROC křivkou (AUC), jejíž hodnota bude vždy mezi 0 a 1. Čím je tato hodnota vyšší než 0.5, tím lépe model předpovídá. Nejhorší situací je, pokud je AUC přibližně 0.5. Model pak nedisponuje diskriminační schopností rozlišovat mezi pozitivní a negativní třídou. Model, ve kterém se hodnota AUC blíží 0, předpovídá opačný výsledek (předpovídá pozitivní jako negativní a naopak).

Na této křivce je možné najít tzv. optimální pracovní bod ROC křivky. Tento bod lze využít k určení optimální prahové hodnoty pro rozeznání dvou různých tříd. K získání optimálního bodu je nejdříve potřeba tzv. sklon  $S$  [42]:

$$S = \frac{Cost(P|N) - Cost(N|N)}{Cost(N|P) - Cost(P|P)} \cdot \frac{N}{P}, \quad (22)$$

kde  $Cost(N|P)$  jsou náklady při nesprávné klasifikaci pozitivní třídy, naopak  $Cost(P|N)$  náklady při nesprávné klasifikaci negativní třídy.  $P$  a  $N$  vyjadřují celkový počet pozorování v pozitivní a negativní třídě.

Posunem přímky se sklonem  $S$  z levého horního rohu ROC (FPR = 0, TPR = 1) dolů a doprava je protnutím s ROC křivkou nalezen optimální pracovní bod. K získání optimálního pracovního bodu byla využita funkce `perfcurve` v Statistickém Toolboxu Matlab [42].



**Obrázek 2.17.** Ukázka ROC křivky s optimálním pracovním bodem (červeně).

# Kapitola 3

## Výsledky

V této kapitole jsou prezentovány výstupy předzpracování 3D aktigrafických a polysomnografických dat. Porovnávají se příznaky fyziologické a parasomní aktivity. Vybrané příznaky byly použity ke klasifikaci parasomní a fyziologické aktivity. Následuje statistické vyhodnocení klasifikace SVM a Random forest modelu.

### 3.1 Předzpracování dat

V rámci předzpracování dat byly použity dva různé přístupy, které se odvíjely od přesnosti časových anotací (čas počátku detekce). K prvnímu zpracování dat s minutovými anotacemi se přistupovalo co nejjednodušším způsobem, jelikož se nebylo možné detailně zabírat záznamy. V druhé části práce s přesnějšími anotacemi se tato možnost naskytla a je tedy více komplexní.

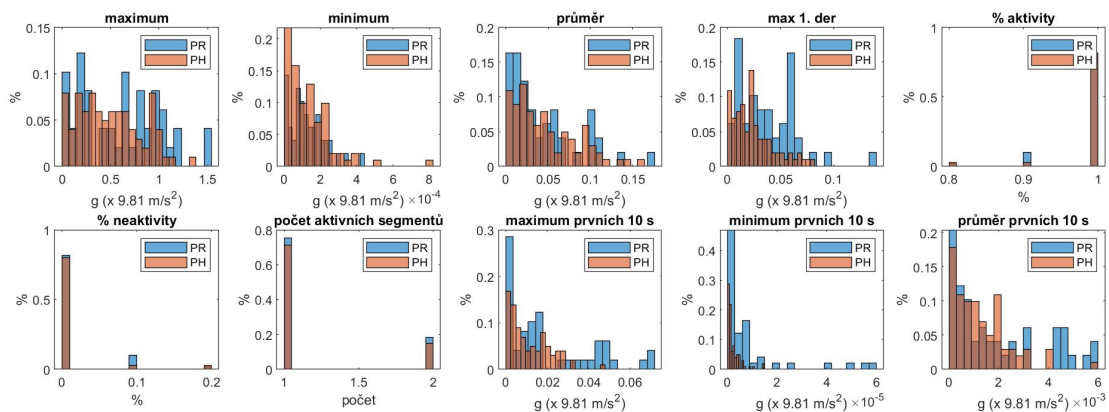
#### 3.1.1 Předzpracování dat s minutovými anotacemi

Pro identifikaci příznaků se nejprve vytvořily události okolo časových anotací (viz Obr. 2.9). Ve vytvořených událostech byly zjištěny oblasti aktivity okolo času detekce (viz Kapitola 2.3.1). Z celkových 150 anotací (49 parasomních a 101 fyziologických) bylo prahováním označeno jako neaktivní celkem 14 fyziologických a 3 parasomní. S ostatními událostmi se tedy nadále dalo pracovat.

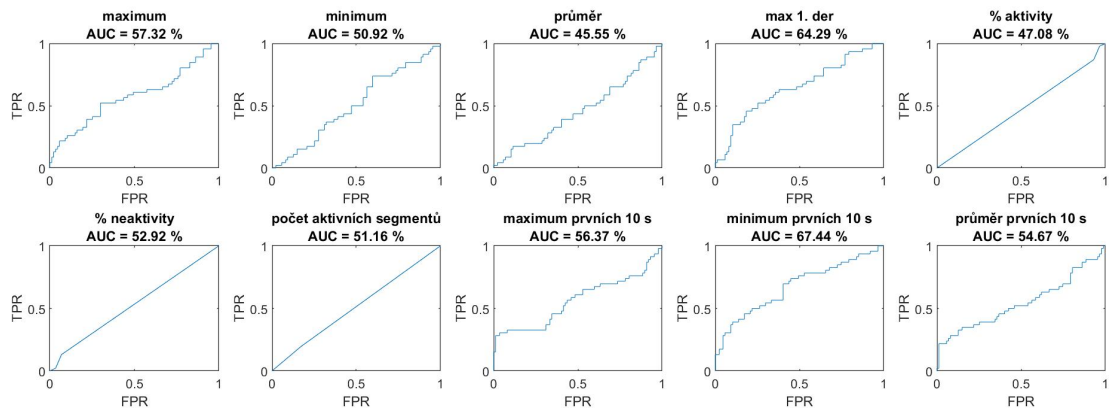
Ze získaných segmentů obsahující aktivitu pacienta, byly vypočteny definované příznaky (viz Kapitola 2.3.2):

- maximum
- minimum
- průměr
- maximum první derivace
- procento aktivity
- procento neaktivity
- počet aktivních oblastí v jednom segmentu
- dynamika pohybu prvních 10 s segmentu: maximum, minimum, průměr.

Z extrahovaných příznaků byly zobrazeny histogramy a ROC křivky pro porovnání parasomní a fyziologické aktivity (Obr. 3.1 a 3.2).



**Obrázek 3.1.** Znормovaný histogram vypočítaných příznaků parasomní a fyziologické aktivity: maximum, minimum, průměr, maximum první derivace, procento aktivity a neaktivity, počet aktivních oblastí v jednom segmentu a dynamika signálu v prvních vteřinách. Zkratky: PR (parasomní aktivita) a PH (fyziologická aktivita).



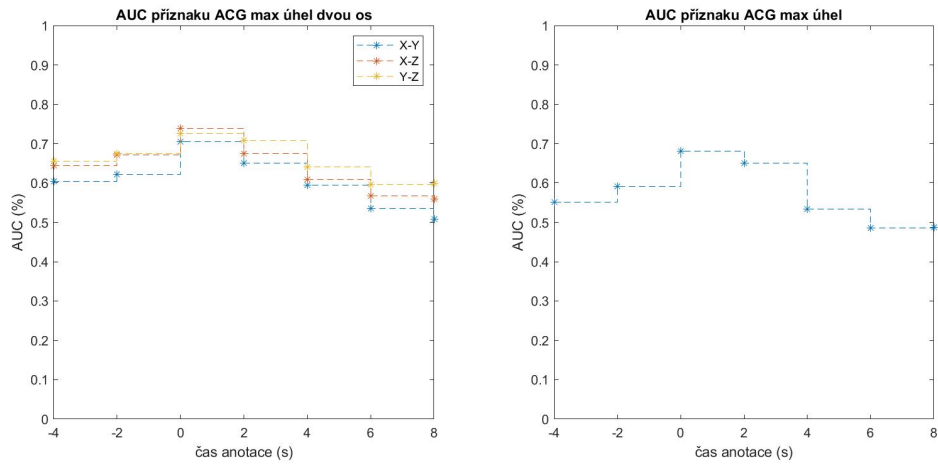
**Obrázek 3.2.** ROC křivky vypočítaných příznaků parasomní a fyziologické aktivity: maximum, minimum, průměr, maximum první derivace, procento aktivity a neaktivity, počet aktivních oblastí v jednom segmentu a dynamika signálu v prvních vteřinách. U každého příznaku je zobrazena vypočtená hodnota AUC.

Na Obr. 3.1 a 3.2 je možné vidět, že ve zvolených příznacích není viditelný rozdíl mezi parasomní a fyziologickou aktivitou. AUC hodnoty všech příznaků jsou velmi blízké hodnotě 50 %, z toho vyplývá velmi malý rozdíl mezi fyziologickou a parasomní aktivitou. Nejvyšší AUC hodnota byla 67 % u příznaku minimum dynamiky prvních 10 s.

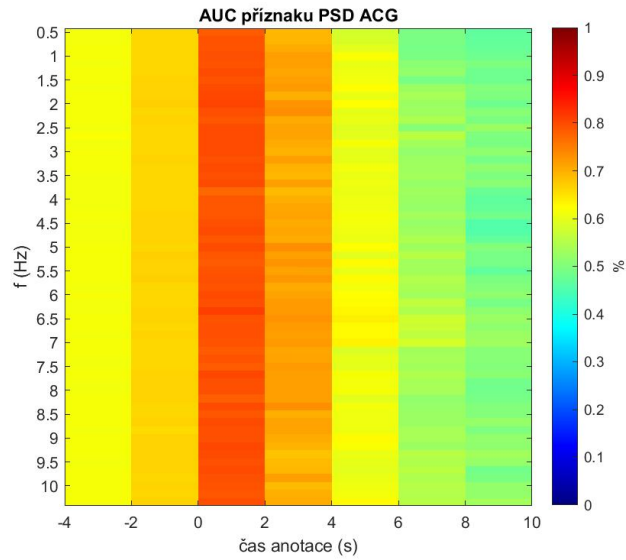
I přesto, že hodnocení příznaků bylo dle AUC hodnot velmi slabé, provedla se klasifikace dat. Výsledná data se statisticky vyhodnotila (viz Kapitola 3.2).

### 3.1.2 Předzpracování dat s vteřinovými anotacemi

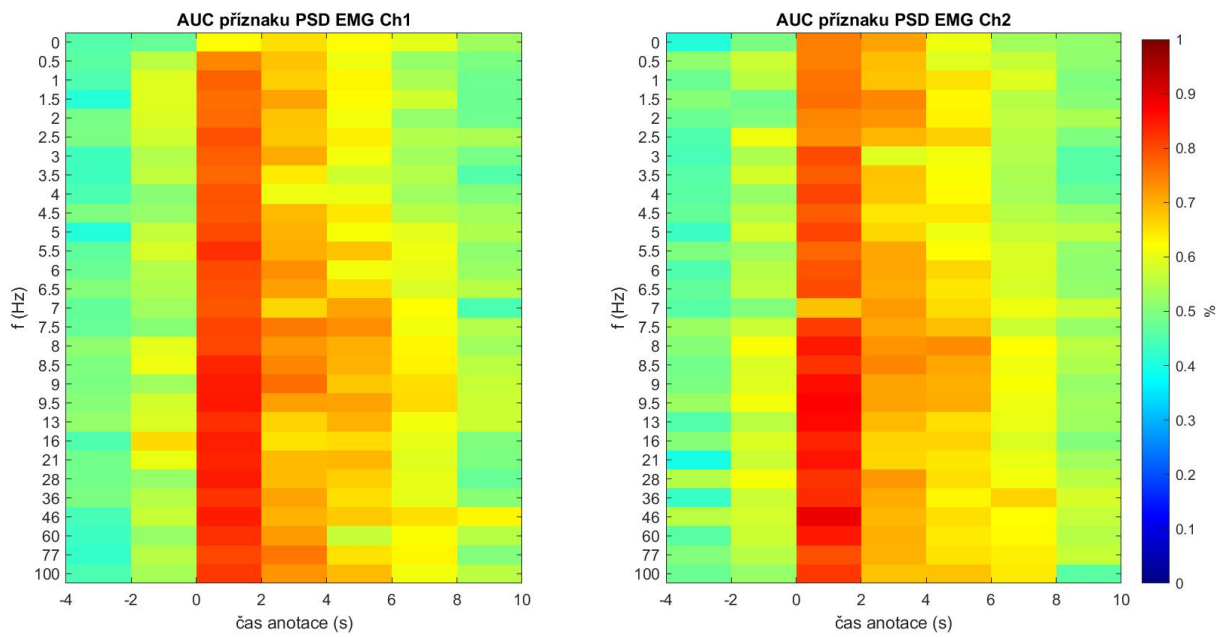
Z příznaků (viz Kapitola 2.3.4) byly vypočteny AUC hodnoty pro dvouvteřinové úseky přes celý 14vteřinový segment a zobrazeny do následujících obrázků: příznak maximální úhel a příznak maximální rozdíl směřů dvou os lze vidět na Obr. 3.3, PSD aktigrafického záznamu na Obr. 3.4, PSD EMG záznamu brady na Obr. 3.5, PSD EMG záznamu končetin na Obr. 3.6, změna RR intervalu na Obr. 3.7.



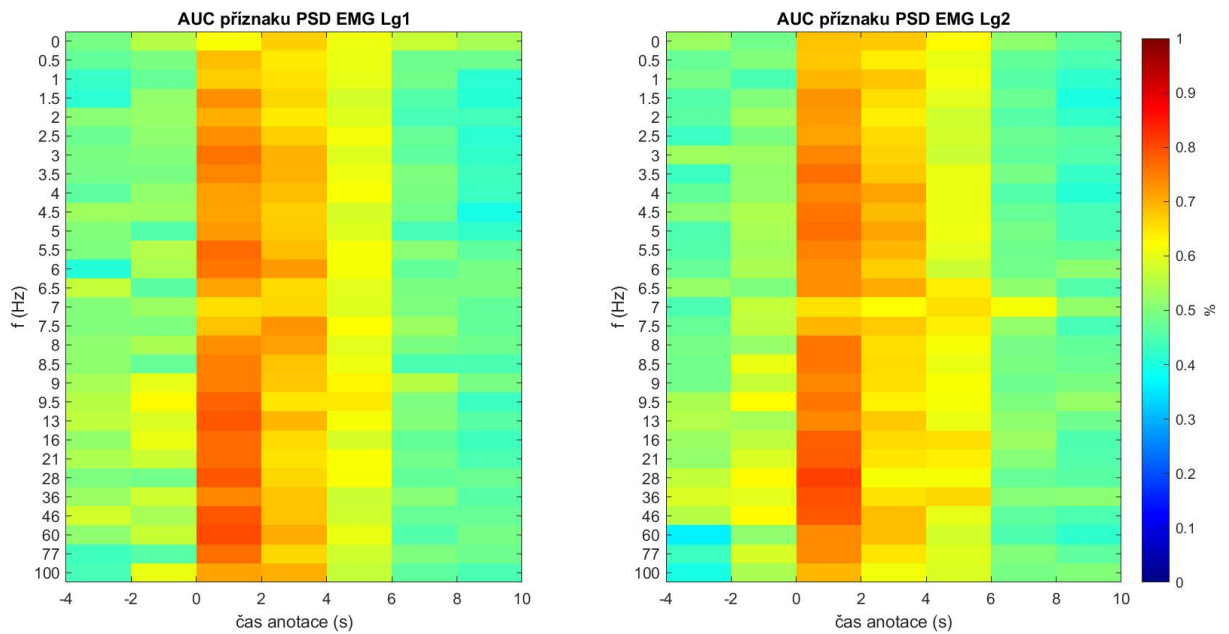
**Obrázek 3.3.** AUC hodnoty příznaků: maximální rozdíl směrů dvou os (vlevo), maximální úhel aktigrafického signálu (vpravo).



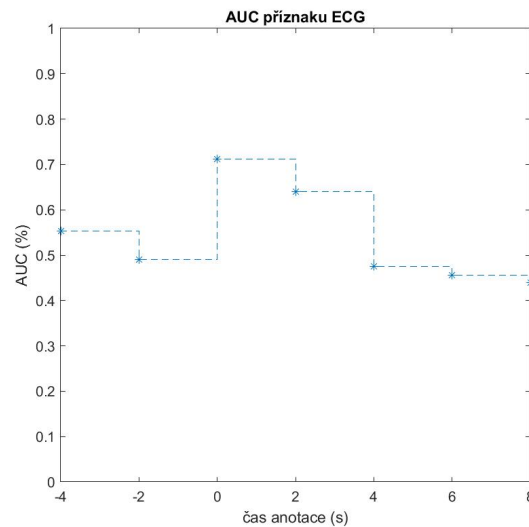
**Obrázek 3.4.** AUC hodnoty příznaku výkonová spektrální hustota aktigrafického signálu v rozmezí 0.5-10 Hz.



**Obrázek 3.5.** AUC hodnoty příznaku výkonová spektrální hustota EMG signálu brady v rozmezí 0-100 Hz.



**Obrázek 3.6.** AUC hodnoty příznaku výkonová spektrální hustota EMG signálu končetin v rozmezí 0-100 Hz.



**Obrázek 3.7.** AUC hodnoty příznaku maximální rozdíl RR intervalu z EKG signálu.

Z obrázků 3.3, 3.4, 3.5, 3.6, 3.7 lze říci, že se AUC hodnota v 0. vteřině zvýší, tedy v čase počátku detekované aktivity. Poté se postupně snižuje. Pro následnou klasifikaci byly z tohoto důvodu vybrány příznaky ze segmentu 0-2 s a 2-4 s. Nejvyšší AUC hodnotu obsahoval biologický signál EMG brady 89 %.

### 3.2 Klasifikace dat s minutovými anotacemi

Jelikož samotné předzpracování dat s minutovými anotacemi nevykazovalo výraznou odlišnost parasomní a fyziologické aktivity, se pro klasifikaci dat použil nejjednodušší SVM model sloužící pro ilustraci. Cílem klasifikace bylo zjistit, jak by model reagoval na tento druh příznaků a statisticky data vyhodnotit.

Data (pozorování z jednotlivých anotací) byla rozdělena do trénovací a testovací sady v poměru 7:3. Statistické vyhodnocení se zpracovalo po 100 opakováních klasifikace (data byla 100krát náhodně rozdělena do trénovací a testovací sady nezávisle na záznamu, zda patří ke stejnému pacientovi, či ne). Výsledky SVM klasifikátoru z trénovací sady jsou vypsány v Tab. 3.1 a z testovací v Tab. 3.2.

Stat. vyhodnocení	Průměr	Směrodatná odchylka
ACC (%)	75.4	11.1
TPR (%)	26.0	34.1
PPV (%)	93.3	9.3

**Tabulka 3.1.** Statistické vyhodnocení SVM klasifikátoru trénovacích dat: vypočtené průměry a rozptyly přesnosti (ACC), senzitivity (TPR) a pozitivní predikce hodnot (PPV).

Výsledkem klasifikace dat z minutových anotací je přesnost 67 %, nicméně se projevila nízká senzitivita 10 %. Z toho vyplývá, že použitím této klasifikace zjistíme pouze desetinu detekcí parasomní aktivity oproti skutečnému výskytu.

Jelikož mají data podprezentovanou pozitivní třídu (poměr parasomních a fyziologických je 1:2), může nastat situace, že bude senzitivita výsledné klasifikace velmi nízká (viz Tab. 3.2). Jedním z řešení by byla možná multiplikace pozitivní třídy. K tomuto řešení je potřeba přidat podmínku, aby nebyla datová sada jednoho záznamu

Stat. vyhodnocení	Průměr	Směrodatná odchylka
ACC (%)	67.2	5.7
TPR (%)	10.0	15.8
PPV (%)	47.9	20.5

**Tabulka 3.2.** Statistické vyhodnocení SVM klasifikátoru testovacích dat: vypočtené průměry a rozptyly přesnosti (ACC), senzitivity (TPR) a pozitivní predikce hodnot (PPV).

v trénovací i testovací sadě zároveň. Důsledkem této cesty však byla nižší senzitivita než u původního datasetu.

Pokud porovnáme výsledky klasifikace trénovací a testovací sady, lze vidět, že jsou hodnoty odlišné. Závislost rozdělení dat do těchto sad může být pro klasifikátor zásadní a ten pak nemusí být schopen stejně dobře vyhodnotit i nová data.

### 3.3 Klasifikace dat s vteřinovými anotacemi

Předzpracování dat s doplněnými přesnějšími anotacemi již vykazovalo vyšší rozlišovací schopnosti parasomní a fyziologické aktivity. *Random forest* klasifikátor byl vybrán kvůli vyššímu počtu příznaků a i k jeho obecnému a přehlednému porozumění datům.

Extrahované příznaky z polysomnografických a aktigrafických dat byly použity pro vyhodnocení parasomní-fyziologická aktivity. Data se stejně jako u SVM klasifikace rozdělila do trénovací a testovací sady v poměru 7:3. Pro vyhodnocení byla tato data rozčleněna klasifikována opět 100krát. Zvláště se použily příznaky podle biologických signálů, tedy aktigrafické, EMG a EKG. Též byly aplikovány všechny zároveň, aby se ověřil jejich vzájemný vliv či nikoliv. Výsledky trénovací sady jsou znázorněny v Tab. 3.3 a testovací v Tab. 3.4.

čas po anotaci		ACG	EMG	EKG	všechny
0-2 s	ACC (%)	93.2	100.0	93.2	100.0
	TPR (%)	55.7	59.8	55.7	59.8
	PPV (%)	53.5	59.8	54.5	59.8
2-4 s	ACC (%)	95.7	100.0	91.8	100.0
	TPR (%)	57.2	59.8	54.9	59.8
	PPV (%)	55.8	59.8	53.5	59.8

**Tabulka 3.3.** Statistické vyhodnocení RF klasifikátoru trénovacích dat: vypočtené průměry přesnosti (ACC), senzitivity (TPR) a pozitivní predikce hodnot (PPV).

Maximální směrodatná odchylka byla u obou tabulek maximálně 6 % s tím, že průměrně se pohybovala kolem 2-3 %. V porovnání s předchozí metodou (viz Kapitola 3.2) lze tedy říci, že výsledky byly méně variabilní a tím pádem stálejší. V porovnání jednotlivých biologických signálů a použití všech najednou se výsledné hodnoty u EMG a ACG výrazně nelišily, u EKG více. U testovacích dat vyšla maximální přesnost segmentu v čase 0-2 vteřina 82.9 %, přičemž od EMG příznaků se lišila pouze 1 %. Senzitivita se výrazně zvýšila, u EMG příznaků na téměř 50 %. V časovém segmentu 2-4 vteřiny se již výsledky klasifikace o něco snížily a výsledky z ACG a EMG dat si byly blízké.

čas po anotaci		ACG	EMG	ECG	všechny
0-2 s	ACC (%)	74.9	81.8	65.4	82.9
	TPR (%)	44.8	48.9	39.1	49.6
	PPV (%)	43.7	47.7	38.3	48.6
2-4 s	ACC (%)	73.8	70.2	57.4	72.3
	TPR (%)	44.1	42.0	34.3	43.2
	PPV (%)	42.5	39.4	33.5	41.0

**Tabulka 3.4.** Statistické vyhodnocení RF klasifikátoru testovacích dat: vypočtené průměry přesnosti (ACC), senzitivity (TPR) a pozitivní predikce hodnot (PPV).

### 3.4 Detekce parasomnie v celonočním záznamu

Na základě detekce aktivity a klasifikace příznaků byla provedena detekce parasomnie v celonočním záznamu. Aby do vyhodnocení nezanášelo bdění nepřesnosti, tyto úseky se ze záznamu vyřadily. Využity k tomu byly vytvořené hypnogramy, které obsahovaly informaci o fázích spánku. Jakmile se v záznamu poprvé zobrazila jiná fáze než bdění, označil se tento bod jako počátek spánku.

Z celkových 35 záznamů obsahovalo hypnogram 32 záznamů, tudíž 3 záznamy se vyřadily, jelikož by nebylo možné odstranit nepřesnosti z fáze bdění pacienta.

Na jeden záznam se průměrně detekovalo přibližně 56 segmentů aktivity, ze kterých průměrně 17 bylo klasifikováno jako parasomnie. Některé záznamy sice spadaly pod pacienty s diagnostikovanou parasomnií, ale v konkrétní noc nebyla expertem detekována. Na těchto záznamech se detekovalo v průměru 16 falešných parasomních aktivit.

Celkové statistické porovnání s reálnými anotacemi se zobrazilo do Tab. 3.5. Statistické vyhodnocení bylo zprůměrováno přes všechny záznamy a vyhodnoceno se směrodatnou odchylkou.

Stat. vyhodnocení	Průměr	Směrodatná odchylka
TPR (%)	42.0	46.0
PPV (%)	5.3	6.3

**Tabulka 3.5.** Statistické vyhodnocení *Random forest* klasifikátoru při detekci v celonočním záznamu: vypočtené průměry a rozptyly senzitivity (TPR) a pozitivní predikce hodnot (PPV).

Z tabulky lze vidět, že směrodatná odchylka senzitivity byla vysoká. Vysokou odchylku způsobuje to, že v některých záznamech je parasomnie detekována ve 100 % případech a v jiných naopak v 0 %.



# Kapitola 4

## Diskuze

### 4.1 Předzpracování dat

Cílem předzpracování aktigrafického a polysomnografického signálu bylo získat příznaky, díky kterým by se dalo klasifikovat parasomní aktivitu a odlišit ji tak od fyziologické. Tento postup se rozdělil dle toho, s jakou přesností anotací bylo možné pracovat.

K prvnímu předzpracování aktigrafických dat bylo možné přistupovat jednoduchým přístupem, jelikož se časy detekcí určily s přesností na minuty. Parametry pro identifikaci různých aktivit byly získány postupem, který je znázorněn na Obr. 2.10. Z porovnání histogramů a ROC křivek na Obr. 3.1 a 3.2 však nebyl vizuálně nalezen znatelný rozdíl mezi fyziologickou a parasomní aktivitou. Tento výsledek mohlo způsobit hned několik faktorů.

Při detekci aktivity z dvouvteřinových epoch mohlo dojít ke ztrátě informace zvoleného časového úseku. Do budoucí práce by mohlo zkrácení epochy vynést přesnější výsledky. Dalším pevně nastaveným parametrem byla toleranční vzdálenost (Obr. 2.9), jejíž hodnota dosahovala 15 s. Takováto doba klidu mezi aktivitou může být příliš vysoká, tím pádem dochází k prodloužení aktivního segmentu a následnému zkreslení vypočítaných parametrů.

Z poznatků získaných během této práce bylo zjištěno, že pro rozeznání různých aktivit je klíčových několik prvních vteřin aktivity pacienta. Proto se časové anotace upravily a expertkou vyhodnotily s přesností na vteřinu. Poté se dalo pozorovat události v detailnějším měřítku. Následnou optimalizací se vypočetly další příznaky. Jelikož jedním z častých projevů doprovázející parasomní epizody jsou rychlé záškuby pacienta, mohla by klasifikaci zlepšit dynamika aktivity na začátku epizody. Z toho důvodu se přidaly další příznaky, které byly vypočteny v prvních několika vteřinách od počátku detekce. Vytvořily se tedy kratší události (4 s před a 10 s po počátku aktivity).

Kromě aktigrafických dat bylo možné do studie zahrnout i jiné signály, které obsahoval polysomnografický záznam. Srdeční aktivita (EKG) během spánku by mohla mít potenciální využití pro rozlišení parasomní a fyziologické aktivity. Dalším možným parametrem zachycující pohyb člověka byla zaznamenaná svalová elektrická aktivita (EMG).

Z těchto tří biologických signálů se extrahovaly příznaky v každém dvouvteřinovém úseku (viz Kapitola 2.3.4) a dle AUC hodnot (viz Obr. 3.3, 3.4, 3.5, 3.6, 3.7) bylo možné pozorovat právě zvolenou dynamiku v čase. Nejvyšší hodnoty se vždy vyskytovaly na počátku aktivity (0-2 s od začátku detekce aktivity). Maximální AUC hodnotou byla z EMG signálu brady 89 %. Proto se vybraly příznaky v úseku 0-2 a 2-4 vteřiny. Tyto příznaky byly použity pro následnou klasifikaci.

Na Obr. 3.5 a 3.6 lze porovnat AUC hodnoty biologických signálů EMG brady a končetin. Nejvyšší AUC hodnota EMG brady dosahovala 89 %, zatímco u EMG končetin 81 %. Z vizuálního porovnání lze říci, že EMG brady obsahuje větší informaci o klasifi-

kaci aktivit než EMG končetin. Nejnižší AUC hodnoty vykazuje příznak z EKG signálu s nejvyšší hodnotou 73 %.

Z EKG signálu lze dále optimalizovat příznaky. Kromě RR intervalu by bylo možné vypočítat variabilitu srdeční aktivity (HRV) nebo případně též využít výkonovou spektrální hustotu.

## 4.2 Klasifikace předzpracovaných dat

Na základě příznaků z předzpracovaných dat se provedla klasifikace. Podle toho, zda byla předzpracována data s minutovými nebo vteřinovými anotacemi, se použil vhodný klasifikátor.

U dat předzpracovaných s minutovými anotacemi byla provedena klasifikace SVM modelem. Přesnost klasifikace se vypočítala po 100 iteracích s hodnotou přibližně 75 % na trénovací sadě (Tab. 3.1). Na testovacích datech (Tab. 3.2) se přesnost snížila na 67 %. Nejvariabilnějším parametrem však byla senzitivita, jejíž průměrná hodnota na trénovacích datech se pohybovala kolem 26 % se směrodatnou odchylkou 34 %. Na testovací sadě klesla průměrná hodnota senzitivity na 10 % se směrodatnou odchylkou 16 %. Tyto rozdílné hodnoty mohou naznačovat, že některé epizody lze detekovat, jiné nelze a pouze na rozdělení trénovací a testovací sady záleží, jaké výsledky pak budou vypočteny. Nepřesnost výsledků mohla vnést také nízká reprezentace pozitivní třídy (parasomní aktivita) oproti negativní (fyziologická aktivita), která byla v poměru 1:2.

Jelikož už samotné předzpracování dat nevykazovalo příliš vysokou odlišnost (AUC hodnoty), neočekávaly se od klasifikace vynikající výsledky. Z toho důvodu byl pro ilustraci zvolen nejjednodušší SVM model. Z výsledků lze vidět, že se model přeučuje v závislosti na trénovací a testovací sadě. Výsledky trénovací sady vyšly vyšší než testovací.

Na základě poznatků a odborných konzultací bylo zřejmé, že je pro klasifikaci různých aktivit zajímavý začátek epizody. Proto se data přeanotovala a zlepšování SVM modelu se dále nevěnovala pozornost.

Druhé předzpracování s vteřinovými anotacemi již vykazovalo vyšší rozlišovací schopnost. Použitím *Random forest* modelu byly využity všechny příznaky ze zvolených časových úseků. Celkový počet příznaků na tento úsek dosahoval 181 (vysoký počet z nich jsou výkonové spektrální hustoty ve vymezeném frekvenčním pásmu). Pozorování byla opět rozdělena na trénovací a testovací sadu (v poměru 7:3) a klasifikována. Aby se ověřila přesnost výsledků, rozdělila se v tomto poměru data a vyhodnotila po 100 opakováních. Klasifikace se rozdělila podle biologických signálů, poté se použila na příznaky všech biologických signálů najednou.

Výsledky klasifikace na trénovací (viz Tab. 3.3) a testovací sadě (viz Tab. 3.4) ukázaly, že nejužitečnější informaci obsahoval počátek aktivity 0-2 s. V tomto úseku měly nejvyšší přesnost příznaky z EMG signálu na testovací sadě a to 82 %. Příznaky z aktigrafického záznamu dosahovaly přesnosti testovací sady 75 %. Příznosným parametrem zde byla senzitivita, která se oproti předchozí metodě zvýšila z 10 % (viz Tab. 3.1) na 45 % u ACG záznamu. Senzitivita z EMG příznaků vyrostla na 49 %. Nejnižší přesnost 65 % vycházela z příznaků EKG signálu na testovacích datech. Tyto zprůměrované hodnoty měly též nižší směrodatnou odchylku než předchozí výsledky. Maximální směrodatná odchylka byla 6 %, průměrně se pohybovala mezi 2-3 %. Již zde tolik nezáviselo na rozdělení dat na trénovací a testovací, jelikož byly výsledky konzistentní s relativně nízkou směrodatnou odchylkou.

Pro budoucí práci by bylo vhodné využít i jiné modely, např. Markovovy modely.

V porovnání klasifikace spánkových fází podle aktigrafie (viz studie [21]) vyšly výsledky senzitivity detekce lepší. Dle výsledků této práce přináší aktigrafie potenciál při klasifikaci parasomnie. Některé epizody lze však podle aktigrafie obtížně zachytit.

### 4.3 Detekce parasomnie v celonočním záznamu

Pro zjištění, jaké výsledky by vycházely na novém, neanotovaném signálu, byla provedena detekce v celonočním záznamu. Tato detekce probíhala nezávisle na anotacích, tedy vyhodnocovala a klasifikovala celý záznam. Nejdříve byla detekována aktivita z aktigrafického záznamu pomocí globálního prahu. Tímto prahem a toleranční vzdáleností neaktivity mezi úseky se získal výsledný segment, který označoval aktivní segment. Parametry globální práh a toleranční vzdálenost byly pevně nastavené konstanty, mohly proto způsobovat při vyhodnocování nepřesnosti. A jelikož byl výpočet příznaků vázaný na počátek segmentu, mohla globální prahová hodnota ovlivnit dobu označení aktivity. Klasifikace segmentů se statisticky vyhodnotila s reálnými anotacemi (viz Tab. 3.5), přičemž průměrná senzitivita byla 42 % se směrodatnou odchylkou 46 %. Pozitivní predikce vyšla 5 % se směrodatnou odchylkou 6 %. Z těchto hodnot lze říci, že aplikovaný postup označí velký počet falešných pozitivit (falešné detekce parasomnie), ale nalezne téměř polovinu reálně se vyskytujících projevů.

Statistické zkreslení mohl být též způsobeno tím, že u některých záznamů pacientů s diagnostikovanou parasomnií, nebyla v daný celonoční záznam detekována parasomnie. Celkem se v datech nachází 11 takovýchto záznamů.

Touto metodou lze detekovat paroxysmální (záchvatovité) projevy, které však prokazatelně nedefinují pouze NREM parasomnií. Tento projev může vypadat obdobně u epilepsie. Nelze tedy pevně říci, že by se u pacientů mohlo jednat o epileptické či parasomní projevy.

Dalším potenciálním zlepšením zpracování dat je použití jiných biologických signálů, jako je např. EEG. Zpracování EEG signálu však vyžaduje náročnější zpracování, proto nebylo v rámci této práce zahrnuto.

Nejdůležitějším poznatkem této práce je, že pro rozlišení parasomní a fyziologické aktivity jsou nejužitečnější první vteřiny aktivity. V porovnání s EMG vykazuje klasifikace z aktigrafických dat o trochu nižší přesnost (82 % a 75 %). Na základě studie podobných přístupů je toto, pokud je mi známo, první prací zabývající se klasifikací parasomních epizod na základě aktigrafických dat.

## Kapitola 5

### Závěr

Explorační analýzou byly předzpracovány aktigrafické a následně i polysomografické signály. Nejdříve se zpracovaly aktigrafické signály s využitím minutových anotací od expertky. Vypočítaly se příznaky aktivních segmentů, přičemž nejvyšší AUC hodnotou příznaku bylo 67 %, ostatní se pohybovaly okolo 50-60 %. Z této sady příznaků se pomocí SVM klasifikátoru rozlišila parasomní aktivita s přesností 67 % na testovací sadě. Senzitivita, která měla na trénovací sadě vysokou směrodatnou odchylku (34 %), vyšla na testovací sadě 10 %. Pozitivní predikce odpovídala na testovacích datech 48 % se směrodatnou odchylkou 21 %.

Získáním vteřinových anotací se bylo možné více zaměřit na okolí počátku události. Proto se vytvořil kratší události, ze kterých se vypočítaly příznaky po dvouvteřinovém okně. Do těchto příznaků byly zahrnuty také další biologické signály: EMG a EKG. Na klasifikaci poté se použily příznaky s nejvyšší AUC hodnotami, tedy první 4 vteřiny od počátku aktivity. Pomocí *Random forest* modelu byla data opět rozdělena na trénovací a testovací sadu a statisticky vyhodnocena. Přesnost testovací sady se pohybovala okolo 82 % se senzitivitou 50 %. Pozitivní predikce vyšla 49 %. Nejvíce informací o rozdílnosti fyziologické a parasomní aktivity nesl EMG signál a těsně pod ním aktigrafický signál.

Nad rámec práce byla provedena detekce parasomnie v celonočním záznamu. Pomocí detekce aktivity aktigrafického záznamu se zjistily segmenty, ze kterých se extrahovaly příznaky několika prvních vteřin. Statistickým porovnáním výsledků s reálnými anotacemi vyšla senzitivita 42 % a pozitivní predikce 5 %.

V případě další práce by bylo dobré se více zaměřit právě na zpracování dat v prvních několika vteřinách počátku aktivity a zahrnout i další biologické signály, jako např. EEG. Pro možnost navázání a pokračování ve zpracování dat se vytvořil dataset aktigrafických i polysomnografických dat okolo anotací. Data jsou k dispozici na vyžádání u vedoucího této práce.

## Literatura

- [1] MUDr. Miroslav Moráň. Parasomnie v NREM spánku. *Neurologie pro Praxi*. 2002, 2002 (3), 3.
- [2] A polysomnographic and clinical report on sleep-related injury in 100 adult patients. *American Journal of Psychiatry*. 1989, 146 (9), 1166-1173. DOI 10.1176/ajp.146.9.1166.
- [3] "Buskova. "Neurology for practice". "2014", "15" ("4"), "186–188".
- [4] M. M. Ohayon, M. W. Mahowald, Y. Dauvilliers, A. D. Krystal a D. Leger. Prevalence and comorbidity of nocturnal wandering in the US adult general population. *Neurology*. 2012-05-14, 78 (20), 1583-1589. DOI 10.1212/WNL.0b013e3182563be5.
- [5] "Moran. "Internal Medicine for Practice". "2009", "11" ("10"), "466–470".
- [6] Christopher P. Derry, A. Simon Harvey, Matthew C. Walker, John S. Duncan a Samuel F. Berkovic. NREM Arousal Parasomnias and Their Distinction from Nocturnal Frontal Lobe Epilepsy. *Sleep*. 2009-12-01, 32 (12), 1637-1644. DOI 10.1093/sleep/32.12.1637.
- [7] Chiara Fois, Mary-Anne S. Wright, GianPietro Sechi, Matthew C. Walker a Sofia H. Eriksson. The utility of polysomnography for the diagnosis of NREM parasomnias. *Journal of Neurology*. 2015, 262 (2), 385-393. DOI 10.1007/s00415-014-7578-2.
- [8] Doc. MUDr. Bartoš Aleš, Ph.D. *Spáněk, paměť a kognitivní poruchy*. <https://docplayer.cz/132047067-Doc-mudr-ales-bartos-phd-ad-centrum-narodni-ustav-dusevniho-zdravi-neurologicka-klinika-uk-3-lf-a-fnkv-praha.html>.
- [9] Avi Sadeh a Christine Acebo. The role of actigraphy in sleep medicine. *Sleep Medicine Reviews*. 2002, 6 (2), 113-124. DOI 10.1053/smr.2001.0182.
- [10] Simone Boerema, Lex van Velsen, Leendert Schaake, Thijs Tönis a Hermie Hermens. Optimal Sensor Placement for Measuring Physical Activity with a 3D Accelerometer. *Sensors*. 2014, 14 (2), 3188-3206. DOI 10.3390/s140203188.
- [11] *GT3X*. 2011. <https://actigraphcorp.com/support/activity-monitors/gt3x/>.
- [12] Jeffer E. Sasaki, Dinesh John a Patty S. Freedson. Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport*. 2011, 14 (5), 411-416. DOI 10.1016/j.jsams.2011.04.003.
- [13] Miroslav Husák. *Mikrosenzory a mikroaktuátory*. 2008 vydání. Praha: Academia, 2008. ISBN 978-80-200-1478-8.
- [14] Marcel Kreidl a Radislav Šmíd. *Technická diagnostika*. 2006 vydání. Praha: BEN - technická literatura, 2006. ISBN 80-730-0158-6.
- [15] Steven M. Thurman, Nick Wasylyshyn, Heather Roy, Gregory Lieberman, Javier O. Garcia, Alex Asturias, Gold N. Okafor, James C. Elliott, Barry Giesbrecht,

- Scott T. Grafton, Sara C. Mednick, Jean M. Vettel a Raffaele Ferri. Individual differences in compliance and agreement for sleep logs and wrist actigraphy. *PLOS ONE*. 2018-1-29, 13 (1), DOI 10.1371/journal.pone.0191883.
- [16] Daniel J. Buysse, Charles F. Reynolds, Timothy H. Monk, Susan R. Berman a David J. Kupfer. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*. 1989, 28 (2), 193-213. DOI [https://doi.org/10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4).
- [17] Yunyoung Nam, Yeesock Kim a Jinseok Lee. Sleep Monitoring Based on a Tri-Axial Accelerometer and a Pressure Sensor. *Sensors*. 2016, 16 (5), DOI 10.3390/s16050750.
- [18] Janna Mantua, Nickolas Gravel a Rebecca Spencer. Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography. *Sensors*. 2016, 16 (5), DOI 10.3390/s16050646.
- [19] Roger J. Cole, Daniel F. Kripke, William Gruen, Daniel J. Mullaney a J. Christian Gillin. Automatic Sleep/Wake Identification From Wrist Activity. *Sleep*. 1992-09-01, 15 (5), 461-469. DOI 10.1093/sleep/15.5.461.
- [20] Avi Sadeh, M. Sharkey a Mary A. Carskadon. Activity-Based Sleep-Wake Identification. *Sleep*. 1994-05-01, 17 (3), 201-207. DOI 10.1093/sleep/17.3.201.
- [21] M. Náhlík. *Detekce cyklu spánku-bdění a spánkových stádií z aktigrafického záznamu*. 2019.
- [22] Elena Beani, Martina Maselli, Elisa Sicola, Silvia Perazza, Francesca Cecchi, Paolo Dario, Irene Braitto, Roslyn Boyd, Giovanni Cioni a Giuseppina Sgandurra. Actigraph assessment for measuring upper limb activity in unilateral cerebral palsy. *Journal of NeuroEngineering and Rehabilitation*. 2019, 16 (1), DOI 10.1186/s12984-019-0499-7.
- [23] Alessandro Zampogna, Alessandro Manoni, Francesco Ascì, Claudio Liguori, Fernanda Irrera a Antonio Suppa. Shedding Light on Nocturnal Movements in Parkinson's Disease. *Sensors*. 2020, 20 (18), DOI 10.3390/s20185171.
- [24] Mikkel Schneller, Mogens Pedersen, Nidhi Gupta, Mette Aadahl a Andreas Holtermann. Validation of Five Minimally Obstructive Methods to Estimate Physical Activity Energy Expenditure in Young Adults in Semi-Standardized Settings. *Sensors*. 2015, 15 (3), 6133-6151. DOI 10.3390/s150306133.
- [25] Md Aktaruzzaman, Massimo Walter Rivolta, Ruby Karmacharya, Nello Scabottolo, Luigi Pugnetti, Massimo Garegnani, Gabriele Bovi, Giovanni Scalerà, Maurizio Ferrarin a Roberto Sassi. Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification. *Computers in Biology and Medicine*. 2017, 89 212-221. DOI 10.1016/j.combiomed.2017.08.006.
- [26] David J. Kupfer, Thomas P. Detre, Gordon Foster, Gary J. Tucker a Jose Delgado. The application of delgado's telemetric mobility recorder for human studies. *Behavioral Biology*. 1972, 7 (4), 585-590. DOI [https://doi.org/10.1016/S0091-6773\(72\)80220-7](https://doi.org/10.1016/S0091-6773(72)80220-7).
- [27] Sonia Ancoli-Israel, Roger Cole, Cathy Alessi, Mark Chambers, William Moorcroft a Charles P. Pollak. The Role of Actigraphy in the Study of Sleep and Circadian Rhythms. *Sleep*. 2003-05-01, 26 (3), 342-392. DOI 10.1093/sleep/26.3.342.

- [28] Avi Sadeh, Peter J. Hauri, Daniel F. Kripke a Peretz Lavie. "The Role of Actigraphy in the Evaluation of Sleep Disorders". *Sleep*. 1995, 18 (4), 288-302. DOI 10.1093/sleep/18.4.288.
- [29] R Dick, T Penzel, I Fietze, M Partinen, H Hein a J Schulz. AASM standards of practice compliant validation of actigraphic sleep analysis from SOMNOwatch™ versus polysomnographic sleep diagnostics shows high conformity also among subjects with sleep disordered breathing. *Physiological Measurement*. 2010-12-01, 31 (12), 1623-1633. DOI 10.1088/0967-3334/31/12/005.
- [30] Olivia Walch, Yitong Huang, Daniel Forger a Cathy Goldstein. "Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device". *Sleep*. 2019, 42 (12), DOI 10.1093/sleep/zsz180. zsz180.
- [31] Emi Yuda, Yutaka Yoshida, Ryujiro Sasanabe, Haruhito Tanaka, Toshiaki Shiomi a Junichiro Hayano. Sleep Stage Classification by a Combination of Actigraphic and Heart Rate Signals. *Journal of Low Power Electronics and Applications*. 2017, 7 (4), DOI 10.3390/jlpea7040028.
- [32] Alexander T.M. Van de Water, Alison Holmes a Deirdre A. Hurley. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography - a systematic review. *Journal of Sleep Research*. 2011, 20 (1pt2), 183-200. DOI 10.1111/j.1365-2869.2009.00814.x.
- [33] Trevor Hastie, Robert Tibshirani a Jerome Friedman. *The elements of statistical learning*. 2nd ed vydání. New York: Springer, c2009. ISBN 978-0-387-84858-7.
- [34] Gareth James, Daniela Witten, Trevor Hastie a Robert Tibshirani. *An introduction to statistical learning*. 1st vydání. New York: Springer, [2013. ISBN 978-1-4614-7137-0.
- [35] Leo Breiman. Bagging predictors. *Machine Learning*. 1996, 24 (2), 123-140. DOI 10.1007/BF00058655.
- [36] Leo Breiman. Random Forests. *Machine Learning*. 2001, 45 (1), 5-32. DOI 10.1023/A:1010933404324.
- [37] Seymour Epstein. Aggregation and beyond. *Journal of Personality*. 1983, 51 (3), 360-392. DOI 10.1111/j.1467-6494.1983.tb00338.x.
- [38] Samuel R. LaMunion, David R Bassett a Scott E. Crouter. Effects of Changing ActiGraph Bandpass Filter Width For Detecting Walking and Running. 2016, 48 546-547. DOI 10.1249/01.mss.0000486644.57610.81.
- [39] Jan Christian Brønd a Daniel Arvidsson. Sampling frequency affects the processing of Actigraph raw acceleration data to activity counts. *Journal of Applied Physiology*. 2016-02-01, 120 (3), 362-369. DOI 10.1152/jappphysiol.00628.2015.
- [40] Jan Christian Brond, Lars Bo Andersen a Daniel Arvidsson. Generating ActiGraph Counts from Raw Acceleration Recorded by an Alternative Monitor. 2017, 49 (11), 2351-2360. DOI 10.1249/MSS.0000000000001344.
- [41] Wilbur R. Knorr. On Heiberg's Euclid. *Science in Context*. 2001, 14 (1-2), 133-143. DOI 10.1017/S0269889701000059.
- [42] *Statistics and Machine Learning Toolbox User's Guide*. 2021.
- [43] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006, 27 (8), 861-874. DOI 10.1016/j.patrec.2005.10.010.

- [44] Davide Chicco a Giuseppe Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020, 21 (1), DOI 10.1186/s12864-019-6413-7.



# Příloha A

## Zkratky a symboly

### A.1 Zkratky

PSG	Polysomnografie (Polysomnography).
ACG	Aktigrafie (Actigraphy).
EEG	Elektroencefalografie (Electroencephalography).
EMG	Elektromyografie (Electromyography).
EKG	Elektrokardiografie (Electrocardiography).
EOG	Elektrookulografie (Electrooculography).
HRV	Srdeční variabilita srdce (Heart rate variability).
ROC	Operační charakteristika přijímače (Receiver operating characteristics).
AUC	Plocha pod křivkou (Area under curve).
TP	Správně odhadnuté pozitivní případy (True positive).
FP	Špatně odhadnuté pozitivní případy (False positive).
FN	Špatně odhadnuté negativní případy (False negative).
SVM	Metoda podpurných vektorů (Support vector machine).
RBF	Radiální základní funkce (Radial basis function).
NÚDZ	Národní ústav duševního zdraví.
OSA	Obstrukční spánková apnoe (Obstructive sleep apnea).
PLMS	Periodické pohyby končetinami ve spánku (Periodic limb movements in sleep).
REM	Rychlý pohyb očí (Rapid eye movement).
non-REM	Žádný rychlý pohyb očí (Non-rapid eye movement).
PSD	Výkonová spektrální hustota (Power spectral density).
SVM	Metoda podpurných vektorů (Support vector machine).

### A.2 Symboly

$G$	Gravitační zrychlení.
$SpO_2$	Saturace krve kyslíkem.

# Příloha B

## Soubory

### B.1 Přílohy diplomové práce

Zdroj_kod.zip	Zdrojový kód v prostředí MATLAB®.
ParasomnDataset.zip	Datový set parasomních dat. Z důvodu většího objemu dat jsou k dispozici na vyžádání u vedoucího práce Ing. Eduarda Bakšteina, Ph.D.

### B.2 Soubory zdrojového kódu

acgrotation.m	Funkce pro výpočet aktigrafických příznaků rotace.
ActivWithTolerance.m	Funkce pro vyhodnocení aktivity s toleranční vzdáleností.
after_max.m	Funkce pro výpočet prahové hodnoty z histogramu událostí.
AVGdiff.m	Funkce pro výpočet dynamických příznaků z několika prvních vteřin segmentu.
classifierStatistics.m	Funkce pro výpočet statistických hodnot.
createActSegment.m	Funkce pro vytvoření aktivního segmentu dle klasifikace.
CreateDataset.m	Skript pro vytvoření datasetu pro budoucí práce.
createEv.m	Funkce pro uložení události před a po anotaci.
ECGRdetection.m	Funkce pro detekci pozic R-špiček EKG signálu.
event_data.m	Funkce pro uložení parametrů k vytvoření události.
featextract.m	Funkce pro uložení příznaků z ACG a PSG dat.
featuresActi.m	Funkce pro výpočet příznaků z agregovaných aktigrafických dat.
fillnan.m	Funkce pro doplnění NaN hodnot do matice.
filtActi.m	Funkce pro filtraci dat pásmovou propustí.
hpg_statistics.m	Skript pro statistické vyhodnocení fází hypnogramu při anotacích.
minmaxnorm.m	Funkce pro normalizaci metodou min-max.
ParasomnDetection.m	Skript pro detekci parasomnie v celonočním záznamu.
plotXX.m	Funkce pro zobrazení zpracovaných signálů.
preprocessActi.m	Skript pro předzpracování aktigrafických dat z minutových anotací.
preprocessRawActi.m	Skript pro předzpracování aktigrafických dat a vybraných PSG signálů z vteřinových anotací.
psd.m	Funkce pro výpočet PSD.
readDFile.m	Funkce pro načtení PSG dat z D-souboru.

<code>RF_classifier.m</code>	Funkce pro klasifikaci dat pomocí <i>Random forest</i> modelu.
<code>RRdiff.m</code>	Funkce pro výpočet maximální změny RR intervalu.
<code>save_ACG.m</code>	Skript pro uložení hrubých aktigrafických dat a výpočet událostí.
<code>save_ECG.m</code>	Skript pro uložení hrubých EKG dat a výpočet událostí.
<code>save_EEG.m</code>	Skript pro uložení hrubých EEG dat.
<code>save_EMG.m</code>	Skript pro uložení hrubých EMG dat a výpočet událostí.
<code>save_hypnogram.m</code>	Skript pro uložení hypnogramů.
<code>save_PSD.m</code>	Funkce pro vytvoření datasetu událostí z PSD dat.
<code>save_RESP.m</code>	Skript pro uložení hrubých dat dýchání a výpočet událostí.
<code>savemy_dataset.m</code>	Skript pro uložení aktigrafických dat.
<code>segmentation.m</code>	Funkce pro vytvoření segmentů ze vstupního signálu.
<code>SVM_classifier.m</code>	Funkce pro klasifikaci a vyhodnocení dat pomocí SVM klasifikátoru.
<code>timedetection.m</code>	Funkce pro uložení časových anotací z excel tabulky.

### **B.3 Soubory vytvořeného datového setu**

<code>Informace datového setu.docx</code>	Informace celého souboru.
<code>ParasomniaData.zip</code>	MAT soubory datasetu pro budoucí práci.
<code>PatientInfo.xlsx</code>	Pacientské informace.
<code>PSGchannels.xlsx</code>	Informace o kanálech PSG.