

Bakalářská práce



České  
vysoké  
učení technické  
v Praze

**F3**

Fakulta elektrotechnická  
Katedra teorie obvodů

## Klasifikace parasomních epizod v celonočním polysomnografickém záznamu

**Marie Turnovcová**

Vedoucí: Ing. Eduard Bakštein, Ph.D.

Studijní program: Lékařská elektronika a bioinformatika

Květen 2022



## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Turnovcová** Jméno: **Marie** Osobní číslo: **483053**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra teorie obvodů**  
Studijní program: **Lékařská elektronika a bioinformatika**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Klasifikace parasomních epizod v celonočním polysomnografickém záznamu**

Název bakalářské práce anglicky:

**Classification of parasomnia episodes in polysomnography recordings**

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 z celonočních polysomnografických (PSG) záznamů.

Pokyny k vypracování:

1. Nastudujte problematiku PSG snímání a proveďte rešerši existujících metod pro detekci či klasifikaci událostí v PSG záznamech.
2. Proveďte explorační analýzu a identifikujte vhodné příznaky či navrhněte jiný vhodný postup pro rozlišení anotovaných parasomních a fyziologických epizod.
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] 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>
- [2] Castillo, M. (2021) Klasifikace parasomních epizod z 3-osých aktigrafických záznamů
- [3] Camaioni M, Scarpelli S, Gorgoni M, Alfonsi V, De Gennaro L. EEG Patterns Prior to Motor Activations of Parasomnias: A Systematic Review. *Nat Sci Sleep*. 2021;13:713-728 <https://doi.org/10.2147/NSS.S306614>

Jméno a pracoviště vedoucí(ho) bakalářské práce:

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

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

Datum zadání bakalářské práce: **02.02.2022**

Termín odevzdání bakalářské práce: **20.05.2022**

Platnost zadání bakalářské práce: **30.09.2023**

Ing. Eduard Bakštein, 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

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

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studentky

## Poděkování

Ráda bych poděkovala vedoucímu mé bakalářské práce Ing. Eduardu Bakšteinovi, Ph.D. za jeho čas a poskytnutí cenných rad a také za velkou ochotu a trpělivost při konzultacích této práce.

Děkuji také mé rodině a mým blízkým za jejich podporu během studia.

## Prohlášení

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

V Praze, 20. května 2022

## Abstrakt

Tato bakalářská práce se věnuje tématu jedné z poruch spánku zvané parasomnie a její detekci z EEG signálů nahraných pomocí polysomnografie. Detekce a klasifikace parasomnie je zatím obvykle prováděna spánkovým expertem, což je časově velmi náročné. Automatická detekce by v této oblasti byla velkým přínosem. V této práci byly během explorační analýzy z kanálů Cz a Fz EEG signálů a také z nezávislých komponent (nalezených pomocí analýzy nezávislých komponent (ICA) v programu EEGLAB) vybrány a extrahovány příznaky. Tyto příznaky byly dále využity ve vytvořených klasifikátorech. V některých z klasifikátorů byly použity metody rozhodovacích stromů a Random Forest. Klasifikátory byly následně evaluovány na zadaném datasetu obsahujícím parasomní a fyziologické epizody. Nejlepší klasifikátor využívající příznaky získané z kanálu Fz a Cz dosáhl přesnosti 81% se senzitivitou 73% a nejlepší klasifikátor využívající příznaky získané z nezávislých komponent dosáhl přesnosti 83% se senzitivitou 78%.

**Klíčová slova:** parasomnie, klasifikace, polysomnografie, EEG

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

## Abstract

This bachelor thesis deals with a topic of sleep disorder called parasomnia and its detection from EEG signals recorded by polysomnography. Detection and classification of parasomnias is so far usually performed by a sleep expert, which is very time consuming. An automatic detection would be of great help in this field. In this work, during exploratory analysis, symptoms were selected and extracted from Cz and Fz channels of EEG signals as well as from independent components (found by independent component analysis (ICA) in EEGLAB software). These features were further used in the developed classifiers. Decision tree and Random Forest methods were used in some of the classifiers. The classifiers were then evaluated on a given dataset containing parasomatic and physiological episodes. The best classifier using features extracted from the Fz and Cz channels achieved an accuracy of 81% with a sensitivity of 73%, and the best classifier using features extracted from independent components achieved an accuracy of 83% with a sensitivity of 78%.

**Keywords:** parasomnia, classification, polysomnography, EEG

**Title translation:** Classification of parasomnia episodes in polysomnography recordings

## Obsah

<b>1 Úvod</b>	<b>1</b>	3.2.2 Použití ICA pro nalezení EEG artefaktů . . . . .	14
<b>2 Úvod do problematiky</b>	<b>3</b>	3.3 Příznaky a klasifikátory . . . . .	15
2.1 Parasomnie . . . . .	3	3.3.1 Kritéria hodnocení klasifikátorů . . . . .	15
2.1.1 NREM parasomnie . . . . .	4	3.3.2 Použité příznaky . . . . .	17
2.1.2 Diagnostika parasomnie . . . . .	4	3.3.3 Použité klasifikátory . . . . .	20
2.1.3 Automatická detekce parasomních epizod . . . . .	5	3.3.4 Naimplementované klasifikátory na základě nezávislých komponent	23
2.2 Polysomnografie . . . . .	5	3.3.5 Klasifikátory na základě záznamu z 2 kanálů EEG . . . . .	30
2.2.1 Užitečnost polysomnografie pro diagnostiku NREM parasomnií . . . . .	7	<b>4 Experimentální výsledky</b>	<b>33</b>
2.2.2 Detekce událostí v PSG . . . . .	7	4.1 Dataset . . . . .	33
<b>3 Použité metody</b>	<b>9</b>	4.2 Explorační analýza . . . . .	34
3.1 Analýza EEG signálu . . . . .	9	4.3 Diskuze výsledků . . . . .	40
3.1.1 EEG artefakty a analýza nezávislých komponent ICA . . . . .	11	<b>5 Závěr práce</b>	<b>43</b>
3.1.2 Projevy parasomnie v EEG signálu . . . . .	12	<b>A Literatura</b>	<b>45</b>
3.2 Analýza nezávislých komponent	12		
3.2.1 Definice ICA . . . . .	13		

## Obrázky

2.1 Grafická reprezentace polysomnografie z uživatelského hlediska, převzato z [11]. . . . .	6	4.1 Krosvalidace přes všechny skupiny. . . . .	34
2.2 Hypnogram, převzato z [23]. . . . .	6	4.2 PSD EEG kanálů Fz a Cz. . . . .	35
3.1 Systém 10-20, převzato z [25]. . . . .	10	4.3 Signál z EEG kanálu Fz a jeho frekvenční pásma na příkladu začátku jedné fyziologické a jedné parasomní epizody. Anotovaný začátek epizody se nachází v čase 0 s. . . . .	36
3.2 Bipolární (A) a unipolární(B) elektrody, převzato z [25]. . . . .	11	4.4 ROC křivky pro příznaky Fz kanálu. . . . .	38
3.3 Vyhodnocení nezávislé komponenty pomocí IClabel. . . . .	15	4.5 Nezávislá komponenta a její frekvenční pásma na příkladu začátku jedné fyziologické a jedné parasomní komponenty. Anotovaný začátek epizody se nachází v čase 0s. . . . .	39
3.4 ROC křivka, převzato z [3]. . . . .	17		
3.5 Diagram znázorňující použité příznaky pro jednu epizodu. . . . .	19		
3.6 Diagram znázorňující použité příznaky pro jednu epizodu. . . . .	20		
3.7 Rozhodovací strom, převzato z [6]. . . . .	21		
3.8 Random Forest, převzato z [28]. . . . .	23		
3.9 Sada příznaků trénovacích epizod použitá k natrénování rozhodovacího stromu pro allEpDT. . . . .	29		
3.10 Sada příznaků trénovacích epizod použitá k natrénování rozhodovacího stromu pro EpDT_3. . . . .	30		



## Tabulky

3.1 Základní frekvenční pásma mozkových vlny dle [25]. . . . .	11
3.2 Frekvenční rozsahy pro jednotlivá frekvenční pásma. . . . .	20
4.1 Počet pacientů v jednotlivých skupinách a počet pacientů s alespoň 1 anotovanou parasomní epizodou. . . . .	34
4.2 Rozlišovací schopnosti příznaků pro různá pásma. . . . .	37
4.3 Výsledné hodnoty všech klasifikátorů. . . . .	41





# Kapitola 1

## Úvod

Poruchy spánku a obtíže s nimi spojené jsou problémy, které se týkají velkého procenta populace a bývají zařazovány mezi civilizační choroby 21. století. Kvalitní spánek je zásadní jak pro fyzické, tak pro psychické zdraví, naopak nedostatečný či nekvalitní spánek zhoršuje schopnost zmírnit a zvládat pocity úzkosti a zároveň způsobuje zvýšení hladiny emocionálního stresu, což může vést ke zdravotním obtížím jako jsou například deprese, srdeční onemocnění nebo obezita. Mezi časté poruchy spánku patří například insomnie, spánková apnoe, hypersomnie nebo parasomnie, jejíž rozpoznávání je tato práce věnována.

Umělá inteligence a strojové učení přináší do zdravotnictví velký potenciál. Nabízí zpracování velkého objemu medicínských dat, které by jinak nebylo možné takto komplexně vyhodnotit, případně by pro odborníka bylo velmi náročné, nejen časově. I klasifikace parasomnie je časově velmi náročná – klasifikaci provádí spánkový expert, a to na základě dat z polysomnografického vyšetření. Aplikace metod strojového učení a datové analýzy by v tomto ohledu mohla značně ulehčit práci.

Polysomnografické vyšetření je velice cenná spánková metoda pro diagnostiku poruch spánku. V rámci tohoto vyšetření je běžně zaznamenáván také signál z elektroencefalogramu (EEG). V rámci této práce má být vytvořen klasifikátor pro rozlišení parasomních epizod právě na základě EEG signálu.

Cíle této práce jsou:

1. Nastudovat problematiku polysomnografického (PSG) snímání a provést rešerši existujících metod pro detekci či klasifikaci událostí v PSG záznamech.
2. Provést explorační analýzu a identifikovat vhodné příznaky pro rozlišení anotovaných parasomních a fyziologických epizod.
3. Vytvořit klasifikátor pro rozlišení parasomních a fyziologických epizod a evaluovat jeho vlastnosti na zadaném datasetu, včetně zhodnocení možnosti reálného nasazení vytvořeného algoritmu.

V kapitole (2) Úvod do problematiky je popsána porucha spánku parasomnie a lékařský postup studia spánku pomocí polysomnografie. V další kapitole (3), jež je věnována metodám použitým v této práci, je popsána analýza EEG signálu, výpočetní metoda analýza nezávislých komponent, také vybrané příznaky a klasifikátory využité při klasifikaci a dále popis implementovaných klasifikátorů. Předposlední kapitola (4) zahrnuje popis datasetu, na němž byla klasifikace provedena, podkapitolu o provedení explorační analýzy a na závěr diskuzi dosažených výsledků. Závěrečná kapitola (5) tuto bakalářskou práci shrnuje a uzavírá.

Klasifikátory byly naimplementovány v programu MATLAB. Stejně tak nástroj EEGLAB použit v této práci k nalezení nezávislých komponent je MATLAB toolbox.

## Kapitola 2

### Úvod do problematiky

Tato kapitola je uvedením do problematiky této bakalářské práce a odpovídá také na otázku motivace, která nás vede ke snaze vytvořit klasifikátor rozlišující parasomní a fyziologické epizody. Předmětem této práce je klasifikace parasomních epizod, a to konkrétně ze signálů EEG naměřených pomocí polysomnografie v Národním ústavu duševního zdraví (dataset je blíže popsán v části 4.1). Parasomnie, její projevy a diagnostika spolu s oddílem o automatické detekci této poruchy je popsána v části 2.1. Polysomnografie je popsána v podkapitole 2.2.

### 2.1 Parasomnie

Parasomnie jsou poruchy spánku, někdy také považované za poruchy probouzení, při kterých dochází k nepředvídatelným epizodám. Tyto epizody probíhají ve smíšeném stavu spánku a bdělosti a člověk při nich může nevědomě vykonávat řadu pohybů a jednání. Nejčastěji se tyto poruchy spánku projevují v hlubokém (delta) spánku a epizody, jenž tyto poruchy způsobují, si pacient obvykle nepamatuje [26].

Mezi parasomnie patří například somnambulismus (náměsíčnictví), noční děs, skřípání zuby ve spánku (bruxismus), spánková opilost, hypnagogické záškuby či poruchy příjmu potravy související se spánkem [26].

Tyto poruchy ovlivňují život postiženého v mnoha oblastech. Většinou ovlivňují negativně jeho celkový výkon, mohou vést k psychickým poruchám či úrazům, v některých případech může dojít až k ohrožení života. Proto je důležité tyto poruchy klasifikovat a následně léčit [26].

### 2.1.1 NREM parasomnie

Parasomnie s nerychlými pohyby očí (NREM parasomnie) jsou jednou ze skupin parasomnií, pocházejících primárně z fáze spánku s pomalými vlnami (NREM3), ke kterým pravděpodobně dochází v důsledku nedokonalého probuzení z NREM spánku. To se děje vlivem inkompletní kortikální aktivity na budivý podnět. Kromě této nestálosti NREM spánku a jeho fragmentace vlivem dalších onemocnění má na vznik NREM parasomnie vliv také genetická výbava: konkrétní parasomnie mohou být spojeny s určitými alelami [8],[13]. Mezi predisponující faktory pro vznik NREM parasomnie patří spánková deprivace, situační stres, spánek v cizím prostředí, horečka, zvukové stimuly nebo užívání některých léků či návykových látek [8].

NREM parasomnie zahrnují tři typická chování: probuzení se zmateností, somnambulismus a noční děs [18]. Vyskytují se převážně v dětském věku, zároveň ale postihují 2-4% dospělých. Typicky se NREM parasomní epizody dostavují v první třetině noci a trvají od desítek sekund i po minuty [8].

### 2.1.2 Diagnostika parasomnie

Diagnóza NREM parasomnie je často stanovena na základě klinické anamnézy pacienta anebo – v případě, že jde o časté epizody, které jsou komplikované, nebezpečné, přetrvávají i v dospělosti nebo v dospělosti vznikají – se provádí vyšetření ve spánkových centrech pomocí video-polysomnografie (vPSG), díky níž můžeme odhalit rysy připomínající NREM parasomni (například náhlé probuzení se zmateností ze spánku NREM3) [13]. V případech existující podobnosti s jinými záchvatovitými nočními příhodami může být klinická diagnóza obzvláště náročná. Polysomnografický obraz pak slouží také k vyloučení výskytu jiných poruch spánku nebo noční epilepsie. Avšak normální polysomnografický nálezní diagnózu NREM parasomnie nevyklučuje [18] [8]. Ačkoliv se v posledních letech pozornost zaměřila na klinické metody (klinický algoritmus nebo dotazník), které by odlišily noční epilepsii čelního laloku od parasomnie, zatím se žádná z metod neprokázala jako spolehlivá [18].

### ■ 2.1.3 Automatická detekce parasomních epizod

V diplomové práci [10] pracovala Mariana Castillo na klasifikaci parasomních epizod ze signálů EMG a EKG získaných polysomnografickým vyšetřením a na klasifikaci z 3-osých aktigrafických záznamů. Při klasifikaci devadesátivteřinových úseků obsahujících epizodu (parasomní nebo fyziologickou) se jí podařilo dosáhnout nejvyšší celkové přesnosti klasifikace modelu 83% se senzitivitou 50%. Avšak implementovaná detekce v celonočním záznamu, pro kterou byly vypočteny stejné příznaky jako při zpracování vteřinových anotací a které byly klasifikovány na stejném modelu (Random Forest) byla celková senzitivita průměrně 42% se směrodatnou odchylkou 46% a pozitivní predikce (tj. pravděpodobnost, že je pozitivně detekované pozorování opravdu z pozitivní třídy) pouze 5% se směrodatnou odchylkou 6%. Použití signálu EEG by mohlo přinést klasifikaci epizod zlepšení, proto je na něj tato práce zaměřena.

Další metody pro automatickou detekci parasomnií se v odborné literatuře nepodařilo dohledat.

## ■ 2.2 Polysomnografie

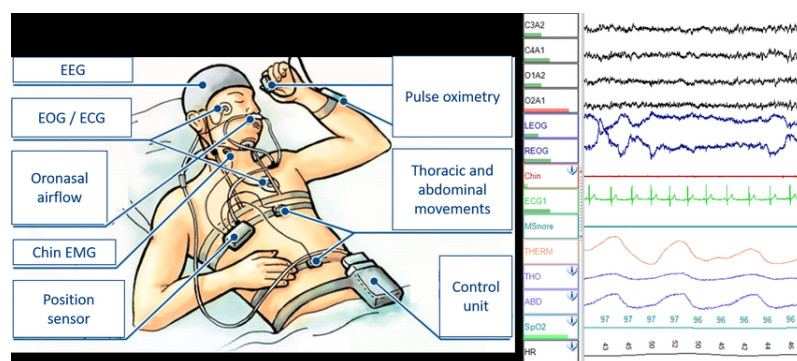
Polysomnografie (PSG) je lékařský postup složený z několika souběžných nezávislých testů, které během spánku monitorují některé tělesné funkce a pro jejich pozdější studium se pomocí různých kanálů zaznamenávají. Mezi hlavní poruchy, které se pomocí polysomnografie diagnostikují, patří spánková apnoe nebo jiné poruchy dýchání související se spánkem, narkolepsie, porucha chování při REM spánku (Rapid Eye Movement - REM spánek - fáze spánku charakterizována rychlými pohyby očí), nevysvětlitelná chronická nespavost nebo periodická porucha pohybu končetin [20].

Mezi polysomnografem zaznamenávané informace patří:

- EEG – elektroencefalogram, který měří a zaznamenává aktivitu mozkových vln pro identifikaci fází spánku a detekci záchvatové aktivity,
- EOG – elektrookulogram, který zaznamenává pohyby očí, což je důležité pro identifikaci různých fází spánku, zejména REM spánku,
- EMG – elektromyogram, který zaznamenává svalovou aktivitu,

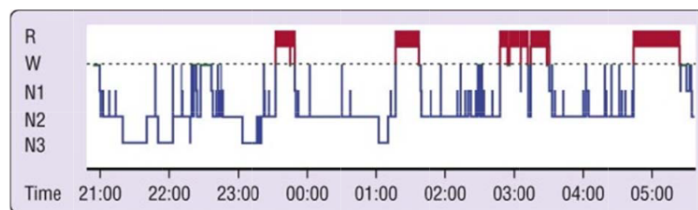
## 2. Úvod do problematiky

- EKG – elektrokardiogram zaznamenávající srdeční frekvenci a rytmus,
- pulzní oxymetrie,
- monitor dechu,
- teploměr,
- monitor krevního tlaku a další.



**Obrázek 2.1:** Grafická reprezentace polysomnografie z uživatelského hlediska, převzato z [11].

Na obrázku 2.2 je příklad hypnogramu, což je typické grafické zobrazení polysomnografie, které představuje stadia lidského spánku v průběhu času. [11]



**Obrázek 2.2:** Hypnogram, převzato z [23].

Podle potřeb konkrétních studií spánku jsou pak vybírány pouze některé specifické testy. Ty jsou také parametrizovány pro konkrétní případy. Například EEG se obvykle skládá z 10-16 elektrod, avšak u pacientů s epilepsi se často používá elektrod 20 [20].



### ■ 2.2.1 Užitečnost polysomnografie pro diagnostiku NREM parasomnií

Video-PSG (vPSG) má vysokou diagnostickou hodnotu u lidí s podezřením na NREM parasomnii. Ve studii [18] byla diagnóza stanovena v 76,5% případů z celkového počtu 124 pacientů. U 60,5% pacientů byla potvrzena NREM parasomnie, u 16% byla stanovena jiná diagnóza nebo se tyto diagnózy překrývaly.

U případů, které byly v klinické diagnóze jasné a anamnéza byla nekomplikovaná, potvrdila vPSG NREM parasomnii u 95% pacientů. U lidí, u kterých nebyla klinická diagnóza zřejmá, byla pravděpodobnost potvrzení NREM parasomnie pomocí vPSG 53%, zároveň ale identifikovala jiné neočekávané nebo nerozeznávané diagnózy (18%) nebo potenciální vyvolávající faktory (24%). Proto je vPSG doporučována právě v případech, kdy jsou v klinické anamnéze neobvyklé rysy.

### ■ 2.2.2 Detekce událostí v PSG

Události v polysomnografických záznamech jsou typicky hledány pomocí softwarových nástrojů vyvinutých speciálně pro spánkové laboratoře. Mezi detekované události patří spánkové vzrušení, delta vlny, pohyb očí, pohyb končetin nebo respirační události (jako je např. obstrukční spánková apnoe nebo hypopnoe [33]) a jiné.

Pro detekci, hodnocení a sledování vývoje poruch spánku mimo tyto laboratoře je zapotřebí automatická analýza zaznamenávaných signálů. V odborné literatuře můžeme nalézt různé přístupy používané pro detekci či klasifikaci událostí. Např. ve studii z roku 1987 [21] byly pro detekci událostí potřebných k rozlišení normálních a abnormálních lidských spánkových vzorců použity biologické signály včetně EEG a Markovovy modely, díky nimž je možné zachytit časové variace biologických signálů. Také v diplomové práci [5] z roku 2016 byly pro detekci událostí ve spánku použity Markovovy modely na signály nahrané pomocí PSG. Metodou klouzavého okna je v oné práci pro každý segment signálu provedena binární klasifikace (obsahuje událost/neobsahuje událost). Stejně tak v článku [4] byly využity Markovovy modely k online segmentaci tepu a klasifikaci elektrokardiogramů EKG.

Ve studii z roku 2009 [24], která byla zaměřena na rozpoznání syndromu obstrukční spánkové apnoe, byla využita metoda Support vector machine (SVM).

V dalších studiích jsou události klasifikovány např. pomocí spektrální a nelineární či kvadratické diskriminační analýzy, metody k-nejbližších sousedů, lineární regrese nebo také na základě srdečního rytmu a úrovně saturace kyslíkem [16]. S využitím metod strojového učení je možné automaticky detekovat události související s poruchami spánku s vysokou přesností [16].

Po nastudování existující literatury lze říci, že existující publikované přístupy se zaměřují zejména na detekci událostí respiračního charakteru. Pro automatickou detekci či klasifikaci parasomních epizod nebyly kromě již výše uvedených referencí nalezeny další relevantní práce.

# Kapitola 3

## Použité metody

V této kapitole jsou popsány použité metody pro klasifikaci parasomních a fyziologických epizod. Nejdříve je v podkapitole 3.1 popsán EEG signál, jeho artefakty a také jak se v něm projevuje parasomnie. Druhá podkapitola 3.2 je věnována analýze nezávislých komponent a jejímu využití k nalezení artefaktů v EEG signálu. Nejzrůsáhlejší část této kapitoly je podkapitola 3.3 popisující použité příznaky a klasifikátory a také kritéria, podle kterých jsou klasifikátory hodnoceny.

### 3.1 Analýza EEG signálu

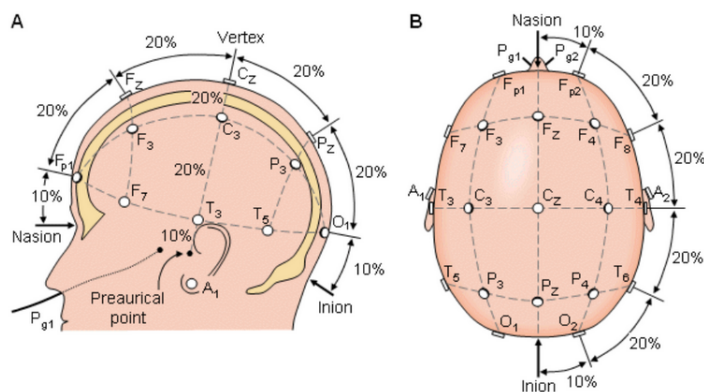
Elektroencefalografie (EEG) je diagnostická metoda, umožňující snímání a záznam elektrické aktivity mozku s vysokým časovým rozlišením.

Spontánní aktivita, která probíhá v živém jedinci nepřetržitě, se měří na pokožce hlavy (skalpu) nebo přímo na mozku (na mozkové kůře neboli kortexu) tzv. elektroencefalogramem. Amplituda EEG je asi 100  $\mu\text{V}$  při měření na pokožce hlavy a asi 1-2 mV při měření na povrchu mozku. Šířka pásma tohoto signálu je od méně než 1 Hz do přibližně 50 Hz [25].

Počet nervových buněk v mozku se odhaduje řádově na  $10^{11}$ . Kortikální neurony (neurony mozkové kůry) jsou silně propojené - povrch jednoho neuronu může být pokryt  $10^3$  až  $10^5$  synapsí [25]. EEG pak vzhledem k

povaze měření vně hlavy zaznamenává zejména synchronní aktivitu velkého počtu neuronů.

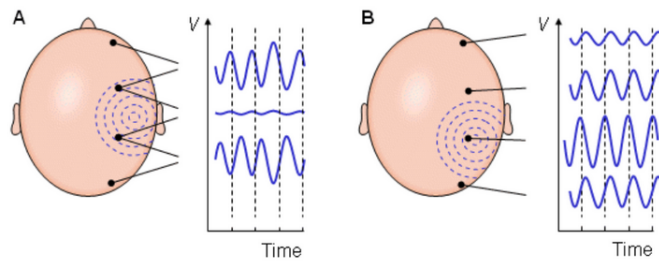
K záznamu spontánního EEG se obvykle používá mezinárodně standardizovaný systém 10-20 [25]. V tomto systému jsou elektrody umístěny na povrchu pokožky hlavy následovně: Referenčními body jsou horní část nosu v úrovni očí (Nasion) a kostěná boule na spodině lebky (Inion). Z těchto referenčních bodů jsou měřeny obvody lebky v předozadní (mediální) a pravolevé (frontální) rovině. Umístění elektrod je pak určeno rozdělením těchto obvodů na 10% a 20% intervaly, tedy vzdálenosti mezi sousedními elektrodami jsou buď 10% nebo 20% celkové mediální nebo frontální vzdálenosti lebky, jak zobrazuje obrázek 3.1.



**Obrázek 3.1:** Systém 10-20, převzato z [25].

Elektrody jsou označeny písmeny: A = Ear lobe (ušní boltec); C = Central (střední); P = Parietal (temenní); F = Frontal (přední); O = Occipital (týlní); T = Temporal (spánkové) a čísla (lichá čísla pro elektrody umístěné nad levou mozkovou hemisférou, sudá čísla pro elektrody nad pravou hemisférou).

Při měření EEG lze použít bipolární nebo unipolární elektrody. Obrázek 3.2 tyto varianty zobrazuje. U bipolárních elektrod se měří potenciálový rozdíl mezi párem elektrod, u unipolárních se potenciál každé elektrody porovnává buď s neutrální elektrodou nebo s průměrem všech elektrod [25].



**Obrázek 3.2:** Bipolární (A) a unipolární(B) elektrody, převzato z [25].

V signálu EEG typicky rozlišujeme pásma aktivity alpha ( $\alpha$ ), beta ( $\beta$ ), delta ( $\delta$ ), theta ( $\theta$ ) a gamma ( $\gamma$ ), jejichž frekvenční rozsahy nejsou zcela kodifikovány a napříč literaturou se mírně liší. Jeden z příkladů zvolených frekvenčních rozsahů pro tyto pásma je uveden v tabulce 3.1

Frekvenční pásmo	Frekvenční rozsah
Delta( $\delta$ )	0,5-4Hz
Theta( $\theta$ )	4-8Hz
Alpha( $\alpha$ )	8-13Hz
Beta( $\beta$ )	13-30Hz
Gamma( $\gamma$ )	>30Hz

**Tabulka 3.1:** Základní frekvenční pásma mozkových vlny dle [25].

### 3.1.1 EEG artefakty a analýza nezávislých komponent ICA

Elektroencefalogram zaznamenává kromě mozkové aktivity vyšetřovaného i další aktivitu jiných zdrojů. Každý elektrodový kanál EEG shromažďuje lineární směs všech vhodně promítnutých elektrických signálů. A to v závislosti na poloze a orientaci generátorů signálu a elektrických polí, které jsou vytvářeny vzhledem ke každému nahrávanému kanálu. Každý z těchto kanálů zaznamenává rozdíl mezi aktivitou na dvou nebo více elektrodách na hlavě. Tento proces míchání se vztahuje i na mozkovou aktivitu: vícekanálový záznam EEG obsahuje elektrické signály vzdáleného pole vycházející z různých oblastí mozku vyšetřovaného, ve kterých kortikální tkáň generuje synchronní elektrické potenciály.

Zaznamenaná aktivita zdrojů majících svůj původ jinde než ve zkoumané oblasti je označována jako artefakt. Mezi biologické artefakty patří potenciály vznikající v očích, elektromyografická aktivita spojená se silnými svalovými

kontrakcemi, které jsou dostatečně blízko snímajícím elektrodám, nebo také elektrocardiografické signály. Dalšími artefakty jsou nebiologické signály vyvolané například elektrickými zařízeními se střídavým proudem a stejně tak samotné elektrody - v případě, že je impedance na rozhraní elektroda-kůže velká nebo nestabilní [27]. Tyto artefakty jsou označovány jako artefakty technické.

Vzhledem ke způsobu snímání EEG jsou v signálech zejména sousedících elektrod značné překryvy. K rozlišení zdrojových procesů, které vedly k zaznamenaným signálům, se používá dekompozice pomocí analýzy nezávislých komponent (ICA), které je věnována část 3.2.

#### 3.1.2 Projevy parasomnie v EEG signálu

Na základě review Camaioni et al. [9] můžeme rozlišit následující projevy parasomnie v EEG signálu bezprostředně před parasomními událostmi. Některé studie, kterými se tento článek zabýval, v EEG bezprostředně před epizodami náměšičnosti, spánkovým děsem nebo zmateným vzrušením pozorovaly tzv. hypersynchronní deltu (HSD). HSD jsou kontinuální delta vlny napětí  $\geq 150 \mu V$ .

Všechny v článku uvedené studie analyzovaly behaviorální epizody NREM parasomnie, ke kterým došlo během spánku s pomalými vlnami (SWS – Slow Wave Sleep). Analýza projevů parasomnie v signálu EEG, jenž předchází nástupu behaviorální epizody, přinesla následující nálezy související s hlubokým spánkem: bezprostředně před motorickou aktivací ve frontálních a centrálních oblastech byla odhalena zvýšená aktivita pomalých vln (SWA – Slow Wave Activity) a pomalé oscilace (SO – Slow Oscillation), zároveň s prokázanou přítomností vysokofrekvenční aktivity EEG (tj. 24,00–30,00 Hz) bezprostředně před NREM parasomní epizodou. Podle [9] jsou tyto skutečnosti, které by v budoucnu mohly představovat prediktivní vzorce EEG pro nástup NREM parasomní epizody.

## 3.2 Analýza nezávislých komponent

Independent Component Analysis (ICA) neboli analýza nezávislých komponent je výpočetní metoda pro oddělení signálů patřící mezi metody slepé separace zdrojů (Blind Source Separation – BSS). Slepou separací zdrojů

je oddělována sada zdrojových signálů od sady smíšených signálů, a to bez informací (nebo pouze s velmi malým množstvím informací) o zdrojových signálech či o procesu jejich míchání. Nejznámějším příkladem, který BSS řeší, je problém koktejlového večírku: v místnosti je mnoho lidí, kteří hovoří současně a posluchač se snaží sledovat jednu z diskusí, která zde probíhá [15]. S tímto problémem si dokáže poradit lidský mozek, při digitálním zpracování signálu je to ale obtížný úkol.

Při měření povrchového potenciálu mozkové aktivity elektroencefalogramem dochází k podobnému problému. Naměřené potenciály jsou pravděpodobně smíchané signály jednotlivých center aktivit v mozku [2].

Metoda ICA může být nápomocná pro separaci artefaktů z EEG. Běžně se nejdříve transformují EEG signály do prostoru nezávislých zdrojových komponent a poté, co jsou ručně vybrány neumělé neurální zdroje a naopak artefakty obsahující komponenty jsou odstraněny, následuje rekonstrukce EEG bez artefaktů.

Metodu ICA pro rozdělení vícerozměrného signálu na jeho aditivní podsložky můžeme použít za předpokladu, že ony podsložky jsou na sobě statisticky nezávislé a zároveň že jsou to signály, které nemají Gaussovo (normální) rozložení hustoty pravděpodobnosti [2].

### ■ 3.2.1 Definice ICA

Předpokládáme  $n$  lineárních kombinací (směsí)  $x_1, \dots, x_n$ , z  $n$  nezávislých komponent

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n \quad \forall j. \quad (3.1)$$

Dále předpokládáme, že každá směs  $x_j$  i každá nezávislá komponenta  $s_k$  je náhodná veličina a také jak směsi tak nezávislé komponenty mají nulový průměr. (V případě, že to tak není,  $x_i$  se vycentrují odečtením výběrového průměru.) Pro předchozí rovnici využijeme maticovo-vektorový zápis:

$$x = As, \quad (3.2)$$

kde  $x$  je náhodný vektor s prvky  $x_1, \dots, x_n$  a  $s$  náhodný vektor s prvky  $s_1, \dots, s_n$  a matice  $A$  obsahuje prvky  $a_{ij}$ . Pozorujeme pouze  $x$ , zatímco  $A$

a  $s$  musíme odhadnout. Z odhadnuté matice  $A$  pak můžeme vypočítat její inverzní matici  $W$  a získat tak nezávislou komponentu:

$$s = Wx. \quad (3.3)$$

Toto je model ICA, který popisuje, jak jsou generována pozorovaná data procesem míchání komponent  $s_i$  [12].

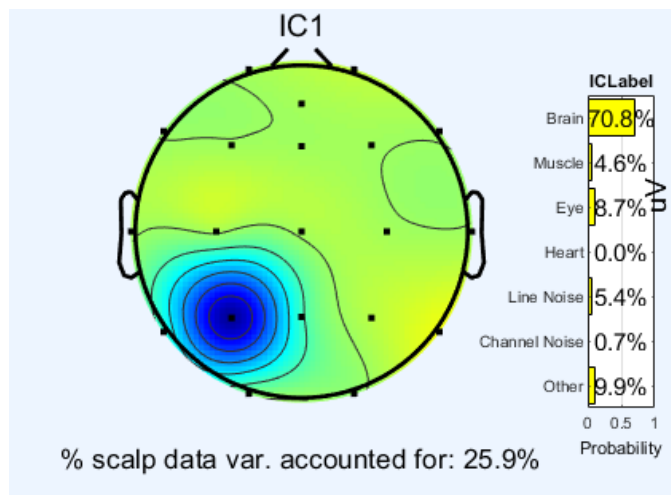
Tedy pokud bychom si aplikaci ICA chtěli představit na dříve zmíněném problému koktejlového večírku, nezávislé komponenty  $s$  jsou jednotlivé hlasy lidí, matice  $A$  představuje míchání hlasu lidí vlivem okolního prostředí a  $x$  jsou jejich lineárními kombinacemi, které zaznamenávají jednotlivé mikrofony [2].

### ■ 3.2.2 Použití ICA pro nalezení EEG artefaktů

Metoda ICA – opět za předpokladu, že zdroje signálu jsou lineární směsí mozkových a umělých zdrojů – může rozložit pozorovaný signál na nezávislé komponenty. Metoda byla testována na simulovaných a experimentálních datech, na nichž se prokázala jako účinná při separaci signálů z jejich lineárních směsí a při extrakci očních informací přítomných v signálech EOG [22].

MATLAB toolbox EEGLAB [14] obsahuje nástroj pro nalezení nezávislých komponent EEG signálu. Obvykle je používán pro odstranění artefaktů, které byly do dat přimíchány, a to tak, že nejdříve automaticky identifikuje šumové komponenty, které je dále možné odstranit, aniž by došlo k poškození originálních signálů EEG. EEGLAB obsahuje také IClab – plugin pro automatickou klasifikaci nezávislých komponent EEG [19]. Na obrázku 3.3 je zobrazeno vyhodnocení jedné nezávislé komponenty parasomní epizody pomocí tohoto nástroje.





Obrázek 3.3: Vyhodnocení nezávislé komponenty pomocí ICLabel.

### 3.3 Příznaky a klasifikátory

Příznak neboli feature je vlastnost pozorovaného jevu, kterou je možno měřit, analyzovat a také použít jako vstup do klasifikátoru [6]. Pokud zdrojová data obsahují velké množství redundantní informace (např. více korelovaných časových signálů), slouží proces extrakce příznaků k výpočtu menšího množství parametrů použitelných pro klasifikaci. Tím také obvykle dochází k účinnému snížení množství i redundance vstupních dat současně s vyšší efektivitou učení následných modelů.

Metody strojového učení pro učení funkce z trénovacích dat se rozdělují na učení s učitelem (supervised learning) a učení bez učitele (unsupervised learning). U učení s učitelem obsahují trénovací data spolu se vstupními hodnotami také požadované výstupy, zatímco u učení bez učitele tyto výstupní hodnoty nejsou známy [6]. V této práci byly použity metody strojového učení s učitelem.

#### 3.3.1 Kritéria hodnocení klasifikátorů

Pro vyhodnocení klasifikátorů byla použita kritéria senzitivita, specifická, přesnost a plocha pod ROC křivkou (AUC), která jsou dále popsána.

Senzitivita (TPR) neboli True Positive Rate je podmíněná pravděpodobnost, která vyjadřuje relativní četnost správně klasifikovaných pozitivních případů. Tedy kdy skutečně pozitivní případy jsou klasifikátorem ohodnoceny jako pozitivní [17].

$$TPR = P(TP|P) = \frac{TP}{TP + FN}, \quad (3.4)$$

kde  $TP$  (True Positive) jsou skutečně pozitivní označeny jako pozitivní a  $FN$  (False Negative) jsou skutečně pozitivní označeny jako negativní.

Specificita (TNR) neboli True Negative Rate je podmíněná pravděpodobnost, která vyjadřuje relativní četnost správně klasifikovaných negativních případů. Tedy kdy skutečně negativní případy jsou klasifikátorem ohodnoceny jako negativní.

$$TNR = P(TN|N) = \frac{TN}{TN + FP}, \quad (3.5)$$

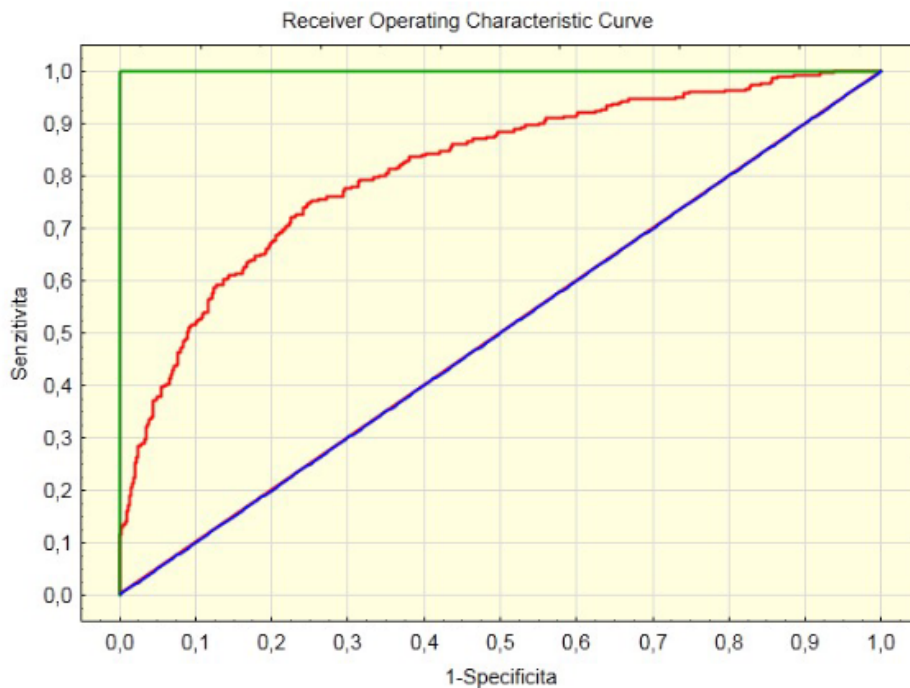
kde  $TN$  (True Negative) jsou skutečně negativní označeny jako negativní a  $FP$  (False Positive) jsou skutečně negativní označeny jako pozitivní.

Přesnost klasifikátoru  $P$ :

$$P = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.6)$$

ROC křivka (Receiver Operating Characteristic Curve) je graf, který znázorňuje vyhodnocení binárního klasifikátoru pro různé hodnoty rozhodovacího prahu. Na ose  $x$  je vykreslena pomocí parametru  $FPR$  (False Positive Rate), jehož výpočet je uveden v rovnici 3.7, a  $TPR$  (True Positive Rate) na ose  $y$  [29].

$$FPR = (1 - \text{specificita}) = \frac{FP}{FP + TN} \quad (3.7)$$



**Obrázek 3.4:** ROC křivka, převzato z [3].

ROC křivka je zobrazena na obrázku 3.4. Pro nějaký model může ROC křivka vypadat například jako křivka znázorněná červenou barvou, optimální model je zde vykreslen barvou zelenou (senzitivita i specificita v něm dosahují hodnoty 1).

AUC (Area Under the Curve) je pak plocha pod ROC křivkou, která vyjadřuje míru schopnosti klasifikátoru rozlišovat mezi třídami a používá se jako souhrn ROC křivky.

### ■ 3.3.2 Použité příznaky

Příznaky používané v klasifikátorech jsou: energie a relativní energie signálu 3.8, rozsah 3.9, maximální diference 3.10 a plocha pod křivkou ROC. Z analýzy některých příznaků v různých časech v bakalářské práci [10] bylo prokázáno, že nejužitečnější informace o rozlišení fyziologické a parasomní aktivity se nachází v několika prvních vteřinách od počátku aktivity. I proto byly všechny zmíněné příznaky kromě plochy pod křivkou extrahovány na základě explorační analýzy z prvních 4 sekund od počátku anotované epizody. Vypočteny byly následující příznaky:

Energie signálu neboli Root Mean Square (RMS)

$$RMS(X) = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}} \quad (3.8)$$

Relativní energie signálu (NRMS) je získána normalizací RMS.

Rozsah neboli Range (RNG)

$$RNG(X) = \max(X) - \min(X) \quad (3.9)$$

Maximální diference (MaxDiff)

$$MaxDiff(X) = \max(x_2 - x_1, x_3 - x_2, \dots, x_m - x_{m-1}) \quad (3.10)$$

### ■ Příznaky použité u klasifikátorů pracujících s nezávislými komponentami

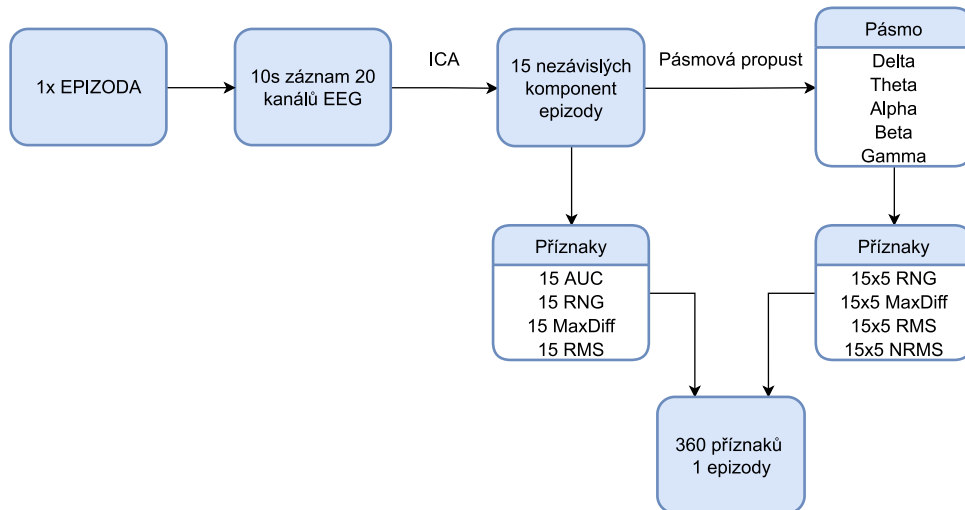
U klasifikátorů, ve kterých figurují nezávislé komponenty, byly použity následující příznaky nezávislých komponent:

- AUC pro rozptyl před epizodou a po začátku epizody,
- rozsah (RNG),
- maximální diference (MaxDiff),
- energie (RMS),
- relativní energie (NRMS).

Nezávislé komponenty byly nalezeny pomocí nástroje EEGLAB. Pro další práci bylo uloženo 15 nezávislých komponent. Toto číslo bylo zvoleno na základě explorační analýzy, která ukázala, že tyto komponenty obsahují užitečné množství artefaktických dat.

Příznaky AUC pro rozptyl, RNG, MaxDiff a RMS byly vypočítány pro originální signál nezávislé komponenty. Nezávislá komponenta byla pásmovou

propustí rozdělena do 5 frekvenčních pásem se zvoleným frekvenčním rozsahem popsaným v tabulce 3.2. Tato tabulka popisuje 6 frekvenčních pásem – pro nezávislé komponenty nebylo použito uvedené pásmo low delta. Pro každé z použitých pásem byly vypočítány příznaky RNG, MaxDiff, RMS a NRMS. Celkově je to tedy 24 příznaků na jednu komponentu, tedy 360 příznaků pro epizodu, což je zobrazeno na obrázku 3.5.



**Obrázek 3.5:** Diagram znázorňující použité příznaky pro jednu epizodu.

### ■ Příznaky použité u klasifikátorů pracujících s EEG kanály

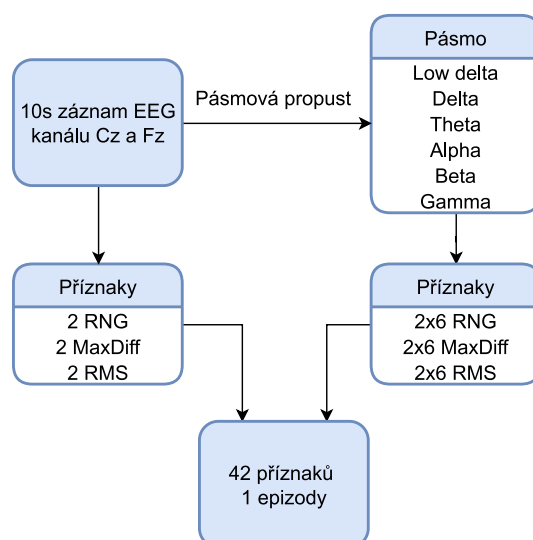
U klasifikátorů pracujících s Cz a Fz EEG kanály, které byly vybrány na základě review Camaioni et al. [9], byly použity následující příznaky:

- rozsah RNG,
- maximální difference MaxDiff,
- energie RMS.

Signály z EEG kanálu Cz a Fz byly pásmovou propustí rozděleny do 6 frekvenčních pásem se zvoleným frekvenčním rozsahem popsaným v tabulce 3.2. Celkově pak připadá 42 příznaků na jednu epizodu, což je zobrazeno na obrázku 3.6.

Frekvenční pásmo	Použitý frekvenční rozsah
Low delta	0.2-2Hz
Delta( $\delta$ )	0.5-4Hz
Theta( $\theta$ )	4-8Hz
Alpha( $\alpha$ )	8-13Hz
Beta( $\beta$ )	13-30Hz
Gamma( $\gamma$ )	30-45Hz

**Tabulka 3.2:** Frekvenční rozsahy pro jednotlivá frekvenční pásma.



**Obrázek 3.6:** Diagram znázorňující použité příznaky pro jednu epizodu.

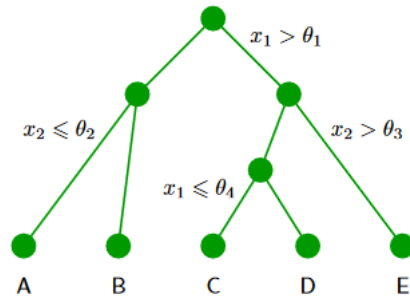
### 3.3.3 Použité klasifikátory

V tomto oddílu jsou popsány klasifikátory Decision Tree a Random Forest, které byly použity v naimplementovaných klasifikátorech, jež jsou popsány v následujících oddílech 3.3.4 a 3.3.5.

#### Decision Tree

Decision Tree neboli rozhodovací strom je klasifikátor predikující příslušnost k dané klasifikační třídě pro jednotlivé záznamy. Tento strom je naučený (zkonstruovaný) analýzou trénovací sady, která obsahuje záznamy včetně

příslušnosti k jednotlivým třídám, a následně je použit na sadě testovací. V tomto případě jde tedy o učení s učitelem. Abychom takový model z trénovací množiny naučili, musíme určit strukturu stromu, včetně toho, která vstupní proměnná je vybrána v každém uzlu k vytvoření kritéria rozdělení a také hodnotu prahového parametru pro rozdělení [6]. Na obrázku 3.7 je zobrazen příklad struktury rozhodovacího stromu.



**Obrázek 3.7:** Rozhodovací strom, převzato z [6].

U binárního rozhodovacího stromu poskytuje každý uzel informaci o rozhodnutí, které má být učiněno. Z každého uzlu vedou dvě výstupní hrany, z nichž jedna představuje splnění podmínky uzlu, druhá naopak nesplnění podmínky [1].

K větvení stromů je používáno různých metod, například metoda Information Gain, Chi-Square nebo Gini Impurity, jenž je dále popsána. Metoda Gini Impurity používá pro rozhodování o větvení stromu Gini Index. Ten lze pro zbývající dataset  $D$  v určité větvi vypočítat pomocí vzorce

$$Gini(D) = 1 - \sum_{i=1}^n p_i^2, \quad (3.11)$$

kde  $p_i$  je pravděpodobnost výskytu třídy  $i$  v datasetu  $D$  v dané větvi a  $n$  je počet tříd.

Pro každý příznak je spočítán vážený součet Gini indexů větví, které by vznikly rozdělením podle tohoto příznaku  $P$ . Následující vzorec zobrazuje výpočet váženého součtu Gini indexů pro binární rozhodovací strom.

$$Gini_P(D) = \frac{N_1}{N} Gini(D_1) + \frac{N_2}{N} Gini(D_2), \quad (3.12)$$

kde  $N$  je velikost zbylého datasetu v právě rozhodovaném uzlu,  $N_1$  resp.  $N_2$  je velikost datasetu v potenciálně vzniklých větvích po rozdělení podle příznaku  $P$ .

Následně je ze všech příznaků vybrán příznak s nejnižší hodnotou  $Gini_P(D)$  a podle něj se strom větví.

Pro oba (binární rozhodovací strom) vzniklé uzly se postup opakuje. Opět se pomocí vzorce 3.12 vypočítají vážené součty Gini indexů jednotlivých příznaků a podle příznaku s nejnižší hodnotou tohoto váženého součtu je strom v příslušném uzlu větven. Tento proces končí, pokud rozvětvení povede ke vzniku listu (list obsahuje data pouze z jedné třídy).

Je zde využito hladové (greedy) optimalizace, v níž je krok po kroku vybírán příznak, který nabízí bezprostřední přínos. Vzhledem k tomu je v rozhodovacím stromu důležité v určité chvíli přestat přidávat uzly, aby nedošlo k přeučení (overfitting). Toho můžeme dosáhnout například stanovením minimální velikosti listu nebo hloubky stromu (prořezáváním hotového stromu nebo zastavením růstu stromu dřív než jsou vyčerpána všechna trénovací data).

## ■ Random Forest

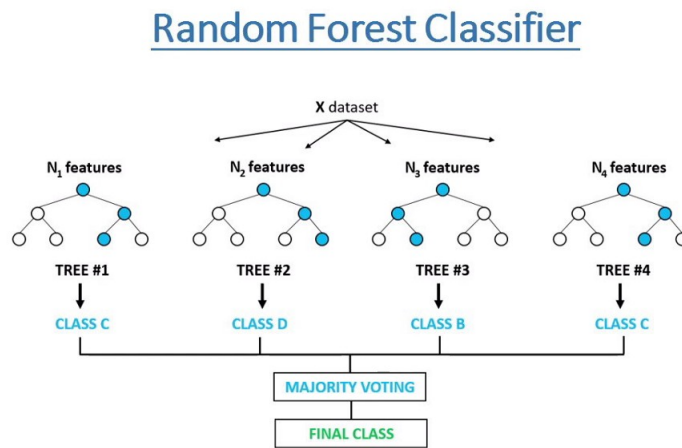
Random Forest neboli náhodný les je metoda strojového učení s učitelem jak pro klasifikaci, tak pro regresi. Pro rozhodnutí využívá velké množství rozhodovacích stromů s použitím různých náhodně vybraných podmnožin z celé trénovací sady. Tento proces vytvoření různých datových sad z jedné datové sady se nazývá bootstrapping [30]. V případě, že jsou pomocí procesu bootstrapping vytvořeny sady použité v rozhodovacích stromech, hovoříme o technice bagging.

Podmnožiny trénovací sady obsahují při trénování jednotlivé příznaky (pozorování) a výslednou třídu. Jejich velikost  $n$  je zvolena takto:  $n = \log_2(N + 1)$ , kde  $N$  je velikost celé trénovací sady [31].



Po sestavení jednotlivých rozhodovacích stromů není prováděno prořezávání, stromy tedy mohou být přeučené. Zároveň jsou nezávislé, každý z nich provede klasifikaci. Celkový výsledek je pak vybrán podle toho, jakou třídu klasifikovala většina stromů [7]. Tento postup znázorňuje obrázek 3.8, ve kterém jsou zobrazeny 4 rozhodovací stromy. Každý z nich klasifikuje na základě vstupní podmnožiny dat (například  $N_1$  features). V tomto případě bude výslednou klasifikací pomocí Random Forest vybrána třída C.

Pro dosažení vyšší efektivity klasifikátoru Random Forest, než jaké je dosaženo samotným rozhodovacím stromem, je potřeba natrénovat stovky rozhodovacích stromů.



**Obrázek 3.8:** Random Forest, převzato z [28].

Množstvím použitých stromů je dána jeho výpočetní složitost a proto jeho natrénování zabírá více času než jiné srovnatelné algoritmy [10].

### ■ 3.3.4 Naimplementované klasifikátory na základě nezávislých komponent

Tento oddíl je věnován klasifikátorům, jenž jsou natrénovány a otestovány na základě příznaků nezávislých komponent EEG signálu. Tyto klasifikátory byly určeny na základě explorační enelýzy. V úvodu jsou popsány dva algoritmy používané v těchto klasifikátorech. Po něm následuje výčet a popis 5 klasifikátorů, které do tohoto oddílu spadají.

V MATLAB toolboxu EEGLab byla pro jednotlivé epizody provedena ICA analýza 10 vteřinových úseků EEG signálu každé epizody zahrnujících 5 vteřin před epizodou a 5 vteřin po anotovaném počátku epizody. Pro každou epizodu bylo uloženo 15 nezávislých komponent. Pro každou z těchto komponent bylo vypočítáno AUC pro rozptyl. Tento výpočet je zobrazen algoritmem 1, který pracuje následovně: Rozptyl je nacházen pro půl vteřinová okna. Protože je vzorkovací frekvence EEG signálu 250Hz, na půl vteřinová okna, připadá 125 hodnot. AUC pro rozptyl je pro každou komponentu vypočítáno pomocí funkce `perfcurve` [29], kde jako třída 2 byly označeny rozptyly vypočítané před epizodou a jako třída 1 rozptyly po anotovaném počátku epizody. Třída 1 byla označena jako pozitivní.

---

**Algorithm 1** `getAUC(comps)`


---

```

1: Inputs: comps ▷ 15 komponent epizody
2: labels  $\leftarrow$  [2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1]
3: AUC  $\leftarrow$  zeros(15, 1)
4: for all  $i \in 1 : \text{length}(\text{comps})$  do
5:   comp  $\leftarrow$  comps(i)
6:   scores  $\leftarrow$  zeros(20, 1)
7:   g  $\leftarrow$  [0, 125, 250, 375, ..., 2375]
8:   l  $\leftarrow$  0
9:   for all  $j \in g$  do
10:    l  $\leftarrow$  l + 1;
11:    scores(l)  $\leftarrow$  var(comp(g : g + 124)) ▷ Rozptyl
12:    AUC(i)  $\leftarrow$  perfcurve(labels, scores, 1)
13: return AUC

```

---

Dále byl pro každou epizodu vypočítán počet komponent, jejichž AUC pro rozptyl bylo větší než 0.95. Tyto komponenty s AUC pro rozptyl větším než 0.95 budou dále označovány jako silné komponenty. Pseudokód 2 zobrazuje nalezení pole indexů silných komponent epizody. Počet silných komponent v epizodě je pak délka pole indexů, tedy `length(strongIdx(comps))`.

---

**Algorithm 2** `strongIdx(comps)`


---

```

1: AUC  $\leftarrow$  getAUC(comps) ▷ getAUC 1
2: idx  $\leftarrow$   $\emptyset$ 
3: for all  $i \in 1 : \text{length}(\text{comps})$  do
4:   if AUC(i) > 0.95 then
5:     idx(end + 1)  $\leftarrow$  i
return idx

```

---

### ■ Klasifikátor prahování AUC pro rozptyl silných nezávislých komponent (StrongComp)

Prvním naimplementovaným klasifikátorem byl jednoduchý klasifikátor StrongComp 3. Ten mezi třídami volí pouze na základě počtu silných komponent následovně: pokud epizoda nemá žádnou silnou komponentu, je označena jako fyziologická, v jiném případě je klasifikována jako epizoda parasomní.

---

#### Algorithm 3 StrongComp(*comps*)

---

```

1: Input: comps ▷ 15 komponent epizody
2: idxs ← strongIdx(comps) ▷ strongIdx 2
3: N ← length(idxs)
4: if N>0 then
5:   return Epizoda je parasomní
6: else
7:   return Epizoda je fyziologická

```

---

### ■ Klasifikátor spojující prahování AUC pro rozptyl silných nezávislých komponent a prahování RMS silných komponent (StrongCompRMS)

StrongCompRMS je klasifikátor, u kterého je epizoda označena jako fyziologická, pokud neobsahuje žádné silné komponenty. Narozdíl od předchozího klasifikátoru jsou v tomto klasifikátoru ostatní epizody (s alespoň jednou silnou komponentou) klasifikovány právě podle jejich silných komponent, konkrétně podle RMS těchto silných komponent ve zvoleném frekvenčním pásmu gamma (30-45Hz). Tento příznak byl pro tuto klasifikaci zvolen, protože při provedení perfcurve pro všechny příznaky ze silných komponent měl AUC - a tedy rozlišovací schopnost mezi třídami - nejvyšší. Tyto hodnoty AUC jsou zobrazeny v tabulce 4.2.

Klasifikátor StrongCompRMS funguje následovně: Nejdříve je na trénovací sadě nalezen optimální klasifikační RMS práh *Th\_RMS*. Nalezení tohoto prahu zobrazuje algoritmus 6, jehož výstup - práh *Th\_RMS* - je použit na testovacích epizodách.

Vyhodnocení jedné testovací epizody je zobrazeno algoritmem 7. V něm můžeme vidět, že epizoda je označena jako fyziologická, pokud neobsahuje žádné silné komponenty. Dále je vypočítán počet silných komponent, jejichž energie gamma pásma je větší než práh *Th\_RMS*. Dále, pokud má většina

silných komponent epizody energii gamma pásma větší než práh  $Th\_RMS$ , je klasifikována jako parasomní, v opačném případě jako fyziologická.

---

**Algorithm 4**  $getRMS(ep, comp)$ 


---

$RMS \leftarrow$  energie pásma 30-45Hz komponenty  $comp$ , epizody  $ep$   
**return** RMS

---



---

**Algorithm 5**  $getLabel(epizoda)$ 


---

1: **if**  $epizoda$  je parasomní epizoda **then**  
 2:      $label \leftarrow 1$   
 3: **if**  $epizoda$  je fyziologická epizoda **then**  
 4:      $label \leftarrow 2$   
**return**  $label$

---



---

**Algorithm 6**  $StrongCompRMS\_trn(trnSet)$ 


---

1: **Inputs:**  $trnSet$   $\triangleright$  trénovací sada obsahující pro každou ze 120 epizod 15 komponent  
 2:  $scores \leftarrow \emptyset$   
 3:  $labels \leftarrow \emptyset$   
 4: **for all**  $i \in 1 : length(trnSet)$  **do**  
 5:      $AUC \leftarrow getAUC(trnSet(i))$   $\triangleright$  getAUC 1  
 6:     **for all**  $j \in 1 : 15$  **do**  
 7:         **if**  $AUC(j) > 0.95$  **then**  
 8:              $scores(end + 1) \leftarrow getRMS(i, j)$   $\triangleright$  getRMS 4  
 9:              $labels(end + 1) \leftarrow getLabel(i)$   $\triangleright$  getLabel 5  
 10:  $Th\_RMS \leftarrow getOPT(perfcurve(labels, scores, 1))$   $\triangleright$  Nalezení optimálního klasifikačního RMS prahu  
 11: **return**  $Th\_RMS$

---

**Algorithm 7** StrongCompRMS\_tst( $ep, Th\_RMS$ )

---

```

1: Input: ep, Th_RMS           ▷ epizoda, klasifikační RMS práh
2:  $idx \leftarrow strongIdx(compsEPI)$            ▷ strongIdx 2
3: if length(idx) = 0 then
4:   return Epizoda je fyziologická;
5:  $F \leftarrow 0$ 
6:  $P \leftarrow 0$ 
7: for all  $i \in idx$  do
8:   if getRMS(ep,i) > Th_RMS then           ▷ getRMS 4
9:      $P \leftarrow P + 1$ 
10:  else
11:     $F \leftarrow F + 1$ 
12: if P > F then
13:   return Epizoda je parasomní
14: else
15:   return Epizoda je fyziologická

```

---

■ **Klasifikátor prahování RMS všech nezávislých komponent (CompRMS)**

Dalším naimplementovaným klasifikátorem je CompRMS. Narozdíl od předchozího uvedeného klasifikátoru StrongCompRMS je optimální klasifikační RMS práh nalezen pro energii gamma pásma **všech** komponent, nejen silných. Energie gamma pásma byla jako příznak všech komponent použita na základě vyhodnocení klasifikační síly jednotlivých příznaků v rámci explorační analýzy. Natrénování tohot klasifikátoru zobrazuje Algoritmus 8. Samotnou klasifikaci jedné testovací epizody pak zobrazuje algoritmus 9. Epizoda je zde opět klasifikována jako parasomní, pokud většina jejích komponent je klasifikována jako parasomní, jinak je označena jako fyziologická.

**Algorithm 8** CompRMS\_trn(*trnSet*)

---

```

1: Inputs: trnSet ▷ trénovací sada obsahující pro každou ze 120 epizod 15
   komponent
2: scores ← zeros(120 × 15, 1)
3: labels ← zeros(120 × 15, 1)
4: for all i ∈ 1 : length(trnSet) do
5:   for all j ∈ 1 : 15 do
6:     l ← (i - 1) × 15 + j
7:     scores(l) ← getRMS(i, j) ▷ getRMS 4
8:     labels(l) ← getLabel(i) ▷ getLabel 5
9: Th_RMS ← getOPT(perfcurve(labels, scores, 1)) ▷ Nalezení
   optimálního klasifikačního RMS prahu
10: return Th_RMS

```

---

**Algorithm 9** Comp&RMS\_tst(*ep*, *Th\_RMS*)

---

```

1: Input: ep, Th_RMS ▷ epizoda, klasifikační RMS práh
2: F ← 0
3: P ← 0
4: for all i ∈ 1 : 15 do
5:   if getRMS(ep, i) > Th_RMS then ▷ getRMS 4
6:     P ← P + 1
7:   else
8:     F ← F + 1
9: if P > F then
10:  return Epizoda je parasomní
11: else
12:  return Epizoda je fyziologická

```

---

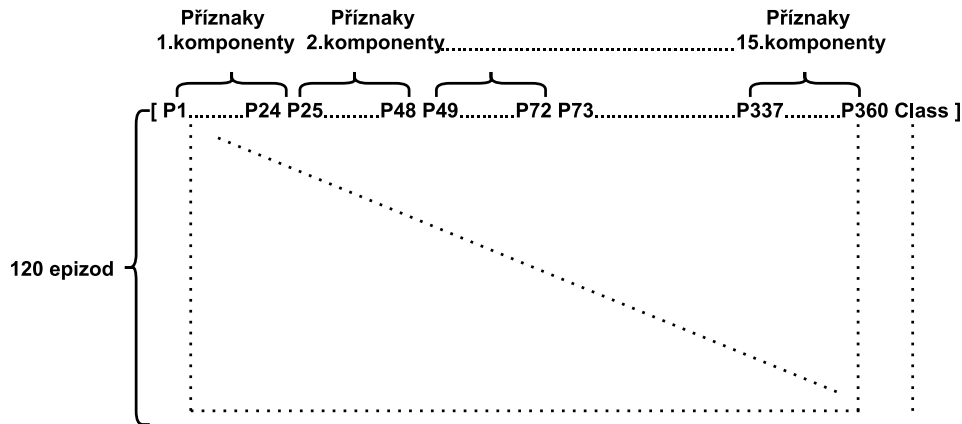
### ■ Klasifikátor Decision Tree použitý na všechny příznaky epizod všech komponent (allEpDT)

Dalším implementovaným klasifikátorem je allEpDT, v němž byl pro klasifikaci použit Decision Tree (DT), tedy rozhodovací strom, jehož princip je blíže popsán v kapitole 3.3.3.

Algoritmus 10 popisuje proces natrénování rozhodovacího stromu pomocí funkce fitctree. Tato funkce vrací binární klasifikační rozhodovací strom založený na vstupních proměnných. Vstupní proměnné jsou u této klasifikace všechny příznaky (360) všech trénovacích epizod (120). Obrázek 3.9 zobrazuje, jak je vstupní proměnná (v algoritmu 10 *trnFeatureTbl*) tvořena. Jedné kom-

ponentě vždy náleží 24 příznaků, které byly popsány v části 3.3.2, komponent je 15 a Class jsou třídy, do kterých epizody patří, tj. parasomní/fyziologická.

V algoritmu 11 je pak binární klasifikační rozhodovací strom využit pro klasifikaci testovací epizody.



**Obrázek 3.9:** Sada příznaků trénovacích epizod použitá k natrénování rozhodovacího stromu pro allEpDT.

---

**Algorithm 10** allEpDT\_trn(*trnFeatureTbl*)

---

- 1: **Inputs:** *trnFeatureTbl* ▷ table trénovací sady obsahující 24 příznaků pro každou z 15 komponent všech 120 trénovacích epizod ( $120 \times 360$ )
  - 2: *labels*  $\leftarrow$  zeros(120, 1)
  - 3: **for all**  $i \in 1 : \text{length}(\text{trnSet})$  **do**
  - 4:     *labels*( $i$ )  $\leftarrow$  getLabel( $i$ ) ▷ getLabel 5
  - 5: *trnFeatureTbl.Class*  $\leftarrow$  *labels*
  - 6: *DT*  $\leftarrow$  fitctree(*trnFeatureTbl*, 'Class')
  - 7: **return** DT
- 

---

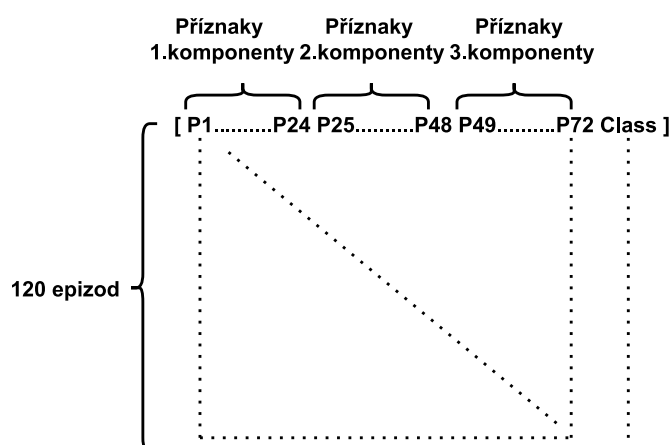
**Algorithm 11** allEpDT\_tst(*tstFeatureTbl*, *DT*)

---

- 1: **Inputs:** *tstFeatureTbl* ▷ testovací epizoda obsahující 24 příznaků pro každou z 15 komponent (celkem 360 příznaků)
  - 2: *C*  $\leftarrow$  predict(*DT*, *tstFeatureTbl*) ▷ C obsahuje hodnotu 1 pro klasifikaci epizody jako parasomní nebo hodnotu 2 pro klasifikaci epizody jako fyziologické
  - 3: **return** C
-

### ■ Klasifikátor Decision Tree použitý na všechny příznaky epizod 3 komponent (EpDT\_3)

Vylepšením předchozího klasifikátoru allEpDT je tento klasifikátor EpDT\_3. Ten pracuje na stejném principu jako allEpDT pouze s tím rozdílem, že vstupními proměnnými jsou příznaky pouze od prvních tří komponent. Toolbox EEGlab totiž nezávislé komponenty vrací seřazené podle toho, kolik dat (artefaktických) obsahují. Tedy čím nižší umístění komponenta v pořadí nezávislých komponent zaujímá, tím více dat obsahuje. Příznaků od tří komponent jedné epizody je 72, celková sada příznaků použitá k natrénování rozhodovacího stromu je tedy 120x72. Tuto sadu společně s třídami epizod zobrazuje obrázek 3.10.



**Obrázek 3.10:** Sada příznaků trénovacích epizod použitá k natrénování rozhodovacího stromu pro EpDT\_3.

### ■ 3.3.5 Klasifikátory na základě záznamu z 2 kanálů EEG

Tento oddíl je věnovaný klasifikátorům, jenž jsou natrénovány a otestovány na základě příznaků EEG signálu z kanálů Cz a Fz. Tyto kanály byly pro klasifikaci parasomnie z EEG signálu vybrány na základě odborné literatury [9]. Příznaky RNG, MaxDiff a RMS, které byly vypočítány pro signály a také pro jednotlivá vybraná frekvenční pásma, jsou blíže popsány v části 3.3.2. Nyní následuje výčet a popis 3 klasifikátorů, které do tohoto oddílu spadají.



### ■ Klasifikátor Decision Tree použitý na příznaky EEG kanálů Cz a Fz (EpDT\_CzFz)

Pro tento klasifikátor bylo použito 21 vybraných příznaků EEG signálu z kanálu Cz a 21 vybraných příznaků z kanálu Fz. Stejně jako u předchozích klasifikátorů založených na použití Decision Tree je nejdříve natrénován binární rozhodovací strom, což je zobrazeno algoritmem 12. Algoritmus 13 pak zobrazuje klasifikaci testovací epizody.

---

#### Algorithm 12 $\text{EpDT\_CzFzTrn}(trnFeatureTbl)$

---

```

1: Inputs: trnFeatureTbl    ▷ table trénovací sady obsahující 42 příznaků
   pro každou epizodu všech 120 trénovacích epizod ( $120 \times 42$ )
2: labels  $\leftarrow$  zeros(120, 1)
3: for all  $i \in 1 : \text{length}(trnSet)$  do
4:   labels( $i$ )  $\leftarrow$  getLabel( $i$ )                                ▷ getLabel 5
5: trnFeatureTbl.Class  $\leftarrow$  labels
6: DT  $\leftarrow$  fitctree(trnFeatureTbl, 'Class')
7: return DT

```

---



---

#### Algorithm 13 $\text{EpDT\_CzFzTst}(tstFeatureTbl, DT)$

---

```

1: Inputs: tstFeatureTbl    ▷ testovací epizoda obsahující 42 příznaků
2: C  $\leftarrow$  predict(DT, tstFeatureTbl)    ▷ C obsahuje hodnotu 1 pro
   klasifikaci epizody jako parasomní nebo hodnotu 2 pro klasifikaci epizody
   jako fyziologické
3: return C

```

---

### ■ Klasifikátor Random Forest použitý na příznaky z EEG kanálů Cz a Fz (RF)

RF neboli Random Forest je učicí metoda, jenž byla v tomto klasifikátoru použita a která je blíže popsána v části 3.3.3. Do Random Forest modelu byly využity stejné příznaky jako u klasifikátoru EpDT\_CzFz, tedy 42 příznaků (21 od Cz, 21 od Fz kanálu EEG).

### ■ Klasifikátor Random Forest použitý na příznaky z EEG kanálů Cz a Fz vylepšený optimalizací hyperparametrů (RFtune)

Posledním uvedeným klasifikátorem je klasifikátor RFtune, což je předchozí klasifikátor RF vylepšen pomocí Quantile Error a Bayesian Optimization [32], díky nimž byl model natrénován s pomocí optimalizovaných hyperparametrů. Příkladem takového hyperparametru je např. složitost (hloubka) rozhodovacího stromu v Random Forest.

V rámci práce byly implementovány i další kombinace těchto klasifikátorů. Avšak nevykazovaly zlepšení, proto v bakalářské práci nejsou popsány.

# Kapitola 4

## Experimentální výsledky

V této kapitole je nejdříve popsán dataset (4.1) obsahující signály EEG, z nichž byly na základě explorační analýzy popsané v části 4.2 vybrány příznaky. Poslední část této kapitoly je věnována vyhodnocení použitých klasifikátorů (4.3).

### 4.1 Dataset

Naměřená data jsou poskytnuta Národním ústavem duševního zdraví (NÚDZ), kde byly EEG signály nahrány pomocí polysomnografie, díky které je možné souběžné nahrávání více biologických signálů najednou, jak bylo popsáno v podkapitole 2.2.

Z vyšetření 32 pacientů ve složení 20 žen a 12 mužů ve věkovém rozmezí 19 až 44 let s diagnostikovanou parasomnií je k dispozici celkem 35 celonočních záznamů. Celkově je to 150 epizod, z toho 49 parasomních. Tento dataset je tedy nevyvážený, což může vést k vyšší míře nesprávné klasifikace. Všechny epizody jsou somnologem (MUDr. Jitkou Buškovou, Ph.D.) doplněny jak o informaci, zda je epizoda fyziologická či parasomní, tak o vteřinovou anotaci začátku epizody, tedy o počátek fyziologického či parasomního probuzení.

Záznamy EEG signálů obsahují informace z 22 EEG kanálů.

Epizody byly rozděleny do 5 skupin (folds) po 30 epizodách. Do každé skupiny byly zařazeny všechny epizody od 6 popř. 7 pacientů. Obrázek 4.1 zobrazuje počet fyziologických a počet parasomních epizod v jednotlivých skupinách. Pro evaluaci klasifikátorů pak byly vždy 4 skupiny složeny do jedné trénovací sady a zbylá skupina byla použita jako sada testovací, tj. 120 epizod vždy tvoří trénovací a 30 epizod testovací sadu. Po každé iteraci byly uloženy výsledky z obou sad. Ve výsledné tabulce 4.3 je pak uváděn jejich průměr.

5 fold - krosvalidace:

	Fyziologické epizody	Parasomní epizody	
1)	23	7	} Trénovací sada
2)	18	12	
3)	19	11	
4)	20	10	
5)	21	9	} Testovací sada

Provedeno 5 iterací, ve kterých je jako testovací sada použit vždy jiný fold

**Obrázek 4.1:** Krosvalidace přes všechny skupiny.

Několik pacientů (11) nemá z celonočního nahrávání anotovanou žádnou parasomní epizodu. V tabulce 4.1 můžeme vidět, že stejně tak jako nejsou skupiny vyrovnané, co se týče počtu parasomních a fyziologických epizod, nejsou vyrovnané ani v počtu pacientů, kteří alespoň jednu parasomní epizodu během zaznamenávané noci měli.

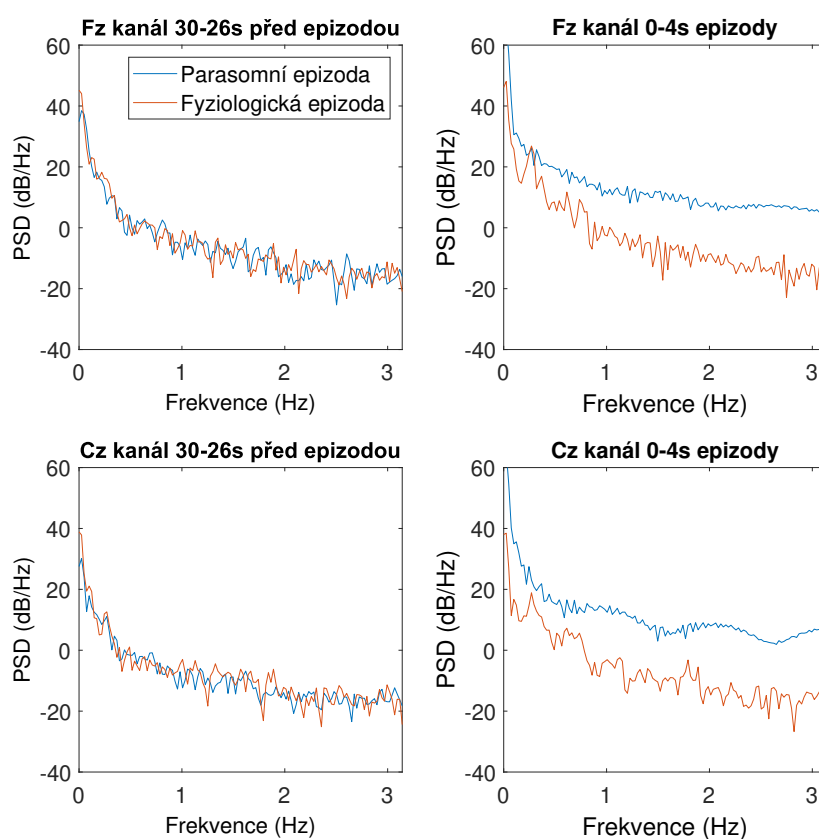
Počet pacientů ve skupině	7	6	6	7	6
Počet pacientů s alespoň 1 parasomní epizodou	5	4	5	3	4

**Tabulka 4.1:** Počet pacientů v jednotlivých skupinách a počet pacientů s alespoň 1 anotovanou parasomní epizodou.

## 4.2 Explorační analýza

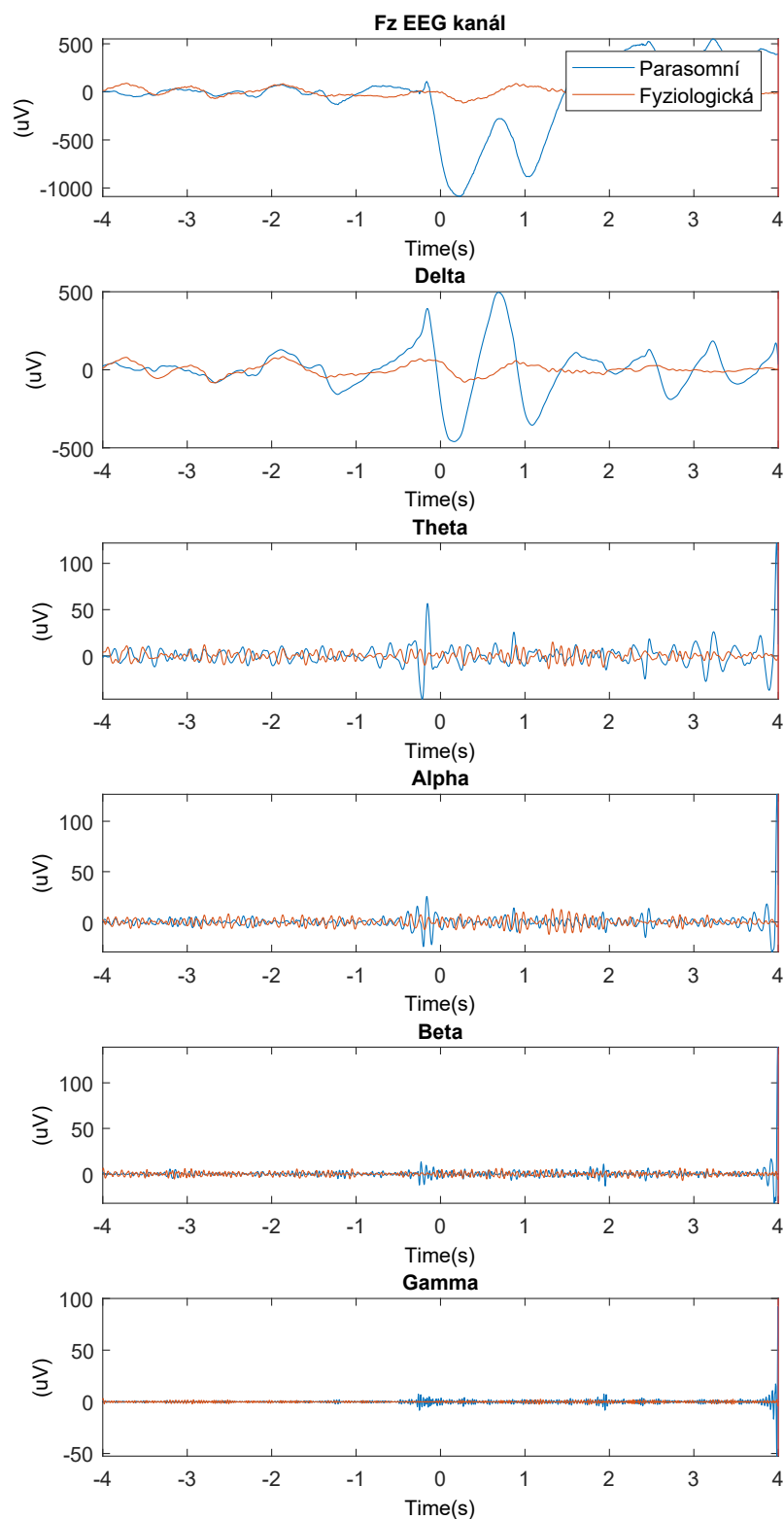
Devadesátivteřinové záznamy s parasomními (49) nebo fyziologickými (101) epizodami v dostupném datasetu obsahují pro každou epizodu hodnoty z 22 ka-

nálů EEG s počátkem epizody anotovaným na 30. vteřinu včetně informace, o jakou epizodu jde. Pro jednotlivé kanály byly vykresleny spektrogramy a výkonové spektrální hustoty (PSD) v porovnání čtyřsekundového úseku třicet sekund před začátkem epizody a čtyřsekundového úseku od začátku epizody. Obrázek 4.2 je ukázkou PSD pro EEG kanál Fz a Cz jedné parasomní a jedné fyziologické epizody a můžeme vidět, že pro tyto epizody byla PSD pro oba kanály u parasomní epizody vyšší než u fyziologické. Na základě této analýzy byly později použity právě čtyři sekundy od počátku epizody, u kterých parasomní epizody častěji vykazovaly vyšší výkonovou hustotu než epizody fyziologické.



**Obrázek 4.2:** PSD EEG kanálů Fz a Cz.

Dále byla pro EEG kanály vykreslena frekvenční pásma. Obrázek 4.3 zobrazuje signál a frekvenční pásma EEG kanálu Fz jedné parasomní a jedné fyziologické epizody. Příznaky pak byly nalezeny jak v originálním signálu z EEG kanálu, tak z jeho pásem.



**Obrázek 4.3:** Signál z EEG kanálu Fz a jeho frekvenční pásma na příkladu začátku jedné fyziologické a jedné parasomní epizody. Anotovaný začátek epizody se nachází v čase 0 s.

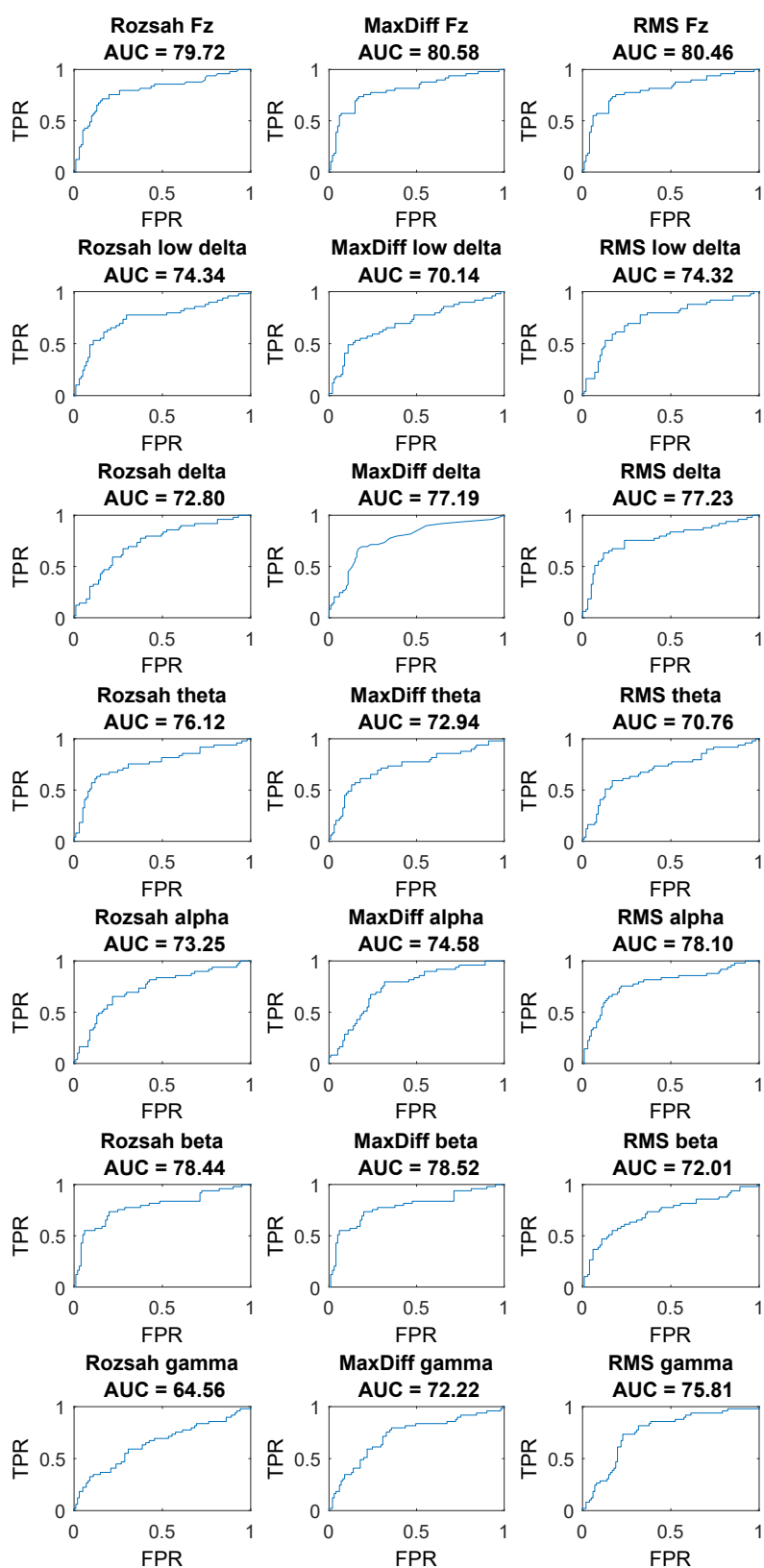
Pro vybrané příznaky (rozsah, maximální diferenci a energii) byla vypočítána hodnota AUC pro různá frekvenční pásma. Například na obrázku 4.4 jsou vykresleny ROC křivky pro všech 21 příznaků kanálu Fz. Nejvyšší AUC je zde 80,58 u maximální diference kanálu Fz. Tento příznak byl spolu s maximální diferencí kanálu Cz a rozsahem beta kanálu Cz a Fz vybírán v klasifikátoru EpDT\_CzFz (3.3.5) rozhodovacím stromem jako jeho práh.

Začátky epizod jsou náhlé a souvisejí s artefakty, které se projeví v některých nezávislých komponentách. Na základě tohoto poznatku byly v EEGLAB v signálech EEG pomocí ICA nalezeny nezávislé komponenty, protože by oproti použití pouze samotného EEG mohly přinést nové informace. A proto pro ně byla také provedena explorační analýza zahrnující vykreslení spektrogramů a PSD jednotlivých komponent, jejich frekvenční pásma (příklad je uveden na obrázku 4.5) a následné extrahování příznaků. Zvlášť byly zkoumány také silné komponenty (s AUC pro rozptyl  $> 0,95$ ).

Tabulka 4.2 zobrazuje hodnoty AUC pro příznaky minimum, maximum, maximální diferenci a energii pro všechny silné komponenty všech dostupných epizod. Nejvyššího AUC jak u silných komponent, tak u všech komponent dosahovala energie gamma pásma.

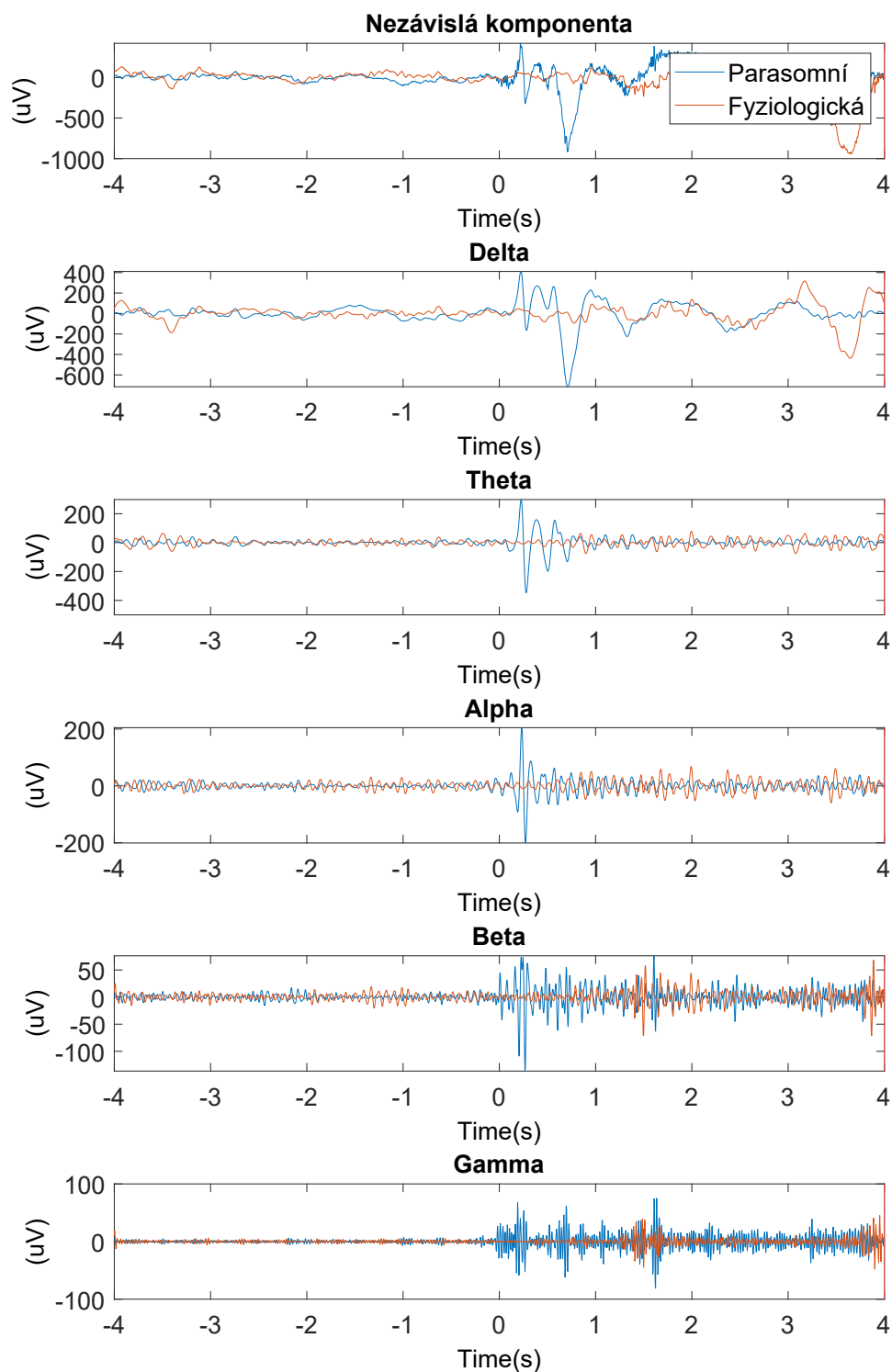
	AUC			
	Min	Max	MaxDiff	RMS
Originální komponenta	0,7517	0,7548	0,7797	0,7522
Delta	0,7549	0,7555	0,7806	0,7580
Theta	0,7868	0,7750	0,7708	0,7861
Alpha	0,7639	0,7541	0,7699	0,7693
Beta	0,7812	0,7747	0,7829	0,7886
Gamma	0,7951	0,7936	0,7899	0,8110

**Tabulka 4.2:** Rozlišovací schopnosti příznaků pro různá pásma.



Obrázek 4.4: ROC křivky pro příznaky Fz kanálu.





**Obrázek 4.5:** Nezávislá komponenta a její frekvenční pásma na příkladu začátku jedné fyziologické a jedné parasomní komponenty. Anotovaný začátek epizody se nachází v čase 0s.

### 4.3 Diskuze výsledků

V rámci této práce bylo vytvořeno 8 klasifikátorů pro rozlišení parasomních a fyziologických epizod a otestováno na zadaném datasetu.

Seznam použitých klasifikátorů:

- StrongComp – klasifikátor prahování AUC pro rozptyl silných nezávislých komponent,
- StrongCompRMS – klasifikátor spojující prahování AUC pro rozptyl silných nezávislých komponent a prahování RMS silných komponent,
- CompRMS – klasifikátor prahování RMS všech nezávislých komponent,
- DTep\_allcomp – klasifikátor Decision Tree použitý na všechny příznaky epizod všech komponent,
- DTep\_3comp – klasifikátor Decision Tree použitý na všechny příznaky epizod 3 komponent,
- DTep\_CzFz – klasifikátor Decision Tree použitý na příznaky EEG kanálů Cz a Fz,
- RF – klasifikátor Random Forest použitý na příznaky z EEG kanálů Cz a Fz ,
- RFtune – klasifikátor Random Forest použitý na příznaky z EEG kanálů Cz a Fz vylepšený optimalizací hyperparametrů.

Výsledné hodnoty všech klasifikátorů jsou popsány v tabulce 4.3. Můžeme v ní vidět, že první klasifikátor založen na klasifikaci z příznaků nezávislých komponent StrongComp mezi třídami rozlišoval s přesností 63% a senzitivitou 70% na testovacích datech. Na něj navázal klasifikátor StrongCompRMS, jehož klasifikaci se přesnost na testovací sadě oproti klasifikátoru StrongComp významně zvýšila a to na 75%, naopak senzitivita poklesla na 63%. U klasifikátoru CompRMS senzitivita ještě poklesla, a to na 54%. DTep\_allcomp rozlišil epizody testovací sady s přesností 73% a senzitivitou 60%. Omezením použitých příznaků z 15 komponent na příznaky pouze z prvních 3 komponent, které obsahují nejvíce artefaktických dat, vznikl klasifikátor DTep\_3comp, který dosáhl na testovací sadě přesnosti 83% se senzitivitou 78%, což je v rámci této práce nejlepší výsledek.

První klasifikátor založen na klasifikaci z příznaků EEG kanálu Cz a Fz, tedy klasifikátor DTap\_CzFz, rozlišil epizody testovací sady s přesností 71% a senzitivitou 67%. Jeho využitím v Random Forest klasifikátoru se přesnost klasifikace testovací sady zvýšila na 78%. Díky vyladění použitých hyperparametrů v Random Forest bylo dosaženo přesnosti klasifikace na testovací sadě 81% se senzitivitou 73%.

Nejlepší klasifikátor pracující s příznaky extrahovanými z nezávislých komponent byl tedy klasifikátor DTep\_3comp a nejlepší klasifikátor pracující s příznaky extrahovanými z EEG kanálů Cz a Fz byl RFtune. Oba tyto klasifikátory dosahují velmi podobných výsledků a také bohužel senzitivita je u obou výrazně nižší než specificita. Můžeme tedy usuzovat, že použití nezávislých komponent nepřináší oproti použití signálu EEG výhodu.

Dosažená přesnost byla velmi podobná přesnosti, jaké dosáhla Mariana Castillo [10] při klasifikaci parasomních epizod z EMG dat (82%) a při klasifikaci, ve které využila jak data z EMG a ECG, tak data z 3D-aktigrafu (83%). Klasifikátory ze signálu EEG v této práci ale vykazují vyšší senzitivitu (pro porovnání senzitivita klasifikátoru z EMG dat byla pouze 49%).

Klasifikátor	Trénovací sada			Testovací sada		
	Přesnost	Senzitivita	Specificita	Přesnost	Senzitivita	Specificita
StrongComp	0,6333	0,7136	0,5937	0,6333	0,7032	0,5877
StrongCompRMS	0,7583	0,6314	0,8189	0,7533	0,6264	0,8066
CompRMS	0,7767	0,5708	0,8763	0,7600	0,5381	0,8603
DTep_allcomp	0,8350	0,7910	0,8564	0,7267	0,5979	0,7896
DTep_3comp	0,8450	0,7756	0,8783	0,8267	0,7757	0,8513
DTep_CzFz	0,8267	0,7619	0,8588	0,7133	0,6724	0,7618
RF	0,8383	0,7352	0,8885	0,7800	0,6724	0,8525
RFtune	0,8700	0,7565	0,9257	0,8067	0,7273	0,8621

**Tabulka 4.3:** Výsledné hodnoty všech klasifikátorů.

Tyto klasifikátory nebyly otestovány na celonočních záznamech. Můžeme předpokládat, že stejně jako v diplomové práci [10] by tato detekce nedosahovala takových výsledků jako u detekce, při které jsou počátky epizod známé.

Abychom si mohli být jistější, jak přesně klasifikátory fungují, bylo by dobré mít anotace epizod od více než jednoho spánkového experta, protože se stává, že ani odborníci se při označení epizody neshodnou. Také použití většího datasetu s vyšším zastoupením pozitivní třídy, než bylo u použitého datasetu, by výsledky mohlo pozitivně ovlivnit. U tohoto datasetu byl poměr pozitivní - negativní třída 1:2.

Kromě provedení detekce parasomních epizod v celonočním záznamu by

v případě další práce bylo dobré zaměřit se na analýzu EEG signálu v několika málo vteřinách před epizodou, konkrétně na aktivitu pomalých vln a vysokofrekvenční aktivitu v EEG signálech s odstraněnými artefakty. Také by bylo dobré zkusit propojit již vytvořené klasifikátory (např. klasifikátor z EMG a EEG dat).

## Kapitola 5

### Závěr práce

Hlavními cíli této práce bylo nastudovat problematiku polysomnografického snímání a parasomnie v odborné literatuře a na jejím základě vybrat vhodné příznaky pro klasifikaci parasomních epizod. Dalším úkolem bylo vytvořit klasifikátor pro rozlišení parasomních a fyziologických epizod a jeho evaluace na signálech EEG nahraných pomocí polysomnografie v Národním ústavu duševního zdraví.

Po provedení rešerše byly na základě explorační analýzy zvoleny a vypočítány příznaky ze 4 sekund od počátku epizody jak z EEG kanálů, tak z nezávislých komponent EEG signálů nalezených pomocí analýzy nezávislých komponent ICA v programu EEGLAB. Nezávislé komponenty byly mezi analyzované signály zařazeny, protože jsme předpokládali, že identifikace komponent, souvisejících se svalovou a další artefaktickou aktivitou, by mohly být pro rozlišení parasomních epizod přínosem. Mezi použité příznaky byl zařazen rozsah, maximální diference, energie signálu a normalizovaná energie signálu, a to jak pro původní signál, tak pro jeho frekvenční pásma.

Příznaky byly dále použity ve vytvořených klasifikátorech, v nichž bylo využito metod prahování, rozhodovacích stromů nebo Random Forest. Pro klasifikaci ze signálu EEG byly příznaky na základě odborné literatury vybrány z EEG kanálu Cz a Fz. Pro klasifikaci z nezávislých komponent byly vybírány příznaky z prvních 15 nebo z prvních 3 nezávislých komponent EEG signálu. (Komponenty jsou v EEGLAB řazeny od největšího množství artefaktických dat, jež obsahují, po nejmenší.)

U klasifikátorů, které používaly příznaky extrahované z EEG kanálu Cz

a Fz, bylo na testovací sadě dosaženo nejvyšší přesnosti 81% se senzitivitou 73%. U klasifikátorů, které používaly příznaky extrahované z nezávislých komponent EEG signálů, byla nejvyšší dosažená přesnost 83% se senzitivitou 78%.

Použití nezávislých komponent v klasifikaci v tomto případě přineslo oproti signálům EEG lepší výsledky bohužel jen s nevelkým rozdílem. Do budoucna by bylo dobré se ještě více zaměřit na projevy parasomnie v signálech EEG. Zejména by bylo zajímavé zaměřit se na detekci charakteristické EEG aktivity, předcházející dle somnologické literatury parasomním epizodám, či kombinacím s automatickou detekcí spánkových stádií z EEG.



## Příloha A

### Literatura

- [1] Fit binary decision tree for multiclass classification - matlab fitctree. URL: [https://www.mathworks.com/help/stats/fitctree.html?searchHighlight=fitctree&s\\_tid=srchtitle\\_fitctree\\_1#bujbrib-6](https://www.mathworks.com/help/stats/fitctree.html?searchHighlight=fitctree&s_tid=srchtitle_fitctree_1#bujbrib-6).
- [2] Ondřej Konopka 1. Analýza nezávislých komponent. URL: <http://amber.feld.cvut.cz/bio/konopka/>.
- [3] 2012\_10\_15\_statsoft\_jaky\_je\_mezi\_nimi\_rozdil. URL: <http://www.statsoft.cz/>.
- [4] Rodrigo V. Andreão, Bernadette Dorizzi, and Jérôme Boudy. Ecg signal analysis through hidden markov models. *IEEE Transactions on Biomedical Engineering*, 53:1541–1549, 8 2006. doi:10.1109/TBME.2006.877103.
- [5] Alakh Sudhir Biniwale, Mina Guirguis, and Byron Gao. Analysis of human polysomnography (psg) for automatic sleep event detection using hidden markov model. Technical report, Texas State University, San Marcos, Texas, 2016.
- [6] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [7] L Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001. doi:10.1023/A:1010950718922.
- [8] MUDr. Jitka Bušková. Nrem parasomnie. *Neurologie pro Praxi*, 2014. 15(4).

- [9] Milena Camaioni, Serena Scarpelli, Maurizio Gorgoni, Valentina Alfonsi, and Luigi De Gennaro. Eeg patterns prior to motor activations of parasomnias: A systematic review. *Nature and Science of Sleep*, 13:713–728, 2021. doi:10.2147/NSS.S306614.
- [10] Mariana Castillo. Klasifikace parasomních epizod z 3-osých aktigrafických záznamů, 2021.
- [11] Antonino Crivello, Paolo Barsocchi, Michele Girolami, and Filippo Palumbo. The meaning of sleep quality: A survey of available technologies. *IEEE Access*, 7:167374–167390, 2019. doi:10.1109/ACCESS.2019.2953835.
- [12] Definition of ica. URL: [http://cis.legacy.ics.tkk.fi/aapo/papers/IJCNN99\\_tutorialweb/node3.html#xAs](http://cis.legacy.ics.tkk.fi/aapo/papers/IJCNN99_tutorialweb/node3.html#xAs).
- [13] Panagis Drakatos, Lucy Marples, Rexford Muza, Sean Higgins, Alexander Nesbitt, Eptehal M. Dongol, Raluca Macavei, Valentina Gnoni, Laura Perez Carbonell, Iain Duncan, Adam Birdseye, Sakina Dastagir, Ivana Rosenzweig, David O’Regan, Adrian J. Williams, Guy D. Leschziner, and Brian D. Kent. Video polysomnographic findings in non-rapid eye movement parasomnia. *Journal of Sleep Research*, 28, 4 2019. doi:10.1111/jsr.12772.
- [14] Eeglab. URL: <https://scn.ucsd.edu/eeglab/index.php>.
- [15] Dmitry Efimov. *Independent Component Analysis*, pages 724–728. Springer New York, New York, NY, 2014. doi:10.1007/978-1-4614-6170-8\_147.
- [16] Hugo Espiritu and Vangelis Metsis. Automated detection of sleep disorder-related events from polysomnographic data. *Proceedings - 2015 IEEE International Conference on Healthcare Informatics, ICHI 2015*, pages 562–569, 12 2015. doi:10.1109/ICHI.2015.105.
- [17] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27:861–874, 6 2006. doi:10.1016/j.patrec.2005.10.010.
- [18] Chiara Fois, Mary Anne S. Wright, Gian Pietro Sechi, Matthew C. Walker, and Sofia H. Eriksson. The utility of polysomnography for the diagnosis of nrem parasomnias: an observational study over 4 years of clinical practice. *Journal of Neurology*, 262:385–393, 2 2015. doi:10.1007/s00415-014-7578-2.
- [19] Github - scn/iclabel: Automatic eeg ic classification plugin for eeglab. URL: <https://github.com/scn/ICLabel>.
- [20] Vanessa Ibáñez, Josep Silva, and Omar Cauli. A survey on sleep assessment methods. *PeerJ*, 2018, 2018. doi:10.7717/peerj.4849.



- [21] Ben H. Jansen and Wei Kang Cheng. Classification of sleep patterns by means of markov modeling and correspondence analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9:707–710, 1987. doi:10.1109/TPAMI.1987.4767968.
- [22] Xiao Jiang, Gui Bin Bian, and Zean Tian. Removal of artifacts from eeg signals: A review. *Sensors (Switzerland)*, 19, 3 2019. doi:10.3390/s19050987.
- [23] Sharon Keenan and Max Hirshkowitz. Chapter 141-monitoring and staging human sleep. Technical report, The University of Arizona, 2011.
- [24] Ahsan H. Khandoker, Marimuthu Palaniswami, and Chanda K. Karmakar. Support vector machines for automated recognition of obstructive sleep apnea syndrome from eeg recordings. *IEEE Transactions on Information Technology in Biomedicine*, 13:37–48, 2009. doi:10.1109/TITB.2008.2004495.
- [25] Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 3 2012. doi:10.1093/acprof:oso/9780195058239.001.0001.
- [26] MUDr. Miroslav Moráň. Parasomnie v nrem spánku. *Neurologie pro Praxi*, 2002. 2002 (3), 3.
- [27] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 9 2019. doi:10.1016/j.neuroimage.2019.05.026.
- [28] Random forest classifier and its hyperparameters | by ankit chauhan | analytics vidhya | medium. URL: <https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6>.
- [29] Receiver operating characteristic (roc) curve or other performance curve for classifier output - matlab perfcurve. URL: <https://www.mathworks.com/help/stats/perfcurve.html>.
- [30] Claude Sammut and Geoffrey I. Webb, editors. *Bootstrap Sampling*, pages 137–137. Springer US, Boston, MA, 2010. doi:10.1007/978-0-387-30164-8\_886.
- [31] Claude Sammut and Geoffrey I. Webb, editors. *Random Forests*, pages 828–828. Springer US, Boston, MA, 2010. doi:10.1007/978-0-387-30164-8\_695.
- [32] Tune random forest using quantile error and bayesian optimization - matlab & simulink. URL: <https://www.mathworks.com/help/stats/tune-random-forest-using-quantile-error-and-bayesian-optimization.html>.

- [33] Minsoo Yeo, Hoonsuk Byun, Jiyeon Lee, Jungick Byun, Hak Young Rhee, Wonchul Shin, and Heenam Yoon. Respiratory event detection during sleep using electrocardiogram and respiratory related signals: Using polysomnogram and patch-type wearable device data. *IEEE Journal of Biomedical and Health Informatics*, 26:550–560, 2 2022. doi: 10.1109/JBHI.2021.3098312.