



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

IDENTIFIKACE SPÁNKOVÝCH PORUCH Z DAT AKTIGRAFIE A SPÁNKOVÝCH DENÍKŮ

IDENTIFICATION OF SLEEP DISORDERS BASED ON ACTIGRAPHY DATA AND SLEEP DIARIES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Miroslav Molík

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Marek Mikulec

BRNO 2022

Diplomová práce

magisterský navazující studijní program **Telekomunikační a informační technika**

Ústav telekomunikací

Student: Bc. Miroslav Molík

ID: 197775

Ročník: 2

Akademický rok: 2021/22

NÁZEV TÉMATU:

Identifikace spánkových poruch z dat aktigrafie a spánkových deníků

POKYNY PRO VYPRACOVÁNÍ:

Poruchy spánku mohou být častým počátečním projevem neurodegenerativních onemocnění. Pokud jsou tato onemocnění identifikována v prvních fázích, může být jejich léčba daleko účinnější. Cílem práce je z dostupných dat aktigrafie, spánkových deníků a z dat polysomnografie identifikovat spánkové poruchy. V rámci diplomové práce bude proveden návrh identifikace spánkových poruch pomocí srovnání s dostupnými normami, zároveň srovnání pomocí metod strojového učení, či dalšími vhodnými metodami zpracování časových řad. Výstup bude validován s výsledky neurologických a neuropsychologických vyšetření. Pokud budou signatury spánkových poruch spolehlivě identifikovatelné, bude výsledkem práce metodologie, která umožní na základě dat aktigrafie a spánkových deníků identifikovat spánkové poruchy a umožní včasnou léčbu případných neurodegenerativních onemocnění.

DOPORUČENÁ LITERATURA:

- [1] SMITH, M. T., et al., Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment, *Journal of Clinical Sleep Medicine*, vol. 14, no. 07, pp. 1209–1230, Jul. 2018, doi:10.5664/jcsm.7228.
- [2] FARRAHI, V., et al., Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches, *Gait & Posture*, vol. 68, pp. 285–299, Feb. 2019, doi:10.1016/j.gaitpost.2018.12.003.

Termín zadání: 7.2.2022

Termín odevzdání: 24.5.2022

Vedoucí práce: Ing. Marek Mikulec

prof. Ing. Jiří Mišurec, CSc.
předseda rady studijního programu

UPOZORNĚNÍ:

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

ABSTRAKT

Tato diplomová práce se zabývá predikcí Parkinsonovy choroby pomocí spánkových parametrů získaných z aktigrafie a spánkových deníků. Cílem je navrhnout model strojového učení, který bude schopen poznat pacienty trpící Parkinsonovou chorobou. K trénování byla použita datová množina dodaná Fakultní nemocnicí u sv. Anny, která byla za dosažením lepších výsledků různě upravována. Následně byly tyto úpravy podle výsledků modelů zhodnoceny a na jejich bázi byly vybrány dva modely, které dosahují testovací přesnosti 85 a 82 %).

KLÍČOVÁ SLOVA

Spánek, spánkové poruchy, neurodegenerativní onemocnění, aktigrafie, spánkové deníky, parkinsonova choroba, predikce parkinsona

ABSTRACT

This master's thesis deals with prediction of Parkinson's disease using sleep parameters from actigraphy and sleep diaries. The goal is to design a machine learning approach, which will be able to recognize patients suffering from Parkinson's disease. For training, dataset supplied by St. Anne's University Hospital Brno was used which was variously modified for achieving best possible results. These adjustments were evaluated according to the results of the trained models and based on these results, two models (achieving test accuracies of 85 and 82 %) were selected.

KEYWORDS

Sleep, sleep disorder, neurodegenerative disease, actigraphy, sleep diary, parkinson's disease, parkinson's prediction

MOLÍK, Miroslav. *Identifikace spánkových poruch z dat aktigrafie a spánkových deníků*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2022, 57 s. Diplomová práce. Vedoucí práce: Ing. Marek Mikulec

Prohlášení autora o původnosti díla

Jméno a příjmení autora: Bc. Miroslav Molík
VUT ID autora: 197775
Typ práce: Diplomová práce
Akademický rok: 2021/22
Téma závěrečné práce: Identifikace spánkových poruch z dat aktigrafie a spánkových deníků

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

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

Brno

.....

podpis autora*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové panu Ing. Marku Mikulcovi za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	11
1 Spánek a jeho parametry	12
1.1 Fáze spánku	12
1.2 Spánkové parametry a jejich měření	12
1.2.1 Spánkové deníky	13
1.2.2 Polysomnografie	14
1.2.3 Aktigrafie	14
1.3 Poruchy spánku a neurodegenerativní nemoci s nimi spojené	16
1.3.1 Insomnie	16
1.3.2 Nykturie	16
1.3.3 Syndrom neklidných nohou (RLS)	16
1.3.4 Periodické pohyby končetin (PLM)	17
1.3.5 Porucha chování v REM spánku	17
1.3.6 Alzheimerova choroba	18
1.3.7 Parkinsonova choroba	18
1.3.8 Mírná kognitivní porucha	18
2 Strojové učení	19
2.1 Proces strojového učení	19
2.2 Souhrn použitých metod strojového učení na data akcelerometrů	20
2.3 Neuronové sítě	20
2.4 Metoda podpurných vektorů	21
3 Návrh řešení	23
3.1 Obecný návrh	23
3.1.1 Predikce Parkinsonovy choroby	23
3.2 Programovací jazyk	23
3.3 Knihovny	24
3.3.1 Pandas	24
3.3.2 Tensorflow	24
3.3.3 Keras	25
3.3.4 Scikit-learn	25
3.3.5 Imbalanced-learn	25
4 Vstupní data	26
4.1 Aktigrafie	26
4.2 Newcastle polysomnography and accelerometer dataset	26

4.3	Data z projektu niCE-life	27
4.4	Data z Fakultní nemocnice u sv. Anny	28
5	Návrh a zpracování datové množiny	30
5.1	Předzpracovaná datová množina	30
5.2	Datová množina Aktigrafie a RBD-SQ	31
5.2.1	Programové řešení	32
5.3	Datová množina Spánkové deníky a RBD-SQ	33
5.4	Datová množina Aktigrafie, Spánkové deníky a RBD-SQ	33
6	Návrh modelu strojového učení	34
6.1	Neuronové sítě	34
6.2	Metoda podpůrných vektorů	34
6.2.1	Přímá aplikace SVM na vytvořené datové množiny	34
6.2.2	Aplikace SVM na datovou množinu PD a MCI	39
6.2.3	Výběr modelu	43
	Závěr	48
	Literatura	49
	Seznam symbolů a zkratk	54
A	Dataset Aktigrafie a RBD-SQ	56

Seznam obrázků

2.1	Učení s učitelem	20
2.2	Metoda podpůrných vektorů pro dvě dimenze.	21
2.3	Ukázka jádrové transformace.	22
3.1	Návrh modelu strojového učení	23
3.2	Příklad struktury dataframe	24
4.1	Výstup GENEActive v csv	26
4.2	Výstup PSG v Newcastle datasetu	27
6.1	Matice záměn nejlepšího modelu pro aktigrafii	36
6.2	Matice záměn nejlepšího modelu pro aktigrafii po aplikaci xscale . . .	36
6.3	Matice záměn nejlepšího modelu pro spánkové deníky	38
6.4	Matice záměn vybraného modelu 1	45
6.5	Matice záměn vybraného modelu 2	46

Seznam tabulek

1.1	Příklad spánkového deníku.	14
1.2	Výstupní hodnoty zařízení GeneActiv Original	15
1.3	RBD-SQ dotazník	17
4.1	Formát dodaných dat spánkového deníku	28
5.1	Příklad předzpracovaných dat pacienta <i>pre-LBD-1</i> z jedné noci	31
6.1	Procentuální výskyt hodnot y v datové množině	34
6.2	Nejpřesnější kombinace vstupních dat z datové množiny Aktigrafie a RBD-SQ	35
6.3	Nejpřesnější kombinace vstupních dat z datové množiny Spánkové deníky a RBD-SQ	37
6.4	Aktigrafie, Spánkové deníky a RBD-SQ	39
6.5	Procentuální výskyt hodnot y v datové množině	40
6.6	Spánkové deníky a RBD-SQ (Probable Parkinson && MCI)	41
6.7	Spánkové deníky a RBD-SQ (Probable Parkinson && MCI) SMOTE 0,4	41
6.8	Spánkové deníky a RBD-SQ (Probable Parkinson && MCI) SMOTE 0,3	42
6.9	Aktigrafie a RBD-SQ (Probable Parkinson && MCI)	42
6.10	Aktigrafie a RBD-SQ (Probable Parkinson && MCI) SMOTE 0,3	43
6.11	Nejpřesnější modely pro náhodný stav 43	44
6.12	Nejpřesnější modely pro náhodný stav 48	44
6.13	Nejpřesnější modely pro náhodný stav 52	44
6.14	Výsledky modelu Sleep onset latency, Wake bouts, Awakening > 5 minutes, Sleep fragmentation, RBD-SQ	46
6.15	Výsledky modelu Sleep onset latency, Awakening > 5 minutes, RBD-SQ .	47
A.1	Ukázka datové množiny Aktigrafie a RBD-SQ.	57

Úvod

Změny ve spánku jsou brány jako běžný projev spojený se stárnutím [1]. Na druhou stranu spánkové poruchy spojené s těmito změnami, jsou často spojovány s neurodegenerativními chorobami, kdy v případě Parkinsonovy choroby trpí některou ze spánkových poruch 74 až 98 % pacientů [2].

Neurodegenerativní onemocnění jsou diagnostikována v pozdním věku, kdy jsou příznaky pozorovatelné a nemoc je rozšířená. Problém neurodegenerativních onemocnění spočívá v jejich neléčitelnosti. Moderní medicína kromě zmírnění následků nabízí metody zpomalující postup nemoci, které ale jsou z důvodu pozdní diagnózy méně účinné.

Řešení problému lze dosáhnout dřívější diagnózou, kdy jednou z možností je využití strojového učení. Aplikace této metody na data z aktigrafie nabízí pro pacienta jednoduché a automatizované řešení. V případě spánkových deníků se jedná o metodu závislou na manuálním vstupu pacienta, ale za to je velmi levná a dostupná.

Úkolem této práce je z dostupných dat aktigrafie a spánkových deníků vytvořit datovou množinu a s její pomocí navrhnout a natrénovat model strojového učení, který bude schopen na bázi zmíněné datové množiny klasifikovat pacienty trpící Parkinsonovou chorobou.

Diplomová práce je rozdělena do šesti kapitol. První kapitola představuje spánkové parametry, poruchy a neurodegenerativní onemocnění s nimi spojené. Druhá kapitola pojednává o strojovém učení. Teoreticky jsou zmíněny neuronové sítě a SVM, o kterých se uvažuje v následujících kapitolách. Třetí kapitola se zabývá obecným návrhem řešení a ve čtvrté jsou představeny dostupné datové množiny. Poslední dvě kapitoly spadají do praktické části. Pátá kapitola pojednává o návrhu a následné úpravě datových množin. V šesté kapitole je představen návrh modelu strojového učení a výběr konkrétního řešení.

1 Spánek a jeho parametry

Spánek je důležitá aktivita pro lidské tělo, při které dochází k fyzické a psychické regeneraci. Význam nachází i v medicíně, kdy z probíhajících poruch spánku lze předpovědět některá z neurodegenerativních onemocnění. Na tyto nemoci existují léčebné metody invazivní i neinvazivní, které ale jen dokáží zpomalit průběh onemocnění. Problémem je, že se většinou na tyto choroby přichází ve středním, v horším případě pozdním stadiu. Cílem je, pokud neuvažujeme lepší medikaci, najít vhodnou metodu, která nám umožní zavčas detekovat a reagovat na tyto nemoci.

1.1 Fáze spánku

Spánek se skládá ze dvou hlavních fází. První fáze se nazývá REM a u zdravého člověka trvá jedna tato fáze 10 - 60 minut. Je charakteristická vysokou mozkovou aktivitou blízkou mozkové aktivitě probuzeného člověka. Samotné tělo je ve stavu nazvaném atonie, což je dočasná paralýza svalů. Výjimkou jsou oči, které se i přes zavřená víčka rychle pohybují a svaly kontrolující dýchání.

Druhá fáze je pojmenována NREM (non REM). Ta se dále dělí na tři stadia, kdy vyšší číslo znamená hlubší spánek:

- N1 je stadium, ve kterém není lidské tělo ještě úplně uvolněné, ale mozková aktivita se začíná pomalu snižovat. Normálně trvá 1 až 5 minut.
- N2 je stadium, při kterém dochází k poklesu teploty těla, svaly jsou uvolněnější a dýchání spolu s tepovou frekvencí pomalejší. Běžně trvá 10 až 60 minut.
- N3 je stadium nazývaný se hluboký spánek. Mozková aktivita je charakterizována delta vlnami o frekvenci 1 – 2 Hz. Jedno toto stadium trvá 20 až 40 minut a je při něm těžké člověka probudit [3].

1.2 Spánkové parametry a jejich měření

Každá spánková porucha se projevuje jiným způsobem. K jejich rozlišování se používají spánkové parametry, které lze získat pomocí různých metod měření (spánkové deníky, polysomnografie, aktigrafie).

- DSE – Trvání spánkové epizody se používá k výpočtu efektivity spánku a počítá se podle vzorce 1.1. Kromě hodnoty TASAFA, která se zapisuje do rozšířeného konsensuálního spánkového deníku (ECSD), se všechny hodnoty pro výpočet DSE dají získat z běžně používaných spánkových deníků [4]. Vysvětlení všech těchto hodnot se nachází v bodech níže.

$$DSE = SL + TST + WASO + TASAFA \quad (1.1)$$

- NA > 5 – Počet probuzení během noci delších jak 5 minut.
- RSWA – REM spánek bez atonie je charakterizován zvýšenou fázickou nebo tonickou svalovou aktivitou. Tento parametr lze získat pouze z polysomnografie (přesněji EMG).
- SE – Efektivita spánku je důležitý parametr pro evaluaci nespavosti, který je definován jako poměr TST k času v posteli TIB. Výhodou je, že tyto hodnoty se zaznamenávají v běžně používaných spánkových denících. Na druhou stranu nevýhodou je, že čas v posteli TIB zahrnuje i čas nesouvisející se spánkem, například aktivity jako čtení, sledování televize nebo odpočívání, což může výslednou hodnotu SE zkreslit. Efektivitu spánku můžeme získat i pomocí tohoto vzorce: $SE = TST / DSE \cdot 100$, kdy parametr DSE se počítá ze vzorce 1.1. Tento výpočet je oproti předchozí variantě přesnější [4].
- SF – Fragmentace spánku je definována jako poměr všech probuzení během noci k celkovému času v posteli TST v hodinách.
- SL nebo SOL – Spánková latence je čas od zhasnutí světla do začátku první epochy jakékoli fáze spánku. Zjednodušeně je to čas, za jak dlouho sledovaná osoba usnula.
- TASAFA – Doba, po kterou se osoba snaží po finálním probuzení usnout.
- TST – Celkový čas spánku je celková doba spánku zaznamenaná během určitého časového intervalu.
- WASF – Čas strávený v posteli po finálním probuzení.
- WASO – Bdění po počátku spánku je doba, po kterou je pacient vzhůru po prvním usnutí.

1.2.1 Spánkové deníky

Vedení spánkového deníku je neinvazivní metoda, která dává pacientovi a případně lékaři přehled o parametrech souvisejících se spánkem pacienta. Pacient si denně tyto parametry zaznamenává do předem připraveného deníku. Formát deníku není standardizovaný, a tudíž je možné se setkat s velkým množstvím variant, které se od sebe liší například formou otázky, parametru nebo větším obsahem otázek [5]. Formát, který bude použit, závisí na tom, na jakou spánkovou poruchu pacienta je třeba se zaměřit. V každém případě jsou dány základní parametry, které se v podobných variantách vyskytují ve většině formátů. Jedná se například o datum, kdy pacient šel do postele, v kolik se pokoušel usnout a v kolik usnul. V jiné variantě se například místo času, kdy pacient usnul, uvádí čas, za jak dlouho pacient přibližně usnul. Stále se ale jedná o zprávu se stejnou informační hodnotou. Níže je uveden základní spánkový deník s běžně používanými parametry.

Tab. 1.1: Příklad spánkového deníku.

Datum	17.10. 2021	18.10. 2021	19.10. 2021
Čas lehnutí do postele:	9:20	10:30	
Čas usnutí:	9:30	10:50	
Počet probuzení:	0	2	
Trvání všech probuzení:	0 minut	10 min	
Čas finálního probuzení:	7:00	6:50	
Čas vstání z postele:	7:15	7:15	
Kvalita spánku	velmi dobrá	dobrá	
Poznámky	-	po večeři 3 piva	

1.2.2 Polysomnografie

Polysomnografie (PSG) je takzvaný zlatý standard pro měření spánku. Standardně se provádí přes noc ve specializovaných spánkových laboratořích za dozoru odborného pracovníka, který by měl být přítomen po celou dobu měření [6]. Ve speciálních případech se provádí ještě měření ve dne (nazývané MSLT). Polysomnografie monitoruje tyto parametry: elektroenceleografie (EEG), elektrokardiografie (EKG), elektromyografie (EMG), elektrookulografie (EOG), dechová aktivita, okysličení krve, poloha těla, esofageální tlak, tělesná teplota a krevní tlak [7].

1.2.3 Aktigrafie

Aktigrafie (ACG) je metoda, která zaznamenává pohybovou aktivitu končetin v čase (některá zařízení snímají i jiné hodnoty, jako je teplota nebo intenzita světla v místnosti) [8]. Měření se provádí pomocí aktigrafického zařízení, které se po určitý čas nosí na zápěstí, kotníku a pase. Ve spánkové medicíně se používají hlavně první dvě zmíněné varianty, kdy se poté na data ze zařízení aplikují algoritmy, ze kterých se následně získá doba bdělosti a přehled, kdy uživatel spal. Doporučená doba, po kterou by se aktigrafické zařízení mělo nosit, je minimálně 72 hodin až 14 po sobě jdoucích dní [9].

Použité zařízení

Měření dat aktigrafie bylo pro tuto diplomovou práci provedeno pomocí zařízení GENEActiv Original od firmy Activinsights. GENEActiv je lehký a voděodolný akcelerometr nošený na zápěstí, který je navržený tak, aby nezasahoval do každodenní činnosti pacienta [8]. K sehnání je ve 4 variantách:

- **GENEActiv Original** je základní zařízení v GENEActiv sérii. Je navržen pro monitorování při normálních životních aktivitách, včetně spánku a fyzické aktivity. Výdrž baterie závisí na frekvenci snímání dat, kdy pro maximum 100 Hz vydrží 7 dní a pro minimum 10 Hz vydrží až 45 dní. Zařízení obsahuje paměť o velikosti 0,5 Gb, která je schopna až do vybití ukládat další měřená data. Ta se při nabíjení přesunou v binárním, nebo za pomoci GENEActiv softwaru v CSV formátu do počítače k analýze [8].
- **GENEActiv Action** je navržen pro profesionální sportovce, kdy je k měření akcelerace potřeba širší rozsah a vyšší vzorkovací frekvenci.
- **GENEActiv Sleep** monitoruje cirkadiánní rytmus, což je biologický rytmus o délce okolo 24 hodin ovlivňující mimo jiné i spánek. Kromě předchozí možnosti ještě umožňuje stálé měření teploty kůže.
- **GENEActiv Wireless** umožňuje streamovat data v reálném čase.

Tab. 1.2: Výstupní hodnoty zařízení GeneActiv Original [8].

Měřená hodnota	Rozsah	Jednotky
Zrychlení v ose X	8 až -8	[g]
Zrychlení v ose Y	8 až -8	[g]
Zrychlení v ose Z	8 až -8	[g]
Intenzita osvětlení	0 až 5000	[lux]
Uživatelské tlačítko	0 nebo 1	
Teplota	0 až 60	[°C]

Porovnání aktigrafie s polysomnografií

Aktigrafie je oproti polysomnografii pro pacienta pohodlnější. Není nutné navštěvovat z důvodu měření nemocniční zařízení a nechat se připojovat na přístroje. Tudíž je na něj vyvinuta menší psychická zátěž, což je faktor, který může spánkové parametry zkreslovat. Další výhodou aktigrafie je schopnost dlouhodobého měření a relativně nízké náklady na provoz [9]. Polysomnografie na druhou stranu nabízí komplexnější a přesnější analýzu spánku.

1.3 Poruchy spánku a neurodegenerativní nemoci s nimi spojené

1.3.1 Insomnie

Insomnie, též nazývaná nespavost, je definována jako problém s usínáním, udržováním spánku a častým probouzením. Je doprovázena sníženým fungováním člověka během dne, které přetrvává minimálně 4 týdny. Jedná se například o sníženou pozornost a problémy se soustředěním. Nespavost může trvat pár dní nebo může být chronická. To závisí na příčinách a symptomech s ní spojených. Může se jednat například o vyšší přísun alkoholu, chvilkový stres, úzkost, špatné spánkové návyky, chronické bolesti, ale i RLS (sekce 1.3.3) a PLM (sekce 1.3.4) [17].

Nespavost lze zhodnotit pomocí parametrů získaných z polysomnografie, spánkových deníků i aktigrafie. Pro aktigrafii jsou považovány za nejvhodnější parametry TST, SOL a $NA > 5$ (počet probuzení delší jak 5 minut) [18]. U spánkových deníků se jedná obecně o parametry SOL, WASO, TST, TIB a SE [5].

1.3.2 Nykturie

Nykturie je definována jako potřeba urinace jednou nebo vícekrát během noci. Je jednou z příčin poruch spánku, což se ve výsledku promítá i do kvality života a zdraví. Častěji se vyskytuje u starších osob [19].

Nykturie je častou příčinou výše zmíněné insomnie (viz sekce 1.3.1), tudíž z hlediska spánkové medicíny lze pozorovat změny na stejných spánkových parametrech 1.3.1, avšak ze spánkových parametrů nykturií s jistotou diagnostikovat nelze [19].

1.3.3 Syndrom neklidných nohou (RLS)

RLS je neurologické onemocnění projevující se nutkáním hýbat s dolními končetinami, které je doprovázené mravenčením, svěděním, pulzujícím pocitem a bolestí končetin. Tyto projevy jsou charakteristické tím, že se dějí v klidu (při sezení a ležení) hlavně ve večerních hodinách a noci. Ve výsledku je negativně ovlivněna kvalita spánku pacienta a s tím spojená insomnie.

RLS se dělí na primární, který se nepojí s žádnou nemocí a jeho příčinou je genetika, a sekundární, který se spojuje například s nedostatkem železa, těhotenstvím a Parkinsonovou chorobou (sekce 1.3.7) [20].

1.3.4 Periodické pohyby končetin (PLM)

PLM je spánková porucha charakterizovaná periodickým pohybem končetin při spánku, což může způsobovat probouzení (insomnie, sekce 1.3.1). Projevuje se i při bdělosti, a to ve formě trhavých pohybů.

PLM je častější s rostoucím věkem a vyskytuje se spolu s RLS (80 až 90 %), RBD (70 %), narkolepsií (45 až 60 %) a insomnií (1 až 5 %) [21].

1.3.5 Porucha chování v REM spánku

Porucha chování v REM spánku (RBD) se dělí na idiopatickou (iRBD) a sekundární. Projevuje se častými nepříjemnými sny, svalovým napětím a energickými pohyby, které mohou být potenciálně škodlivé. Navíc je velká šance, že u pacientů diagnostikovaných s iRBD se vyvine některá z neurodegenerativních onemocnění ze skupiny alfa synukleionopatie (Parkinsonova nemoc, Demece s Lewyho tělísky a Mnohočetná atrofie systému) [11]. K diagnóze se používá polysomnografie, pomocí které se získá spánkový parametr RSWA [22]. Podle studie z roku 2019 lze k diagnóze použít i aktigrafie spolu s RBD-SQ dotazníkem, který se skládá ze 13 otázek s odpovědí ano/ne (viz tab. 4.1) [25].

Tab. 1.3: RBD-SQ dotazník [25].

Otázka		Odpověď
1.	Někdy mívám velmi živé sny.	ano/ne
2.	Moje sny mají často agresivní nebo akční obsah.	ano/ne
3.	Obsah snu většinou odpovídá mému nočnímu chování.	ano/ne
4.	Vím, že se mé ruce nebo nohy ve spánku pohybují.	ano/ne
5.	Stalo se, že jsem v posteli (skoro) ublížil partnerovi nebo sobě.	ano/ne
6.	Během svých snů se děly následující jevy:	
6.1.	mluvení, křik, nadávání, hlasitý smích.	ano/ne
6.2.	náhlé pohyby končetin „rvačky“	ano/ne
6.3.	gesta, komplexní pohyby, které jsou ve spánku zbytečné (například mávat, zdravit, plašit komáry, padat z postele)	ano/ne
6.4.	věci, které spadly okolo postele (například lampa, kniha, brýle)	ano/ne
7.	Stává se, že mě můj pohyb budí.	ano/ne
8.	Po probuzení si často dobře pamatuju obsah svých snů.	ano/ne
9.	Můj spánek je často narušený.	ano/ne
10.	Mám/měl jsem nemoc nervového systému (například mrtvice, poranění hlavy a mozku, parkinsonismus, RLS, narkolepsie, deprese, epilepsie, zánětlivé onemocnění mozku). Které?	ano/ne

1.3.6 Alzheimerova choroba

Alzheimerova choroba je nejběžnější neurodegenerativní chorobou. Byla objevena Aloisem Alzheimerem v roce 1907 a stále trápí společnost do dnešní doby [12]. Projevuje se ztrátou mozkových buněk, což má za následek ztrátu paměti a stavy s ní související. Léčba Alzheimerovy choroby se potýká s několika problémy:

1. Neexistuje lék, který by tuto chorobu léčil.
2. Ve většině případů se na chorobu přichází, až po projevu viditelných příznaků, a to ve věku okolo 60 až 70 let.
3. Dostupná medikace je brána jako podpůrná (zmírňuje příznaky nemoci) [13].

Alzheimerova choroba, ale i další neurodegenerativních onemocnění, se často projevuje poruchami spánku. Tyto projevy se projevují dříve než samotné onemocnění, čehož lze využít k včasné diagnóze. Mezi tyto poruchy patří: potíže s usínáním, opakované probouzení, brzké probouzení, ospalost a časté zdřímnutí během dne [14], [15].

1.3.7 Parkinsonova choroba

Parkinsonova choroba je druhou nejběžnější neurodegenerativní chorobou. Je zapříčiněna nadměrnou ztrátou nervových buněk a projevuje se například sníženou hybností, třesem končetin a problémy se spánkem. Léčba Parkinsonovy choroby není v současné době možná. Existují pouze prostředky, které umožňují mírnit průběh nemoci. Proto je vhodné co nejdříve tuto nemoc detekovat a zpomalit průběh nemoci. Jeden z projevů, který se s Parkinsonovou chorobou pojí a je využíván k predikci, jsou právě spánkové poruchy [16]. Přesněji se jedná o RBD, insomni, nykturii, syndrom neklidných nohou (RLS), periodické pohyby končetin (PLM), poruchy dýchání ve spánku, excesivní denní spavost a poruchy cirkadiálního rytmu [22].

Nejlépe se pro predikci Parkinsonovy choroby jeví RBD (sekce 1.3.5), které se po vyšetření pacientů pomocí PSG vyskytovalo s pravděpodobností 39 až 46 % [22].

1.3.8 Mírná kognitivní porucha

Mírná kognitivní porucha (zkratka MCI) je syndrom, který je definovaný jako pokles kognitivních funkcí¹, který se jeví u pacienta s daným věkem a vzděláním jako abnormální. Samotná porucha značně nezasahuje do života a někteří pacienti se po čase této poruchy zbaví. Více jak polovině pacientů (okolo 65 % [24]) je ale do pěti let diagnostikována demence [23].

¹Mezi kognitivní funkce patří paměť, pozornost, rychlost zpracování informací, řeč a prostorová orientace.

2 Strojové učení

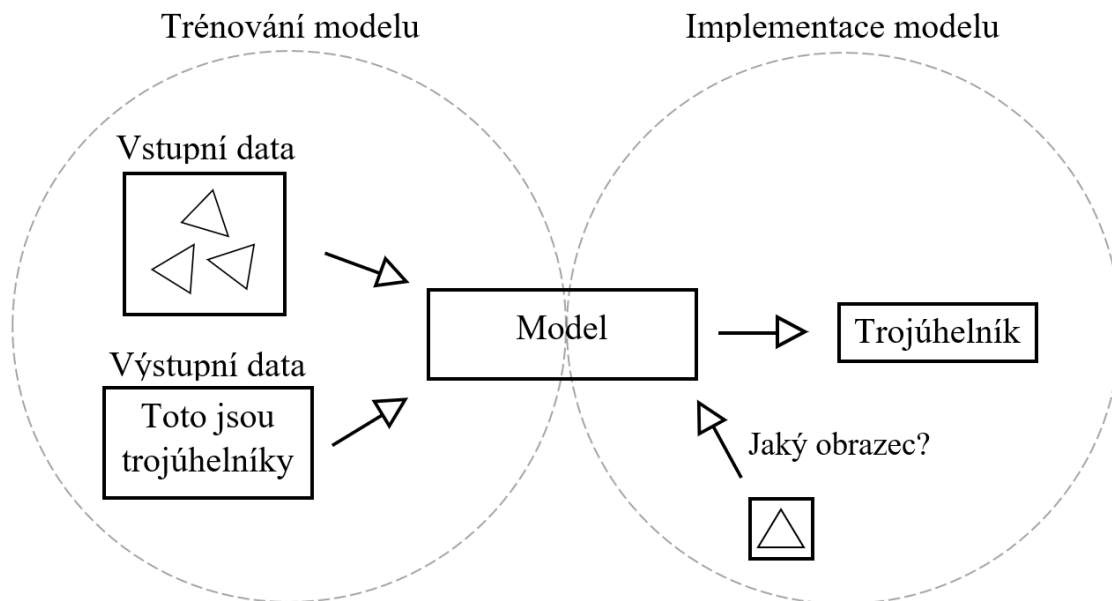
Strojové učení (*machine learning*) spadá pod oblast umělé inteligence a lze obecně popsat jako výpočetní metody, algoritmy, techniky, které používají zkušenosti ke zlepšení výkonu a přesnosti (systém se tzv. „učí“). Zkušenosti jsou prezentovány jako minulé informace, které jsou ve formě elektronických dat dostupné učiteli [26]. Strojové učení se používá v mnoha odvětvích. Jedná se například o bankovníctví, ekonomiku, kvalitu dat, zdravotnictví, ve kterém nachází uplatnění právě v diagnostice nemocí, jejíž součástí jsou neurodegenerativní onemocnění.

Kromě oblasti použití je možné strojové učení rozdělit podle použitých modelů a procesu učení (sekce 2.1).

2.1 Proces strojového učení

Proces strojového učení lze rozdělit do 4 kategorií.

- Učení s učitelem (*supervised learning*), kdy modelu jsou dodána vstupní a očekávaná výstupní data, která mu zadá učitel (viz obr. 2.1).
- Učení bez učitele (*unsupervised learning*) je trénovací model, který dostává pouze vstupní data.
- Kombinace učení s učitelem a bez učitele (*semi-supervised learning*) je kombinací vstupních dat s výstupními a pouze vstupních dat, která jsou většinou větší jak předchozí skupina dat.
- Zpětnovazebné učení (*reinforcement learning*) funguje na principu „odměn“ a „trestů“, které dostává agent podle provedených akcí v daném prostředí [26].



Obr. 2.1: Příklad jednoduchého modelu učení s učitelem.

2.2 Souhrn použitých metod strojového učení na data akcelerometrů

Studie z roku 2019 [27], která systematicky shrnuje 62 studií týkajících se kalibrace a validace měření akcelerometru, prezentuje použité metody strojového učení. Nejčastěji používané byly neuronové sítě, následované metodou podpůrných vektorů (viz níže).

- Neuronové sítě (ANN) - 32 studií.
- Metoda podpůrných vektorů (SVM) - 18 studií.
- Metoda náhodný les (RF) - 12 studií.
- Metoda rozhodovacích stromů (DT) - 11 studií.
- Lineární regrese (LR) - 7 studií.

Dále byly v 17 studiích zmíněny Bayesovské sítě, algoritmus k-nejbližších sousedů, metoda Naive Bayes a skrytý Markovův model [27].

2.3 Neuronové sítě

Strojové učení nabízí různé varianty modelů (rozhodovací stromy, podpůrné vektory a další). Mezi tyto modely patří i umělé neuronové sítě, které se staly populárním a užitečným nástrojem pro klasifikaci, rozpoznávání vzorů a predikci [28]. Základ

neuronových sítí položili v roce 1943 vědci Warren McCulloch a Walter Pitts. Ve své práci popsali funkci neuronu a sestavení neuronové sítě z jednoduchých elektrických obvodů [29]. V 50. letech minulého století se vědci snažili převést neuronové sítě do výpočetních systémů. V roce 1958 sestavil Frank Rosenblatt jednoduchou síť. Ta obsahuje jednu vrstvu s neuronem, která provádí matematickou operaci a pomocí ní rozřídí vstupy do dvou tříd.

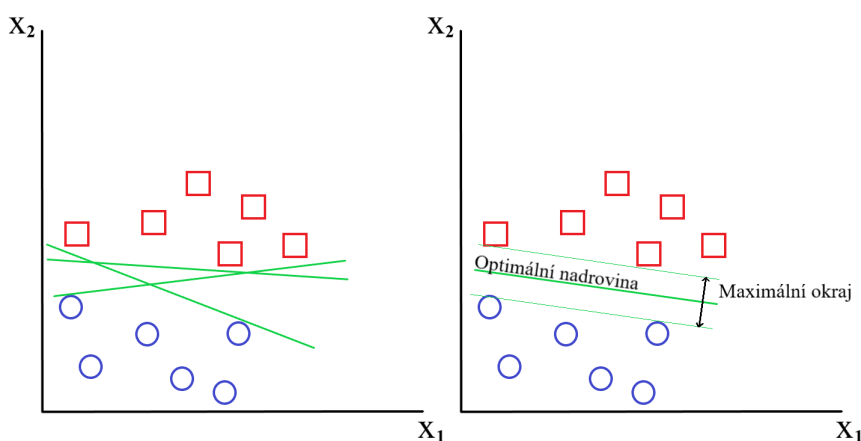
Funkce umělé neuronové sítě se podobá biologické variantě. Biologická neuronová síť se skládá z velkého počtu neuronů (pro lidský mozek jsou řády v bilionech). Ty mezi sebou, vysíláním elektrických impulzů, neustále komunikují pomocí výběžků (zvané dendrity). Dendrity neuronů jsou mezi sebou spojeny přes spoj zvaný synapse, který se opakující komunikací posiluje (učí se). Následnou funkcí je zapamatování, které vzniká stimulací výběžků zvaných axony [30].

2.4 Metoda podpůrných vektorů

Metoda podpůrných vektorů SVM je metoda strojového učení, která spadá pod učení s učitelem a používá se pro klasifikaci, regresi a detekci odlehlých hodnot [36]. Jejím cílem je nalézt optimální nadrovinu v n -dimenzionálním prostoru, kde n je počet datových vstupů (tzv. features). Tato nadrovina slouží jako prahová hodnota, pomocí které se klasifikují jednotlivé datové body [31].

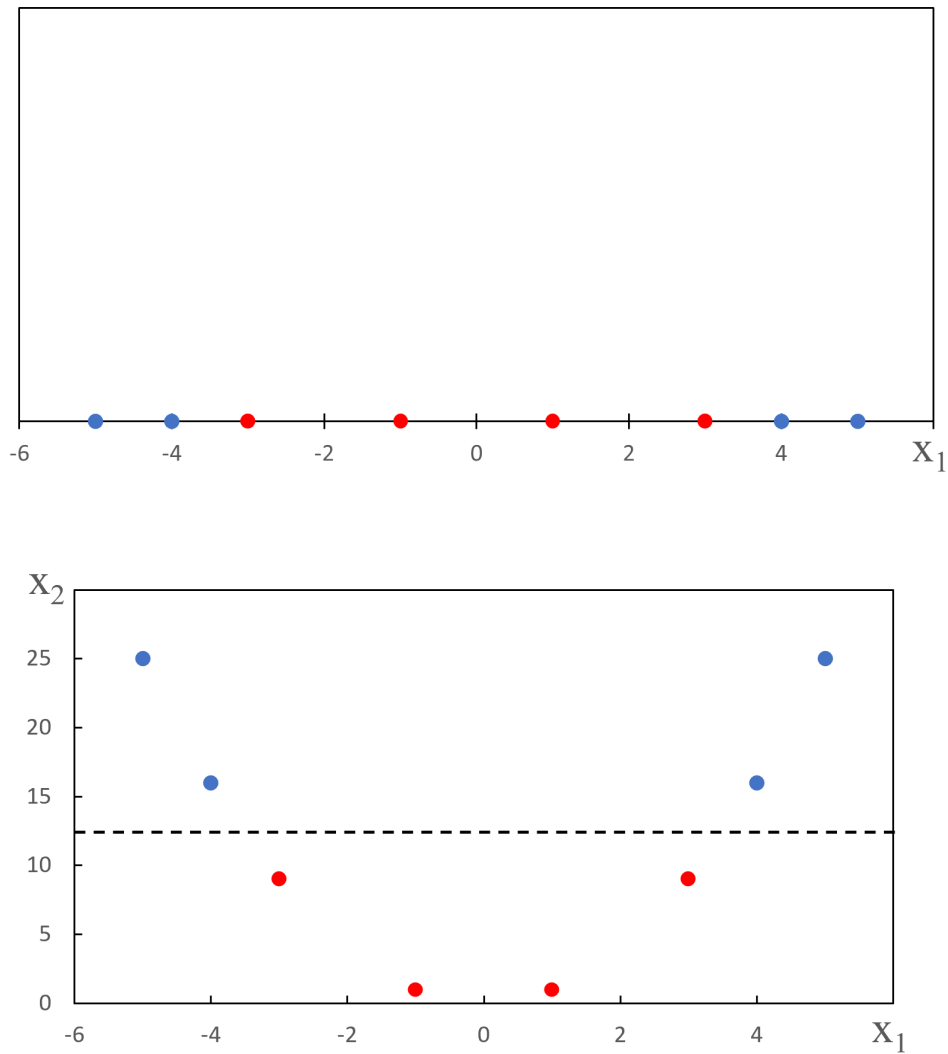
SVM nabízí dvě varianty:

- Lineární – jedná se o jednoduchou metodu, u které nedochází k jádrové transformaci (kernel transformation). Data by ale měla být lineárně separovatelná. Obecný příklad je znázorněn na obr. 2.2.



Obr. 2.2: Výběr optimální roviny lineárního modelu SVM, kde kolečkem a čtvercem jsou znázorněny klasifikované datové body a zelenou přímkou nadroviny.

- Nelineární – s nelineární variantou se lze setkat častěji (výskyt lineárně separovatelných datových množin v praxi je nízký). Ke zpracování nelineárních dat se používá jádrová transformace, která má za úkol tato data převést do lineárně separovatelné podoby.



Obr. 2.3: Zisk lineárně separovatelných dat po aplikaci kvadratické jádrové transformace $\phi(x) = x^2$ do 2D [32].

3 Návrh řešení

V této kapitole bude představen návrh řešení problematiky identifikace spánkových poruch. Ten obsahuje základní myšlenku za touto problematikou, výběr programovacího jazyka a jeho krátký popis. Dále jsou uvedeny knihovny, které jsou v diplomové práci použity a splňují mezi sebou požadavek kompatibility.

3.1 Obecný návrh

K identifikaci spánkových poruch bude v následující diplomové práci použito učení s učitelem (tzv. supervised learning, sekce 2.1). Takto sestavený a naučený model (viz obr. 3.1) bude mít za úkol rozhodnout, zda pacient trpí nebo v budoucnu bude trpět některým z neurodegenerativních onemocnění.

Každé neurodegenerativní onemocnění se projevuje odlišně. Proto bude pro trénování modelu zapotřebí sestavit pro každé neurodegenerativní onemocnění vlastní datové množiny.



Obr. 3.1: Obecný postup při návrhu modelu strojového učení, kdy frází „best“ je myšlen proces vytváření nejlepšího možného modelu.

3.1.1 Predikce Parkinsonovy choroby

Pro predikci Parkinsonovy choroby vychází nejlépe RBD (sekce 1.3.5) a insomnie (sekce 1.3.1) [22]. Z těchto poznatků bude vycházet návrh a zpracování jednotlivých datových množin, které budou kombinovat data z aktigrafie, RBD-SQ, spánkových deníků a případně polysomnografie. Hlavní myšlenkou je vytvoření datové množiny, která bude využívat hlavně data z aktigrafie a RBD-SQ dotazníku případně doplněná daty ze spánkových deníků, tak aby byla dosažena co největší přesnost predikce.

3.2 Programovací jazyk

Jako programovací jazyk byl vybrán Python, který je uživatelsky přívětivý a nabízí velkou škálu kvalitních knihoven. Nejnovější verze je 3.10.0 (neuvažuje se pre-release

verze). Pro potřeby práce ale byla vybrána verze 3.9.7, která je kompatibilní s knihovnou Tensorflow (verze 3.6 - 3.9 pro Tensorflow 2) a je zároveň nejvyšší verzí Python 3.9 [33].

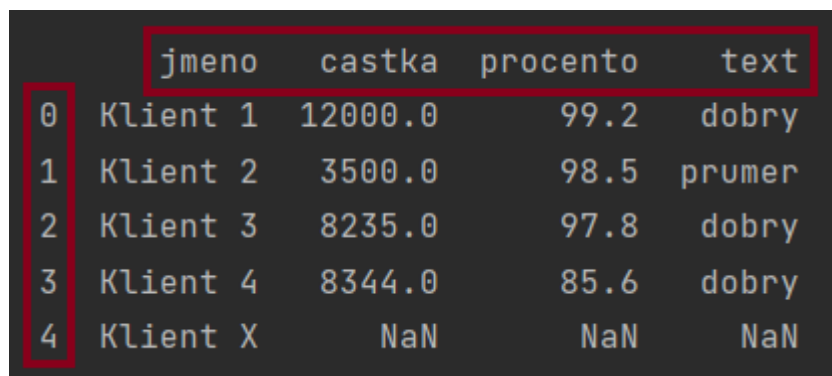
3.3 Knihovny

Tato sekce se zabývá popisem a výběrem vhodných knihoven. V případě této diplomové práce se jedná o knihovny pro:

1. Jednoduché a přehledné zpracování dat.
2. Návrh, sestavení a učení modelů strojového učení.

3.3.1 Pandas

Pandas je knihovna sloužící k jednoduché analýze a manipulaci dat. V našem případě byla vybrána ke zpracování CSV datové množiny pomocí funkce `pandas.readcsv()`, která oproti funkci `csv.reader()` nabízí víc možností. Výstupem funkce je `DataFrame`. Jedná se o 2D tabulkovou datovou strukturu, která je označena osami (tmavě červené obdélníky na obr. 3.2).



	jmeno	castka	procento	text
0	Klient 1	12000.0	99.2	dobry
1	Klient 2	3500.0	98.5	prumer
2	Klient 3	8235.0	97.8	dobry
3	Klient 4	8344.0	85.6	dobry
4	Klient X	NaN	NaN	NaN

Obr. 3.2: Příklad datové struktury `DataFrame`.

3.3.2 Tensorflow

Tensorflow je otevřená platforma, jejíž první verze byla vydána Googlem v roce 2017. V nynější době se používá Tensorflow 2, který se průběžně aktualizuje (verze 2.6.0 vydaná 11. srpna 2021). Obsahuje řadu flexibilních nástrojů a knihoven, které patří k nejmodernějším v ML. Tensorflow se používá ve známých firmách jako je Google, Intel nebo Twitter [34].

3.3.3 Keras

Keras je otevřená knihovna napsaná v jazyce `Python` a je nástavbou na platformu `Tensorflow`, od které vyžívá výhod jako je škálovatelnost a multiplatformní možnosti. Další výhodou je možnost kompilovat Keras na TPU nebo velkých souborech GPU. Keras nabízí rozhraní pro řešení problémů strojového učení se zaměřením na hluboké učení [35].

3.3.4 Scikit-learn

Scikit-learn je volně dostupná knihovna pro jazyk `Python`, která je vyvíjena od roku 2007. Nabízí různá řešení v oblasti strojového učení (klasifikace, regrese a shlukování). Součástí je například implementace algoritmů SVM, RF a k-nejbližších sousedů [36].

3.3.5 Imbalanced-learn

Imbalanced-learn je volně dostupná, MIT licensovaná knihovna, která je závislá na knihovně `scikit-learn` a nabízí nástroje na řešení klasifikace nevyvážených tříd (nadvzorkování a podvzorkování tříd) [37].

4 Vstupní data

K této diplomové práci jsou k dispozici vstupní data ze tří různých zdrojů. Jedná se o Newcastle polysomnography and accelerometer dataset, data aktigrafie a spánkových deníků z projektu niCE-life¹, data aktigrafie a spánkových deníků z Fakultní nemocnice u sv. Anny v Brně.

4.1 Aktigrafie

Data aktigrafie byla ve všech datových množinách nasbírána pomocí zařízení GENEActiv Original (bližší popis v sekci 1.2.3). Sběr dat probíhá tak, že subjekt musí po určitou dobu zařízení nosit (běžně jeden týden). Poté je třeba zařízení připojit k počítači. Data se následně extrahují do binárního souboru. Software GENEActiv nabízí možnost převést tento binární soubor do CSV formátu [8], kdy výstupem je několik hodnot, které jsou od sebe odděleny čárkami. Jedná se o čas, zrychlení ve třech osách, teplota, stav tlačítka a osvětlení v prostředí (obr. 4.1).

```
2020-06-24 23:20:04:900,-0.6650,0.3566,-0.7007,0,0,23.1
2020-06-24 23:20:04:940,-0.6490,0.3526,-0.6967,0,0,23.1
2020-06-24 23:20:04:980,-0.6690,0.3646,-0.6808,0,0,23.1
2020-06-24 23:20:05:020,-0.6650,0.3606,-0.6729,0,0,23.1
2020-06-24 23:20:05:060,-0.6490,0.3486,-0.6848,0,0,23.1
2020-06-24 23:20:05:100,-0.6530,0.3646,-0.6769,0,0,23.1
```

Obr. 4.1: Ukázka výstupních hodnot GENEActiv v CSV formátu.

4.2 Newcastle polysomnography and accelerometer dataset

Jedná se o volně přístupnou datovou množinu, která obsahuje CSV soubor s informacemi o pacientech, 55 záznamů aktigrafie (CSV soubory), 28 záznamů vyhodnocení spánku pomocí polysomnografie (textový soubor). Měření proběhlo na 28 pacientech ve věkovém rozmezí 21 až 72 let při poměru pohlaví 17 : 11, kdy kromě 8 pacientů měli všichni zúčastnění alespoň jednu spánkovou poruchu (u čtyř pacientů byly diagnostikovány 2 poruchy). Jedná se o spánkové poruchy: hypersomnie, insomnie,

¹Celé jméno projektu: *Development of an integrated concept for the deployment of innovative technologies and services allowing independent living of frail elderly (niCE-life)*, označení projektu: CE1581.

porucha chování v REM spánku, spánková apnoe, narkolepsie, parasomnie, syndrom neklidných nohou, spánková paralýza a nykturie [38]. Data polysomnografie byla naměřena standardní procedurou (obsahující videonahrávky, EEG, EMG, monitorování pohybu očí a pohybů končetin ve spánku [39]) ve specializované spánkové klinice [40]. Výstupem jsou textové soubory (viz obr. 4.2), které v třicetisekundových intervalech zaznamenávají fázi spánku a polohu pacienta [38].

Sleep Stage	Position	Time [hh:mm:ss]	Event	Duration[s]
N1	Right	22:57:03	N1	30
N1	Right	22:57:33	N1	30
W	Right	22:58:03	W	30
W	Right	22:58:33	W	30
W	Right	22:59:03	W	30
N1	Right	22:59:33	N1	30
W	Right	23:00:03	W	30
W	Right	23:00:33	W	30
W	Right	23:01:03	W	30
W	Right	23:01:33	W	30
W	Right	23:02:03	W	30
W	Right	23:02:33	W	30

Obr. 4.2: Výstup PSG v Newcastle datasetu.

Aktigrafie byla změřena pomocí dvou akcelerometrů, které měli mít pacienti nasazené na obou zápěstích po jednu noc (výjimkou je pacient číslo 10, u kterého je záznam pouze z levého zápěstí). Akcelerometr byl nastaven na snímání při frekvenci 85,7 Hz [40]. Ostatní parametry akcelerometru jsou popsány v subsekcí 1.2.3 a formát výstupních dat v sekci 4.1.

4.3 Data z projektu niCE-life

Tato datová množina se skládá z aktigrafie (viz sekce 4.1) a spánkových deníků 7 subjektů ve věkovém rozmezí 66 až 83 let. Aktigrafie byla změřena pomocí jednoho akcelerometru, který byl nastavený na snímání při frekvenci 25 Hz. Ostatní parametry akcelerometru jsou popsány v subsekcí 1.2.3 a formát výstupních dat v sekci 4.1. Spánkový deník byl použit v německém jazyce ve formátu znázorněném v tabulce 4.1. Nikomu ze zúčastněných nebyla prozatím diagnostikována žádná spánková porucha, jedná se ovšem o seniory z domu s pečovatelskou službou, nikoliv o pacienty s profesionálně stanovenou diagnózou neurologickými experty. Pro potřeby této práce budou považováni za zdravé kontroly.

Tab. 4.1: Formát použitého spánkového deníku pro datovou množinu niCE-life a Fakultní nemocnici u sv. Anny a Fakultní nemocnici Olomouc.

Datum	22.11.2021
1a. Kolikrát jste si šel/šla během dne lehnout?	1x
1b. Kolik minut jste během dne spasal/a?	50 min.
2a. Kolik alkoholických nápojů jste přes den vypil/a?	1
2b. V kolik hodin jste vypil/a poslední?	21:00
3a. Kolik nápojů s kofeinem (kafe, čaj, energy drinků) jste přes den vypil/a?	2
3b. V kolik hodin jste vypil/a poslední?	18:32
4. Užil/a jste lék na spaní? (pokud ano jaký, jakou dávku a v jaký čas)	
5. V kolik hodin jste se snažil/a usnout?	22:15
6. Jak dlouho vám přibližně trvalo usnout?	15 min.
7. Kolikrát jste se během noci probudil/a? (pro každé probuzení: přibližně od kdy - do kdy)	1x 2:02 - 2:10
8. V kolik hodin jste se probudil/a?	6:28
9. Jak dlouho po probuzení (přibližně) jste zůstal/a ležet v posteli?	15 min.
10. Jak byste ohodnotil/a kvalitu spánku?	Velmi špatný Špatný Průměr <u>Dobrý</u> Velmi dobrý
11. Jak odpočínutě jste se po probuzení cítil/a?	Neodpočínutě Trochu odpočínutě Docela odpočínutě <u>Odpočínutě</u> Velmi odpočínutě
12. Komentář (nepovinné)	nachlazení

4.4 Data z Fakultní nemocnice u sv. Anny

Datová množina od FN u sv. Anny je podobná předchozí množině (datová množina niCE-life v sekci 4.3). Skládá se též z aktigrafie (viz sekce 4.1) a spánkových deníků. Aktigrafie byla změřena pomocí jednoho akcelerometru (frekvence 25 Hz). Ostatní parametry akcelerometru jsou popsány v subsekci 1.2.3 a formát výstupních

dat v sekci 4.1. Spánkový deník byl použit ve formátu znázorněném v tabulce 4.1. Rozdílem oproti předchozí datové množině je počet pacientů, který je v tomto případě 95. Věkové rozmezí pacientů se pohybuje od 55 po 81 let. Dalším podstatným rozdílem je fakt, že pacienti podstoupili řadu vyšetření na přítomnost neurodegenerativních onemocnění, a byla jim předběžně stanovena diagnóza profesionálními neurology.

Bohužel ne všechna data jsou použitelná. U pacienta *pre-LBD-46* a *pre-LBD-59* jsou data poškozena. U pacientů *pre-LBD-31*, *pre-LBD-33*, *pre-LBD-48*, *pre-LBD-54* a *pre-LBD-75* chybí záznam spánkového deníku a u pacientů *pre-LBD-65*, *pre-LBD-75* a *pre-LBD-89* chybí všechna data. Celkem je k dispozici 85 záznamů aktigrafie spolu se spánkovými deníky.

5 Návrh a zpracování datové množiny

Pro strojové učení je důležitým prvkem datová množina. Ta by, pro natrénování co nejlepšího modelu, měla být dostatečně obsáhlá, kvalitně zpracovaná a obsahovat relevantní data. V této kapitole budou přiblíženy jednotlivé databáze a následně bude rozebráno jejich zpracování.

Na data z FN u sv. Anny a projektu niCE-life byla aplikovaná heuristická metoda (článek: *A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer*) [39].¹ Příkladem je vzorec 5.1, který se používá k vypočítání úhlu ze zrychlení os x , y a z . Tento úhel se pak porovnává s nastavenou mezní hodnotou, která rozhoduje, zda daná osoba spí. Následně se aplikací algoritmů získaly spánkové parametry. Výstupem je datová množina popsaná v následující sekci 5.1.

$$\varphi = \left(\tan^{-1} \frac{a_z^2}{\sqrt{a_x^2 + a_y^2}} \right) \cdot 180/\pi \quad (5.1)$$

5.1 Předzpracovaná datová množina

Tato datová množina obsahuje 18 typů hodnot. Jedná se o:

- Subject – Název subjektu (pre-LBD-X, HC-X, AUX-Y).
- Date – Datum provedeného měření ve formátu den DD.MM.RRRR.
- Probable Parkinson Disease – Pacient trpící pravděpodobně Parkinsonovou chorobou (PRAVDA, NEPRAVDA).
- Probable Mild Cognitive Impairment – Pacient trpící pravděpodobně mírnou kognitivní poruchou (PRAVDA, NEPRAVDA).
- Sleep Apnea – Pacient pravděpodobně trpící spánkovou apnoe (PRAVDA, NEPRAVDA).
- Wake bouts – Počet probuzení.

Dále se jedná o TIB, SOL, SOL norm, WASO, WASO norm, WASF, TST, NA > 5, NA > 5 norm, SE a SF. Hodnoty norm nabývají hodnot 1 (appropriate), 0 (uncertain) a -1 (inappropriate) a určují se ze spánkových norem [42].

Do stejného formátu byly exportovány i spánkové deníky. Hodnoty WASO, TST, Wake bouts, NA > 5, SE a SF se ve většině případů znatelně liší od hodnot aktigrafie (červeně znázorněné hodnoty v tabulce 5.1). U spánkových deníků je chybovým faktorem pacient, který si ráno neuvědomí, kolikrát se reálně vzbudil a jak dlouho tato probuzení trvala. Tyto hodnoty jsou pravděpodobně nižší než by měly reálně být. Otázkou je, jak přesné a relevantní jsou data aktigrafie.

¹Implementace je dostupná na: <https://github.com/xsigmu06/actigraphy-processing> [41].

Tab. 5.1: Příklad předzpracovaných dat pacienta pre-LBD-1 z jedné noci, kdy červeně jsou označena rozdílná data.

	Aktigraf	Deník
Čas pokoušení se usnout		23:35:00
Čas začátku spánku	23:41:00	23:45:00
Čas probuzení	5:55:00	05:45:00
Čas vstání z postele		5:55:00
Čas strávený v posteli (TIB)		6:20:00
Trvání usnutí (SOL)	0:06:00	0:10:00
Trvání všech probuzení (WASO)	2:17:30	0:02:00
Čas vstávání z postele (WASF)	0:00:00	0:10:00
Celkové trvání spánku (TST)	3:56:30	5:58:00
Počet probuzení za noc	17	2
Počet probuzení delší jak 5 minut (NA>5)	13	0
Efektivita spánku (SE)	62,6%	94,2%
Fragmentace spánku	4,31	0,34

5.2 Datová množina Aktigrafie a RBD-SQ

Ideální myšlenkou je vytvoření a použití datové množiny, která se bude sestávat pouze z dat aktigrafie a výstupní hodnoty RBD-SQ dotazníku, a zároveň bude po natrénování vhodného modelu dosahovat vysoké přesnosti při predikci Parkinsonovy choroby.

Z předzpracované datové množiny byly odstraněny hodnoty Subject, Date, Sleep Apnea, SOL norm, WASO norm a NA > 5 norm. Většina těchto dat je pro trénování predikčního modelu irelevantní (nesou nulovou nebo chybnou hodnotu). Výjimkou je hodnota WASO norm, která se jeví jako potenciální náhrada za hodnotu WASO a bude vyzkoušena.

Po přidání výsledné hodnoty RBD-SQ z centrální databáze vznikla datová množina, která se skládá z hodnot Probable Parkinson Disease (očekávaná výstupní hodnota neuronové sítě y), Probable Mild Cognitive Impairment, TIB, WASO, WASF, TST, Wake bouts (WB), NA > 5, SE, SF a RBD-SQ. V příloze se přiložená tabulka A.1, která představuje strukturu vytvořené datové množiny.

5.2.1 Programové řešení

Tato datová množina byla zpracována s využitím knihovny `pandas` (sekce 3.3.1). Načtené datové množiny aktigrafie a RBD-SQ byly spojeny dohromady. Po spojení vznikly prázdné hodnoty, které bylo potřeba odstranit. Následně byla implementována funkce `df.drop(columns=dropped_columns)`, která umožňuje odstranit libovolné sloupce definované v proměnné `dropped_columns`. Nakonec bylo potřeba nahradit hodnoty PRAVDA a NEPRAVDA za číselnou variantu 1 a 0. Celý postup je proveden pomocí následujícího zdrojového kódu.

```
1 import pandas as pd
2
3 #Variables
4 dropped_columns = ['Subject', 'Date', 'Sleep Apnea', 'Sleep onset
    latency - norm', 'Wake after sleep onset - norm', 'Awakening > 5
    minutes - norm', 'Sleep efficiency - norm']
5
6 #Reading csv files from path variable
7 df = pd.read_csv("C:\\Data\\dataset.csv", encoding='utf-8', sep=";"
    , header=0, index_col=0, decimal=',')
8 df_rbd = pd.read_csv("C:\\Data\\RBD_SQ.csv", encoding='utf-8', sep=
    ";;", header=0, index_col=0)
9
10 #Converting type of Date column variables to Datetime
11 df['Date'] = pd.to_datetime(df['Date'], format='%Y-%m-%d')
12
13 #Printing all types of columns
14 print(df.dtypes)
15
16 #Sorting Dataframe by Subject column
17 df = df.sort_values('Subject')
18
19 #Reset indexes and deleting patients without RBD-SQ
20 df = df.reset_index(drop=True)
21 df = df.drop(df.loc[0:80].index)
22 df = df.reset_index(drop=True)
23
24 #Adding RBD_SQ and dropping RBD-SQs without patients
25 df = df.merge(right=df_rbd, on='Subject', how='right', copy=False)
26 df = df.drop(df.loc[597:613].index)
27
28 #Dropping irrelevant columns
29 df = df.drop(columns=dropped_columns)
30
31 #Replacing 'PRAVDA' and 'NEPRAVDA' values as 1 and 0
```

```

32 df['Probable Parkinson Disease'] = df['Probable Parkinson Disease']
    .replace(['PRAVDA', 'NEPRAVDA'], [1, 0])
33 df['Probable Mild Cognitive Impairment'] = df['Probable Mild
    Cognitive Impairment'].replace(['PRAVDA', 'NEPRAVDA'], [1, 0])
34
35 #Saving Dataframe back to csv file
36 df.to_csv(path_or_buf="C:\\Data\\dataset_parkinson_actigraphy.csv",
    encoding='utf-8', sep=";", decimal=',')

```

5.3 Datová množina Spánkové deníky a RBD-SQ

K získání této datové množiny byly použity stejné algoritmy jako na datovou množinu Aktigrafie a RBD-SQ (viz subsekcce 5.2.1). Výsledná datová množina obsahuje stejné prvky (sekce 5.2), rozdílné jsou pouze hodnoty, které jsou získány ze spánkových deníků.

5.4 Datová množina Aktigrafie, Spánkové deníky a RBD-SQ

Poslední kombinací je datová množina, která obsahuje vstupy z aktigrafie, spánkových deníků a RBD-SQ.

6 Návrh modelu strojového učení

Tato kapitola se zabývá výběrem a návrhem klasifikačního modelu strojového učení, který má za úkol co nejlépe klasifikovat pacienty s Parkinsonovou chorobou.

6.1 Neuronové sítě

První myšlenkou této diplomové práce bylo vytvoření klasifikační neuronové sítě s hodnotou Probable Parkinson jako výstupní hodnotou y . Po úpravě datové množiny je ve finále hodnot pro učení neuronové sítě málo a výstupní hodnoty jsou nevyvážené (viz tabulka 6.1).

Tab. 6.1: Procentuální výskyt výstupních hodnot y (Probable Parkinson), kdy číslu 1 připadá hodnota PRAVDA a číslu 0 NEPRAVDA.

Výskyt Parkinsona	1	0	Celkem
Počet datových buněk	152	445	597
Procentuální zastoupení [%]	25,46	74,53	-

6.2 Metoda podpůrných vektorů

Metoda podpůrných vektorů (sekce 2.4) naopak umožňuje pracovat s nevyváženou a znatelně menší datovou množinou. K naučení modelu je použita knihovna scikit-learn (viz subsekce 3.3.4), která k řešení SVM nabízí tři jádrové transformace: radiální bázovou funkci, polynomiální a sigmoidní funkci [36].

6.2.1 Přímá aplikace SVM na vytvořené datové množiny

K natrénování modelu jsou k dispozici tři upravené datové množiny (kapitola 5), které se skládají z potenciálně užitečných dat. Úkolem je zjistit, která tato data jsou reálně přínosná a jaký model SVM použít.

Vyzkoušet všechny datové vstupy a jejich kombinace manuálně by bylo časově náročné. Pomocí modulu `itertools`¹, který je součástí jazyka `Python`, lze tento proces automatizovat. Konkrétní implementace je znázorněna ve zdrojovém kódu níže.

```
1 #Columns used for iteration (input values x except for value y ->
   Probable Parkinsons Disease)
2 iter_columns = df.loc[:, df.columns != 'Probable Parkinson Disease']
```

¹Modul `itertools` implementuje iterační stavební bloky, které jsou rychlé a efektivně pracují s pamětí [43].

```

3
4 #Setting up iterator for combinations
5 score_saved = 0
6 min_size = 2
7 max_size = df.shape[1]
8 column_subsets = itertools.chain(*map(lambda x: itertools.
    combinations(iter_columns, x), range(min_size, max_size)))

```

Aktigrafie a RBD-SQ

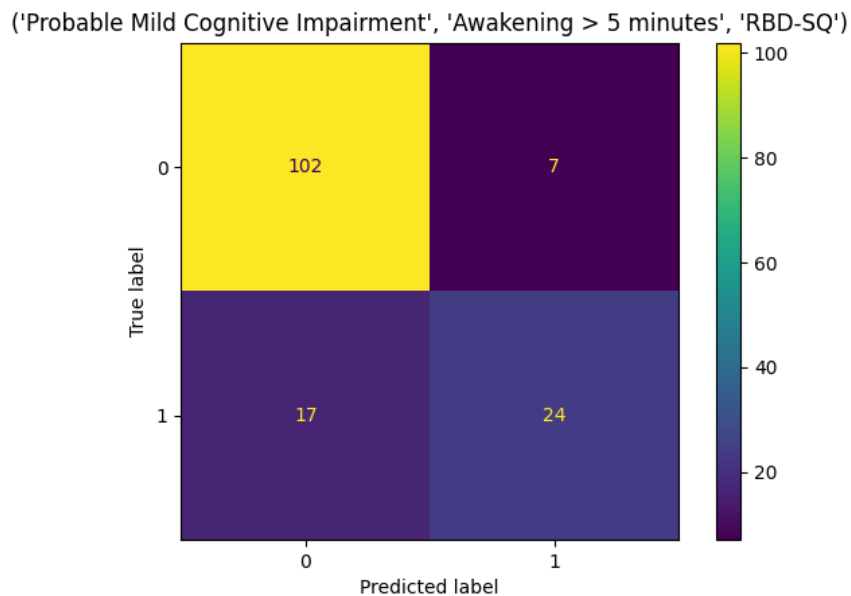
Jako první byla vyzkoušena datová množina Aktigrafie a RBD-SQ. Data byla rozdělena na testovací a trénovací množinu v poměru 1 : 3 s náhodným stavem² číslo 42. Výsledky jsou znázorněny v tabulce níže.

Tab. 6.2: Nejpresnější kombinace vstupních dat z datové množiny Aktigrafie a RBD-SQ, kdy po 19 kombinacích se začaly vyskytovat kombinace bez prvku Probable Mild Cognitive Impairment.

Přesnost	Vstupní data
0,840	Probable Mild Cognitive Impairment, Wake bouts, Awakening >5 minutes, Sleep efficiency, Sleep fragmentation, RBD-SQ
0,833	Probable Mild Cognitive Impairment, Wake after sleep offset, RBD-SQ
0,820	Probable Mild Cognitive Impairment, Sleep efficiency, RBD-SQ
0,806	Probable Mild Cognitive Impairment, Time in bed, Sleep fragmentation, RBD-SQ
⋮	
0,760	Sleep fragmentation, RBD-SQ

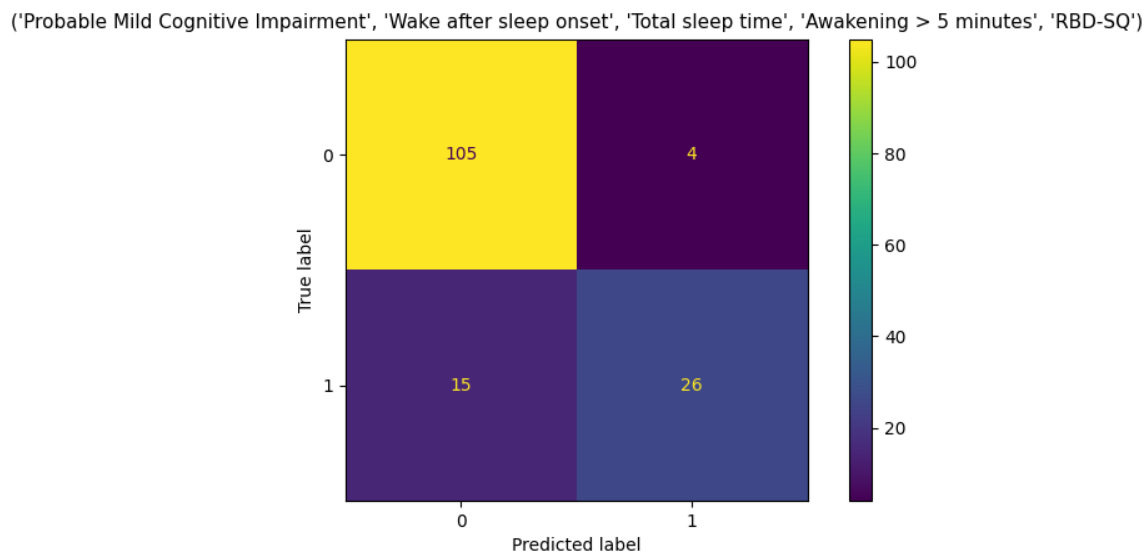
V predikci Parkinsonovy choroby vycházely nejlépe prvky Probable Mild Cognitive Impairment a RBD-SQ. Kromě těchto prvků se vyskytovaly i kombinace zbylých prvků, ty se ale změnou náhodného stavu znatelně lišily, a tudíž nešla jednoznačně určit jejich užitečnost. Přesnost se pohybovala okolo $(84 \pm 2) \%$.

²Náhodný stav (v knihovně `scikit-learn` parametr `random-state`) zajišťuje náhodné míchání dat před rozdělením datové množiny na trénovací a testovací data [36]. Tato náhodnost lze kontrolovat pomocí volitelného parametru, který akceptuje hodnotu `int`.



Obr. 6.1: Matice záměn nejlepšího modelu pro aktigrafii.

Po aplikaci metody scale z knihovny scikit-learn na vstupní data byl natrénován model s přesností 0,86. Celkově se zvýšilo rozpětí přesností při různých náhodných stavech, kdy hodnoty nabývaly $(84 \pm 6) \%$.



Obr. 6.2: Matice záměn nejlepšího modelu pro aktigrafii s aplikací scale metody na vstupní data.

Spánkové deníky a RBD-SQ

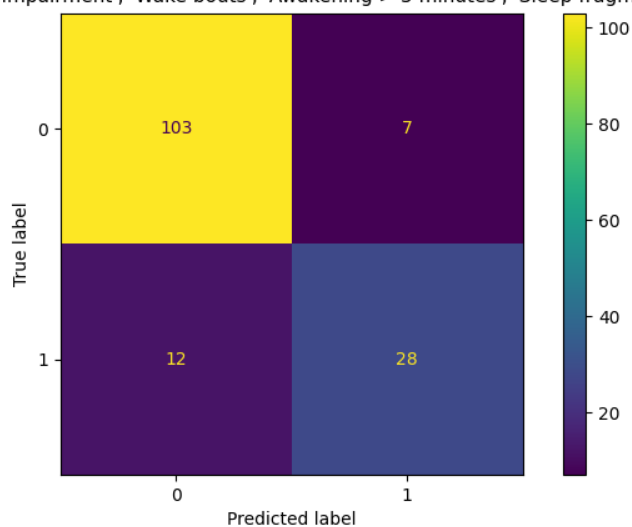
Metoda byla dále vyzkoušena na data ze spánkových deníků, která byla stejně jako u aktigrafie rozdělena na testovací a trénovací množinu v poměru 1 : 3 s náhodným stavem číslo 42.

Tab. 6.3: Nejpresnější kombinace vstupních dat z datové množiny Spánkové deníky a RBD-SQ, kdy po 24 kombinacích se začaly vyskytovat kombinace bez prvku Probable Mild Cognitive Impairment.

Přesnost	Vstupní data
0,873	Probable Mild Cognitive Impairment, Wake bouts, Awakening >5 minutes, Sleep fragmentation, RBD-SQ
0,866	Probable Mild Cognitive Impairment, Wake bouts, Awakening >5 minutes, RBD-SQ
0,846	Probable Mild Cognitive Impairment, Awakening >5 minutes, Sleep fragmentation, RBD-SQ
0,846	Probable Mild Cognitive Impairment, Awakening >5 minutes, RBD-SQ
0,84	Probable Mild Cognitive Impairment, Wake bouts, Sleep fragmentation, RBD-SQ
⋮	
0,793	Wake after sleep offset, RBD-SQ
0,793	Sleep onset latency, Wake bouts, RBD-SQ
0,786	Sleep onset latency, Wake bouts, Sleep fragmentation, RBD-SQ

V případě spánkových deníků se kromě Probable Mild Cognitive Impairment a RBD-SQ jeví jako důležitý prvek Awakening > 5 minutes. Dále se v závislosti na náhodném stavu objevuje prvek Wake bouts a Sleep fragmentation. Přesnost modelu natrénovaném na datech ze spánkových deníků je oproti modelu natrénovaném na datech z aktigrafie obecně lepší (přibližně 3 %).

('Probable Mild Cognitive Impairment', 'Wake bouts', 'Awakening > 5 minutes', 'Sleep fragmentation', 'RBD-SQ')



Obr. 6.3: Matice záměn nejlepšího modelu pro spánkové deníky.

Aktigrafie, Spánkové deníky a RBD-SQ

Jako poslední byl vyzkoušen model naučený na datech z Aktigrafie i spánkových deníků. Pro odlišení byly jednotlivé prvky aktigrafie označeny písmenem x a prvky spánkových deníků písmenem y (v tomto případě se nejedná o označení vstupních a výstupních dat). Data byla rozdělena na testovací a trénovací množinu v poměru 1 : 3 s náhodným stavem číslo 42.

Tab. 6.4: Nejpresnější kombinace vstupních dat z datové množiny Aktigrafie, Spánkové deníky a RBD-SQ, kdy po 106 kombinacích se začaly vyskytovat kombinace bez prvku Probable Mild Cognitive Impairment.

Přesnost	Vstupní data
0,865	Probable Mild Cognitive Impairment_x, Sleep fragmentation_x, Wake bouts_y, RBD-SQ
0,865	Probable Mild Cognitive Impairment_x, Wake bouts_y, Awakening >5 minutes_y, RBD-SQ
0,859	Probable Mild Cognitive Impairment_x, Sleep fragmentation_x, Awakening >5 minutes_y, RBD-SQ
0,859	Probable Mild Cognitive Impairment_x, Sleep fragmentation_x, Wake bouts_y, Sleep fragmentation_y, RBD-SQ
0,859	Probable Mild Cognitive Impairment_x, Wake bouts_y, Awakening >5 minutes_y, Sleep fragmentation_y, RBD-SQ
⋮	
0,778	Sleep onset latency_y, RBD-SQ

Shrnutí

Přímá aplikace SVM na vytvořené datové množiny není pro predikci Parkinsonovy choroby dostatečná a je potřeba stávající datovou množinu podrobit dalším úpravám nebo ji rozšířit o další měření pacientů. Na druhou stranu lze pozorovat i několik pozitivních výsledků. Obecně se SVM jeví jako validní řešení problematiky Parkinsonovy choroby. RBD-SQ je důležitý prvek pro predikci Parkinsonovy choroby, což se shoduje s teorií (viz subsekcce 1.3.7 a 3.1.1).

6.2.2 Aplikace SVM na datovou množinu PD a MCI

Kromě RBD-SQ se pro predikci Parkinsonovy choroby jeví jako důležitý prvek MCI (viz subsekcce 1.3.8). MCI je stejně jako ostatní neurodegenerativní onemocnění diagnostikováno v pozdním věku, tudíž jeho použití není zatím pro predikci reálné. Nese však nezanedbatelnou informační hodnotu, která lze potenciálně využít pro natrénování modelu.

V této subsekcce budou rozebrány výsledky získané aplikací SVM na danou datovou množinu. Prvky datové množiny Probable Parkinson a Probable MCI byly spojeny pomocí logického operátoru AND a použity jako výstupní data y. Tento krok změnil zastoupení jednotlivých tříd, kdy počet datových buněk Výskyt Parkinsona a MCI klasifikovaný pozitivně je o 41 menší a počet negativně klasifikovaných

o toto číslo větší. Procentuálně (viz tab. 6.5) se jedná o velmi nevyváženou datovou množinu, kdy kromě vhodné metody ML bude třeba uvažovat o vyvážení poměru tříd (sběr dalších dat, nadvzorkování/podvzorkování tříd).

Tab. 6.5: Procentuální výskyt výstupních hodnot y (Probable Parkinson & MCI), kdy číslu 1 připadá hodnota PRAVDA a číslu 0 NEPRAVDA.

Výskyt Parkinsona a MCI	1	0	Celkem
Počet datových buněk	111	486	597
Procentuální zastoupení [%]	18,59	81,40	-

Výsledky této kapitoly byly získány pomocí algoritmu zmíněného v subsekcí 6.2.1, který byl aplikován na jednotlivé datové množiny s náhodným stavem 42 až 52. Cílem je zmírnit dopad náhodnosti a získat všeobecně nejpresnější a nejstabilnější model.

Kromě samotné datové množiny PD a MCI, budou vyzkoušeny její varianty upravené pomocí knihovny `imblearn`.

Spánkové deníky a RBD-SQ

Podle výsledků ze subsekcce 6.2.1 vycházejí spánkové deníky jako stabilnější a přesnější volba. Z tohoto důvodu byly vybrány jako první volba na testování upravených datových množin.

V tab. 6.6 jsou zaznamenány nejpresnější kombinace pro datovou množinu PD a MCI. Tyto přesnosti byly získány zprůměrováním přesností jednotlivých náhodných stavů a dosahují hodnoty 85,50 %. Přesnost testování v některých případech dosahuje vyšších hodnot (maximálně 85,81 %) jako přesnost učení, což je následkem přeučení modelu. Při výběru vhodného modelu bude třeba poupravit vstupní parametry SVM tak, aby nedošlo k přetrénování a velkému poklesu přesnosti.

Pro predikci se (kromě RBD-SQ) jeví jako velmi užitečný parametr SOL. Další užitečné parametry jsou WB, $NA > 5$ a SF^3 .

³Parametry WB a $NA > 5$ a SF získané ze spánkových deníků sdílejí určité podobnosti. Parametr WB má v 75% případů totožnou hodnotu jako parametr $NA > 5$ a lze pomocí něj získat parametr SF (viz sekce 1.2).

Tab. 6.6: Nejpresnější kombinace vstupních dat bez vyvažování tříd.

Vstupní data	Přesnost učení	Přesnost testování	Odchylka
SOL, NA>5, SF, RBD-SQ	0.8550	0.8545	0.0366
SOL, WB, NA>5, SF, RBD-SQ	0.8527	0.8581	0.0333
SOL, WB, NA>5, RBD-SQ	0.8521	0.8581	0.0366
SOL, NA>5, RBD-SQ	0.8517	0.8448	0.0266
SOL, WB, SF, RBD-SQ	0.8468	0.8490	0.0300
SOL, WB, RBD-SQ	0.8450	0.8472	0.0266
WB, SF, RBD-SQ	0.8446	0.8303	0.0200

Za zmínku stojí kombinace WASO, SF, RBD-SQ, která má pro všechny zapsané náhodné stavy stejné vstupní parametry SVM modelu. Tato kombinace je proti kombinacím s parametrem SOL méně přesná, ale nabízí se jako stabilní řešení.

Úprava datové množiny pomocí knihovny imblearn

Datová množina PD a MCI je velmi nevyvážená (viz subsekcce 6.2.2). SVM dokáže s takovou množinou pracovat, ale otázkou je, zda pomocí syntetického vyvážení nedosáhne lepších výsledků. Následující výsledky se týkají aplikace SVM na synteticky upravenou datovou množinu pomocí tříd SMOTE a ADASYN z knihovny imblearn. Vstupní parametr náhodný stav byl nastaven na hodnotu 42.

Prvně byly vyzkoušeny obě varianty bez dalších úprav vstupních parametrů. U obou variant dosahovaly přesnosti testování buď nízké hodnoty nebo v porovnání s přesností učení vysoké hodnoty. Celkově se přesnost predikce znatelně snížila.

Dále byla vyzkoušena metoda SMOTE s nastaveným parametrem vzorkovací strategie (`sampling_strategy`) na hodnotě 0,4, kdy zastoupení minoritní třídy bylo zvýšeno na 40 % [37] (viz tab.6.7). Přesnosti učení a testování se u modelů zhoršily a dochází k přetrénování.

Tab. 6.7: Nejpresnější kombinace vstupních dat s vyvážením SMOTE se zastoupením minoritní množiny 40 %.

Vstupní data	Přesnost učení	Přesnost testování	Odchylka
SOL, NA>5, SF, RBD-SQ	0.8094	0.8363	0.0300
WB, NA>5, SF, RBD-SQ	0.8056	0.8115	0.0666
SOL, WB, NA>5, SF, RBD-SQ	0.8029	0.8333	0.0500
SOL, NA>5, RBD-SQ	0.8025	0.8533	0.0400
WASF, SF, RBD-SQ	0.8018	0.8133	0.0533
WB, SE, RBD-SQ	0.7991	0.7848	0.0633
WB, SE, SF, RBD-SQ	0.7985	0.7836	0.0666

Nakonec byla vyzkoušena metoda **SMOTE** s nastaveným parametrem vzorkovací strategie na hodnotě 0,3 (viz tab. 6.8). V tomto případě se zvýšila přesnost testování a snížila přesnost učení. Jako potenciálně použitelný model se jeví kombinace SOL, WB, NA>5, RBD-SQ, která má ve většině případů stejné vstupní parametry SVM.

Tab. 6.8: Nejpresnější kombinace vstupních dat s vyvážením SMOTE se zastoupením minoritní množiny 30 %.

Vstupní data	Přesnost učení	Přesnost testování	Odchylka
SOL, NA>5, SF, RBD-SQ	0.8311	0.8551	0.0333
SOL, NA>5, RBD-SQ	0.8268	0.8430	0.0266
SOL, WB, NA>5, SF, RBD-SQ	0.8261	0.8503	0.0366
SOL, WB, NA>5, RBD-SQ	0.8228	0.8600	0.0366
WASF, SF, RBD-SQ	0.8223	0.8224	0.0300
SOL, WB, SF, RBD-SQ	0.8214	0.8503	0.0233
SOL, SF, RBD-SQ	0.8212	0.8436	0.0366

Aktigrafie a RBD-SQ

Na datovou množinu Aktigrafie a RBD-SQ byly aplikovány stejné algoritmy jako v případě Spánkových deníků a RBD-SQ (viz 6.2.2). Vyvážení tříd pomocí **SMOTE** a **ADASYN** s výchozím vstupem mělo taktéž negativní vliv na přesnosti modelů.

Tab. 6.9: Nejpresnější kombinace vstupních dat bez vyvažování tříd.

Vstupní data	Přesnost učení	Přesnost testování	Odchylka
SE, SF, RBD-SQ	0.8340	0.8299	0.0266
NA>5, SE, SF, RBD-SQ	0.8303	0.8249	0.0233
NA>5, RBD-SQ	0.8259	0.8230	0.0266
SE, RBD-SQ	0.8244	0.8257	0.0400
SF, RBD-SQ	0.8222	0.8282	0.0366
NA>5, SE, RBD-SQ	0.8214	0.8212	0.0300
WB, SE, SF, RBD-SQ	0.8200	0.8227	0.0200

Nižší přesnost mají modely po aplikaci **SMOTE** s nastaveným parametrem vzorkovací strategie na 0,3. První dva modely (viz tab. 6.10) ale nabízejí stabilní řešení, kdy vstupní parametry SVM jsou v 91 % případů totožné.

Tab. 6.10: Nejpresnější kombinace vstupních dat s vyvážením SMOTE se zastoupením minoritní množiny 30 %.

Vstupní data	Přesnost učení	Přesnost testování	Odchylka
WB, NA>5, SE, RBD-SQ	0.8303	0.8208	0.0266
WB, NA>5, SE, SF, RBD-SQ	0.8284	0.8208	0.0300
WB, SE, RBD-SQ	0.8200	0.8153	0.0266
SE, RBD-SQ	0.8181	0.8006	0.0366
SOL, WASO, TST	0.8169	0.7790	0.0233
SOL, WASO, TST, NA>5	0.8169	0.7788	0.0233
SOL, WASO, TST, NA>5, RBD-SQ	0.81697	0.7788	0.0233

Shrnutí

Modely natrénované na datové množině Probable Parkinson & MCI jeví znaky přetrénování a bude třeba upravit vstupní parametry u vybraných modelů (což se projeví na přesnostech). Bylo ale nalezeno několik modelů, které jsou s rozdílnými stavy stabilní.

6.2.3 Výběr modelu

Tato subsekcce bude zaměřena na výběr finálního modelu natrénovaného na Probable Parkinson & MCI. Potenciál jeví modely ze spánkových deníků (viz tab. 6.6). Dále se jedná o jeden model deníků a dva modely aktigrafie s rozšířenou datovou množinou pomocí SMOTE.

První myšlenkou bylo vzít jednotlivé modely a definovat jim mezi náhodnými stavy jeden neměnný vstupní parametr, který bude řešit případné přetrénování s přijatelným snížením přesnosti. Po manuálních testech bylo zjištěno, že nalezení tohoto parametru je nemožné a tudíž bude třeba zvolit jiný přístup.

Většina modelů jeví znaky přetrénování. Po detailním zkoumání bylo zjištěno, že toto přetrénování se vyskytuje u určitých náhodných stavů. Otázkou je, zda lze tohoto faktu využít a vybrat vhodný model.

Spánkové deníky

Pro predikci Parkinsonovy choroby byly v této diplomové práci vybrány spánkové deníky. Nevýhodou je, že si potenciální pacient bude muset zapisovat každé ráno poznámky o spánku (blíže o problematice pojednává subsekcce 1.2.1), ale v případě SVM nabízí oproti aktigrafii znatelně lepší přesnost predikce. Z výsledků ze subsekcce 6.2.2 bylo zjištěno, že modely se učí znatelně lépe na specifickém rozdělení datové

množiny (rozdělení na trénovací a testovací množinu pomocí `train_test_split`). Konkrétně se jedná o rozdělení s náhodným stavem `random_state` 43, 48 a 52⁴.

V tab. 6.14, 6.12 a 6.13 jsou prezentovány nejlepší modely pro zmíněné náhodné stavy. Pro natrénování modelu jsou důležité spánkové parametry SOL, NA>5 a RBD-SQ. Vstupní parametry SVM nejsou úplně jednotné, ale s jistotou lze říci, že parametr `kernel` je pro všechny případy roven `'rbf'`, což je radiální bázová funkce.

Tab. 6.11: Nejpresnější modely pro náhodný stav 43.

	Trénování	Testování	Vstupní parametry
SOL, NA>5, SF, RBD-SQ	0.8814	0.8470	{1, 10,'rbf'}
SOL, WB, NA>5, SF, RBD-SQ	0.8747	0.8400	{50, 1,'rbf'}
SOL, WB, NA>5, RBD-SQ	0.8747	0.8400	{10, 1,'rbf'}
SOL, NA>5, RBD-SQ	0.8747	0.8330	{1, 10,'rbf'}
SOL, WB, SF, RBD-SQ	0.8792	0.8330	{50, 1,'rbf'}
SOL, WB, RBD-SQ	0.8769	0.8330	{10, 1,'rbf'}

Tab. 6.12: Nejpresnější modely pro náhodný stav 48.

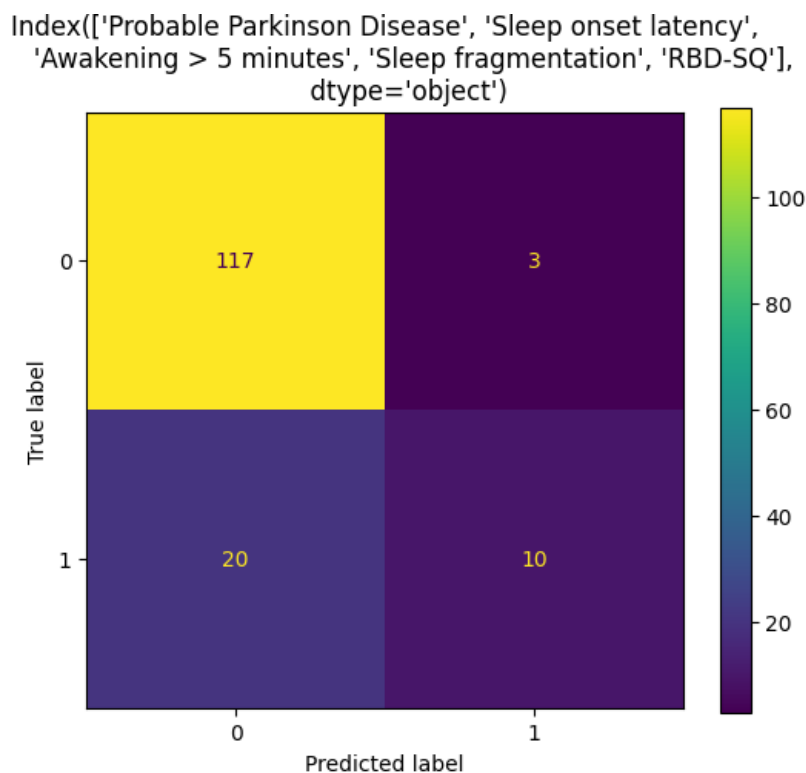
	Trénování	Testování	Vstupní parametry
SOL, NA>5, SF, RBD-SQ	0.8591	0.8466	{1, 10,'rbf'}
SOL, WB, NA>5, SF, RBD-SQ	0.8613	0.8400	{1, 10,'rbf'}
SOL, WB, NA>5, RBD-SQ	0.8613	0.8400	{1, 10,'rbf'}
SOL, NA>5, RBD-SQ	0.8523	0.8600	{1, 10,'rbf'}
SOL, WB, SF, RBD-SQ	0.8501	0.8266	{1, 1,'rbf'}
SOL, WB, RBD-SQ	0.8479	0.8266	{1, 1,'rbf'}

Tab. 6.13: Nejpresnější modely pro náhodný stav 52.

	Trénování	Testování	Vstupní parametry
SOL, NA>5, SF, RBD-SQ	0.8658	0.8133	{100, 1,'rbf'}
SOL, WB, NA>5, SF, RBD-SQ	0.8591	0.8200	{1, 10,'rbf'}
SOL, WB, NA>5, RBD-SQ	0.8591	0.8200	{1, 10,'rbf'}
SOL, NA>5, RBD-SQ	0.8658	0.8200	{10, 1,'rbf'}
SOL, WB, SF, RBD-SQ	0.8456	0.8466	{10, 10,'rbf'}
SOL, WB, RBD-SQ	0.8434	0.8333	{1, 10,'rbf'}

⁴Hodnota `random_state` přijímá datový typ integer, tudíž použitelných hodnot náhodných stavů bude mnohem víc, než je zmíněno v této práci

První model byl vybrán z tab. 6.14. Jedná se o model se spánkovými parametry Sleep onset latency, Wake bouts, Awakening > 5 minutes, Sleep fragmentation, RBD-SQ a vstupními parametry 'C': 1, 'gamma': 10, 'kernel': 'rbf'. Datová množina byla rozdělena podle parametrů 'random_state': 43, 'test_size': 0.25.



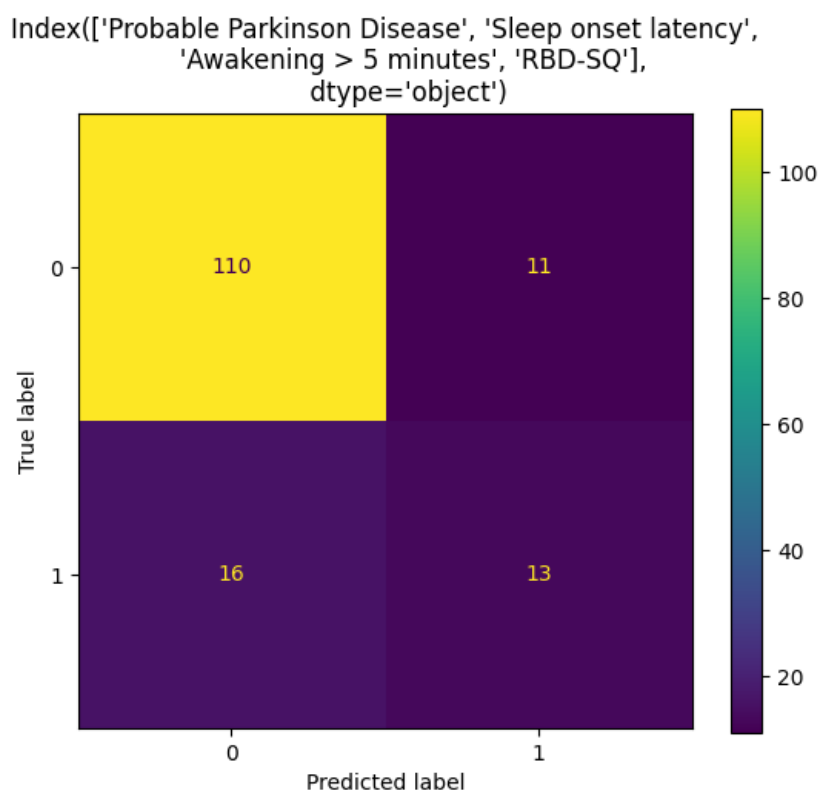
Obr. 6.4: Matice záměn modelu Sleep onset latency, Wake bouts, Awakening > 5 minutes, Sleep fragmentation, RBD-SQ.

Tento model je s přesností 88 % nepřesnějším modelem této diplomové práce. Kromě toho nabízí vysoké procento specificity (99 %), což by se dalo teoreticky využít v případě, kdy by se uvažovala invazivní léčba (například operace). Senzitivita je u trénování 70 %, ale u testování klesla na 33 %, kdy se v oblasti medicíny jedná o nízkou hodnotu.

Tab. 6.14: Výsledky modelu Sleep onset latency, Wake bouts, Awakening > 5 minutes, Sleep fragmentation, RBD-SQ.

	Trénování	Testování
Přesnost [%]	88	85
Senzitivita [%]	70	33
Specifita [%]	99	98
F1 [%]	80	47
MCC	0,79	0,44

Druhý model byl vybrán z tab.6.13. Jedná se o model se spánkovými parametry Sleep onset latency, Awakening > 5 minutes, RBD-SQ a vstupními parametry 'C': 10, 'gamma': 1, 'kernel': 'rbf'. Datová množina byla rozdělena podle parametrů 'random_state': 52, 'test_size': 0.25.



Obr. 6.5: Matice záměn modelu Sleep onset latency, Awakening > 5 minutes, RBD-SQ.

Tento model dosahoval při trénování přesnosti 87 %, senzitivity 70 % a specifity 98 %. Při testování došlo jako u předchozího modelu k poklesu zmíněných hodnot.

Senzitivita je ale v tomto případě větší (45 %). U specifity došlo k poklesu na 91 %.

Tab. 6.15: Výsledky modelu Sleep onset latency, Awakening > 5 minutes, RBD-SQ.

	Trénování	Testování
Přesnost [%]	87	82
Senzitivita [%]	70	45
Specifita [%]	98	91
F1 [%]	78	49
MCC	0,75	0,38

Shrnutí

Navržené modely dosahují vysoké specifity, ale zaostávají v senzitivě. Celkově tato predikční metoda zaostává za ostatními výzkumy (například [8], [6]), ale minimálně se nabízí jako levné řešení, použitelné například pro prevenci. Pro predikci PD byl důležitý parametr RBD-SQ. V nejpřesnějších modelech se vyskytovaly parametry Sleep onset latency, Awakening > 5 minutes, Wake bouts a Sleep fragmentation.

Závěr

Cílem práce bylo z dostupných dat aktigrafie a spánkových deníků vytvořit datovou množinu a s její pomocí navrhnout a natrénovat model strojového učení, který bude schopen na bázi zmíněné datové množiny klasifikovat pacienty trpící Parkinsonovou chorobou.

Teoretická část se v první kapitole zabývá spánkovými parametry a jejich měřením. Následně jsou rozebrány spánkové poruchy a s nimi spojené neurodegenerativní onemocnění. Druhá kapitola je zaměřená na strojové učení. Přesněji se jedná o neuronové sítě a SVM, které se uvažují v praktické části. Nakonec je proveden obecný návrh řešení a popis dostupných datových množin.

Praktická část se ve dvou kapitolách zaměřuje na predikci Parkinsonovy choroby. Pátá kapitola popisuje úpravy provedené v datových množinách, kdy nejprve byl k jednotlivým datovým množinám přiřazen podle názvu subjektů datový sloupec RBD-SQ a MCI. Následně byly odstraněny redundantní spánkové parametry a chybné vstupy. Poslední kapitola pojednává o návrhu modelu strojového učení.

Upravené datové množiny obsahují 597 prvků a třídy jsou v poměru 1 : 3. Jedná se o poměrně malé a nevyvážené datové množiny, proto byl vybrán algoritmus SVM, který dokáže s takovými daty pracovat. Cílem bylo experimentálně zjistit, jaké vstupní kombinace spánkových parametrů a MCI vycházejí z hlediska přesnosti nejlépe. V případě prvního experimentu se jevily nejlépe právě parametry RBD-SQ a MCI, což souhlasí s teorií. MCI se ale stejně jak PD projevuje až v pozdním věku, a tudíž není vhodná jako vstupní parametr pro predikci.

Pro druhý experiment byl v datové množině spojen parametr PD a MCI spojen pomocí logického operátoru AND a výsledek byl použit jako výstupní hodnota. Tento postup využívá informaci, kterou nese parametr MCI a model bude tudíž klasifikovat pravděpodobnější případy Parkinsonovy choroby. Počet pozitivně klasifikovaných případů se po úpravě snížil na 18,6 %, a proto bylo třeba uvážit metody pro syntetickou úpravu dat. V rámci diplomové práce byly vyzkoušeny metody SMOTE a ADASYN z knihovny `imblearn`, které měly ve finále negativní vliv na trénované modely.

Výstupem práce jsou dva modely natrénované na datových množinách PD&&MCI. Výběr probíhal ze spánkových deníků, které byly oproti aktigrafii nezanedbatelně přesnější. Modely dosahují přesností 85 a 82 %, vysoké specificity, ale zaostávají v senzitivitě. Pro predikci se v případě SVM jeví jako užitečné parametry RBD-SQ, Sleep onset latency, Awakening > 5 minutes, Wake bouts a Sleep fragmentation.

Literatura

- [1] K. Crowley. *Sleep and Sleep Disorders in Older Adults* Neuropsychology Review vol. 21. 41–53 [cit. 2022-05-11]. Dostupné z doi: 10.1007/s11065-010-9154-6.
- [2] K. Crowley. *Sleep disorder related to Parkinson's disease* J Neurol vol 244. [cit. 2022-05-11]. Dostupné z doi: 10.1007/BF03160564.
- [3] CHOKROVERTY, Sudhansu, 2010. *Overview of sleep & sleep disorders*. The Indian journal of medical research. 131(2), 126–140.
- [4] REED, David L. a William P. SACCO. *Measuring Sleep Efficiency: What Should the Denominator Be?* Journal of Clinical Sleep Medicine [online]. 2016, 12(02), 263-266 [cit. 2021-10-19]. ISSN 1550-9389. Dostupné z: doi:10.5664/jcsm.5498
- [5] CARNEY, Colleen E., Daniel J. BUYSSE, Sonia ANCOLI-ISRAEL, Jack D. EDINGER, Andrew D. KRYSTAL, Kenneth L. LICHSTEIN a Charles M. MORIN. *The Consensus Sleep Diary: Standardizing Prospective Sleep Self-Monitoring*. Sleep [online]. 2012, 35(2), 287-302 [cit. 2021-10-15]. ISSN 0161-8105. Dostupné z: doi:10.5665/sleep.1642
- [6] MARINO, Miguel, Yi LI, Michael N. RUESCHMAN, et al. *Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography*. Sleep [online]. 2013, 36(11), 1747-1755 [cit. 2021-12-10]. ISSN 0161-8105. Dostupné z: doi:10.5665/sleep.3142
- [7] IBÁÑEZ, Vanessa, Josep SILVA a Omar CAULI. *A survey on sleep assessment methods*. PeerJ [online]. 2018, 6 [cit. 2021-12-10]. ISSN 2167-8359. Dostupné z: doi:10.7717/peerj.4849
- [8] Activinsights [online]. [cit. 26. 10. 2021]. Dostupné z URL: <<https://www.activinsights.com>>.
- [9] SMITH, Michael T., Christina S. MCCRAE, Joseph CHEUNG, Jennifer L. MARTIN, Christopher G. HARROD, Jonathan L. HEALD a Kelly A. CARDEN. *Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline*. Journal of Clinical Sleep Medicine [online]. 2018, 14(07), 1231-1237 [cit. 2021-10-13]. ISSN 1550-9389. Dostupné z: doi:10.5664/jcsm.7230
- [10] BLACKWELL, Terri, Susan REDLINE, Sonia ANCOLI-ISRAEL, Jennifer L. SCHNEIDER, Susan SUROVEC, Nathan L. JOHNSON, Jane A. CAULEY a

- Katie L. STONE. *Comparison of Sleep Parameters from Actigraphy and Polysomnography in Older Women: The SOF Study*. *Sleep* [online]. 2008, 31(2), 283-291 [cit. 2021-10-13]. ISSN 1550-9109. Dostupné z: doi:10.1093/sleep/31.2.283
- [11] SCHENCK, Carlos H. a Mark W. MAHOWALD. *REM Sleep Behavior Disorder: Clinical, Developmental, and Neuroscience Perspectives 16 Years After its Formal Identification in SLEEP*. *Sleep* [online]. 2002, 25(2), 120-138 [cit. 2021-10-17]. ISSN 1550-9109. Dostupné z: doi:10.1093/sleep/25.2.120
- [12] LANE, C. A., J. HARDY a J. M. SCHOTT. *Alzheimer's disease*. *European Journal of Neurology* [online]. 2018, 25(1), 59-70 [cit. 2021-10-28]. ISSN 13515101. Dostupné z: doi:10.1111/ene.13439
- [13] WELLER, Jason a Andrew BUDSON. *Current understanding of Alzheimer's disease diagnosis and treatment*. *F1000Research* [online]. 2018, 7 [cit. 2021-10-28]. ISSN 2046-1402. Dostupné z: doi:10.12688/f1000research.14506.1
- [14] BRZECKA, Anna, Jerzy LESZEK, Ghulam Md ASHRAF, et al. *Sleep Disorders Associated With Alzheimer's Disease: A Perspective*. *Frontiers in Neuroscience* [online]. 2018, 12 [cit. 2021-11-17]. ISSN 1662-453X. Dostupné z: doi:10.3389/fnins.2018.00330
- [15] PETER-DEREX, Laure, Pierre YAMMINE, Hélène BASTUJI a Bernard CROISILE. *Sleep and Alzheimer's disease*. *Sleep Medicine Reviews* [online]. 2015, 19, 29-38 [cit. 2021-12-03]. ISSN 10870792. Dostupné z: doi:10.1016/j.smrv.2014.03.007
- [16] HANSEN, Ingeborg H., Mikkel MARCUSSEN, Julie A. E. CHRISTENSEN, Poul JENNUM a Helge B. D. SORENSEN. *Detection of a sleep disorder predicting Parkinson's disease*. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* [online]. IEEE, 2013, 2013, s. 5793-5796 [cit. 2021-10-17]. ISBN 978-1-4577-0216-7. Dostupné z: doi:10.1109/EMBC.2013.6610868
- [17] BAGLIONI, Chiara, Kai SPIEGELHALDER, Caterina LOMBARDO a Dieter RIEMANN. *Sleep and emotions: A focus on insomnia*. *Sleep Medicine Reviews* [online]. 2010, 14(4), 227-238 [cit. 2021-11-12]. ISSN 10870792. Dostupné z: doi:10.1016/j.smrv.2009.10.007
- [18] NATALE, Vincenzo, Giuseppe PLAZZI a Monica MARTONI. *Actigraphy in the Assessment of Insomnia: A Quantitative Approach*. *Sleep* [online]. 2009, 32(6), 767-771 [cit. 2021-11-12]. ISSN 1550-9109. Dostupné z: doi:10.1093/sleep/32.6.767

- [19] ANCOLI-ISRAEL, Sonia, Donald L. BLIWISSE a Jens Peter NØRGAARD. *The effect of nocturia on sleep*. Sleep Medicine Reviews [online]. 2011, 15(2), 91-97 [cit. 2021-11-13]. ISSN 10870792. Dostupné z: doi:10.1016/j.smr.2010.03.002
- [20] ALLEN, Richard P, Daniel PICCHIETTI, Wayne A HENING, Claudia TRENT-KWALDER, Arthur S WALTERS a Jacques MONTPLAISIR. *Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology*. Sleep Medicine [online]. 2003, 4(2), 101-119 [cit. 2021-11-14]. ISSN 13899457. Dostupné z: doi:10.1016/S1389-9457(03)00010-8
- [21] NATARAJAN, R. *Review of periodic limb movement and restless leg syndrome*. Journal of Postgraduate Medicine [online]. 2010, 56(2) [cit. 2021-11-14]. ISSN 0022-3859. Dostupné z: doi:10.4103/0022-3859.65284
- [22] Lama M. Chahine, Amy W. Amara, Aleksandar Videnovic. *A systematic review of the literature on disorders of sleep and wakefulness in Parkinson's disease from 2005 to 2015* [online]. 2017 [cit. 2021-10-17]. ISBN 978-1-4577-0216-7. Dostupné z: doi:10.1016/j.smr.2016.08.001
- [23] Serge Gauthier, Barry Reisberg, Michael Zaudig, Ronald C Petersen, Karen Ritchie, Karl Broich, Sylvie Belleville, Henry Brodaty, David Bennett, Howard Chertkow, Jeffrey L Cummings, Mony de Leon, Howard Feldman, Mary Ganguli, Harald Hampel, Philip Scheltens, Mary C Tierney, Peter Whitehouse, Bengt Winblad. *Mild cognitive impairment*. The Lancet, Volume 367, Issue 9518 [online]. 2006 [cit. 2021-04-11]. ISSN 0140-6736. Dostupné z: doi:10.1016/S0140-6736(06)68542-5
- [24] B. Guarnieri et al., *Prevalence of Sleep Disturbances in Mild Cognitive Impairment and Dementing Disorders: A Multicenter Italian Clinical Cross-Sectional Study on 431 Patients*. Dementia and Geriatric Cognitive Disorders, vol. 33, no. 1, pp. 50–58. 2012 [cit. 2021-04-11]. Dostupné z: doi:10.1159/000335363.
- [25] Šandala Kristián, Dostálová Simona, Nepožitek Jiří, Ibarburu Lorenzo y Losada Veronika, Dušek Petr, Růžička Evžen, Šonka Karel, Kemlink David. *Actigraphic Screening for Rapid Eye Movement Sleep Behavior Disorder* [online]. 2019 [cit. 2021-11-13]. Dostupné z: doi:10.1101/19001867
- [26] Mohri, M., Rostamizadeh, A. and Talwalkar, A., 2018. *Foundations of machine learning*. MIT press. [cit. 2021-11-17]
- [27] FARRAHI, Vahid, Maisa NIEMELÄ, Maarit KANGAS, Raija KORPELA-NEN a Timo JÄMSÄ. *Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches*. Gait &

- Posture [online]. 2019, 68, 285-299 [cit. 2021-11-27]. ISSN 09666362. Dostupné z: doi:10.1016/j.gaitpost.2018.12.003
- [28] ABIODUN, Oludare Isaac, Aman JANTAN, Abiodun Esther OMOLARA, Kemi Victoria DADA, Nachaat AbdElatif MOHAMED a Humaira AR-SHAD. *State-of-the-art in artificial neural network applications: A survey*. Heliyon [online]. 2018, 4(11) [cit. 2021-11-19]. ISSN 24058440. Dostupné z: doi:10.1016/j.heliyon.2018.e00938
- [29] MCCULLOCH, Warren S. a Walter PITTS. *A logical calculus of the ideas immanent in nervous activity*. The Bulletin of Mathematical Biophysics [online]. 1943, 5(4), 115-133 [cit. 2021-11-19]. ISSN 0007-4985. Dostupné z: doi:10.1007/BF02478259
- [30] Ch. Stergiou, D. Siganos. *Neural networks* [online]. [cit. 2021-11-19]. Dostupné z URL: <<http://srii.sou.edu.ge/neural-networks.pdf>>.
- [31] R. Gandhi. *Support Vector Machine — Introduction to Machine Learning Algorithms* [online]. [cit. 2022-01-03]. Dostupné z URL: <<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>>.
- [32] D. Wilimitis. *The Kernel Trick in Support Vector Classification* [online]. [cit. 2022-02-03]. Dostupné z URL: <<https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>>.
- [33] *Python* [online]. [cit. 2021-11-12]. Dostupné z URL: <<https://www.python.org>>.
- [34] *Tensorflow* [online]. [cit. 2021-26-10]. Dostupné z URL: <<https://www.tensorflow.org>>.
- [35] *Keras* [online]. [cit. 2021-26-10]. Dostupné z URL: <<https://keras.io/about/>>.
- [36] *Scikit-learn* [online]. [cit. 2022-23-2]. Dostupné z URL: <<https://scikit-learn.org>>.
- [37] *Imbalanced-learn* [online]. [cit. 2022-26-4]. Dostupné z URL: <<https://imbalanced-learn.org>>.
- [38] VAH HEES, Vincent, Sarah CHARMAN a Kirstie ANDERSON. *Newcastle polysomnography and accelerometer data* [online]. 2018 [cit. 2021-11-20]. Dostupné z: doi:10.5281/zenodo.1160410

- [39] VAN HEES, Vincent T., Séverine SABIA, Kirstie N. ANDERSON, et al. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. PLOS ONE [online]. 2015, 10(11) [cit. 2021-11-21]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0142533
- [40] VAN HEES, Vincent Theodoor, S. SABIA, S. E. JONES, et al. *Estimating sleep parameters using an accelerometer without sleep diary*. Scientific Reports [online]. 2018, 8(1) [cit. 2021-11-21]. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-018-31266-z
- [41] SIGMUND, Jan. *Automatizovaná diagnóza poruch spánku využívající nositelná zařízení*. Brno, 2020, 52 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Marek Mikulec
- [42] OHAYON, Maurice, Emerson M. WICKWIRE, Max HIRSHKOWITZ, et al. *National Sleep Foundation's sleep quality recommendations: first report*. Sleep Health [online]. 2017, 3(1), 6-19 [cit. 2021-12-11]. ISSN 23527218. Dostupné z: doi:10.1016/j.sleh.2016.11.006
- [43] *Itertools — Functions creating iterators for efficient looping* [online]. [cit. 2022-16-3]. Dostupné z URL: <<https://docs.python.org/3/library/itertools.html>>.

Seznam symbolů a zkratek

ACG	Actigraphy – aktigrafie
ADASYN	Oversample using Adaptive Synthetic – Nadvzorkování pomocí adaptivní syntetiky
ANN	Artificial Neural Network – umělá neuronová síť
CSV	Comma-Separated Values – soubor s hodnotami oddělenými čárkou
DSE	Duration of the Sleep Episode – trvání spánkové epizody
DT	Decision Tree – metoda rozhodovacích stromů
ECSD	Expanded Consensus Sleep Diary – rozšířený spánkový deník
EEG	Elektroencefalografie
EKG	Elektrokardiogram
EMG	Elektromyografie
EOG	Elektrookulografie
GPU	Graphics Processing Unit – grafický procesor
LR	Linear Regression – lineární regrese
MCI	Mild Cognitive Impairment – Mírná kognitivní porucha
ML	Machine Learning – Strojové učení
NA>5	Number of awakenings greater than 5 minutes – počet probuzení delších než 5 minut
PLM	Periodic Limb Movement – periodické pohyby končetin
PSG	Polysomnografie – polysomnografie
RBD	REM Behaviour Disorder – porucha chování v REM spánku
RBD-SQ	RBD Screening Questionnaire – RBD dotazník
REM	Rapid Eye Movement – fáze spánku
RF	Random Forest - metoda náhodný les
RLS	Restless Leg Syndrom – syndrom neklidných nohou

RSWA	REM Sleep Without Atonia – REM spánek bez atonie
SE	Sleep Efficiency – efektivita spánku
SF	Sleep Fragmentation – fragmentace spánku
SMOTE	Synthetic Minority Over-sampling Technique – syntetická nadvzorkovací technika minority
SOL	Sleep Onset Latency – doba, za kterou osoba usnula
SRBD	Sleep-Related Breathing Disorder – porucha dýchání spojená se spánkem
SVM	Support Vector Machine – metoda podpůrných vektorů
TASAFA	Time Attempting to Sleep After Final Awakening – doba, za kterou se osoba pokouší po finálním probuzení spát
TIB	Time In Bed – čas strávený v posteli
TPU	Tensor Processing Unit – tenzorový procesor
TST	Total Sleep Time – celková doba spánku
WASF	Wake After Sleep Offset – doba, za jakou osoba vstala po probuzení
WASO	Wake After Sleep Onset – celková doba bdění po prvním usnutí
WB	Wake Bouts – počet probuzení

A Dataset Aktivografie a RBD-SQ

Tab. A.1: Ukázka datové množiny Aktigrafie a RBD-SQ.

PPD	PMCI	TIB	SOL	WASO	WASF	TST	WB	NA>5	SE	SF	RBD-SQ
0	0	32400,0	150,0	9510,0	0,0	22740,0	24,0	19,0	70,18518519	3,799472296	1,0
0	0	33000,0	0,0	12930,0	0,0	20070,0	24,0	17,0	60,81818182	4,304932735	1,0
0	0	34200,0	0,0	13440,0	480,0	20280,0	28,0	19,0	59,29824561	4,970414201	1,0
0	0	29400,0	420,0	8730,0	0,0	20250,0	23,0	17,0	68,87755102	4,088888889	1,0
0	0	41400,0	0,0	11970,0	0,0	29430,0	26,0	23,0	71,08695652	3,180428135	1,0
0	0	34080,0	90,0	11430,0	0,0	22560,0	31,0	21,0	66,1971831	4,946808511	1,0
0	0	36600,0	0,0	8550,0	0,0	28050,0	20,0	16,0	76,63934426	2,56684492	1,0
0	0	23400,0	330,0	6750,0	0,0	16320,0	14,0	10,0	69,74358974	3,088235294	2,0
0	0	26400,0	0,0	7140,0	1170,0	18090,0	17,0	15,0	68,52272727	3,383084577	2,0
0	0	36900,0	1020,0	17250,0	630,0	18000,0	32,0	20,0	48,7804878	6,4	2,0
0	0	26100,0	360,0	11130,0	300,0	14310,0	15,0	8,0	54,82758621	3,773584906	2,0
0	0	29700,0	0,0	10230,0	2610,0	16860,0	18,0	13,0	56,76767677	3,84341637	2,0
0	0	22800,0	0,0	8430,0	3120,0	11250,0	17,0	11,0	49,34210526	5,44	2,0
0	0	25800,0	0,0	9840,0	210,0	15750,0	23,0	17,0	61,04651163	5,257142857	2,0
0	0	28800,0	180,0	7470,0	30,0	21120,0	20,0	16,0	73,33333333	3,409090909	3,0
0	0	31800,0	0,0	8550,0	420,0	22830,0	21,0	13,0	71,79245283	3,311432326	3,0
0	0	31800,0	480,0	7410,0	270,0	23640,0	18,0	15,0	74,33962264	2,741116751	3,0
0	0	32580,0	270,0	9030,0	270,0	23010,0	26,0	22,0	70,62615101	4,06779661	3,0
0	0	33720,0	180,0	7740,0	120,0	25680,0	19,0	17,0	76,15658363	2,663551402	3,0