

Introduction

Sny prožíváme během fáze spánku, které se říká REM. Lidé s Rapid Eye Movement Sleep Behaviour Disorder, zkratka RBD). Normálně je tělo během snů nehýbné, ale pro pacienty s touto nemocí to neplatí, mohou sebou kopat, házet a křičet, či být náměsíční. Také je u nich velké riziko rozvinutí Parkinsonovi nemoci (PD). [CITACE]

Záměrem této práce je porovnat pacienty s PD a RBD oproti zdravým kontrolám (HC) na základě řečových parametrů a dotazníkového šetření a zjistit, zda a podle čeho je možné tyto skupiny od sebe odlišit.

Methods

Hlavní cíle této práce jsou: Zjistit zda je rozdíl v motorice mezi skupinami PD a RBD na základě dotazníku UPDRS III. Zjistit zda lze skupiny rozlišit na základě řečových parametrů. A pokud ano, tak prozkoumat na základě kterých a zjistit, které jsou důležité.

Dataset obsahuje data od 130 pacientů. Z toho 30 (9 žen, 21 mužů) je v raném stádiu neléčeného PD, 50 (9 žen, 41 mužů) trpí RBD a 50 (8 žen, 42 mužů) jsou zdravé kontroly. Z experimentu jsme se rozhodli vyřadit věk, protože je podobný napříč skupinami, data o medikaci, která je nasazena u příliš malého počtu pacientů a klinické informace o začátku a době trvání nemoci. Použili jsme data z dotazníku UPDRS III, který testuje motoriku pacientů a data o sadě 12 řečových parametrů, která byla naměřena během čtení a monologu. Z datasetu byli také odstraněni outlieři.

Pro otestování rozdílů na základě dotazníku UPDRS III byl použit t-test, zhoršení motoriky je zjevným symptomem PD a vede k její diagnóze. Selektce parametrů proběhla pomocí multinomiální logistické regrese s L1 regularizací, ta byla použita i jako klasifikační model. Nejlepší parametr C pro sílu regularizace byl nalezen pomocí křížové validace.

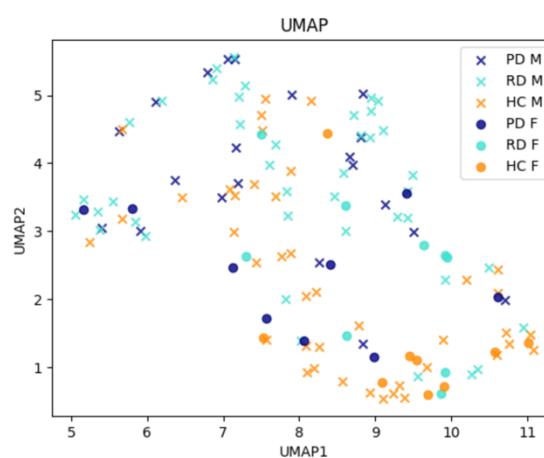
Osm nejvýznamnějších řečových parametrů bylo otestováno pomocí metody ANOVA a rozdílné skupiny byly nalezeny pomocí post-hoc testů.

Results

Je vždy dobré se na data podívat, zde je pro vizualizaci použito PCA a UMAP.



Obrázek 1 PCA



Obrázek 2 UMAP

V obou případech se manifestovaly clustery, i když v případě lineárního PCA je separace o něco horší než u nelineárního UMAPu. Clustery PD a RBD se navzájem překrývají více než s HC.

Pro další vizualizaci lze také využít LDA, které nám oproti PCA dá i informaci o tom, jak dobře lze clustery separovat.

Pro ověření rozdílu mezi motorikou RBD a PD pacientů byl použit t-test. Nejprve byly ověřeny předpoklady t-testu, normalita a homogenita



Obrázek 3 LDA

rozptylů. Normalita byla testována Shapirovým-Wilkovým testem, ten potvrdil normalitu u obou skupin: PD: $p = 0.019$, RBD: $p < 0.001$. Homogenita rozptylů byla testována Bartlettovým testem, ten ji potvrdil: $p < 0.001$.

H_0 – Mezi výsledky z dotazníku pro UPDRS III pro skupiny PD a RBD není významný rozdíl.

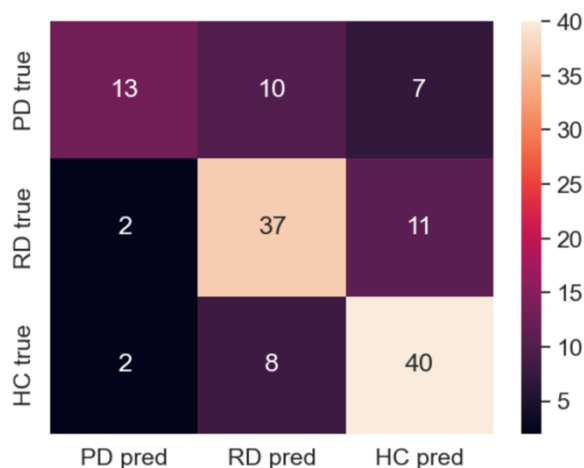
H_1 – Mezi výsledky z dotazníku pro UPDRS III pro skupiny PD a RBD je významný rozdíl.

Mezi PD ($M = 20.4$, $SD = 12.4$) a RBD ($M = 5.2$, $SD = 4.1$) byl nalezen významný rozdíl; $t(78) = 7.94$, $p\text{-value} < 0.001$.

Další částí bylo vybrat na základě kterých parametrů jsou třídy nejlépe separovatelné, pro tento účel jsme využili logistickou regresi s L1 regularizací. Logistická regrese dosáhla přesnosti 70 %.

Duration of pause intervals, monologue	1.4
Rate of speech timing, reading	0.55
Gaping in-between voiced intervals, monologue	0.5
Acceleration of speech timing, monologue	0.4
Rate of speech timing, monologue	0.37
Duration of unvoiced stops, monologue	0.34
Pause intervals per respiration, monologue	0.34
Rate of speech respiration, reading	0.30

Tabulka 1 Váhy z logistické regrese pro 8 nejvýznamnějších řečových parametrů



Obrázek 4 confusion matrix pro logistickou regresi

Pro tyto vybrané parametry jsme ověřili jejich statistickou významnost. Hlavním nástrojem pro testování mezi skupinami byla jednocestná, respektive dvoucestná ANOVA. Nejprve byly ověřeny předpoklady pro ANOVu, normalita a homogenita rozptylů. Normalita byla testována Shapirovým-Wilkovým testem:

		Řečové parametry – čtení												
		'Group'	'EST_R'	'RST_R'	'AST_R'	'DPI_R'	'DVI_R'	'GVI_R'	'DUS_R'	'DUF_R'	'RLR_R'	'PIR_R'	'RSR_R'	'LRE_R'
p-value	PD		0.343	0.404	0.534	0.421	0.042	0.757	0.076	0.300	0.206	0.438	0.308	0.064
	RBD		0.042	0.261	0.043	0.022	0.796	0.949	0.005	0.589	0.458	0.812	0.736	0.120
	HC		0.039	0.627	0.391	0.072	0.004	0.241	0.005	0.479	0.216	0.002	0.186	0.203

		Řečové parametry – monolog												
		'Group'	'EST_M'	'RST_M'	'AST_M'	'DPI_M'	'DVI_M'	'GVI_M'	'DUS_M'	'DUF_M'	'RLR_M'	'PIR_M'	'RSR_M'	'LRE_M'
p-value	PD	0.226	0.295	0.798	0.255	0.102	0.380	0.033	0.283	0.451	0.029	0.783	0.822	
	RBD	0.215	0.083	0.829	0.095	0.156	0.500	0.005	0.677	0.167	0.015	0.670	0.046	
	HC	0.026	0.764	0.070	0.079	0.066	0.882	0.002	0.992	0.696	0.035	0.672	0.072	

Tabulka 2 p-values pro shapiro-wilk test

Homogenita rozptylů byla testována Bartlettovým testem:

		Řečové parametry – čtení											
		'EST_R'	'RST_R'	'AST_R'	'DPI_R'	'DVI_R'	'GVI_R'	'DUS_R'	'DUF_R'	'RLR_R'	'PIR_R'	'RSR_R'	'LRE_R'
p-value		0.121	0.936	0.208	0.077	0.054	0.631	0.641	0.950	0.292	0.680	0.243	0.903

	Řečové parametry – monolog											
	'EST_M'	'RST_M'	'AST_M'	'DPI_M'	'DVI_M'	'GVI_M'	'DUS_M'	'DUF_M'	'RLR_M'	'PIR_M'	'RSR_M'	'LRE_M'
p-value	0.123	0.707	0.063	0.024	0.345	0.794	0.568	0.344	0.225	0.882	0.883	0.547

Tabulka 3 p-values pro Bartlettův test

Při hladině statistické významnosti 0.05 můžeme vyřadit parametry DUS a PIR, ostatní parametry lze přibližně považovat za normálně rozdělené s homogenními rozptyly.

U jednotlivých subjektů byl přítomen i údaj o pohlaví, proto byla použita dvoucestná ANOVA pro odhalení možného vlivu pohlaví na výsledky. Korekce hladiny statistické významnosti byla realizována Bonferonniho metodou, původní hodnota byla 0.05, nová hodnota je 0.0083. Z tabulky vyplývá, že pohlaví má vliv na hodnoty řečových parametrů a deformuje možný vliv nemoci. Při dalším testování budou vzorky dále separovány podle pohlaví. Obě pohlaví nejsou ve vzorku zastoupeny rovnoměrně. Žen je v každé skupině jen 9, mužů pak 21, 41 a 41 ve skupinách PD, RBD a HC.

		Monologue				Reading	
		RST	AST	DPI	GVI	RST	RSR
Groups	F-statistic	5.49	0.16	5.63	0.22	2.4	0.11
	d.f.	2	2	2	2	2	2
	p-value	0.005	0.851	0.005	0.800	0.095	0.893
Gender	F-statistic	8.21	0.25	8.93	14.05	14,78	0
	d.f.	1	1	1	1	1	1
	p-value	0.005	0.620	0.003	0.000	0.000	0.961
Groups*gender	F-statistic	0.25	3.8	0.09	0.24	1.54	0.63
	d.f.	2	2	2	2	2	2
	p-value	0.776	0.025	0.918	0.788	0.219	0.533

Pro testování rozdílů mezi skupinami žen byla použita jednocestná ANOVA. Korekce hladiny statistické významnosti byla realizována Bonferonniho metodou, původní hodnota byla 0.05, nová hodnota je 0.0083. Z výsledků vyplývá, že se nepodařilo prokázat rozdíl mezi skupinami žen. Výsledek může být dán malým vzorkem, nebo nemoci PD a RBD postihují řečové parametry u žen méně, než u mužů.

	Monologue				Reading	
	RST	AST	DPI	GVI	RST	RSR
F-statistic	2.46	1.94	2.1	0.02	0.11	0.58
d.f.	2	2	2	2	2	2
p-value	0.107	0.165	0.145	0.981	0.895	0.568

Pro skupinu mužů byl proveden stejný test, který odhalil rozdíl mezi skupinami u parametrů RST (v obou měřeních) a DPI. Pro ověření, mezi kterými skupinami je rozdíl byl proveden Tukeyho-Kramerův post-hoc test.

	Monologue				Reading	
	RST	AST	DPI	GVI	RST	RSR
F-statistic	7.83	2.62	7.52	1.02	8.05	0.6
d.f.	2	2	2	2	2	2
p-value	0.001	0.078	0.001	0.365	0.001	0.553

Discussion

Testy ukazují, že v případě měření monologu je u parametrů RST a DPI statisticky významný rozdíl mezi dvojicí skupin PD a RBD a skupinou HC. Z toho vyplývá, že u těchto parametrů dochází při onemocnění RBD ke zhoršení srovnatelnému s onemocněním PD. U měření čtení je u parametru PIR nepatrný překryv konfidenčních intervalů, přesto je efekt skupin viditelný. To ukazuje možnost předpovídat možné rozvinutí PD ještě před zřetelnými motorickými symptomy. Taková predikce by se mohla udělat jednoduchým testem řeči a vypočítáním uvedených parametrů. Jejich odchylka od zdravé populace by mohla být záminka pro dlouhodobé sledování parametrů a jejich případných změn, až k nasazení medikace.

Conclusion