



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSVÍ

Katedra biomedicínské techniky

Klasifikace EEG založená na Riemannově metrice

EEG classification based on Riemannian manifold

Diplomová práce

Studijní program: Biomedicínská a klinická technika

Studijní obor: Biomedicínský inženýr

Vedoucí práce: Ing. Marek Piorecký

Bc. Barbora Balcarová

Kladno 2020

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Balcarová** Jméno: **Barbora** Osobní číslo: **456615**
Fakulta: **Fakulta biomedicínského inženýrství**
Garantující katedra: **Katedra biomedicínské techniky**
Studijní program: **Biomedicínská a klinická technika**
Studijní obor: **Biomedicínský inženýr**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Klasifikace EEG založená na Riemannově metrice

Název diplomové práce anglicky:

EEG classification based on Riemannian manifold

Pokyny pro vypracování:

Předzpracujte spánkové EEG záznamy (aplikujte frekvenční filtry, proveďte vhodnou montage, potlačte oční a svalové artefakty). EEG data popište pomocí příznaků. Příznakový prostor klasifikujte pomocí vybraného učícího a neučícího se klasifikátoru. Pro oba klasifikátory porovnejte různé druhy metrik a jejich vliv na výslednou klasifikaci EEG signálu.

Seznam doporučené literatury:

- [1] P. Van Hese ; W. Philips ; J. De Koninck ; R. Van de Walle ; I. Lemahieu, Automatic detection of sleep stages using the EEG, IEEE Engineering in Medicine and Biology Society, ročník 23, číslo 1, 2002
[2] Vinay K. Ingle, John G. Proakis, Digital signal processing using MATLAB, ed. Third edition, CENGAGE Learning, 2012, ISBN 978-1-111-42737-5

Jméno a příjmení vedoucí(ho) diplomové práce:


Ing. Marek Piorecký

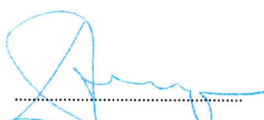
Jméno a příjmení konzultanta(ky) diplomové práce:

Ing. Jan Štrobl, PhDr. Jana Kopřivová, Ph.D.

Datum zadání diplomové práce: **10.02.2020**

Platnost zadání diplomové práce: **19.09.2021**


prof. Ing. Peter Kneppo, DrSc., dr.h.c.
podpis vedoucí(ho) katedry


prof. MUDr. Ivan Dylevský, DrSc.
podpis děkana(ky)

Prohlášení

Prohlašuji, že jsem diplomovou práci s názvem „Klasifikace EEG založená na Riemannově metrice“ vypracovala samostatně a použila k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k diplomové práci.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 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.

V Kladně dne 21. 5. 2020

.....

Bc. Barbora Balcarová

Poděkování

Děkuji svému vedoucímu Ing. Markovi Pioreckému za vedení práce, cenné rady a trpělivost. Celému BRAIN Teamu děkuji za podporu a podnětné připomínky. Dále děkuji lékařům z Národního ústavu duševního zdraví za poskytnutí velkého množství záznamů včetně jejich klasifikace. Také děkuji svým rodičům za všeobecnou podporu během celého studia i vypracování diplomové práce a mému bráchovi za tu pohodlnou židli, co mi na psaní práce půjčoval.

Abstrakt

Klasifikace EEG založená na Riemannově metrice

Práce se zabývá automatickou klasifikací spánkových fází v elektroencefalografických (EEG) záznamech. Cílem práce bylo porovnat vlivy použité metriky na výsledky klasifikace jednoho učícího a jednoho neučícího klasifikátoru se zaměřením na metriku Riemannovu. V programovém prostředí MATLAB[®] jsem předzpracovala a segmentovala spánkové EEG záznamy 20 zdravých subjektů, na extrahované příznaky jsem aplikovala klasifikátory k-NN a k-means, v nichž byly implementovány porovnávané metriky: Riemannova, Euklidova, Minkowského a Mahalanobisova. Klasifikaci jsem vyhodnotila pomocí statistických charakteristik: správnosti, senzitivity a specificity a pro porovnání metrik jsem provedla Wilcoxonův test. Výsledky naznačují, že použitá metrika má pro různé klasifikátory různý vliv na klasifikaci. Pro algoritmus k-NN byla nejvhodnější Mahalanobisova metrika, zatímco Riemannova metrika byla nejvhodnější pro algoritmus k-means.

Klíčová slova

Klasifikace EEG, Riemannova metrika, spánkové EEG, k-NN, k-means.

Abstract

EEG classification based on Riemannian manifold

The thesis deals with automated classification of sleep stages in electroencephalographic (EEG) recordings. The objective of the thesis was to compare effects of used metric on the classification results of one supervised and one unsupervised classifier with focus on Riemannian metric. EEG recordings of 20 healthy subjects was preprocessed and segmented in programming environment MATLAB[®], then the k-NN and k-means classifiers with Riemannian, Euclidean, Minkowski and Mahalanobis metric implemented were applied on extracted features. The classification was evaluated by statistical characteristics like accuracy, sensitivity and specificity and Wilcoxon test was used for comparing the metrics. The results indicate that the metric used has different impact on classification for different classifiers. Mahalanobis metric was the most suitable for k-NN classifier while Riemannian metric was the most suitable for k-means.

Key words

EEG classification, Riemannian manifold, sleep EEG, k-NN, k-means.

Obsah

Seznam použitých symbolů a zkratk	9
1 Úvod	10
2 Přehled současného stavu	12
2.1 Fáze spánku	12
2.2 Automatická klasifikace	13
2.3 Metriky	14
2.3.1 Riemannova metrika	15
3 Cíle práce	17
4 Metody	18
4.1 Data	18
4.2 Program MATLAB	20
4.3 Předzpracování a segmentace	21
4.4 Použité příznaky	22
4.5 Implementace klasifikačních algoritmů	23
4.5.1 Euklidova, Mahalanobisova a Minkowského metrika	24
4.5.2 Riemannova metrika	25
4.5.3 Algoritmus k-NN	27
4.5.4 Algoritmus k-means	29
4.6 Vyhodnocení klasifikace	31
5 Výsledky	32
5.1 Ověření algoritmu k-NN na testovacích datech	32
5.2 Výsledky klasifikace algoritmem k-NN	32
5.3 Statistické porovnání metrik pro k-NN klasifikátor	35
5.4 Ověření algoritmu k-means na testovacích datech	35
5.5 Výsledky klasifikace algoritmem k-means	37
5.6 Statistické porovnání metrik pro k-means klasifikátor	40
6 Diskuze	42
7 Závěr	46
Seznam použité literatury	47
Příloha A: Výsledky klasifikace testovaných postupů	51

Příloha B: Seznam skriptů a funkcí	52
Příloha C: Obsah přiloženého CD	53

Seznam použitých symbolů a zkratek

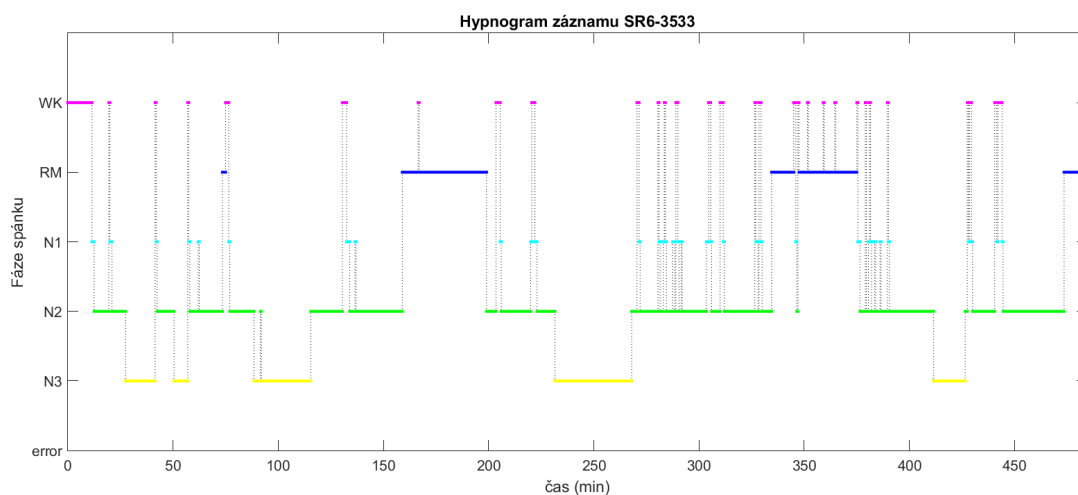
Zkratka	Význam
EEG	Elektroencefalogram
EOG	Elektrookulogram
EMG	Elektromyogram
EKG	Elektrokardiogram
R&K	Rechtschaffen and Kales
AASM	American Academy of Sleep Medicine
REM nebo RM	Rapid eye movement
NREM	Non-rapid eye movement
N1, N2, N3	Fáze NREM spánku
WK	Fáze bdění
SVM	Support Vector Machines
k-NN	K Nearest Neighbours
BCI	Brain-Computer Interface
NUDZ	Národní ústav duševního zdraví
FIR	Finite impulse response
ASD	Amplitudová spektrální hustota

1 Úvod

Spánek je neodmyslitelnou součástí života, přibližně jednu třetinu života každý člověk prospí. Spánek je stav fyziologického snížení bdělosti, základní obnovovací proces důležitý pro duševní i fyzické zdraví [1, 2]. Na tělesném cyklu dne a noci se významně podílí hormon melatonin produkovaný endokrinní žlázou šišinkou (epifýzou), která reaguje na světlo a tmu [3]. Zvýšená hladina melatoninu během noci nejen způsobuje ospalost, ale také napomáhá regeneraci těla [3].

Spánek je charakterizován specifickými změnami v celém organismu jako např. snížením dechové frekvence, krevního tlaku a tělesné teploty, dále uvolněním svalstva, snížením celkové citlivosti na podněty a v neposlední řadě charakteristickými změnami mozkové činnosti [1].

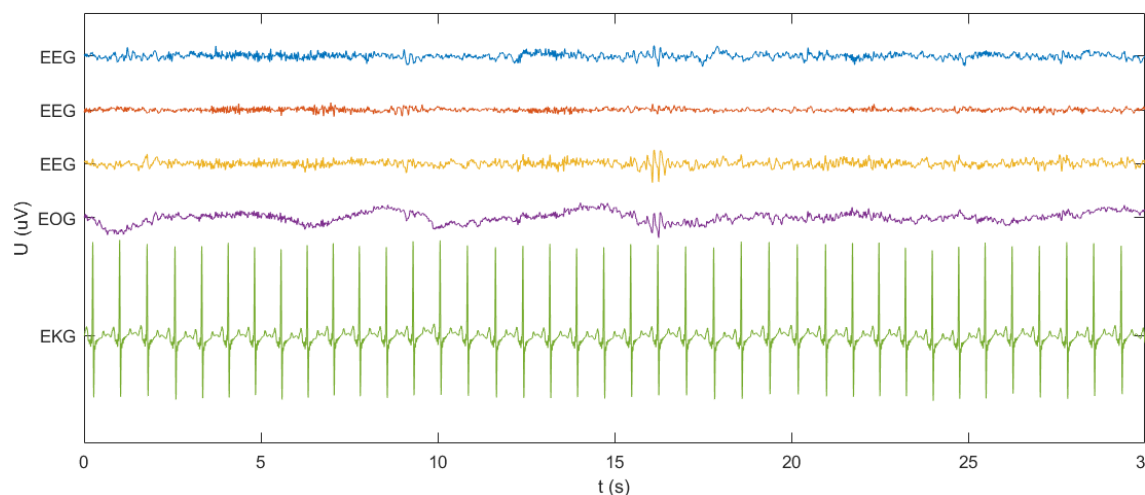
Přestože je pocitově vnímán jako jednolitý, spánek není po celou dobu svého trvání stejný. Skládá se z několika od sebe odlišných fází, které se během noci objevují v opakujících se cyklech [4]. Příklad takového střídání spánkových fází je uveden na obrázku 1.1 na tzv. hypnogramu. Sledování spánku, konkrétní skladby a průběhu jednotlivých fází je nezbytné nejen pro porozumění mechanismům spánku, ale i diagnostiku spánkových poruch [2, 4].



Obrázek 1.1: Hypnogram spánkového záznamu vytvořený v programu MATLAB.

Pro monitorování a hodnocení spánku se velmi často používá tzv. polysomnografie, tj. současný záznam několika fyziologických signálů během spánku [2]. Mezi nahrávané signály patří elektroencefalogram (EEG), elektrookulogram (EOG), elektromyogram (EMG), elektrokardiogram (EKG) apod., používá se také záznam dechu nebo i videozáznam spícího pacienta [2]. Příklad současného záznamu více signálů během spánku je uveden na obrázku 1.2.

EEG, tedy snímání elektrických potenciálů způsobených aktivitou mozku z povrchu lebky [5], je nejdůležitějším signálem pro klasifikaci jednotlivých fází spánku. Pro určení fází spánku je ovšem výhodné sledovat ho v kombinaci s dalšími signály jako EOG (snímání elektrických potenciálů vyvolaných pohybem očí) a EMG (snímání elektrické aktivity svalů, zde konkrétně svalů na bradě). Tyto signály jsou stejně jako EEG součástí standardizovaných manuálů pro určování fází spánku. [4, 6]



Obrázek 1.2: Příklad jednoho segmentu spánkového záznamu několika signálů.

Spánkový záznam je dlouhý typicky okolo 8 hodin, při zavedeném vizuálním hodnocení se záznam rozdělí na 30 sekund dlouhá okna (segmenty), které pak jeden po druhém lékař hodnotí, aby určil příslušné fáze spánku a měl nakonec celkovou představu o pacientově spánku. U takto dlouhých záznamů je výhodné použít digitální analýzu signálu, která může významně zjednodušit a urychlit práci lékaři. Aby i lékaři měli více času na spánek.

Digitální zpracování a vyhodnocování signálu je komplexní proces sestávající se z mnoha článků, jako předzpracování, rozdělení na segmenty a klasifikace různými automatickými nebo poloautomatickými metodami. Tyto články je navíc často nutno přizpůsobovat konkrétním parametrům signálu nebo požadovaným výstupům. Klasifikace spánkových EEG záznamů je např. specifická tím, že jejím prvním cílem obvykle nebývá přímo hledání typických EEG aktivit nebo patologických grafoelementů v signálu, ale určení spánkové fáze celých poměrně dlouhých segmentů.

Snaha o využití automatické klasifikace není ničím novým. Metody klasifikace spánkových záznamů mají ale podle studií, které budou rozebrány dále, stále relativně nízkou spolehlivost nebo omezené využití. Přestože mohou už teď lékaři značně usnadnit práci a celý proces klasifikace urychlit, stále je zde velký prostor pro zlepšování a hledání vhodnějších metod.

2 Přehled současného stavu

2.1 Fáze spánku

První pokusy o určení odlišných mozkových aktivit během spánku se objevily již velmi brzy po vyvinutí EEG. Postupně byly vytvářeny různé modifikace prvotních schémat a objeveny další fáze. To vyústilo v roce 1968, kdy Rechtschaffen a Kales (R&K) publikovali první mezinárodně uznávaný manuál vytvořený na základě konsenzu odborníků. Manuál definuje celkem 5 fází a pro klasifikaci spánku se využíval dlouhá léta. [4]

Teprve v roce 2007 publikovala American Academy of Sleep Medicine (AASM) manuál nový, který z původních R&K zásad vychází, ale reviduje je a přidává také další pravidla pro hodnocení událostí během spánku (tj. ne pouze pro rozlišení spánkových fází) [4]. Jelikož byl R&K manuál velmi zavedený, lze nalézt i studie publikované po roce 2007, ve kterých je klasifikace spánku provedena ještě podle něj, např. [6] nebo [7].

AASM manuál popisuje ve spánku fáze REM (rapid eye movement - rychlé pohyby očí) a NREM (non-rapid eye movement), NREM spánek se potom dělí na další tři fáze označované jako N1, N2 a N3. K těmto celkem čtyřem fázím spánku musíme ještě přidat stav bdělosti, obvykle označovaný jako WK (z anglického wakefulness). [4]

Zestručněné charakteristiky jednotlivých spánkových fází dle AASM manuálu jsou uvedeny v tabulce 2.1. V tabulce jsou zmíněny elementy EEG jako spánková vřeténka a K-komplexy. Spánkové vřeténko je úsek signálu s frekvencí 12–14 Hz trvající alespoň 0,5 sekundy, obálka signálu má obvykle tvar vřetena. K-komplex je bifazická vlna s velkou amplitudou, jedná se o zápornou výchylku následovanou kladnou výchylkou, trvající alespoň 0,5 sekundy a lze ji dobře odlišit od aktivity na pozadí. [4, 6]

Jednou z nejvýraznějších změn v nové klasifikaci je, že hluboký spánek (přibližně fáze N3) byl podle R&K manuálu rozdělen na fáze dvě, podle procentuálního výskytu pomalých vln. To bylo v AASM klasifikaci potlačeno, jelikož se ukázalo, že toto dělení nemá biologické podklady. Je také vhodné zmínit, že mozková aktivita se u malých dětí vyvíjí a charakteristiky jednotlivých fází spánku se pro ně liší. [4]

Tabulka 2.1: Charakteristiky fází spánku dle AASM manuálu [4].

Fáze	EEG	EOG	EMG
WK	alfa aktivita (8 – 13 Hz) při zavřených očích	specifické pohyby	vysoký tonus
REM	nízká amplituda, směs frekvencí	rychlé pohyby	nízký tonus
N1	nízká amplituda, směs frekvencí (dominuje 4 – 7 Hz)	pomalé pohyby	nízký tonus
N2	K-komplexy a spánkové vřeténka, směs frekvencí na pozadí	–	–
N3	pomalé vlny (< 2 Hz) s velkou amplitudou, spánková vřeténka	–	–

2.2 Automatická klasifikace a klasifikační algoritmy

Automatickou klasifikací spánkových fází se v nedávné době zabývala řada studií, které využily různé metody klasifikace a široké spektrum klasifikačních algoritmů. Např. ve studii [6] použili pro klasifikaci třívrstvou neuronovou síť a v [8] byla použita modifikace klasifikátoru SVM (metoda podpůrných vektorů). V další studii [2] se zabývali větším spektrem klasifikátorů, konkrétně: k-NN, SVM, rozhodovací strom, Bayesovský klasifikátor a Random forests, nejlepších výsledků dosáhli s posledně jmenovaným. Článek [2] uvádí i další studie, které většinou využily metodu SVM.

Dosavadní metody automatické klasifikace mají podle uvedených studií stále relativně nízkou spolehlivost, zvláště pokud je použit pouze EEG záznam a je snaha rozlišit všechny spánkové fáze podle AASM manuálu [2, 4, 6].

Algoritmy používané pro automatickou klasifikaci se dají rozdělit do dvou základních typů: klasifikátory využívající učení s učitelem (také označované jako učící se) a klasifikátory využívající učení bez učitele (neučící se). Neučící se algoritmy nepotřebují předem žádnou informaci o klasifikovaných datech a hledají jejich přirozenou strukturu. Učící se algoritmy pro svoji funkci naopak potřebují předem klasifikovaná data (tzv. trénovací množinu), na jejichž základě pak provádějí klasifikaci neznámých dat (tzv. testovací množiny). [5]

Mezi neučící se metody patří např. shluková analýza, konkrétně metoda k-means (k-středů). Metoda funguje tak, že se nejprve nějakým způsobem určí k středů v prostoru a data potom přiřadí do shluku, jehož střed je nejbližší. U vzniklých shluků spočte jejich středy a znovu zkontroluje u všech dat, který střed je nejbližší. Takto postupuje dále a zároveň kontroluje, jak se změnila pozice středů oproti předchozím, když se pozice přestanou měnit, algoritmus se ukončí. [5]

Omezením metody je, že může fungovat pouze pro offline klasifikaci, tj. teprve když máme k dispozici všechna data, nemusí být také vždy zřejmý optimální počet shluků, který chceme v datech hledat [5]. Metoda také nedokáže odhalit různě překrývající se shluky. A po provedeném shlukování je ještě třeba nějakou metodou (zpravidla ručně) přiřadit shlukům jejich význam např. třídu, kterou reprezentují. Její nespornou výhodou je ovšem snadná aplikace na data, o nichž toho předem moc nevíme.

Jednoduchým příkladem učícího se klasifikátoru je algoritmus k -NN (k nearest neighbours), který pro každý klasifikovaný objekt hledá k nejbližších sousedů z trénovací množiny. Třídu klasifikovaného objektu potom určí podle převažující třídy nalezených nejbližších sousedů. Tento algoritmus umožňuje online klasifikaci. [5]

Vliv na dobrou funkčnost k -NN klasifikátoru může mít počet hledaných nejbližších sousedů, tedy parametr k , přesto v mnoha studiích ani není uvedeno, jakou jeho hodnotu použili. Pro výběr vhodného parametru jsem našla následující studie týkající se zpracování EEG. Ve studii [9] věnující se hodnocení příznaků pro automatickou klasifikaci spánku např. použili různé hodnoty k v rozsahu 5 – 26 pro různé sady příznaků. Ve studii [7] použili rovněž pro klasifikaci spánkových fází algoritmus k -NN a nejlepších výsledků dosáhli s hodnotou $k = 5$. Ve studii [10] věnující se rozlišení zdravých a dyslektických dětí pomocí EEG testovali hodnoty k v rozmezí 1 až 13, nejlepších výsledků dosáhli s hodnotou 5. Také ve studii [11] pozorovali nejlepší výsledky s hodnotou 5, když testovali hodnoty od 2 do 10 pro k -NN klasifikátor použitý pro odlišení epileptické aktivity.

2.3 Metriky

Výše uvedené metody klasifikace jsou příznakově orientované, tzn. že klasifikovaný segment signálu je popsán parametry, které ho charakterizují v časové, frekvenční nebo prostorové oblasti a na základě těchto příznaků má své místo v tzv. příznakovém prostoru. Algoritmus pak určuje jeho podobnost nebo odlišnost od ostatních segmentů podle vzájemné vzdálenosti v příznakovém prostoru. Tuto vzdálenost lze určit pomocí vícero různých metrik a způsob jejího určení může mít na funkci algoritmu vliv. [5]

Metrik existuje nepřeberné množství, nejpřirozenější a také nejpoužívanější je Euklidova vzdálenost. Další používanou metrikou je Mahalanobisova, která do svého výpočtu zahrnuje kovarianční matici. Další metriky jsou např. bloková, čtvercová,

Minkowského či Manningova vzdálenost. Většina z nich je již implementována v programovém prostředí MATLAB[®]. [5, 12]

2.3.1 Riemannova metrika

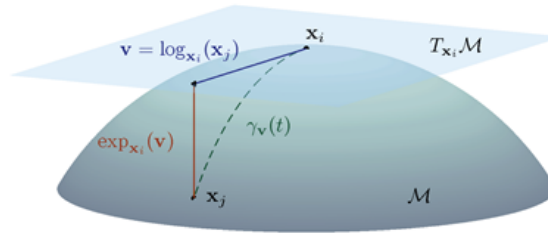
Mezi méně známé patří Riemannova metrika, ta počítá vzdálenost objektů v příznakovém prostoru podobně, jako kdyby byly umístěny na kulovité ploše, což ovšem velmi připomíná měření EEG na povrchu hlavy. Tato spojitost vybízí k použití právě Riemannovy metriky pro zpracování EEG záznamu. V poslední době se objevují různé úspěchy s aplikací Riemannovy metriky při zpracování EEG [7, 13], navíc data umožňuje zpracovávat ve formě matic, díky čemuž jsme schopni zachytit v nich prostorovou souvislost mezi jednotlivými měřeními kanály.

Riemannova metrika byla použita např. ve studii [14] pro segmentaci obrazu nebo v nedávných studiích [13] a [15] pro multikanálové odstraňování artefaktů a další práci s EEG signálem. V případech [13] a [14] byla použita mj. v kombinaci s algoritmem k-means. Také ve studii [16] byl metodou založenou na této metrice zpracováván EEG signál a to pro aplikaci v BCI (brain-computer interface, což je rozhraní pro komunikaci mezi lidským mozkem a počítačem).

Dále studie [7] se zabývala přímo klasifikací spánkových fází, použili zde multikanálová data popsaná pomocí spektrální výkonové hustoty v různých frekvenčních pásmech. Ke klasifikaci potom použili algoritmus k-NN s využitím Riemannovy vzdálenosti.

Riemannova metrika je vzdálenost v Riemannově prostoru, který je definován jako hladká varieta (anglicky differentiable manifold) s danou metrikou. Varieta (hladká varieta) je topologický prostor, který je lokálně podobný obecně n-rozměrnému Euklidovskému prostoru a jsou na něm definovány tečné vektory [17]. V Riemannově prostoru lze měřit vzdálenosti bodů a úhly tečných vektorů. Jak bylo uvedeno výše, Riemannův prostor lze ilustrovat jako kulovitou plochu. To je znázorněno i na obrázku 2.1.

Riemannova vzdálenost mezi body x_i a x_j , které leží na varietě, je definována jako nejmenší délka všech možných hladkých křivek na varietě mezi body x_i a x_j . Hladká křivka s nejkratší délkou se nazývá geodetika. Geodetika je tedy zobecněná přímka představující nejkratší cestu mezi dvěma body na povrchu (respektive na Riemannově varietě). [14]



Obrázek 2.1: Ilustrace Riemannova prostoru, kde M je 2D varieta, $T_{x_i}M$ je tečná rovina v bodě x_i , $\gamma_v(t)$ je geodetika mezi body x_i a x_j . Převzato z [14].

Tečný vektor v z bodu x_i do x_j je definován jako $v = \log_{x_i}(x_j)$. Exponenciální mapa (způsob převodu z tečného prostoru zpět na varietu) zobrazí v do bodu x_j vztahem $x_j = \exp_{x_i}(\log_{x_i}(x_j))$, z toho vychází Riemannova vzdálenost, která je pak dána vztahem [14]:

$$\delta(x_i, x_j) = \|\log_{x_i}(x_j)\|_{x_i}. \quad (1)$$

Ekvivalentní forma vyjádření délky geodetiky, tj. nejkratší vzdálenosti mezi body, může být dána pomocí zobecněných vlastních čísel. Tato druhá forma je vhodnější pro práci v maticově zaměřeném programovém prostředí MATLAB[®] a její implementace je detailně popsána v metodách (kapitola 4.5.2). [17]

3 Cíle práce

Hlavním cílem této diplomové práce je ověřit vhodnost použití principu Riemannovy metriky pro automatickou klasifikaci fází spánku ve spánkových EEG záznamech, zejména její vliv na přesnost automatické klasifikace.

Prvním dílčím cílem je provedení kompletního předzpracování surového EEG záznamu v programovém prostředí MATLAB[®] s využitím toolboxu FieldTrip. Toto předzpracování zahrnuje mimo jiné potlačení artefaktů, provedení vhodné montáže a segmentaci signálu. Bezprostředně potom je cílem provést extrakci vhodných příznaků popisujících jednotlivé segmenty signálu.

Dalším dílčím cílem je samotná klasifikace segmentů signálu s pomocí dvou vybraných klasifikačních algoritmů, konkrétně k-NN a k-means, při využití principu Riemannovy metriky a dalších metrik pro srovnání (Euklidova, Mahalanobisova, Minkowského).

Klasifikaci chci opět provést v programovém prostředí MATLAB[®], kde pro výše zmíněné algoritmy sice již existují funkce, které dokonce umožňují výběr z několika metrik, není mezi nimi ovšem metrika Riemannova, která celkově vyžaduje mírně odlišný přístup k datům. Z toho důvodu je nutné tyto algoritmy s vloženou Riemannovou metrikou znovu implementovat.

Posledním dílčím cílem je statistické vyhodnocení výsledků klasifikace spánkových fází v závislosti na použité metrice a z toho vyplývající zhodnocení, zda je použití Riemannovy metriky a s tím související přístup k datům pro klasifikaci spánkových EEG záznamů výhodnější než klasičtější metriky.

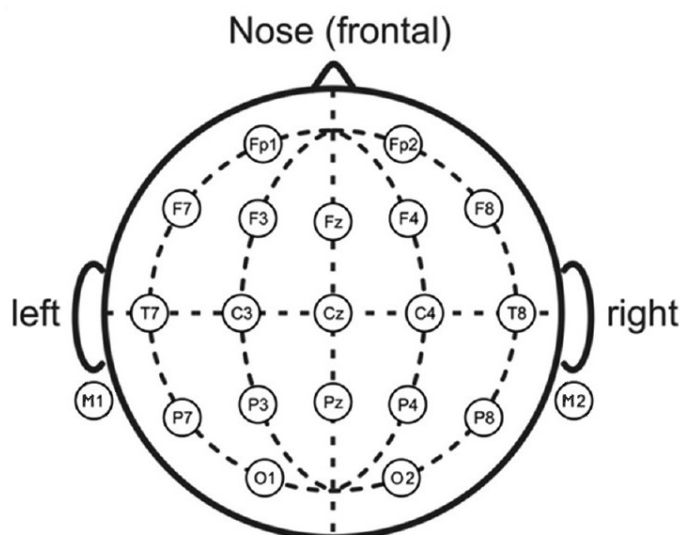
4 Metody

4.1 Data

Do této práce byly použity reálné noční spánkové záznamy naměřené ve spolupráci s Národním ústavem duševního zdraví (NUDZ). Měření bylo schváleno etickou komisí NUDZ. Sběr dat probíhal od roku 2016 v rámci studia spánkových poruch, mnou použitá data jsou měření kontrolní skupiny zdravých subjektů.

Byla použita data od celkem 20 subjektů, jednalo se o muže a ženy ve věkovém rozmezí 24 – 52 let. Spánkové záznamy byly dlouhé 7 až 8,5 hodiny, od každého subjektu bylo tak k dalšímu zpracování k dispozici přibližně 900 – 1000 třicetisekundových segmentů signálu.

Ze spánkových záznamů jsem použila pouze EEG signál (přibližně u poloviny subjektů byl k dispozici pouze EEG, u zbytku i další polysomnografické signály). Měření EEG bylo provedeno přístrojem Brainscope se vzorkovací frekvencí 1 kHz, záznam byl z důvodu velikosti datových souborů a zrychlení automatického zpracování následně podvzorkován na frekvenci 250 Hz. Bylo použito 19 EEG elektrod (+ 2 referenční) rozmístěných podle mezinárodního systému 10-20, byla použita montáž s referenčními elektrodami M1 a M2 (umístěnými na mastoideu), která je doporučovaná AASM manuálem a k vizuálnímu hodnocení spánku se standardně používá. Rozmístění elektrod je znázorněno na obrázku 4.1.



Obrázek 4.1: Rozmístění elektrod podle mezinárodního systému 10-20. Převzato z [18] a upraveno.

V případě, kdy bylo kvůli lepšímu fungování algoritmů s Riemannovou metrikou potřeba vybrat pouze data z menšího množství kanálů, jsem použila EEG kanály doporučené AASM manuálem: F4, C4 a O2 [4].

Ke každému segmentu signálu jsem pro účely ověření automatického určení spánkové fáze měla k dispozici jeho vizuální klasifikaci zkušených lékařů z NUDZ (každý záznam ohodnotili pro kontrolu dva).

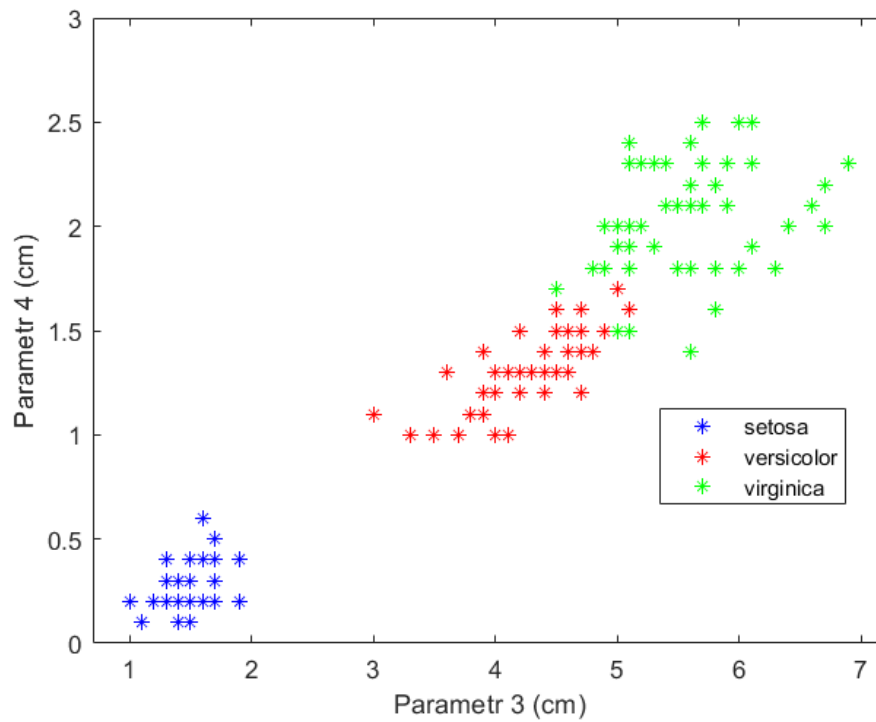
Po provedeném kompletním předzpracování a segmentaci (popsané v kapitole 4.3) záznamy obsahovaly pro každou spánkovou fázi určitý počet segmentů, který se lišil jak pro jednotlivé fáze, tak pro jednotlivé subjekty. Souhrn těchto počtů je uveden v tabulce 4.1. Tabulka uvádí průměrný, minimální a maximální počet segmentů u jednoho subjektu a celkový počet segmentů u všech 20 subjektů.

Tabulka 4.1: Počty segmentů náležejících k jednotlivým spánkovým fázím obsažených v EEG záznamech jednotlivých subjektů.

	Průměrně	Minimálně	Maximálně	Celkem
WK	61	15	204	1219
RM	188	124	259	3751
N1	28	10	47	561
N2	381	252	494	7621
N3	213	138	314	4269

Pro testování funkčnosti vytvořených algoritmů jsem použila testovací sadu dat obsažených přímo v programovém prostředí MATLAB[®]. Jednalo se o tzv. Fisher's iris data set, což je soubor dat z roku 1936 obsahující naměřené délky a šířky dvou druhů okvětních lístků u tří druhů kosatečů [19]. Tato data nelze přirovnávat přímo k segmentům signálu, ale již vytvářejí 4D příznakový prostor se třemi shluky odpovídajícími jednotlivým druhům, přičemž při zobrazení do dvou nejvhodnějších dimenzí je jeden shluk poměrně dobře izolovaný a další dva se mírně překrývají, jak je vidět i na obrázku 4.2. Soubor dat obsahuje i vektor s odpovídajícím druhem kosatce ke každé čtveřici příznaků.

Tato data mi svou podstatou umožnila vytvořit z nich matice příznaků (ne pouze vektory). Mohla jsem jako jeden řádek použít délku a šířku jednoho druhu lístků a druhý druh lístků jako druhý řádek. Takto jsem měla každý objekt v prostoru popsany maticí o rozměrech 2x2. Vytvořené algoritmy jsem tak mohla otestovat na jednodušším prostoru s méně dimenzemi než má klasifikovaný EEG prostor. Testovací data nebylo potřeba normovat, jelikož všechny parametry jsou stejné povahy a ve stejných jednotkách (délky v centimetrech).



Obrázek 4.2: Zobrazení testovacích dat ve 2D prostoru. Barevně jsou odlišeny jednotlivé třídy.

4.2 Program MATLAB[®]

Zpracování spánkových záznamů, implementaci klasifikačních algoritmů a vyhodnocení klasifikace jsem provedla v programu MATLAB[®] společnosti MathWorks [20]. Používala jsem jeho verzi R2019b (Campus-Wide License). Pro předzpracování a segmentaci signálu jsem použila MATLAB[®] open source toolbox FieldTrip [21].

4.3 Předzpracování a segmentace

Nejprve jsem ze spánkového záznamu každého subjektu vybrala pouze EEG kanály (ostatní signály nebyly využity). Kanály jsem seřadila tak, aby byly ve všech záznamech shodně uspořádané (to je důležité později, kdy se při použití Riemannovy metriky s daty pracuje v maticích).

Hlavním krokem předzpracování byla aplikace frekvenčních filtrů. K tomu jsem s výhodou použila funkci obsaženou ve výše zmíněném toolboxu FieldTrip s názvem `ft_preprocessing`. V rámci této funkce jsem aplikovala filtry typu horní a dolní propust s mezními frekvencemi 0,5 a 30 Hz. Dolní mezní frekvence 0,5 Hz pomáhá potlačit stejnosměrnou složku signálu a zároveň nezasahuje do frekvenčního pásma delta, které standardně začíná právě na 0,5 Hz [5]. Horní mezní frekvence 30 Hz zase spolehlivě potlačuje vysokofrekvenční artefakty a je těsně nad horní hranicí frekvenčního pásma beta (resp. sigma) [5, 22], přestože jeho horní mez (i pojmenování) není úplně ustálená a v některých studiích se může lišit, např. [2, 6].

Filtry jsem použila typu FIR (finite impulse response), jelikož tento typ má výhodnou charakteristiku díky, které nezkrsluje tvar signálu, použité filtry byly řádu 1500. Dále jsem použila nastavení funkce `demean` a `detrend`, které pomáhá potlačit průměrnou složku signálu a ustálit plovoucí elektrody. Jako poslední nastavený parametr jsem v rámci `ft_preprocessing` místo původně unipolárně snímaných kanálů použila montáž s referenčními elektrodami M1 a M2, výsledkem této montáže bylo celkem 19 signálových kanálů pro další použití.

Po frekvenční filtraci jsem provedla segmentaci signálu opět s využitím toolboxu FieldTrip. Podle standardního postupu v oblasti klasifikace spánku jsem signál rozdělila na segmenty o fixní délce 30 sekund.

Nakonec jsem na záznam aplikovala ještě prahový filtr využívající násobek mezikvartilového rozpětí signálu pro odstranění odlehlých hodnot napříč elektrodami. Detekoval hodnoty o N -násobek mezikvartilového rozpětí vyšší než horní kvartil a o N -násobek mezikvartilového rozpětí nižší než dolní kvartil. Na základě pozorování jsem zvolila hodnotu $N = 9$. Vzhledem k dostatečnému množství dat jsem segmenty obsahující detekované odlehlé hodnoty pro další zpracování vyřadila celé.

4.4 Použité příznaky

Předzpracovaný signál z každého segmentu a kanálu jsem popsala celkem 24 příznaky, jejich kompletní seznam včetně krátkého popisu je uveden v tabulce 4.2.

Tabulka 4.2: Příznaky použité pro popis segmentů.

Číslo	Zkratka	Jednotka	Popis
1	SIGM	μV	směrodatná odchylka hodnot amplitudy
2	APOS	μV	maximální pozitivní hodnota v segmentu
3	ANEG	μV	minimální negativní hodnota v segmentu
4	Delt1	$\mu\text{V}/\sqrt{\text{Hz}}$	amplitudová spektrální hustota (ASD) v pásmu 0,5 – 1,9 Hz
5	Delt2	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 2,0 – 3,9 Hz
6	Thet1	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 4,0 – 5,9 Hz
7	Thet2	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 6,0 – 7,9 Hz
8	Alph1	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 8,0 – 10,4 Hz
9	Alph2	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 10,5 – 12,9 Hz
10	Sigma	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 18,0 – 29,0 Hz
11	Beta	$\mu\text{V}/\sqrt{\text{Hz}}$	ASD v pásmu 13,0 – 17,9 Hz
12	MAX1D	$\mu\text{V}/\text{s}$	maximální absolutní hodnota první derivace
13	MAX2D	$\mu\text{V}/\text{s}^2$	maximální absolutní hodnota druhé derivace
14	mf	Hz	střední frekvence
15	MD1	$\mu\text{V}/\text{s}$	průměrná hodnota první derivace
16	MD2	$\mu\text{V}/\text{s}^2$	průměrná hodnota druhé derivace
17	mob	s^{-1}	mobilita (Hjorthův parametr)
18	comp	—	komplexita (Hjorthův parametr)
19	act	μV^2	aktivita (Hjorthův parametr)
20	LOfC	μV	délka křivky
21	NLinE	μV^2	nelineární energie
22	ZC	—	počet průchodů nulou
23	Peaks	Hz	dominantní spektrální vrchol
24	Infle	—	počet inflexních bodů

Jednalo se konkrétně o soubor příznaků používaný v programu Wavefinder [23]. Jednotlivé příznaky ze souboru se v různých kombinacích často využívají ve studiích věnujících se zpracování EEG, při klasifikaci spánkových fází jsou oblíbené zejména příznaky ve spektrální oblasti [2, 6, 24]. Pro výpočet příznaků z předzpracovaných dat jsem v programu MATLAB[®] implementovala vlastní funkci vytvořenou na základě obecných poznatků a literatury, především [5].

Z důvodu časové náročnosti a problematiky konvergence implementovaných algoritmů v kombinaci s Riemannovou metrikou při použití všech příznaků bylo potřeba z vypočtených příznaků vybrat pouze malou část. Při výběru jsem vycházela

z významu příznaků pro popis jednotlivých spánkových fází a rozdílů mezi nimi, zejména z frekvenčních charakteristik uvedených výše v tabulce 2.1, a dále také z poznatků své bakalářské práce [25], kde jsem se mimo jiné zabývala redukcí počtu příznaků na základě korelací mezi nimi. Podle těchto kritérií jsem vybrala 6 příznaků, a to příznaky číslo: 1, 4, 9, 10, 14 a 19 (číslované dle tabulky 4.2).

Jelikož použité příznaky mají různé jednotky a jejich hodnoty se pohybují v různých řádech, bylo potřeba je normalizovat, aby měl každý příznak stejnou váhu při umístění segmentu v příznakovém prostoru. Normalizaci jsem prováděla podle vztahu:

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (2)$$

kde x_n je vypočtená normalizovaná hodnota příznaku, x je jeho původní hodnota, x_{\max} a x_{\min} jsou maximální a minimální hodnota tohoto příznaku v rámci celého spánkového záznamu subjektu. Tím jsem získala výsledný příznakový prostor, kde jsou hodnoty všech příznaků v intervalu $\langle 0, 1 \rangle$.

4.5 Implementace klasifikačních algoritmů

Příznakový prostor jsem klasifikovala pomocí dvou vybraných algoritmů. Jeden algoritmus byl tzv. učící se a druhý neučící se, vybrala jsem algoritmy s jednoduchým principem: k-NN a k-means. Oba klasifikátory jsou k dispozici implementované v programovém prostředí MATLAB[®], ale tyto funkce nenabízí použití Riemannovy metriky. Z toho důvodu jsem potřebovala modifikované klasifikátory implementovat znovu.

Pro oba klasifikační algoritmy jsem vytvořila dvě verze, jedna byla pouze pro Riemannovu metriku (dále značená jako \mathcal{R}), druhá pro metriky ostatní (značená jako \mathcal{E}). K tomu jsem přistoupila hlavně z toho důvodu, že (jak bylo uvedeno výše) Riemannova metrika umožňuje pracovat v maticové formě se všemi kanály v jednom časovém úseku najednou a celkově vyžaduje odlišný přístup k datům než u ostatních použitých metrik.

Klasifikátor k-NN s Euklidovou, Minkowského a Mahalanobisovou metrikou nebylo teoreticky bezpodmínečně nutné ručně implementovat, jelikož tyto možnosti nabízí prostředí MATLAB[®]. Funkce pro klasifikátor k-means už měla o dost menší výběr metrik a z mnou vybraných nabízela pouze Euklidovu. Proto jsem pro všechny metriky sjednotila do mnou vytvořených algoritmů. Měla jsem tak jistotu, že klasifikátory s různými metrikami se budou lišit pouze těmito sledovanými metrikami

a ničím jiným, a také mi to umožnilo mít pro obě verze shodný formát vstupních dat a odlišně s nimi pracovat až v rámci funkce.

Formát vstupních dat do klasifikačních algoritmů vypadal následovně. Využila jsem možnost, kterou nabízí prostředí MATLAB[®], a sice práci s tzv. buňkami. Vytvořila jsem vektor těchto buněk a do každé jsem uložila jeden segment spánkového záznamu ve formě matice, kde řádky reprezentovaly jednotlivé kanály záznamu a sloupce vypočítané příznaky.

Dále jsem ke každému segmentu měla označenou jeho spánkovou fázi na odpovídající pozici v samostatném obyčejném vektoru, ve kterém jsem fáze WK, RM, N1, N2 a N3 (v tomto pořadí) označila číslicemi 1 – 5. To posloužilo jednak jako vstupní data k trénovací množině pro algoritmus k-NN, jednak pro vyhodnocení výsledků klasifikace.

4.5.1 Euklidova, Mahalanobisova a Minkowského metrika

Jak bylo už zmíněno, zahrnuje jsem do klasifikačních algoritmů několik metrik. Euklidovu, Mahalanobisovu a Minkowského metriku nebylo nutné implementovat ručně, protože pro ně v prostředí MATLAB[®] již existují funkce, konkrétně jsou zahrnuty do funkce `pdist2`, která umožňuje vypočítat vzájemné vzdálenosti mezi dvěma sadami dat [12]. Pro použité metriky níže uvádím základní matematické vztahy.

Euklidova metrika je definována následovně. Pokud máme v kartézské souřadné soustavě dva vektory (body n -rozměrném prostoru) A a B , pak jejich vzájemnou Euklidovu vzdálenost spočteme podle vztahu [5]:

$$\delta_{\text{Eu}}(A, B) = \sqrt{\sum_{i=1}^n [A(i) - B(i)]^2} \quad (3)$$

Mahalanobisova vzdálenost dvou vektorů A a B se vypočte podle následujícího vztahu [5]:

$$\delta_{\text{Ma}}(A, B) = \sqrt{(A(i) - B(i))W^{-1}(A(i) - B(i))^T}, \quad (4)$$

kde W je výběrová kovarianční matice. Funkce tuto kovarianční matici vypočte z vektoru A . Při výpočtu vzdálenosti tím pádem hraje roli i rozptyl dat v jednotlivých dimenzích. [12]

Výpočetní vztah Minkowského metriky má zajímavé vlastnosti. Vzájemnou Minkowského vzdálenost výše uvedených vektorů A a B v n -rozměrném prostoru vypočteme jako [12]:

$$\delta_{\text{Mi}}(A, B) = \sqrt[p]{\sum_{i=1}^n |A(i) - B(i)|^p}, \quad (5)$$

kde parametr p je kladná skalární hodnota. Existuje několik speciálních případů zvolené hodnoty p . Pokud zvolíme $p = 1$, dostaneme tzv. blokovou vzdálenost, v případě $p = 2$ dostaneme výše uvedený vztah 3, tedy Euklidovu vzdálenost [12]. Pro implementaci vzdálenosti do klasifikačních algoritmů v této práci jsem zvolila $p = 3$.

4.5.2 Riemannova metrika

Dle [13] a [17] lze pro reprezentaci segmentu použít prostorovou kovarianční matici, která zahrne rozptyl příznaků, korelace a jejich prostorové uspořádání. Přirozený prostor vzniklé symetrické a pozitivně definitní matice je Riemannův [13, 17]. Běžné operace jako výpočty vzdálenosti a průměru pak vyžadují speciální přístup [17].

Odhad prostorové kovarianční matice C jsem spočítala podle vzorce [13]:

$$C = \frac{1}{m-1} X \cdot X^T, \quad (6)$$

kde X je matice příznaků o rozměrech $n \times m$. V mém případě je m počet příznaků a n počet kanálů signálu v jednom časovém úseku. Aby vzniklá kovarianční matice C splňovala výše uvedenou vlastnost, že je pozitivně definitní, musí pro rozměry matice příznaků X platit vztah: $m \geq n$. To jsem dodržela tím, že jsem použila celkem 24 příznaků pro popis signálu z 19 kanálů.

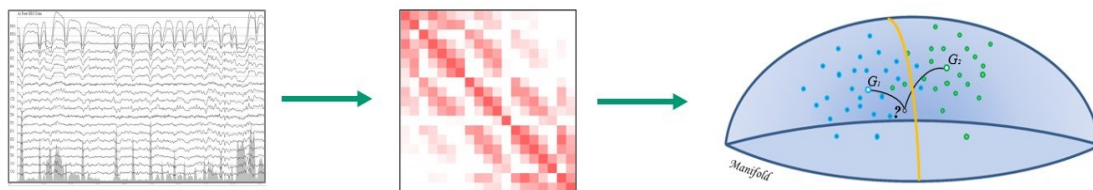
Vzdálenost takovýchto kovariančních matic (např. C_i a C_j) už lze v Riemannově prostoru vypočítat pomocí dříve zmíněných zobecněných vlastních čísel jako [13]:

$$\delta_{\text{Ri}}(C_i, C_j) = \sqrt{\left[\sum_{i=1}^n \ln^2 \lambda_i \right]}, \quad (7)$$

kde λ_i jsou zobecněná vlastní čísla matic C_i a C_j , která obecně odpovídají maticovému vztahu, který lze dále upravit [17]:

$$\begin{aligned}\lambda_i C_i x - C_j x &= 0 \\ C_j x &= \lambda_i C_i x \\ C_i^{-1} C_j x &= \lambda_i x\end{aligned}\tag{8}$$

Rovnice 8 pak koresponduje i s článkem [13], kde je λ_i uvedeno jako vlastní čísla maticového součinu $C_i^{-1} C_j$. Stručné a názorné shrnutí klasifikace s pomocí Riemannovy metriky, které jsem provedla, ilustruje obrázek 4.3.



Obrázek 4.3: Postup klasifikace segmentů signálu v Riemannově prostoru. Ze segmentu se vytvoří kovarianční matice, tou je dáno umístění v prostoru, kde je potom možné hledat hranice mezi shluky. Převzato z [26].

V Riemannově prostoru je kromě specifického přístupu k výpočtu vzdálenosti potřeba podobně postupovat i při výpočtu geometrického středu sady kovariančních matic. Tento výpočet byl využit u algoritmu k-means, kde se geometrické středy shluků počítají v každé iteraci. Geometrický střed M jsem počítala pomocí iterativního algoritmu, kdy se M inicializuje jako aritmetický průměr všech k matic v sadě. Dále ho v každé iteraci počítáme jako [13]:

$$M = M^{1/2} \exp \left[\frac{1}{k} \sum_k \ln(M^{-1/2} C_k M^{-1/2}) \right] M^{1/2}\tag{9}$$

K ukončení iterací posloužil výpočet Frobeniovy normy následujícího výrazu [13]:

$$\left\| \sum_k \ln(M^{-1/2} C_k M^{-1/2}) \right\|_F\tag{10}$$

Hodnota normy konverguje k určité konstantě, pokud byl rozdíl předcházející a aktuální hodnoty normy menší než zvolená prahová hodnota $\varepsilon = 0,01$, byly iterace ukončeny.

4.5.3 Algoritmus k-NN

Pro algoritmus k-NN jsem musela vstupní data nejprve rozdělit na trénovací a testovací množinu. Rozdělení bylo provedeno náhodně pomocí funkce `crossvalind` tak, aby 2/3 segmentů z každé třídy (spánkové fáze) byly zařazeny do trénovací množiny a zbývající 1/3 do testovací množiny. Celou klasifikaci pomocí k-NN jsem pro každý záznam provedla vždy 20x, pokaždé s jiným náhodným výběrem testovací a trénovací množiny a výsledky jednotlivých klasifikací jsem potom zprůměrovala.

Vstupy do vytvořených funkcí `kNN_E` a `kNN_R` tedy jsou:

- vektor buněk obsahující trénovací množinu dat
- vektor obsahující třídy segmentů v trénovací množině
- vektor buněk obsahující testovací množinu dat
- parametr k
- metrika použitá pro výpočet vzdálenosti

Posledně uvedený vstupní parametr obsahuje pouze funkce `kNN_E`, kde je možné si vybrat mezi Euklidovskou, Minkowského a Mahalanobisovou metrikou. Funkce `kNN_R` je určena pouze pro Riemannovu metriku, proto tam tento parametr nemá smysl. Parametr k určující počet nejbližších sousedů, které algoritmus hledá, jsem na základě literatury volila $k = 5$.

Výstupem je vektor obsahující určené třídy segmentů v testovací množině.

Funkce `kNN_E` nejprve každou buňku z trénovací i testovací množiny rozdělí podle kanálů na 19 nezávislých segmentů a odpovídajícím způsobem rozšíří i vektor s třídami segmentů v trénovací množině. U takto upravených dat již funkce postupně vybírá jeden segment testovací množiny po druhém, vždy u něj spočítá podle požadované metriky vzdálenost ke všem segmentům z trénovací množiny a vybere k nejbližších. Výslednou třídu určí podle toho, do které třídy náleží nejvíce z těchto nejbližších sousedů. Pokud by nastala rovnost, segment se přiřadí do třídy, kde je součet vzdáleností nejbližších sousedů nejmenší.

Funkce `kNN_R` musí data z trénovací i testovací množiny také nejprve upravit, z matic v buňkách vytvořit kovarianční matice podle vztahu 6 popsaného v kapitole 4.5.2. Dále už se podle stejného principu jako u `kNN_E` počítá s těmito maticemi. Tj. funkce vezme jednu matici z testovací množiny, spočítá podle vztahu 7 Riemannovu vzdálenost ke všem maticím z trénovací množiny a na základě nejbližších

sousedů určí výslednou třídu, resp. spánkovou fázi, časového úseku. Níže uvádím pseudokód vytvořených funkcí k-NN, který shrnuje jejich princip.

```
Input: příznaky trénovací množiny (data), třídy segmentů trénovací množiny (class), příznaky testovací množiny (Pvec), počet nejbližších sousedů (k), metrika pro výpočet vzdálenosti (distType)  
i = 1;  
for každý segment testovací množiny do  
    P = příznaky segmentu i testovací množiny;  
    spočti vzdálenosti testovacího segmentu od segmentů v trénovací množině dle požadované metriky;  
    NN = vyber k segmentů s nejnižší vzdáleností od P (nejbližší sousedé);  
    classNN = třídy NN segmentů;  
    spočti kolik NN, je z které třídy;  
    if existuje právě jedno maximum počtu NN z jedné třídy then  
        classP = třída, z které je maximální počet segmentů ;  
    else  
        sečti vzdálenosti NN z každé maximální třídy;  
        classP = třída NN, které mají nejnižší součet vzdáleností;  
    end  
    end  
    i = i+1;  
end  
Output: třídy segmentů testovací množiny(classNew)
```

Algoritmus 1: Funkce k-NN

4.5.4 Algoritmus k-means

Pro algoritmus k-means nebylo nutné data podobně jako u k-NN dělit na testovací a trénovací. **Vstupy** pro funkce `kmeans_E` a `kmeans_R` jsou tak pouze:

- vektor buněk se všemi časovými úseky
- parametr k
- metrika použitá pro výpočet vzdálenosti

Podobně jako u k-NN obsahuje posledně uvedený vstupní parametr pouze funkce `kmeans_E`. Hodnotu parametru k určujícího počet vyhledaných shluků jsem volila 5, což je počet spánkových fází, které jsem chtěla rozlišit.

Funkce `kmeans_E` musí data upravit stejně jako `kNN_E` a rozdělit jednotlivé kanály z časových úseků na nezávislé segmenty. Potom funkce nejprve z dat náhodně vybere k segmentů, které použije jako iniciální středy shluků. Následně spočte podle požadované metriky vzdálenost segmentů od středů a každý segment přiřadí do shluku, jehož střed je nejbližší. Dále vypočte nové středy takto vzniklých shluků a opět vzdálenosti segmentů od těchto středů. Celý cyklus se opakuje do doby, dokud se pozice středů shluků nepřestanou měnit. **Výstupy** funkce `kmeans_E` jsou:

- vektor indexů určujících shluk, do kterého byl segment přiřazen
- souřadnice středů shluků
- vektor buněk se všemi segmenty

Funkce `kmeans_R` musí opět soubor dat nejprve upravit a vypočítat podle zmíněného vztahu 6 kovarianční matice. Potom stejně jako verze `kmeans_E` vybere náhodně iniciální středy shluků, ovšem ty jsou v tomto případě již reprezentovány právě kovariančními maticemi. Následně podle vztahu 7 vypočte vzdálenost matic všech časových úseků od středů a každou matici přiřadí do shluku, jehož střed je nejbližší. Nové středy vzniklých shluků se potom vypočítají iterativně a to podle vztahu 9 uvedeného v kapitole 4.5.2 s ukončovací podmínkou vypočtenou podle rovnice 10. Postup výpočtu nových shluků a jejich středů se opět opakuje dokud se nepřestanou měnit. **Výstupy** funkce `kmeans_R` jsou:

- vektor indexů určujících shluk, do kterého byl časový úsek přiřazen
- kovarianční matice odpovídající středům shluků
- kovarianční matice odpovídající časovým úsekům

Funkčnost uvedené implementace jsem ověřila na testovacích datech, pro použití v reálném příznakovém prostoru jsem ovšem z důvodů špatné funkčnosti popsaných v kapitole 5.5 musela přistoupit k následujícím změnám.

Vynechala jsem nestabilní výpočet středů shluků podle rovnice 9 a středy jsem počítala pouze způsobem, kterým se tento výpočet inicializuje tj. průměrem všech matic ve shluku. Dále jsem provedla změnu ve výpočtu 6 prostorové kovarianční matice, čímž se upravilo rozložení dat v prostoru. Inspirovala jsem se váhováním těchto matic pomocí logaritmu ve studii [15]. Upravený výpočet kovariančních matic C potom vypadal takto:

$$C = \frac{1}{m-1} \ln(X) \cdot \ln(X)^T, \quad (11)$$

kde X zůstalo jako matice příznaků o rozměrech $n \times m$.

Pro funkci k-means níže rovněž uvádím pseudokód pro shrnutí jejího principu.

Input: soubor příznaků (*data*), počet center (*k*), metrika pro výpočet vzdálenosti (*distType*)
výběr iniciačních center (náhodný výběr bodu a jeho určení jako centra)
while *aktuální pozice center je jiná než předešlá* **do**
 for $i=1, \dots, N$ **do**
 | dle požadované metriky spočítej vzdálenost každého bodu od
 | každého centra
 end
 přiřaď bod k centru, ke kterému má nejmenší vzdálenost
 for $j=1, \dots, k$ **do**
 | vypočítej centra nových shluků
 end
end
Output: indexy výsledných shluků (*idx*)

Algoritmus 2: Funkce k-means

4.6 Vyhodnocení klasifikace

Vyhodnocení klasifikace spánkových fází v závislosti na použitém algoritmu a metrice jsem provedla pomocí tzv. **klasifikačních matic**. Klasifikační matice zaznamená, jaká fáze byla u segmentu určena automatickým algoritmem vůči tomu, do jaké fáze skutečně patří (určeno lékařem). Shrnuje tedy počty správně klasifikovaných segmentů a kolik segmentů bylo klasifikováno do které nesprávné třídy. U algoritmu k-means výpočtu klasifikačních matic předcházelo ruční přiřazení automaticky vytvořených shluků k nejlépe odpovídajícím spánkovým fázím na základě známé klasifikace segmentů od lékařů. Pro samotný výpočet klasifikační matice jsem si vytvořila vlastní funkci.

Dále jsem kvalitu klasifikace ohodnotila pomocí standardních parametrů vypočtených z klasifikační matice: **senzitivita**, **specifita** a **správnost**. Správnost jsem určila jako podíl správně klasifikovaných segmentů a všech klasifikovaných segmentů v záznamu. Senzitivitu a specifitu bylo (vzhledem ke klasifikaci do více než dvou tříd) potřeba určit pro každou spánkovou fázi zvlášť. Senzitivita pro určitou fázi se v tomto případě definuje jako počet segmentů klasifikovaných jako tato fáze dělený počtem segmentů, které do této fáze skutečně patří. Obdobně se specifita pro určitou fázi definuje jako počet segmentů klasifikovaných jako jiné fáze dělený počtem segmentů, které skutečně patří do jiných fází. Pro výpočet uvedených parametrů jsem si opět vytvořila vlastní funkce.

Abych mohla uvést, zda na klasifikaci má nebo nemá vliv použitá metrika, provedla jsem statistické otestování rozdílů ve výsledcích klasifikace s různými metrikami. Nejprve jsem otestovala normalitu dat, pro to jsem zvolila jednovýběrový Smirnov-Kolmogorův test s nulovou hypotézou, že data pocházejí z normálního rozdělení a alternativní, že z normálního rozdělení nejsou.

Na základě toho, že test na hladině významnosti 5 % zamítl, že by hodnoty zkoumaných parametrů pocházely z normálního rozdělení, jsem pro porovnání metrik zvolila neparametrický Wilcoxonův párový test. Testovala jsem nulovou hypotézu, že hodnoty parametrů pro srovnávané metriky se neliší, proti alternativní jednostranné hypotéze, že rozdíl středních hodnot těchto parametrů je větší než nula neboli že výsledky klasifikace pomocí první metriky jsou lepší než pomocí druhé. Hladina významnosti pro zamítnutí nulové hypotézy byla opět 5 %.

Pro statistické testy jsem využila funkce nabízené programovým prostředím MATLAB[®] a to funkci `kstest` pro Smirnov-Kolmogorův test a `signrank` pro Wilcoxonův test.

5 Výsledky

5.1 Ověření algoritmu k-NN na testovacích datech

Na testovacích datech popsaných v kapitole 4.1 jsem provedla otestování vytvořeného klasifikátoru k-NN s použitím všech porovnávaných metrik. Po provedení 20 různých náhodných výběrů testovací a trénovací množiny a vyhodnocení klasifikace dle parametrů popsaných v kapitole 4.6 jsem získala výsledky uvedené v tabulce 5.1.

Tabulka 5.1: Výsledky ověření k-NN na testovacích datech.

Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		82,7 %	96,4 %	82,8 %	81,6 %
Senzitivita	třída 1	99,1 %	100 %	99,1 %	99,1 %
	třída 2	72,7 %	93,8 %	73,0 %	74,4 %
	třída 3	76,3 %	95,6 %	76,4 %	71,3 %
Specifická	třída 1	100 %	100 %	100 %	100 %
	třída 2	87,9 %	97,8 %	87,9 %	85,2 %
	třída 3	86,2 %	96,9 %	86,3 %	87,2 %

5.2 Výsledky klasifikace algoritmem k-NN

Nejprve zde uvádím výsledek klasifikace EEG spánkového záznamu pomocí algoritmu k-NN s použitím Riemannovy metriky za původně zamýšlených podmínek, což znamená, že každý segment byl popsán 24 příznaky a bylo použito všech 19 kanálů signálu. Zjevně neuspokojivý výsledek je znázorněn pomocí klasifikační matice v tabulce 5.2.

Tabulka 5.2: Výsledek klasifikace algoritmem k-NN s Riemannovou metrikou při použití 19 kanálů a 24 příznaků.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	0	0	0	0	0
	RM	0	0	0	0	0
	N1	0	0	0	0	0
	N2	0,7	0,4	0,1	4,5	0,7
	N3	30,3	40,6	11,9	104,5	59,3

Správnost a senzitivita klasifikace včetně všech ostatních metrik za těchto podmínek jsou potom uvedeny v příloze A v tabulce 1.

Dále už uvádím výsledky klasifikace v redukovaném příznakovém prostoru vybraných 6 příznaků a 3 kanálů. Za těchto podmínek klasifikace s Riemannovou metrikou vykazovala výsledky s podobnou přesností jako ostatní metody, jak je vidět v tabulce 5.3, kde jsou uvedeny kompletní výsledky použité pro porovnání vlivu různých metrik na klasifikaci algoritmem k-NN. Uvedené hodnoty správnosti, senzitivity a selektivity jsou vždy aritmetickým průměrem výsledků klasifikace záznamů všech 20 subjektů.

Tabulka 5.3: Výsledky klasifikace spánkových záznamů algoritmem k-NN s využitím zredukovaného počtu kanálů a příznaků.

Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		72,8 %	69,8 %	72,6 %	77,7 %
Senzitivita	WK	45,7 %	44,8 %	45,5 %	56,4 %
	RM	72,3 %	71,0 %	71,9 %	79,0 %
	N1	7,1 %	3,9 %	7,0 %	9,6 %
	N2	75,8 %	72,6 %	75,7 %	80,0 %
	N3	78,5 %	71,6 %	77,9 %	82,0 %
Specifita	WK	98,2 %	97,7 %	98,2 %	98,3 %
	RM	90,3 %	90,2 %	90,2 %	91,7 %
	N1	98,5 %	98,8 %	98,5 %	98,8 %
	N2	79,7 %	76,5 %	79,4 %	83,7 %
	N3	92,8 %	91,3 %	92,8 %	94,6 %

Pro detailnější představu o výsledku klasifikace uvádím v tabulkách 5.4 – 5.7 průměrné klasifikační matice pro jednotlivé použité metriky. V klasifikačních maticích jsou důležité hodnoty na diagonále, které ukazují počty správně klasifikovaných segmentů, v jednotlivých sloupcích pak lze vidět, jak bylo klasifikováno kolik segmentů náležejících do určité spánkové fáze.

Je zde například vidět, že segmenty náležející ke spánkové fázi N1, u které byla velmi nízká senzitivita, byly nejčastěji klasifikovány jako fáze REM nebo N2. Otestovala jsem některé upravené parametry a postupy klasifikace ve snaze zvýšit senzitivitu u fáze N1, žádný z nich ale nepřinesl zlepšení bez negativních vlivů na klasifikaci ostatních fází. Výsledky testovaných postupů jsou uvedeny v příloze A.

Tabulka 5.4: Klasifikační matice při využití Euklidovy metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	34,2	4,8	3,0	4,8	0,6
	RM	12,3	136,2	9,5	40,5	3,4
	N1	2,4	3,3	2,1	5,2	1,5
	N2	9,3	37,4	10,7	291,1	40,2
	N3	1,6	4,7	1,7	38,5	166,8

Tabulka 5.5: Klasifikační matice při využití Riemannovy metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	11,2	2,2	1,1	2,2	0,4
	RM	4,4	45,3	3,0	13,7	0,9
	N1	0,5	0,9	0,4	1,6	0,4
	N2	3,2	12,6	3,8	93,2	17,3
	N3	0,7	1,2	0,7	16,1	51,7

Tabulka 5.6: Klasifikační matice při využití Minkowského metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	34,1	5,0	3,0	5,0	0,6
	RM	12,3	135,6	9,6	40,8	3,4
	N1	2,4	3,4	2,1	5,3	1,6
	N2	9,4	37,8	10,6	290,9	41,1
	N3	1,6	4,6	1,7	38,2	165,6

Tabulka 5.7: Klasifikační matice při využití Mahalanobisovy metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	40,1	4,6	3,6	4,5	0,6
	RM	9,8	148,6	9,0	35,7	1,9
	N1	2,4	3,0	2,9	4,1	0,9
	N2	6,4	27,4	10,1	306,5	34,3
	N3	1,1	3,0	1,3	29,4	174,8

5.3 Statistické porovnání metrik pro k-NN klasifikátor

Po provedení klasifikace jsem měla k dispozici sady hodnot správnosti, senzitivity a specificity od 20 hodnocených subjektů pro každou použitou metriku. Nejprve jsem otestovala, zda tato data pocházejí z normálního rozdělení. Pro všechny sady hodnot byla nulová hypotéza na hladině významnosti 5 % zamítnuta, hodnoty tedy nepocházely z normálního rozdělení.

Dále jsem tedy provedla Wilcoxonův párový test. V tabulce 5.8 uvádím výsledky testů pro jednotlivé parametry a pro jednotlivé dvojice metrik. V každé dvojici je vždy vlevo uvedena metrika, jejíž hodnoty parametrů byly vyšší (naznačovaly lepší kvalitu klasifikace). Dvojice metrik jsou pro přehlednost v tabulce seřazeny podle postupně klesající hodnoty rozdílu průměrných správností. Hodnota 1 v příslušném poli znamená, že byla na hladině významnosti 5 % zamítnuta nulová hypotéza a přijata alternativní, která říká, že se parametry klasifikace pro tuto dvojici metrik liší. Hodnota 0 znamená, že nulová hypotéza byla přijata.

Tabulka 5.8: Výsledky statistického vyhodnocení vlivu použité metriky pro algoritmus k-NN Wilcoxonovým testem.

Dvojice metrik		Ma – Ri	Ma – Mi	Ma – Eu	Eu – Ri	Mi – Ri	Eu – Mi
Správnost		1	1	1	1	1	1
Senzitivita	WK	1	1	1	0	0	0
	RM	1	1	1	0	0	1
	N1	1	1	1	1	1	0
	N2	1	1	1	1	1	0
	N3	1	1	1	1	1	1
Specificita	WK	1	0	0	0	0	1
	RM	0	1	1	0	0	0
	N1	0	1	1	0	0	0
	N2	1	1	1	1	1	1
	N3	1	1	1	1	1	0

Hodnota 0 znamená přijetí nulové hypotézy, hodnota 1 znamená zamítnutí nulové hypotézy na hladině významnosti 5 %. Porovnávané metriky: Ma – Mahalanobisova, Ri – Riemannova, Mi – Minkowského, Eu – Euklidova.

5.4 Ověření algoritmu k-means na testovacích datech

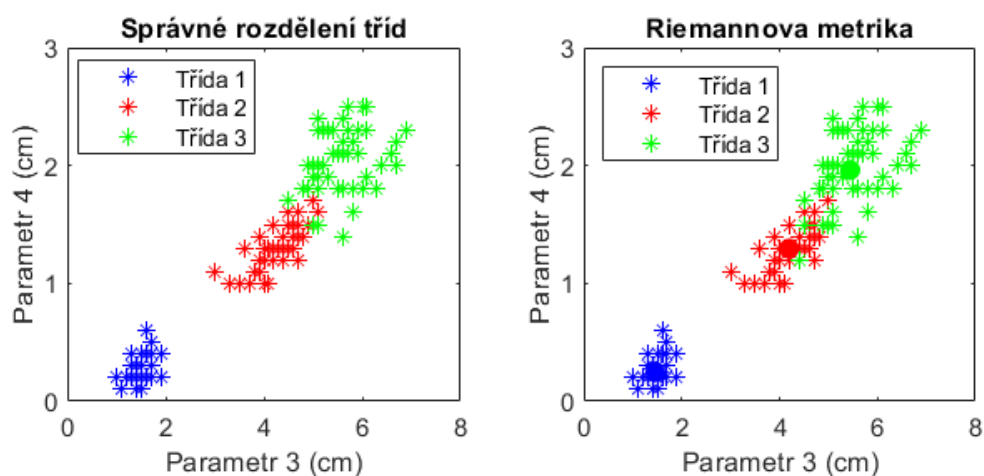
Na stejných testovacích datech jako k-NN jsem rovněž otestovala vytvořený klasifikátor k-means s využitím čtyř porovnávaných metrik, výsledky této klasifikace jsou uvedeny v následující tabulce 5.9.

Tabulka 5.9: Výsledky ověření *k-means* na testovacích datech.

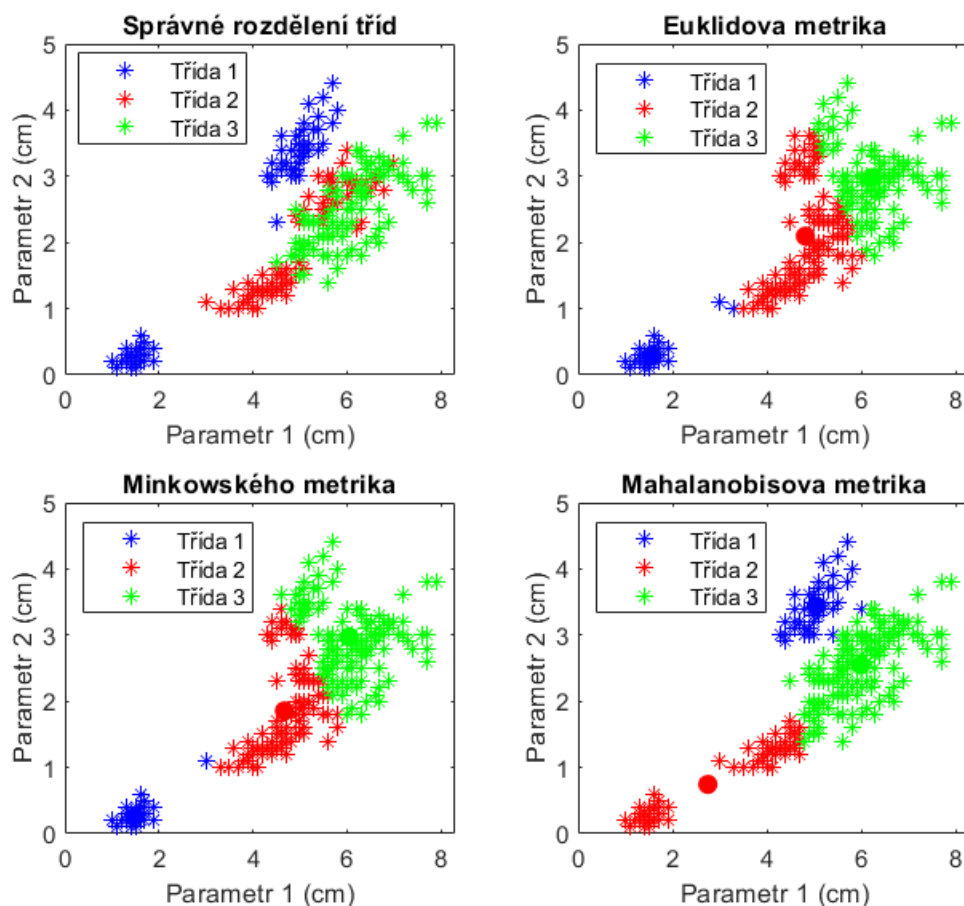
Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		56,7 %	98,0 %	58,0 %	64,0 %
Senzitivita	třída 1	50,0 %	100 %	50,0 %	49,0 %
	třída 2	58,0 %	96,0 %	55,0 %	44,0 %
	třída 3	62,0 %	98,0 %	69,0 %	99,0 %
Specifická	třída 1	98,5 %	100 %	99,5 %	99,0 %
	třída 2	65,0 %	99,0 %	76,0 %	74,5 %
	třída 3	71,5 %	98,0 %	61,5 %	72,5 %

Na výsledcích v tabulce 5.9 je vidět, že kvalita klasifikace při použití Riemannovy metriky se značně liší od klasifikace s ostatními metrikami. V tomto případě je důležité poznamenat, že příznakový prostor, který byl klasifikován Riemannovou metrikou se liší od prostoru klasifikovaného ostatními metrikami. Příčinou je postup nakládání s maticemi příznaků v rámci algoritmu `kmeans_E` popsáný v kapitole 4.5.4, tj. rozdělení matice na jednotlivé řádky, které v případě reálných dat odpovídají jednotlivým EEG kanálům, v případě těchto testovacích dat odpovídají dvěma druhům lístků.

Tímto postupem vznikl pro Euklidovu, Minkowského a Mahalanobisovu metriku 2D prostor znázorněný na obrázku 5.2, se kterým si poměrně dobře uměl poradit algoritmus *k-NN*, ale už byl méně vhodný pro klasifikaci algoritmem *k-means*. Příznakový prostor pro Riemannovu metriku zůstal čtyřdimenzionální a tím i vhodněji uspořádaný pro klasifikaci pomocí *k-means*, což je vidět i na obrázku 5.1.



Obrázek 5.1: Výsledek klasifikace testovacích dat algoritmem *k-means* při použití Riemannovy metriky v porovnání se správným rozdělením do tříd.



Obrázek 5.2: Výsledek klasifikace testovacích dat algoritmem k-means při použití Euklidovy, Minkowského a Mahalanobisovy metriky v porovnání se správným rozdělením do tříd.

5.5 Výsledky klasifikace algoritmem k-means

Funkčnost implementace algoritmu k-means s Riemannovou metrikou popsané v první části kapitoly 4.5.4 a vytvořené na základě literatury (zejména [13]) jsem ověřila na testovacích datech, kde nedocházelo k problému s konvergencí a klasifikace probíhala s vysokou přesností, jak bylo prezentováno v předchozí kapitole 5.4.

Při použití na reálném příznakovém prostoru spánkových dat, který byl hůře separabilní a také vícedimenzionální došlo ale k tomu, že se iterativní výpočet středů shluků (rovnice 9) stával nestabilní a nebyl schopen dokonvergovat k zadané podmínce. Navíc se vypočtené středy dostávaly mimo původní souřadnice normovaného prostoru (všechny použité příznaky byly normovány na hodnoty v rozmezí 0 až 1). Tím pádem byly v dalším cyklu k-means algoritmu všechny klasifikované segmenty přiřazeny pouze ke středům, které byly normovanému prostoru nejbližší. Výsledkem algoritmu, jehož iniciační počet center byl $k = 5$, tak byla data rozdělená

např. do dvou shluků, v extrémních případech data nebyla rozdělena vůbec a všechna náležela k jednomu, nejbližšímu středu.

Níže uvedené výsledky klasifikace spánkových dat v tabulce 5.10 jsou tedy v případě Riemannovy metriky získány pomocí upraveného algoritmu, jak bylo popsáno v kapitole 4.5.4 a podobně jako u k-NN jsem i pro k-means klasifikaci využila redukovaný prostor 6 příznaků a 3 kanálů, který umožnil dynamičtější konvergenci.

V tabulce 5.10 lze vidět, že správnost klasifikace je obecně až o 30 % nižší než byla u algoritmu k-NN, a na rozdíl od něj nejlepší výsledky ve většině parametrů vykazuje klasifikace s využitím Riemannovy metriky (výraznou výjimkou je senzitivita u fáze REM). Nejvyšší senzitivitu většinou vykazovala fáze REM, což napovídá, že její shluky byly poměrně kompaktní. Fáze, které vykazovaly vyšší hodnotu senzitivity, měly obvykle naopak nižší hodnotu specifity.

Tabulka 5.10: Výsledky klasifikace spánkových záznamů algoritmem k-means.

Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		43,0 %	50,2 %	42,7 %	46,2 %
Senzitivita	WK	17,2 %	23,4 %	16,0 %	28,8 %
	RM	65,8 %	48,7 %	67,4 %	58,1 %
	N1	3,5 %	5,5 %	7,6 %	2,7 %
	N2	38,6 %	56,2 %	38,3 %	47,0 %
	N3	41,5 %	48,0 %	40,7 %	46,3 %
Specifita	WK	90,3 %	95,6 %	91,5 %	95,5 %
	RM	72,7 %	86,2 %	72,1 %	73,4 %
	N1	94,5 %	97,9 %	91,1 %	93,0 %
	N2	77,1 %	67,1 %	78,3 %	71,6 %
	N3	88,8 %	80,5 %	91,3 %	92,2 %

V tabulkách 5.11 až 5.14 opět uvádím klasifikační matice, pro informaci o tom, jak byly klasifikovány chybně určené segmenty. Je zde vidět velké množství segmentů fáze N2, které byly klasifikovány jako REM spánek a často i naopak byly segmenty REM klasifikovány jako fáze N2. Také segmenty fází N2 a N3 byly často zaměňovány. A segmenty fáze N1, která měla stejně jako pro klasifikátor k-NN velmi nízkou senzitivitu, byly často klasifikovány jako fáze REM nebo fáze N2.

Tabulka 5.11: Klasifikační matice při využití Euklidovy metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	45,2	33,5	7,9	119,0	69,7
	RM	87,8	375,6	47,3	354,2	71,2
	N1	3,6	2,5	3,1	43,8	86,7
	N2	35,0	137,7	20,4	434,9	145,7
	N3	10,8	13,5	6,0	191,2	266,7

Tabulka 5.12: Klasifikační matice při využití Riemannovy metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	17,1	12,3	2,6	14,1	5,9
	RM	24,4	97,0	10,1	51,3	7,9
	N1	1,6	4,9	1,6	9,6	2,0
	N2	11,6	54,3	8,7	213,8	88,0
	N3	6,1	19,2	5,4	92,4	109,6

Tabulka 5.13: Klasifikační matice při využití Minkowského metriky.

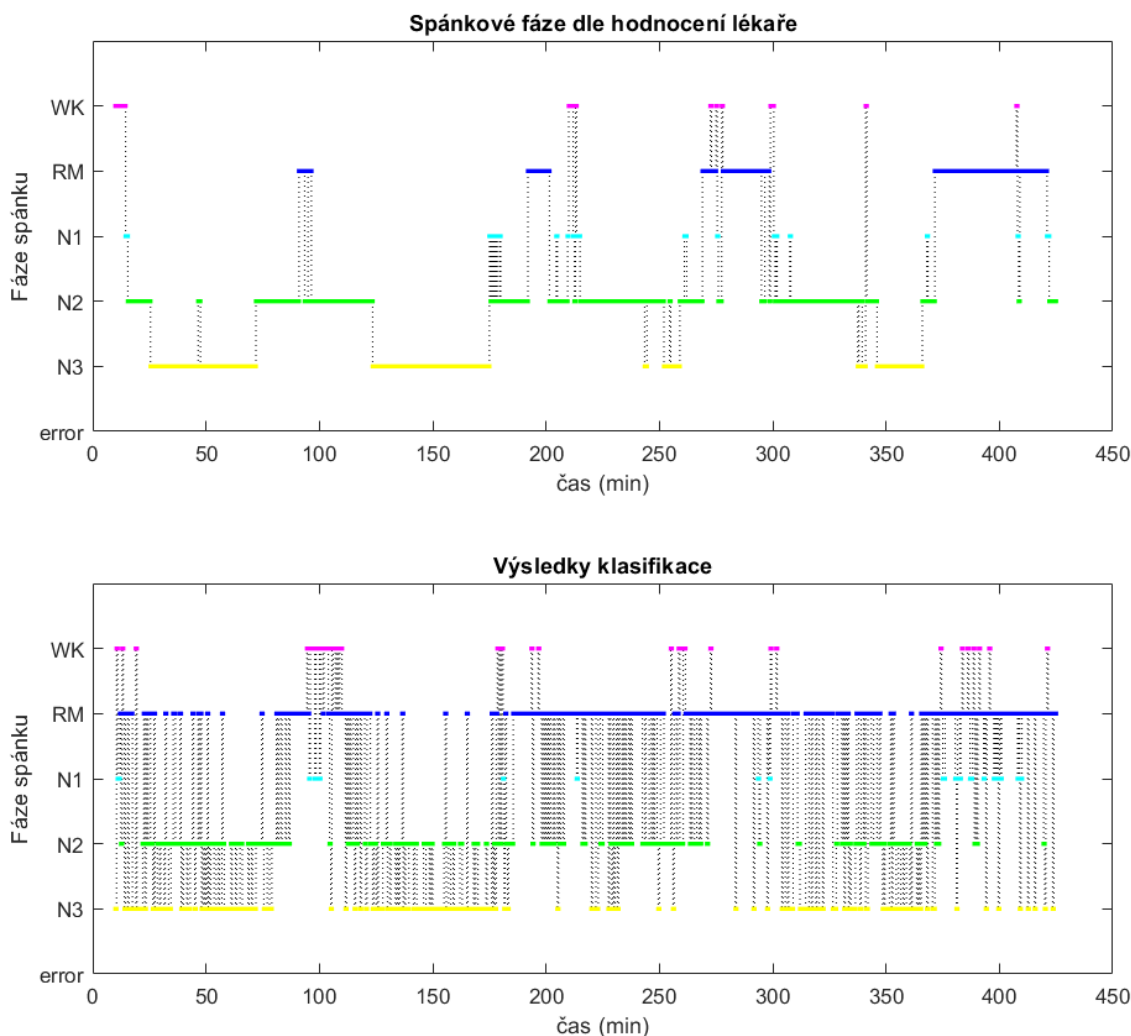
		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	33,4	31,9	5,9	103,2	64,6
	RM	88,9	382,4	48,0	362,4	74,2
	N1	14,5	17,2	7,1	99,4	88,0
	N2	33,6	118,3	18,7	431,9	150,8
	N3	12,0	13,0	5,0	146,2	262,4

Tabulka 5.14: Klasifikační matice při využití Mahalanobisovy metriky.

		Správná fáze				
		WK	RM	N1	N2	N3
Výsledek klasifikace	WK	54,1	24,8	6,9	60,2	19,3
	RM	67,8	319,6	37,3	336,7	106,9
	N1	5,7	9,0	2,2	78,0	85,1
	N2	48,6	201,3	32,4	536,8	138,8
	N3	6,2	8,0	5,9	131,4	290,0

Obecně shluky, které nejlépe odpovídaly fázím REM a N2, často obsahovaly i mnoho segmentů fází ostatních, což vysvětluje relativně nízkou specificitu klasifikace těchto fází.

Níže uvádím obrázek 5.3, na kterém jsou výsledky klasifikace vybraného záznamu algoritmem k-means s použitím Riemannovy metriky porovnány pomocí hypnogramů s hodnocením spánkových fází dle lékaře.



Obrázek 5.3: Výsledky automatické klasifikace algoritmem k-means při využití Riemannovy metriky v porovnání s hodnocením lékaře.

5.6 Statistické porovnání metrik pro k-means klasifikátor

U parametrů popisujících kvalitu klasifikace: správnosti, senzitivity a specificity pro každou použitou metriku jsem nejprve pomocí Smirnov-Kolmogorova testu na hladině významnosti 5 % zamítla, že by hodnoty pocházely z normálního rozdělení.

Potom jsem provedla Wilcoxonův párový test. V tabulce 5.15 jsou uvedeny výsledky testů pro jednotlivé parametry a pro jednotlivé dvojice metrik. V každé dvojici je vždy vlevo uvedena metrika, jejíž hodnoty parametrů byly vyšší (na-

značovaly lepší kvalitu klasifikace). Dvojice metrik jsou v tabulce opět seřazeny podle postupně klesající hodnoty rozdílu průměrných správností. Hodnota 1 v příslušném poli znamená, že byla na hladině významnosti 5 % zamítnuta nulová hypotéza a říká tak, že se parametr klasifikace pro tuto dvojici metrik lišil.

Statistický rozdíl ve správnosti klasifikace se na hladině významnosti 5 % provedlo prokázat pouze ve dvou případech, a to že klasifikace s Riemannovou metrikou byla lepší než s Euklidovou a Minkowského metrikou.

Tabulka 5.15: Výsledky statistického vyhodnocení vlivu použité metriky pro algoritmus k-means Wilcoxonovým testem.

Dvojice metrik		Ri – Mi	Ri – Eu	Ri – Ma	Ma – Mi	Ma – Eu	Eu – Mi
Správnost		1	1	0	0	0	0
Senzitivita	WK	0	0	0	1	1	0
	RM	0	0	0	0	0	0
	N1	0	0	0	0	0	0
	N2	1	1	0	1	1	0
	N3	0	0	0	0	0	0
Specificita	WK	0	1	0	1	1	0
	RM	1	1	1	0	0	0
	N1	1	0	1	0	0	1
	N2	0	0	0	0	0	0
	N3	0	0	0	0	0	0

Hodnota 0 znamená přijetí nulové hypotézy, hodnota 1 znamená zamítnutí nulové hypotézy na hladině významnosti 5 %. Porovnávané metriky: Ma – Mahalanobisova, Ri – Riemannova, Mi – Minkowského, Eu – Euklidova.

6 Diskuze

Hlavním zjištěním této práce je, že použití různých metrik při klasifikaci spánkového EEG ovlivňuje její výsledky v závislosti na použitém klasifikátoru různou měrou. Pro každý klasifikační algoritmus dávala staticky významně lepší výsledky jiná metrika. Mimo jiné jsem tak na Euklidově, Riemannově, Minkowského a Mahalanobisově metrice ověřila předpoklad, že zvolená metrika má na výsledky automatické klasifikace nějaký vliv.

Modifikovala jsem algoritmy k-NN a k-means pro použití s Riemannovou metrikou při klasifikaci spánkového EEG. Nepovedlo se prokázat, že by pro klasifikaci spánkových fází byly principy Riemannovy metriky obecně výhodnější, přestože nabízely výhodu zpracování multikanálových segmentů a vykazovaly nejlepší výsledky na testovacích datech, naopak se ukázalo, že Riemannova metrika klade na způsob zpracování a data různé omezující požadavky.

Pro klasifikátor k-NN byla nejvhodnější Mahalanobisova metrika, která měla i statisticky významně lepší senzitivitu než všechny ostatní metriky pro všechny spánkové fáze, poté následovala metrika Euklidova, Minkowského a nejnižší kvalitu klasifikace vykazovala Riemannova metrika. Pro klasifikátor k-means se povedlo prokázat, že za daných podmínek byla Riemannova metrika výhodnější než Euklidova a Minkowského. Jenom v několika případech se u některých spánkových fází a některých dvojic metrik povedlo prokázat rozdíl v senzitivitě a specificitě klasifikace, rozdíly mezi ostatními metrikami ve správnosti klasifikace se prokázat nepodařilo.

Zajímavým zjištěním byly velice dobré výsledky získané při použití Mahalanobisovy vzdálenosti v kombinaci klasifikátorem k-NN. V literatuře bývá bez dalšího vysvětlení brána jako první volba Euklidova metrika, pravděpodobně z důvodu její snadné aproximace. I v mé práci jsem v případě klasifikátoru k-NN dosáhla s Euklidovou metrikou druhých nejlepších výsledků a např. v článku [10], kde se věnovali optimalizaci k-NN klasifikátoru pro rozlišení EEG zdravých a dyslektických dětí, dosáhli nejlepších výsledků s Euklidovou metrikou, ovšem porovnávali ji s Kosinovou a korelační. Mahalanobisova vzdálenost má oproti Euklidově složitější výpočetní vztah, a i když ne o mnoho, v některých aplikacích to může hrát roli. V článku [27] (věnuje se algoritmům pro BCI systémy) lze o Mahalanobisově metrice dohledat poznámku, že klasifikátory založené na této metrice se i přes dobré poskytované výsledky objevují spíše zřídka.

Klasifikátor k-NN dosahoval správnosti klasifikace v závislosti na metrice 69,7 až 77,7 %. Pro srovnání uvádím dva roky starou studii [2], kde použili pro klasifikaci spánkových fází také pouze EEG signál, byl popsáný frekvenčními příznaky, klasifikaci prováděli na 5 fází stejně jako já a mezi použitými klasifikačními algoritmy byl i k-NN. S ním dosáhli správnosti klasifikace 64,8 %, bohužel zde ke k-NN klasifikátoru neuvádějí další podrobnosti, dá se ale předpokládat, že použili Euklidovu metriku. Nejlepších výsledků zde dosáhli s klasifikátorem Random forests a to 75,3 %. Lze nalézt i studie dosahující vyšší správnosti v klasifikaci spánkových fází, např. 93 % ve studii [6], kde ovšem použili neuronové sítě, navíc hodnotili fáze REM a N1 jako jednu třídu, nebo 88,5 % ve studii [8] používající SVM klasifikátor pro klasifikaci všech 5 fází.

Jak je vidět z výsledku srovnání na dvou klasifikačních algoritmech, pro každý z nich byla výhodnější jiná metrika. Dá se proto předpokládat, že pro jiné klasifikátory nebo i jiný typ dat by mohla být pro lepší kvalitu klasifikace výhodná zase jiná metrika. Na základě této práce nelze určit metriku, která by byla obecně pro klasifikaci EEG dat nejlepší, to by také přinejmenším vyžadovalo otestování širšího spektra metrik, klasifikátorů a ideálně i ještě větší počet spánkových záznamů. Přesto, když odmyslíme klasifikaci na odlišném principu Riemannovy metriky, ostatní tři metriky vykazovaly u obou algoritmů stejné pořadí dle dosahované správnosti klasifikace: nejlepší byla Mahalanobisova, po té Euklidova a nakonec Minkowského metrika s koeficientem $p = 3$.

Během řešení diplomové práce vznikly některé odlišnosti od původně zamýšlených postupů. Ve výsledcích v kapitole 5.1 jsem již nastínila, že jsem klasifikaci algoritmem k-NN s použitím Riemannovy metriky nemohla provést podle původního záměru v příznakovém prostoru se všemi 24 příznaky a 19 kanály záznamu. Algoritmus s Riemannovou metrikou v takovém prostoru totiž selhával. Řešení tohoto problému, na který měl zřejmě vliv vysoký počet dimenzí v příznakovém prostoru, se našlo v postupu, který byl popsán v kapitole 4.1 a 4.4, tj. výběrem pouze 6 příznaků a 3 kanálů na základě literatury.

Toto řešení je výhodné z vícero důvodů. Zaprvé výsledky této práce by mohli být do budoucna využity v Národním ústavu duševního zdraví a tam v současné době při monitorování spánku upouštějí od snímání EEG všemi 19 elektrodami a přecházejí na systém 4 EEG elektrod + 2 referenční. Zadruhé méně příznaků a méně zpracovávaných dat vůbec způsobuje menší výpočetní náročnost a tím i potenciál pro využití v real-time klasifikaci.

Po provedení klasifikace se zredukovaným počtem použitých příznaků a kanálů jsem se bez ohledu na použitou metriku potýkala s velmi malou senzitivitou u spánkové fáze N1 a to mezi 3 a 10 %. To byl částečně negativní důsledek právě redukce počtu příznaků, ale zásadní rozdíl senzitivity této fáze v porovnání s ostatními byl z velké části způsoben malým počtem segmentů náležejících k této fázi napříč všemi subjekty (viz také tabulka 4.1).

Chybějící množství segmentů nebylo pochopitelně možné nějakým způsobem do záznamu dodat, proto jsem se pokusila alespoň tento vliv potlačit. Jednou z možností bylo spojit fázi N1 do jedné třídy s jinou fází, rozhodla jsem se spojit ji s fází REM, se kterou má podobné charakteristiky EEG signálu (viz tabulka 2.1), tento postup lze najít i v literatuře [6] a také z klasifikačních matic bylo vidět, že poměrně velká část N1 segmentů byla klasifikována jako REM fáze.

Výsledky klasifikace ukázaly, že spojená třída N1 + REM má sice hodnotu senzitivity přes 70 %, tj. odpovídá hodnotě původní třídy REM, ale celková správnost klasifikace vzrostla pouze o méně než 2 %. Spojení tříd mělo tedy mírně pozitivní vliv na celkový výsledek, ale za cenu ztráty rozlišení dvou fází.

Druhá možnost byla nevybírat do trénovací množiny algoritmu k-NN 2/3 segmentů z každé spánkové fáze, ale stejný absolutní počet segmentů pro každou fázi. Tento počet se musel odvíjet od nejméně početné skupiny. Výsledky klasifikace byly velmi příznivé pro fázi N1, jejíž senzitivita vzrostla na hodnoty okolo 30 % a i senzitivita u fáze WK přibližně o 10 % vzrostla, ale na druhé straně asi o 10 % klesla senzitivita fáze REM a o více než 30 % klesla i senzitivita fáze N2, takže došlo i k významnému poklesu celkové správnosti klasifikace. Proto jsem se rozhodla ani jeden z těchto testovaných postupů dále nevyužít. Kompletní popisované výsledky testovaných postupů jsou uvedeny v tabulkách v příloze A.

Menší průměrná senzitivita (44 až 57 %) byla také u fáze WK, zde rovněž záleželo na počtu segmentů v záznamu a u této fáze se počty napříč subjekty velmi lišily. U záznamů, které obsahovaly větší počet segmentů, byla i senzitivita klasifikace úměrně vyšší a dosahovala i stejných hodnot jako pro jiné fáze.

I algoritmus k-means musel být upraven a to přímo jeho implementace oproti původní verzi vytvořené na základě studie [13]. Iterativní výpočet středů shluků zahrnutý v algoritmu a fungující na testovacích datech nebyl na reálných datech schopen dokonvergovat, jak je uvedeno v kapitole 5.5. Nejprve jsem se pokusila snížit počet příznaků i kanálů podobně jako u k-NN klasifikátoru, což ale pouze mírně snížilo četnost výskytu nestabilních výpočtů.

Omezila jsem tedy maximální počet iterací, nastal ovšem další rovněž v kapitole 5.5 popsáný jev, že se středy dostávaly mimo prostor dat a výsledkem byla data roztržena do méně shluků než byla zadaná hodnota k . Pokoušela jsem se najít nějaký způsob normalizace, kterým by se vypočtené pozice středů ve stejném poměru vrátily do rozměrů původního příznakového prostoru. Nepodařilo se mi ale najít postup, který by zaručil, že pozice středů nebudou touto normalizací zkresleny, navíc zde byl i problém se standardním způsobem ukončení algoritmu k-means ve chvíli, kdy se pozice středů již nemění, toho se při použití normalizace nepovedlo zcela docílit.

Nakonec zcela nepomohlo ani úplné odstranění iterativního výpočtu nových středů z algoritmu a pro spolehlivou funkčnost algoritmu jsem ještě musela upravit vztah pro výpočet kovarianční matice (viz rovnice 11).

I přes tyto překážky a úpravy byla kvalita klasifikace algoritmem k-means s Riemannovou metrikou nejlepší. Celkově byla ale kvalita klasifikace algoritmem k-means slabá a výsledky algoritmu k-NN byly oproti němu i o desítky procent lepší. Také výsledky klasifikace pomocí k-means napříč hodnocenými spánkovými záznamy různě kolísaly a pravděpodobně z toho důvodu se nepodařilo prokázat statisticky významné rozdíly správnosti klasifikace mezi všemi metrikami.

Stejně jako u algoritmu k-NN byla i u k-means velmi nízká hodnota senzitivity u spánkové fáze N1 (méně než 8 %), zde také hrál svoji roli malý počet segmentů této fáze v záznamech. Toto je ve spánkových záznamech poměrně častý problém [2], pokud bychom se mu chtěli vyhnout, lze tuto fázi buď spojit s jinou, jak bylo zmíněno u k-NN nebo např. ve článku [13] dokonce tuto spánkovou fázi pro svoji studii úplně vynechali.

Kvalitě klasifikace pro k-means algoritmus by mohlo pomoci i zvýšení hodnoty k na hodnotu větší než je počet fází, do teoreticky nadbytečných shluků by se mohly dostat např. segmenty obsahující artefakty, které se nepodařilo odstranit provedenou filtrací.

Cílem této práce sice nebylo hledat optimální nastavení k-means, ale je pravděpodobné, že pro lepší kvalitu klasifikace algoritmem k-means, by bylo vhodné použít některou z metod redukce dimenze příznakového prostoru jako je PCA nebo t-SNE. V budoucnu by bylo dobré vyzkoušet, zda by v takto upraveném příznakovém prostoru i algoritmus s Riemannovou metrikou nebyl stabilnější podobně jako na testovacích datech.

7 Závěr

Provedla jsem kompletní předzpracování spánkových EEG záznamů 20 subjektů. Na záznamy jsem v programovém prostředí MATLAB[®] pomocí toolboxu Fieldtrip aplikovala frekvenční filtry typu FIR, s mezními frekvencemi 0,5 Hz a 30 Hz. Pro EEG signál jsem zvolila montáž s referenčními elektrodami M1 a M2 na základě standardu pro hodnocení spánkových dat. Poté jsem segmentovala signál na segmenty o délce 30 sekund. Nakonec jsem aplikovala ještě prahový detektor pro detekci velmi odlehlých hodnot a odstranila segmenty, které tyto hodnoty obsahovaly. S pomocí uvedených filtrací jsem v záznamech částečně potlačila oční a svalové artefakty.

Z předzpracovaných EEG záznamů jsem pro jednotlivé segmenty extrahovala příznaky. Pro popis každého segmentu jsem s pomocí vytvořené funkce vypočítala celkem 24 příznaků. Pro klasifikaci jsem následně dle literatury z 24 vypočítaných příznaků vybrala pouze 6 a z celkem 19 kanálů EEG signálu jsem vybrala pouze 3.

V programovém prostředí MATLAB[®] jsem implementovala vybraný učící se klasifikátor k-NN a neučící se k-means. Oba klasifikátory jsem implementovala včetně zkoumané Riemannovy metriky a rovněž jsem do nich zakomponovala další tři srovnávací metriky: Euklidovu, Minkowského (parametr $p = 3$) a Mahalanobisovu metriku. Vytvořené algoritmy jsem nejprve ověřila na testovacích datech, poté jsem jejich prostřednictvím provedla klasifikaci segmentů EEG do pěti spánkových fází.

Prokázala jsem, že zvolená metrika má vliv na výsledky automatické klasifikace spánkového EEG a že se tyto výsledky ve většině případů od sebe statisticky významně liší. Pro klasifikátor k-NN byla nejvhodnější Mahalanobisova metrika, která dosáhla průměrné správnosti klasifikace 77,7 % a její senzitivita byla i přes 80 %. Výjimkou byla obecně problematická spánková fáze N1, jejíž senzitivita pro oba klasifikátory a všechny metriky nepřesáhla 10 %. Dosažené hodnoty správnosti klasifikace algoritmem k-NN se všemi metrikami byly srovnatelné s výsledky uváděnými v literatuře.

Princip Riemannovy metriky má výhodu a potenciál v tom, že v časovém úseku udržuje kontext všech kanálů, stejně jako i lékař hodnotí všechny signálové kanály najednou. Prokázala jsem, že Riemannova metrika byla za stanovených podmínek nejvhodnější metrikou pro klasifikátor k-means. Význam tohoto zjištění snižuje fakt, že správnost této klasifikace byla pouze 50,2 % a senzitivita dosáhla hodnoty maximálně 56 %. Klasifikátor k-means se celkově nejevil jako vhodný pro klasifikaci spánkových fází v EEG záznamech.

Seznam použité literatury

- [1] Spánek; In: *Nové universum A-Ž*. V Praze: Knižní klub, první vydání, 2003, ISBN 80-242-1069-x.
- [2] Tzimourta, K.: EEG-Based Automatic Sleep Stage Classification. *Biomedical Journal of Scientific & Technical Research*, ročník 7, č. 4, 2018, ISSN 25741241, doi:10.26717/BJSTR.2018.07.001535, [vid. 2019-06-25]. Dostupné z: <https://biomedres.us/fulltexts/BJSTR.MS.ID.001535.php>
- [3] Roberts, A. M.: *Kompletní lidské tělo*. Praha: Knižní klub, vyd. 1 vydání, 2012, ISBN 978-80-242-2958-4.
- [4] Silber, M.: Staging Sleep. *Sleep Medicine Clinics*, ročník 4, č. 3, 2012, doi:<http://dx.doi.org/10.1016/j.jsmc.2012.06.009>, [vid. 2019-06-25]. Dostupné z: [https://www.sleep.theclinics.com/article/S1556-407X\(12\)00074-4/fulltext](https://www.sleep.theclinics.com/article/S1556-407X(12)00074-4/fulltext)
- [5] Krajča, V.; Mohylová, J.: *Číslicové zpracování neurofyziologických signálů*. Praha: České vysoké učení technické v Praze, první vydání, 2011, ISBN 9788001047217.
- [6] Ebrahimi, F.: Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2008, ISBN 978-1-4244-1814-5, s. 1151–1154, doi:10.1109/IEMBS.2008.4649365, [vid. 2019-06-25]. Dostupné z: <http://ieeexplore.ieee.org/document/4649365/>
- [7] Li, Y.; Wong, K. M.: Riemannian Distances for Signal Classification by Power Spectral Density. *IEEE Journal of Selected Topics in Signal Processing*, ročník 7, č. 4, 2013: s. 655–669, ISSN 1932-4553, doi:10.1109/JSTSP.2013.2260320. Dostupné z: <http://ieeexplore.ieee.org/document/6509394/>
- [8] Bajaj, V.; Pachori, R. B.: Automatic classification of sleep stages based on the time-frequency image of EEG signals. *Computer Methods and Programs in Biomedicine*, ročník 112, č. 3, 2013: s. 320–328, doi:<http://dx.doi.org/10.1016/j.cmpb.2013.07.006>, [vid. 2019-06-25]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0169260713002265?via%3Dihub>

- [9] Gharbali, A. A.; Najdi, S.; Fonseca, J. M.: Investigating the contribution of distance-based features to automatic sleep stage classification. In *Computers in Biology and Medicine*, 2018, s. 8 – 23, doi:<https://doi.org/10.1016/j.compbimed.2018.03.001>. Dostupné z: <https://www.sciencedirect.com/science/article/abs/pii/S0010482518300520?via%3Dihub>
- [10] Zainuddin, A. Z. A.; Lee, K. Y.; Mansor, W.; aj.: Optimized KNN classify rule for EEG based differentiation between capable dyslexic and normal children. In *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, 2016, ISBN 978-1-4673-7791-1, s. 685–688, doi:10.1109/IECBES.2016.7843537, [vid. 2017-12-26]. Dostupné z: <http://ieeexplore.ieee.org/document/7843537/>
- [11] Chatterjee, S.; Choudhury, N. R.; Bose, R.: Detection of epileptic seizure and seizure-free EEG signals employing generalised S -transform. *IET Science, Measurement & Technology*, 2017: s. 847–855, ISSN 1751-8822, doi:10.1049/iet-smt.2016.0443, [vid 2017-12-26]. Dostupné z: <http://digital-library.theiet.org/content/journals/10.1049/iet-smt.2016.0443>
- [12] THE MATHWORKS, I.: MATLAB Documentation: pdist2. [vid. 2020-04-26]. Dostupné z: <https://www.mathworks.com/help/stats/pdist2.html>
- [13] Saifutdinova, E.; Congedo, M.: An Unsupervised Multichannel Artifact Detection Method for Sleep EEG Based on Riemannian Geometry. *Sensors*, ročník 19, č. 3, 2019, ISSN 1424-8220, doi:10.3390/s19030602, [vid. 2020-01-23]. Dostupné z: <http://www.mdpi.com/1424-8220/19/3/602>
- [14] Goh, A.; Vidal, R.: Clustering and dimensionality reduction on Riemannian manifolds. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2008, ISBN 978-1-4244-2242-5, s. 1–7, doi:10.1109/CVPR.2008.4587422, [vid. 2019-06-25]. Dostupné z: <http://ieeexplore.ieee.org/document/4587422/>
- [15] Salsabilian, S.; Slavakis, K.: Riemannian-geometry-based modeling and clustering of network-wide non-stationary time series: The brain-network case. [vid. 2020-04-21]. Dostupné z: arxiv.org/abs/1701.07767
- [16] Congedo, M.; Barachant, A.; Bhatia, R.: Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, ročník 4, č. 3, 2017-03-27: s. 155–174, ISSN 2326-263X, doi:

- 10.1080/2326263X.2017.1297192. Dostupné z: <http://dx.doi.org/10.1080/2326263X.2017.1297192>
- [17] Gong, S.; Cristani, M.: *Person re-identification*. New York: Springer, 2014, ISBN 978-1-44716295-7.
- [18] Lukasz Piatek; Fiedler, P.; Haueisen, J.: Eye state classification from electroencephalography recordings using machine learning algorithms. *Digital Medicine*, ročník 4, č. 2, 2018: s. 84 – 95, ISSN 2226-8561, doi:10.4103/digm.digm_41_17, [vid. 2020-04-23]. Dostupné z: <http://www.digitmedicine.com/text.asp?2018/4/2/84/239672>
- [19] Iris flower data set. In: Wikipedia: the free encyclopedia [online]. Wikimedia Foundation, 2003. [vid. 2020-05-09]. Dostupné z: https://en.wikipedia.org/wiki/Iris_flower_data_set
- [20] Grant, M.; Boyd, S.: MATLAB, version R2019b. Campus-Wide License, 2019.
- [21] Oostenveld, R.; Fries, P.: FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, ročník 2011, 2011, doi:10.1155/2011/156869, [vid. 2019-10-23].
- [22] Krajca, V.; Piorecka, V.; Schaabova, H.; aj.: Detection of sleep stages in neonatal EEG records. In *EMBECE & NBC 2017*, Singapore: Springer Singapore, 2017, ISBN 978-981-10-5121-0, s. 250–253, doi:10.1007/978-981-10-5122-7_63, [cit. 2019-12-27]. Dostupné z: http://link.springer.com/10.1007/978-981-10-5122-7_63
- [23] Krajča, V.; Petránek, S.; Pietielä, T.; aj.: Wave-finder: a new system for an automatic processing of long-term EEG recordings. *Quantitative EEG Analysis-Clinical Utility and New Methods*, 1993: str. 103–106, [vid. 2018-01-03].
- [24] Hese, P. V.; Philips, W.; Koninck, J. D.; aj.: Automatic detection of sleep stages using the EEG. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2001, ISBN 0-7803-7211-5, s. 1944–1947, doi:10.1109/IEMBS.2001.1020608, [vid. 2020-04-03]. Dostupné z: <http://ieeexplore.ieee.org/document/1020608/>

- [25] Balcarová, B.: Analýza příznaků automatické klasifikace epileptických EEG záznamů za pomoci algoritmu k-NN. Praha, 2018. Bakalářská práce. ČVUT v Praze, Fakulta biomedicínského inženýrství, Katedra biomedicínské techniky.
- [26] Machine Learning on Manifold. Marco Congedo, 2019, [vid. 2020-04-22]. Dostupné z: <https://github.com/Marco-Congedo/PosDefManifoldML.jl>
- [27] Lotte, F.; Congedo, M.; L'ecuyer, A.; aj.: A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, ročník 4, 2007.

Příloha A: Výsledky klasifikace testovaných postupů

Tabulka 1: Výsledky klasifikace k-NN při použití 19 kanálů a 24 příznaků.

Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		76,9 %	25,2 %	76,8 %	73,8 %
Senzitivita	WK	62,7 %	0 %	67,9 %	62,7 %
	RM	74,1 %	0 %	70,3 %	74,1 %
	N1	18,2 %	0 %	20,4 %	18,2 %
	N2	81,9 %	4,2 %	83,2 %	81,9 %
	N3	75,6 %	98,8 %	85,3 %	75,6 %

Tabulka 2: Výsledky klasifikace k-NN při spojení fází RM a N1 do jedné třídy.

Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		74,6 %	71,1 %	74,4 %	79,2 %
Senzitivita	WK	45,3 %	42,4 %	45,2 %	55,9 %
	RM + N1	71,7 %	70,0 %	71,5 %	76,9 %
	N2	75,7 %	72,2 %	75,6 %	80,0 %
	N3	78,2 %	71,1 %	77,7 %	81,3 %

Tabulka 3: Výsledky klasifikace k-NN při výběru stejně velkých trénovacích množin.

Metrika		Euklidova	Riemannova	Minkowského	Mahalanobisova
Správnost		56,9 %	50,2 %	56,4 %	61,8 %
Senzitivita	WK	55,3 %	56,4 %	55,2 %	63,9 %
	RM	58,9 %	56,2 %	58,2 %	67,6 %
	N1	29,7 %	29,8 %	29,3 %	33,9 %
	N2	46,5 %	36,1 %	46,1 %	51,7 %
	N3	76,8 %	71,7 %	76,1 %	77,9 %

Příloha B: Seznam vytvořených skriptů a funkcí

Název skriptu	Popis
<code>predzpracovani_komplet</code>	Předzpracování záznamu obsahující filtraci, segmentaci a výpočet příznaků
<code>klasifikace_kNN</code>	Skript sloužící pro aplikaci klasifikátoru k-NN s výběrem různých metrik a vyhodnocení klasifikace
<code>klasifikace_kMeans</code>	Skript sloužící pro aplikaci klasifikátoru k-means s výběrem různých metrik a vyhodnocení klasifikace

Název funkce	Popis
<code>kNN_E</code>	Klasifikátor k-NN umožňující výběr použité metriky
<code>kNN_R</code>	Klasifikátor k-NN s Riemannovou metrikou
<code>kMeans_E</code>	Klasifikátor k-means umožňující výběr použité metriky
<code>kMeans_R</code>	Klasifikátor k-means s Riemannovou metrikou
<code>compFeat</code>	Výpočet 24 příznaků
<code>deftrl</code>	Funkce sloužící pro definici 30 s dlouhých segmentů na základě značek spánkových fází
<code>error_matrix</code>	Výpočet klasifikační matice
<code>accuracy</code>	Výpočet správnosti klasifikace
<code>sensitivity</code>	Výpočet senzitivity klasifikace pro jednotlivé třídy
<code>specificity</code>	Výpočet specifity klasifikace pro jednotlivé třídy

Příloha C: Obsah přiloženého CD

- Abstrakt práce v ČJ
- Abstrakt práce v AJ
- Klíčová slova v ČJ
- Klíčová slova v AJ
- Kompletní diplomová práce
- Naskenované zadání diplomové práce
- Příklad příznaků jednoho záznamu
- Skripty a funkce vytvořené v programu MATLAB[®] ve složce
- Ukázková struktura zpracovávaných dat