



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
Katedra biomedicínské techniky

**Vliv normalizace příznaků na klasifikaci
EEG segmentů**

**The influence of the feature's normalisation
on segments of EEG recordings**

Bakalářská práce

Studijní program: Biomedicínská a klinická technika

Studijní obor: Biomedicínský technik

Autor bakalářské práce: Barbora Popelková

Vedoucí bakalářské práce: Ing. Václava Piorecká

Kladno 2018/2019



ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Popelková** Jméno: **Barbora** Osobní číslo: **465461**
Fakulta: **Fakulta biomedicínského inženýrství**
Garantující katedra: **Katedra biomedicínské techniky**
Studijní program: **Biomedicínská a klinická technika**
Studijní obor: **Biomedicínský technik**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Vliv normalizace příznaků na klasifikaci EEG segmentů

Název bakalářské práce anglicky:

The influence of the feature's normalisation on segments of EEG recordings

Pokyny pro vypracování:

Metody automatické a semi-automatické klasifikace jsou založeny na výpočtu příznaků a jejich normalizaci. Záznamy EEG předzpracujte a vypočítejte příznaky jednotlivých segmentů. Analyzujte EEG záznamy prostřednictvím různých metod normalizace. Normalizovaná data klasifikujte. Vyhodnoťte vliv normalizace na výslednou klasifikaci.

Seznam doporučené literatury:

- [1] AMIN, Hafeez, Wajid MUMTAZ, Ahmad SUBHANI, Mohamad SAAD a Aamir MALIK, Classification of EEG Signals Based on Pattern Recognition Approach, Frontiers in Computational Neuroscience, ročník 11, číslo 103, 2017
- [2] BOGAARTS, J., D. HILKMAN, E. GOMMER, V. VAN KRANEN-MASTENBROEK a J. REULEN, Improved epileptic seizure detection combining dynamic feature normalization with EEG novelty detection, Medical & Biological Engineering & Computing, ročník 54, číslo 12, 2016
- [3] Sebastian Raschka, About feature scaling and normalization, 2014, http://sebastianraschka.com/Articles/2014_about_feature_scaling.html
- [4] Krajča V., Mohylová J., Číslíkové zpracování neurofyziologických signálů, ed. 1st, ČVUT Praha, 2011, ISBN 9788001047217

Jméno a příjmení vedoucí(ho) bakalářské práce:

Ing. Václava Piorecká

Jméno a příjmení konzultanta(ky) bakalářské práce:

Ing. Marek Piorecký

Datum zadání bakalářské práce: **18.02.2019**

Platnost zadání bakalářské práce: **20.09.2020**

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

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

PROHLÁŠENÍ

Prohlašuji, že jsem bakalářskou práci s názvem „Vliv normalizace příznaků na klasifikaci EEG segmentů“ vypracovala samostatně a použila k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k bakalářské 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).

V dne

.....

Barbora Popelková

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucí své bakalářské práce Ing. Václavě Piorecké za přínosné konzultace, cenné rady a odborné vedení při psaní této práce. Dále bych ráda poděkovala konzultantovi Ing. Marku Pioreckému za podnětné návrhy k práci. Mé poděkování patří také BRAIN Teamu za užitečné připomínky.

ABSTRAKT

Vliv normalizace příznaků na klasifikaci EEG segmentů

Pro zjednodušení vyhodnocování dlouhodobých elektroencefalografických záznamů se stále častěji využívá automatických či semi-automatických klasifikátorů, založených na výpočtu příznaků. Příznaky se do klasifikátoru vkládají normalizované z důvodu jejich různých dynamických rozsahů a jednotek. Cílem této práce je vyhodnocení vlivu normalizačních metod na výsledné klasifikaci. V programovém prostředí Wave-Finder bylo předzpracováno šest záznamů s epileptickou aktivitou. Po využití adaptivní segmentace bylo na datech vypočteno 23 příznaků. Na tyto příznaky byly v programovém prostředí MATLAB[®] implementovány dvě normalizační metody. Jedná se o normalizaci Z-skóre a metodu minima a maxima. Pomocí Kolmogorov-Smirovova testu normality bylo zjištěno, že na hladině významnosti 5 % data po normalizaci nepochází z normálního rozložení. Statisticky bylo pomocí Wilcoxonova testu zjištěno, že na datech normalizovaných pomocí dvou normalizačních metod existuje rozdíl na hladině významnosti 5 %. Normalizovaná data byla v programovém prostředí MATLAB[®] klasifikována pomocí k-means klasifikace do šesti tříd. Procentuálně byly vyjádřeny vztahy mezi třídami vzniklými po klasifikaci dat normalizovaných pomocí obou metod. Bylo provedeno vyhodnocení expertem a identifikace jednotlivých tříd. Pomocí normalizované hodnoty vnitrotřídního rozptylu bylo zjištěno, že třídy normalizované pomocí Z-skóre mají vyšší homogenitu. Na vybraném souboru dat se tedy Z-skóre normalizace jeví vhodnější.

Klíčová slova

Příznakový prostor, normalizace, Z-skóre, normalizace minimem a maximem, klasifikace, k-means

ABSTRACT

The influence of the feature's normalisation on segments of EEG recordings

To simplify the evaluation of long-term electroencephalographic records, automatic or semi-automatic classifiers based on feature extraction are increasingly used. Features entered to the classifier were normalised due to their different dynamic ranges and units. The aim of this work is to evaluate the influence of normalisation methods on the final classification. Six records with epileptic activity have been preprocessed in Wave-Finder. After using adaptive segmentation, 23 symptoms were calculated on the data. Two normalisation methods have been implemented on the data in the MATLAB[®] program. In this thesis, Z-score normalisation and a minimum and maximum method were used. Using the Kolmogorov-Smirnov test of normality, it was found that the data does not come from a normal distribution on a significance level of 5 %. By a statistical Wilcoxon test, it was found that there is a difference in the significance level of 5 % on data normalised by two normalisation methods. Normalised data were classified by k-means classification into six classes in the MATLAB[®] program. The relations between resulting classes from the classification of data normalised by both methods were expressed as percentages. An expert evaluation was carried out and individual classes were identified. Using normalised intraspecific variance, classes normalised by Z-score were found to have higher homogeneity. On the selected data set, Z-score normalisation appears to be more appropriate.

Keywords

Feature space, normalisation, z-score, min-max normalisation, classification, k-means

Obsah

Seznam symbolů a zkratk.....	9
1 Úvod	10
1.1 Přehled současného stavu.....	10
1.2 Cíle práce.....	11
2 Metody	12
2.1 Elektroencefalografie	12
2.1.1 Snímání signálu	12
2.1.2 Signál EEG	13
2.2 Programové prostředí	14
2.3 Data	15
2.3.1 Předzpracování	15
2.3.2 Segmentace.....	15
2.3.3 Výpočet příznaků.....	17
2.4 Normalizace	18
2.4.1 Z-skóre.....	18
2.4.2 Metoda minima a maxima	18
2.5 Test normality a statistické vyhodnocení	18
2.6 Klasifikace.....	19
2.6.1 K-means.....	20
2.7 Vyhodnocení klasifikace	22
3 Výsledky.....	23
3.1 Předzpracování dat	23
3.2 Normalizace dat.....	26
3.3 Normalita dat.....	31
3.4 Statistické zhodnocení.....	33
3.5 Klasifikace.....	33
3.6 Vyhodnocení klasifikace	34
4 Diskuse	43
5 Závěr	46
Seznam použité literatury	47

Příloha A: Obsah přiloženého CD.....	50
Příloha B: Boxplot zobrazení prvního záznamu s odlehlými hodnotami.....	51

Seznam symbolů a zkratek

Seznam symbolů

Symbol	Jednotka	Význam
<i>SIGM</i>	μV	Variabilita signálu
<i>APOS</i>	μV	Maximální pozitivní hodnota v daném segmentu
<i>ANEG</i>	μV	Maximální negativní hodnota v daném segmentu
<i>DELT1</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části delta frekv. pásma (0,5 Hz – 1,5 Hz)
<i>DELT2</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části delta frekv. pásma (2,0 Hz – 3,5 Hz)
<i>THET1</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části theta frekv. pásma (4,0 Hz – 5,5 Hz)
<i>THET2</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části theta frekv. pásma (6,0 Hz – 7,5 Hz)
<i>ALPH1</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části alfa frekv. pásma (8,0 Hz – 10,0 Hz)
<i>ALPH2</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části alfa frekv. pásma (10,5 Hz – 12, 5 Hz)
<i>SIGMA</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části beta frekv. pásma (18,0 Hz – 29 Hz)
<i>BETA</i>	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části beta frekv. pásma (13,5 Hz – 29 Hz)
<i>MAX1D</i>	$\mu\text{V}/\text{s}$	Maximální hodnota první derivace
<i>MD1</i>	$\mu\text{V}/\text{s}$	Průměrná hodnota první derivace
<i>MAX2D</i>	$\mu\text{V}/\text{s}^2$	Maximální hodnota druhé derivace
<i>MD2</i>	$\mu\text{V}/\text{s}^2$	Průměrná hodnota druhé derivace
<i>act</i>	μV^2	Aktivita (1. Hjorthův parametr)
<i>NLinE</i>	μV^2	Nelineární energie
<i>mf</i>	Hz	Střední frekvence
<i>Peaks</i>	Hz	Dominantní spektrální vrchol
<i>mob</i>	–	Mobilita (2. Hjorthův parametr)
<i>comp</i>	–	Komplexita (3. Hjorthův parametr)
<i>LOfC</i>	–	Délka křivky
<i>ZC</i>	–	Počet průchodů nulou

Seznam zkratek

Zkratka	Význam
ROC	Receiver operating characteristic
MDM	Median decaying memory
FB	Fixed baseline buffer
EEG	Elektroencefalografie
SVM	Support vector machines
RAPM	Raven's advance progressive matrix
DWT	Discrete wavelet transform
k-NN	k-nearest neighbours
MLP	Multilayer Perceptron
WF	Wave-Finder
AVG	Average reference
PPV	Positive predictive value

1 Úvod

Elektroencefalografie (EEG) je diagnostická metoda zaznamenávající elektrickou aktivitu mozku. Využívá se k monitorování a diagnóze v neurologii a psychiatrii. Jelikož se poruchy nemusí objevit během ambulantního záznamu, je často žádoucí dlouhodobé monitorování. Vyhodnocování těchto záznamů je však často složité z důvodu vysokého počtu dat a vyžaduje velké množství času. Z tohoto důvodu je výhodné využití automatických či semi-automatických klasifikací dlouhodobých záznamů. Tyto klasifikace využívají k zařazování do tříd vypočtené příznaky, které se před samotnou klasifikací musí normalizovat. Normalizace se provádí z důvodu různých dynamických rozsahů hodnot příznaků, které také mají vliv na funkci klasifikátoru. Nejenom výběr správných příznaků, ale i vybraný typ normalizační metody ovlivňují výkon klasifikátoru.

1.1 Přehled současného stavu

Existuje více metod normalizace příznaků, které slouží ke klasifikaci. Autoři článku [1] se zabývali automatickou detekcí epileptických záchvatů. Hlavní motivací autorů k napsání této práce byl fakt, že se mnoho pozornosti ukládá na výpočet různých příznaků, ale minimum pozornosti dostává normalizace těchto příznaků. Je zásadní, aby byla použita normalizace, která nemá negativní vliv na proces detekce záchvatů. Jedná se o první publikaci zabývající se vlivem různých technik normalizace na detekci záchvatů. Autoři porovnávali pět normalizačních metod, které byly založeny na střední hodnotě, mediánu, standardní odchylce, vrcholu signálu a rozsahu signálu. Demonstrovali rozdíly v přesnosti detekce záchvatů až o 52 %. Přesnost byla posuzována jako plocha pod křivkou prahové operační charakteristiky (Receiver operating characteristic, ROC). Výsledky zdůrazňují mediánovou rozkládací paměť (median decaying memory, MDM) jako nejlepší současnou metodu normalizace. [1]

Na tento článek reagovali autoři publikace [2]. V článku se autoři také zabývali automatickou detekcí epileptických záchvatů. V této práci porovnávali tři typy normalizace. Byly jimi MDM, novodobá metoda MDM (Novelty-MDM) a fixní základní vyrovnávací paměť (fixed baseline buffer, FB). MDM využívá posuvnou základní vyrovnávací paměť (sliding baseline buffer) elektroencefalografických epoch k výpočtu normalizačních konstant. Tato metoda zahrnuje části EEG bez záchvatu, zahrnuje však také EEG aktivitu, která může mít škodlivý účinek na normalizaci a následnou detekci. Proto byla do této studie zahrnuta také nová metoda Novelty-MDM, která nejprve zabraňuje zahrnování chybných epoch a tím zlepšuje normalizaci příznaků a následnou detekci záchvatů. Normalizované příznaky autoři klasifikovali pomocí metody podpůrných vektorů (support vector machines, SVM), což je metoda strojového učení s učitelem. Autoři zjistili, že metoda MDM nezlepšila celkovou výkonost klasifikátoru

ve srovnání s FB. Nejlepší normalizační metodou se však stala Novelty-MDM, která významně převýšila FB i MDM. [2]

V článku [3] autoři využili Ravenovy pokročilé progresivní matice (Raven's Advance Progressive Matric, RAPM), která je obvykle využívána pro měření interindividuálních rozdílů v kognitivní výkonnosti. Autoři navrhli přístup „rozpoznávání vzoru“, který zahrnoval extrakci příznaků, normalizaci, selekci, klasifikaci a křížovou validaci. Koeficienty byly extrahovány pomocí diskrétní vlnkové transformace (discrete wavelet transform, DWT). Standardizovány byly pomocí Z-skóre na nulovou střední hodnotu s jednotkovou odchylkou. Pro klasifikaci byly využity algoritmy k-nejbližších sousedů (k-nearest neighbours, k-NN), SVM, vícevrstvá neuronová síť (multi-layer perceptron, MLP) a Naïve Bayes. Výsledky byly porovnány s již existujícími kvantitativními metodami k potvrzení výsledků. [3]

1.2 Cíle práce

Cílem této bakalářské práce je vyhodnocení vlivu normalizace na výslednou klasifikaci. EEG záznamy budou předzpracovány v programovém prostředí Wave-Finder pomocí filtrace a segmentace. Ze segmentů bude vypočítáno 23 příznaků, které se využívají pro klasifikaci. Na tyto příznaky budou v programovém prostředí MATLAB[®] implementovány dvě normalizační metody. Využitými normalizačními metodami v této práci jsou Z-skóre a metoda minima a maxima. Bude otestována normalita normalizovaných dat pomocí Kolmogorov-Smirnovova testu a statisticky pomocí Wilcoxonova testu bude zhodnoceno, zda mezi normalizačními metodami existuje rozdíl na vybraném souboru dat. Normalizované hodnoty příznaků budou klasifikovány pomocí klasifikátoru k-means do šesti tříd. Vliv normalizace na data bude vyhodnocen konzultací s expertem a pomocí hodnoty vnitrotřídního rozptylu.

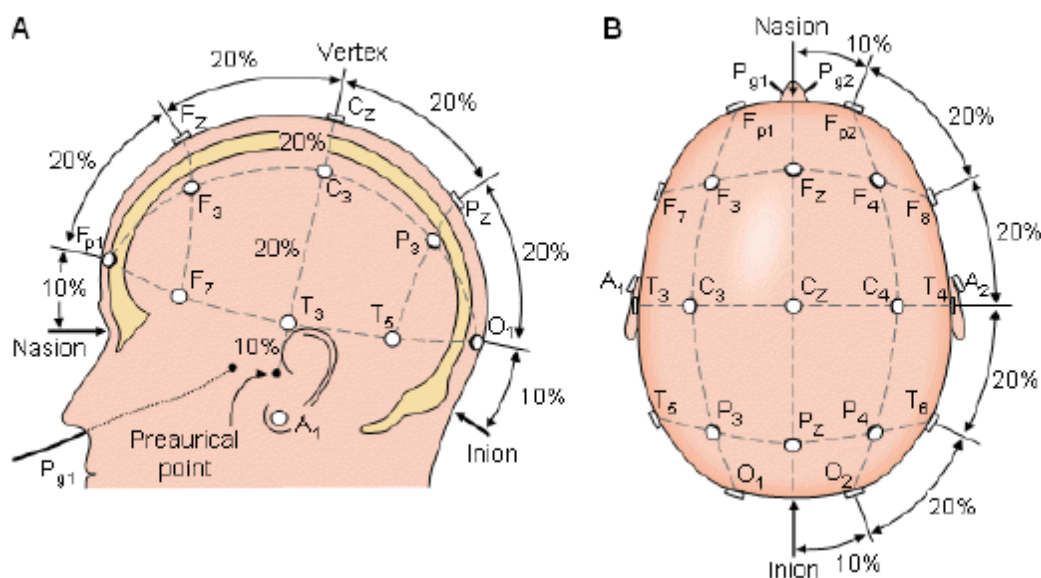
2 Metody

2.1 Elektroencefalografie

První záznam elektrické aktivity mozku byl vytvořen německým psychiatrem Hansem Bergerem v roce 1924, který tento záznam pojmenoval jako elektroencefalogram. Elektroencefalografie (EEG) je metoda měření elektrického potenciálu pomocí elektrod na povrchu hlavy. Měří se spontánní aktivita, jejíž amplituda dosahuje přibližně 100 μV . EEG je neinvazivní diagnostickou metodou. [4; 5]

2.1.1 Snímání signálu

Ke snímání elektrického potenciálu se využívá mezinárodních standardizovaných systémů. Nejčastěji využívaným je systém 10-20. Jedná se o rozložení 21 elektrod na povrchu hlavy. Název tohoto systému je odvozen od rozložení elektrod na povrchu hlavy, které je vidět na Obr. 2.1. [5; 6]



Obr. 2.1: Mezinárodně používaný systém 10-20, pohled na levou hemisféru (A) a pohled shora (B). Zdroj [5].

Elektrody se značí písmenem a číslem. Počáteční písmeno elektrod na obrázku značí oblast umístění elektrody: F – frontální, C – centrální, T – temporální, P – parietální, O – okcipitální. Označení A nesou elektrody referenční, nacházející se na uších. Písmeno Z označuje elektrody ve středu lebky. Elektrody nacházející se v levé hemisféře se značí lichým číslem, elektrody na pravé hemisféře číslem sudým. [7; 8]

Nasnímaný signál se poté zobrazuje pomocí montáží. Unipolární montáž je základní zapojení elektrod vůči referenčním elektrodám A₁ a A₂, které jsou umístěny na uších.

Dalším zapojením je bipolární, při kterém se odečítají hodnoty dvou sousedních elektrod. Další z možností je zapojení se společnou elektrodou (AVG), které vzniká určením aritmetického průměru všech elektrod v daném časovém okamžiku. Tato hodnota je následně odečtena od všech hodnot ve všech kanálech. Již volba zapojení vede k částečné filtraci signálu. [7]

V této práci byla využita poslední ze zmíněných montáží, AVG. Signál z AVG montáže je dynamický a umožňuje pozorovat velikost a tvar vln lépe než bipolární režim, který naopak umožňuje přesnější lokalizaci [8].

2.1.2 Signál EEG

K popisu EEG signálu se využívá frekvenčních spekter. Rozeznáváme čtyři základní frekvenční pásma, která se označují podle písmen řecké abecedy. [9]

Delta rytmus se vyskytuje ve frekvenčním rozmezí od 0,5 do 4,0 Hz. U dospělého člověka v bdělém stavu se jedná o patologický projev. Čím má vlna větší amplitudu a čím je spektrálně čistší, tím její patologický význam nabývá. Delta vlny se vyskytují především v hlubokém spánku, kdy jejich amplituda dosahuje i 100 μV . [9]

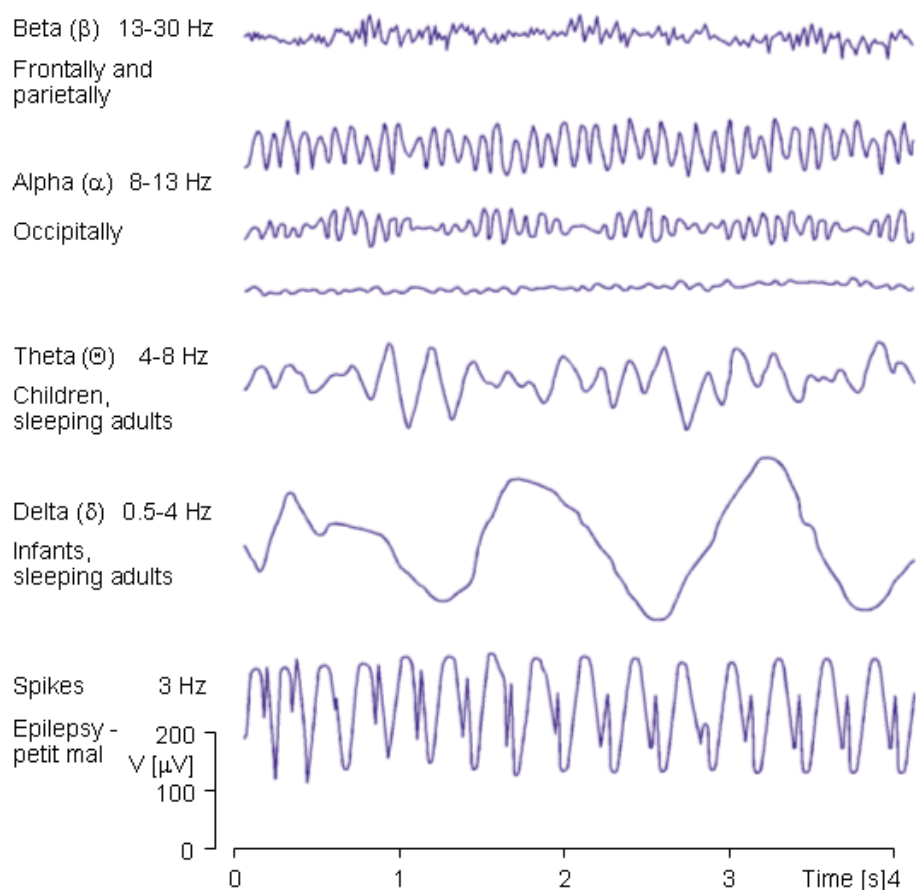
Frekvence theta rytmu se nachází mezi 4,0–8,0 Hz. U zdravých osob se objevuje v centrální, temporální a parietální oblasti. Jestliže je amplituda theta vln alespoň dvakrát vyšší než aktivita alfa (popřípadě 30 μV , jestliže alfa aktivita chybí), hovoříme o patologii. [9]

Alfa rytmus je nejmarkantnějším pásmem ze všech. Jeho frekvenční rozmezí je od 8,0 do 13,0 Hz a amplituda vln bývá 20–50 μV . Nejlépe je vyjádřena alfa aktivita při zavřených očích a relaxaci, potlačuje se otevřením očí a duševní činností (přemýšlení, počítání). Je také charakteristická pro stadium těsně před usnutím. [9; 10]

Frekvence beta rytmu se vyskytuje mezi 13,0–30,0 Hz a převládá nad frontálními laloky. Pásmo bývá charakteristické pro soustředění se na vnější podněty, logicko-analytické myšlení či pro pocity strachu. Amplituda nabývá nejčastěji hodnot 10–30 μV . [9]

Kromě těchto základních rytmů se v záznamu objevují také grafoelementy a artefakty. Grafoelementy jsou typické tvarové charakteristiky. V EEG záznamech se nejčastěji vyskytují v podobě hrotu či komplexu hrot vlna (petit mal), které patří mezi epileptické abnormality. Artefakty mohou být technické či biologické. Mezi technické artefakty se řadí nejčastěji síťový artefakt (v ČR 50 Hz), elektrostatické potenciály nebo impulsní rušení. Biologické artefakty jsou způsobeny například mrkáním či žvýkáním pacienta, pohyby hlavy nebo rušením vlivem činnosti srdce. [7]

Na Obr. 2.2 je znázorněn průběh základních čtyř pásem EEG signálu a vzhled petit mal záchvatu.



Obr. 2.2: Vzhled vln základních EEG pásem, komplex hrot vlna. Zdroj [5].

2.2 Programové prostředí

V práci bylo využito programových prostředí Wave-Finder a MATLAB[®].

Předzpracování dat, segmentace signálu a výpočet příznaků byl proveden v softwaru Wave-Finder (WF). Jedná se o počítačový program realizovaný v jazyce C++, který slouží pro automatické zpracování EEG a byl navržen na neurologickém oddělení v Praze na Bulovce [11].

Ke zpracování dat bylo využito programu MATLAB[®] společnosti MathWorks. Byla použita verze R2017a (akademická licence pro ČVUT).

2.3 Data

Použitá data jsou 19kanálové ambulantní EEG záznamy naměřené v Nemocnici Na Bulovce při standardních podmínkách za použití čepice se systémem elektrod 10-20 s referenčními elektrodami na ušních lalůčkách. Jedná se o šest záznamů pacientů trpících epilepsií. Nahrávky byly měřeny na digitálním systému Brain-Quick (Micromed S.p.A.). Délka nahrávek se pohybovala od 15 do 54 minut. Vzorkovací frekvence záznamů EEG u pacientů s epilepsií je 128 Hz.

Studijní protokol a informovaný souhlas pacienta byly schváleny Etickou komisí Nemocnice Na Bulovce.

2.3.1 Předzpracování

Předzpracování dat je důležité zejména z důvodu odstranění šumu a neúčinné složky signálu. Podstatným krokem předzpracování je filtrace signálu. [12]

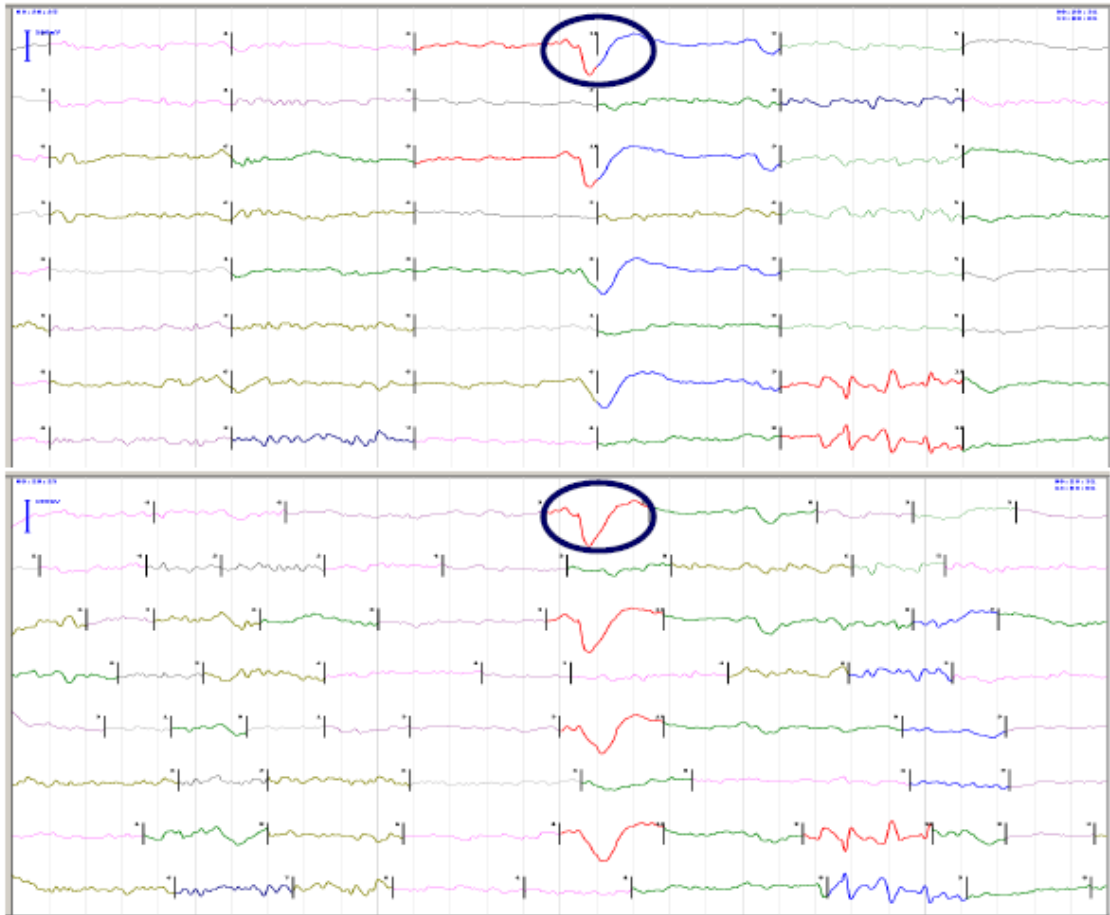
V programovém prostředí WF byla na celý signál použita pásmová propust 0,5 až 30,0 Hz.

2.3.2 Segmentace

EEG signál se z důvodu fyziologických či psychických změn liší. Důsledkem jsou měnící se statické charakteristiky EEG s časem. Z tohoto důvodu je signál považován za nestacionární, což představuje významný problém při jeho analýze. Základním předpokladem použití klasických metod analýzy je stacionární signál. Pomocí správné segmentace lze dosáhnout kvazi-stacionárního signálu. [13]

Existují dva základní typy segmentace. Fixní segmentace, která rozdělí signál do úseků konstantní délky. Hranice těchto úseků nemají žádný vztah k charakteru signálu. Druhým typem je adaptivní segmentace, která signál rozděljuje na základě výskytu nestacionarit na úseky různých délek. [7]

Na Obr. 2.3 je graficky znázorněn princip rozdílu mezi fixní (nahore) a adaptivní segmentací (dole). Na horním obrázku v zakroužkované části je patrné, že fixní segmentace rozdělí signál v pevně stanoveném čase, na rozdíl od adaptivní segmentace vyznačené v zakroužkované části na dolním obrázku.



Obr. 2.3: Princip rozdílu fixní a adaptivní segmentace. Zdroj [7].

Adaptivní segmentace využívá dvou spojených oken, které klouzají po signálu. Pro signál v oknech je odhadnuta míra diference signálu. Jakmile se signál mění, dochází k růstu diference. Pro stacionární signál je míra rozdílu téměř nulová. Maximální hodnota diference je právě na hranicích segmentů. Hranice segmentů jsou umístěny tam, kde je zjištěno lokální maximum míry diference. K eliminaci vlivu nepodstatných kolísání míry rozdílu je využito meze pro segmentaci. Hranice mezi segmenty je detekována pouze v případě, že lokální maximum překročilo tuto mez. Segmentace díky tomuto postupu umožňuje zpracovávání vícekanálových signálů, kdy postupuje v každém kanálu nezávisle na ostatních. [7]

Vzhledem k dalšímu zpracování signálu v této práci (extrakce příznaků), byla v prostředí WF využita vícekanálová adaptivní segmentace. Díky tomuto postupu není do výpočtu příznaků zanesena chyba a pro následnou klasifikaci je výrazně lepší.

2.3.3 Výpočet příznaků

Příznaky jsou charakteristické vlastnosti objektu, jehož hodnoty by měly být stejné pro příznaky ve stejné třídě a odlišné pro příznaky v různých třídách. Tyto příznaky se poté normalizují a předávají do klasifikátoru, který rozhoduje o zařazení objektů do tříd. Optimalizovaná extrakce příznaků je z důvodu přímého ovlivnění výkonu klasifikátoru nezbytná. Kromě toho výraznější příznaky podporují přesnost klasifikátoru. Z těchto důvodů se extrakce příznaků stala nejkritičtější a nejvýraznějším krokem v klasifikaci EEG dat. [3; 14]

V programu WF je implementován výpočet až 24 příznaků [11]. V této práci bylo využito 23 příznaků vypočtených v programu WF. Jejich seznam včetně zkratk a jednotek je vypsán v Tabulce 2.1. Příznaky nejsou v tabulce seřazeny tak, jak je u softwaru WF běžné, ale dle jednotek.

Tabulka 2.1: Seznam použitých příznaků k normalizaci.

Příznak	Zkratka	Jednotka	Popis
1	SIGM	μV	Variabilita signálu
2	APOS	μV	Maximální pozitivní hodnota v daném segmentu
3	ANEG	μV	Maximální negativní hodnota v daném segmentu
4	DELT1	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části delta frekv. pásma (0,5 Hz – 1,5 Hz)
5	DELT2	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části delta frekv. pásma (2,0 Hz – 3,5 Hz)
6	THET1	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části theta frekv. pásma (4,0 Hz – 5,5 Hz)
7	THET2	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části theta frekv. pásma (6,0 Hz – 7,5 Hz)
8	ALPH1	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části alfa frekv. pásma (8,0 Hz – 10,0 Hz)
9	ALPH2	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části alfa frekv. pásma (10,5 Hz – 12,5 Hz)
10	SIGMA	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části beta frekv. pásma (18,0 Hz – 29,0 Hz)
11	BETA	$\mu\text{V}^2/\text{Hz}$	Hodnoty signálu v části beta frekv. pásma (13,5 Hz – 29,0 Hz)
12	MAX1D	$\mu\text{V}/\text{s}$	Maximální hodnota první derivace
13	MD1	$\mu\text{V}/\text{s}$	Průměrná hodnota první derivace
14	MAX2D	$\mu\text{V}/\text{s}^2$	Maximální hodnota druhé derivace
15	MD2	$\mu\text{V}/\text{s}^2$	Průměrná hodnota druhé derivace
16	act	μV^2	Aktivita (1. Hjorthův parametr)
17	NLinE	μV^2	Nelineární energie
18	mf	Hz	Střední frekvence
19	Peaks	Hz	Dominantní spektrální vrchol
20	mob	–	Mobilita (2.Hjorthův parametr)
21	comp	–	Komplexita (3. Hjorthův parametr)
22	LofC	–	Délka křivky
23	ZC	–	Počet průchodů nulou

Všechny příznaky byly spočteny z každého segmentu EEG záznamu. Vztahy pro výpočet jednotlivých příznaků jsou popsány v [7]. Po vypočtení vytvoří jeden příznak n -rozměrný vektor dle počtu segmentů v záznamu.

2.4 Normalizace

V praxi se často stává, že hodnoty příznaků mají různé dynamické rozsahy. Příznaky s vyššími hodnotami mohou mít větší vliv na funkci klasifikátoru než příznaky s nižšími hodnotami. Tento problém bývá řešen využitím normalizace, která způsobuje, že hodnoty všech příznaků leží v podobných rozpětích. [15]

2.4.1 Z-skóre

První normalizační metoda použitá v této práci se nazývá Z-skóre. Přepočtením příznaky nabývají hodnot s nulovou střední hodnotou a jednotkovou standardní odchylkou.

Normalizace příznaků z_i se vypočítá dle vzorce [16]:

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (2.1)$$

kde x_i je i -tý příznak před normalizací, μ je střední hodnota příznaků před normalizací a σ je směrodatná odchylka příznaků před normalizací.

K normalizaci pomocí Z-skóre byla využita implementovaná funkce `zscore` v programovém prostředí MATLAB[®]. Při použití této funkce na matici vzniká matice Z_{norm} normalizovaných hodnot o stejné velikosti. Každý sloupec matice Z byl normalizován na střední hodnotu nula a jednotkovou střední odchylku. [17]

2.4.2 Metoda minima a maxima

Druhou metodou, kterou byla normalizace provedena, je často používaná metoda normalizace minimem a maximem. V tomto případě mají přepočtená data pevný rozsah $\langle 0; 1 \rangle$. Normalizované příznaky X_{i-norm} se počítají dle vzorce [16]:

$$X_{i-norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}}, \quad (2.2)$$

kde X_i je i -tý příznak před normalizací, X_{min} a X_{max} jsou minimální a maximální hodnoty v celé množině dat.

Výstupem byla vzniklá matice normalizovaných hodnot X_{norm} .

2.5 Test normality a statistické vyhodnocení

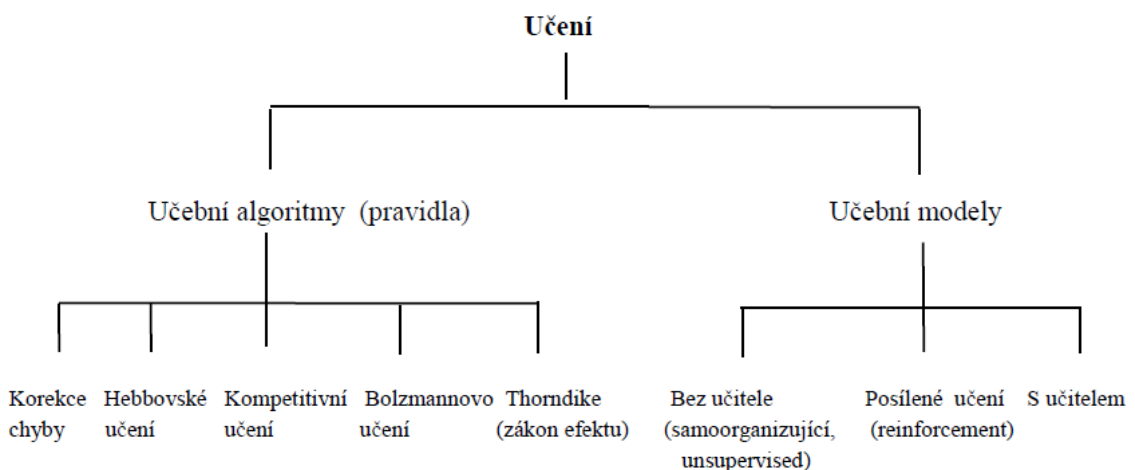
Jelikož některé statistické metody předpokládají normální rozdělení hodnot, bylo potřeba nejprve otestovat normalitu dat. Normalita byla testována na každém příznaku zvlášť. Byl použit jednovýběrový Kolmogorov-Smirnovův test, pro který existuje v programovém prostředí MATLAB[®] implementovaná funkce `kstest`. Předpokladem nulové hypotézy jsou data z normálního rozdělení. Alternativou této hypotézy je, že data z normálního rozložení nepochází. Výsledkem testu jsou hodnoty nula nebo jedna. Při hodnotě

výsledku jedna dochází k zamítnutí nulové hypotézy. Hladina významnosti pro zamítnutí nulové hypotézy je 5 %. [18]

Od testu normality se odvíjel výběr použité statistické metody. Bylo využito neparametrického testu, který nevyžaduje data z normálního rozdělení. Konkrétně byl využit dvouvýběrový Wilcoxonův test, který má v programu MATLAB[®] implementovanou funkci *signrank*. Nulová hypotéza předpokládá, že jsou data obou výběrů totožná (jejich medián je nula). Alternativa proti této hypotéze je, že data nejsou totožná. Výsledkem testu jsou parametry P a H . Hodnota parametru P je v rozsahu $<0; 1>$ a udává pravděpodobnost, se kterou zamítáme nulovou hypotézu. Parametrem H jsou, stejně jako u testu normality, hodnoty nula a jedna. Pokud se $H = 1$, dochází k zamítnutí nulové hypotézy. Hladina významnosti pro zamítnutí nulové hypotézy je 5 %. [19]

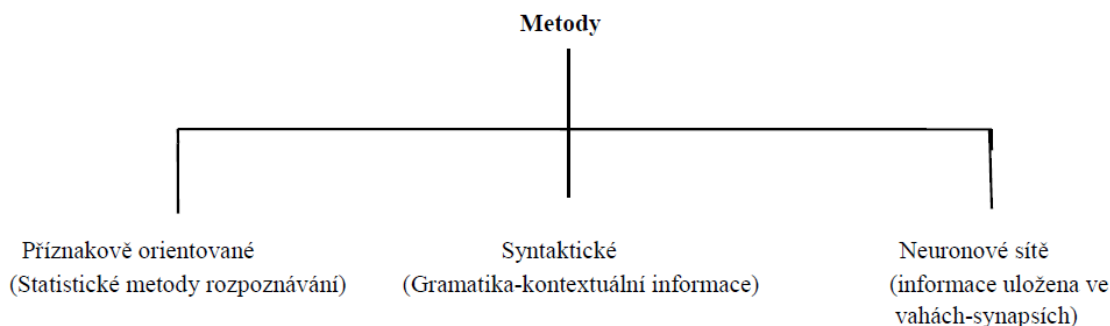
2.6 Klasifikace

Klasifikátory jsou v podstatě funkce, které berou hodnoty různých příznaků ze vzorku a na základě nezávislých proměnných určují třídu, ke které příznak patří [20]. Pojem, který s klasifikací úzce souvisí, je pojem učení. Jejich strukturované rozdělení na učební algoritmy a modely je zobrazeno na Obr. 2.4. Zejména důležité jsou pro nás pojmy učení s učitelem a bez učitele. V případě učení s učitelem dochází k trénování klasifikátoru z vnějšku pomocí etalonů. Tzv. učitel zde zadává vstupní a výstupní data, provádí na nich změny i samotné vyhodnocení. Při využití učení bez učitele systém nemá vnější podporu. Klasifikace dat je založena pouze na informacích, které sít' získá během samotného procesu učení. [7]



Obr. 2.4: Metody učení. Zdroj [7].

Metody klasifikace se rozdělují do tří základních tříd, jejichž strukturované rozdělení je zobrazeno na Obr. 2.5. Mezi způsoby rozdělení patří příznakově orientované metody, syntaktické metody a neuronové sítě. [7]



Obr. 2.5: Metody klasifikace. Zdroj [7].

Mezi příznakově orientované metody učení bez učitele se řadí shluková analýza (Cluster Analysis). K identifikaci zkoumaných objektů do tříd využívá míry podobnosti (resp. nepodobnosti) mezi individuálními pozorováními a hledá přirozenou strukturu dat. Míra podobnosti je určena pomocí metriky, tedy vzdálenosti mezi vzorem a třídami. Nejběžněji používanou vzdáleností shlukové analýzy je Euklidovská vzdálenost. [7; 21]

V Euklidovské vzdálenosti pro dva vektory A a B v kartézském souřadném systému platí vztah [7]:

$$E = \sqrt{\sum_{i=1}^N (A(i) - B(i))^2}, \quad (2.3)$$

kde N je dimenze prostoru.

Metody shlukové analýzy lze rozdělit na hierarchické a nehierarchické. Hierarchické metody jsou založené na přetváření matice vzdálenosti do posloupnosti hierarchicky seřazených rozdělení. Jejich grafické znázornění se nazývá dendrogram. [7]

V této práci je využita nehierarchická metoda shlukové analýzy, kterou je k-means.

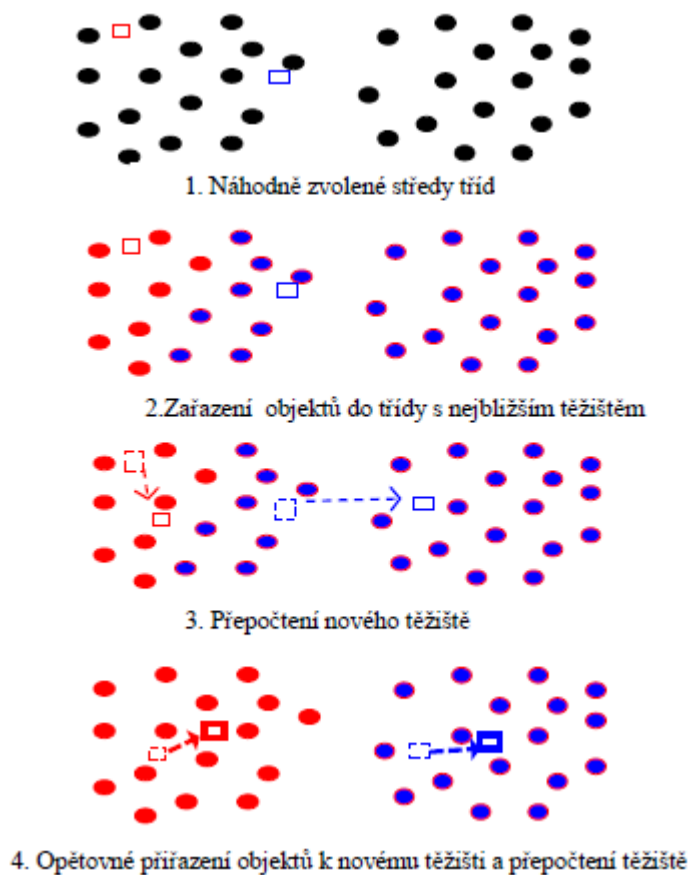
2.6.1 K-means

Jedná se o jednu z nejjednodušších používaných metod pro klasifikaci příznaků, při níž jsou vztahy mezi body počítány pomocí vzdálenosti. Jejím cílem je vytvoření co nejpodobnějších shluků, které jsou současně od ostatních shluků dostatečně vzdálené. [7]

Postup při klasifikaci [7]:

1. Počáteční rozdělení objektů do K shluků.
2. Přiřazení každého zbývajících objektu do shluku s nejbližším těžištěm.
3. Přepočítání těžišť tříd, jež se změnily po přiřazení nových objektů.
4. Opakování kroku 2 dokud celý cyklus přes všechna data nezaznamená změnu členů tříd.

Na Obr. 2.6 je znázorněn graficky celý postup, jak se tvoří shluky při použití metody k-means.



Obr. 2.6: Princip metody k-means. Zdroj [7].

V programovém prostředí MATLAB[®] byla použita funkce *kmeans* na normalizovaná data pomocí obou normalizačních metod.

Vstupním parametrem testu jsou normalizovaná data (Z_{norm} a X_{norm}) a parametr K , což je počet tříd, do kterých chceme data roztrždit. V této práci bylo použito rozčlenění příznaků do šesti tříd. Výstupními parametry testu jsou hodnoty idx , C , $sumd$ a D . Parametr idx je vektor hodnot o stejném počtu řádků, kolik bylo segmentů. Udává, do které třídy byl segment zařazen. Parametrem C je matice udávající umístění těžiště daných tříd. $Sumd$ je parametr vnitrotřídních rozptylů, který udává sumu vzdáleností

všech bodů jedné třídy ke svému těžišti. Parametrem D je matice udávající vzdálenost každého bodu ke každému těžišti. [22]

2.7 Vyhodnocení klasifikace

Klasifikace obou normalizačních metod byla vyhodnocena pomocí parametru vnitrotřídního rozptylu $sumd$ vypočteného v softwaru MATLAB[®]. Jelikož hodnoty normalizované pomocí Z -skóre dosahují jiných rozsahů než hodnoty normalizované minimem a maximem, bylo potřeba ještě normalizovat hodnoty $sumd$, aby byla při porovnání jeho hodnota vypovídající.

K normalizaci byl využit parametr D , resp. nejvzdálenější hodnota příznaku v jedné třídě. Normalizace hodnoty $sumd$ jedné třídy byla vydělena nejvzdálenější (maximální) hodnotou vzdálenosti D pro stejnou třídu. Výsledkem byl vektor $normsumd$ o stejné velikosti, jakou měl původní vektor $sumd$.

3 Výsledky

3.1 Předzpracování dat

Data byla v programovém prostředí WF předzpracována pomocí filtrace. Bylo využito pásmové propusti 0,5 až 30,0 Hz. Záznamy byly rozčleněny pomocí vícekanálové adaptivní segmentace. Počet segmentů pro každý záznam je zapsán v Tabulce 3.1.

Tabulka 3.1: Počet segmentovaných částí v záznamech.

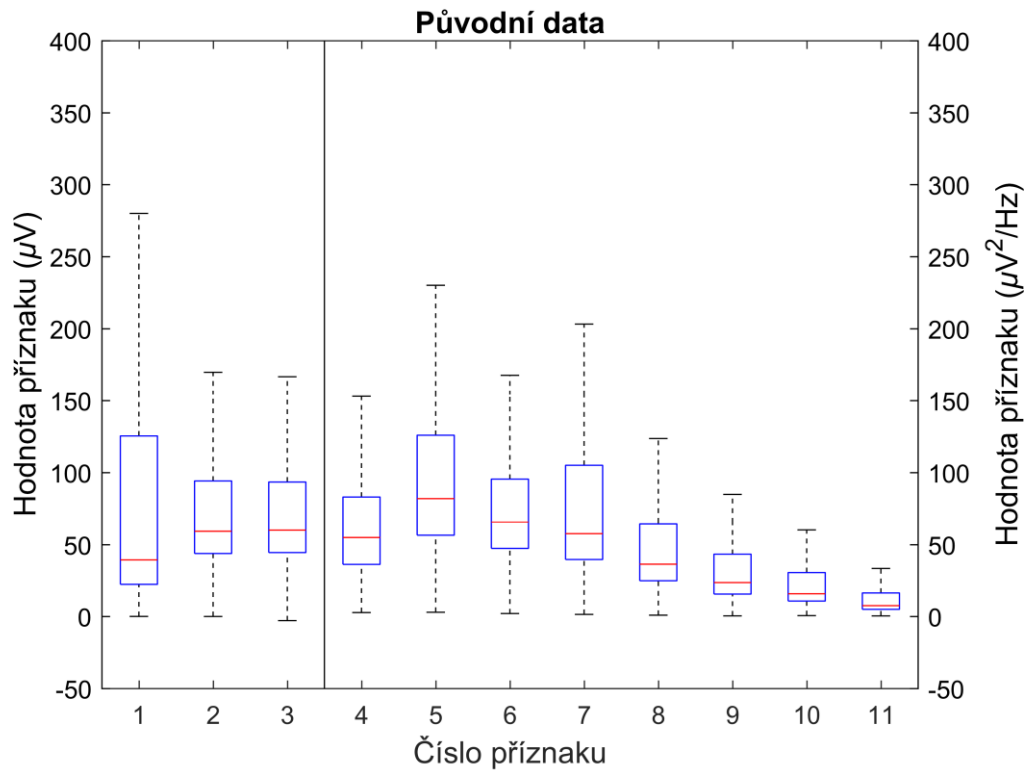
Číslo záznamu	Délka záznamu (min)	Počet segmentů (-)
1	54:13	14221
2	14:57	7241
3	32:03	15975
4	30:54	21776
5	14:51	9806
6	40:33	22007

Z každého segmentu bylo pomocí softwaru Wave-Finder vypočteno 23 příznaků vypsaných v Tabulce 2.1. Výsledkem každého záznamu byla matice o velikosti počtu segmentů \times počet příznaků (23).

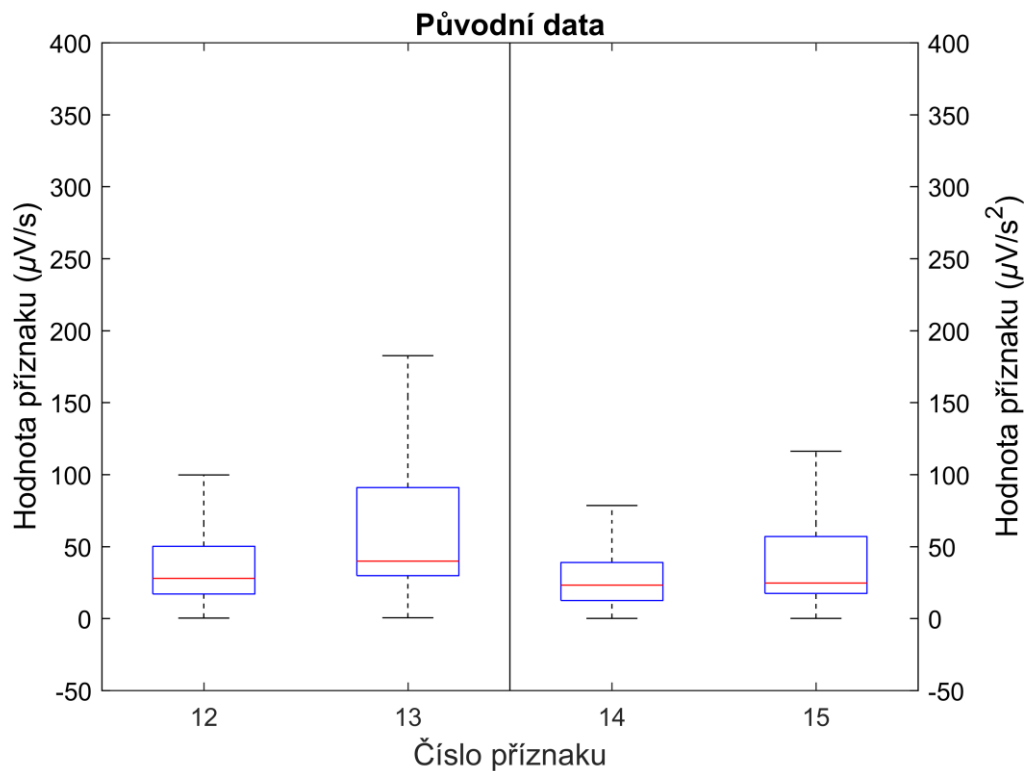
Pro zobrazení hodnot, kterých příznaky nabývaly, bylo využito boxplot zobrazení. Vzhledem k počtu příznaků a jejich různým jednotkám bylo nutné rozdělit jejich boxplot zobrazení do čtyř grafů pro lepší přehlednost. Na ukázkou byl použit první záznam, jehož prvních 11 příznaků je zobrazeno na Obr. 3.1. Příznaky 12–15 lze pozorovat na Obr. 3.2 a příznaky 16–19 na Obr. 3.3. Poslední příznaky 20–23 jsou zobrazeny na Obr. 3.4. Ve všech grafech je využito stejných rozsahů, aby nedošlo ke zkreslení.

Krabicový graf neboli boxplot obsahuje box s dolním a horním ohraničením, které představují první a třetí kvartil hodnot. Uprostřed boxu se nachází linie označující medián hodnot. Vousy, linie vycházející z boxu kolmo nahoru a dolů, představují body, které nejsou považovány za odlehlé. Odlehlé body, tzv. outliers, jsou v grafech vyznačeny jednotlivě pomocí symbolu „+“. [23]

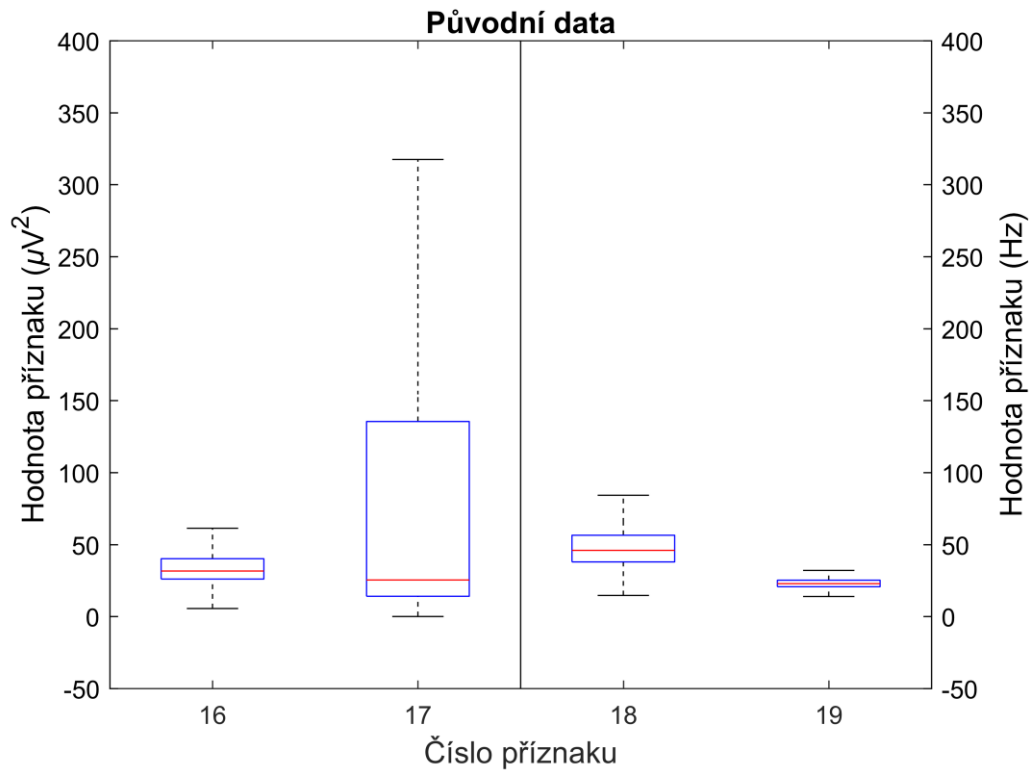
Pro následující Obr. 3.1–3.4 byly odlehlé body odstraněny z důvodu lepší přehlednosti grafů. Původní obrázky i s odlehlými hodnotami jsou v Příloze B. Pro ukázkou je na Obr. 3.5 znázorněn rozdíl při zobrazování boxplot grafů s odlehlými hodnotami a bez odlehlých hodnot.



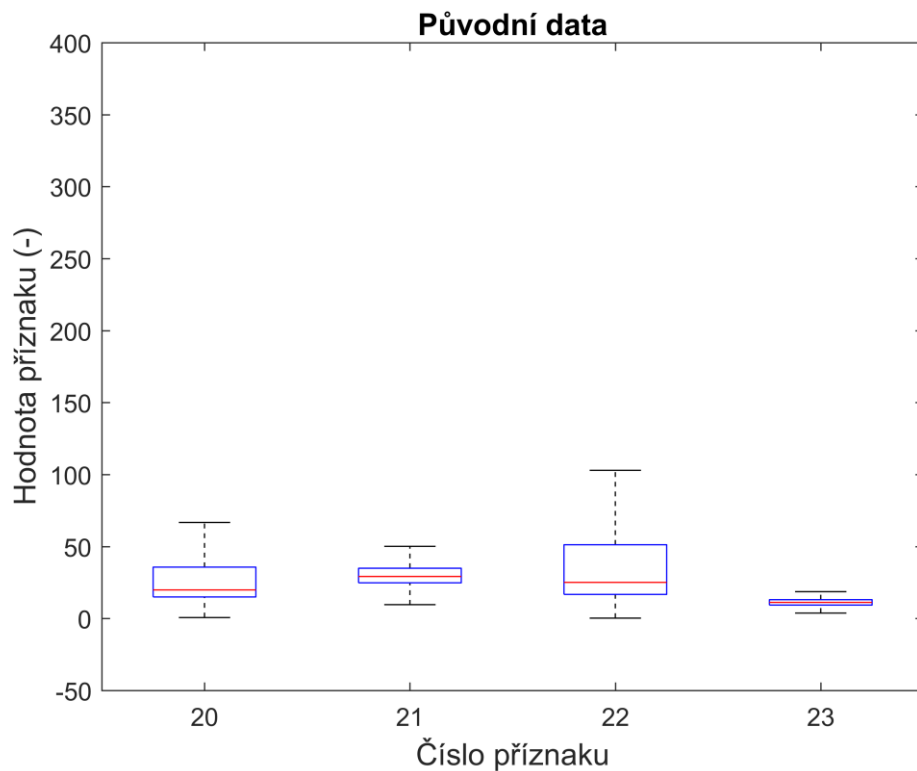
Obr. 3.1: Boxplot zobrazení prvních 11 extrahovaných příznaků z prvního záznamu bez odlehlých hodnot, příznaky jsou seřazeny dle Tabulky 2.1.



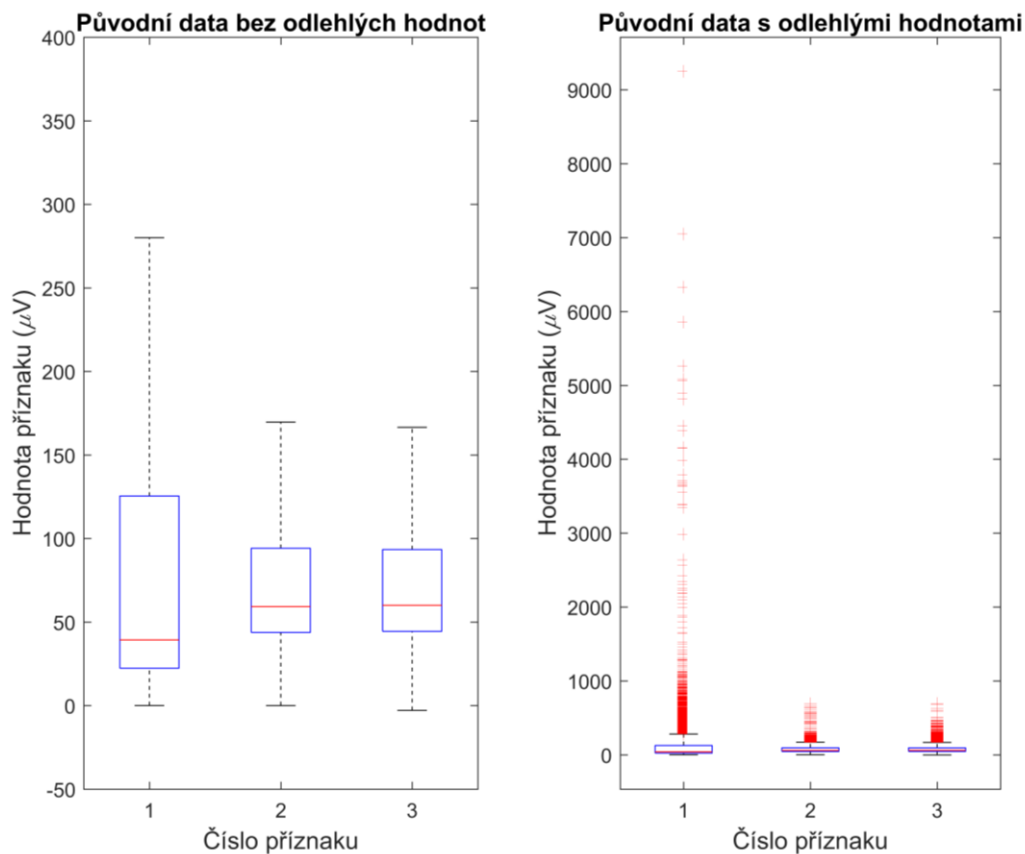
Obr. 3.2: Boxplot zobrazení extrahovaných příznaků 12–15 z prvního záznamu bez odlehlých hodnot, příznaky jsou seřazeny dle Tabulky 2.1.



Obr. 3.3: Boxplot zobrazení extrahovaných příznaků 16–19 z prvního záznamu bez odlehlých hodnot, příznaky jsou seřazeny dle Tabulky 2.1.



Obr. 3.4: Boxplot zobrazení extrahovaných příznaků 20–23 z prvního záznamu bez odlehlých hodnot, příznaky jsou seřazeny dle Tabulky 2.1.

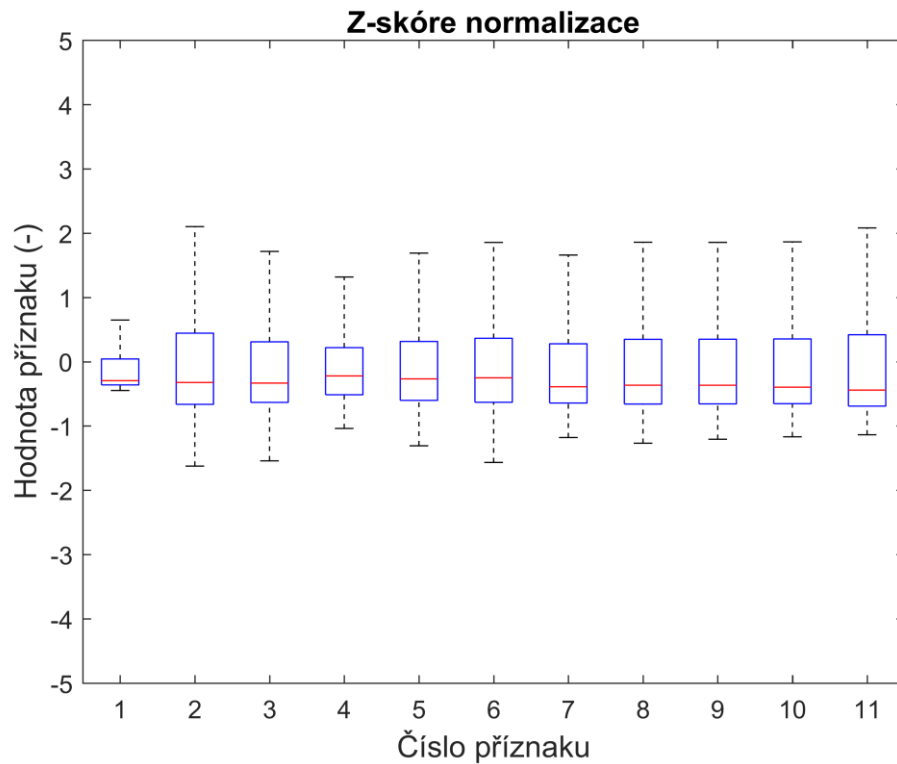


Obr. 3.5: Znázornění rozdílu zobrazení původních hodnot bez odlehlých příznaků a s odlehlými příznaky, příznaky jsou seřazeny dle Tabulky 2.1.

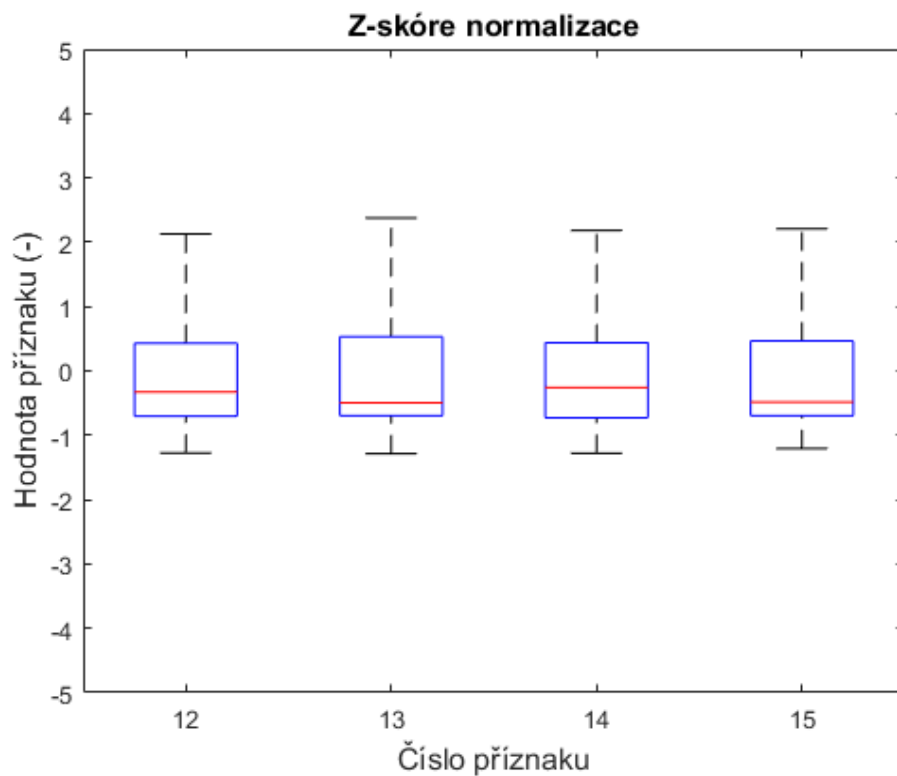
3.2 Normalizace dat

Nejprve bylo k eliminování extrémních hodnot příznaků využito normalizace pomocí Z-skóre, kdy na data byla aplikována již implementovaná funkce v softwaru MATLAB®. Normalizace byla použita na všechny záznamy.

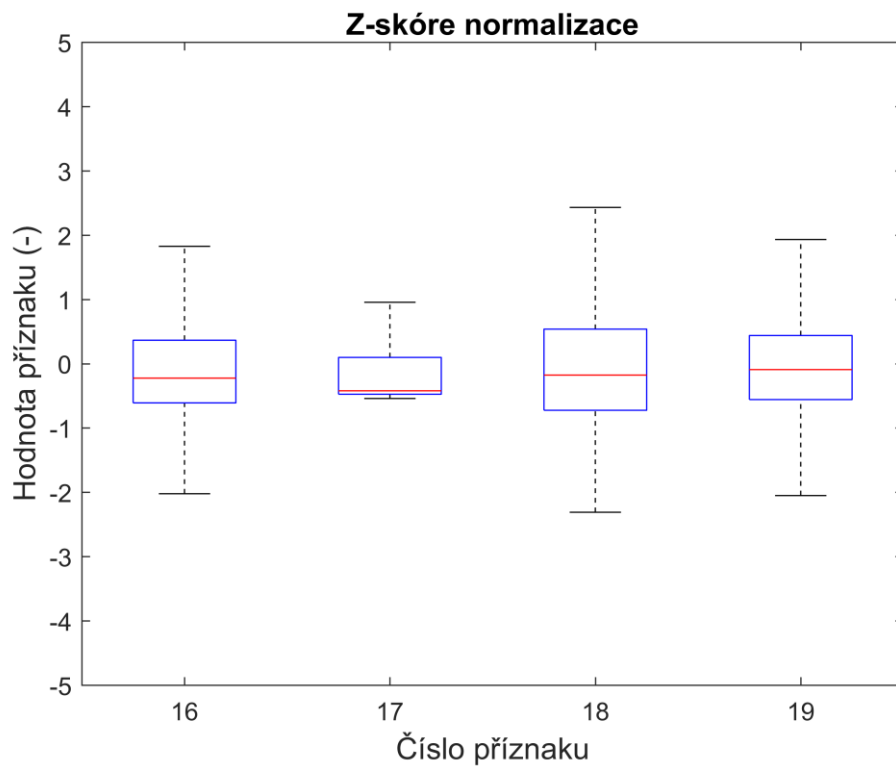
Ze stejného důvodu jako u původních hodnot v kapitole 3.1 bylo i zde využito boxplot zobrazení ve čtyřech grafech. Hodnoty všech grafů opět dosahují stejných rozsahů, aby nedošlo ke zkreslení záznamů. K zobrazení zde bylo znovu využito prvního záznamu. Na Obr. 3.5 jsou zobrazeny rozsahy prvních 11 příznaků normalizovaných pomocí metody Z-skóre. Normalizované příznaky 12–15 lze pozorovat na Obr. 3.6 a další normalizované příznaky 16–19 na Obr. 3.7. Poslední čtyři normalizované příznaky 20–23 jsou zobrazeny na Obr. 3.8.



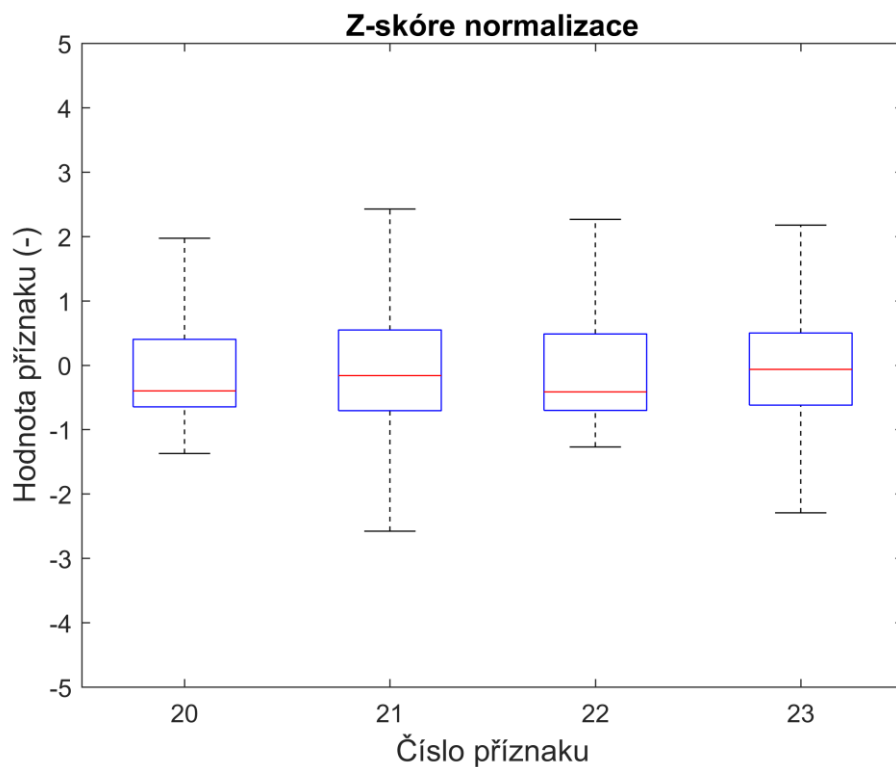
Obr. 3.6: Boxplot zobrazení prvních 11 příznaků prvního záznamu, normalizované pomocí Z-skóre, čísla příznaků jsou seřazena dle Tabulky 2.1.



Obr. 3.7: Boxplot zobrazení příznaků 12–15 prvního záznamu, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1.

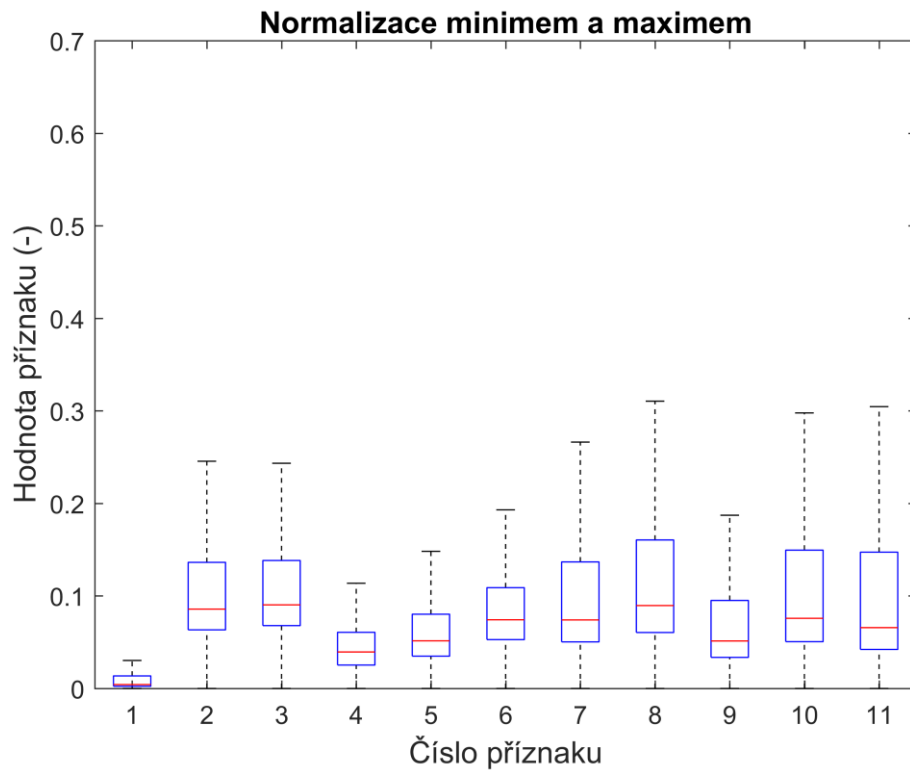


Obr. 3.8: Boxplot zobrazení příznaků 16–19 prvního záznamu, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1.

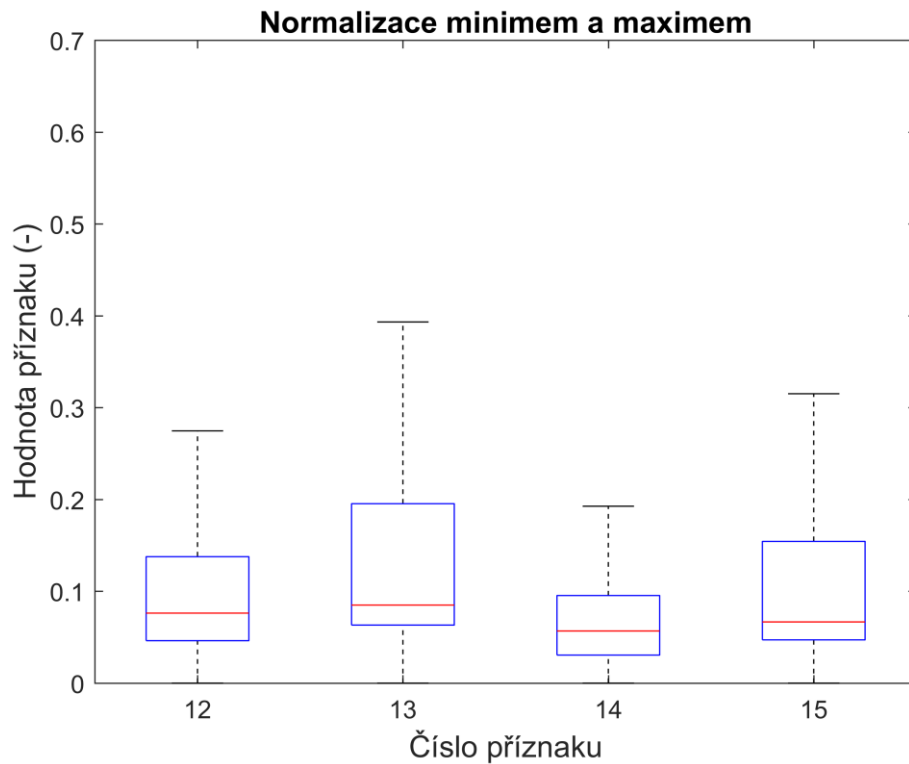


Obr. 3.9: Boxplot zobrazení příznaků 20–23 prvního záznamu, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1.

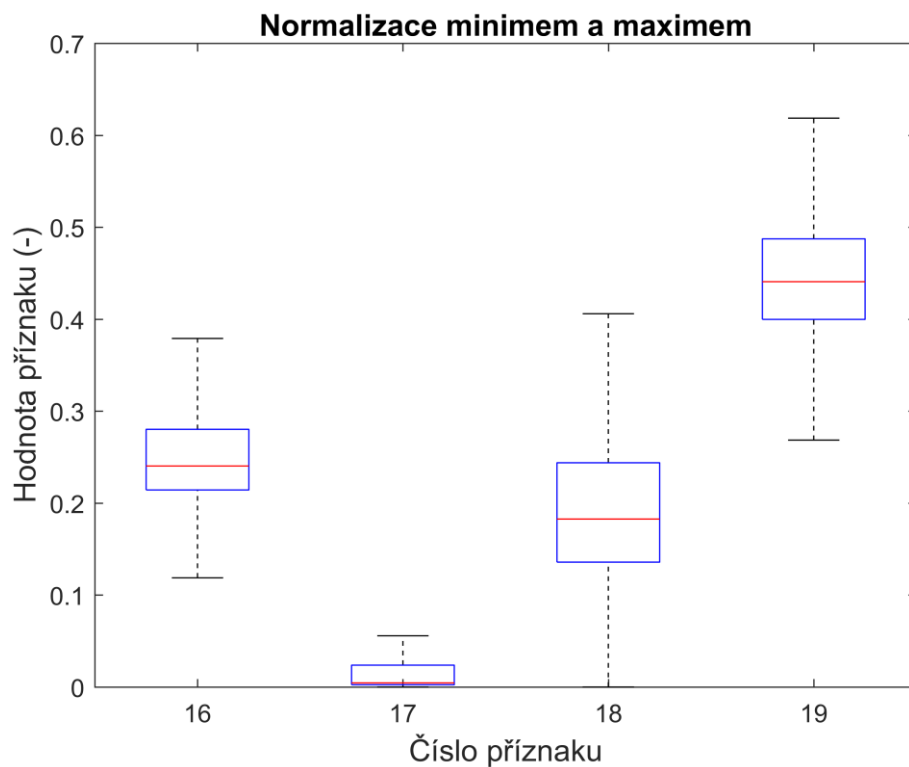
Pro normalizaci minimem a maximem byla v programovém prostředí MATLAB[®] vytvořena vlastní funkce *MMnorm* dle vzorce 2.2. I tato funkce byla použita na všechny záznamy a na Obr. 3.9, Obr. 3.10, Obr. 3.11 a Obr. 3.12 je pomocí krabicových grafů zobrazeno rozložení příznaků prvního záznamu po normalizaci minimem a maximem.



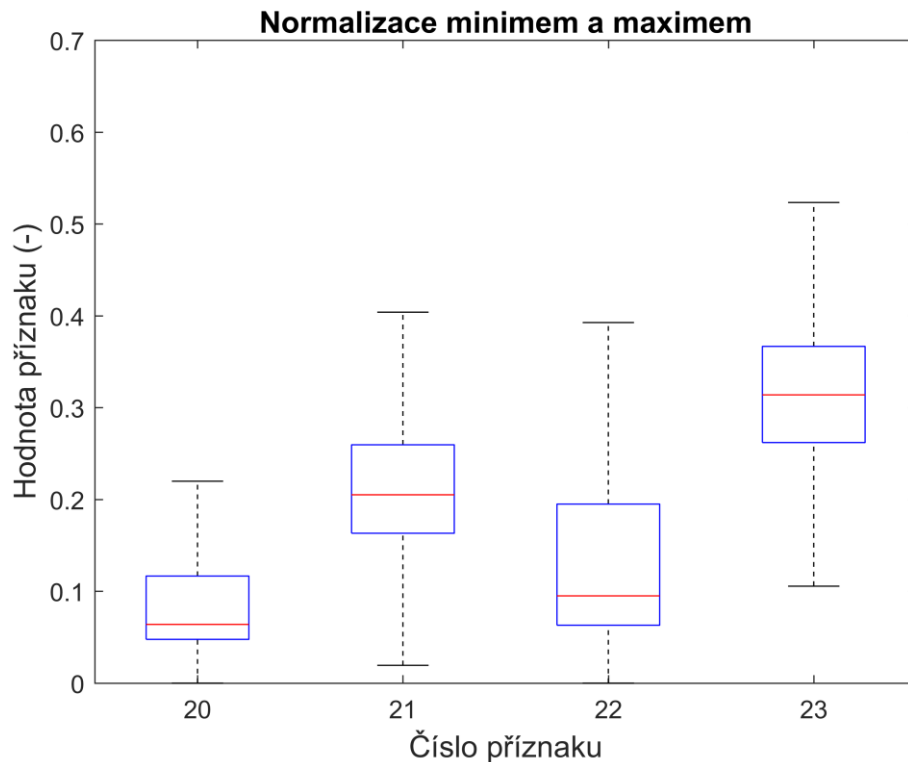
Obr. 3.10: Boxplot zobrazení prvních 11 příznaků prvního záznamu, normalizované pomocí metody minima a maxima, příznaky jsou seřazeny dle Tabulky 2.1.



Obr. 3.11: Boxplot zobrazení příznaků 12–15 prvního záznamu, normalizované pomocí metody minima a maxima, příznaky jsou seřazeny dle Tabulky 2.1.



Obr. 3.12: Boxplot zobrazení příznaků 16–19 prvního záznamu, normalizované pomocí metody minima a maxima, příznaky jsou seřazeny dle Tabulky 2.1.

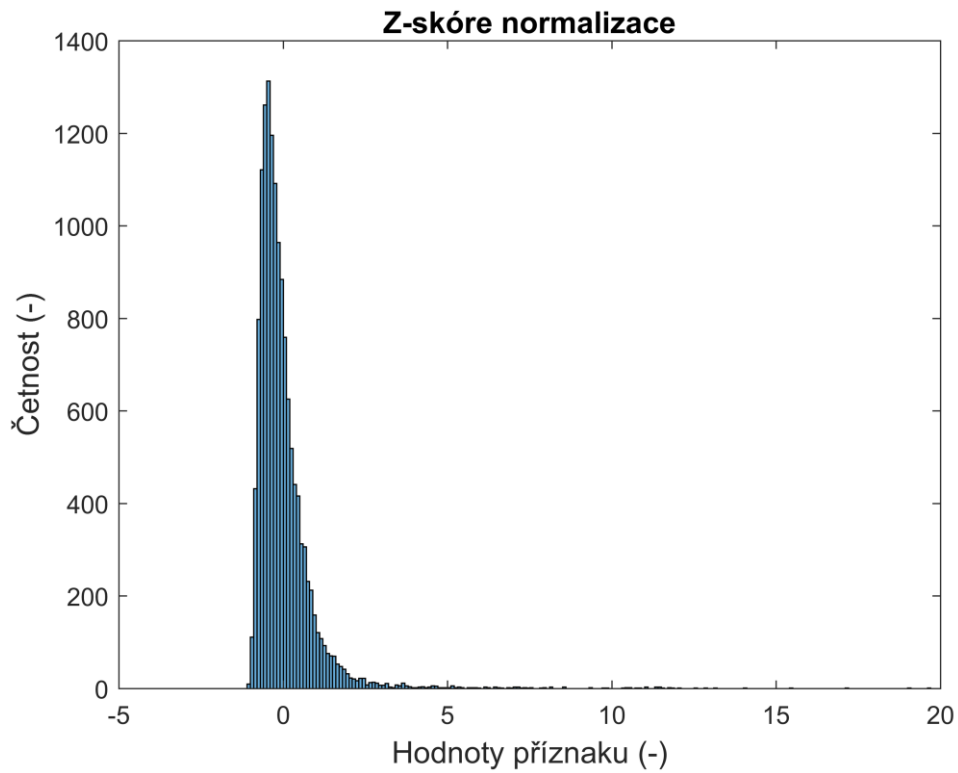


Obr. 3.13: Boxplot zobrazení příznaků 20–23 prvního záznamu, normalizované pomocí metody minima a maxima, příznaky jsou seřazeny dle Tabulky 2.1.

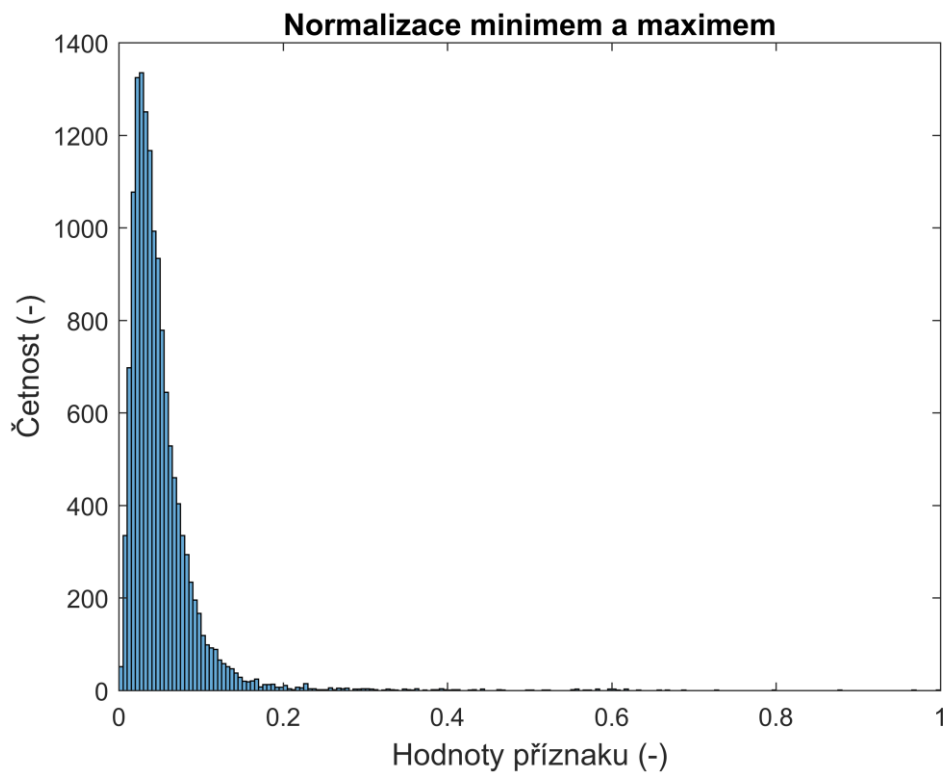
3.3 Normalita dat

Normalita dat byla testována u všech záznamů na obou normalizačních metodách. Každý příznak v záznamu byl testován zvlášť pomocí Kolmogorov-Smirnovova testu. U všech záznamů i obou normalizačních metod byla zamítnuta nulová hypotéza testu u všech příznaků na hladině významnosti 5 %. Výsledkem testu tedy bylo, že hodnoty normalizované pomocí obou normalizačních metod nepochází z normálního rozdělení.

Že hodnoty nepochází z normálního rozdělení, bylo ověřeno také graficky. V programu MATLAB[®] bylo využito zobrazení histogramu, který graficky znázorňuje četnost veličin, resp. jejich rozložení. Na Obr. 3.7 a Obr. 3.8 jsou zobrazeny histogramy obou normalizačních metod. V obou případech se jedná o histogram příznaku *DELTI* z prvního záznamu.



Obr. 3.14: Histogram příznaku DELT1 prvního záznamu, normalizace pomocí Z-skóre.



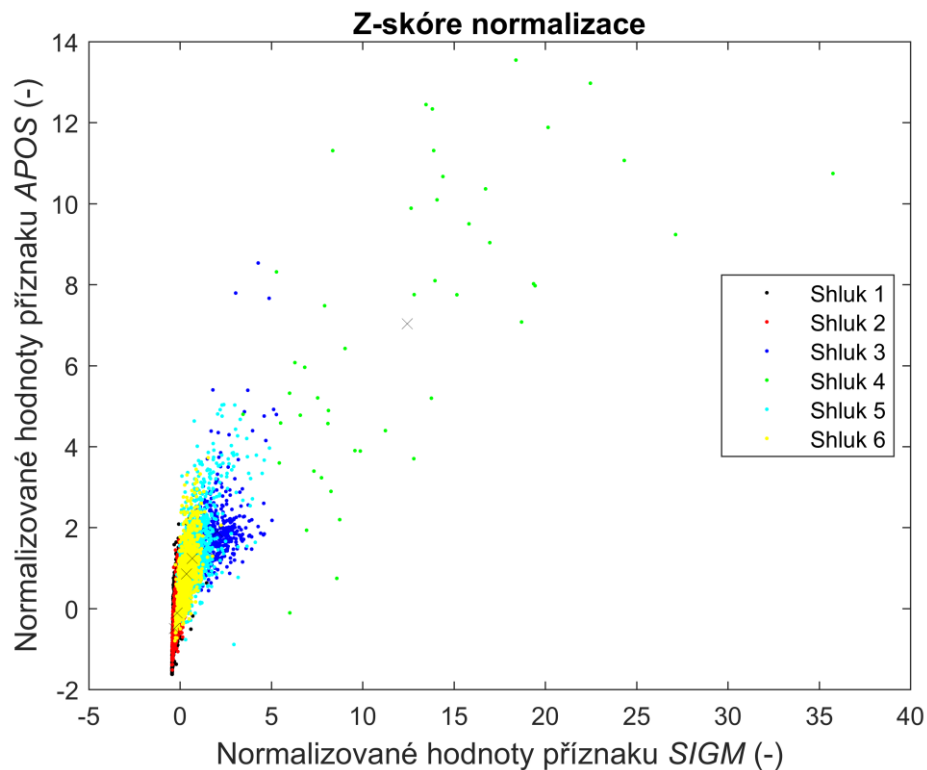
Obr. 3.15: Normalizace příznaku DELT1 prvního záznamu, normalizace minimem a maximem.

3.4 Statistické zhodnocení

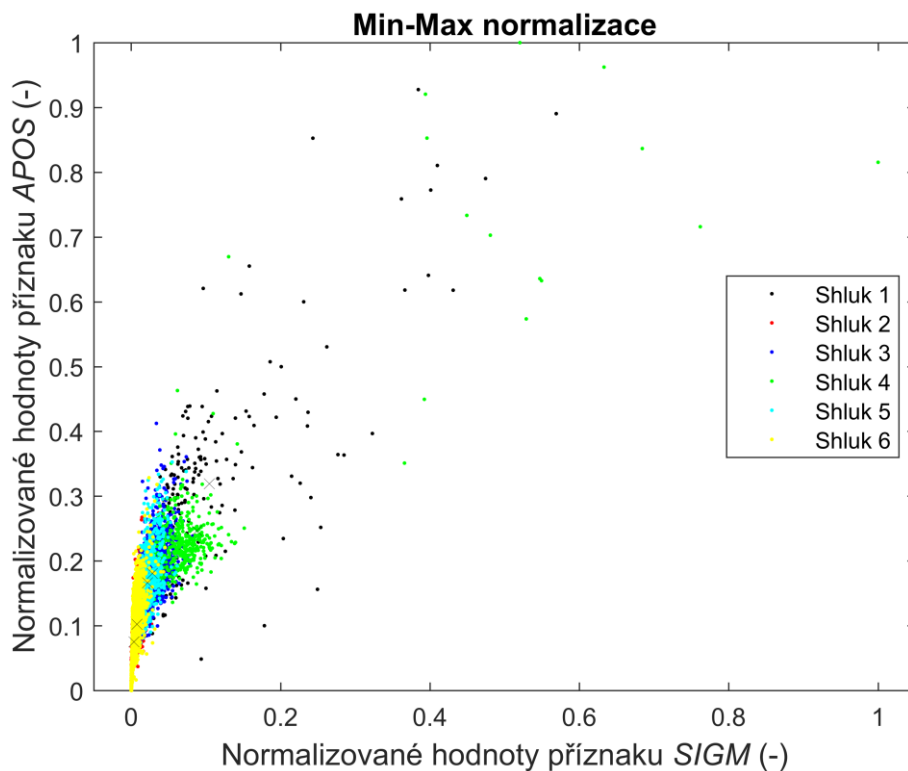
Rozdíl v datových souborech byl testován pomocí neparametrického Wilcoxonova testu. Byly testovány vždy stejné příznaky normalizované pomocí obou metod. U všech příznaků všech záznamů byla zamítnuta nulová hypotéza na hladině významnosti 5 %. Výsledkem testu tedy bylo, že se datové soubory normalizované pomocí dvou normalizačních metod Z-skóre a metodou minima a maxima statisticky liší.

3.5 Klasifikace

Data normalizovaná oběma metodami byla klasifikována pomocí k-means do šesti shluků. Na Obr. 3.16 je zobrazena klasifikace dat normalizovaných pomocí Z-skóre. Na Obr. 3.17 lze pozorovat klasifikaci dat normalizovaných pomocí metody minima a maxima. V obou případech se jedná o klasifikaci prvního záznamu. Jelikož normalizované hodnoty v obou případech dosahují jiných hodnot, jsou i rozsahy obou grafů rozdílné. V případě sjednocení rozsahů by však došlo ke znehodnocení Obr. 3.17. Z toho důvodu byly rozsahy ponechány v původním měřítku.



Obr. 3.16: K-means klasifikace příznaků normalizovaných pomocí Z-skóre.



Obr. 3.17: K-means klasifikace příznaků normalizovaných pomocí metody minima a maxima.

3.6 Vyhodnocení klasifikace

Shluky vytvořené klasifikátorem k-means v softwaru MATLAB® pro data normalizovaná pomocí obou metod si však nemusí odpovídat (viz. Obr. 3.16 a 3.17). Z tohoto důvodu bylo pro každý záznam testováno, jaké shluky si odpovídají po normalizacích oběma metodami. V programu MATLAB® byla vybrána zvlášť každá třída po klasifikaci normalizační metody Z-skóre a ke každé byly přiřazeny do ní spadající segmenty. Pro tyto segmenty byly zjištěny odpovídající hodnoty tříd při klasifikaci normalizační metody minima a maxima. V Tabulkách 3.2–3.7 je vždy vypsán celkový počet segmentů vyskytujících se v třídách po klasifikaci normalizační metody Z-skóre a k nim vypsané zastoupení odpovídajících segmentů v třídách 1–6 při klasifikaci normalizační metody minima a maxima. Tabulky jsou vytvořeny pro všechny záznamy.

Tabulka 3.2: Odpovídající si segmenty po klasifikaci normalizačními metodami, první záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – počet segmentů (-)					
Třída	Počet segmentů (-)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	5441	5344	0	95	0	2	0
Třída 2	1896	13	2	309	0	403	1169
Třída 3	1521	48	98	2	24	1313	36
Třída 4	48	0	0	0	48	0	0
Třída 5	4374	1075	0	3277	0	3	19
Třída 6	621	0	593	0	11	5	12

Tabulka 3.3: Odpovídající si segmenty po klasifikaci normaliz. metodami, druhý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – počet segmentů (-)					
Třída	Počet segmentů (-)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	1707	319	471	22	0	886	9
Třída 2	621	0	473	147	0	0	1
Třída 3	2984	1768	0	0	0	1175	41
Třída 4	48	0	0	0	48	0	0
Třída 5	441	0	2	3	5	0	431
Třída 6	1857	0	34	1124	0	688	11

Tabulka 3.4: Odpovídající si segmenty po klasifikaci normalizačními metodami, třetí záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – počet segmentů (-)					
Třída	Počet segmentů (-)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	4485	2795	381	78	0	866	365
Třída 2	3837	0	2652	129	0	1055	1
Třída 3	840	615	0	24	0	201	41
Třída 4	6970	18	1425	777	0	0	4750
Třída 5	973	5	0	551	413	0	4
Třída 6	17	2	0	0	2	13	0

Tabulka 3.5: Odpovídající si segmenty po klasifikaci normaliz. metodami, čtvrtý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – počet segmentů (-)					
Třída	Počet segmentů (-)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	7062	0	3966	0	356	0	2740
Třída 2	692	111	0	580	0	1	0
Třída 3	2380	43	0	213	0	2124	0
Třída 4	2197	207	16	0	1796	178	0
Třída 5	3412	0	2487	1	91	833	0
Třída 6	5999	0	0	0	677	0	5322

Tabulka 3.6: Odpovídající si segmenty po klasifikaci normalizačními metodami, pátý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – počet segmentů (-)					
Třída	Počet segmentů (-)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	2615	0	0	17	871	32	1695
Třída 2	738	32	0	688	11	0	7
Třída 3	4457	2786	0	5	1663	0	0
Třída 4	165	0	162	3	0	0	0
Třída 5	1140	90	0	19	830	167	34
Třída 6	518	0	29	7	0	421	61

Tabulka 3.7: Odpovídající si segmenty po klasifikaci norm. metodami, šestý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – počet segmentů (-)					
Třída	Počet segmentů (-)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	9678	7641	0	0	249	1210	578
Třída 2	1622	0	0	779	1	196	646
Třída 3	3273	1	0	569	0	1941	762
Třída 4	990	0	3	0	897	77	13
Třída 5	83	0	83	0	0	0	0
Třída 6	7240	3674	0	0	393	441	2732

Tabulky 3.8–3.13 byly vytvořeny obdobným způsobem jako předchozích šest tabulek. Jedná se o procentuální vyjádření zastoupení jednotlivých tříd minima a maxima v pevně stanovené třídě Z-skóre. V tabulkách jsou také barevně vyznačeny třídy po klasifikaci metodou minima a maxima s nejvyšší procentuální shodou s třídou po klasifikaci metody Z-skóre.

Tabulka 3.8: Procentuální vyjádření odpovídajících si tříd obou norm. metod, první záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – procentuální vyjádření (%)					
Třída	Procentuální zastoupení (%)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	100,00	98,22	0,00	1,75	0,00	0,04	0,00
Třída 2	100,00	0,69	0,11	16,30	0,00	21,26	61,66
Třída 3	100,00	3,16	6,44	0,13	1,58	86,32	2,37
Třída 4	100,00	0,00	0,00	0,00	100,00	0,00	0,00
Třída 5	100,00	24,58	0,00	74,92	0,00	0,07	0,43
Třída 6	100,00	0,00	95,49	0,00	1,77	0,81	1,93

Tabulka 3.9: Procentuální vyjádření odpovídajících si tříd obou norm. metod, druhý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – procentuální vyjádření (%)					
Třída	Procentuální zastoupení (%)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	100,00	18,69	27,59	1,29	0,00	51,90	0,53
Třída 2	100,00	0,00	76,17	23,67	0,00	0,00	0,16
Třída 3	100,00	59,25	0,00	0,00	0,00	39,38	1,37
Třída 4	100,00	0,00	0,00	0,00	100,00	0,00	0,00
Třída 5	100,00	0,00	0,45	0,68	1,13	0,00	97,73
Třída 6	100,00	0,00	1,83	60,53	0,00	37,05	0,59

Tabulka 3.10: Procentuální vyjádření odpovídajících si tříd obou norm. metod, třetí záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – procentuální vyjádření (%)					
Třída	Procentuální zastoupení (%)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	100,00	62,32	8,49	1,74	0,00	19,31	8,14
Třída 2	100,00	0,00	69,12	3,36	0,00	27,50	0,03
Třída 3	100,00	73,21	0,00	2,86	0,00	23,93	0,00
Třída 4	100,00	0,26	20,44	11,15	0,00	0,00	68,15
Třída 5	100,00	0,51	0,00	56,63	42,45	0,00	0,41
Třída 6	100,00	11,76	0,00	0,00	11,76	76,47	0,00

Tabulka 3.11: Procentuální vyjádření odpovídajících si tříd obou norm. metod, čtvrtý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – procentuální vyjádření (%)					
Třída	Procentuální zastoupení (%)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	100,00	0,00	56,16	0,00	5,04	0,00	38,80
Třída 2	100,00	16,04	0,00	83,82	0,00	0,14	0,00
Třída 3	100,00	1,81	0,00	8,95	0,00	89,24	0,00
Třída 4	100,00	9,42	0,73	0,00	81,75	8,10	0,00
Třída 5	100,00	0,00	72,89	0,03	2,67	24,41	0,00
Třída 6	100,00	0,00	0,00	0,00	11,29	0,00	88,71

Tabulka 3.12: Procentuální vyjádření odpovídajících si tříd obou norm. metod, pátý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – procentuální vyjádření (%)					
Třída	Procentuální zastoupení (%)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	100,00	0,00	0,00	0,65	33,31	1,22	64,82
Třída 2	100,00	4,34	0,00	93,22	1,49	0,00	0,95
Třída 3	100,00	62,51	0,00	0,11	37,31	0,00	0,00
Třída 4	100,00	0,00	98,18	1,82	0,00	0,00	0,00
Třída 5	100,00	7,89	0,00	1,67	72,81	14,65	2,98
Třída 6	100,00	0,00	5,60	1,35	0,00	81,27	11,78

Tabulka 3.13: Procentuální vyjádření odpovídajících si tříd obou norm. metod, šestý záznam.

Klasifikace normalizační metody Z-skóre		Klasifikace normalizační metody minima a maxima – procentuální vyjádření (%)					
Třída	Procentuální zastoupení (%)	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6
Třída 1	100,00	78,95	0,00	0,00	2,57	12,50	5,97
Třída 2	100,00	0,00	0,00	48,03	0,06	12,08	39,83
Třída 3	100,00	0,03	0,00	17,38	0,00	59,30	23,28
Třída 4	100,00	0,00	0,30	0,00	90,61	7,78	1,31
Třída 5	100,00	0,00	100,00	0,00	0,00	0,00	0,00
Třída 6	100,00	50,75	0,00	0,00	5,43	6,09	37,73

Po konzultaci s expertem došlo k identifikaci jednotlivých tříd z klasifikace dat normalizovaných pomocí metody minima a maxima. Identifikace byla provedena pro všech šest záznamů a její výsledky jsou vypsány v Tabulkách 3.14–3.19.

Tabulka 3.14: Identifikace tříd, první záznam.

Třída	Identifikace třídy
Třída 1	Technické artefakty, fyziologická aktivita s vysokou amplitudou
Třída 2	Fyziologická aktivita s vyšší amplitudou a epileptická aktivita
Třída 3	Epileptická aktivita
Třída 4	Epileptická aktivita a elektroodové artefakty
Třída 5	Epileptická aktivita s šumem a EMG aktivita
Třída 6	Fyziologická aktivita s nižší amplitudou

Tabulka 3.15: Identifikace tříd, druhý záznam.

Třída	Identifikace třídy
Třída 1	Fyziologická aktivita
Třída 2	Fyziologická aktivita s vyšší frekvencí
Třída 3	Fyziologická aktivita s nižší amplitudou
Třída 4	Fyziologická aktivita s nižší amplitudou a vyšší frekvencí
Třída 5	EMG aktivita a fyziologická aktivita s šumem
Třída 6	EMG aktivita s vyšší amplitudou

Tabulka 3.16: Identifikace tříd, třetí záznam.

Třída	Identifikace třídy
Třída 1	Pomalovlnná fyziologická aktivita a oční artefakty
Třída 2	Epileptická aktivita a fyziologická aktivita s vyšší frekvencí
Třída 3	Fyziologická aktivita a epileptická aktivita s šumem
Třída 4	EMG aktivita
Třída 5	Epileptická aktivita, oční a elektroodové artefakty
Třída 6	Fyziologická aktivita

Tabulka 3.17: Identifikace tříd, čtvrtý záznam.

Třída	Identifikace třídy
Třída 1	Oční artefakty a epileptická aktivita
Třída 2	Fyziologická aktivita a epileptická aktivita s nižší amplitudou
Třída 3	Epileptická aktivita, elektroodové a EMG artefakty
Třída 4	Epileptická aktivita s vysokou amplitudou
Třída 5	Epileptická aktivita
Třída 6	Pomalovlnná fyziologická aktivita a oční artefakty

Tabulka 3.18: Identifikace tříd, pátý záznam.

Třída	Identifikace třídy
Třída 1	Fyziologická aktivita
Třída 2	Fyziologická aktivita s nižší amplitudou
Třída 3	Epileptická aktivita a EMG artefakty
Třída 4	Epileptická aktivita a fyziologická aktivita
Třída 5	Oční artefakty a epileptická aktivita
Třída 6	Fyziologická aktivita

Tabulka 3.19: Identifikace tříd, šestý záznam.

Třída	Identifikace třídy
Třída 1	Epileptická aktivita a fyziologická aktivita
Třída 2	Elektrodové artefakty
Třída 3	Pomalovlnná fyziologická aktivita
Třída 4	Epileptická aktivita a EMG artefakty
Třída 5	Epileptická aktivita s šumem
Třída 6	Epileptická aktivita a oční artefakty

Klasifikace byla dále vyhodnocována pomocí normalizovaného vnitrotřídního rozptylu *sumd*. Hodnoty normalizovaného vnitrotřídního rozptylu po klasifikaci dat normalizovaných pomocí Z-skóre jsou zapsané v Tabulce 3.20. Hodnoty *sumd* pro data po klasifikaci dat normalizovaných pomocí metody minima a maxima jsou v Tabulce 3.21. V obou tabulkách je také spočtený průměr normalizovaných hodnot pro každý záznam.

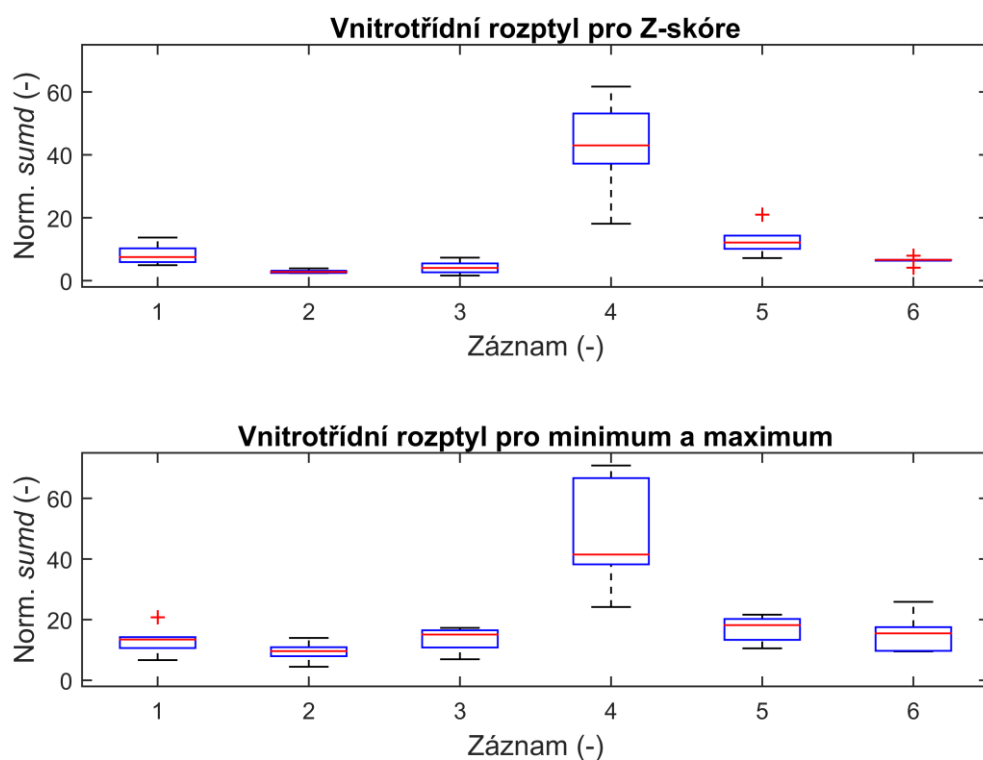
Tabulka 3.20: Vnitrotřídní rozptyly po klasifikaci dat normalizovaných Z-skóre.

Záznam	Normalizovaná hodnota vnitrotřídního rozptylu (-)						Průměr
	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6	
Záznam 1	13,700	7,022	4,919	7,988	10,270	5,903	8,300
Záznam 2	2,456	2,406	2,854	2,703	3,892	3,148	2,910
Záznam 3	5,409	2,696	7,316	2,649	5,485	1,614	4,195
Záznam 4	53,145	18,098	37,202	44,553	41,445	61,705	42,691
Záznam 5	10,117	20,981	11,734	12,512	7,182	14,320	12,808
Záznam 6	4,131	6,343	7,967	6,623	6,649	6,606	6,386

Tabulka 3.21: Vnitrotřídní rozptyly po klas. dat normalizovaných minimem a maximem.

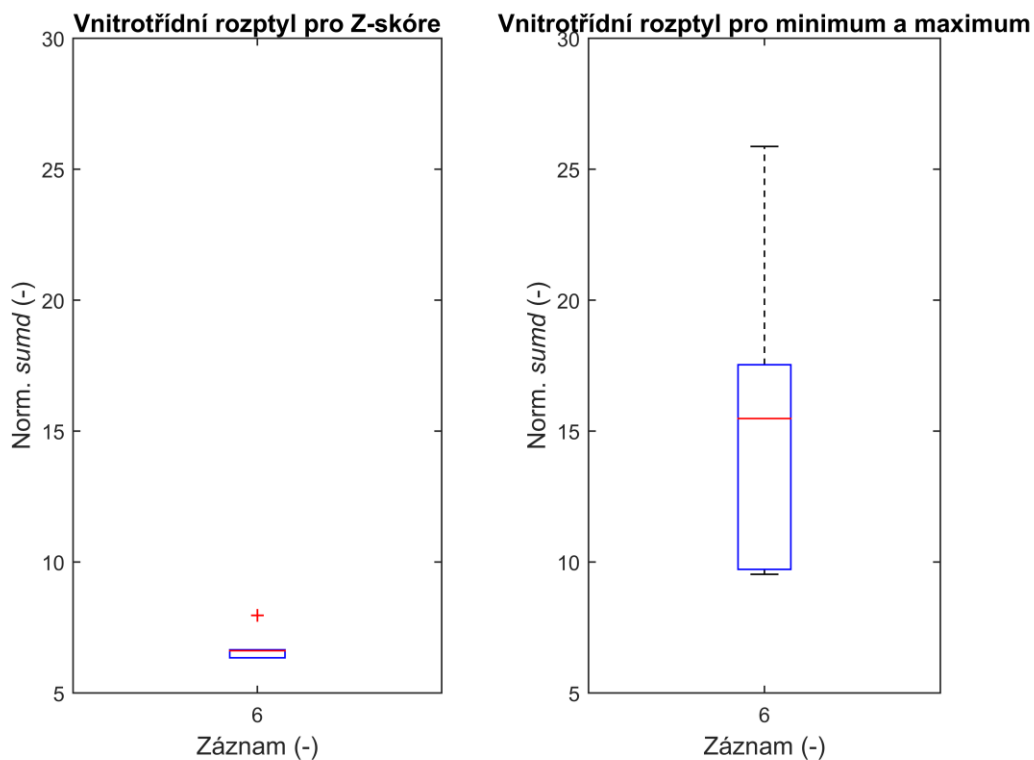
Záznam	Normalizovaná hodnota vnitrotřídního rozptylu (-)						
	Třída 1	Třída 2	Třída 3	Třída 4	Třída 5	Třída 6	Průměr
Záznam 1	6,666	13,061	13,851	14,242	10,635	20,774	13,205
Záznam 2	9,220	10,912	10,024	13,975	7,951	4,464	9,424
Záznam 3	15,356	17,288	10,817	6,942	16,498	14,833	13,622
Záznam 4	70,828	40,569	24,172	38,232	66,671	42,367	47,140
Záznam 5	10,529	13,344	20,218	18,426	17,973	21,635	17,021
Záznam 6	14,907	9,720	17,533	9,531	16,050	25,869	15,602

Z hodnot vnitrotřídních rozptylů byly napříč záznamy vytvořeny dva boxplot grafy. V horní části Obr. 3.18 lze pozorovat boxplot rozložení vnitrotřídních rozptylů po klasifikaci dat normalizovaných pomocí Z-skóre. V dolní části obrázku se nachází boxplot rozložení vnitrotřídních rozptylů po klasifikaci dat normalizovaných pomocí metody minima a maxima.



Obr. 3.18: Boxplot zobrazení vnitrotřídních rozptylů obou normalizačních metod na všech záznamech.

Jelikož normalizované hodnoty vnitřitřídniho rozptylu čtvrtého záznamu na Obr. 3.17 a 3.18 zkreslují hodnoty zbylých záznamů, bylo pro detailnější zobrazení rozdílu využito Obr. 3.19. Na obrázku je detailnější boxplot zobrazení normalizovaných hodnot vnitřitřídniho rozptylu pro šestý záznam. Na levé straně lze pozorovat data po normalizaci Z-skóre. Na pravé straně se nachází boxplot zobrazení dat po normalizaci minimem a maximem.



Obr. 3.19: Detailní boxplot zobrazení normalizovaných hodnot vnitřitřídniho rozptylu šestého záznamu.

4 Diskuse

V této bakalářské práci byl otestován a vyhodnocen vliv dvou normalizačních metod na výsledné klasifikaci dat. Vliv normalizace byl testován na šesti záznamech s epileptickou aktivitou. Délky záznamů se pohybovaly v rozmezí od 14 do 54 minut.

Nejprve byly všechny záznamy předzpracované a segmentované pomocí adaptivní segmentace v programovém prostředí Wave-Finder. Jak lze pozorovat v Tabulce 3.1, dle předpokladu počet segmentů částečně závisel na délce záznamu. Například nejkratší záznamy dva a pět obsahovaly nejmenší množství segmentů. Z každého segmentu bylo vypočítáno 23 příznaků, jejichž výpočet software WF umožňuje. Hodnoty příznaků byly normalizovány pomocí metody Z-skóre a metody minima a maxima.

Hodnoty příznaků byly zobrazeny pomocí boxplot grafů. Původní data prvního záznamu před normalizací jsou zobrazena na Obr. 3.1–3.4. Z grafů byly odstraněny odlehlé hodnoty, aby byl viditelný rozsah boxplotů. Jak dojde k potlačení zobrazení hodnot příznaků odlehlými hodnotami, lze pozorovat na Obr. 3.5. Bez odlehlých hodnot jsou zobrazeny také boxploty pro normalizovaná data, opět se jedná o první záznam. Boxplot zobrazení příznaků normalizovaných pomocí Z-skóre lze pozorovat na Obr. 3.6–3.9. Na datech je zřejmé ustálení mediánové hodnoty kolem nuly u všech příznaků. Zobrazení dat normalizovaných pomocí metody minima a maxima lze pozorovat na Obr. 3.10–3.13. Při normalizaci minimem a maximem můžeme téměř u všech příznaků pozorovat vycházející linie od nulové hodnoty. Většina příznaků také nepřesahuje hodnotu 0,4. Zobrazení původních i normalizovaných dat s odlehlými hodnotami lze pozorovat v Příloze B.

Takto normalizovaná data byla dále testována Kolmogorov-Smirnovovým testem normality. Na hladině významnosti 5 % bylo zjištěno, že data ze všech záznamů po normalizaci nepochází z normálního rozložení. Tento výsledek byl také ověřen graficky pomocí zobrazení histogramů. Na ukázkou byl vybrán příznak *DELTI* z prvního záznamu. Na Obr. 3.14 lze pozorovat tento příznak po normalizaci Z-skóre a na Obr. 3.15 je tento příznak normalizován metodou minima a maxima.

Jelikož normalizovaná data nepochází z normálního rozložení, byl ke statistice vybrán neparametrický test. Na vybraném souboru dat byl pomocí Wilcoxonova testu zjištěn významný rozdíl mezi normalizačními metodami. Na hladině významnosti 5 % došlo k zamítnutí nulové hypotézy u všech záznamů. Bylo tedy zjištěno, že se datové soubory normalizované pomocí dvou normalizačních metod mezi sebou liší.

Normalizovaná data byla vložena do klasifikátoru k-means. V klasifikátoru byla nastavena hodnota tříd na šest. Jedná se o empiricky zjištěnou nejvhodnější hodnotu, při které dochází k relativně úspěšné klasifikaci. Klasifikace dat prvního záznamu normalizovaných pomocí metody Z-skóre je zobrazena na Obr. 3.16, data prvního záznamu normalizovaná pomocí metody minima a maxima jsou zobrazena na Obr. 3.17.

Vzhledem k různým rozsahům normalizačních metod jsou i osy v grafech v jiných rozsazích. Vzhledem k tomu, že funkce k-means přiřazuje shluky k datům náhodně, bylo testováno, jak si shluky obou normalizačních metod odpovídají.

Pro všechny třídy každého záznamu byl určen odpovídající si počet segmentů při obou normalizačních metodách. Tyto hodnoty pro všechny záznamy lze pozorovat v Tabulkách 3.2–3.3.7. Z těchto hodnot bylo vypočítáno procentuální vyjádření odpovídajících si tříd při klasifikaci dat normalizovaných oběma metodami, které lze pozorovat v Tabulkách 3.8–3.13. V tabulkách jsou také barevně zvýrazněny třídy z klasifikace dat normalizovaných minimem a maximem s nejvyšší procentuální shodou s třídami z klasifikace pomocí Z-skóre. Ve všech záznamech a všech třídách můžeme pozorovat shodu vyšší než 50 %. Jedinou výjimkou je Třída 2 v šestém záznamu (normalizace Z-skóre), kdy nejvyšší shoda dosahovala pouze 48 %.

Třídy klasifikované prostřednictvím metody minima a maxima byly identifikovány pomocí experta. Výsledky identifikace tříd jsou zapsány v Tabulkách 3.14–3.19. Bylo zjištěno, že klasifikace prostřednictvím normalizační metody Z-skóre oddělí technické artefakty do samostatné třídy (Třída 4). Při použití normalizace minimem a maximem v této třídě byla zahrnuta také epileptická aktivita. Při identifikaci epileptické aktivity se expert zaměřil na hledání komplexu hrot-vlna. Tento jev například nebyl patrný ve druhém záznamu. Expert dále určil, že klasifikovaná data po normalizaci Z-skóre mají nejspíše vyšší homogenitu, než data po normalizaci minimem a maximem. Zařazení do tříd bylo pro Z-skóre normalizaci pozorováno přesněji.

Tato domněnka byla ověřena v této práci pomocí normalizovaného parametru vnitrotřídního rozptylu. Hodnoty normalizovaného vnitrotřídního rozptylu ve všech záznamech a třídách jsou pro obě metody vypsány v Tabulkách 3.20 a 3.21. V tabulkách je také vypsán průměr normalizovaného parametru *sumd* napříč třídami v záznamu. Ve všech záznamech byly průměrné hodnoty vnitrotřídního rozptylu nižší při normalizaci prostřednictvím Z-skóre. Tyto parametry pro obě metody byly zobrazeny pomocí boxplot grafů v Obr. 3.18. Horní graf zobrazuje vnitrotřídní rozptyl pro normalizaci Z-skóre a na dolním grafu jsou zobrazeny vnitrotřídní rozptyly všech záznamů při normalizaci minimem a maximem. Detailní zobrazení rozdílu normalizovaných vnitrotřídních rozptylů získaných po normalizaci oběma metodami lze pozorovat na Obr. 3.19. I přesto, že se jedná o zobrazení stejného záznamu, lze pozorovat při použití dvou normalizačních metod rozdíl v rozptylu normalizovaných hodnot.

Po konzultaci s expertem a s ohledem na homogenitu tříd bylo zjištěno, že Z-skóre normalizace byla v této práci při klasifikaci výhodnější. Omezením této práce může být například výběr klasifikační metody. Metoda k-means je limitovaná počtem tříd jako vstupním parametrem. Pro přesnější určení počtu tříd, do kolika je záznam nutné rozdělit, existují vzorce pro výpočet. Další možností by bylo využití klasifikační metody, jež ke klasifikaci počet tříd jako vstupní parametr nepotřebuje. Takovouto metodou by

mohl být například klasifikátor DBSCAN. Další variantou by bylo využití učícího se klasifikátoru. Učící se klasifikátory mohou být výhodnější pro data s lineárními vztahy v prostoru na rozdíl od klasifikátorů bez učitele [24].

Samotnou limitací může být také využití již implementované metody k-means v softwaru MATLAB[®]. Metoda může být napříč různými programovacími jazyky naprogramovaná jinak. V této práci například při každém spuštění skriptu s k-means klasifikací docházelo k nové inicializaci a při každém spuštění byly jinak vyhodnoceny středy shluků.

Způsobem, jak zdokonalit tuto práci, by bylo rozšíření statistického vyhodnocení klasifikovaných dat. K vyhodnocení by mohla posloužit ROC analýza a výpočet statistických charakteristik. Mezi relevantní parametry sloužící k vyhodnocení by mohly být využity senzitivita a pozitivní prediktivní hodnota (positive predictive value, PPV). Pro tuto analýzu by bylo potřeba předem označit všechny segmenty záznamu. Na označená data by se dala využít klasifikace s učitelem. Jelikož se jedná pouze o pilotní studii, bylo by vhodné dále provádět testování na větším souboru dat.

5 Závěr

V této bakalářské práci byl otestován a vyhodnocen vliv dvou metod normalizace na výsledné klasifikaci. Bylo předzpracováno šest záznamů v programovém prostředí Wave-Finder. Záznamy byly filtrovány a pomocí adaptivní segmentace rozčleněny. Ze všech segmentů bylo vypočteno 23 příznaků, jež se v softwaru WF dají vypočítat. V programovém prostředí byly na příznaky implementovány normalizační metody, kterými byla Z-skóre normalizace a metoda minima a maxima. Pomocí testu normality bylo zjištěno, že data nepochází z normálního rozdělení. Proto byl pro vyhodnocení rozdílu na datech normalizovanými pomocí dvou metod využit neparametrický Wilcoxonův test. Výsledkem tohoto testu bylo zjištění, že na zkoumaných datech existuje statistický rozdíl na hladině významnosti 5 %. Normalizovaná data byla klasifikována pomocí k-means do šesti tříd. Procentuálně byly vyhodnoceny vztahy mezi třídami vzniklými po klasifikaci Z-skóre a třídami, které vznikly z klasifikace dat normalizovanými minimem a maximem. Procentuální vyjádření u všech záznamů a všech tříd bylo vyšší než 48 %. Po konzultaci s expertem byly identifikovány všechny třídy vzniklé klasifikací dat normalizovaných pomocí minimem a maximem pro všechny záznamy. Na základě vizuálního hodnocení expertem i na základě testování vnitrotřídního rozptylu bylo zjištěno, že homogenita tříd po normalizaci Z-skóre je vyšší než homogenita po normalizaci minimem a maximem. Je tedy možné říci, že s ohledem na homogenitu se na datovém souboru využitém v této práci jeví Z-skóre normalizace jako výhodnější.

Seznam použité literatury

- [1] LOGESPARAN, Lojini, Esther RODRIGUEZ-VILLEGAS a Alexander CASSON. *The impact of signal normalization on seizure detection using line length features*. 2015, **53**(10), 929-942. DOI: 10.1007/s11517-015-1303-x. ISSN 0140-0118. Dostupné také z: <http://link.springer.com/10.1007/s11517-015-1303-x>
- [2] BOGAARTS, J., D. HILKMAN, E. GOMMER, V. VAN KRANEN-MASTENBROEK a J. REULEN. *Improved epileptic seizure detection combining dynamic feature normalization with EEG novelty detection*. 2016, **54**(12), 1883-1892. DOI: 10.1007/s11517-016-1479-8. ISSN 0140-0118. Dostupné také z: <http://link.springer.com/10.1007/s11517-016-1479-8>
- [3] AMIN, Hafeez, Wajid MUMTAZ, Ahmad SUBHANI, Mohamad SAAD a Aamir MALIK. Classification of EEG Signals Based on Pattern Recognition Approach. *Frontiers in Computational Neuroscience*. 2017, **11**(103). DOI: 10.3389/fncom.2017.00103. ISSN 1662-5188. Dostupné také z: <http://journal.frontiersin.org/article/10.3389/fncom.2017.00103/full>
- [4] NUNEZ, Paul a Ramesh SRINIVASAN. *Electric fields of the brain: the neurophysics of EEG*. 2nd ed. New York: Oxford University Press, 2006. ISBN 978-0-19-505038-7.
- [5] MALMIVUO, Jaakko a Robert PLONSEY. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. New York: Oxford University Press, 1995. ISBN 978-0195058239.
- [6] JURCAK, Valer, Daisuke TSUZUKI a Ippaita DAN. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*. 2007, **34**(4), 1600-1611. DOI: 10.1016/j.neuroimage.2006.09.024. ISSN 10538119. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/S1053811906009724>
- [7] KRAJČA, Vladimír a Jitka MOHYLOVÁ. *Číslíkové zpracování neurofyziologických signálů*. 1. vydání. V Praze: České vysoké učení technické, 2011. ISBN 978-80-01-04721-7.
- [8] KOZIC, Zlatko. *Zpracování EEG záznamů dětských pacientů s vývojovou dysfázií*. Praha, 2013. Diplomová práce. České vysoké učení technické v Praze.

- [9] RIEGER, J., L. LHOTSKÁ a V. KRAJČA. Zpracování dlouhodobých EEG záznamů. *Advances in electrical and electronic engineering*. 2005, **4**(3), 151-156. ISSN 1336-1376.
- [10] TEPLAN, Michal. Fundamentals of EEG Measurement. *Measurement Science Review*. 2002, **2**(2).
- [11] KRAJČA, Vladimír a Svojmil PETRÁNEK. "Wave-Finder": A new System for an Automatic Processing of Long-Term EEG Recordings. *Quantitative EEG analysis-clinical utility and new methods*. 1993, , 103-106.
- [12] GERLA, Václav. *Automated Analysis of Long-Term EEG Signals*. Praha, 2012. Disertační práce. České vysoké učení technické v Praze.
- [13] AUFRICHTIGL, R., S.B. PEDERSEN a P. JENNUM. Adaptive Segmentation Of EEG Signals. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society Volume 13: 1991*. IEEE, 1991, **13**(1), 453-454. DOI: 10.1109/IEMBS.1991.684023. ISBN 0-7803-0216-8. Dostupné také z: <http://ieeexplore.ieee.org/document/684023/>
- [14] DOUGHERTY, Geoff. *Pattern recognition and classification: an introduction*. New York: Springer, 2013. ISBN 978-1-4614-5322-2.
- [15] THEODORIDIS, Sergios a Konstantinos KOUTROUMBAS. *Pattern recognition*. 2nd ed. Boston: Academic Press, 2003. ISBN 01-268-5875-6.
- [16] RASCHKA, Sebastian. About Feature Scaling and Normalization: and the effect of standardization for machine learning algorithms. *Sebastianraschka* [online]. 2014 [cit. 2018-12-28]. Dostupné z: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html
- [17] Zscore: Standardized z-scores. *MathWorks®* [online]. b.r. [cit. 2018-12-29]. Dostupné z: <https://www.mathworks.com/help/stats/zscore.html>
- [18] Kstest: One-sample Kolmogorov-Smirnov test. *MathWorks®* [online]. b.r. [cit. 2018-12-28]. Dostupné z: https://www.mathworks.com/help/stats/kstest.html#responsive_offcanvas
- [19] Signrank: Wilcoxon signed rank test. *MathWorks®* [online]. b.r. [cit. 2018-12-28]. Dostupné z: <https://www.mathworks.com/help/stats/signrank.html>
- [20] PEREIRA, Francisco, Tom MITCHELL a Matthew BOTVINICK. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*. 2009, **45**(1), 199-209. DOI: 10.1016/j.neuroimage.2008.11.007. ISSN

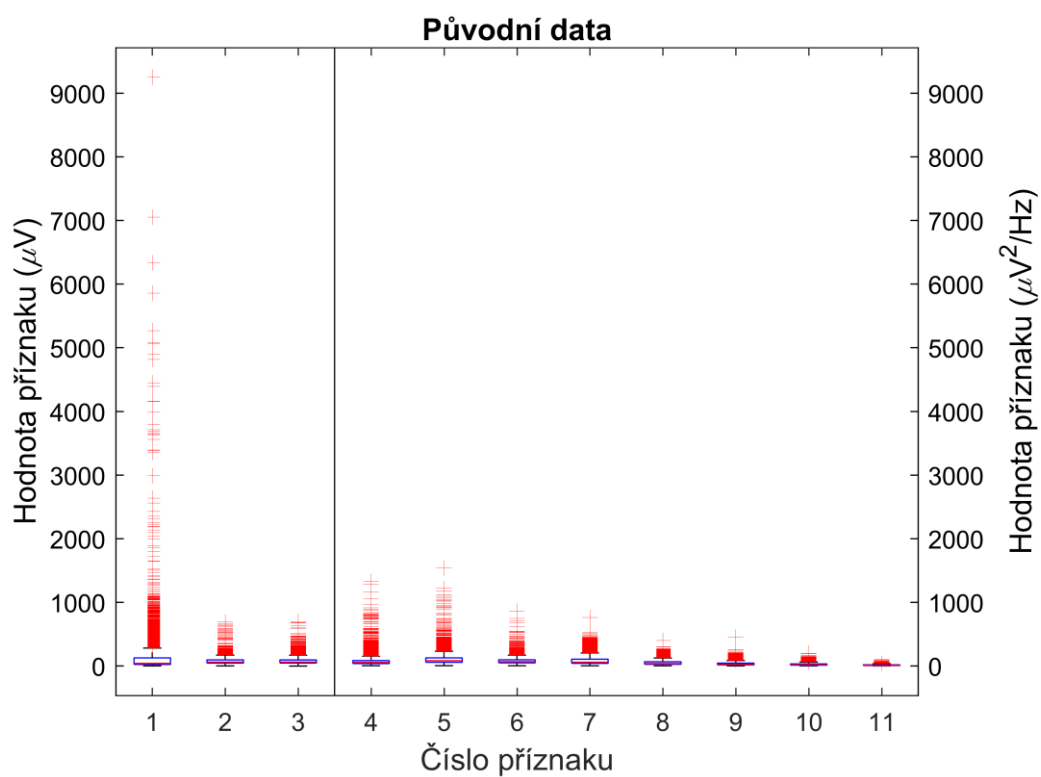
10538119. Dostupné také z:
<https://linkinghub.elsevier.com/retrieve/pii/S1053811908012263>

- [21] WILKS, D.S. Cluster Analysis. *Statistical Methods in the Atmospheric Sciences*. Elsevier, 2011, , 603-616. International Geophysics. DOI: 10.1016/B978-0-12-385022-5.00015-4. ISBN 9780123850225. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/B9780123850225000154>
- [22] Kmeans. *MathWorks®* [online]. b.r. [cit. 2019-04-24]. Dostupné z: <https://www.mathworks.com/help/stats/kmeans.html>
- [23] Boxplot. *MathWorks®* [online]. b.r. [cit. 2018-12-29]. Dostupné z: <https://www.mathworks.com/help/stats/boxplot.html>
- [24] LOVE, Bradley. Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review* [online]. 2002, **9**(4), 829-835 [cit. 2019-05-13]. DOI: 10.3758/BF03196342. ISSN 1069-9384. Dostupné z: <http://www.springerlink.com/index/10.3758/BF03196342>

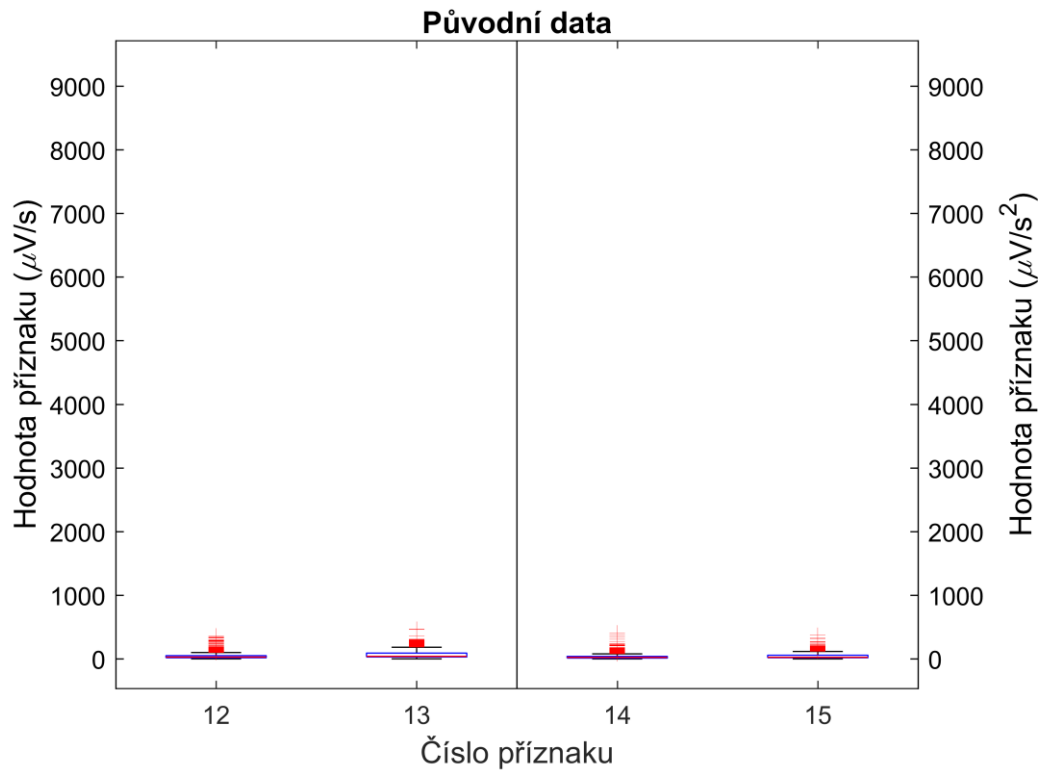
Příloha A: Obsah přiloženého CD

- klíčová slova v ČJ i AJ
- abstrakt v ČJ
- abstrakt v AJ
- naskenované zadání bakalářské práce
- kompletní bakalářská práce
- vytvořené skripty v programu MATLAB[®]

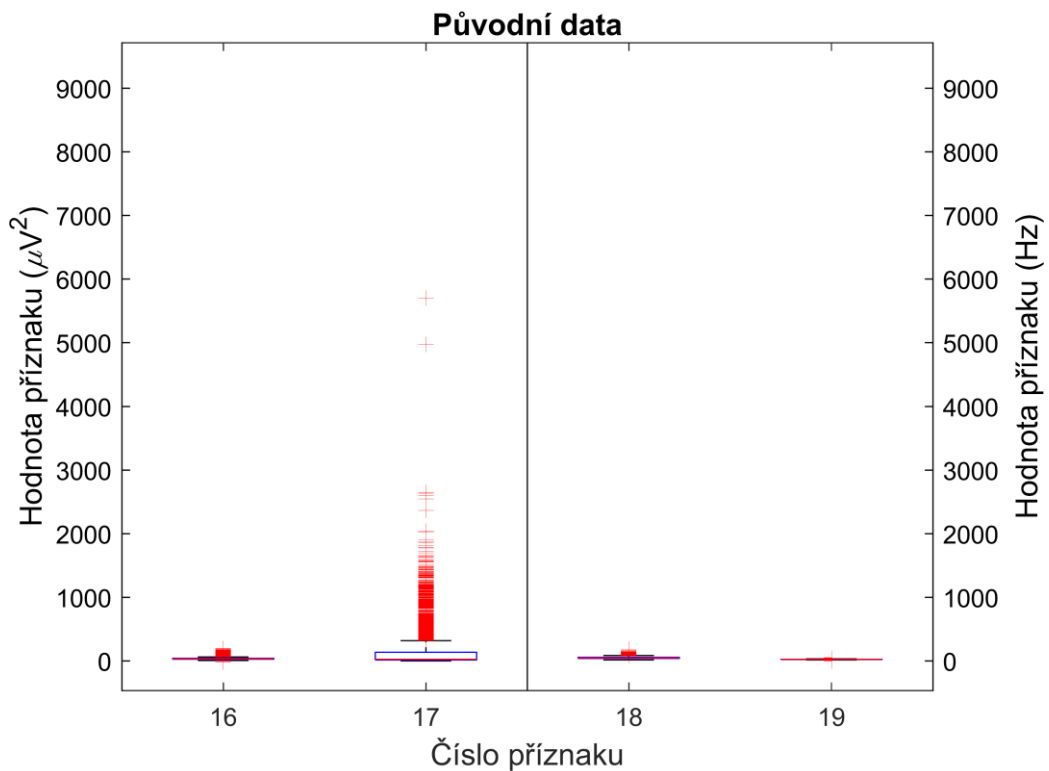
Příloha B: Boxplot zobrazení prvního záznamu s odlehlými hodnotami



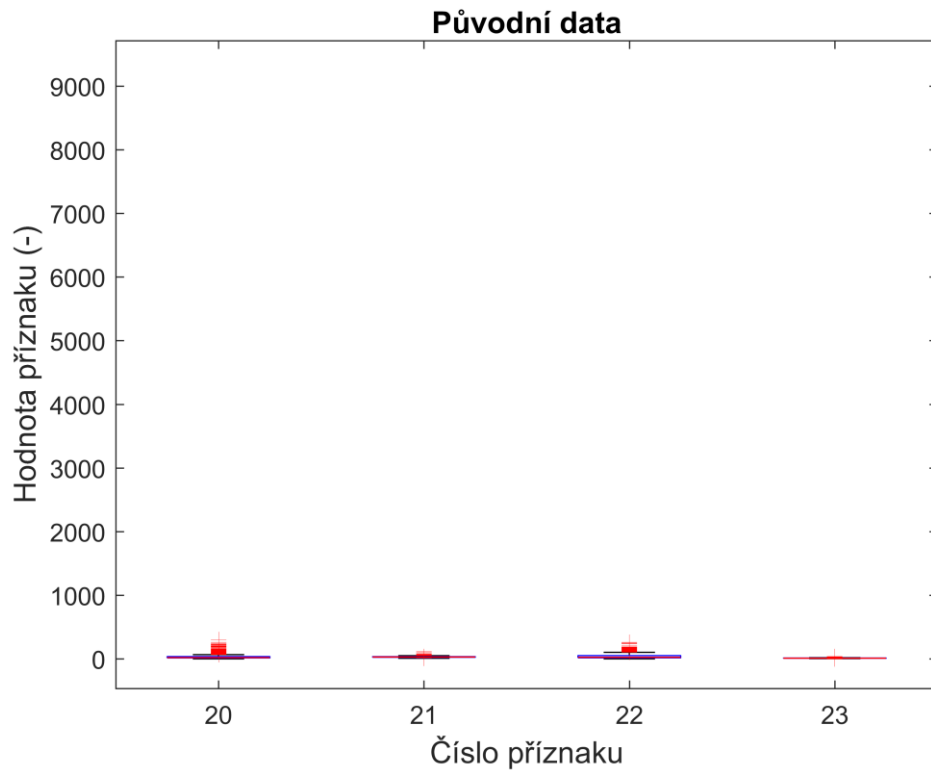
Obr. B.1: Boxplot zobrazení prvních 11 extrahovaných příznaků z prvního záznamu s odlehlými hodnotami, příznaky jsou seřazeny dle Tabulky 2.1



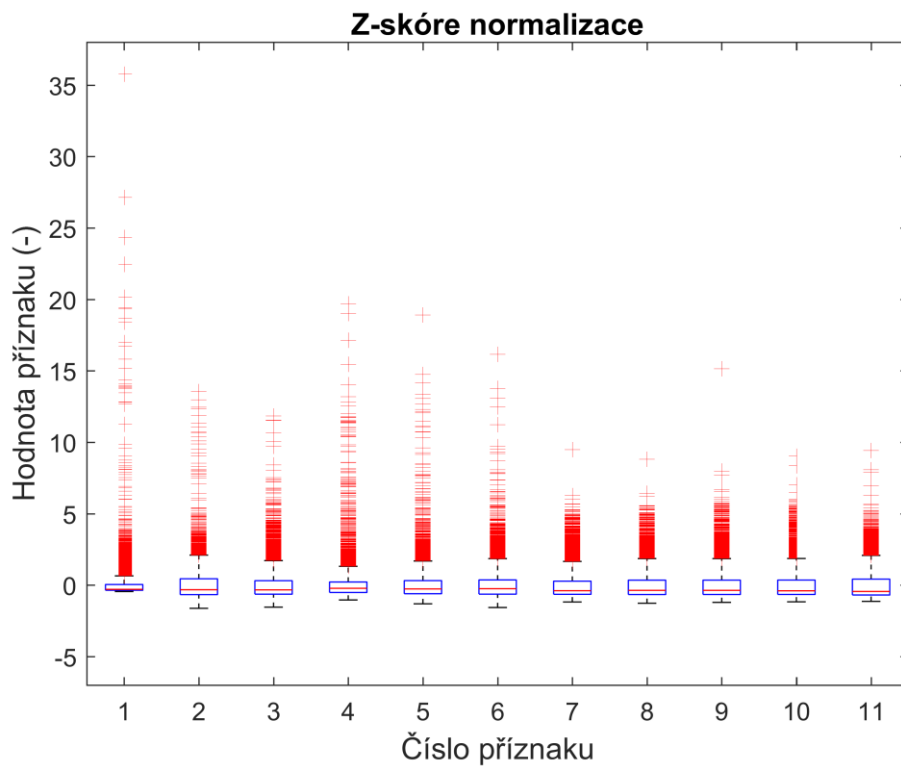
Obr. B.2: Boxplot zobrazení příznaků 12–15 prvního záznamu s odlehlými hodnotami, příznaky jsou seřazeny dle Tabulky 2.1



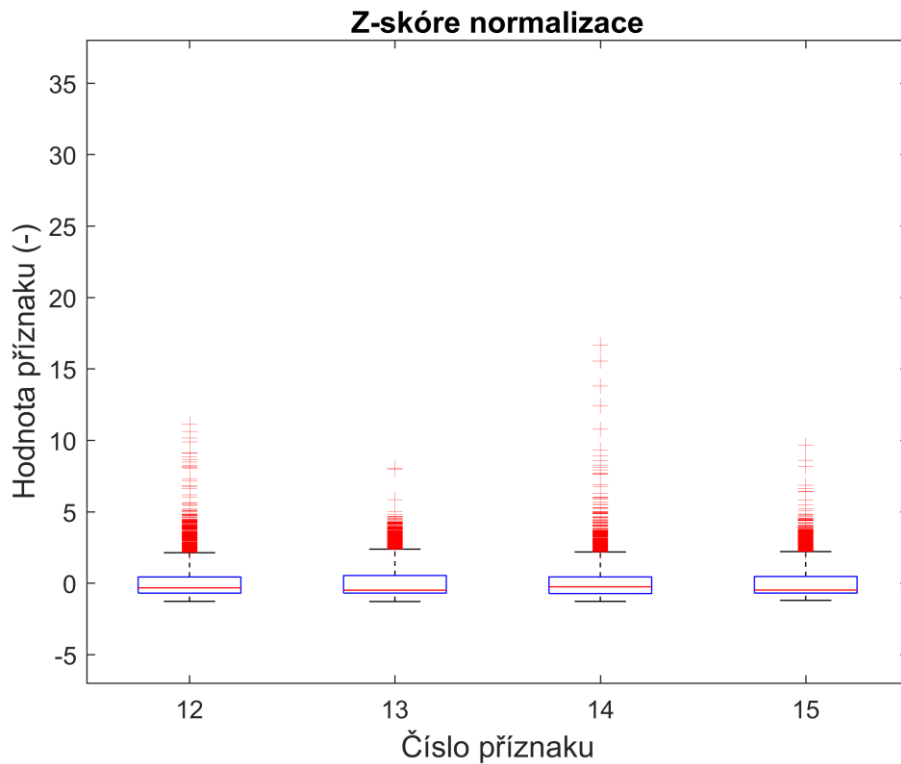
Obr. B.3: Boxplot zobrazení příznaků 16–19 prvního záznamu s odlehlými hodnotami, příznaky jsou seřazeny dle Tabulky 2.1



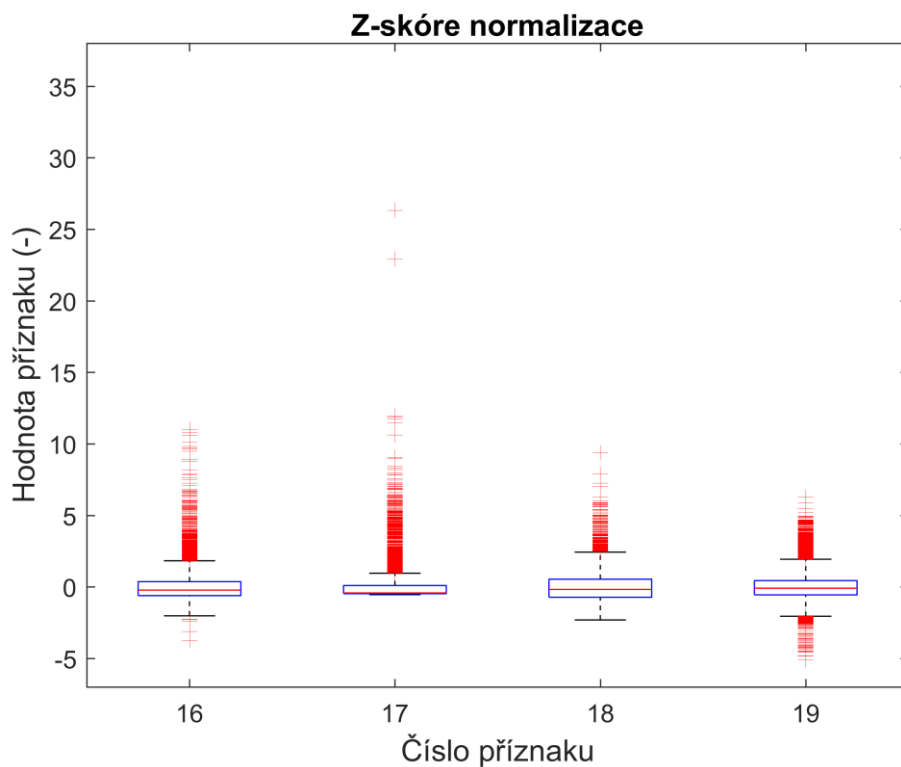
Obr. B.4: Boxplot zobrazení příznaků 20–23 prvního záznamu s odlehlými hodnotami, příznaky jsou seřazeny dle Tabulky 2.1



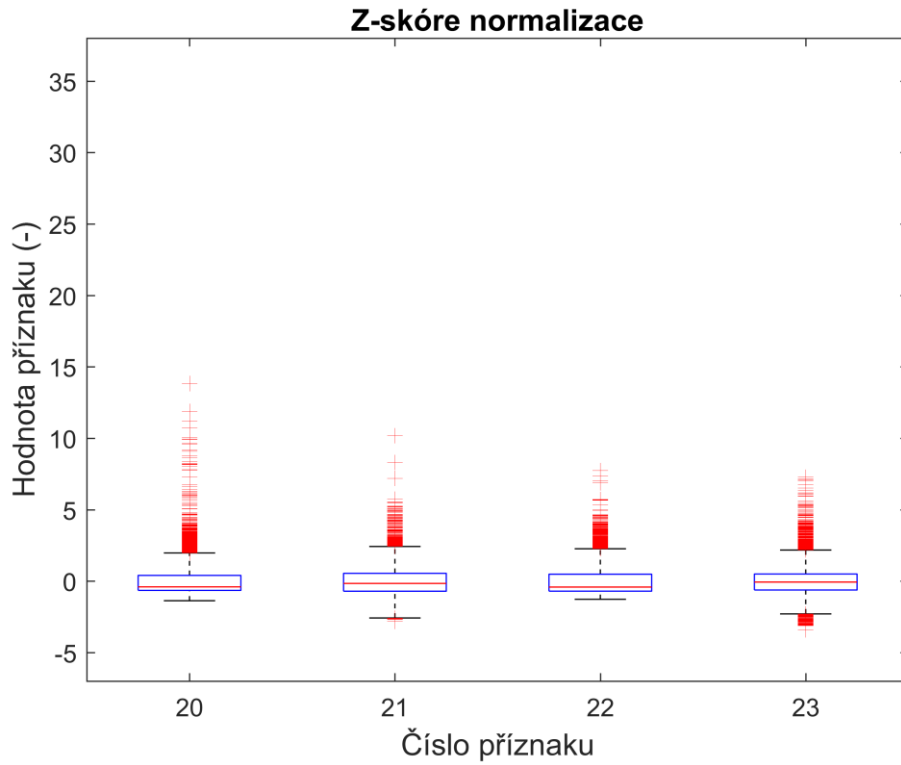
Obr. B.5: Boxplot zobrazení prvních 11 příznaků z prvního záznamu s odlehlými hodnotami, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1



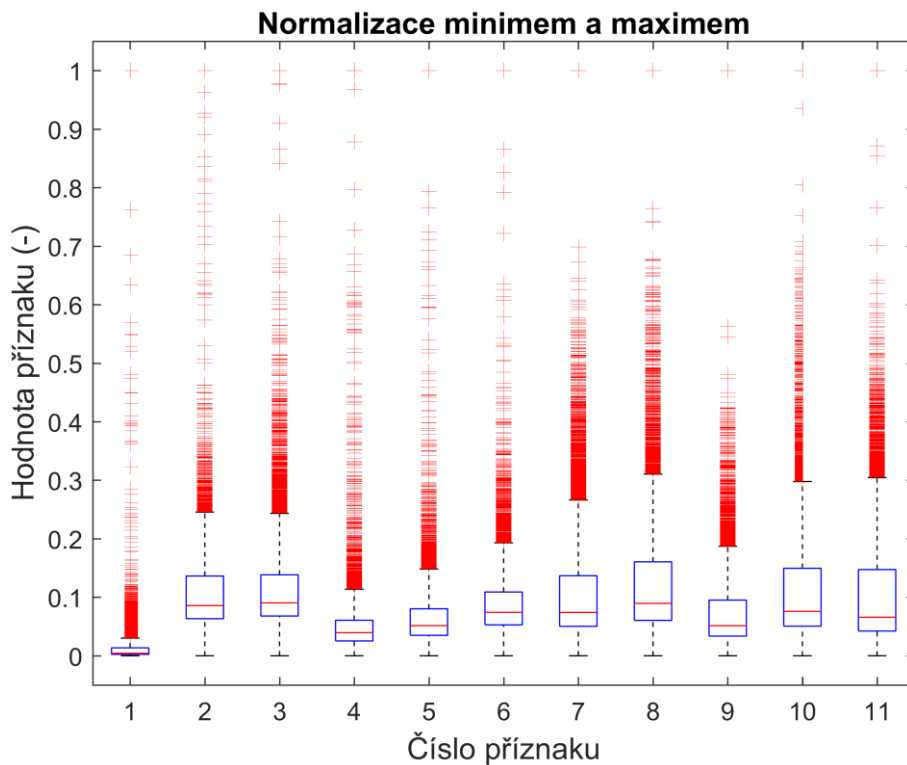
Obr. B.6: Boxplot zobrazení příznaků 12–15 prvního záznamu s odlehlými hodnotami, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1



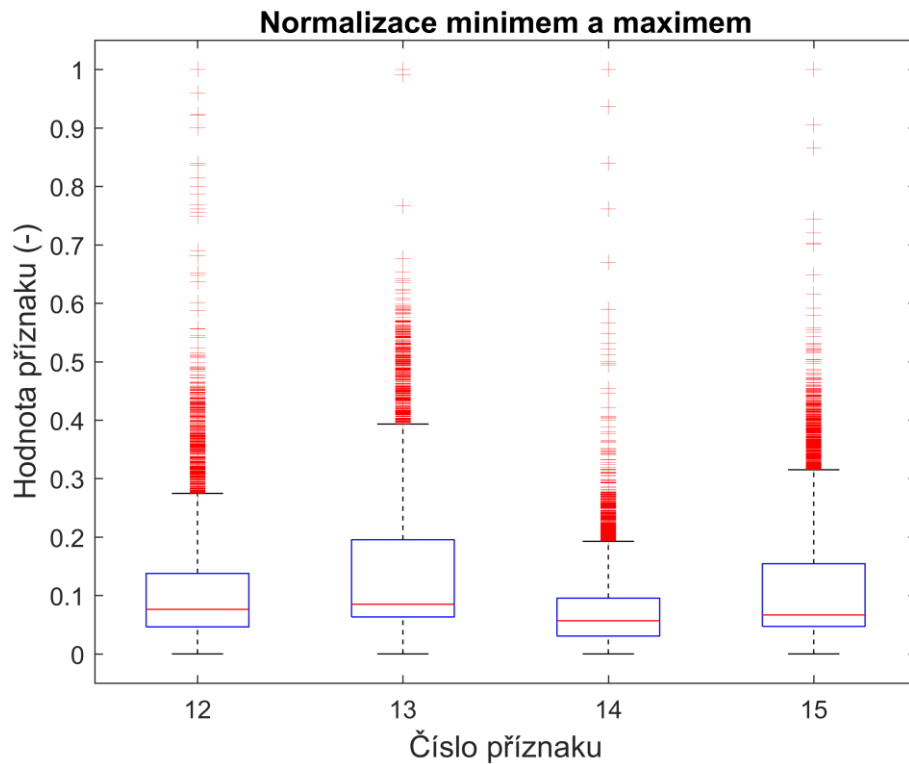
Obr. B.7: Boxplot zobrazení příznaků 16–19 prvního záznamu s odlehlými hodnotami, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1



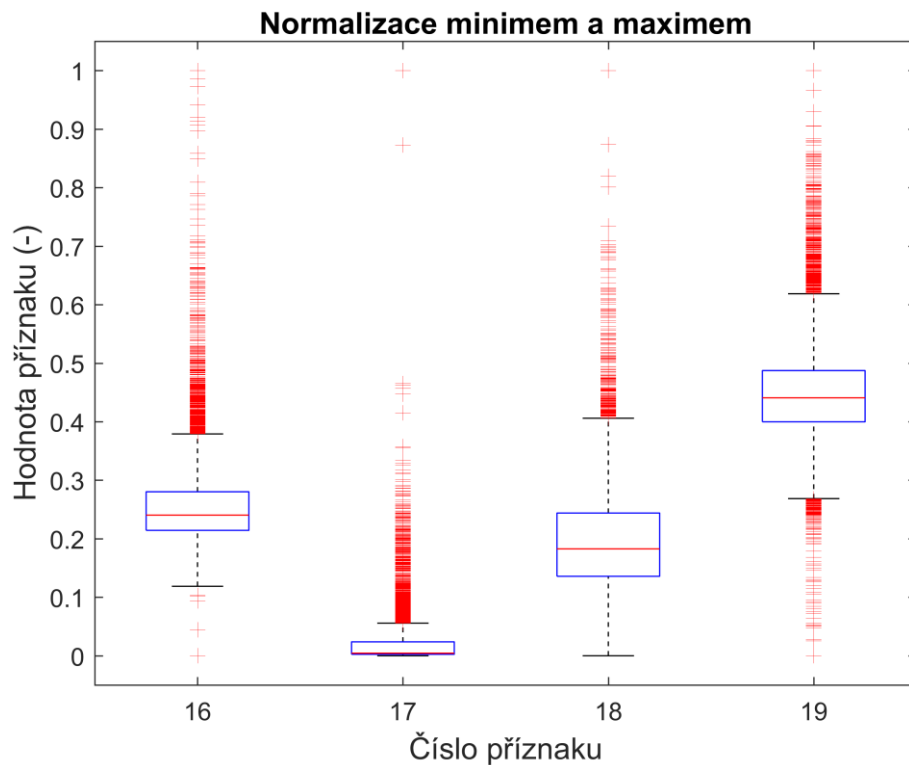
Obr. B.8: Boxplot zobrazení příznaků 20–23 prvního záznamu s odlehlými hodnotami, normalizované pomocí Z-skóre, příznaky jsou seřazeny dle Tabulky 2.1



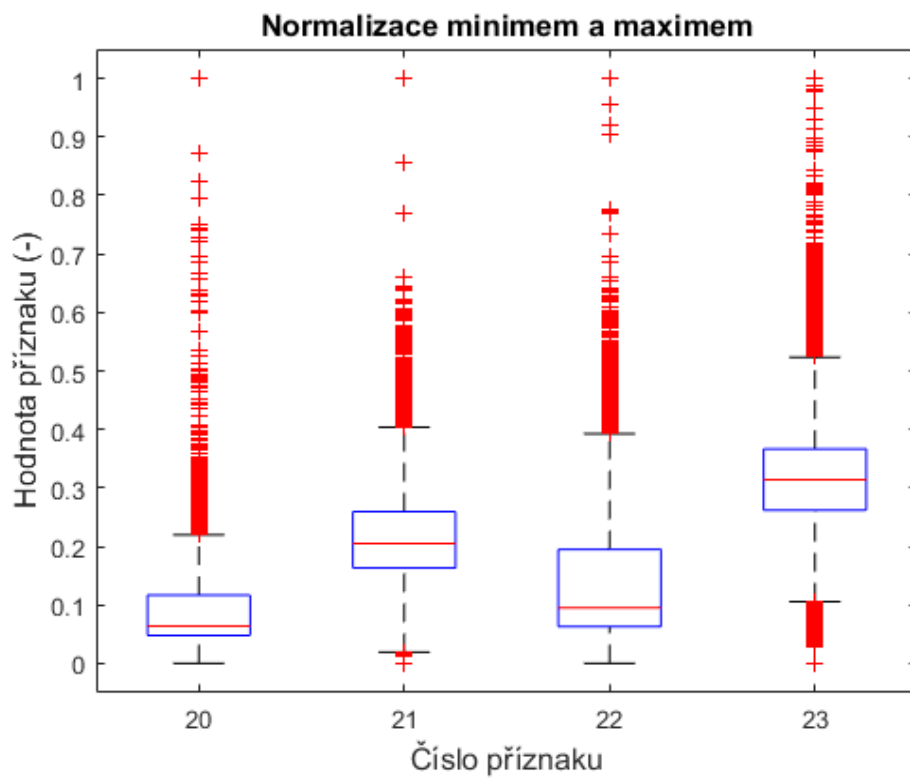
Obr. B.9: Boxplot zobrazení prvních 11 příznaků z prvního záznamu s odlehlými hodnotami, normalizované minimem a maximem, příznaky jsou seřazeny dle Tabulky 2.1



Obr. B.10: Boxplot zobrazení příznaků 12–15 prvního záznamu s odlehlými hodnotami, normalizované minimem a maximem, příznaky jsou seřazeny dle Tabulky 2.1



Obr. B.11: Boxplot zobrazení příznaků 16–19 prvního záznamu s odlehlými hodnotami, normalizované minimem a maximem, příznaky jsou seřazeny dle Tabulky 2.1



Obr. B.12: Boxplot zobrazení příznaků 20–23 prvního záznamu s odlehlými hodnotami, normalizované minimem a maximem, příznaky jsou seřazeny dle Tabulky 2.1