

II. Seminární práce
Počítačové zpracování signálu (KI/PZS)

Karel Murgač

1 Zadání

Cílem této seminární práce je analýza a klasifikace hlasových záznamů z databáze VOICED (PhysioNet). Databáze obsahuje celkem 208 záznamů hlasu (samohláska "a") od různých subjektů.

Hlavní body zadání jsou definovány následovně:

1. Načíst data a orientovat se ve formátu databáze PhysioNet.
2. Implementovat zpracování signálu pomocí vybraných technik v časové nebo frekvenční oblasti.
3. Provést extrakci příznaků s využitím Fourierovy transformace a kepstrální analýzy.
4. Klasifikovat záznamy do dvou skupin: **dobré (zdravé)** a **patologické**.
5. V případě patologických záznamů dále klasifikovat **konkrétní poruchy** (diagnózy).
6. Porovnat úspěšnost automatické klasifikace s expertními anotacemi.
7. Graficky prezentovat výsledky analýzy a úspěšnost klasifikace.

Pro zpracování je zakázáno využívat předpřipravené analytické nástroje z balíku WFDB, signál musí být zpracován pomocí obecných matematických knihoven (např. NumPy, SciPy).

2 Postup řešení

Řešení bylo implementováno v programovacím jazyce Python s využitím prostředí Jupyter Notebook. Postup lze shrnout do následujících kroků:

2.1 Načítání a čištění dat

Data byla načtena pomocí knihovny `wfdb` (pouze pro I/O operace). Klíčovým krokem bylo parsování hlavičkových souborů (`.hea`), které obsahují nestrukturované komentáře s diagnózami. Byla implementována robustní funkce pro extrakci diagnózy, která rozdělila dataset na:

- **Healthy:** 57 záznamů.
- **Pathological:** 151 záznamů.

Celkem bylo zpracováno 208 záznamů.

	record_name	signal	fs	diagnosis	label_binary
0	voice001	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	8000	hyperkinetic dysphonia	Pathological
1	voice002	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	8000		Healthy
2	voice003	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	8000	hyperkinetic dysphonia	Pathological
3	voice004	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	8000	hypokinetic dysphonia	Pathological
4	voice005	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	8000	hypokinetic dysphonia	Pathological

Obrázek 1: Ukázka načtených dat a metadat po zpracování hlaviček.

2.2 Zpracování signálu a extrakce příznaků

Signály byly vzorkovány na frekvenci 8000 Hz. Pro každý záznam byly vypočteny následující příznaky:

- **Časová oblast:** RMS (Root Mean Square) pro energii signálu a ZCR (Zero Crossing Rate) pro míru šumu.
- **Frekvenční oblast:** Spektrální centroid vypočtený z Fourierovy transformace (FFT).
- **Kepstrální oblast:** 13 MFCC koeficientů (Mel-Frequency Cepstral Coefficients), zprůměrovaných v čase.

	diagnosis	label_binary	rms	zcr	spectral_centroid	mfcc_1	mfcc_2	mfcc_3	mfcc_4	mfcc_5	mfcc_6	mfcc_7	mfcc_8	mfcc_9	mfcc_10	mfcc_11	mfcc_12	mfcc_13
0	hyperkinetic dysphonia	Pathological	0.130236	0.229307	1158.195462	-89.878323	53.312515	-66.393846	-16.300682	-6.631975	17.999434	-27.778969	15.623700	6.169481	-34.712455	-3.559567	-6.395851	-8.561320
1	Healthy	Healthy	0.291000	0.218776	1058.374423	-115.652658	36.073274	-52.771162	-33.146484	-15.068009	3.447450	-22.321822	10.138316	13.726444	-27.952345	-21.330649	-14.298918	-6.626558
2	hyperkinetic dysphonia	Pathological	0.259768	0.209086	1505.803696	-57.845436	-3.434850	-42.967608	-35.288694	3.767250	5.081971	-20.880665	13.056897	-4.589232	-28.704122	-10.099688	-0.145197	-10.543688
3	hypokinetic dysphonia	Pathological	0.051169	0.157169	1370.842079	-206.126466	13.042708	-27.005930	32.183496	19.796031	-24.004164	11.081125	4.832909	-22.760615	-3.648753	-25.758860	-8.976304	-23.258990
4	hypokinetic dysphonia	Pathological	0.185829	0.238314	1327.631938	-104.151934	8.180843	-49.468008	-17.298141	28.267973	4.029834	-3.836589	-4.155939	-23.870635	1.722772	-17.541468	-1.895349	-22.703729

Obrázek 2: Ukázka vypočítaných příznaků (RMS, ZCR, Centroid, MFCC) pro jednotlivé záznamy.

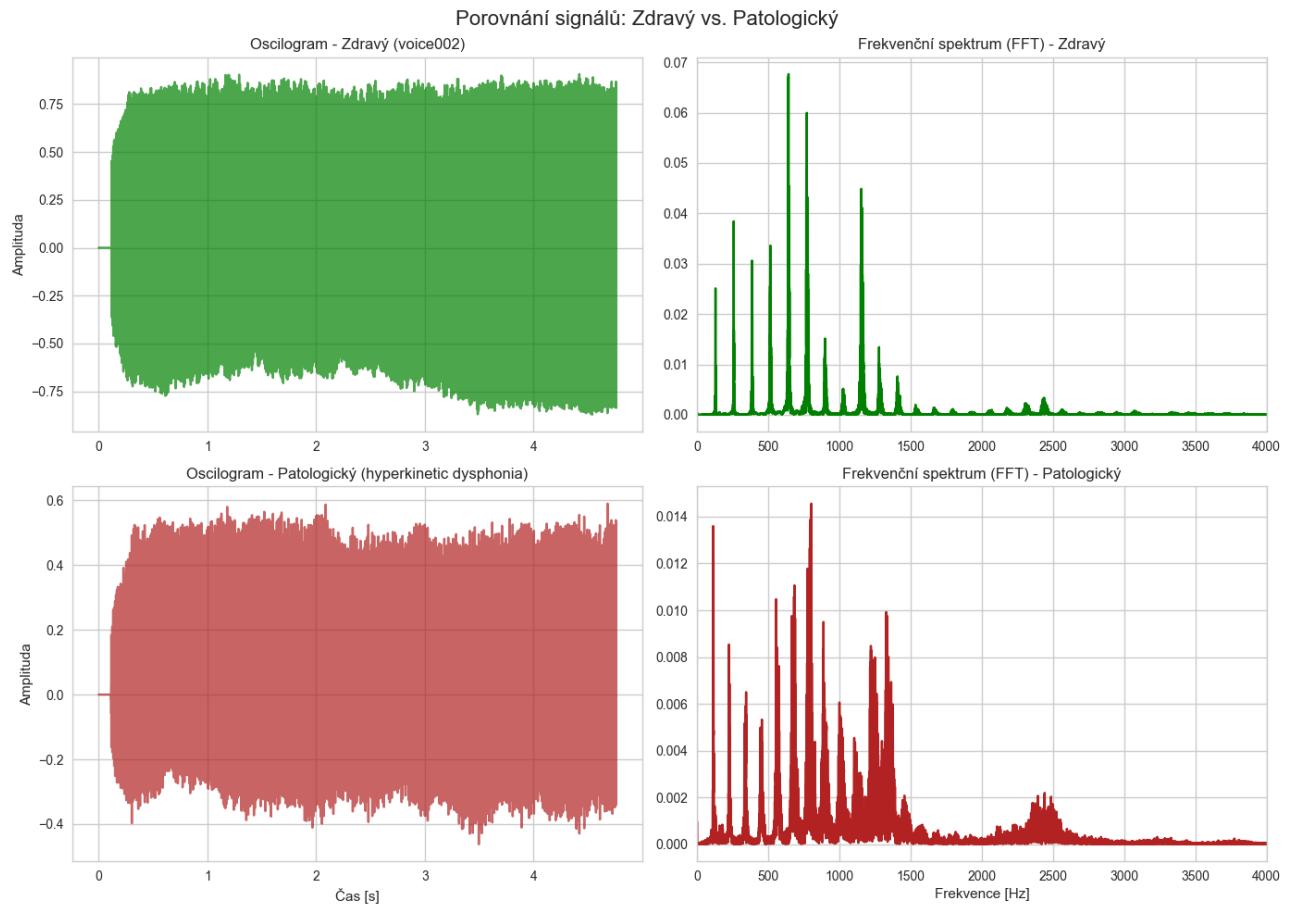
2.3 Klasifikace

Pro klasifikaci byl zvolen algoritmus **Random Forest Classifier** (100 stromů). Data byla rozdělena na trénovací a testovací sadu v poměru 80:20 se stratifikací. Proběhly dva experimenty: 1. **Binární klasifikace:** Zdravý vs. Patologický. 2. **Multiclass klasifikace:** Rozlišení konkrétních diagnóz (např. hyperkinetická dysfonie) pouze na podmnožině patologických dat.

3 Výsledky

3.1 Analýza signálu

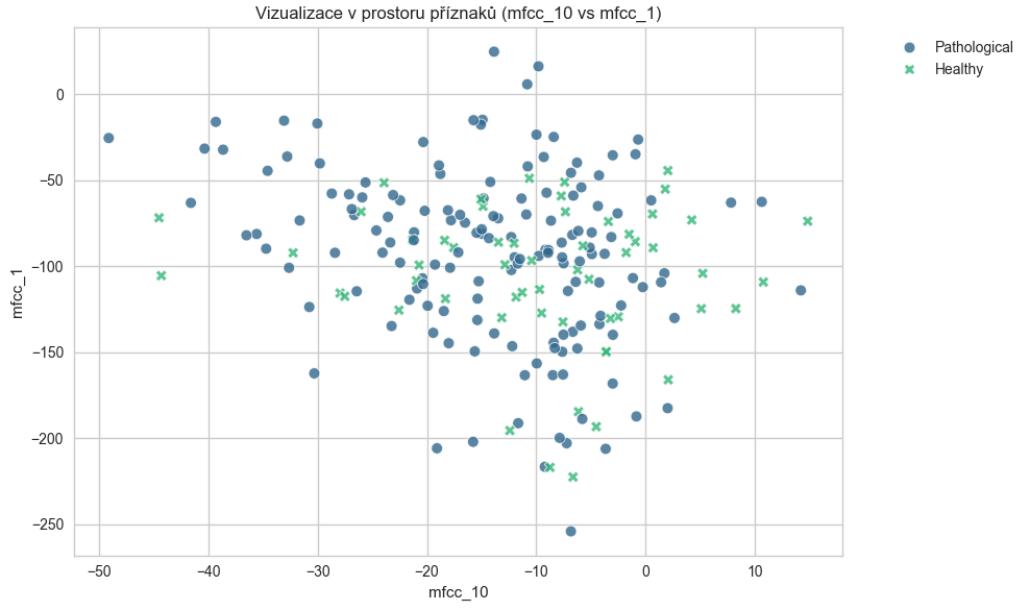
Následující grafy demonstруjí rozdíl mezi zdravým a patologickým hlasem. U patologického signálu (červeně) je patrná vyšší nepravidelnost v časové oblasti a "zašumnější" spektrum ve frekvenční oblasti oproti harmonickému zdravému hlasu.



Obrázek 3: Porovnání oscilogramu a frekvenčního spektra (FFT) zdravého a patologického hlasu.

3.2 Vizualizace v prostoru příznaků

Analýza hlavních příznaků ukázala, že třídy nejsou lineárně separabilní. Jak je vidět na scatterplotu níže (MFCC_10 vs MFCC_1), zdravé a patologické vzorky se výrazně překrývají, což indikuje náročnost úlohy.



Obrázek 4: Vizualizace klasifikace v prostoru dvou vybraných MFCC koeficientů.

3.3 Úspěšnost klasifikace

3.3.1 Binární klasifikace (Zdravý / Nemocný)

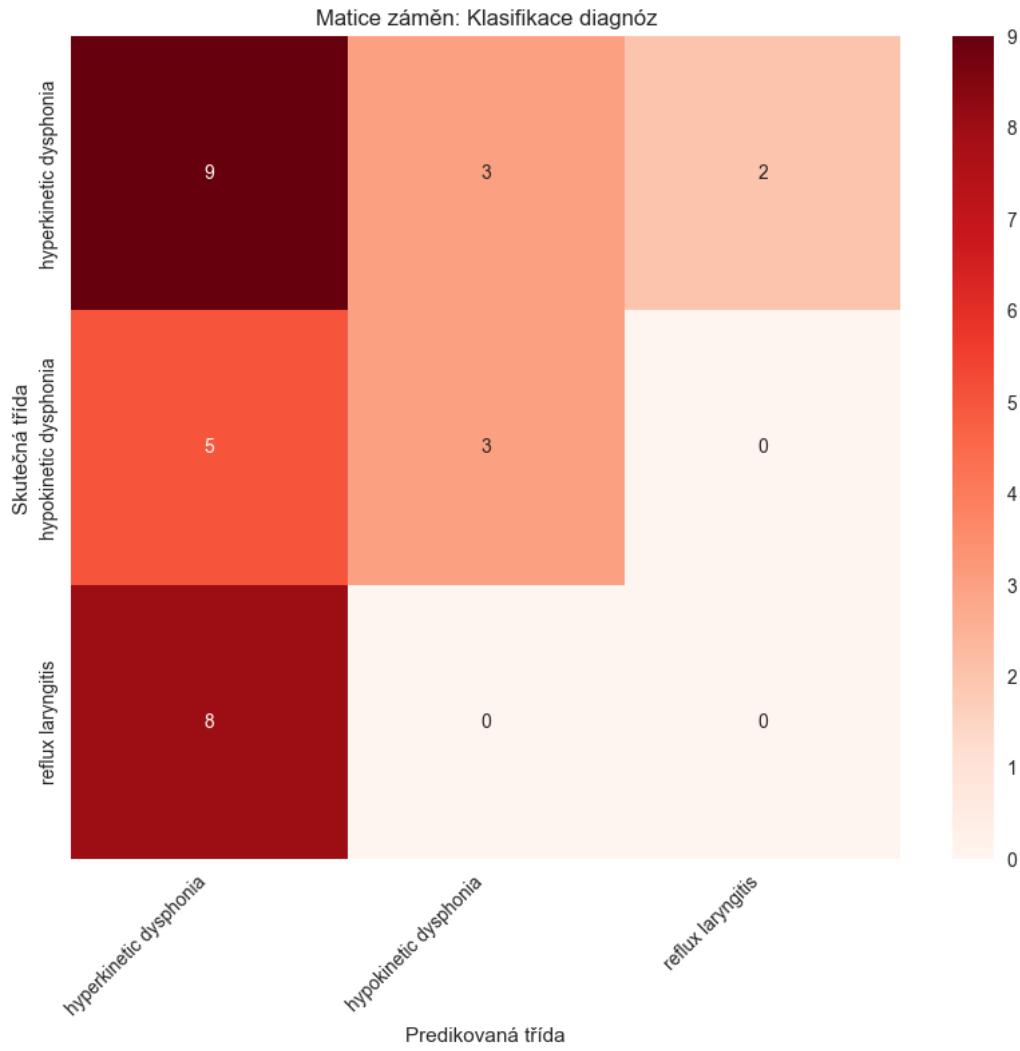
Model dosáhl celkové přesnosti (Accuracy) **64.29 %**. Z tabulky je patrné, že model lépe detekuje patologické hlasy (Recall 0.90), ale má problém se správným určením zdravých hlasů (Recall 0.00), což může být způsobeno nevyvážeností datasetu.

--- Výsledky BINÁRNÍ klasifikace ---					
Přesnost (Accuracy): 64.29%					
	precision	recall	f1-score	support	
Healthy	0.00	0.00	0.00	12	
Pathological	0.69	0.90	0.78	30	
accuracy			0.64	42	
macro avg	0.35	0.45	0.39	42	
weighted avg	0.49	0.64	0.56	42	

Obrázek 5: Detailní report úspěšnosti binární klasifikace.

3.3.2 Multiclass klasifikace (Konkrétní diagnózy)

Při rozlišování konkrétních nemocí (např. refluxní laryngitida, dysfonie) klesla přesnost na **40.00 %**. Matice záměn ukazuje časté záměny mezi podobnými diagnózami, např. mezi hyperkinetickou a hypokineticou dysfonií.



Obrázek 6: Matice záměn pro klasifikaci konkrétních patologií.

3.4 Závěrečné shrnutí

Souhrn dosažených výsledků je uveden v tabulce níže.

==== ZÁVĚREČNÉ VYHODNOCENÍ ===		
	Typ úlohy	Dosažená přesnost (Accuracy)
0	Binární (Zdravý / Nemocný)	64.29%
1	Multiclass (Konkrétní diagnóza)	40.00%

Obrázek 7: Finální porovnání úspěšnosti obou klasifikačních úloh.

4 Závěr a zhodnocení

V rámci této práce byla provedena kompletní analýza databáze VOICED. Podařilo se úspěšně načíst data, extrahovat příznaky včetně MFCC a natrénovat klasifikátor.

Dosažená přesnost v binární klasifikaci (cca 64 %) převyšuje náhodnou volbu, ale ukazuje na limity použitých příznaků pro takto komplexní signál. Hlavním problémem se ukázal být

výrazný překryv charakteristik zdravých a nemocných hlasů v prostoru MFCC koeficientů.

V případě klasifikace konkrétních nemocí (přesnost 40 %) se potvrdilo, že rozlišit specifické typy dysfonií pouze na základě základních spektrálních vlastností je velmi obtížné, zejména při malém počtu trénovacích vzorků pro jednotlivé diagnózy. Pro zlepšení výsledků by bylo vhodné v budoucnu využít pokročilejší metody, jako jsou neuronové sítě (CNN/RNN), nebo rozšířit sadu příznaků o jitter a shimmer.