



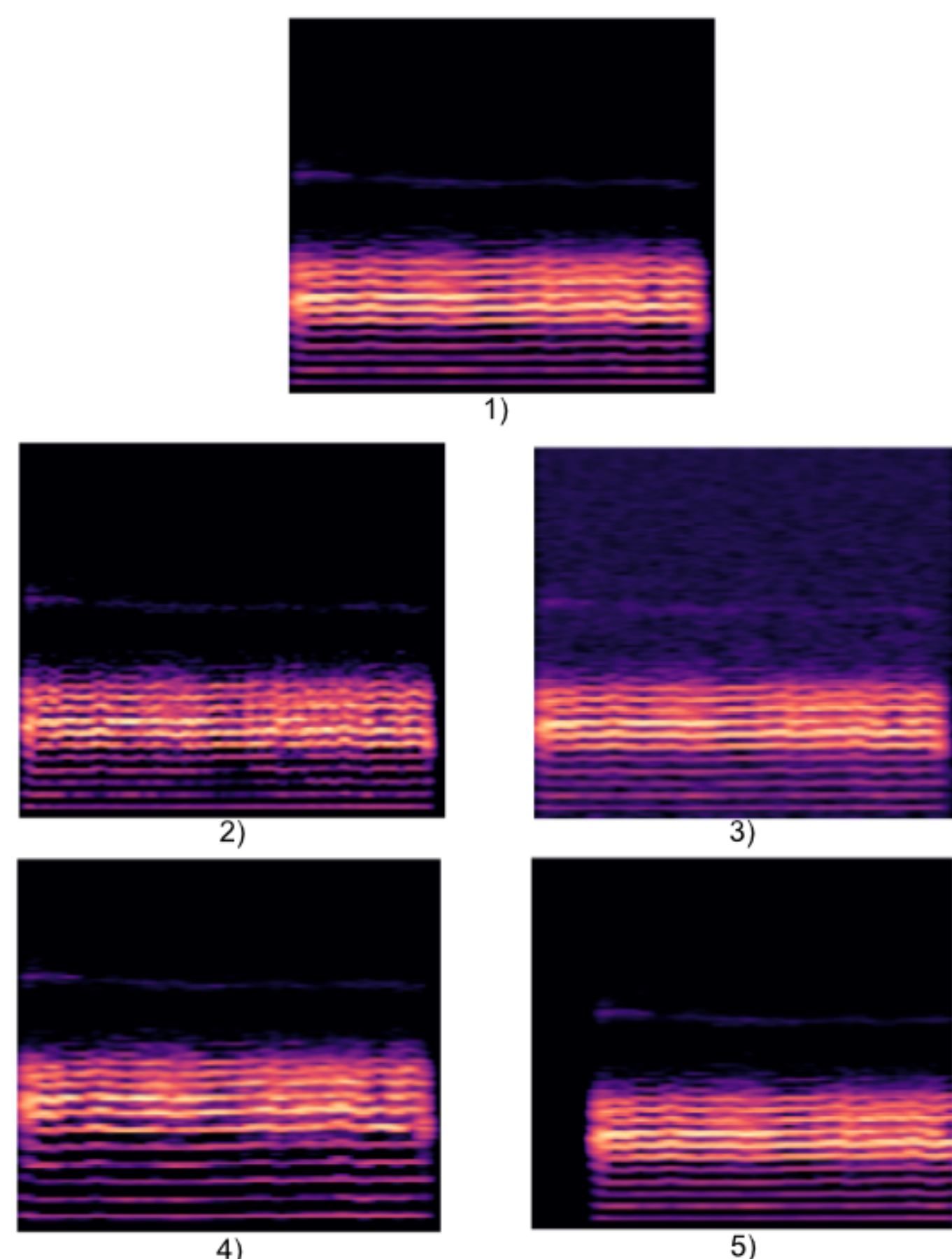
Motivácia a ciele

Spracovanie medicínskych dát je v poslednej dobe jednou z atraktívnych aplikácií neurónových sietí. Reálne využitie experimentálnych riešení je však častokrát limitované nedostatkom dát pre trénovanie modelov strojového učenia. Predpoklady nášho výskumu vychádzajú z možnosti predikcie Parkinsonovej choroby z rečových dát pomocou konvolučných neurónových sietí. Takýto model je trénovaný na dátach v určitom jazyku, pričom predikcia vzoriek v iných jazykoch vedie k rapídemu poklesu výkonnosti.

Pre spomínaný nedostatok dát je potrebné uvažovať o spôsoboch na zvýšenie robustnosti predikčného modelu. Hlavným cieľom práce bol návrh a realizácia spôsobu automatickej dátovej augmentácie.

Zvukové augmentácie

Jednotlivé techniky môžu mať potenciálne pozitívny vplyv na zvýšenie úspešnosti predikcie. V našom výskume sme overovali prínosy štyroch konkrétnych augmentácií s rôznymi magnitúdami. Na obrázku je ilustrovaný efekt techník: 1) pôvodný signál, 2) zmena rýchlosti, 3) pridanie šumu, 4) zmena výšky tónu a 5) časový posun na vzorku z úlohy vyslovenia samohlásky **a** po dobu niekoľkých sekúnd.



Po aplikovaní augmentácie je zvukový signál pre potreby modelu prevedený do obrázkovej podoby ako Mel spektrogram.

Prínosy navrhovaného riešenia

- **Predikcia PD z rečových dát.** Konverzia dát do Mel spektrogramov
- **Zvýšenie robustnosti modelu.** Nájdenie optimálnych augmentačných techník pre poskytnuté dáta, ktoré zvýšia efektívnosť modelu naprieč rôznymi jazykovými prostrediami.
- **Využitie pri iných úlohách.** Identifikácia emócie z reči, detekcia únavy z hlasu, či ďalšie úlohy strojového učenia zamerané na analýzu zvukových signálov.

Navrhnuté riešenie je unikátne tým, že ešte pred spracovaním signálu aplikuje augmentačné techniky priamo na zvukový signál. Existujúce riešenia doposiaľ boli zamerané len na obrázkové dáta a príslušné augmentácie.

Experimentálne vyhodnotenie

Počas realizácie riešenia boli použité štyri dátové sady v rôznych jazykoch - španielska, česká, anglická a talianska. Každá z nich obsahuje rôzny počet zvukových vzoriek v dvoch triedach - zdravý (HC) a pacient s Parkinsonovou chorobou (PD), pričom zaznamenané sú rozličné úlohy, najmä dlhé vyslovenia samohlások (napr. a, e, i, o, u).

Na overenie navrhnutého prístupu boli vykonané tri scenáre. Jednotlivé scenáre sa odlišujú použitím rôznych častí dostupných dátových sád, pričom každý scenár zahŕňa viacero sád experimentov s rôznou dátovou sadou použitou ako zdroj pre trénovanie modelu (inferencia prebehla zakaždým na všetkých štyroch sádach).

Pred začiatkom sady experimentov bola zadaná množina prehľadávaných augmentačných techník spolu s rôznymi magnitúdami (napr. pri technikách ako zmena rýchlosti má význam overovať zrýchlenie aj spomalenie), vrátane jedného experimentu bez aplikácie akejkoľvek augmentácie slúžiaci ako referenčná hodnota.

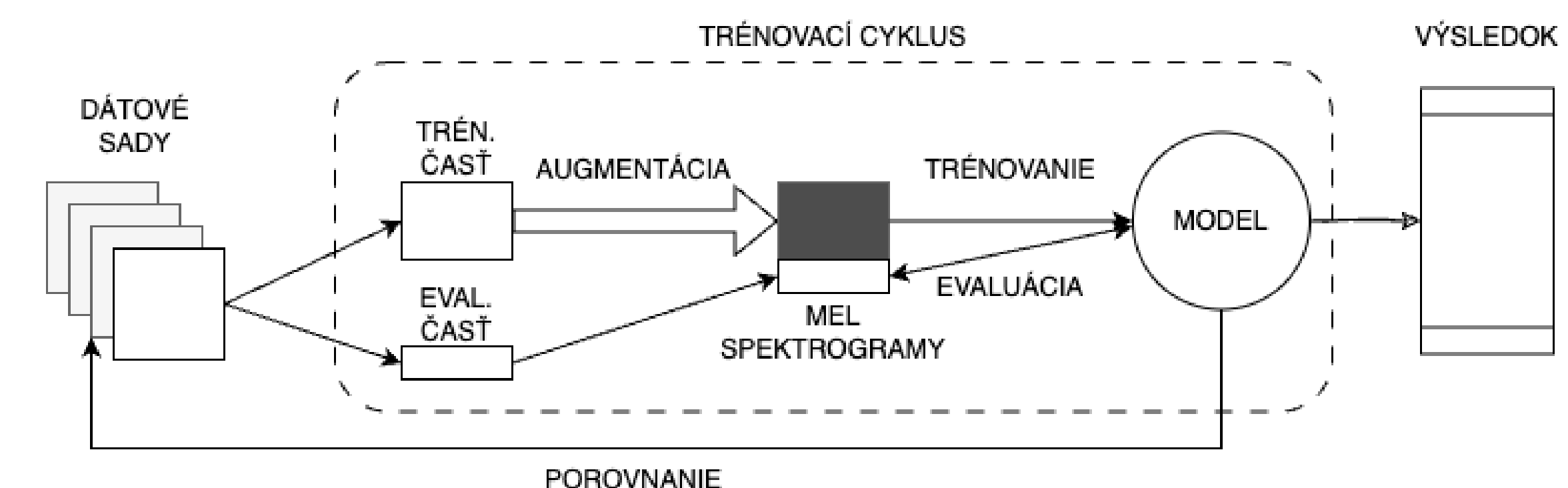
Augmentácia	Magnitúdy
Zmena rýchlosti	0,65, 2,5
Časový posun	3, 6
Pridanie šumu	8, 1
Zmena výšky tónu	5, 9

Pre ilustráciu sú uvedené výsledky experimentov, kde dátové sady sú tvorené úlohami spoločnými pre všetky dátové sady (konkrétne vyslovenie samohlások **a**, **u**) a zdrojovou sadou bola česká dátová sada. Nasledujúca tabuľka je výstupom programu, kde pre každú dátovú sadu je uvedená augmentácia, ktorá spôsobila najvýraznejšie zlepšenie spomedzi overovaných techník, vrátane hodnoty sledovanej metriky (v tomto prípade *binary accuracy*) a dodatočne vypočítaného percentuálneho zlepšenia.

Dátová sada	Najlepšia augmentácia	ACC	Zlepšenie (%)
SP	Add noise (1)	0,781	37,5
CZ	Time Stretch (2.5)	0,688	25,0
EN	Add Noise (1)	0,75	34,4
IT	Time Stretch (0.65)	0,719	31,3

Automatická augmentácia

Keďže ručné hľadanie najefektívnejších augmentácií pre danú dátovú sadu je náročné z časového hľadiska aj neefektívne z pohľadu ľudských zdrojov, vhodným riešením je automatizovaný prístup.



Pred spustením programu je zo strany používateľa špecifikovať sadu augmentačných techník s príslušnými magnitúdami, zvoliť sledovanú metriku na porovnanie (dostupné: *binary accuracy*, *AUC*, *loss*) a určiť zdrojovú dátovú sadu. Počet augmentácií definuje počet experimentov, pre ktoré sa vykoná nasledujúci postup.

V prvom kroku sa zdrojová sada rozdelí v určenom pomere na trénovaciu a evaluačnú časť. Na vzorky v trénovej časti je aplikovaná jedna z augmentačných techník a vzápätí sú obe časti prevedené do Mel spektrogramov.

Zo spektrogramov je vytvorená nová dátová sada na trénovanie predikčného modelu. Po ukončení je vyhodnotená efektívnosť modelu na všetkých štyroch sádach a sú zaznamenané výsledky metrik.

Po ukončení všetkých experimentov je okrem kompletného záznamu metrik dostupný hlavný výsledok - výpis najlepších augmentačných techník pre jednotlivé sady obsahujúci aj hodnotu metriky.

Architektúra modelu

Spektrogramy v obrázkovej podobe spracováva konvolučná neurónová sieť na báze architektúry modelu Xception - využíva sa prenosové učenie. Predikčný model je zložený z týchto vrstiev:

1. **vrstva na predspracovanie vstupu,**
2. **model Xception** bez klasifikačných vrstiev,
3. **združovacia vrstva,**
4. **hlboko-prepojená vrstva** (400 neurónov, ReLU),
5. **Dropout vrstva** (0.2),
6. **predikčná vrstva** (1 neurón, sigmoid).

Trénovanie prebieha maximálne po dobu 100 epoch (využitie techniky skorého zastavenia pokiaľ nedochádza k zlepšeniu metriky počas 15 epoch), s veľkosťou dávky 32.