

VYUŽITIE HLBOKÝCH NEURÓNOVÝCH SIETÍ PRE ROZPOZNANIE PATOLOGICKEJ REČI



Ing. Lukáš Vavrek | doc. Ing. Peter Drotár PhD.

PROBLÉM

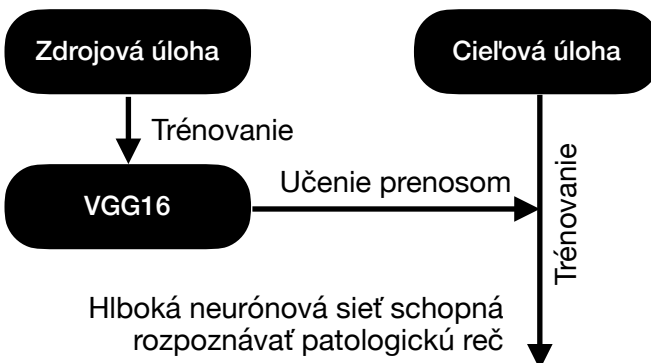
Detekcia patologickej reči je časovo a finančne náročný proces. Diagnostiku vykonáva lekár, špecialista, za pomoci špecializovaných medicínskych nástrojov.

Včasná a presná detekcia patologického hlasu je dôležitá, nakoľko patologický hlas môže byť indikátorom viacerých vážnych zdravotných problémov, napríklad rakoviny hrtana a rovnako má negatívny vplyv aj na duševné zdravie pacienta.

NÁVRH RIEŠENIA

Pre všetky vykonané experimenty sme využívali dátovú sadu Saarbrücken Voice Database, ktorá obsahuje hlasové nahrávky viac ako 2000 ľudí. Dáta sme redukovali iba na tzv. organické poruchy hlasu, spôsobené štruktúrnymi zmenami v hlasivkách.

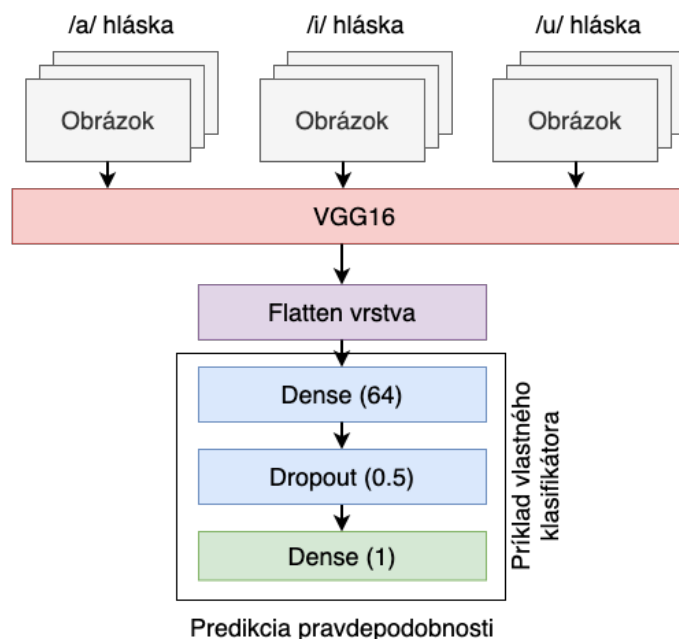
Jednotlivé hlasové vzorky sme spracovávali vo forme spektrogramov.



Analyzovali sme existujúce publikované metódy a navrhli vlastné riešenia na báze využitia hlbokých neurónových sietí.

Ako základ sme použili metódu učenia prenosom (transfer learning). Táto metóda nám umožňuje využívať predtrénované hlboké neurónové siete, naučené riešiť zdrojový problém a pomocou rôznych techník učenia prenosom siete adaptovať na riešenie cieľového problému. Môžeme tak využiť existujúcu znalosť použitých sietí a dosiahnuť dobré výsledky aj v prípade, keď máme k dispozícii len obmedzené množstvo dát.

Multi-vstupová sieť spracúva jednotlivé vzorky subjektov oddelene, samostatnými VGG16 sieťami.



Multi-dátová sieť naopak jednotlivé vzorky vo forme spektrogramov kóduje do jediného obrázka, na úrovni RGB vrstiev.

VYHODNOTENIE

Počas overovania našich experimentov sme dosahovali výsledky porovnateľné s najnovšími publikovanými článkami. Overili sme, že sieť VGG16, ale aj iné podobné CNN hlboké neurónové siete, sú schopné sa adaptovať na detekciu patologickej reči.

Porovnanie výsledkov experimentov

	Presnosť	Špecifickosť	Citlivosť
VGG 16 jednohlásková	79.14 %	0,775	0,807
VGG 16 multi-vstupová	76.26 %	0,714	0,812
VGG 16 multi-data	74.82 %	0,686	0,812
VGG 16 odladená multi-data	76.98 %	0,714	0,826
DenseNet multi-data	71.22 %	0,671	0,754
NASNet multi-data	72.66 %	0,714	0,739
VGG16 súbor modelov	82.01 %	0,843	0,797

Najlepší dosiahnutý výsledok predstavuje zlepšenie o 2,5% v porovnaní s publikovanými výsledkami^[1], za použitia menšieho množstva dát a dosahuje presnosť detekcie patologickej reči na úrovni 82%.

[1] David Martínez et al. "Voice Pathology Detection on the Saarbrücken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit"