

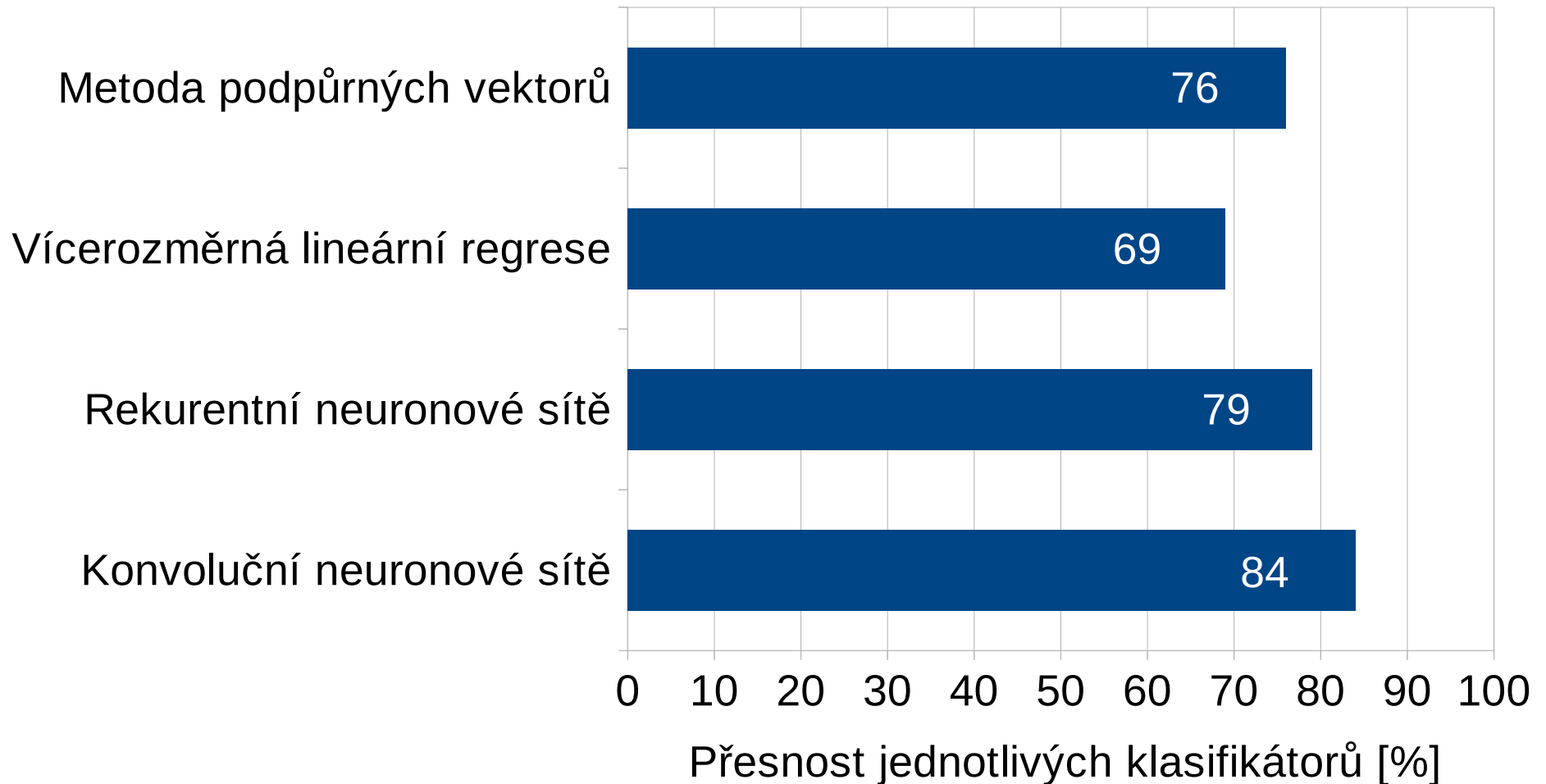
# **Expertní systém pro rozpoznávání dominantní projevované emoce z hlasu**

Jakub Šmíd  
Kybernetika a robotika  
2022

# Cíle práce

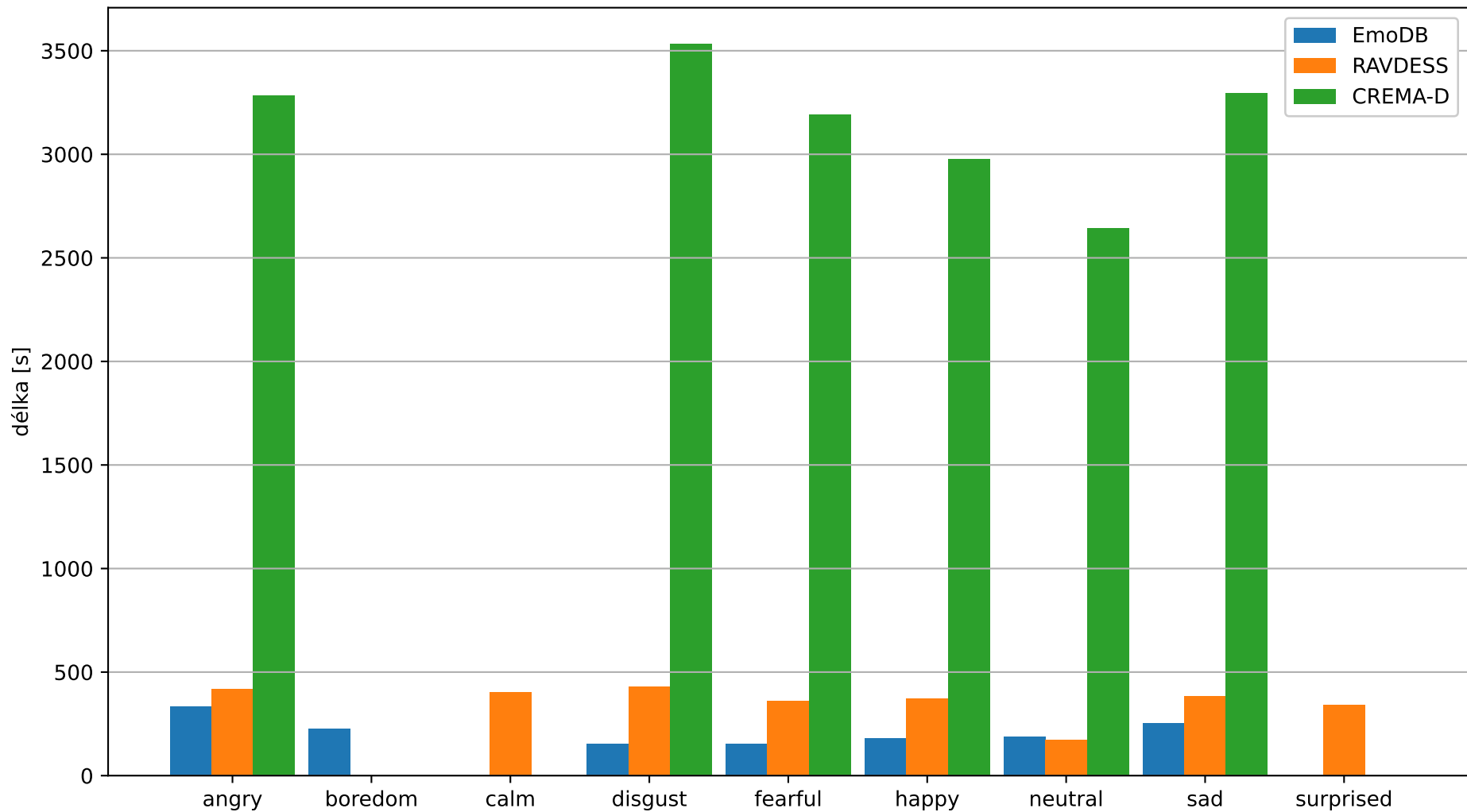
- Analýza dostupných datasetů
- Návrh metody vč. architektury, augmentace a extrakce příznaků
- Návrh parametrů systému a srovnání výsledků s ostatními řešeními
- Testování na větě, která není součástí trénovací a validační množiny
- Testování systému na simulaci UHF/VHF vysílaček
- Implementace demonstrační aplikace

# Používané metody a jejich výsledky



# Dostupné datasety

Rozdělení časů nahrávek



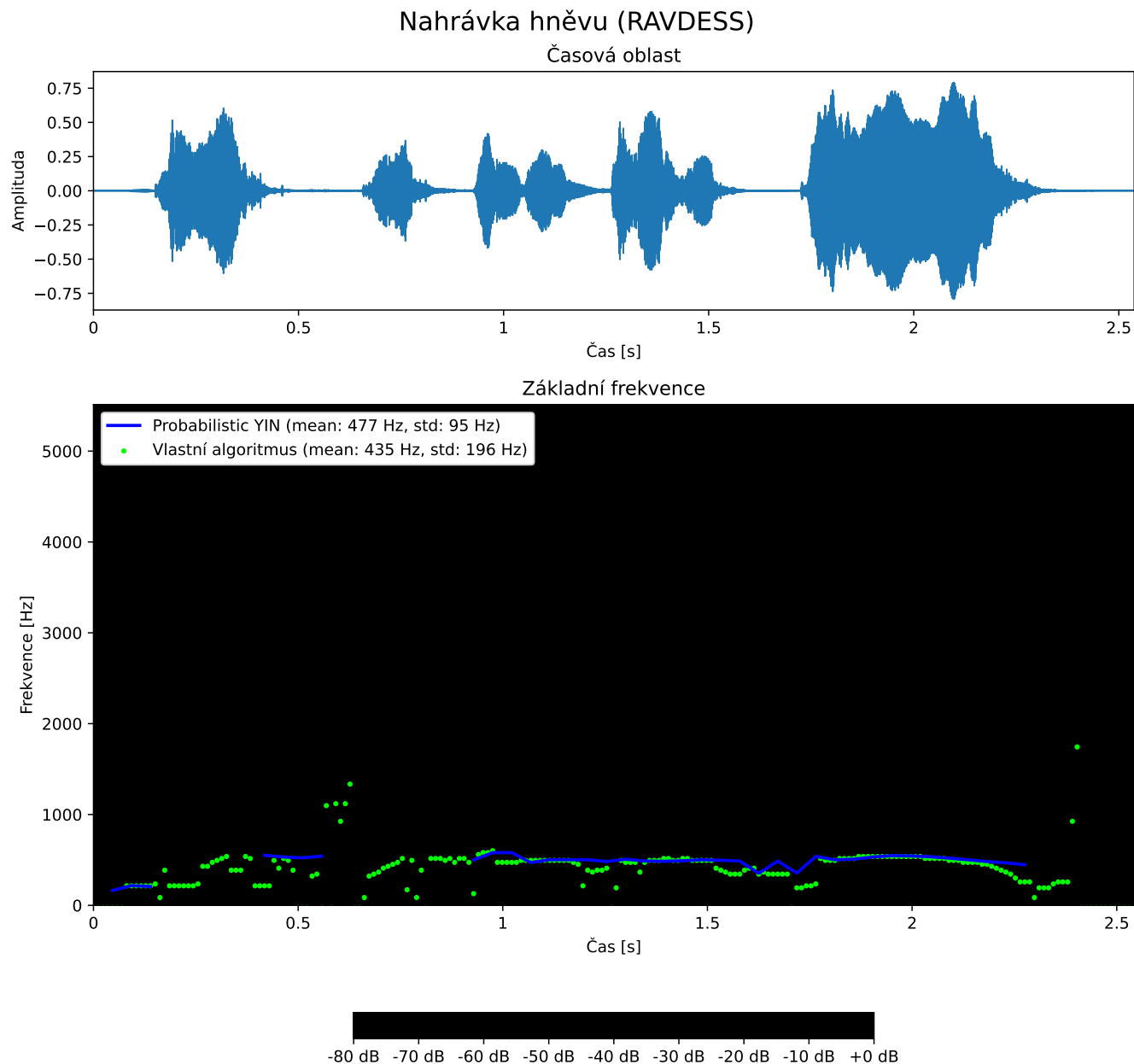
# Augmentace

- Přidání Gaussovského šumu
- Posunutí nahrávky v čase (roll)
- Posunutí nahrávky ve frekvenční oblasti
- Změna rychlosti nahrávky

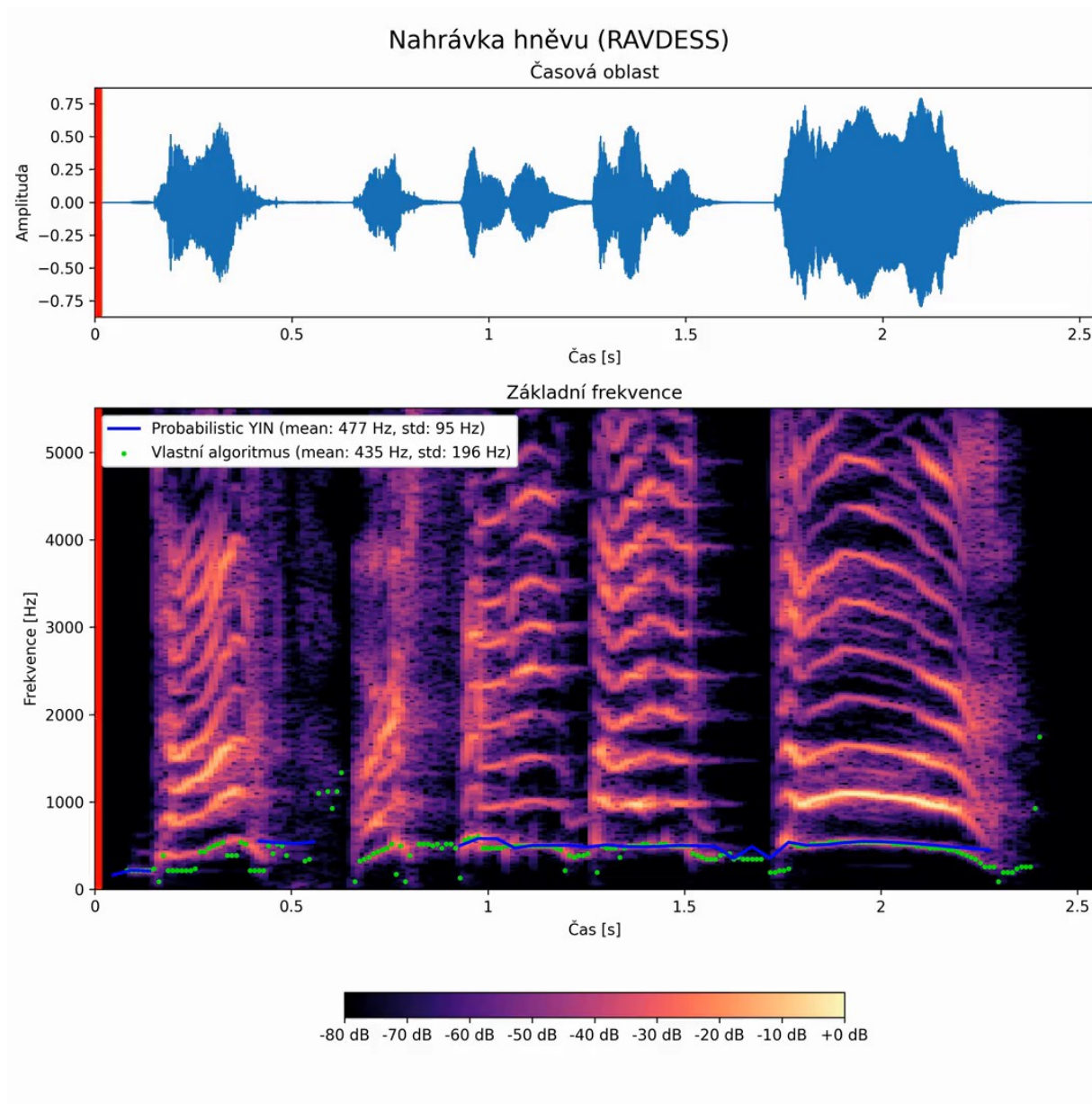
# Extrahované příznaky

- Zero Crossing Rate (ZCR)
- Root Mean Square (RMS)
- Spektrogram a Mel spektrogram
- Chromagram
- Mel frekvenční keprální koeficienty (MFCC)
- Průměr a směrodatná odchylka základní frekvence

# Příklad příznaku – zákl. frekvence



# Příklad příznaku – zákl. frekvence





# Použitý software



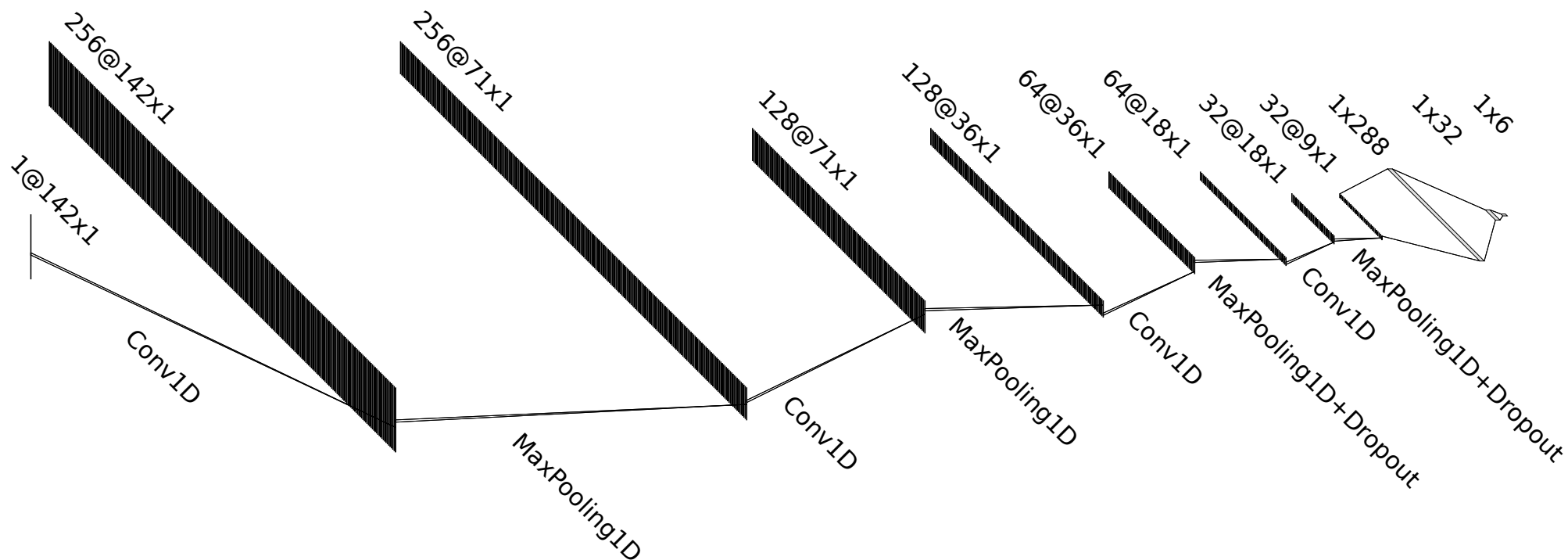
# Výběr příznaků – 1D konv. síť

- Příznaky: **ZCR, RMS, Mel spektrogram, Chromagram** (vektor o velikosti 142)
- Počet parametrů sítě: 226 246

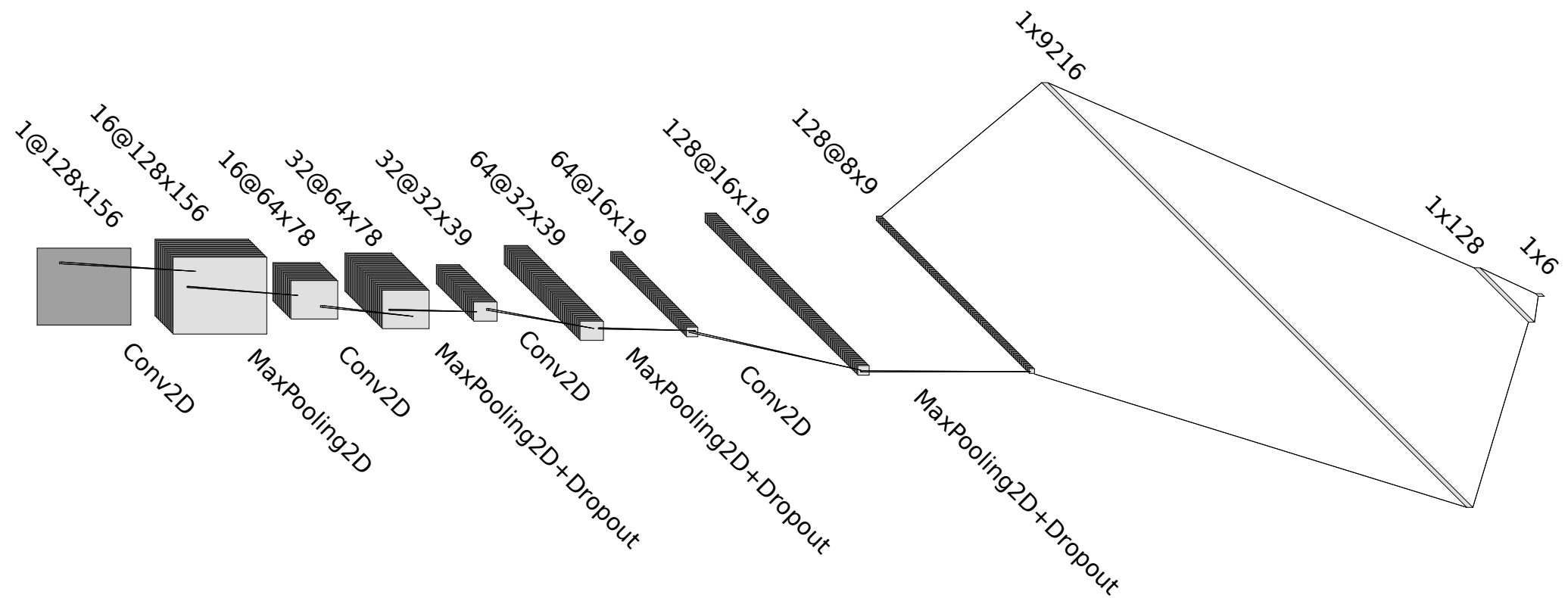
## Spektrogram – 2D konv. síť

- Příznaky: **amplitudový spektrogram** (vektor o velikosti 128x156)
- Počet parametrů sítě: 1 277 702

# Výběr příznaků – 1D konv. síť



# Spektrogram – 2D konv. síť

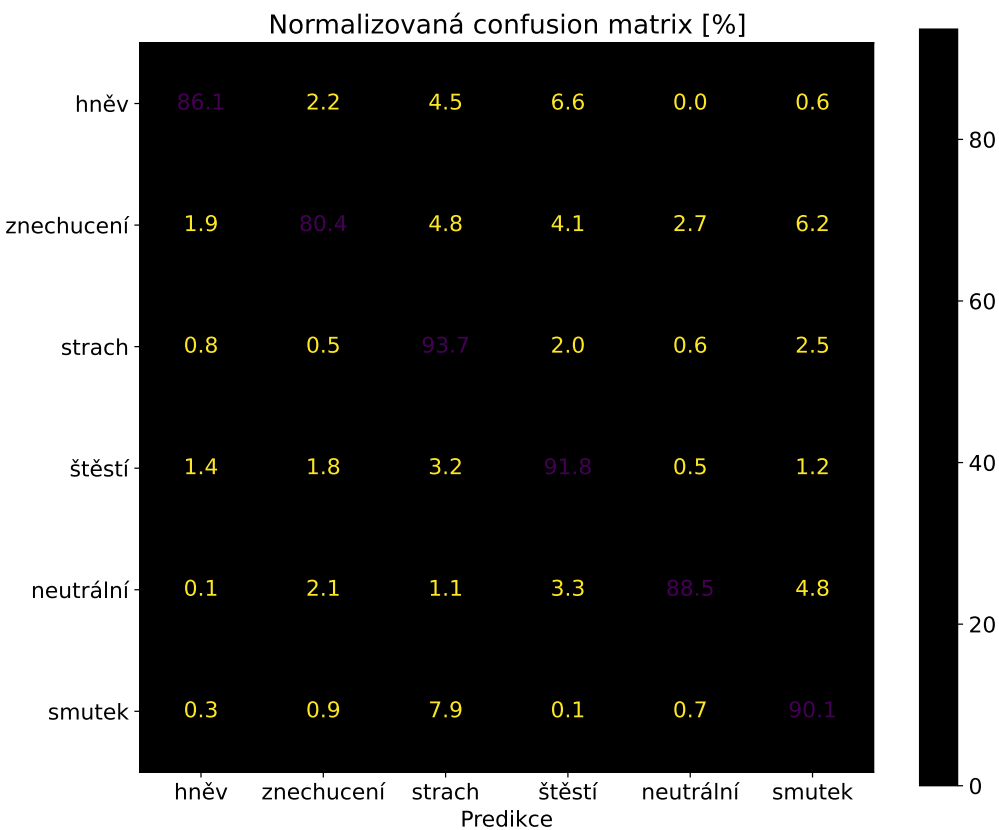


# Výsledky

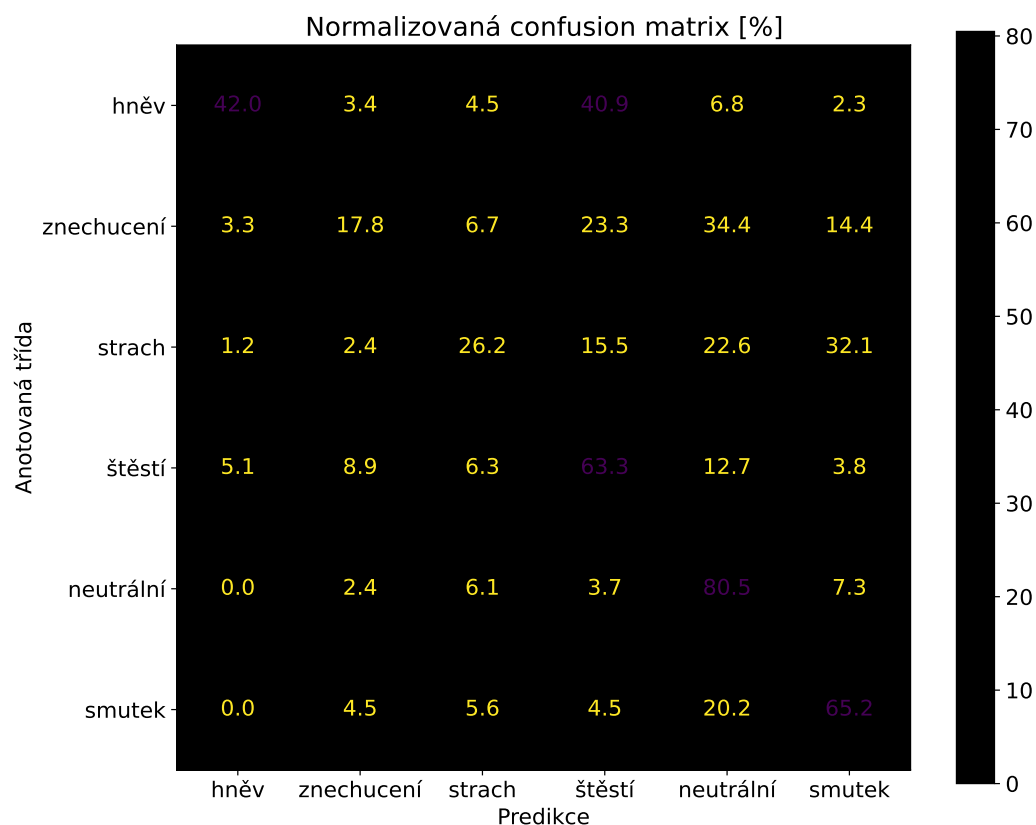
Metoda	Dataset	Přesnost sítě [%]
Výběr příznaků	validace	88.23
Výběr příznaků	testování bez šumu	48.83
Výběr příznaků	testování se šumem	49.41
Spektrogram	validace	76.50
Spektrogram	testování bez šumu	52.54
Spektrogram	testování se šumem	51.17

# Výsledky – výběr příznaků

Validace

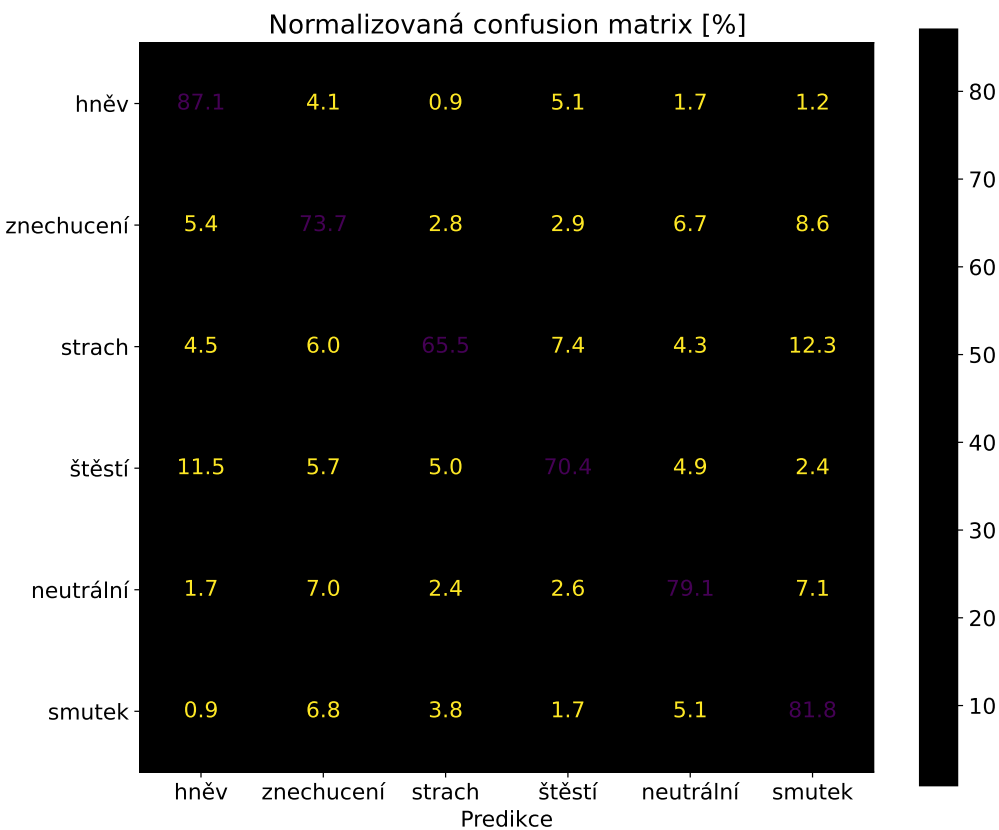


Testování  
(bez šumu)

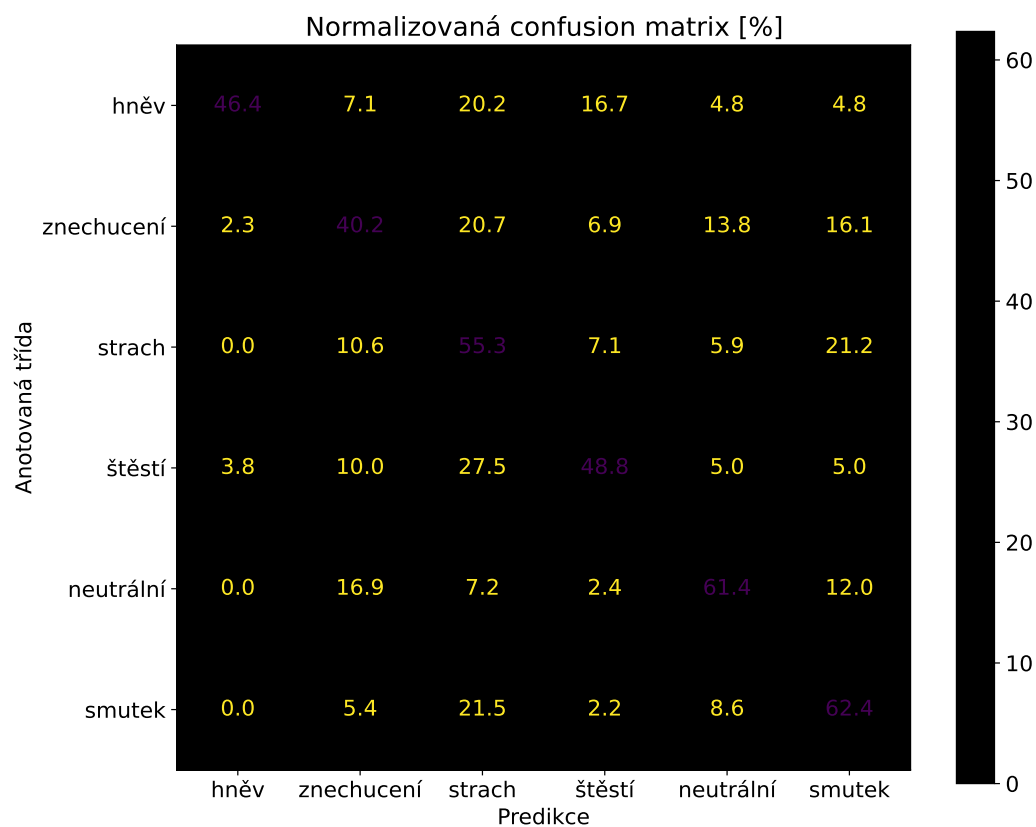


# Výsledky – spektrogram

Validace



Testování  
(bez šumu)



# Závěr

- Provedena analýza 3 datasetů
- Provedena analýza používaných klasifikátorů
- Provedena analýza příznaků a augmentace zvuku
- Navrženy a otestovány 2 architektury konvoluční NN
- Přesnost validace sítě až 88,2 %
- Přesnost otestované sítě až 52,5 %
- Implementována demonstrační aplikace

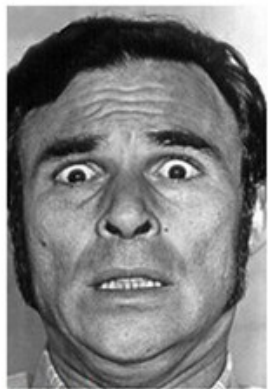




**Děkuji za pozornost**

# Přenositelnost na jiné jazyky

- Klasifikace do tříd základních emocí
- Základní emoce jsou shodné napříč kulturami
- Paul Ekman a Wallace V. Friesen. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology. 1971



Fearful



Angry



Sad



Happy



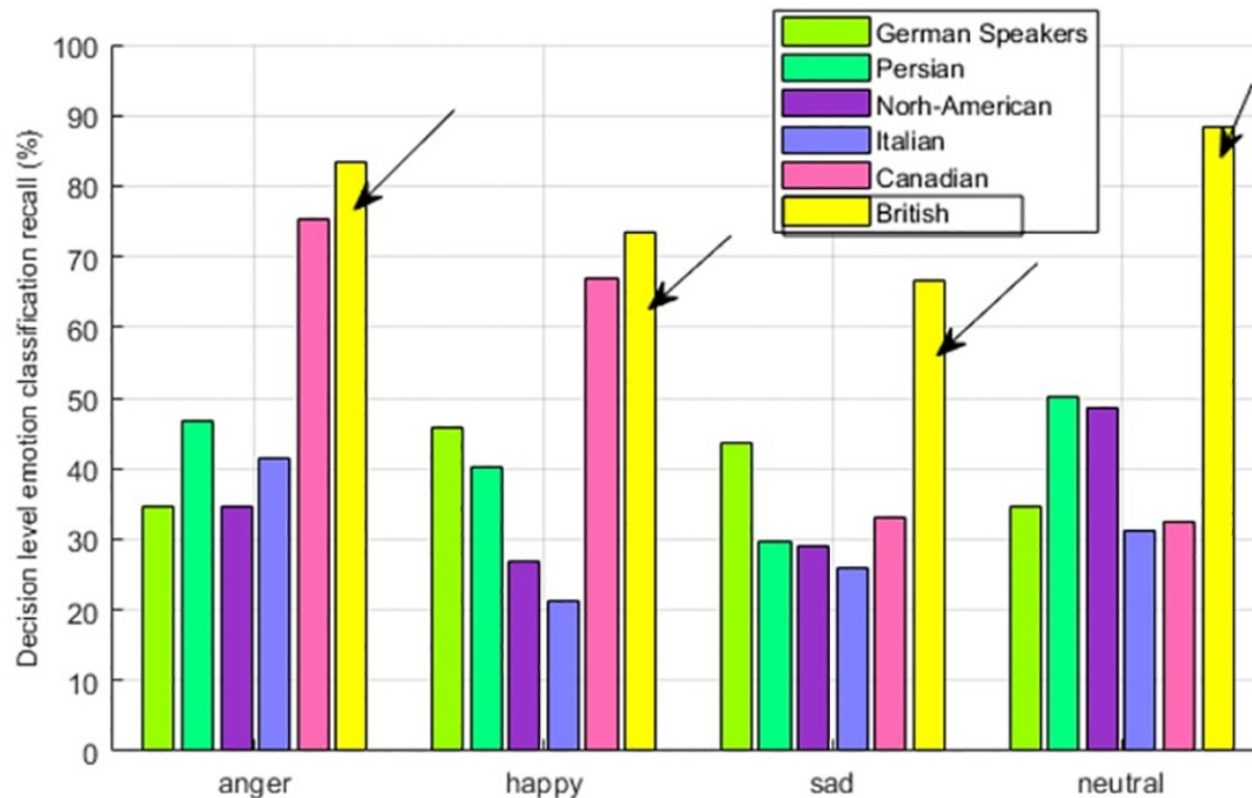
Disgusted



Surprised

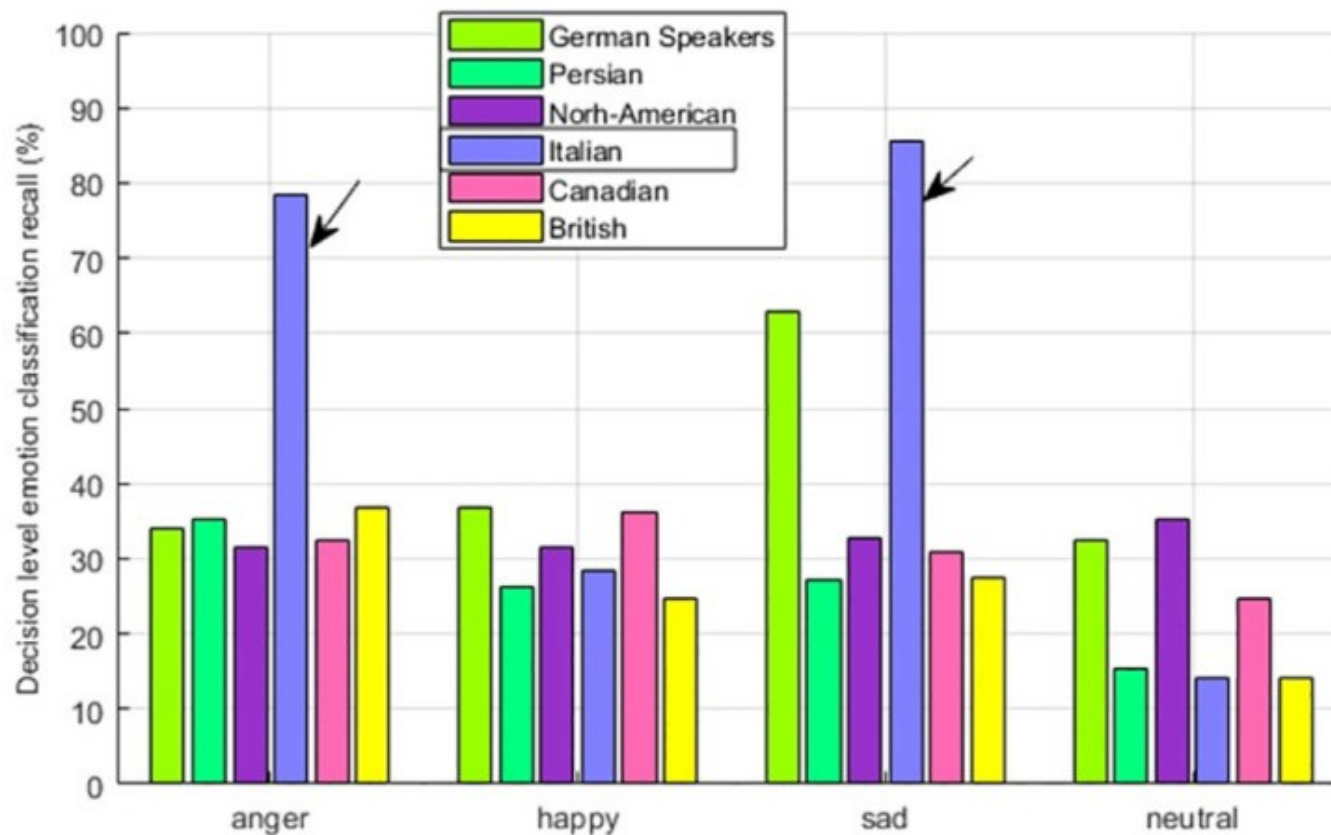
# Přenositelnost na jiné jazyky

- Kanwal S, Asghar S, Hussain A, Rafique A (2022) Identifying the evidence of speech emotional dialects using artificial intelligence: A cross-cultural study.
- $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$



# Přenositelnost na jiné jazyky

- Kanwal S, Asghar S, Hussain A, Rafique A (2022) Identifying the evidence of speech emotional dialects using artificial intelligence: A cross-cultural study.
- Recall =  $TP / (TP + FN)$



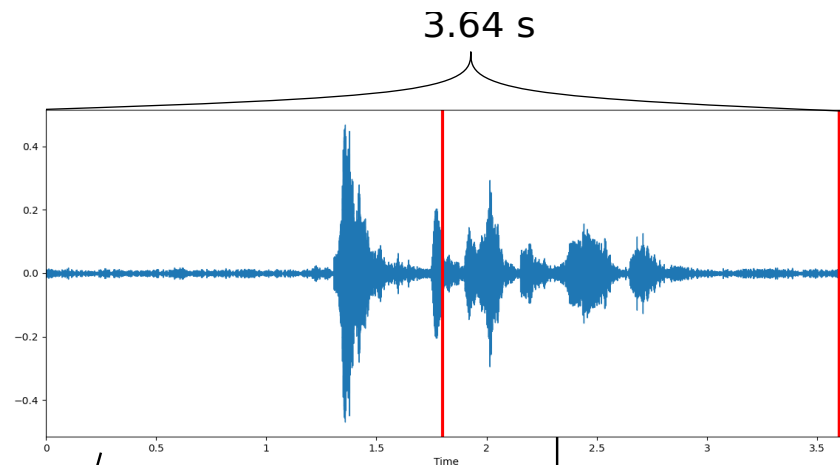
# Zvolení výsledné architektury

- Odladění parametrů při extrakci příznaků
- Inspirace sítí AlexNet
- Validace klasifikátoru s různými kombinacemi vrstev a různými parametry vrstev
- Dosažení největší přesnosti při validaci
- Testování na finální uvedené architektuře

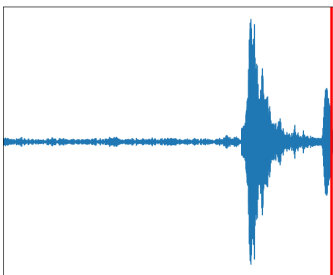
# Nahrávky delší než 1,8 s

Dataset: CREMA-D

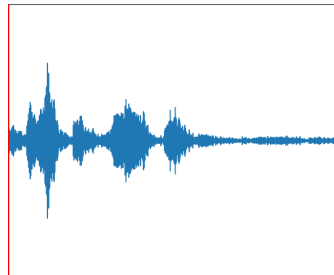
Věta: Don't forget a jacket



1.8 s



1.8 s



0.4 s



generování vstupního  
vektoru

generování vstupního  
vektoru

zahození