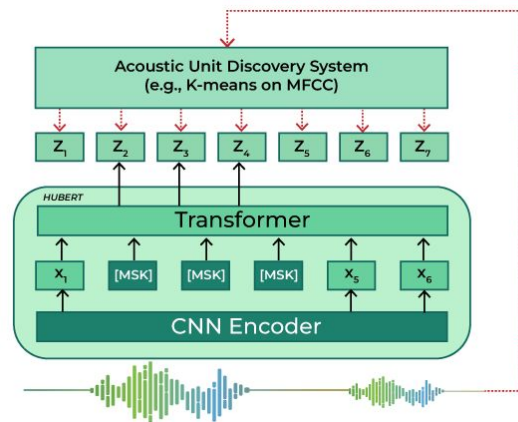
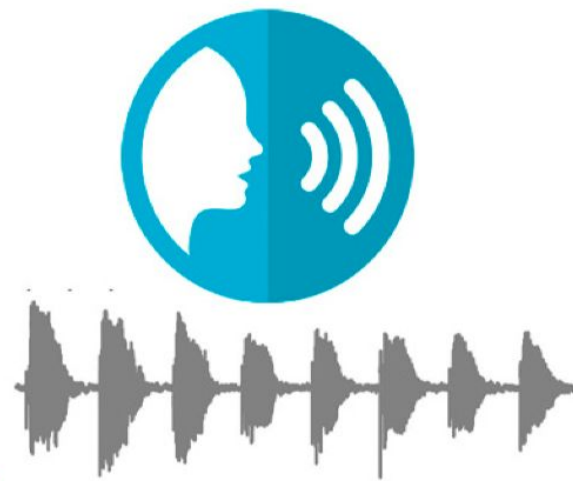


# AI v diagnostice neurodegenerativních onemocnění s fokusem na předtrénované řečové modely



klempond@fbmi.cvut.cz



## Ondřej Klempíř

### CEEOR / FBMI ČVUT

- Medicínská informatika
- Analýza patologické řeči
- Elektronické zdravotní záznamy
- Bioinformatika/genomika

### Klempir, Ondrej

[Czech Technical University in Prague](#), Prague, Czech Republic • Scopus ID: 57195508277 •  [0000-0003-0773-5360](#) ↗

[Show all information](#)

430

Citations by **406** documents

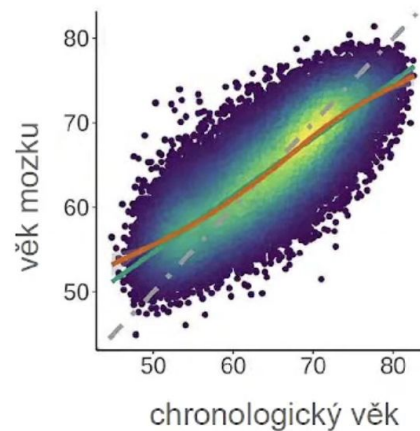
28

Documents

8

[h-index](#)

# JAK STARÝ JE VÁŠ MOZEK?





**Valerio Capraro**  · 3rd+

Associate Professor at the University of Milan Bicocca

13h · 

[+ Follow](#)

Just out in Nature Medicine.

A new AI model can predict the risk of 130 diseases from a single night of sleep.


The authors trained a foundation model on 585,000 hours of overnight sleep recordings collected from 65,000 individuals.

These data were exceptionally rich, combining brain activity, cardiac signals, muscle tone, and respiratory patterns, and were linked to participants' electronic health records.

Impressively, the model predicted the risk of a wide range of conditions, including dementia, heart failure, kidney disease, and stroke...

... YEARS before clinical diagnosis.

Another great example of how AI can transform early disease detection and preventive medicine.






**nature medicine** 

Article <https://doi.org/10.1038/s41591-025-04133-4>

## A multimodal sleep foundation model for disease prediction

---




Received: 3 February 2025  
 Accepted: 18 November 2025  
 Published online: 06 January 2026

**Rahul Thapa**<sup>1,2,10</sup>, **Magnus Ruud Kjaer** <sup>3,4,5,10</sup>, **Bryan He**<sup>2</sup>, **Ian Covert**<sup>2</sup>, **Hyatt Moore IV**<sup>3,4</sup>, **Umaer Hanif**<sup>5,7</sup>, **Gauri Ganjoo**<sup>2</sup>, **M. Brandon Westover**<sup>8</sup>, **Poul Jennum**<sup>9,9</sup>, **Andreas Brink-Kjaer**<sup>8</sup>, **Emmanuel Mignot** <sup>3,11</sup>  & **James Zou** <sup>1,2,11</sup> 

[Check for updates](#)

---

Sleep is a fundamental biological process with broad implications for physical and mental health, yet its complex relationship with disease remains poorly understood. Polysomnography (PSG)—the gold standard for sleep analysis—captures rich physiological signals but is underutilized due to challenges in standardization, generalizability and multimodal integration. To address these challenges, we developed SleepFM, a multimodal sleep foundation model trained with a new contrastive learning approach that accommodates multiple PSG configurations. Trained on a curated dataset of over 585,000 hours of PSG recordings from approximately 65,000 participants across several cohorts, SleepFM produces latent sleep representations that capture the physiological and temporal structure of sleep and enable accurate prediction of future disease risk. From one night of sleep, SleepFM accurately predicts 130 conditions with a C-Index of at least 0.75 (Bonferroni-corrected  $P < 0.01$ ), including all-cause mortality (C-Index, 0.84), dementia (0.85), myocardial infarction (0.81), heart failure (0.80), chronic kidney disease (0.79), stroke (0.78) and atrial fibrillation (0.78). Moreover, the model demonstrates strong transfer learning performance on a dataset from the Sleep Heart Health Study—a dataset that was excluded from pretraining—and performs competitively with specialized sleep-staging models such as U-Sleep and YASA on common sleep analysis tasks, achieving mean  $F_1$  scores of 0.70–0.78 for sleep staging and accuracies of 0.69 and 0.87 for classifying sleep apnea severity and presence. This work shows that foundation models can learn the language of sleep from multimodal sleep recordings, enabling scalable, label-efficient analysis and disease prediction.

   142 11 comments · 13 reposts

# Co nás čeká? Matematické problémy / zadání

- Patologická řeč
- Velké jazykové modely vs. řečové modely
- Klasifikační úlohy
- Regresní úlohy
- Přepis řeči na text

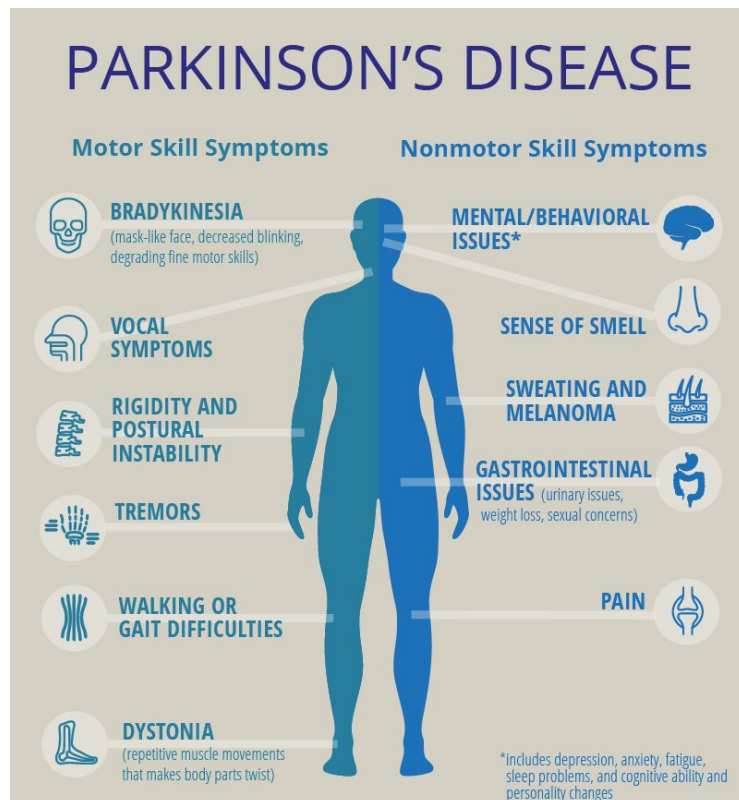
otevřené výzvy included... ?!

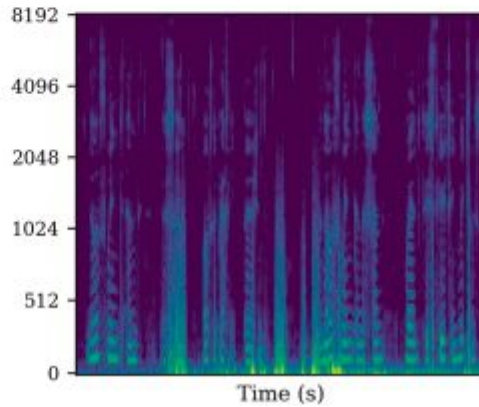
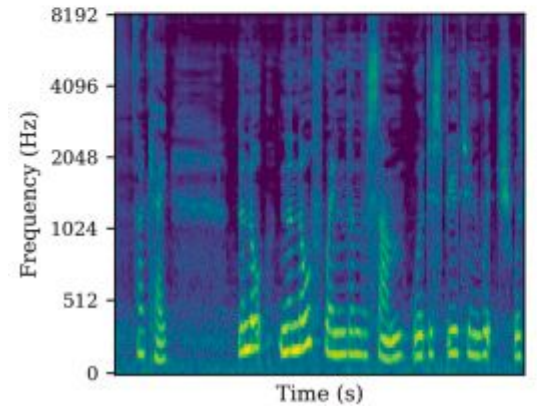
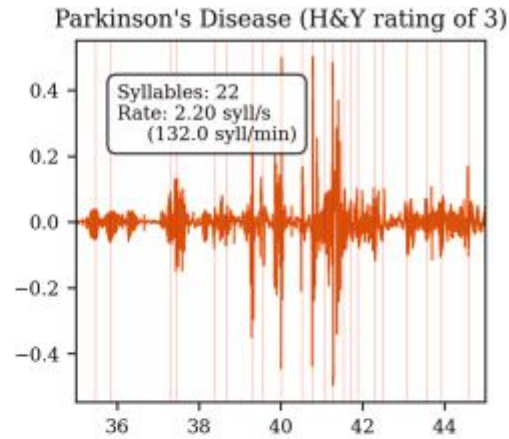
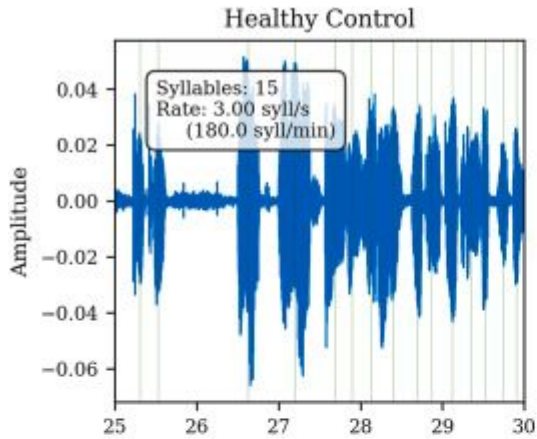
# Parkinsonova nemoc a poruchy řeči

- Poruchy pohybu
  - Stuhlost
  - “Zamrzání”
- Poruchy řeči
- Prodromální fáze

## Standardizovaná vyšetření

- Spontánní
- Čtený text (fixní)
- Pataka
- /A/
- Interview

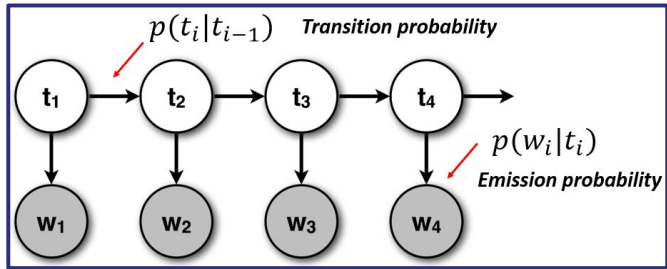




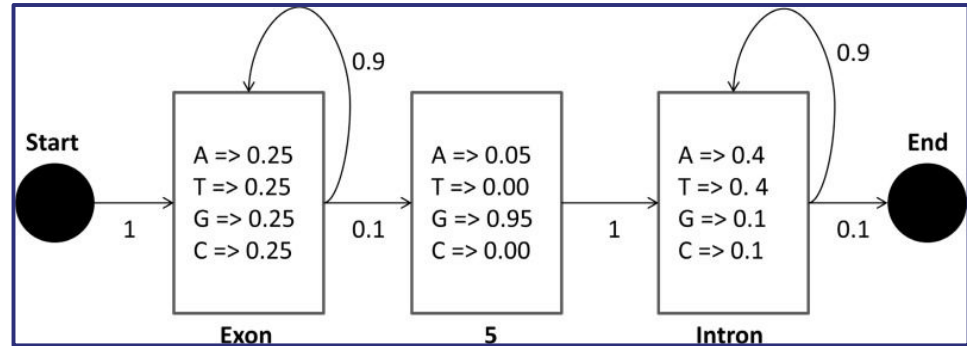
Feature	Statistic	Domain
EnergyEntropy	std	time
SignalEnergy	std/mean	time
ZeroCrossingRate	std	time
SpectralRolloff	std	spectral
SpectralCentroid	std	spectral
SpectralFlux	std/mean	spectral

O. Klempir, J.G. Mullerova, R. Krupicka, Statistical, multi-scale and attention-based layer pooling of Wav2Vec-2 speech embeddings for Parkinson's disease detection, *Computers In Biology And Medicine*. 200 (2026) 111368.  
<https://doi.org/10.1016/j.combiomed.2025.111368>.

# “Language” model pro řeč?

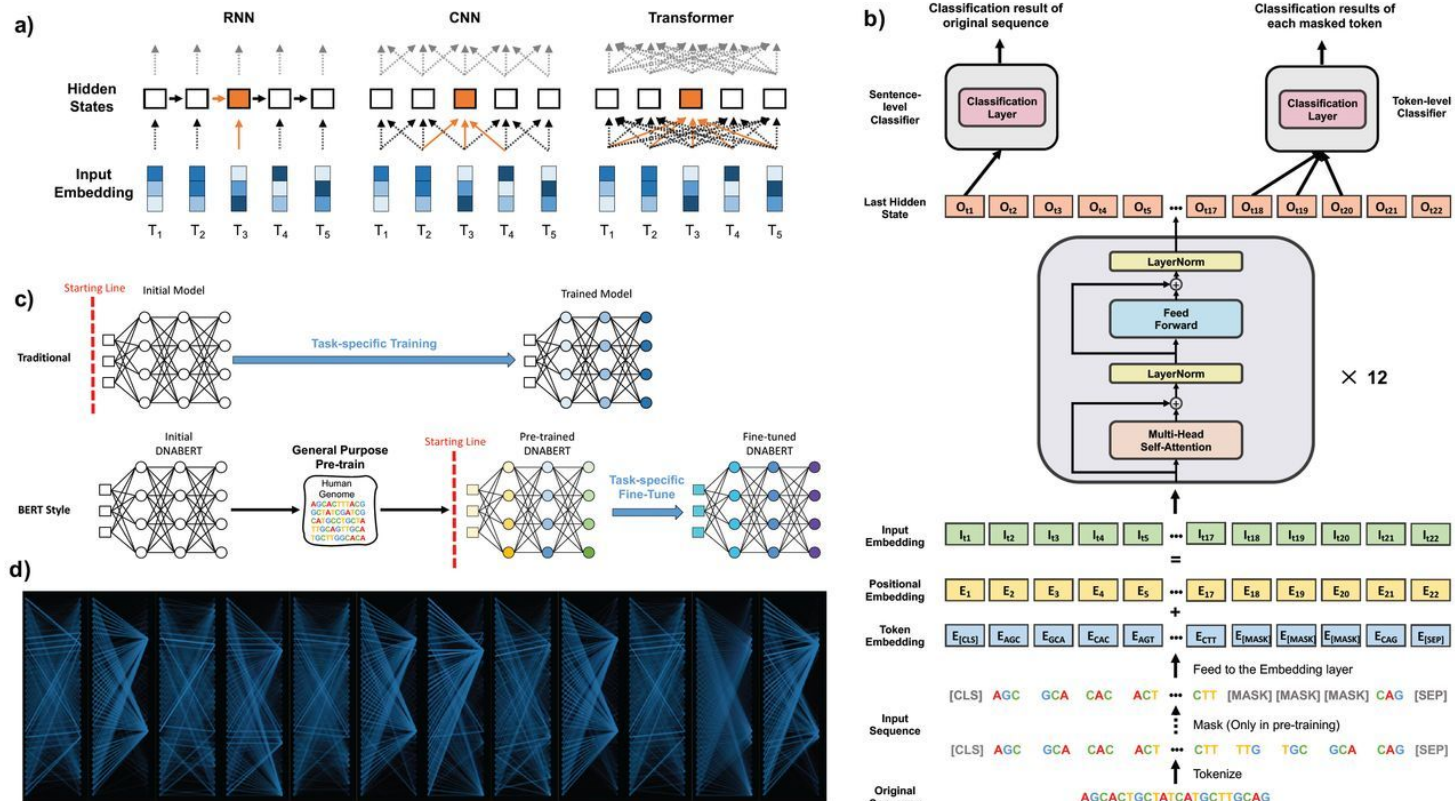


[https://www.davidsbatista.net/blog/2017/11/11/HM\\_and\\_Naive\\_Bayes/](https://www.davidsbatista.net/blog/2017/11/11/HM_and_Naive_Bayes/)

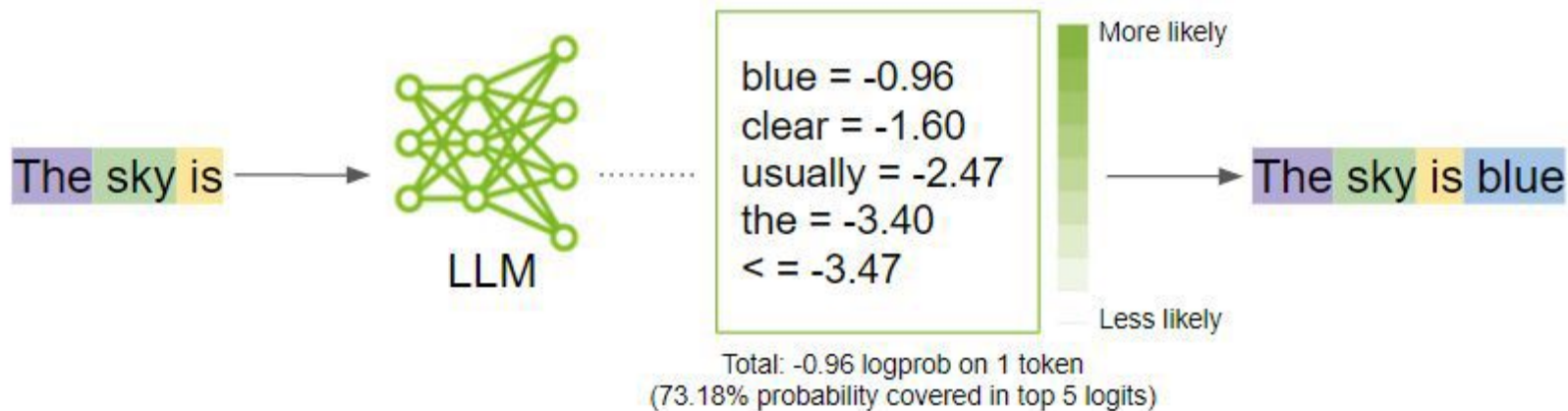


<https://www.ncbi.nlm.nih.gov/books/NBK550342/figure/Ch4-f0002/>

# DNABERT: DNA language in genome



# Next Token Prediction



## Next-token-prediction

The model is given a sequence of words with the goal of predicting the next word.

Example:  
Hannah is a \_\_\_\_\_

Hannah is a *sister*  
Hannah is a *friend*  
Hannah is a *marketer*  
Hannah is a *comedian*

## Masked-language-modeling

The model is given a sequence of words with the goal of predicting a 'masked' word in the middle.

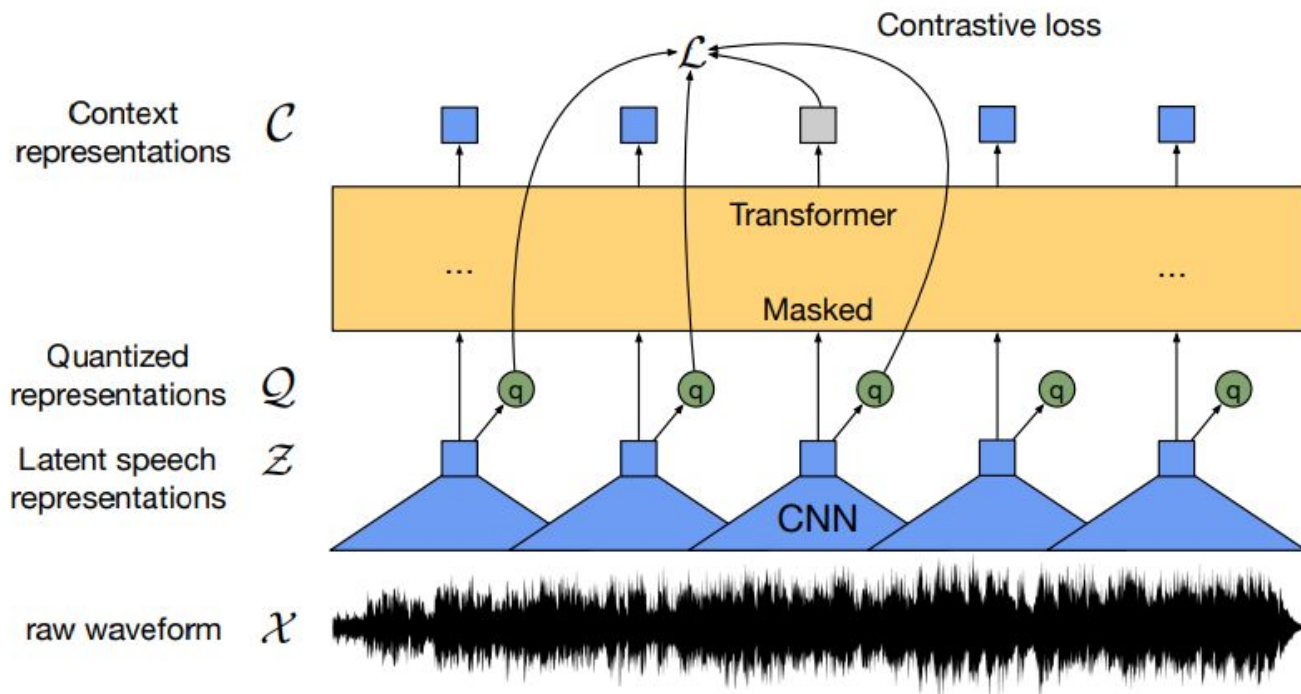
Example  
Jacob [mask] reading

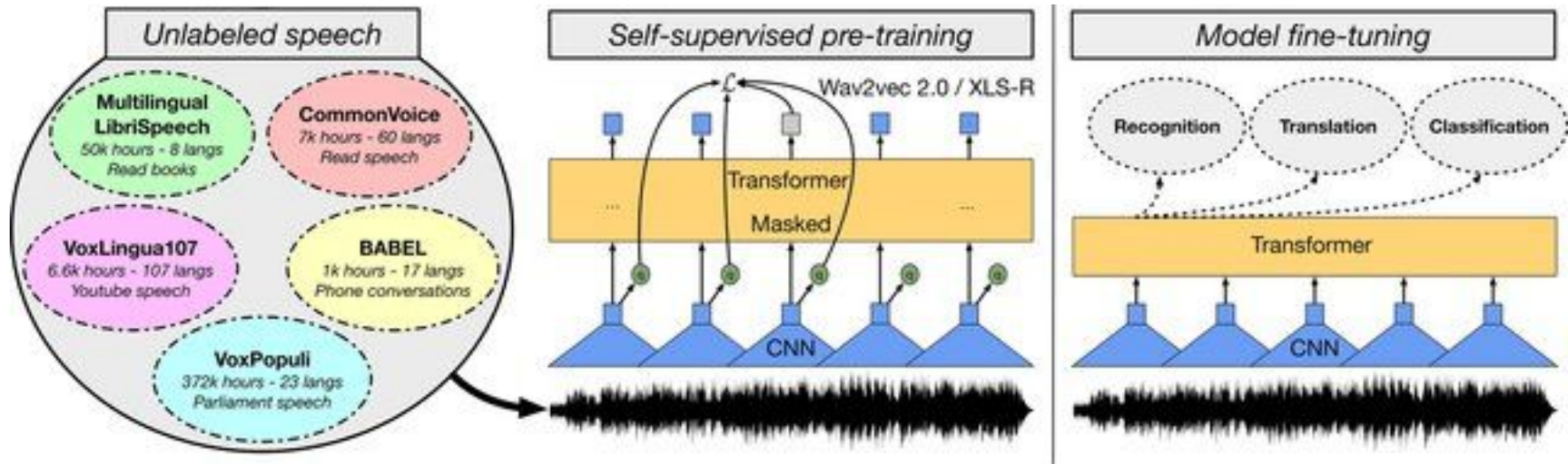
Jacob *fears* reading  
Jacob *loves* reading  
Jacob *enjoys* reading  
Jacob *hates* reading




**Alex Banks**   
@thealexbanks

# Kvantizace “jazyka”





[10.1007/s10791-024-09489-8](https://arxiv.org/abs/2106.04456)

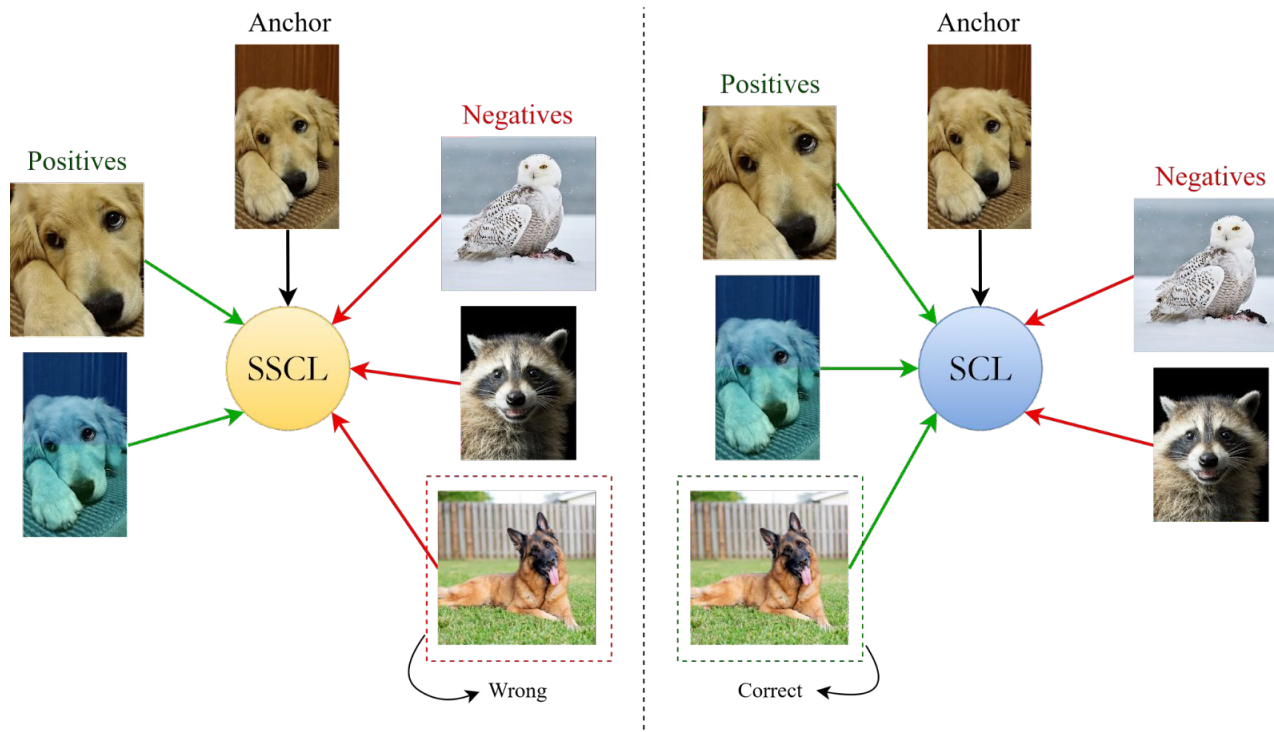
	960 hours	→	1.8 / 3.3% WER
	100 hours	→	2.0 / 4.0% WER
	10 minutes	→	4.8 / 8.2% WER

## Foundation model

<https://towardsdatascience.com/wav2vec-2-0-a-framework-for-self-supervised-learning-of-speech-representations-7d3728688cae/>

# Kontrastivní učení

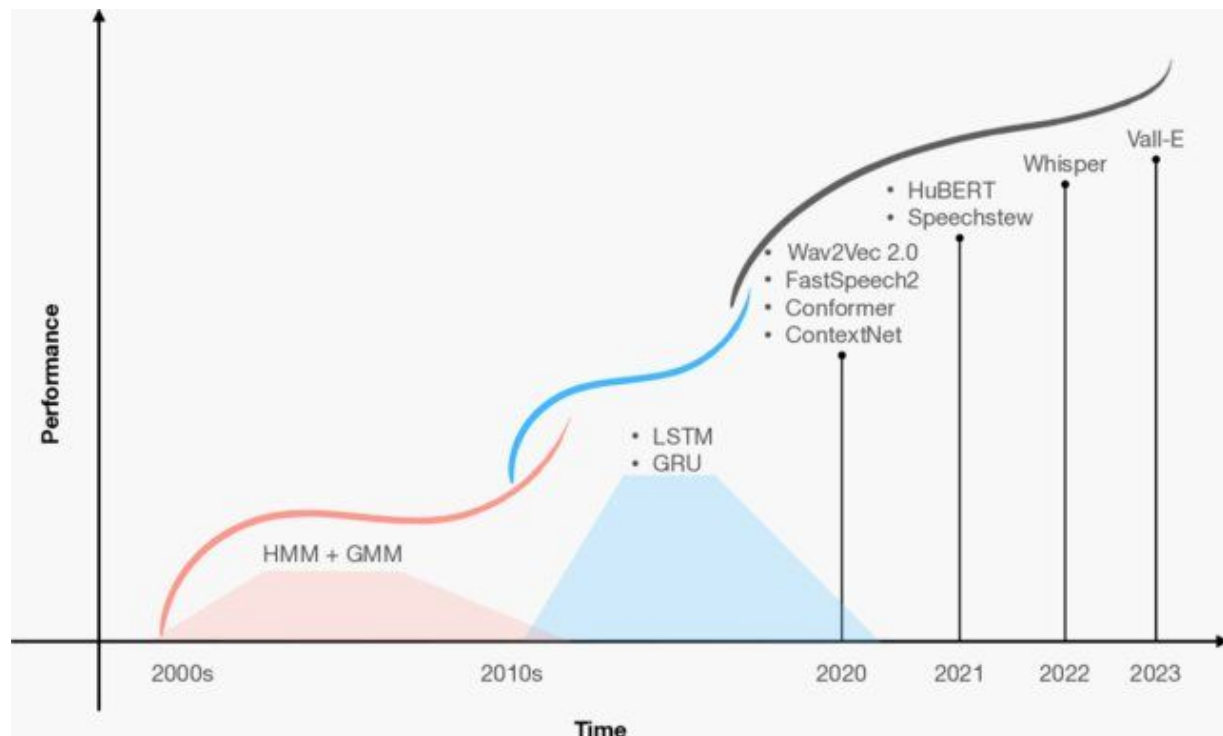
NPFL138

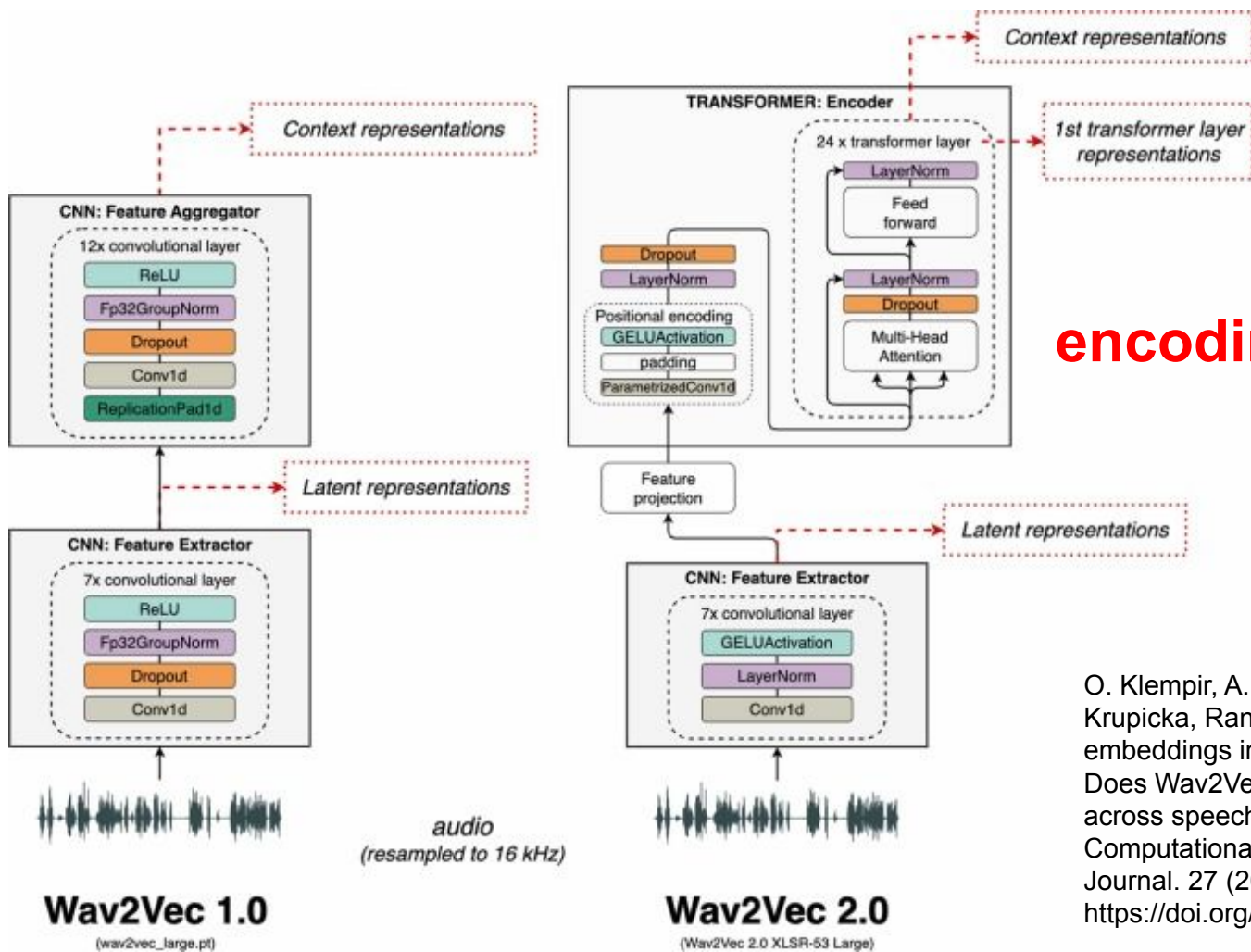


<https://www.v7labs.com/blog/contrastive-learning-guide>

# Seznam vybraných předtrénovaných modelů

- **Wav2Vec (2019)**
- **Wav2Vec 2.0 (2020)**
  - Base
  - Large
  - XLSR
  - ...
- HuBERT
- Whisper
- Data2Vec
- WavLM



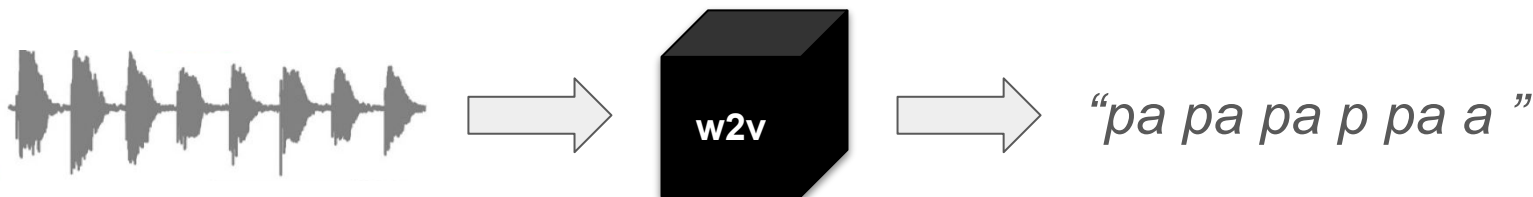


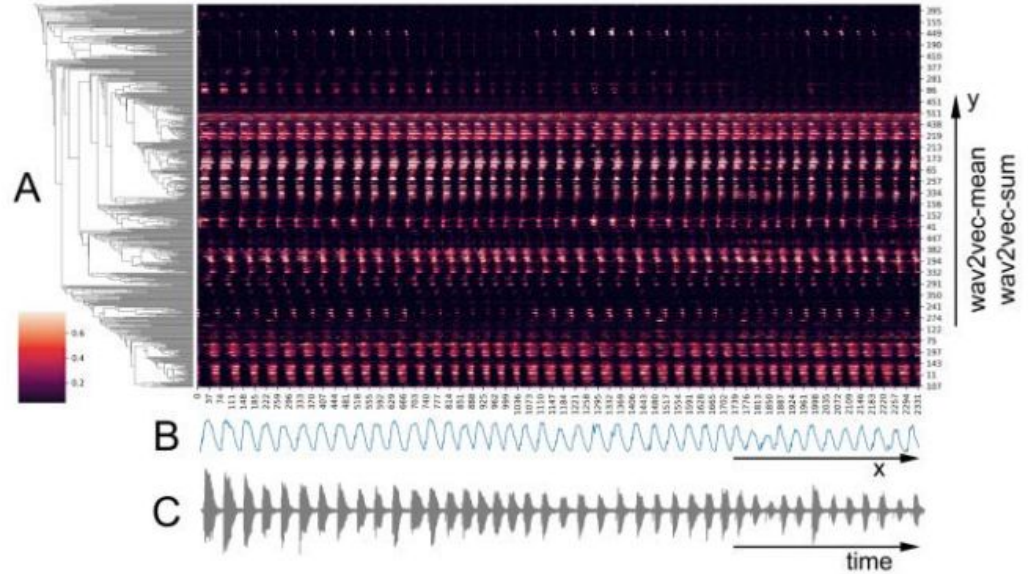
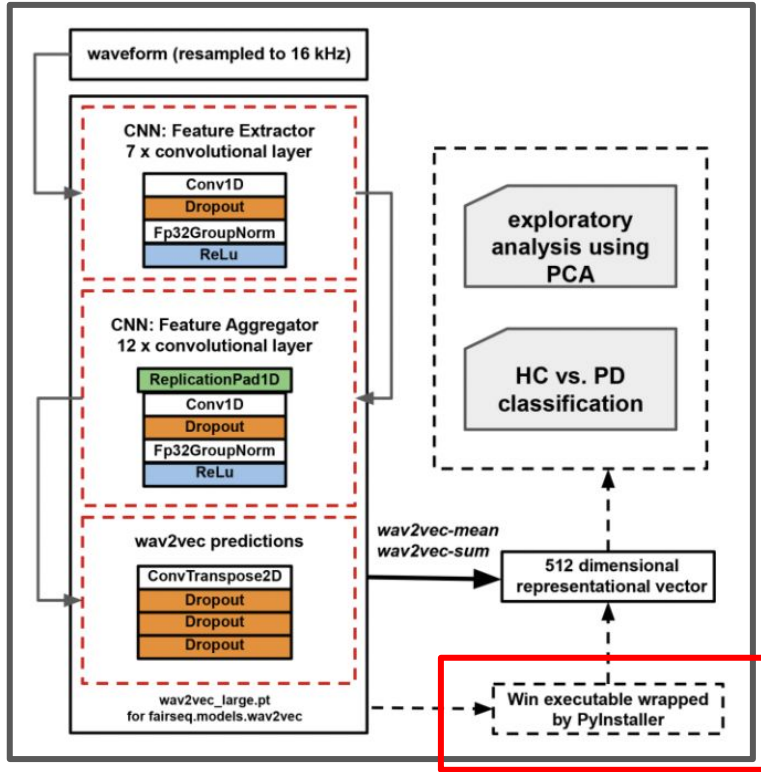
# encoding vs. decoding

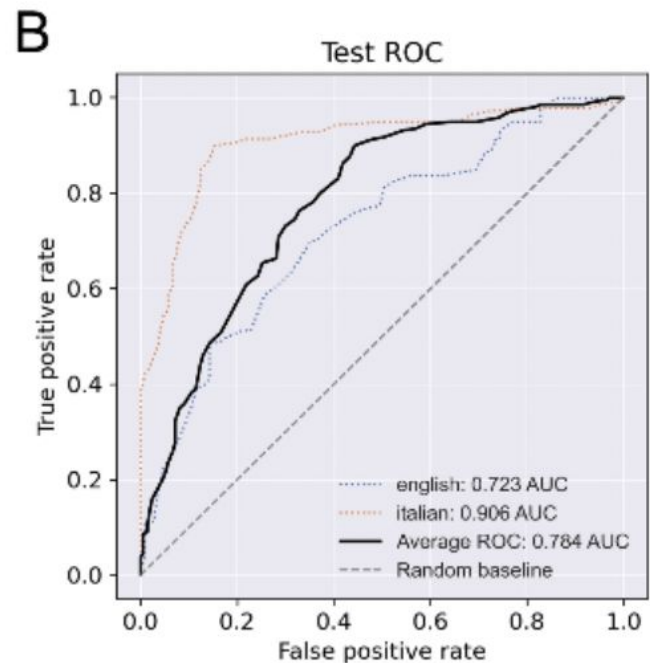
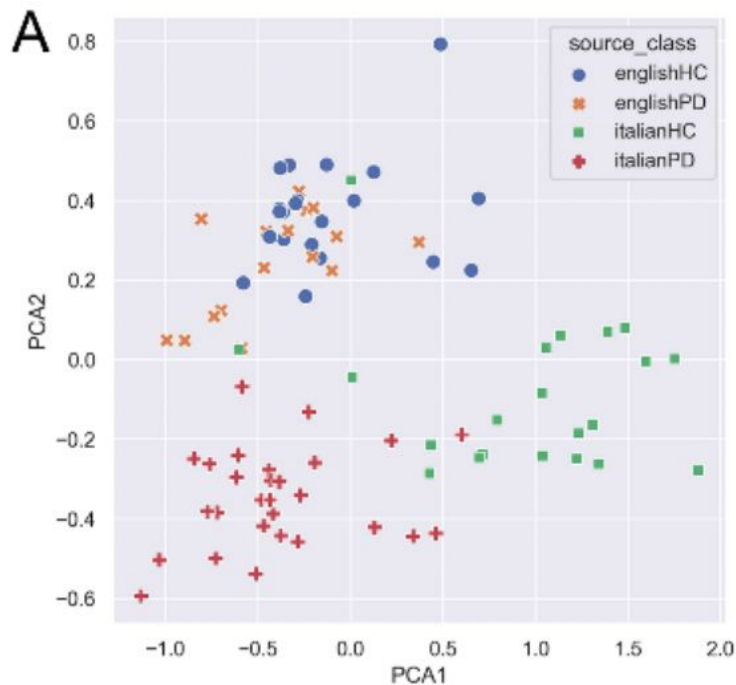
O. Klempir, A. Skryjova, A. Tichopad, R. Krupicka, Ranking pre-trained speech embeddings in Parkinson's disease detection: Does Wav2Vec 2.0 outperform its 1.0 version across speech modes and languages?, Computational And Structural Biotechnology Journal. 27 (2025) 2584-2601. <https://doi.org/10.1016/j.csbj.2025.06.022>.

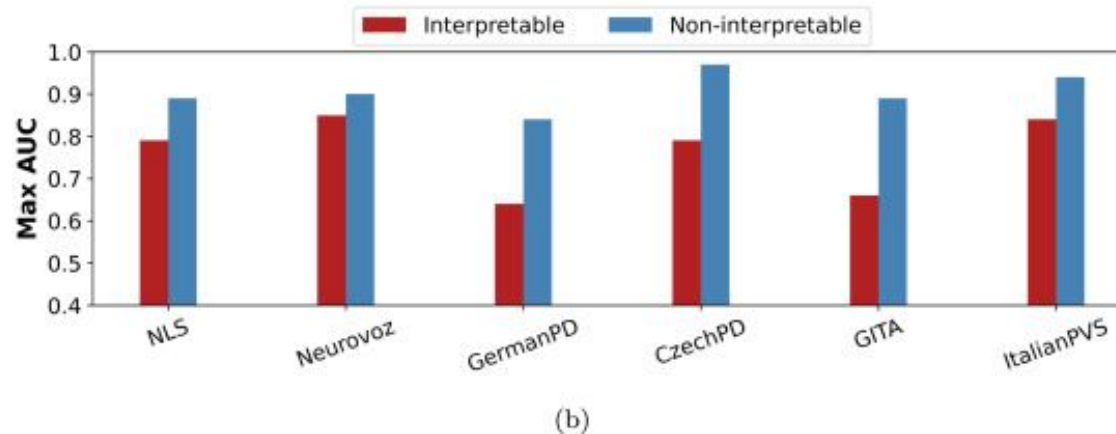
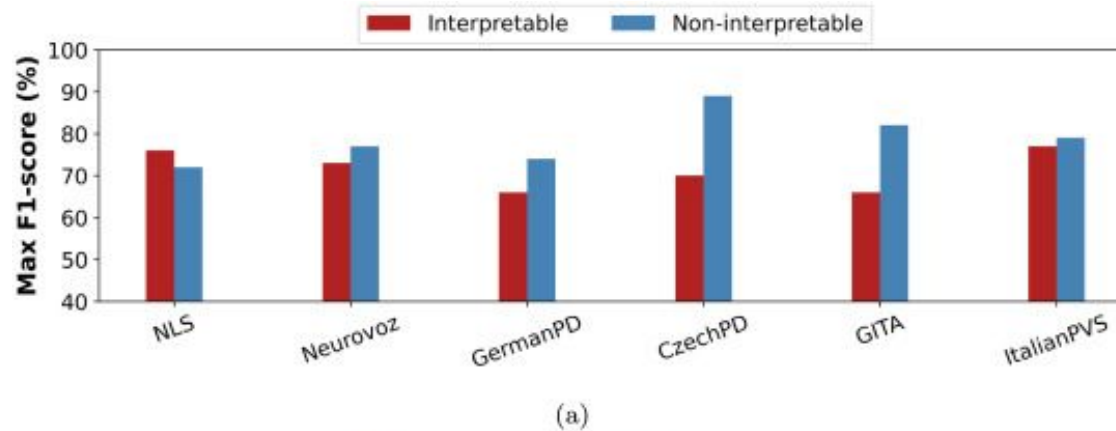
# Úlohy (na příkladech analýzy patologické řeči)

- Klasifikace (detekce nemocného)
- Regrese (odhadni tíži, věk, počet znaků, trvání pauz ...)
  - *embeddings: vektor fixní délky (typicky dimenze 512, 768, 1024)*
  - *modely jako feature generátor*
- Automatické rozpoznání a přepis audio na text (vyhodnot' srozumitelnost)





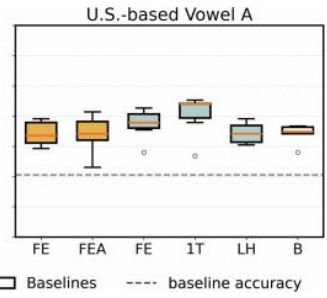
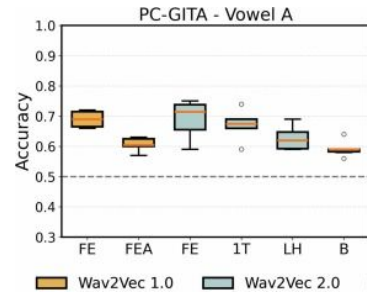
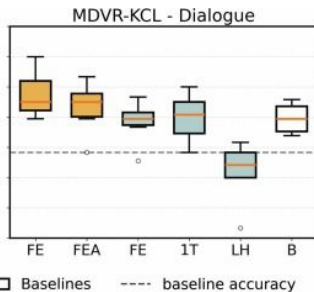
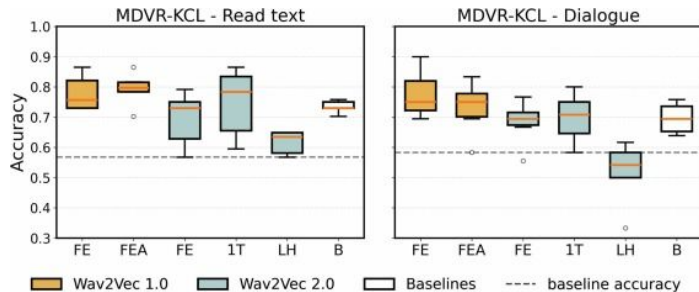
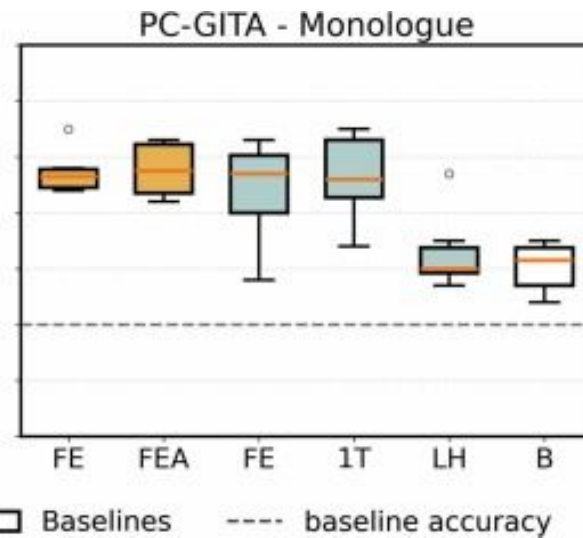
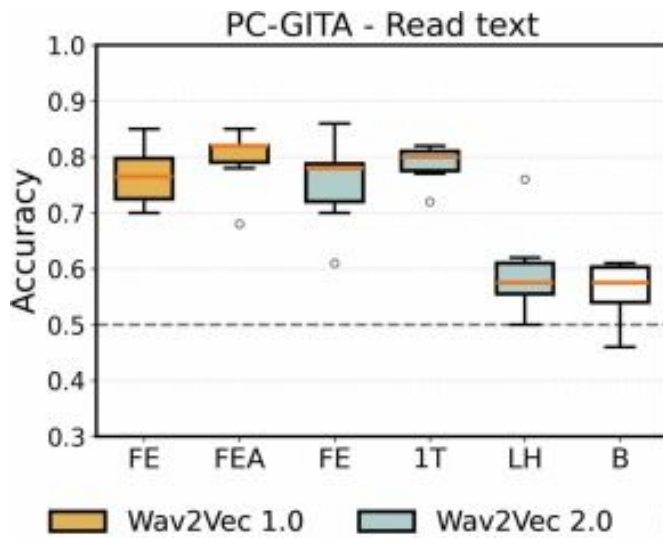




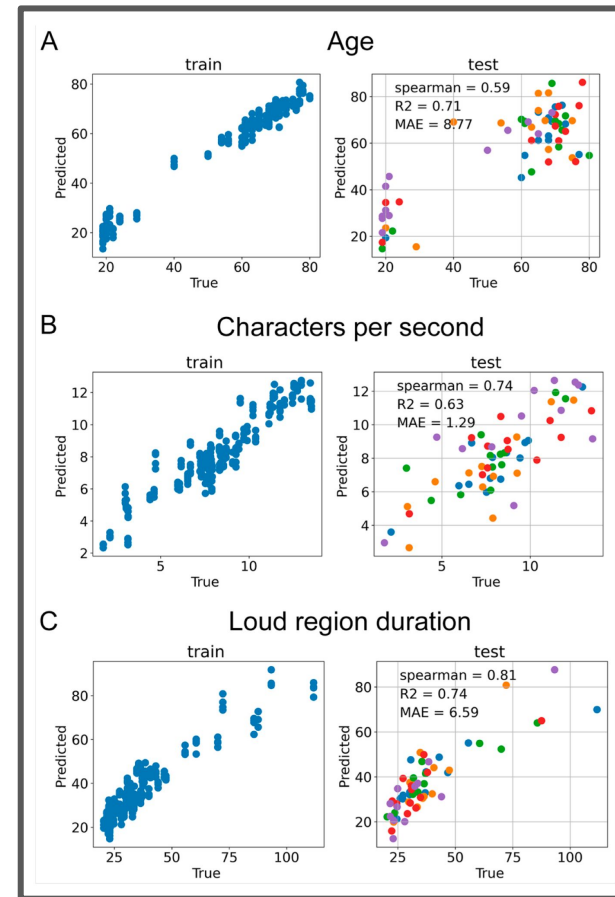
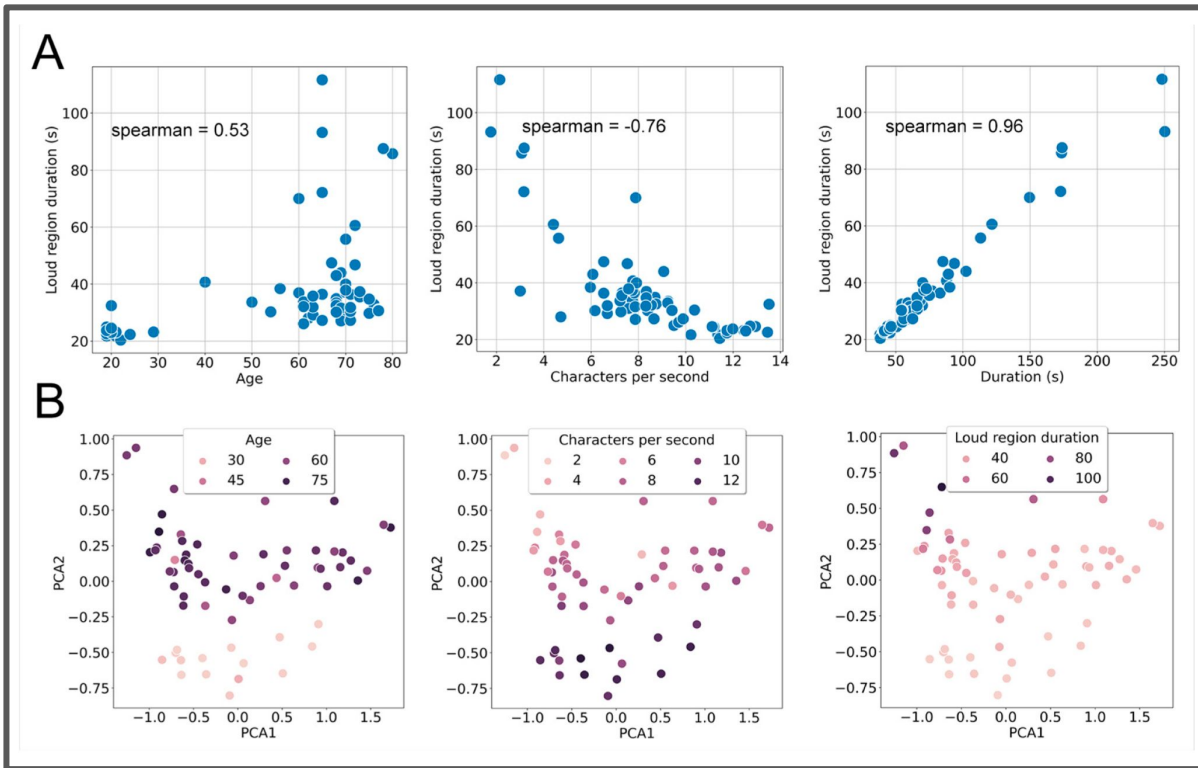
# Menší model, porovnatelný výkon?

- V1 úplně ignorována v literatuře - i mimo patologie
- Nesprávné odkazování na naši práci z 2023, snažili jsme se to fixnou
- Systematicky jsme porovnali v1 vs. v2
- Malý (menší!) netransformerový model kompetitivní a rychlejší

dataset	mode	rank (score)					
		1	2	3	4	5	6
English [24]	read text	w2v2-1T (0.88)	w2v1-FEA (0.85)	baselines (0.51)	w2v1-FE (0.41)	w2v2-FE (0.18)	w2v2-LH (0.07)
	dialogue	baselines (0.95)	w2v2-1T (0.81)	w2v1-FEA (0.80)	w2v1-FE (0.66)	w2v2-FE (0.43)	w2v2-LH (0.02)
PC-GITA [25]	read text	w2v2-1T (0.96)	w2v1-FEA (0.89)	w2v1-FE (0.81)	w2v2-FE (0.68)	baselines (0.20)	w2v2-LH (0.10)
	monologue	w2v2-1T (0.94)	w2v1-FEA (0.93)	w2v2-FE (0.87)	w2v1-FE (0.77)	w2v2-LH (0.27)	baselines (0.13)
	vowels	w2v2-FE (0.95)	w2v2-1T (0.93)	w2v1-FE (0.75)	w2v1-FEA (0.44)	w2v2-LH (0.41)	baselines (0.05)
vowels /a/ [26]	vowels	w2v2-FE (0.89)	w2v2-1T (0.78)	w2v1-FE (0.40)	baselines (0.30)	w2v1-FEA (0.30)	w2v2-LH (0.20)



O. Klempir, A. Skryjova, A. Tichopad, R. Krupicka, Ranking pre-trained speech embeddings in Parkinson's disease detection: Does Wav2Vec 2.0 outperform its 1.0 version across speech modes and languages?, Computational And Structural Biotechnology Journal. 27 (2025) 2584-2601. <https://doi.org/10.1016/j.csbj.2025.06.022>.



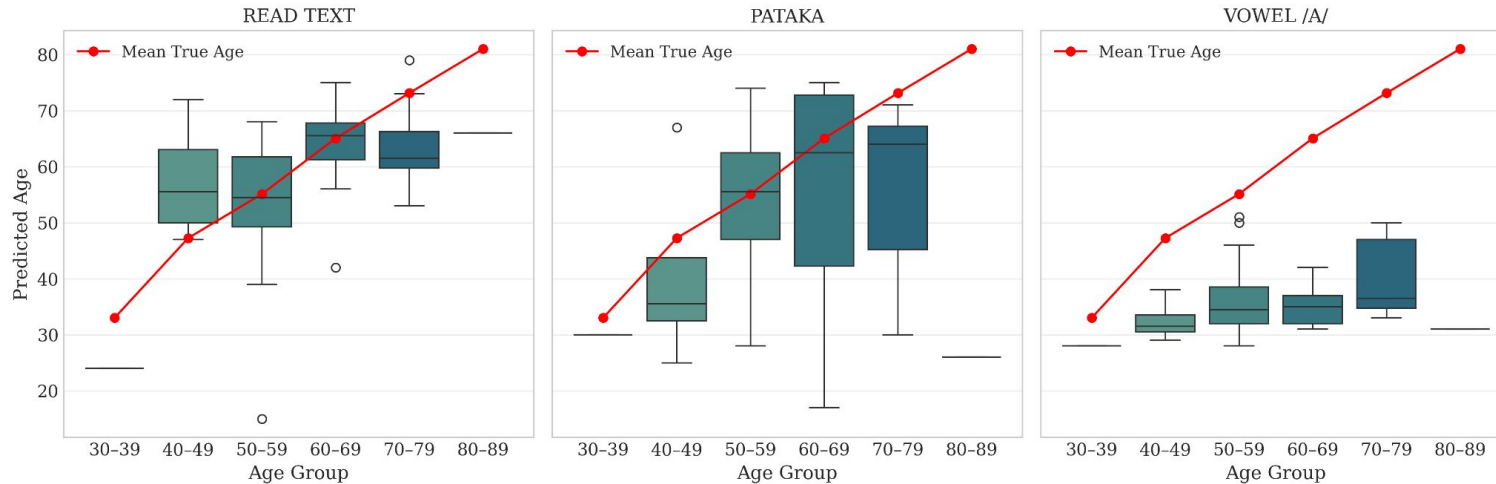
Tady se trénoval downstream regresní model.

O. Klempíř, R. Krupička, Analyzing Wav2Vec 1.0 Embeddings for Cross-Database Parkinson's Disease Detection and Speech Features Extraction, Sensors. 24 (2024) 5520. <https://doi.org/10.3390/s24175520>.

Group	Task	Accuracy for Gender Estimation	Correlation with True Age (p-value)	Chi <sup>2</sup> Goodness of Fit (p-value)	MAD* years
HC (N = 50)	read text	1.00	0.52 ( <b>p &lt; 0.001</b> )	14.12 ( <b>p = 0.003</b> )	5.0
	pataka	0.96	0.36 ( <b>p = 0.01</b> )	16.60 ( <b>p &lt; 0.001</b> )	9.0
	/a/ vowel	0.94	0.05 (p = 0.72)	575.00 ( <b>p &lt; 0.001</b> )	27.0
PD (N = 50)	read text	0.98	0.44 ( <b>p = 0.001</b> )	5.58 (p = 0.13)	6.0
	pataka	0.94	0.20 (p = 0.17)	75.63 ( <b>p &lt; 0.001</b> )	9.5
	/a/ vowel	0.98	0.23 (p = 0.10)	392.30 ( <b>p &lt; 0.001</b> )	25.0

\*MAD (Median Absolute Deviation)

1. K čemu je to dobré?
2. Jaké jsou problémy?



# Interpretace u Wav2Vec

wav2vec 2.0 jako nástroj přepisu řeči na text

$$CER = \frac{S + D + I}{N}$$

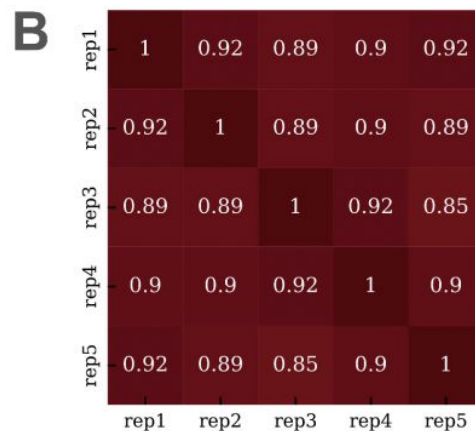
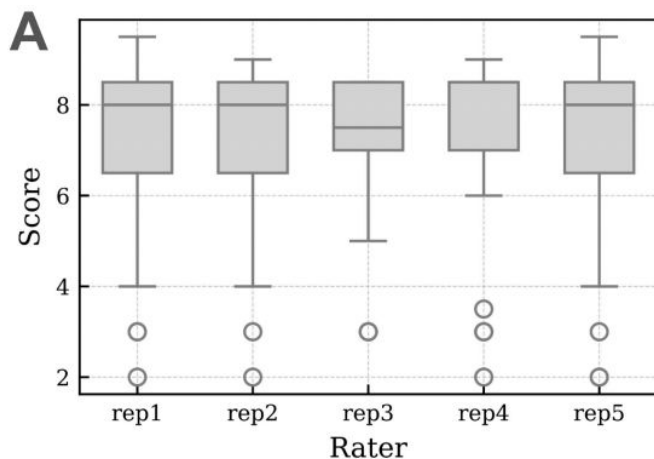
	<i>YHC</i>	<i>EHC</i>	<i>PD</i>
dado	9 (69%)	9 (41%)	14 (58%)
luna	8 (62%)	7 (32%)	9 (38%)
scala	4 (31%)	6 (27%)	9 (38%)
casa	5 (39%)	4 (18%)	3 (13%)
frate	2 (15%)	3 (14%)	6 (25%)
brodo	8 (62%)	1 (5%)	7 (29%)
monotono	6 (46%)	1 (5%)	6 (25%)
montagnoso	6 (46%)	2 (9%)	1 (4%)

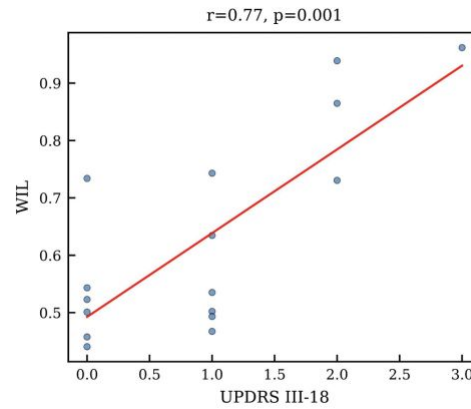
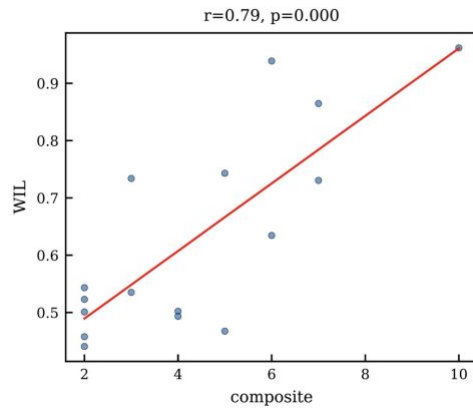
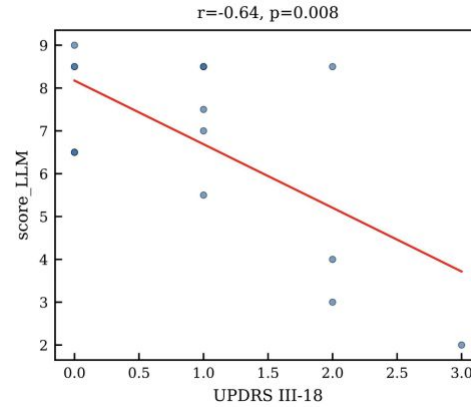
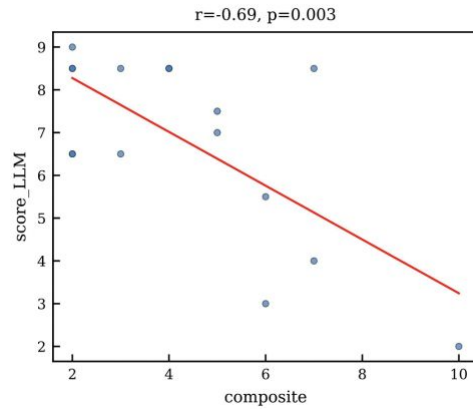
<i>Z-score Classification</i>	<i>CER</i>	<i>UPDRS</i>	<i>Top 5 most unintelligible words</i>
Outlier (> 3)	1.29	4	gatto, buco, iuta, flotta, scia
Outlier (> 3)	1.18	2	plagio, brodo, monotono, pozzo, frate
Moderately Unusual (> 2)	1.07	3	iuta, pozzo, scia, toppo, rete
-	0.95	4	pericoloso, glielo, giro, spruzzo, muro
-	0.91	0	sdraio, brodo, prova, slitta, drago

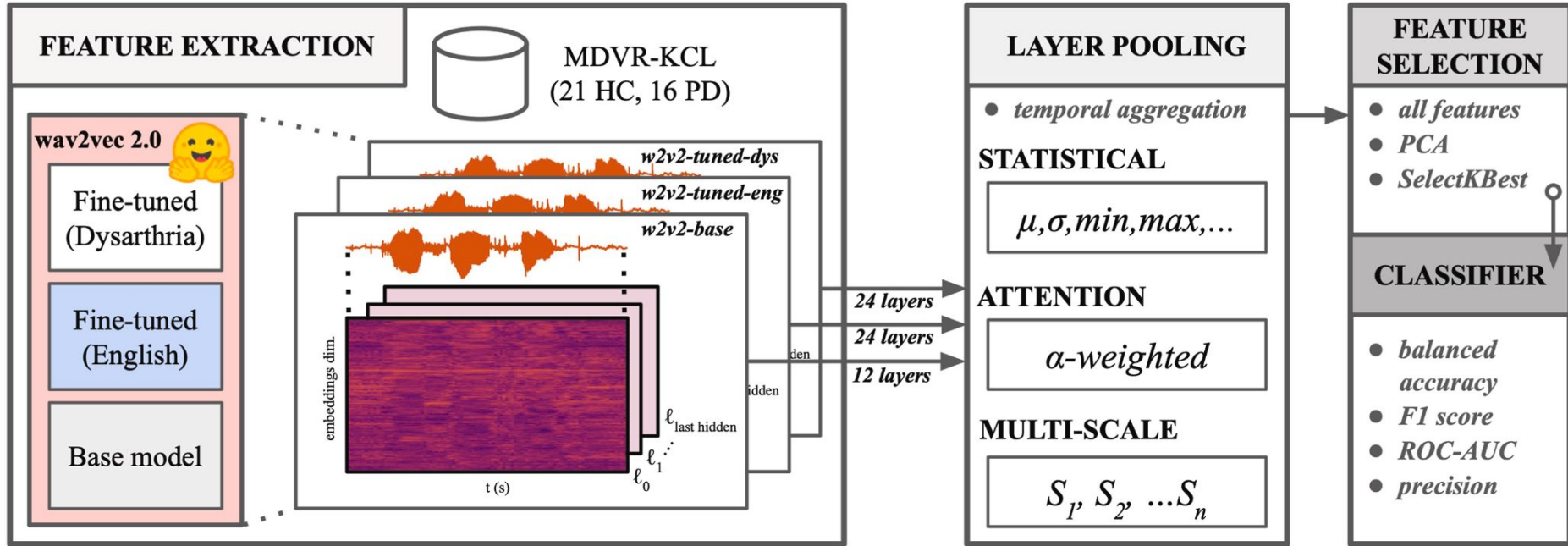
T. Smolik, R. Krupicka, O. Klempir, Assessing Speech Intelligibility and Severity Level in Parkinson's Disease Using Wav2Vec 2.0, in: 2024 47Th International Conference On Telecommunications And Signal Processing (Tsp), IEEE, 2024: pp. 231-234. <https://doi.org/10.1109/tsp63128.2024.10605915>.

# LLM pro návrh skóre srozumitelnosti u Parkinsonovy nemoci

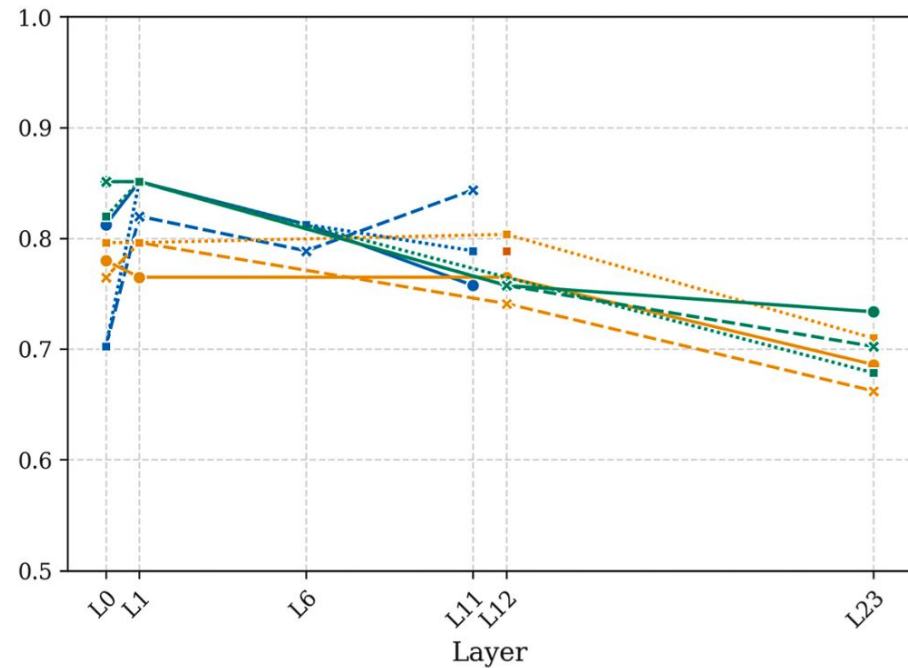
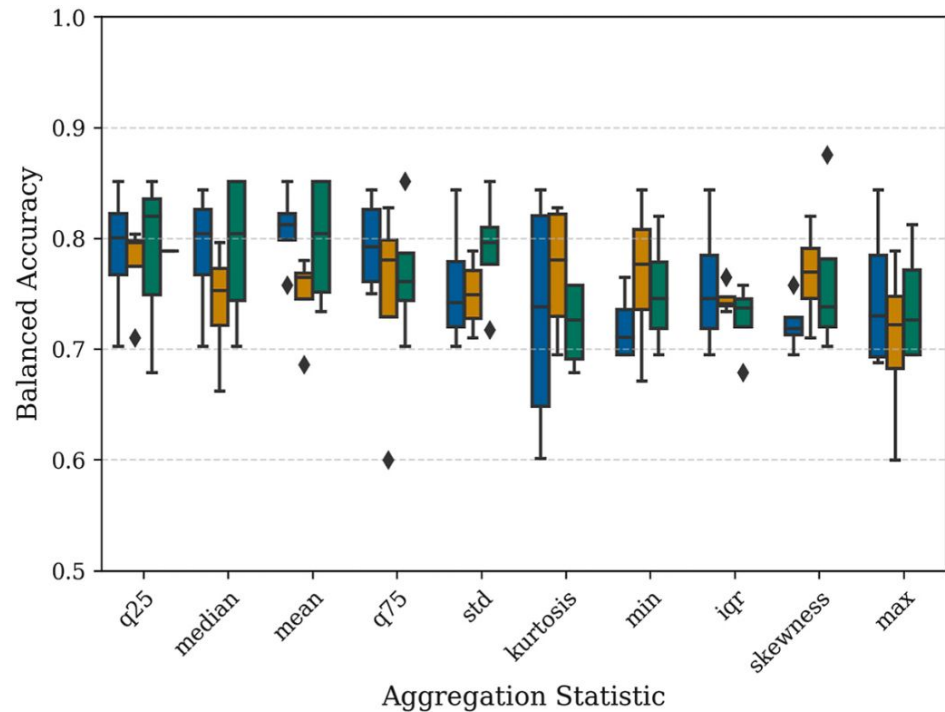
- wav2vec 2.0 jako nástroj přepisu řeči na text
- vůči referenci
- Github typos (BERT - GitHub Typo Corpus)
- LLM - Mistral - lokálně







O. Klempir, J.G. Mullerova, R. Krupicka, Statistical, multi-scale and attention-based layer pooling of Wav2Vec-2 speech embeddings for Parkinson's disease detection, Computers In Biology And Medicine. 200 (2026) 111368. <https://doi.org/10.1016/j.combiomed.2025.111368>.



Model: w2v2-base (blue), w2v2-tuned-dysarthria (orange), w2v2-tuned-eng (green). Aggregation (TOP3): mean (solid line with circles), median (dashed line with stars), q25 (dotted line with squares).

O. Klempir, J.G. Mullerova, R. Krupicka, Statistical, multi-scale and attention-based layer pooling of Wav2Vec-2 speech embeddings for Parkinson's disease detection, Computers In Biology And Medicine. 200 (2026) 111368. <https://doi.org/10.1016/j.compbimed.2025.111368>.

Myslím, že bychom foundation modely neměli ignorovat...



Díky všem!



Přeji úspěšný rok 2026!

## **Mrazivé proroctví Baby Vangy pro rok 2026: Zašli jsme příliš daleko**

Věštbu Baby Vangy pro rok 2026 připomněl portál [Sky History](#). Je z ní, že nás čeká náročný rok. Pokud se totiž proroctví známé bulharské věstkyně vyplní, bude se tento rok nést v duchu mezinárodních konfliktů i zlomového období v oblasti umělé inteligence.