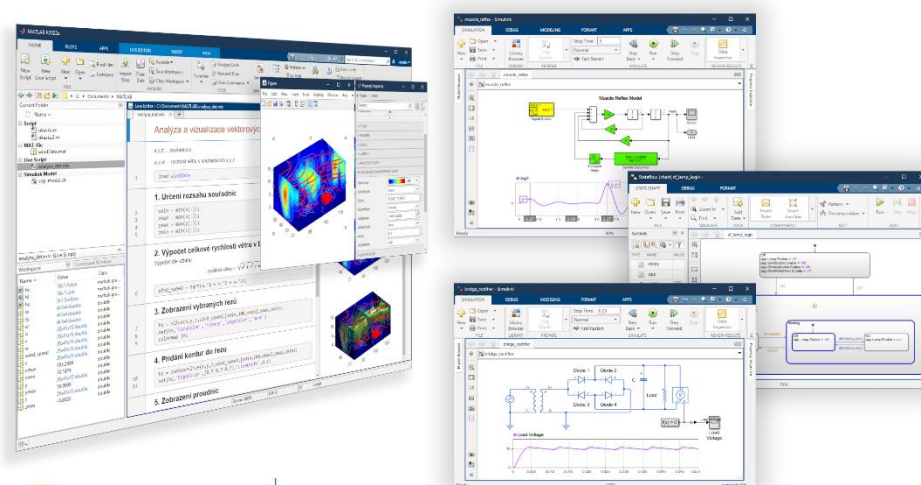


Data centric AI

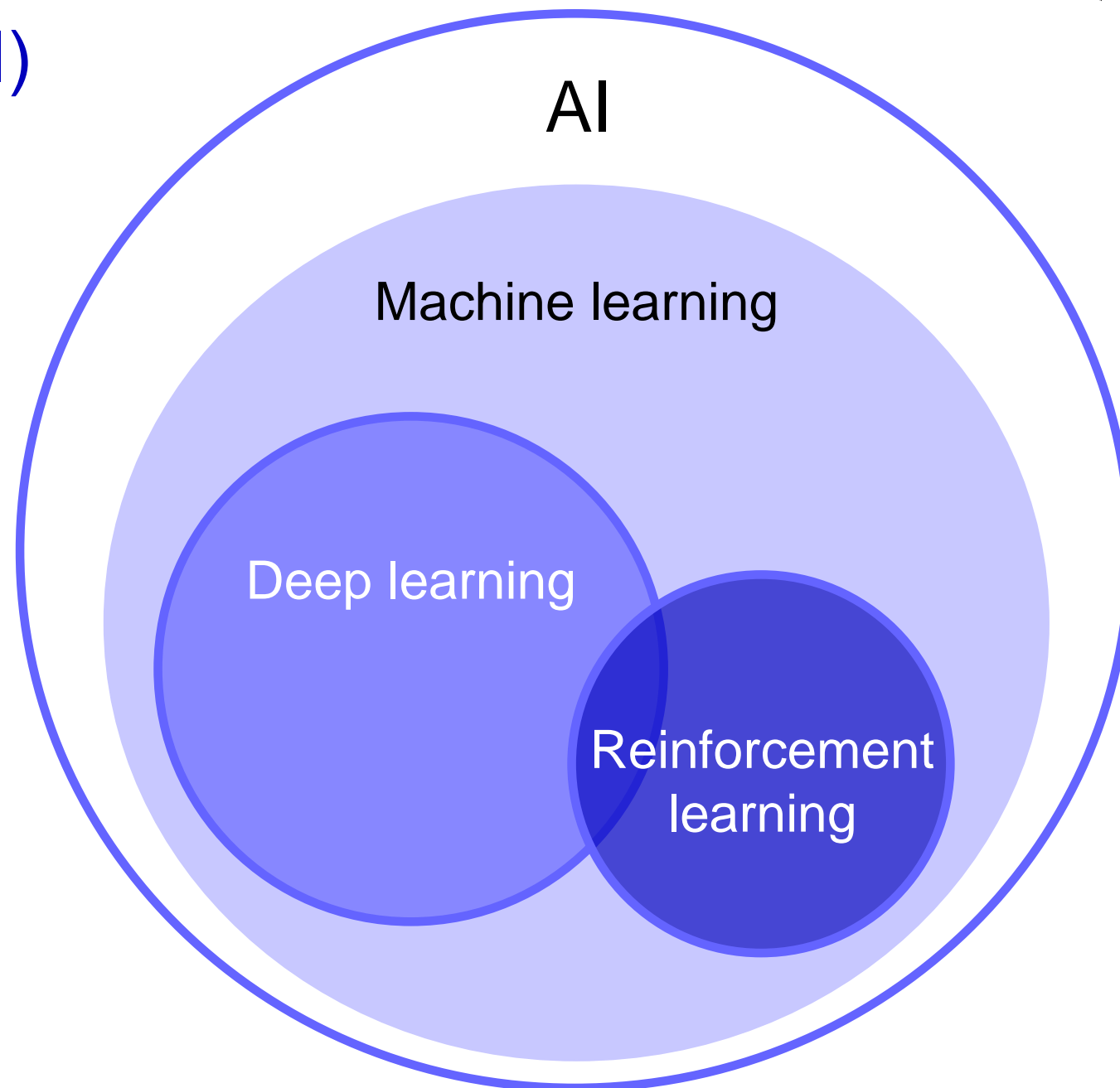


Jaroslav Jirkovský
jirkovsky@humusoft.cz

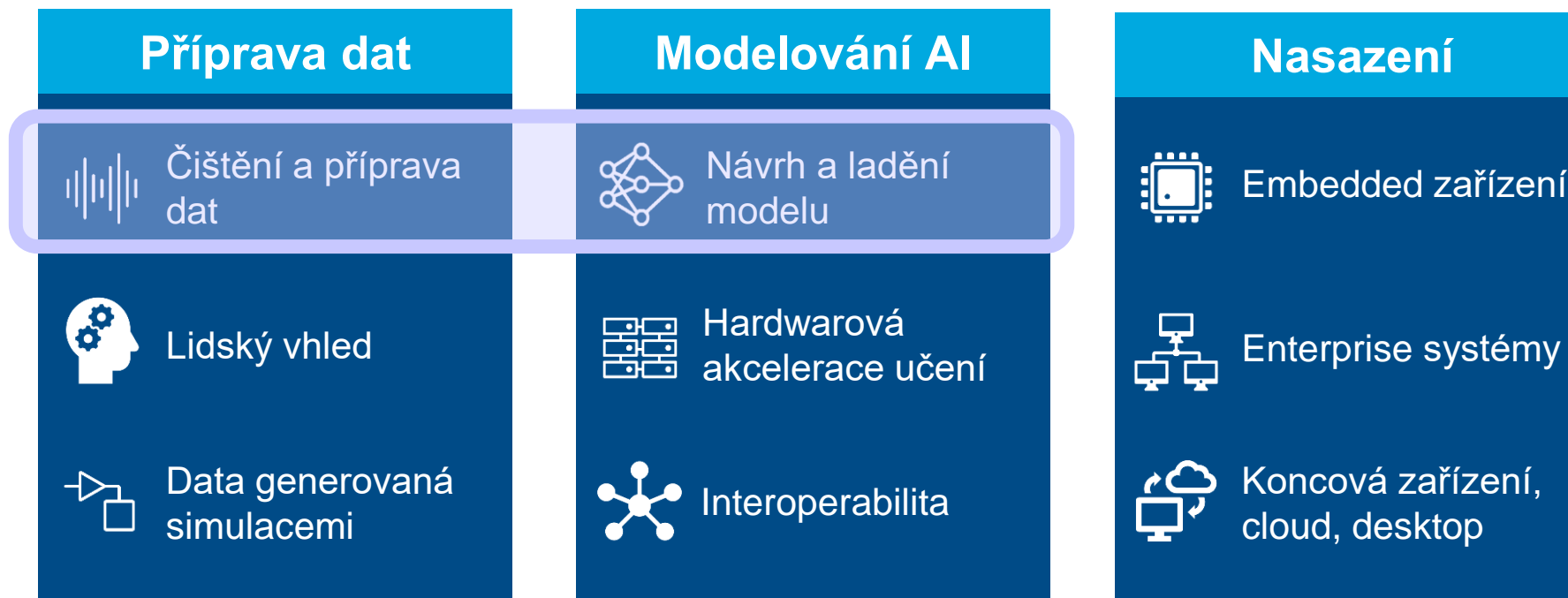
www.humusoft.cz
info@humusoft.cz

www.mathworks.com

Umělá inteligence (AI)



Návrh systémů s využitím AI



Odlišný rozsah ve výzkumných a průmyslových aplikacích

Množství práce věnované AI modelům a datům

Lepší modely nebo Lepší data?



Model-Centric vs. Data-Centric AI

Modelově-orientovaný vývoj AI

- zaměřeno na vývoj AI modelu
- zejména deep learning
 - struktura neuronové sítě
 - hyperparametry jednotlivých vrstev
- data v surové podobě (obraz)
- upravená data (signál)
 - základní úpravy a čištění
 - jednorázová transformace

Datově-orientovaný vývoj AI

- využití známých ověřených modelů
 - machine learning
 - předučené sítě pro deep learning
- zaměřeno na přípravu dat pro učení
 - extrakce, analýza a výběr prediktorů
 - transformace dat různými metodami
 - techniky založené na waveletech
- feature engineering

Nejvíce existujících zdrojů v oblasti AI pro několik málo aplikací



autonomní řízení



modelování jazyka



počítačové vidění



rozpoznávání řeči

Většina aplikací (pro zpracování signálu) nemůže počítat s mnoha zdroji



analýza vibrací



seismické analýzy



prediktivní údržba



digitální zdravotnictví



sledování stavu
technických zařízení

...



Tři praktické přístupy v datově orientovaném vývoji AI systému

- Extrakce prediktorů (features) pro jednodušší a menší AI modely



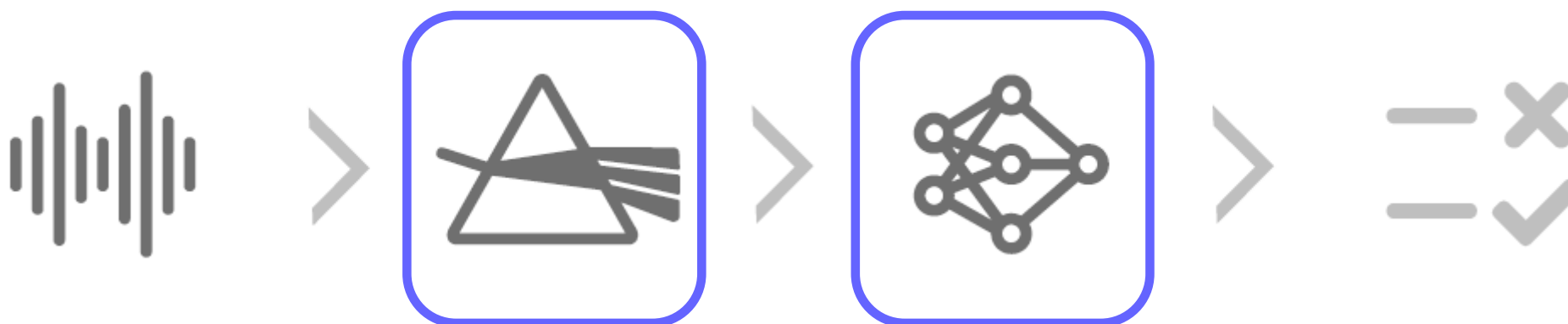
- Transfer learning s předučenými modely



- Lepší signálová data, reálná či simulovaná

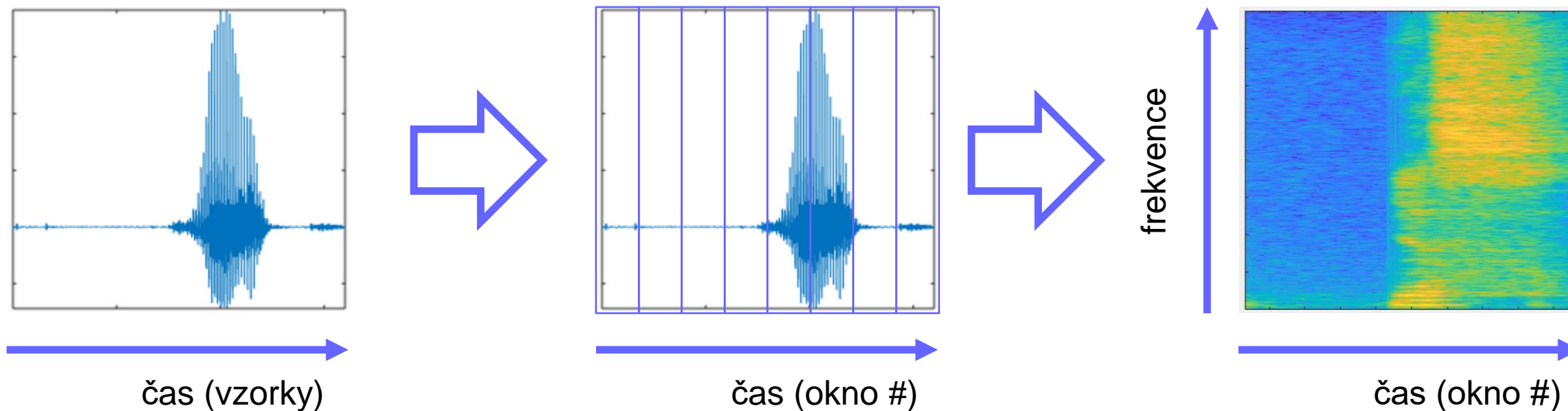


Neuronové sítě se nejčastěji neučí přímo ze surových signálů



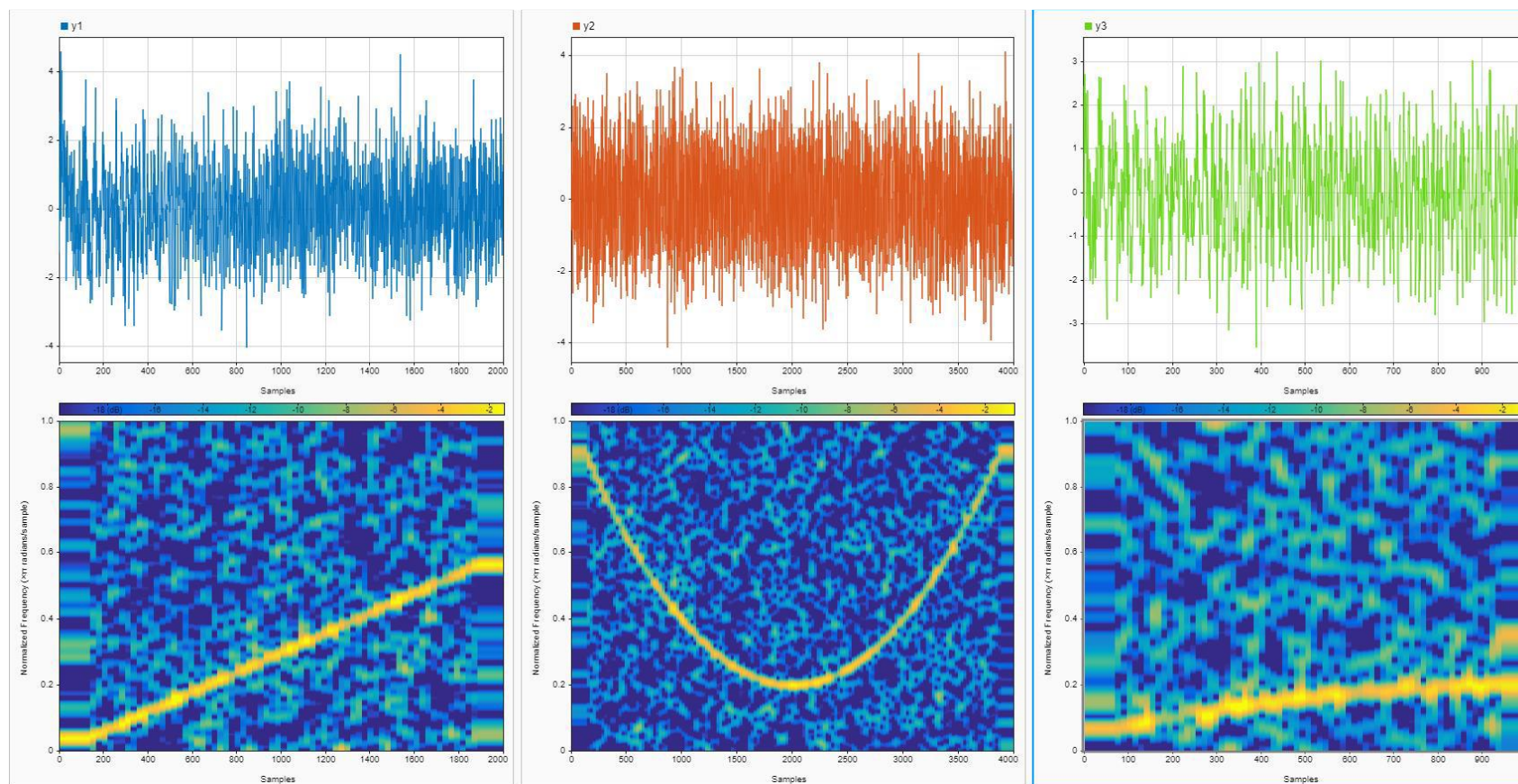
Časově-frekvenční transformace

- Populární metoda extrakce prediktorů

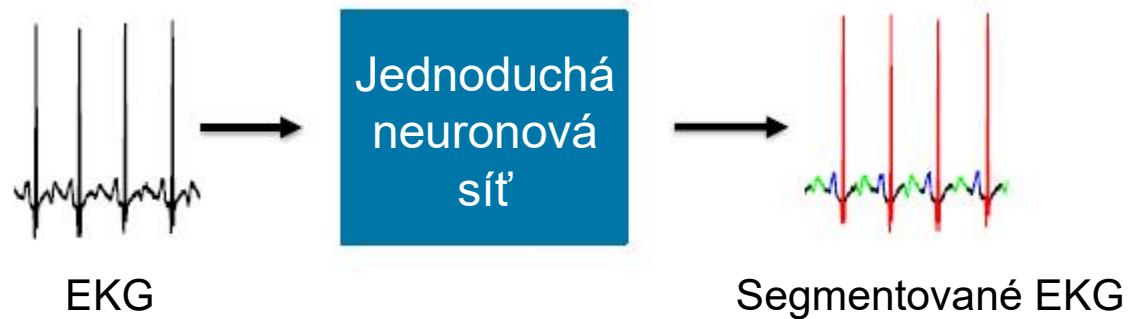


Časově-frekvenční transformace

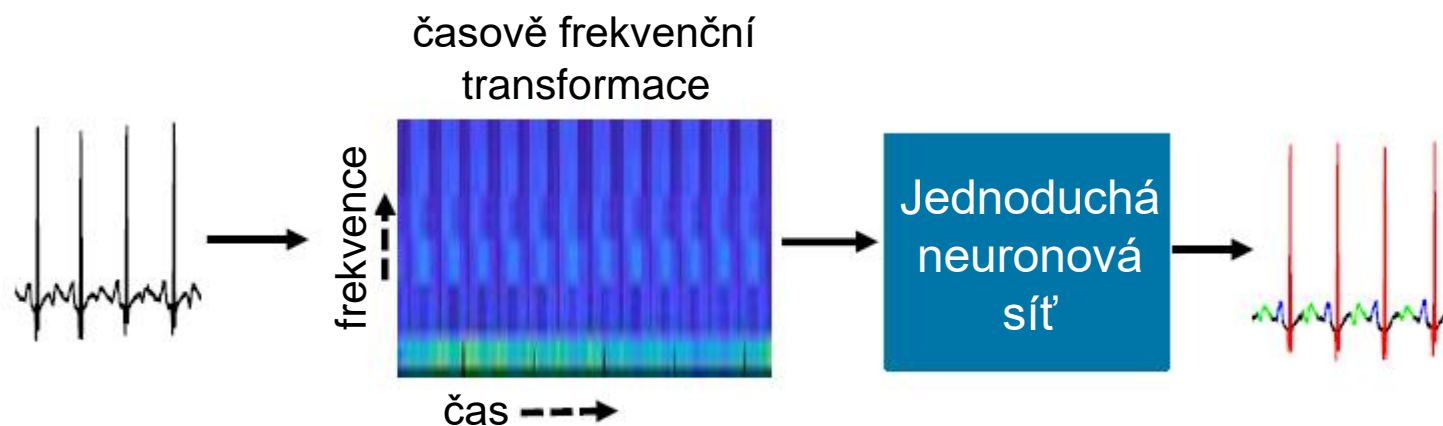
- Zvýrazní charakteristiky signálu



Extrakce prediktorů umožní zvýšit přesnost jednoduchých modelů

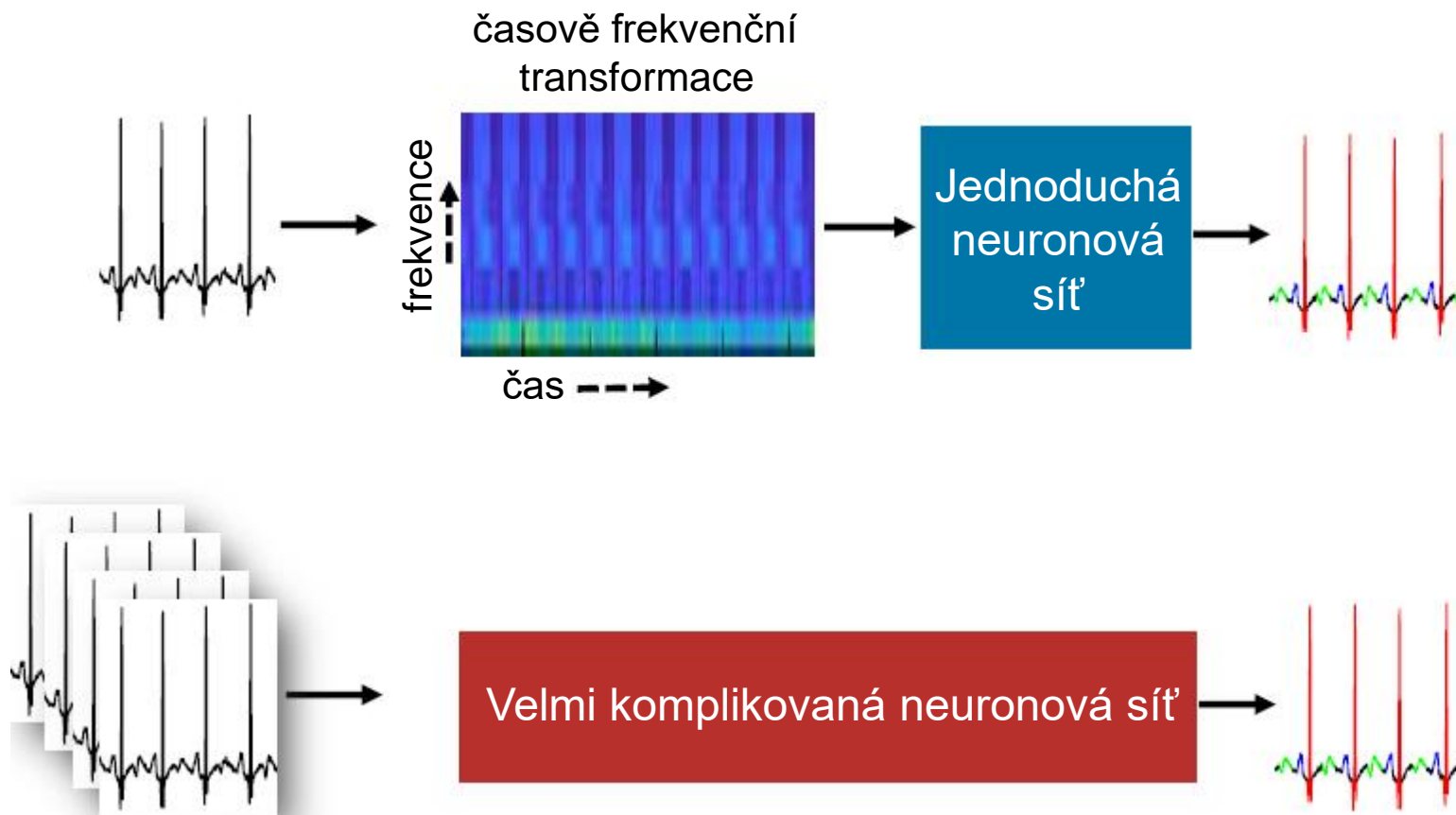


True Class \ Predicted Class	P	QRS	T	n/a
P	37.4%	2.3%	1.1%	2.1%
QRS	4.1%	61.4%	0.6%	4.3%
T	2.5%	1.4%	58.7%	7.3%
n/a	56.0%	34.8%	39.6%	86.2%



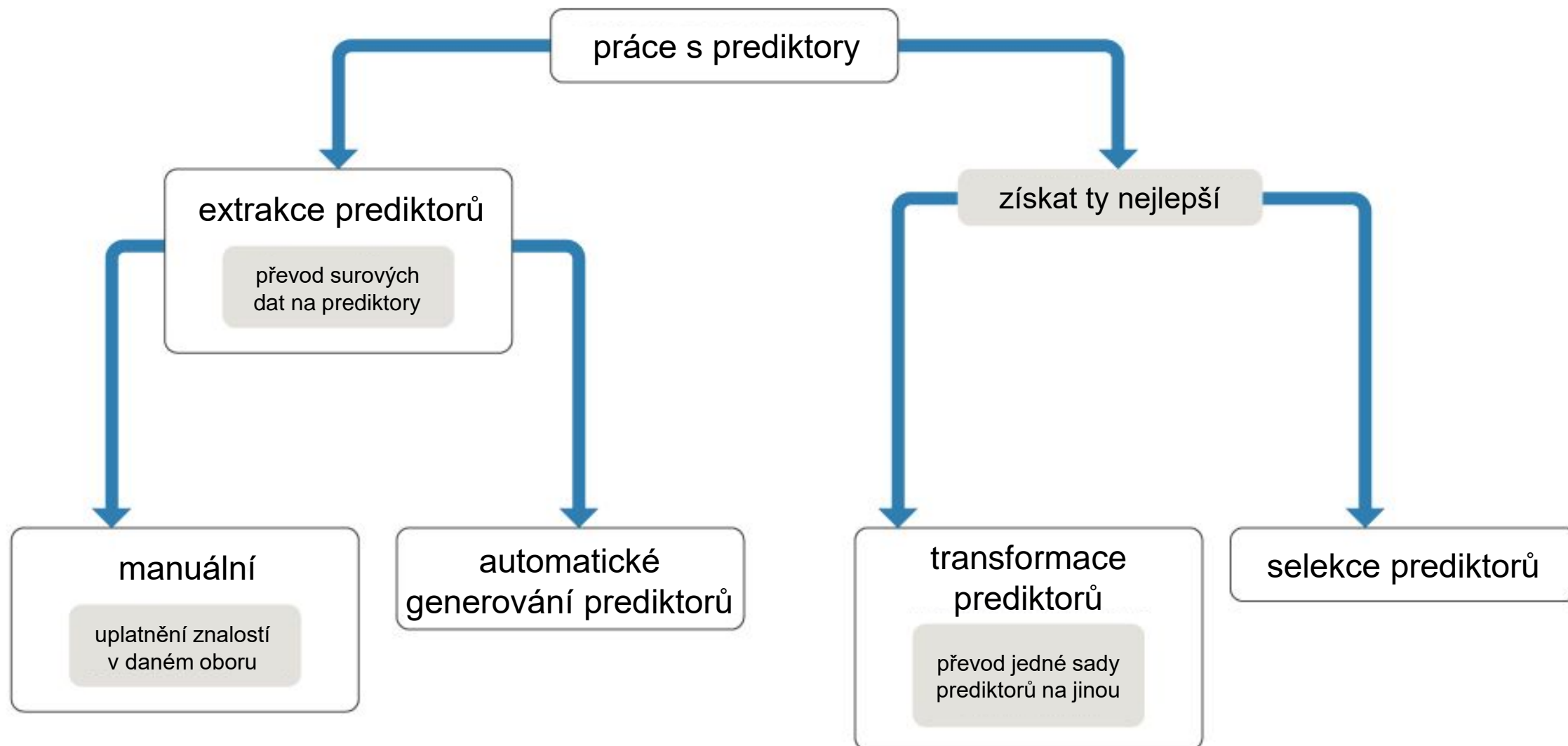
True Class \ Predicted Class	P	QRS	T	n/a
P	80.5%	0.4%	0.3%	3.2%
QRS	0.7%	90.7%	0.3%	2.1%
T	1.0%	0.3%	82.2%	7.7%
n/a	17.8%	8.7%	17.2%	87.1%

Extrakce prediktorů redukuje komplexní data a modely



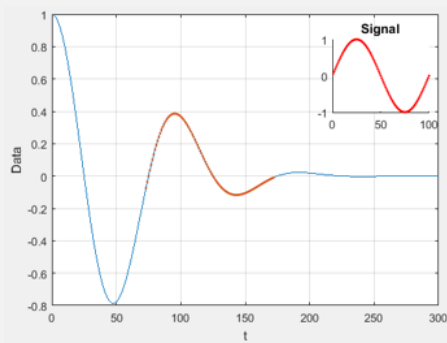
	P	QRS	T	n/a
P	80.5%	0.4%	0.3%	3.2%
QRS	0.7%	90.7%	0.3%	2.1%
T	1.0%	0.3%	82.2%	7.7%
n/a	17.8%	8.7%	17.2%	87.1%
	Predicted Class			

Získání prediktorů, které AI model potřebuje



Extrakce prediktorů (features) ze signálů

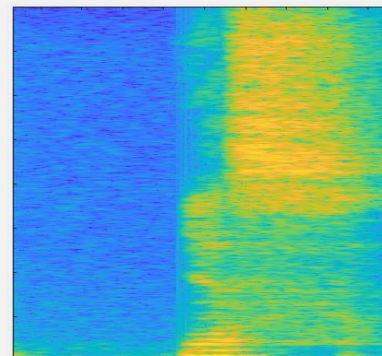
V časové oblasti



- vzory
- body zvratu
- vrcholky
- obálka

.....

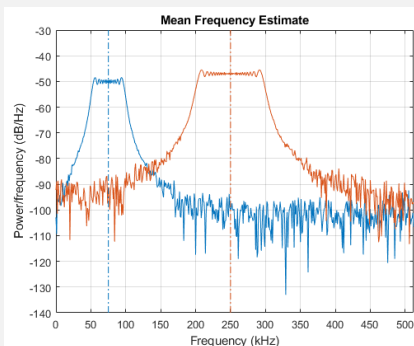
V časově-frekvenční oblasti



- spektrogram
- MFCC
- constant-Q
- scalogram

.....

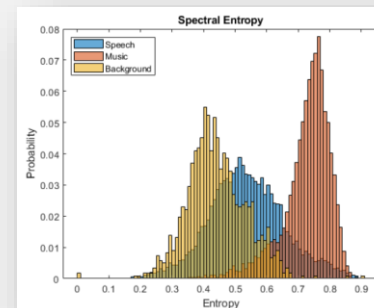
Ve frekvenční oblasti



- měření šířky pásma
- spektrální statistiky

.....

Specifické příznaky



- řeč a zvuk
- navigace a senzorická fúze
- radar
- komunikace

.....

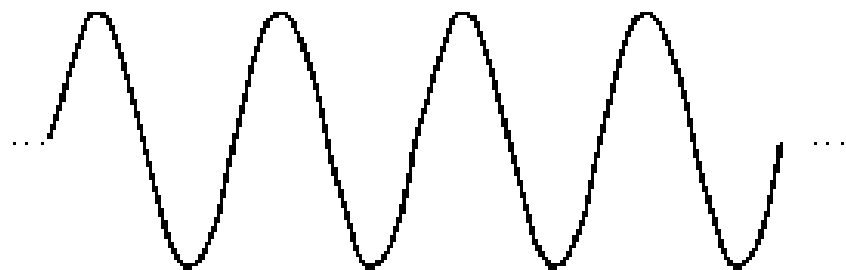
Extrakce prediktorů ze signálů – jak začít?

- Grafická aplikace
 - Diagnostic Feature Designer app
- Souhrnné funkce pro extrakci prediktorů
 - signalTimeFeatureExtractor
 - signalFrequencyFeatureExtractor
 - audioFeatureExtractor
- Přehledy funkcí v dokumentaci
 - Time-Frequency Gallery
 - <https://www.mathworks.com/help/signal/ug/time-frequency-gallery.html>

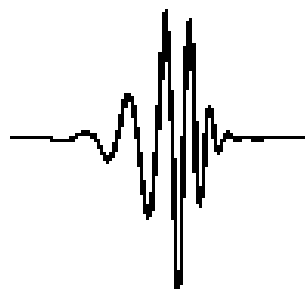


Využití waveletů v AI

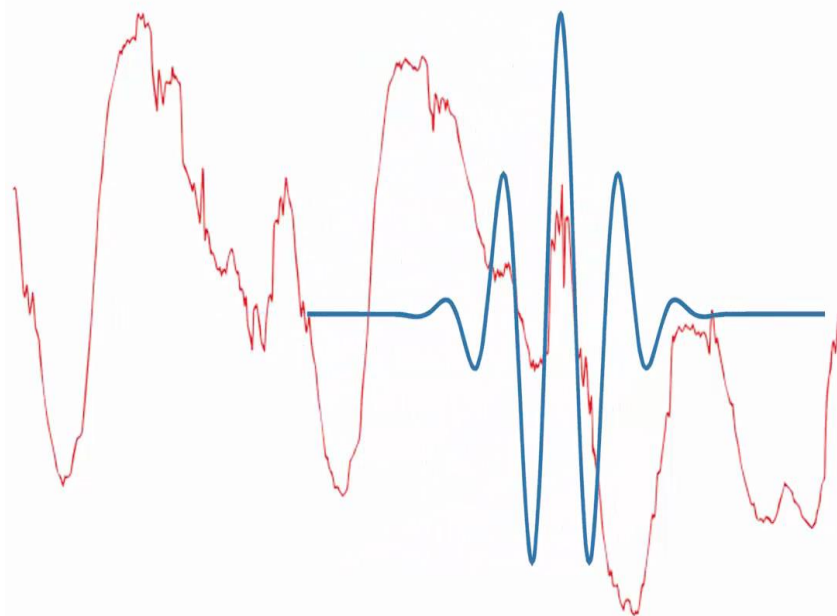
Proč jsou wavelety užitečné?



Sine Wave



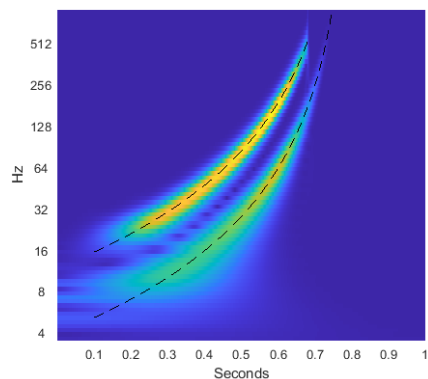
Wavelet (db10)



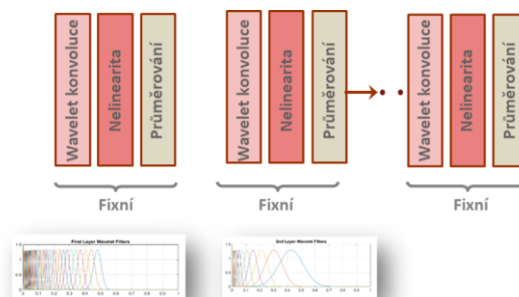
- Efektivní reprezentace dat charakterizovaných přechodovými jevy a přetrvávajícími trendy nebo oscilacemi
- Aktuální výsledky v oblastech:
 - detekce anomálií a monitorování stavu systému
 - analýza biomedicínských signálů
 - seismická analýza
 - radarové systémy a komunikace
 - finanční analýza

Které techniky založené na waveletech jsou vhodné pro AI?

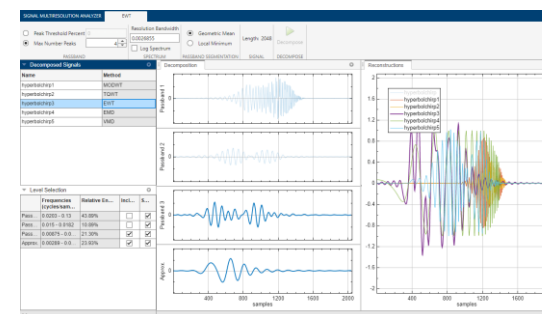
- Můžete využít ...



Continuous Wavelet Transform



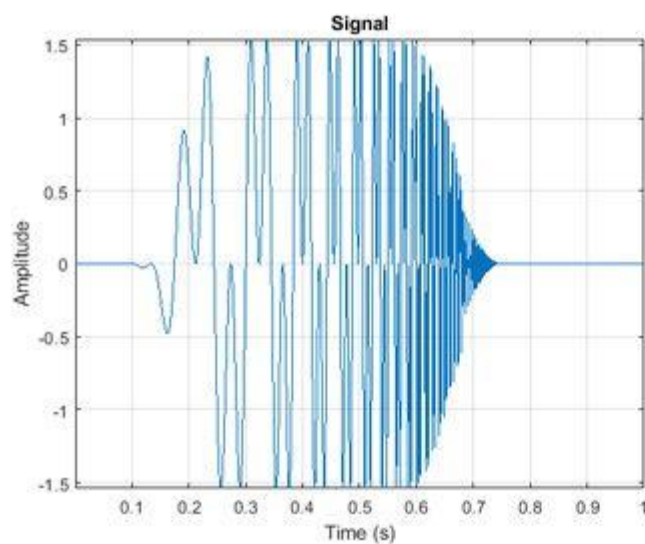
Wavelet Scattering



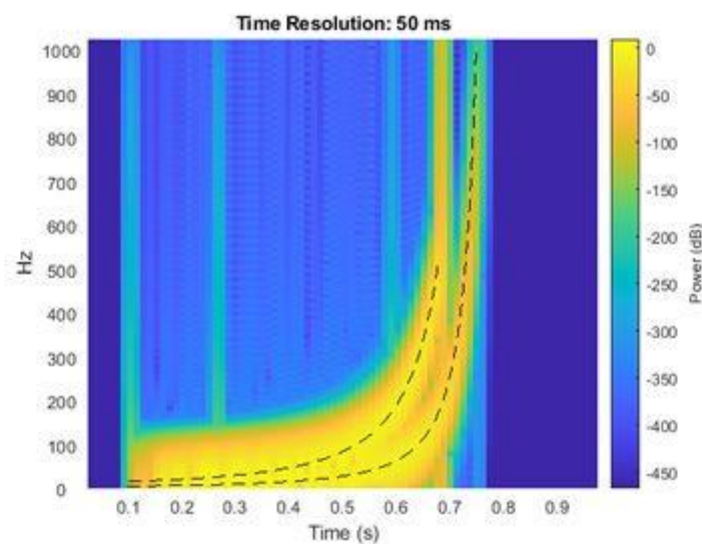
Multiresolution analysis

... a spoustu dalších

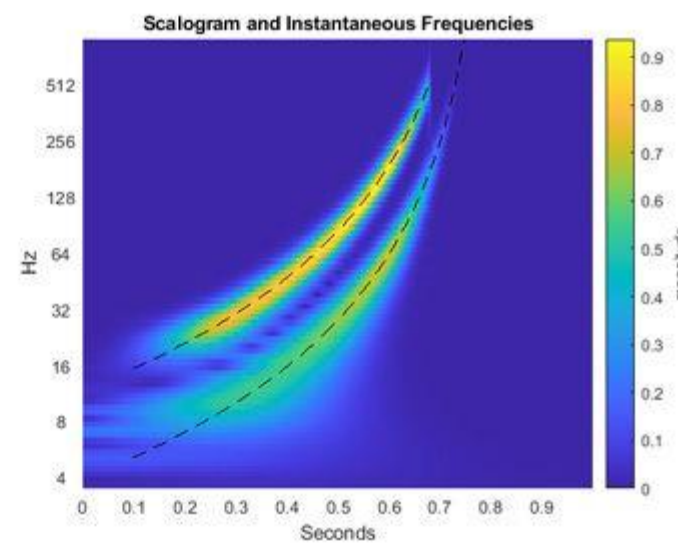
Continuous Wavelet Transform



časová oblast



spectrogram



CWT

Wavelet Scattering

CNN

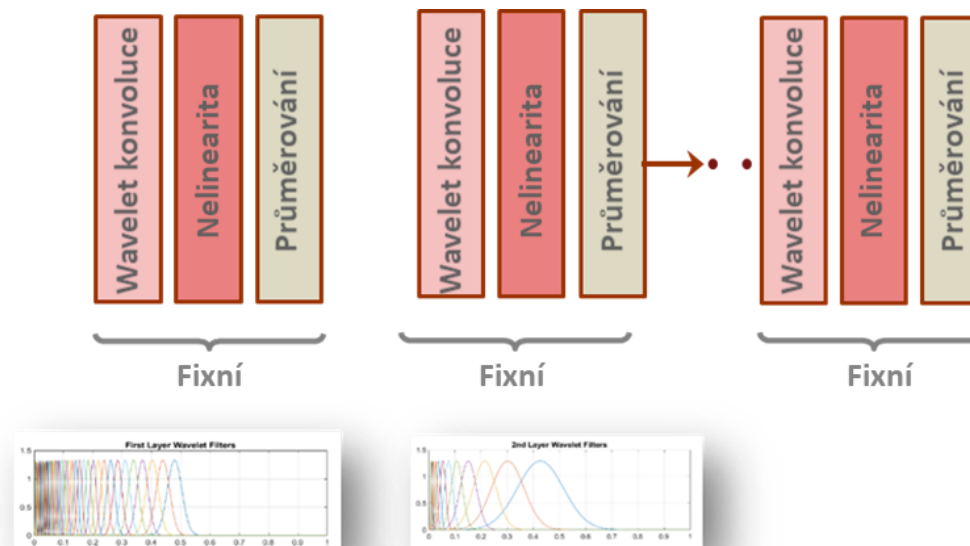


Pseudekód:

```

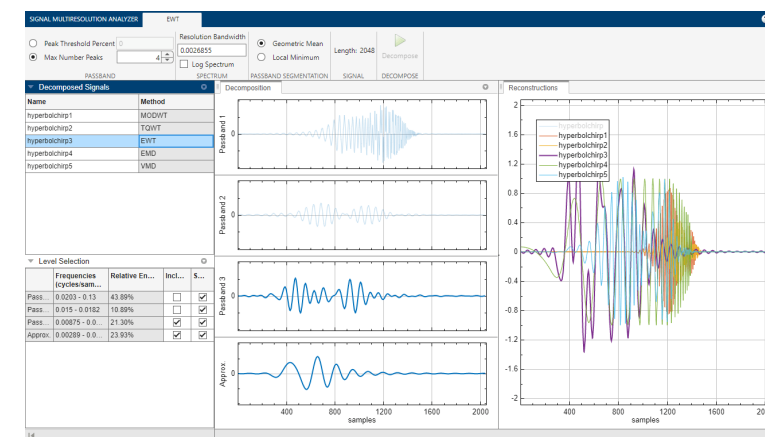
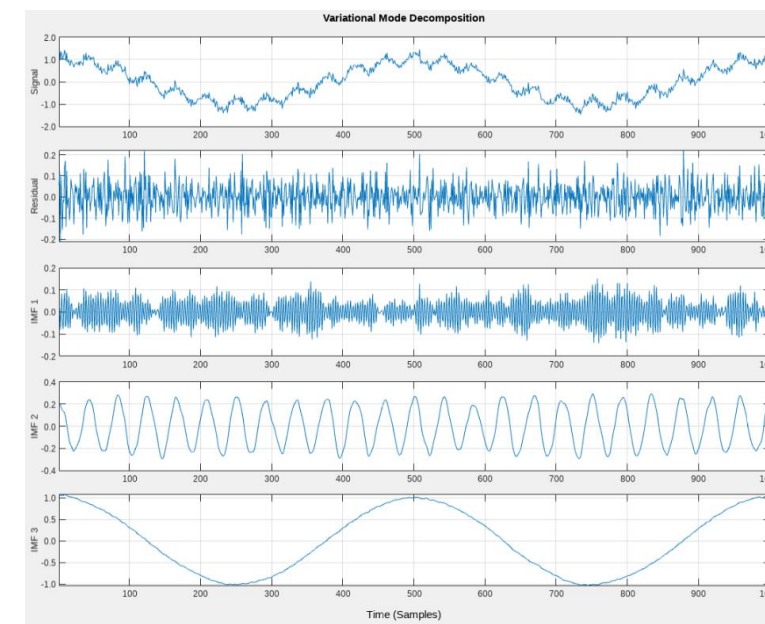
sf = waveletScattering(SignalLength) ;
Cyklus přes signal
  waveletFeature = featureMatrix(sf,signal)
  Přidej waveletFeature k tabulce příznaků
  Přiřaď popisek
end
  
```

Wavelet Scattering



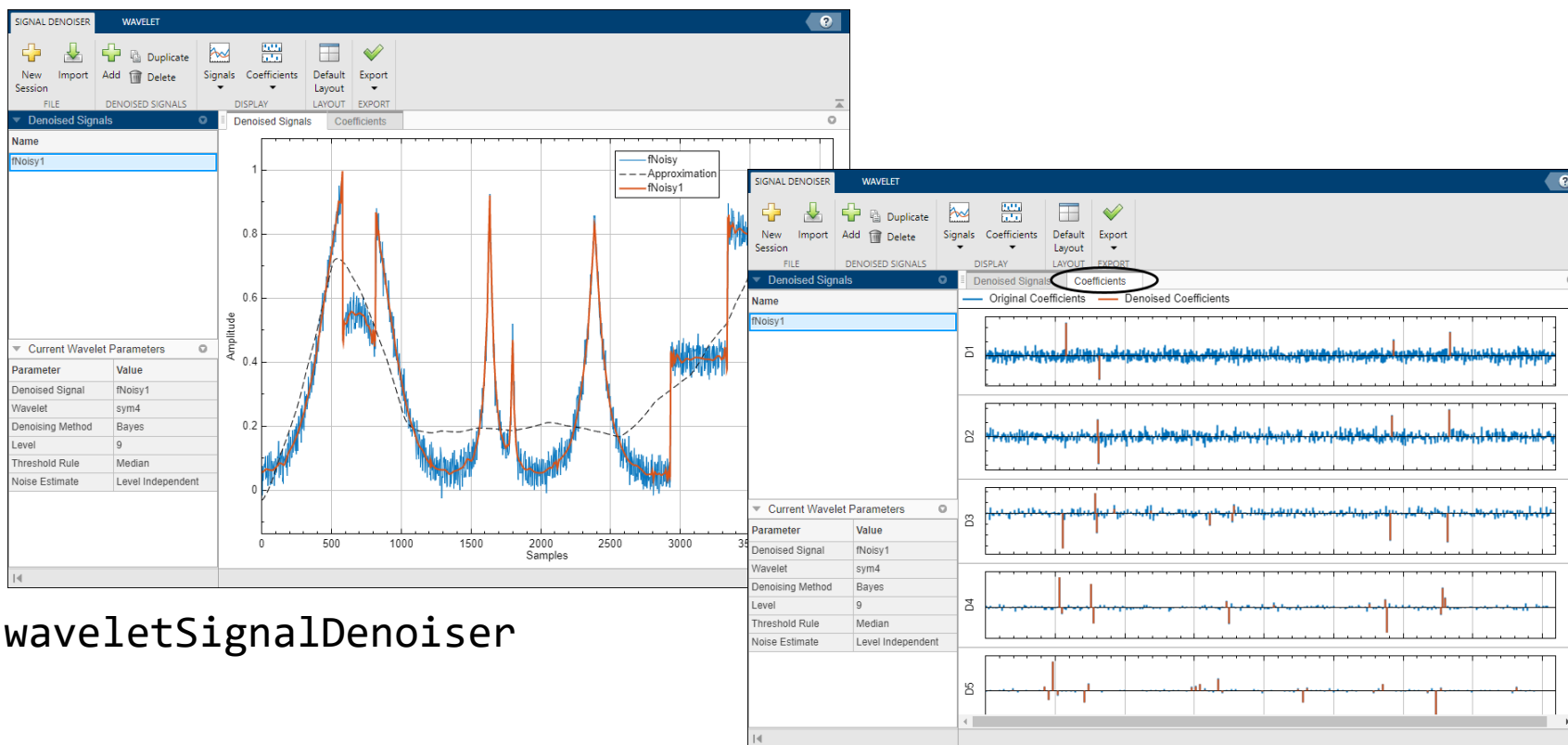
Multiresolution analysis (MRA)

- Dekompozice reálného signálu do několika složek
 - umožní dekompozici nestacionárních a nelineárních signálů
- Různé metody
 - maximal overlap discrete wavelet transform (MODWT)
 - empirical mode decomposition (EMD)
 - empirical wavelet transform (EWT)
 - tunable Q-factor wavelet transform (TQWT)
 - variational mode decomposition (VMD)
- Signal Multiresolution Analyzer app



Wavelet denoising

- Umožní odstranění negausovského šumu
 - signál, který má šum ale také prudké změny charakterizující abnormalitu.



Výběr nejlepších prediktorů – feature selection

- **Výběr vhodných prediktorů k odstranění redundance**
- Iterativní záměna k dosažení optimálního chování modelu
 - sequential feature selection, stepwise regression
- Hodnocení dle vzájemných charakteristik
 - MRMR, chi2, ftest
- NCA, ReliefF
- Vestavěná hodnocení prediktorů při učení
 - např. linear SVM, boosted decision trees, ...

Algorithm	Training	Types of Models	Accuracy
NCA	Moderate	Better for distance-based models	High
MRMR	Fast	Any	High
ReliefF	Moderate	Better for distance-based models	Medium
Sequential	Slow	Any	High
F test	Fast	Any	Medium
Chi-square	Fast	Any	Medium

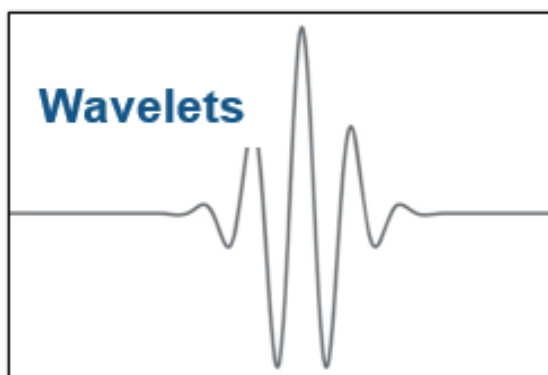
Transformace prediktorů – feature transformation

- **Alternativa k výběru příznaků**
- Méně významné prediktory také potlačeny
- PCA
- Factor analysis
- Nonnegative matrix factorization

Příklad využití – AutoML

- Automated Machine Learning (**AutoML**)
 - Generování příznaků ze signálových (obrazových) dat
 - Výběr skupiny nejlepších příznaků
 - Výběr nejlepšího modelu a hyperparametrů

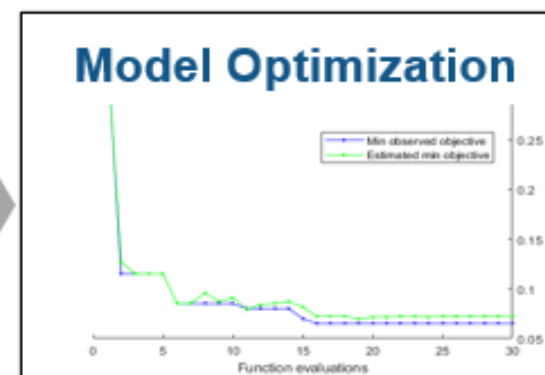
Wavelet Scattering



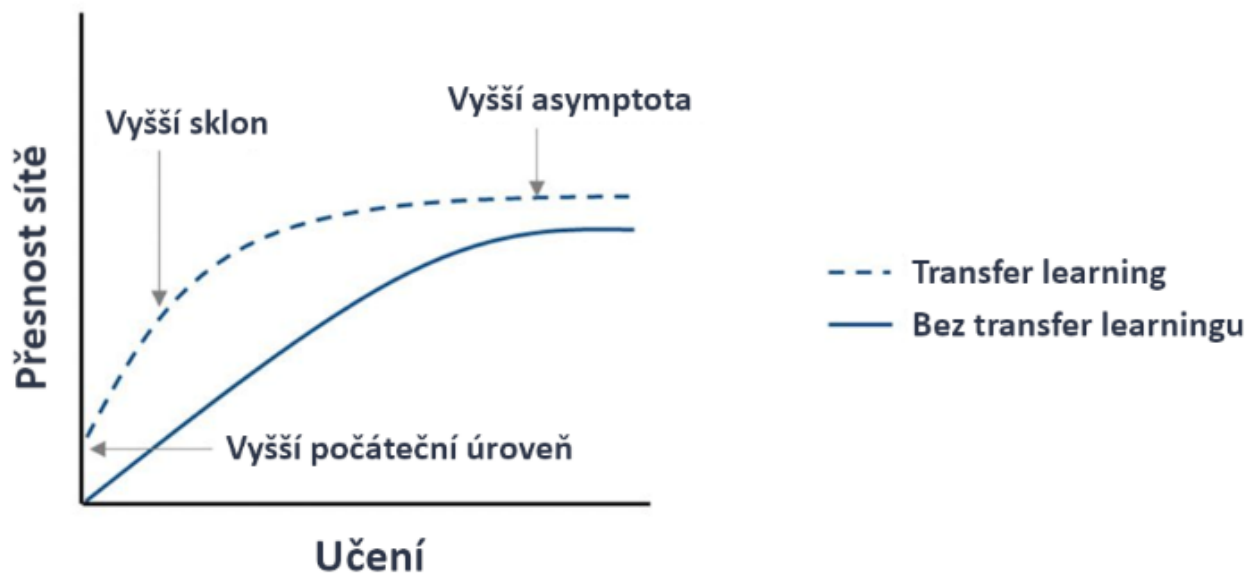
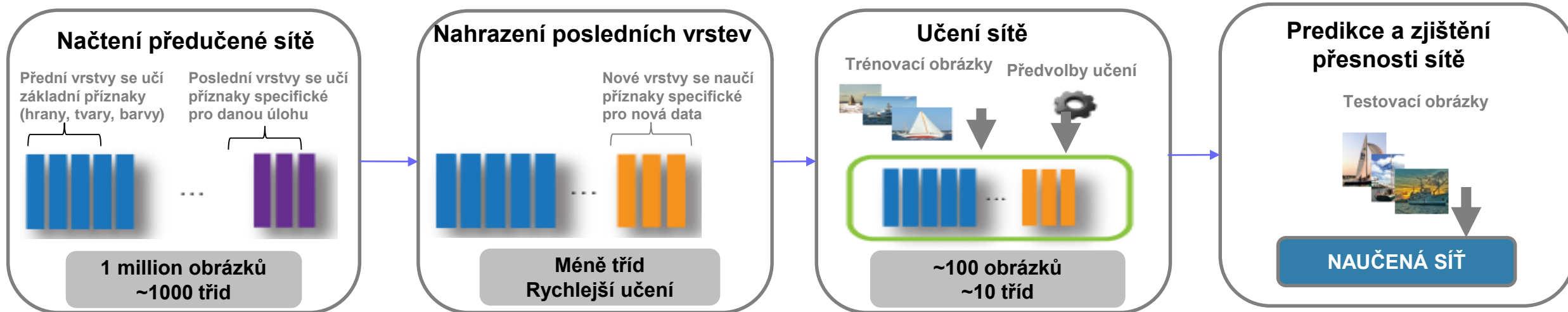
MRMR



fitcauto / fitrauto

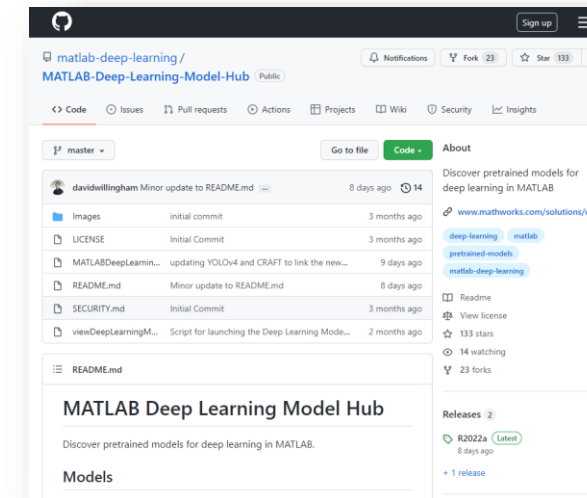
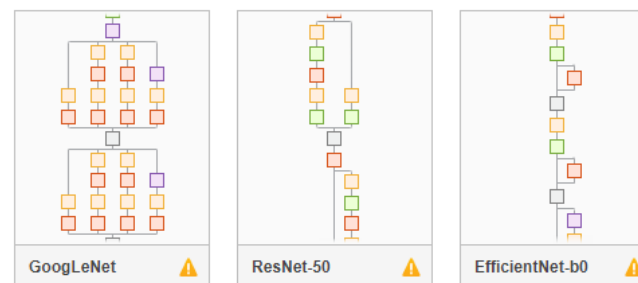


Využití metody „transfer learning“ s předučenými modely



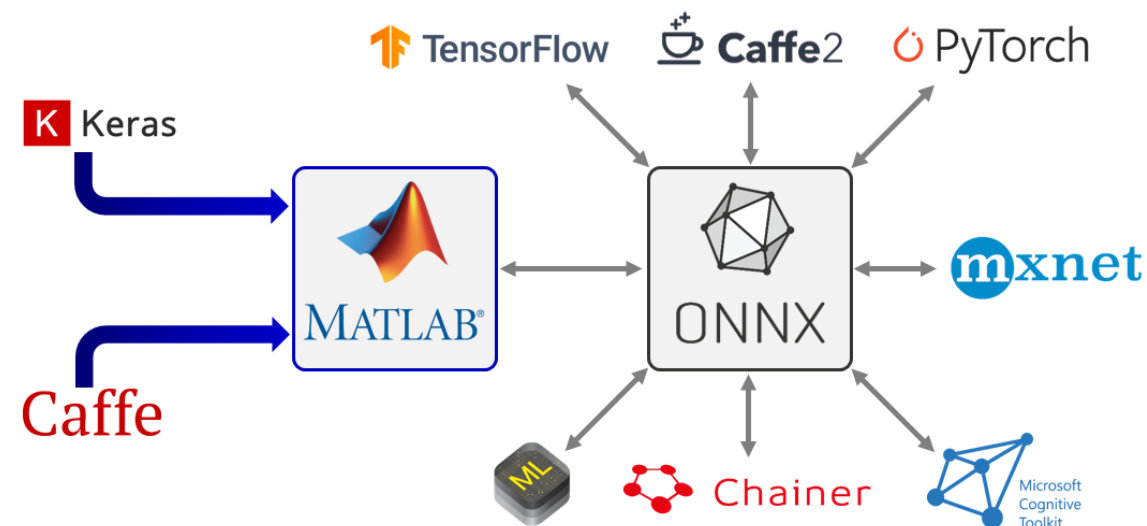
Kde vzít předučené modely?

- Mnoho modelů dostupných přímo v prostředí MATLAB



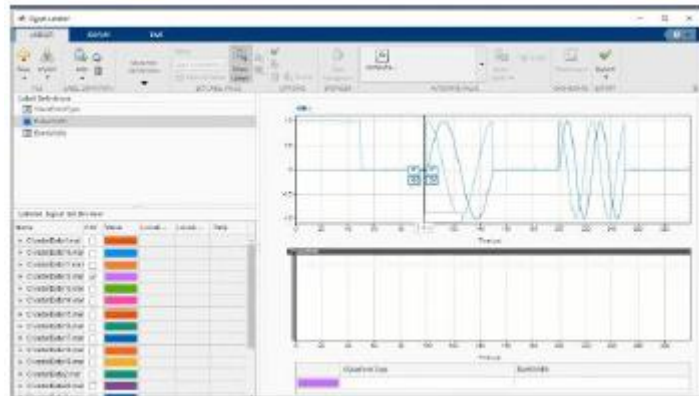
<https://github.com/matlab-deep-learning/MATLAB-Deep-Learning-Model-Hub>

- Import z jiných prostředí



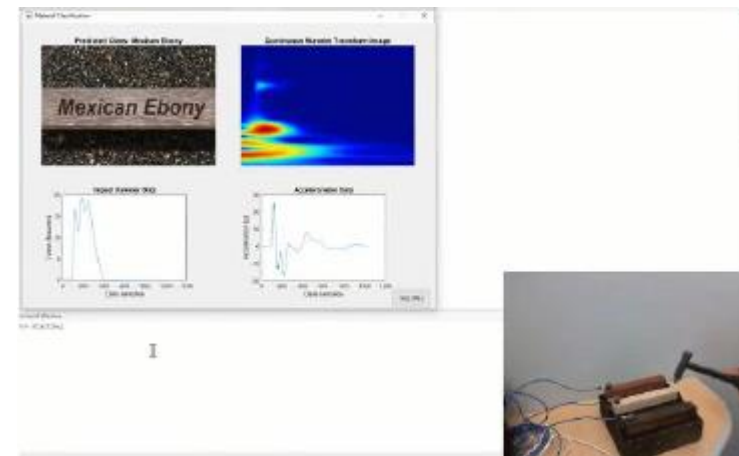
Jak zlepšit kvalitu dat určených pro učení?

Přesné označení vzorků dat

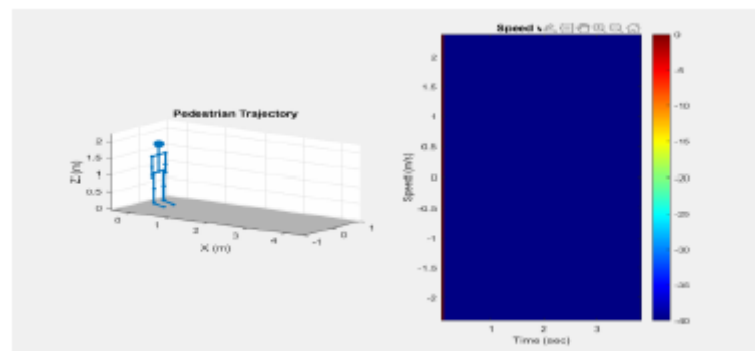


signalLabeler

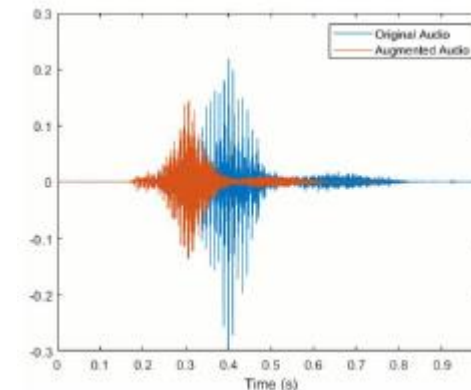
Nová data z připojeného hardware



Syntéza dat pomocí simulací



Umělé rozšíření variability dat



audioDataAugmenter

Děkuji za pozornost