

# Ventilator Pressure Prediction

Google Brain Kaggle competition

Michaela Mašková, Vít Pánek

České vysoké učení technické

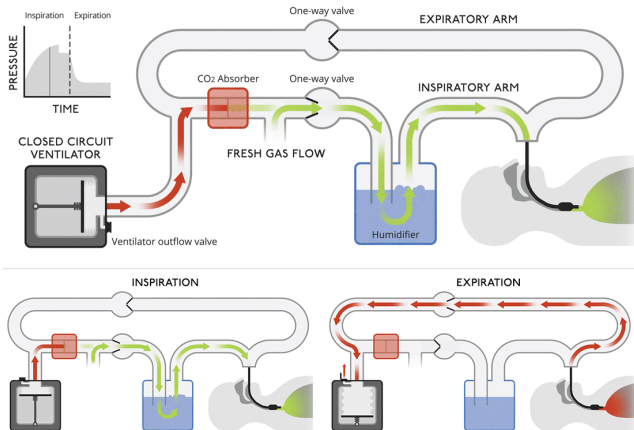
FJFI

Strojové učení 2

6.1.2022



**Cíl:** predikovat tlak na senzoru připojeného k plíci pacienta na základě parametrů (animace).



Trénovací soubor: train.csv – 75k sekvencí/6mil bodů,  
testovací soubor: test.csv – 50k sekvencí/4mil bodů.

**Input:** sekvence 80 bodů s proměnnými

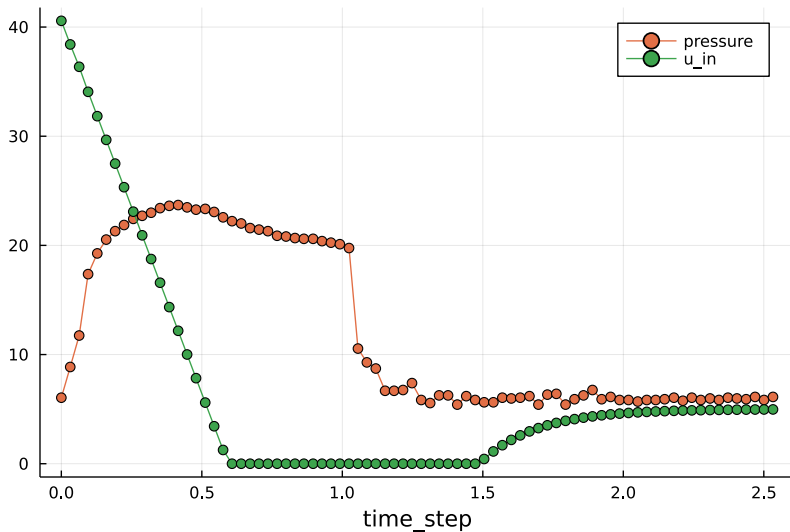
- time\_step – čas,
- u\_in – míra otevřenosti ventilu (0-100),
- R, C – plicní atributy rezistence a poddajnost (3 úrovně),
- u\_out – indikátor, jestli se jedná o nádech nebo výdech.

**Output:** sekvence 80 bodů predikovaného tlaku.

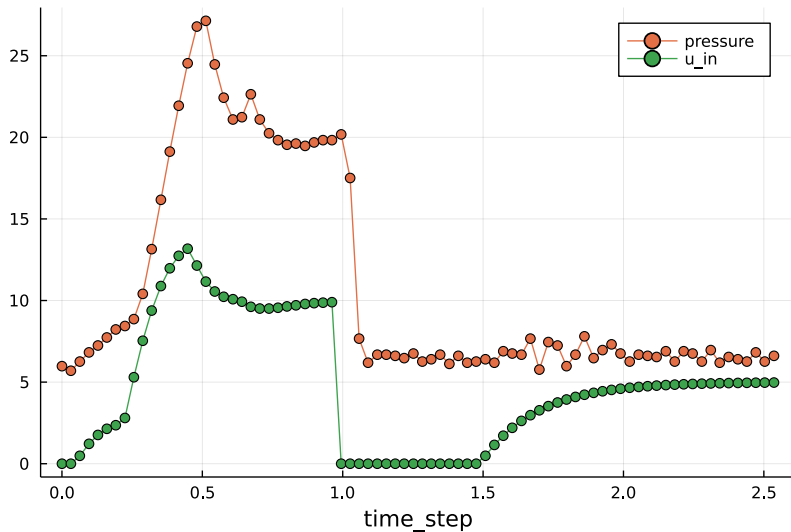
(jednotlivé sekvence by měly být nezávislé)



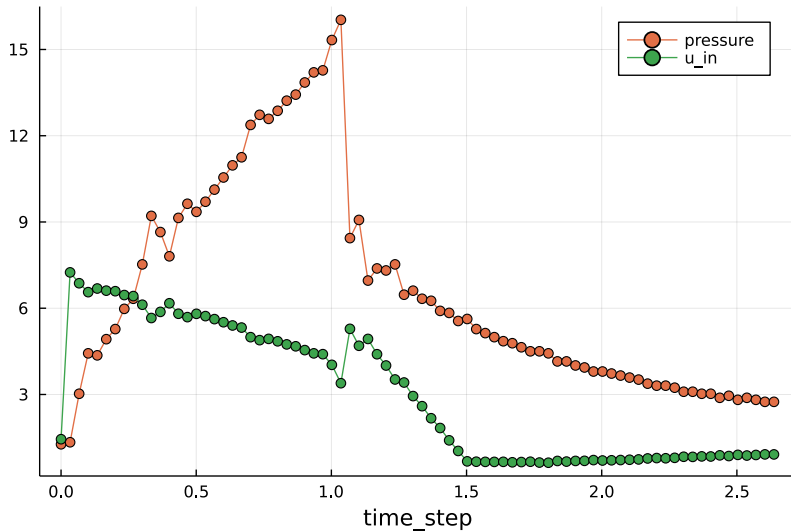
# Data – vizualizace sekvencí



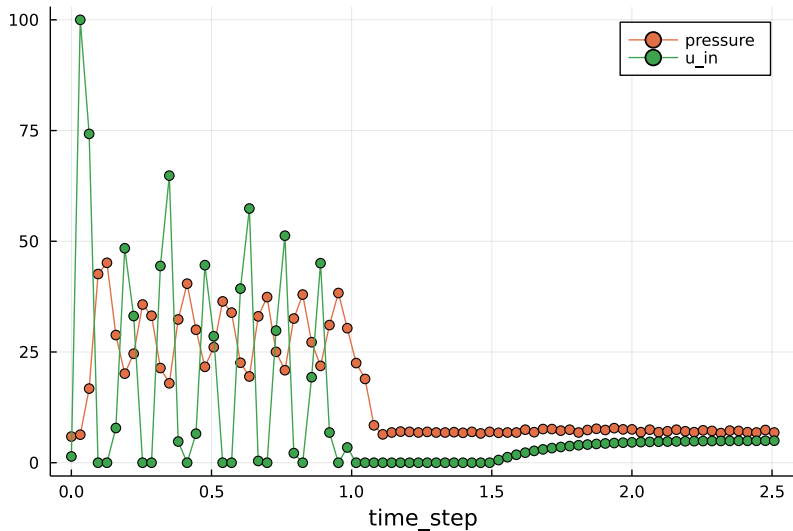
# Data – vizualizace sekvencí



# Data – vizualizace sekvencí

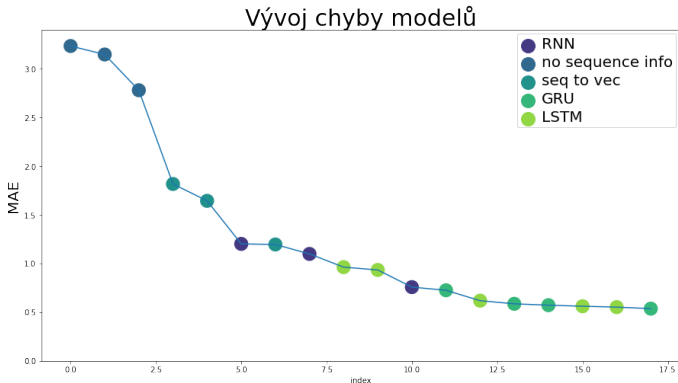


# Data – vizualizace sekvencí



# Vývoj řešení

- zahodit informaci o sekvencích a predikovat jednotlivé tlaky nezávisle
- sekvenci převést na vektor (dimenze  $\approx 162$ ) a predikovat vektor tlaku o dimenzi 80
- použít rekurentní neuronové sítě + feature engineering
- trénovat 9 rekurentních modelů + feature engineering





## Použité features

- features z dat: time\_step, u\_in, případně R, C,
- statistiky: průměr, minimum, maximum z proměnné u\_in zvlášť pro nádech a výdech,
- rozdíly (lagy) v u\_in,
- statistiky rozdílových vektorů (lagů).

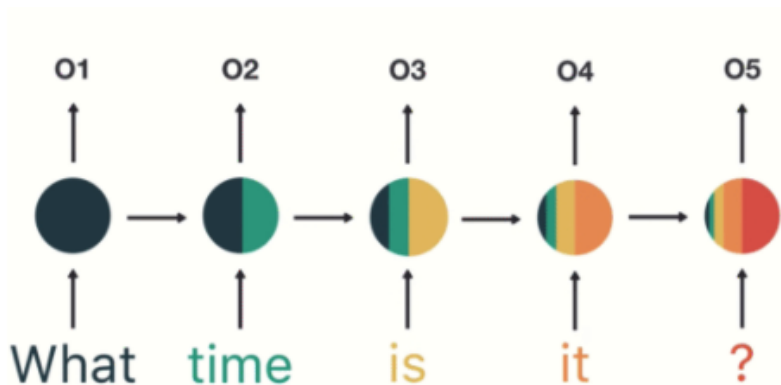
Celkem 42 features jako input do modelu.



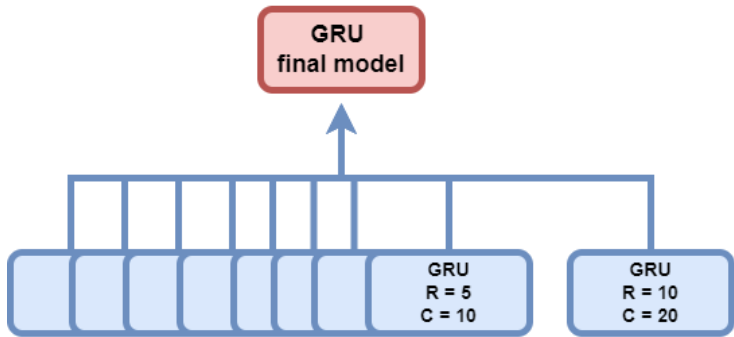
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z  
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A

E





- Colab + vlastní PC,
- RNN ne, LSTM a GRU ano,
- ne jeden model... ale hned devět modelů!



## Trénink

- HELIOS cluster (FJFI),
- random search na parametrech,
- minibatching na 64 sekvencích,
- využití validačních dat k uložení nejlepších modelů po každé epoše.

## Shrnutí

- natrénováno 54 finálních modelů (LSTM, GRU) – zdroje: 50G RAM a 5 CPU pro každý z modelů, 24 hodin
- celkem natrénováno  $\pm 500$  RNN modelů a  $\pm 500$  vektorových modelů.



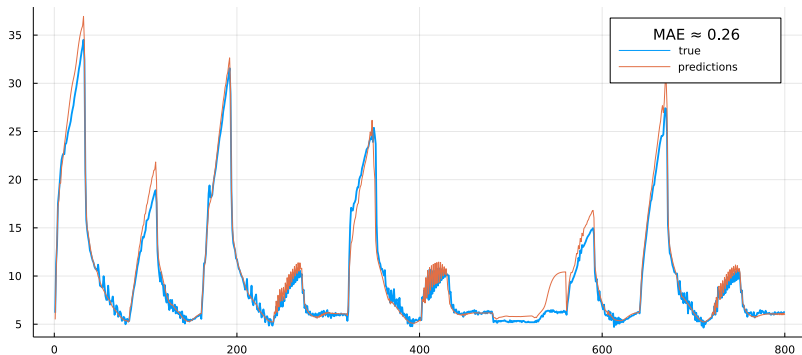
## Parametry nejlepšího modelu

- GRU buňka,
- 3 rekurentní vrstvy, 1 lineární vrstva,
- 32 neuronů ve skrytých vrstvách,
- ADAM optimizer s  $\lambda = 0.001$ ,
- L1 loss (MAE)

⇒ celkem 19 809 parametrů.



Nejlepší model dosáhl chyby  $MAE = 0.5365$  (oficiální Kaggle skóre).



Nejlepší model dosáhl chyby  $MAE = 0.5365$  (oficiální Kaggle skóre).

