KP LABS

Machine learning-powered algorithms for predicting the risk of satellite collisions

Michał Myller, Łukasz Tulczyjew, Michał Kawulok, Daniel Kostrzewa, Jakub Nalepa
{mmyller, ltulczyjew, mkawulok, dkostrzewa, jnalepa}@kplabs.pl

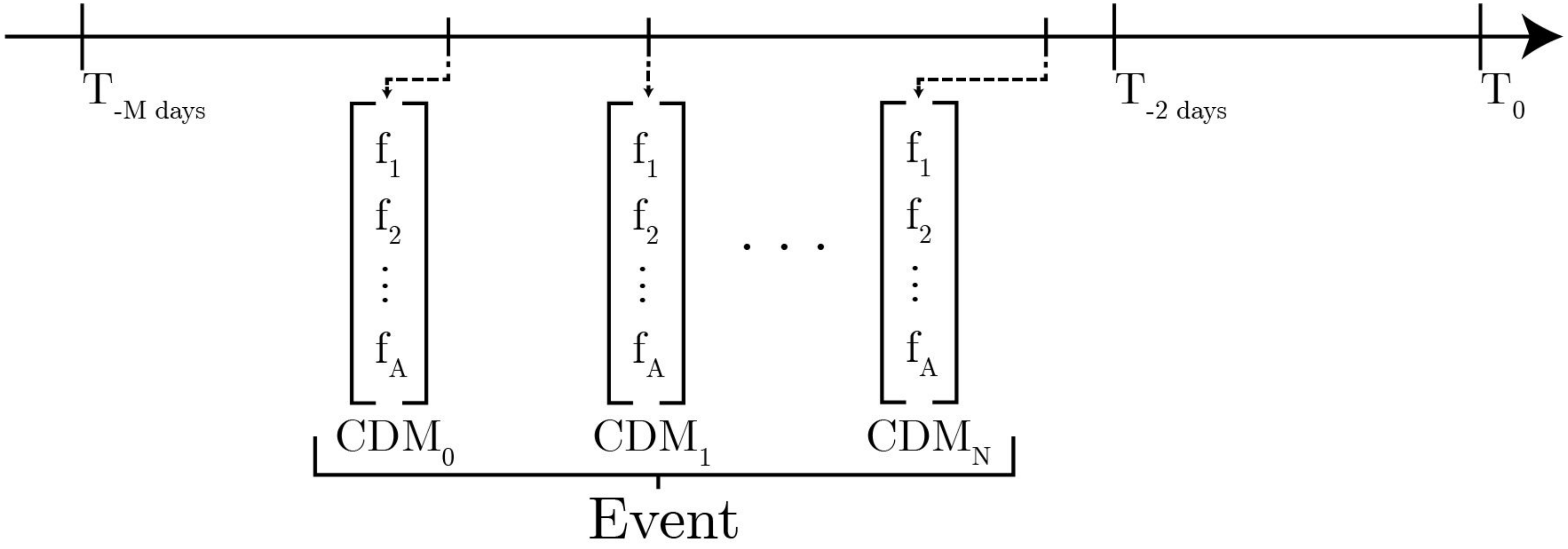
2021

ESA Collision Avoidance Challenge - goals and motivation

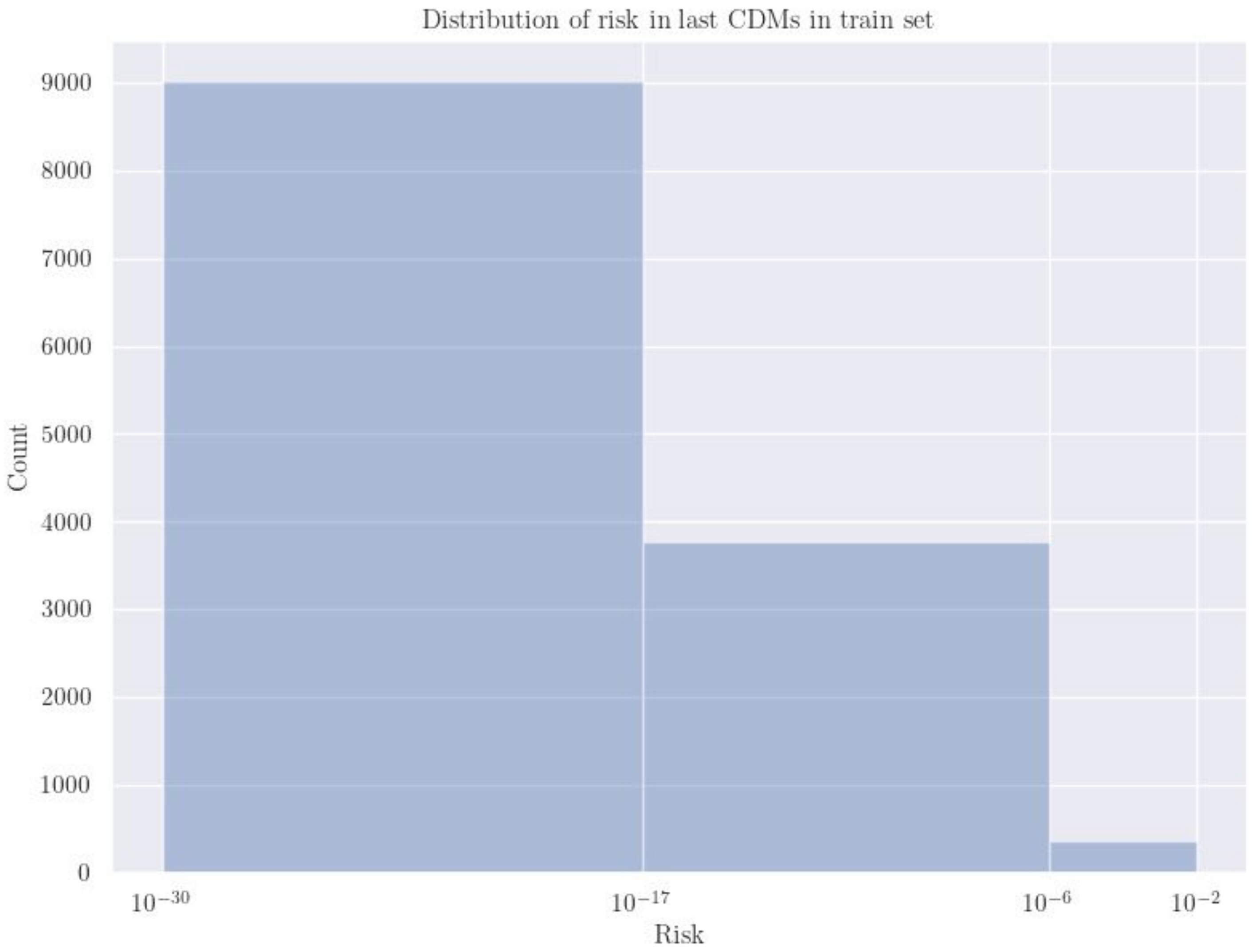
- Over 22300 objects tracked and catalogued by the Space Surveillance Network
- Hundreds of alerts concerning close encounters of a satellite and other space objects are issued every day
- Each year, more than one collision avoidance manoeuvre is performed per satellite
- <u>Design a model predicting final collision risk between a satellite and any other space object 2 days prior</u>

ESA Collision Avoidance Challenge - the data

- 13154 and 2167 unique events in the training and test sets
- Each event contains varying number of conjunction data messages (CDMs)
- 103 features/CDM (identity of the satellite in question and potential collider, time of closest approach, uncertainty, etc.)
- Each event is labeled as either high- or low-risk one



ESA Collision Avoidance Challenge - the data



KP LABS

ESA Collision Avoidance Challenge - scoring

Final score/loss is given by:

$$L(r, \hat{r}) = \frac{1}{F_2} MSE(r, \hat{r}),$$

where F_2 is computed over the whole dataset, using two classes, (high final risk: $r \geq 10^{-6}$, low final risk: $r < 10^{-6}$) and the MSE is only computed for the events that belong to the first class.

$$F_\beta = (1 + \beta^2) \frac{precision \times recall}{(\beta^2 \times precision) + recall},$$

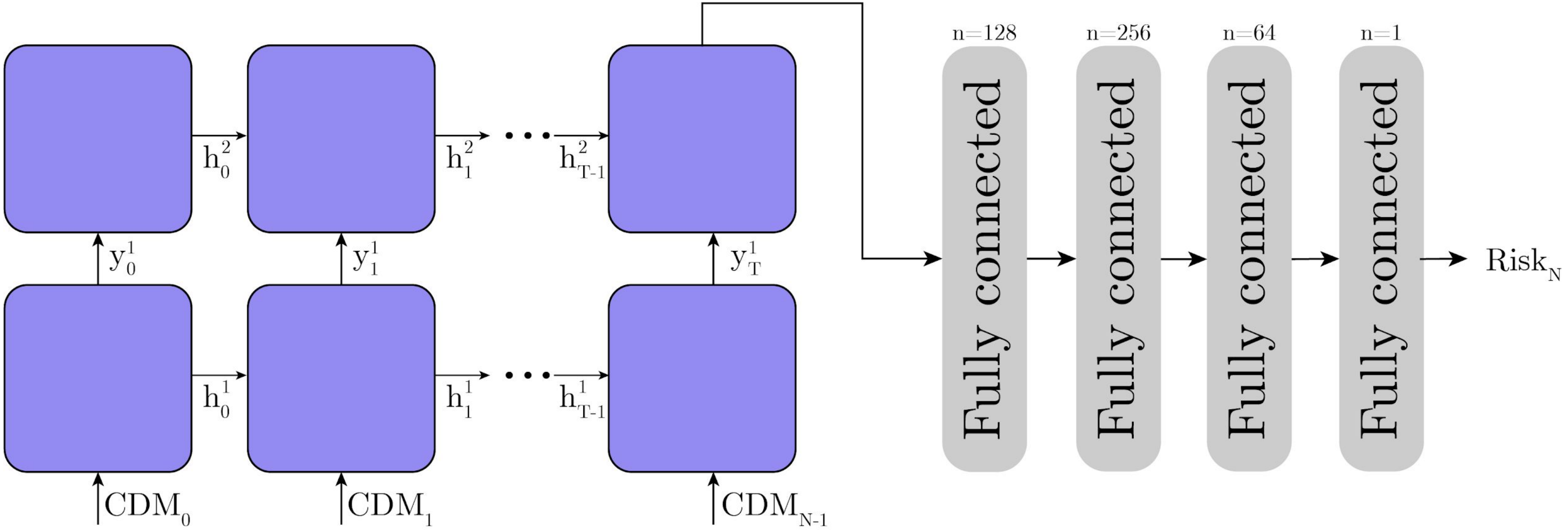
$$MSE(r, \hat{r}) = \frac{1}{N} \sum_{i=1}^{N} (r_i - \hat{r}_i)^2, \{i \mid r_i \geq 10^{-6}\}.$$

Baseline score (predicting always 10^{-5}): **2.502**

Approach 1 - RNN regresor

- Treat each event as a time series and utilize recurrent neural network as a latent vector extractor
- Predict the final risk based on all available CDMs
- Both LSTM and GRU cells tested
- Best score (GRU cell): **1.337**

Approach 1 - RNN regresor

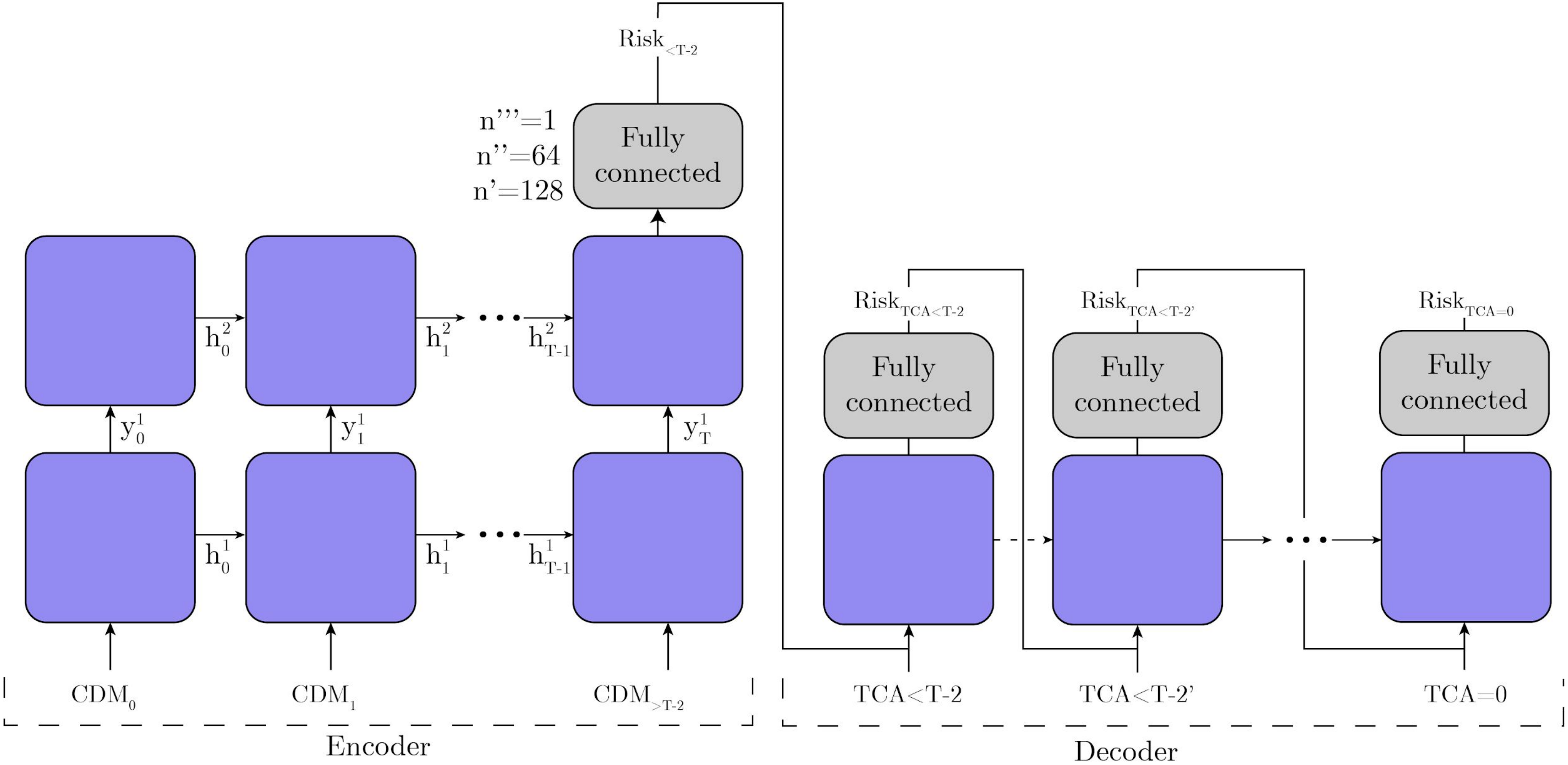


KP LABS

Approach 2 - Sequence to sequence

- Treat each event as a time series
- Encoder predicts the risk based on CDMs up to 2 days prior to the TCA
- Time steps after that point are unknown
- Decoder predicts the risk for the next time step until TCA=0, based on previously predicted risk and current TCA
- Number of decoder time steps has to be picked manually
- GRU cell used
- Best score: **1.999**

Approach 2 - Sequence to sequence



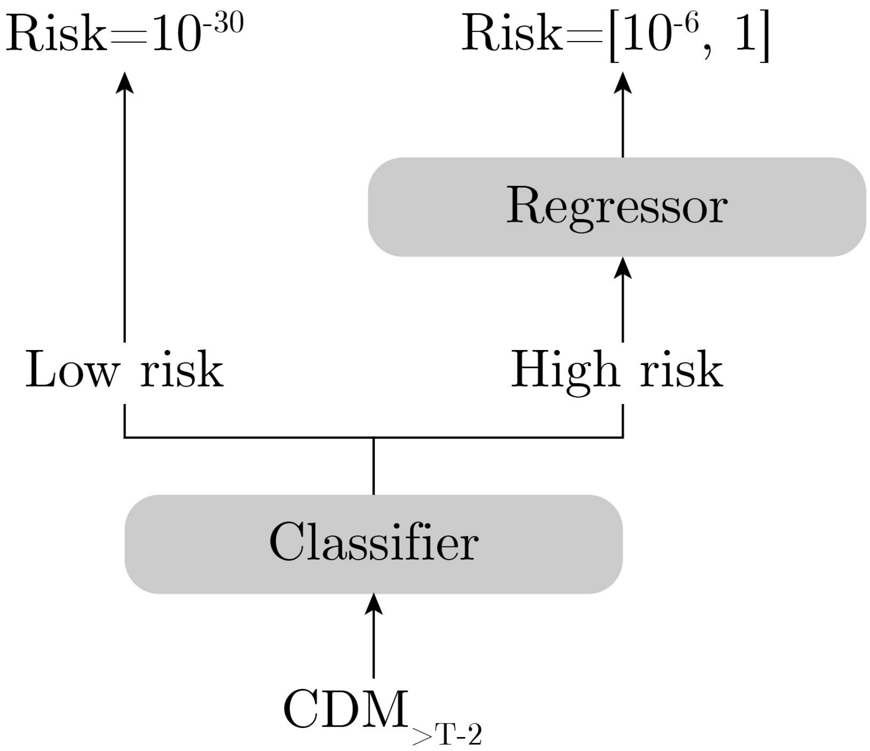
KP LABS

Approach 3 - classifier and regressor

- Do not treat each event as a time series
- CDM from each event closest to TCA=2 is selected as a training sample
- Its label/regressor value is the risk from the last CDM from the corresponding event
- Classifier and regressor are trained separately
- Regressor trained only on high-risk samples

Approach 3 - classifier and regressor

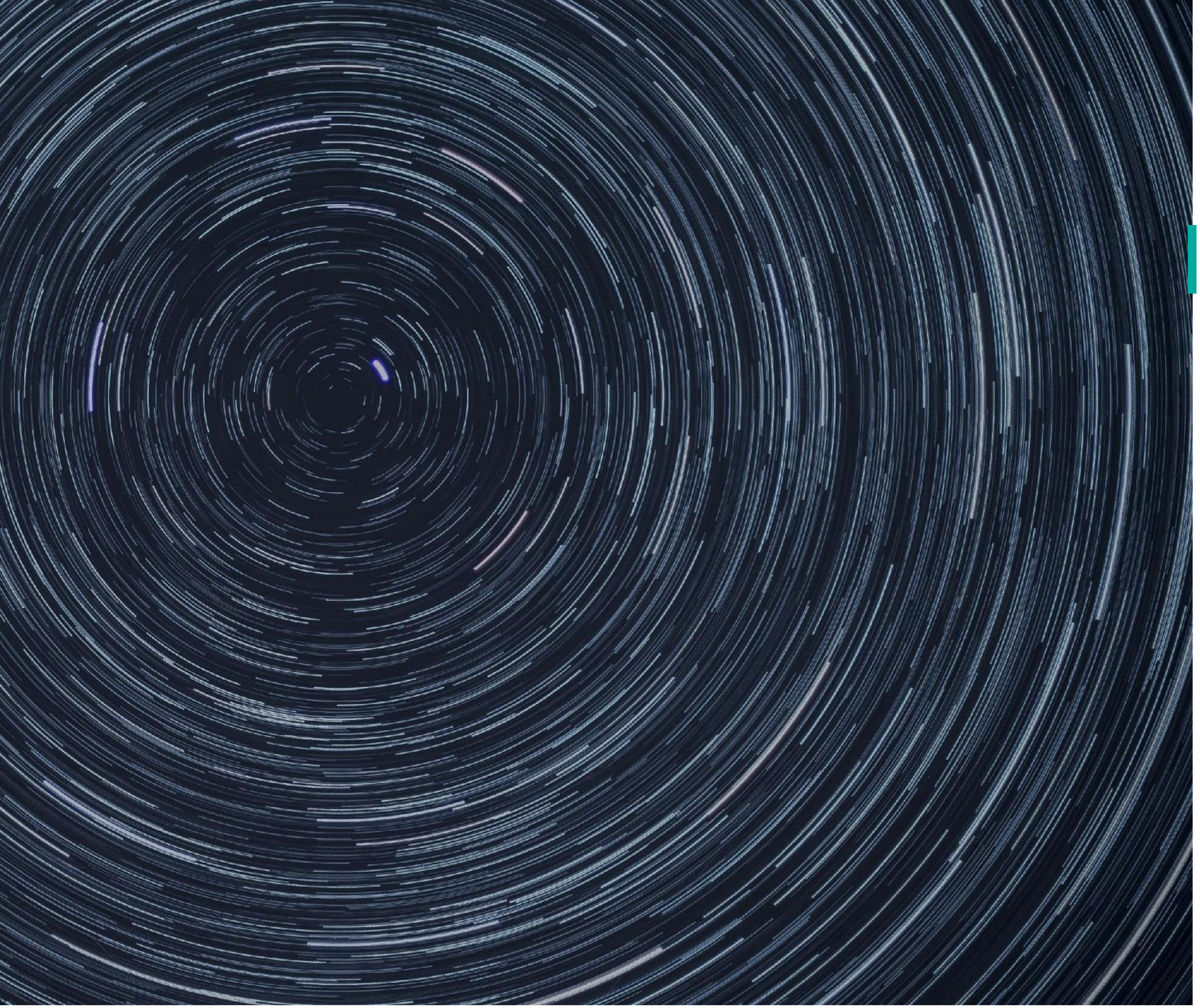
- For prediction, use only the last available CDM from the whole event
- Samples classified as high risk are then evaluated by the regressor to estimate the final risk
- Multilayer perceptron utilized as a classifier, random forest used for regression
- Best score: **0.628**



KP LABS

Conclusions

- Careful data preprocessing is crucial
- Very limited number of high-risk events
- More complex models simply overfitted
- Deep-learning is not always the way to go
- First place score: **0.555**
- Our best score (7th place): **0.628**



THANK YOU

mmyller@kplabs.pl
ltulczyjew@kplabs.pl
mkawulok@kplabs.pl
dkostrzewa@kplabs.pl
jnalepa@kplabs.pl

KP Labs Sp. z o.o,
ul. Stanisława Konarskiego 18C,
44-100 Gliwice,

kontakt@kplabs.pl

www.kplabs.pl