# Plans for ML Applications in AMPEL

Nicolas Miranda (HU Berlin)

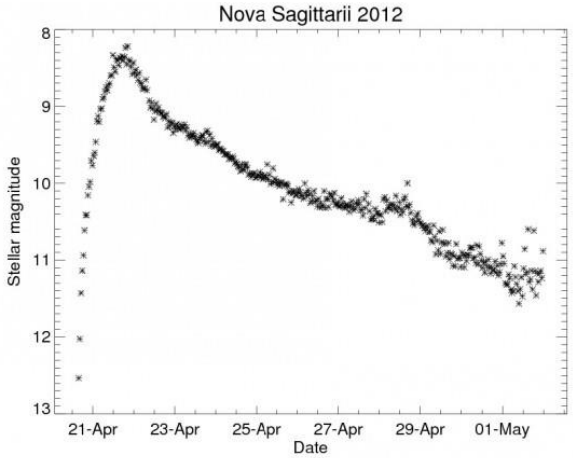
# HEIBRiDS program

- *Helmholtz Einstein International Research School in Data Science*
- Joint graduate program between the Einstein Center Digital Future (ECDF) together with the Helmholtz Association
- Cooperation between 3 Berlin universities, Charité and 6 Helmholtz centers
- Interdisciplinary program for training scientists in Data Science and other scientific disciplines at the same time
- Supervisors: Prof. Johann-Christoph Freytag and Prof. Marek Kowalski

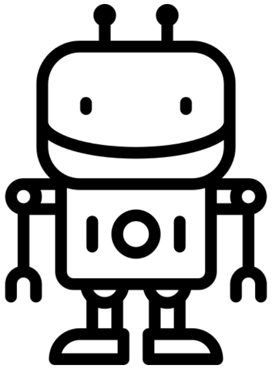
# Why use ML in AMPEL?

- New telescopes like LSST will greatly increase transient detection capabilities.
- However, spectroscopic follow up capabilities of these sources has not increased at the same rate.
- ML automatization of the detection process from photometric data is a good solution for this problem.
- Supernova cosmology surveys with LSST should provide a large amount of samples for ML model training.

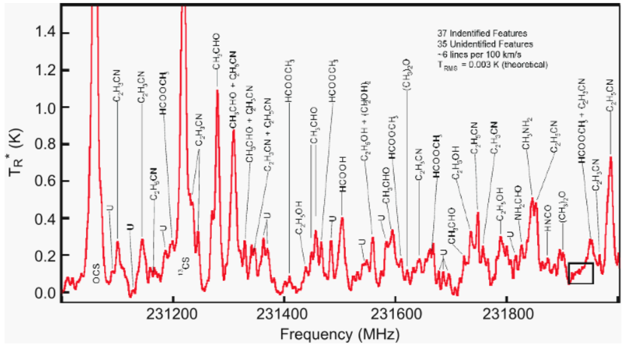
# Use cases of ML in AMPEL




Classify full light curves using a ML model



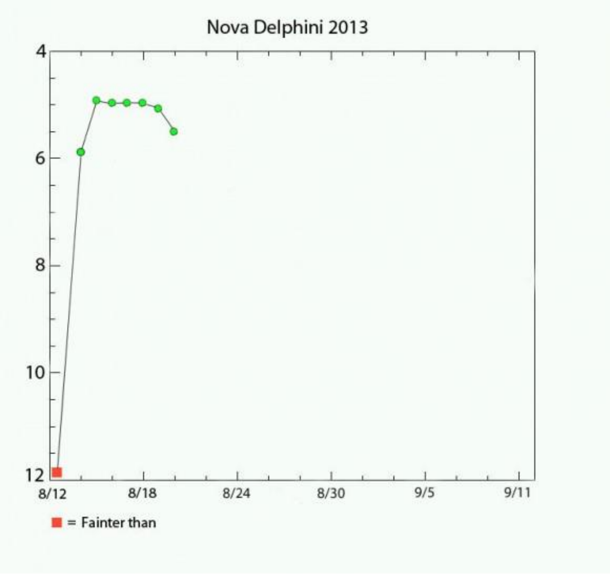
Use spectroscopic information to verify if the classification is correct

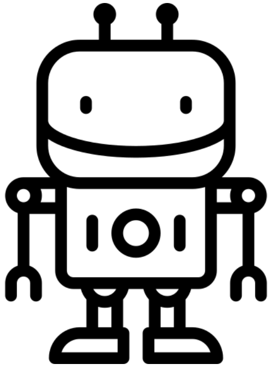



If the classification done by the model is correct enough, then we can use the model directly to identify objects by their photometric data without the need for spectroscopic follow up.

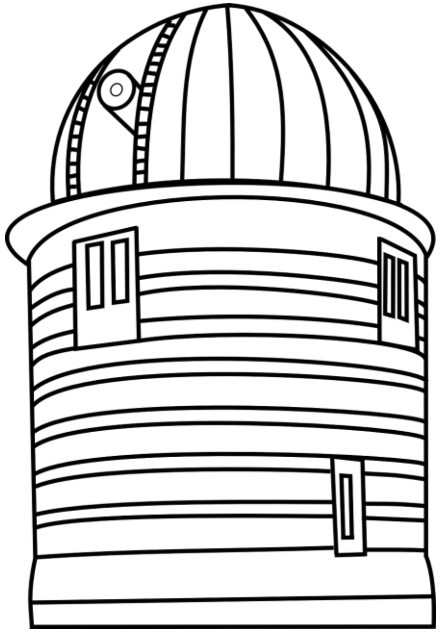
Image credits: NASA, http://www.flaticon.com/authors/freepik , http://cv.nrao.edu

# Use cases of ML in AMPEL



Identify early if a light curve doesn't belong to a known class



Trigger a photometric and spectroscopic follow up of this unusual object

If the classification done by the model is correct enough, then we can use the model directly to identify objects by their photometric data without the need for spectroscopic follow up.

Image credits: https://www.universetoday.com/ , http://www.flaticon.com/authors/freepik , https://pixabay.com/

# Use cases of ML in AMPEL

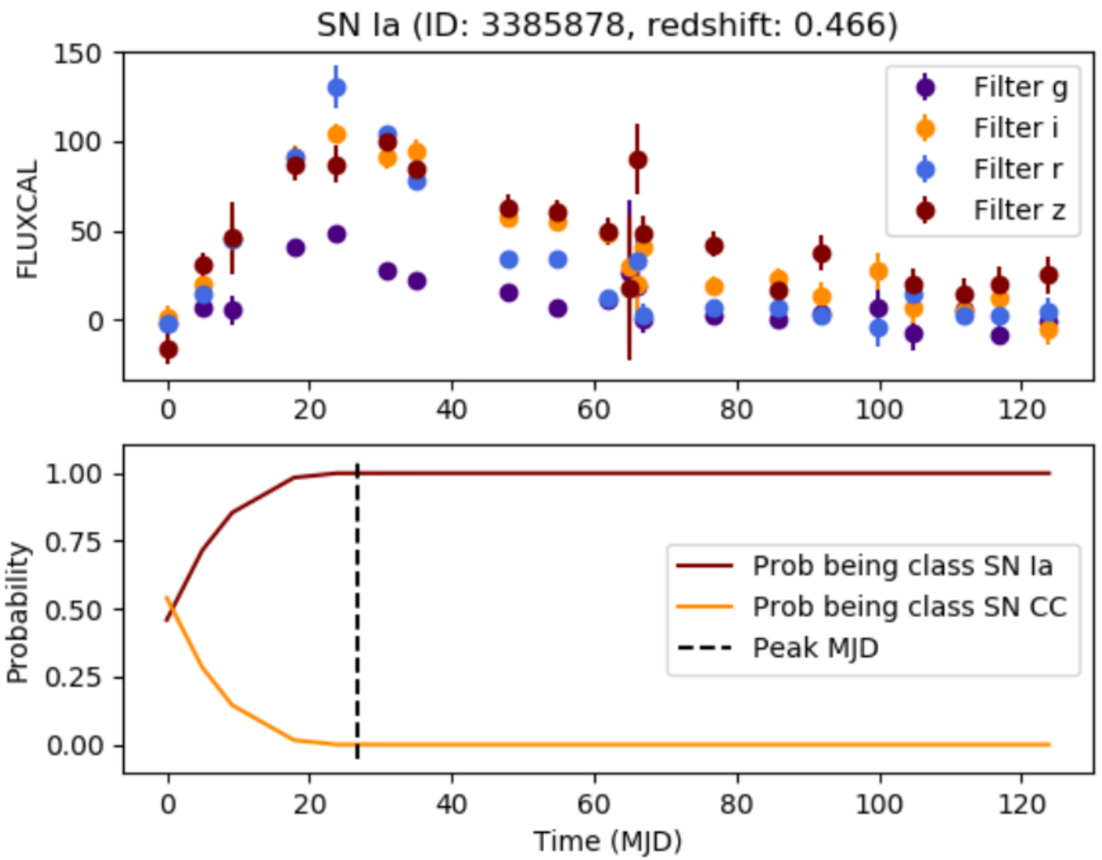
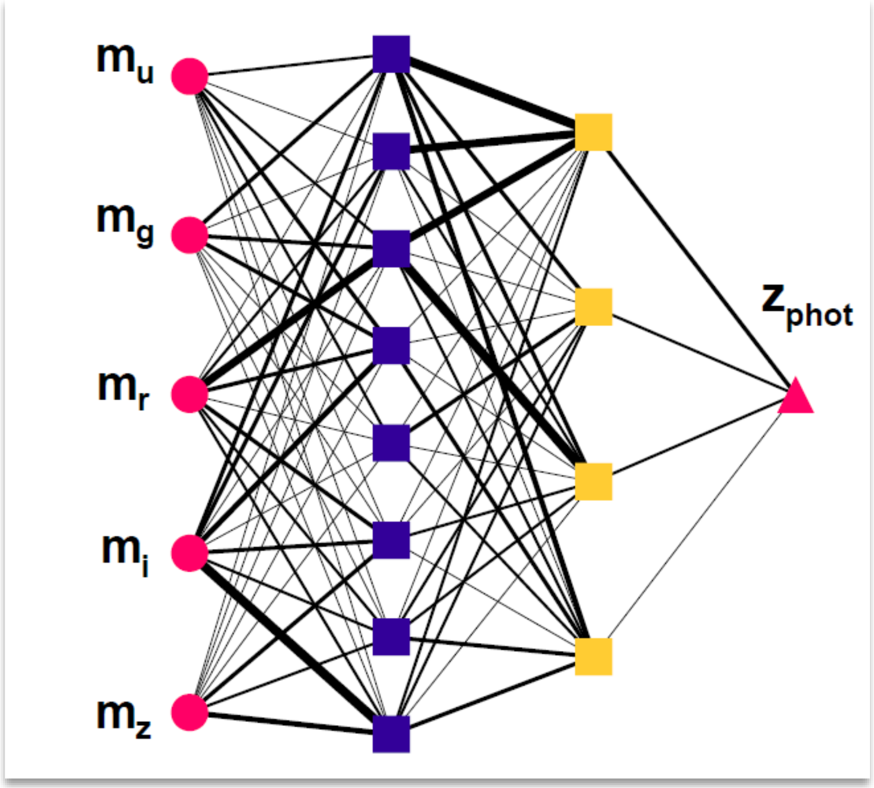



Image credits: Anais Möller, Thibault de Boissière (2019)

# Current use of ML in AMPEL: ANNz2



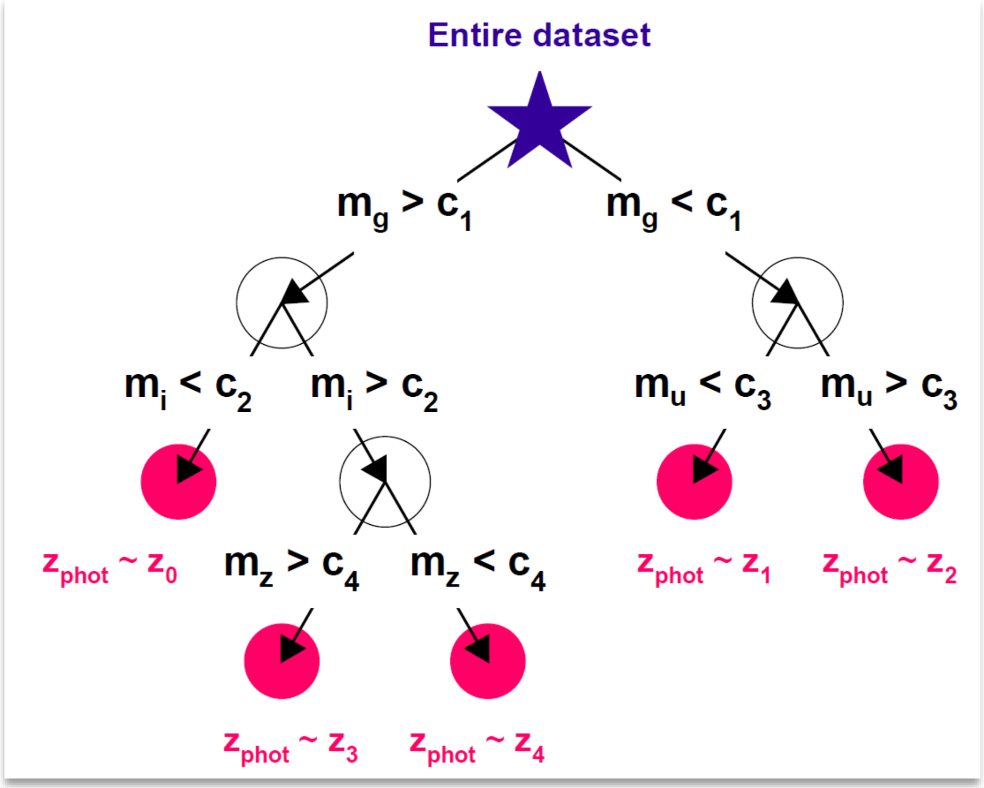
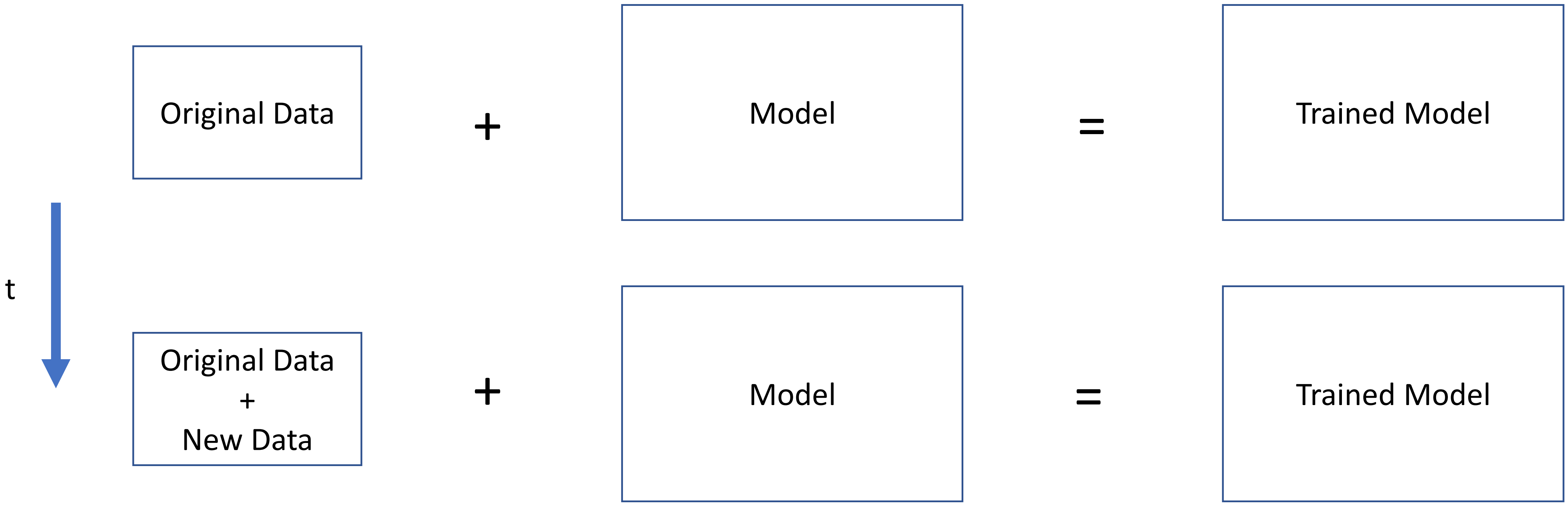




Image credit: I. Sadeh, F. B. Abdalla and O. Lahav (2016)

# Future applications: ML and streams

- The most direct and information-rich way to interpret astronomical transient data is as alert streams.
- This would imply giving a ML model dynamic data as input, instead of a set of static data.
- Training a model in this setup becomes the challenge.
  - One could “un-stream” the data after receiving and saving it, and later treat it as if were static training data.
  - However, it would be desirable to train a model directly over the data stream by using a process which involves incremental learning.

# Naïve approach for ML training with streams



# ML and streams: batch vs example

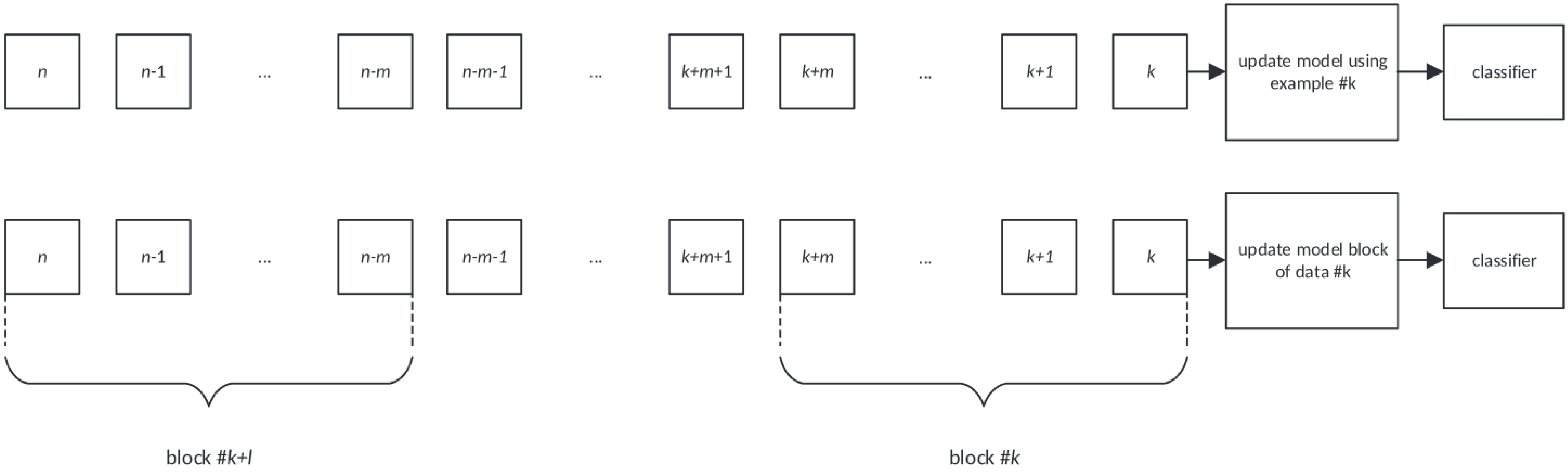



Image credit: B.Krawczyketal. /Information Fusion 37 (2017)

# ML and streams: ADAIN

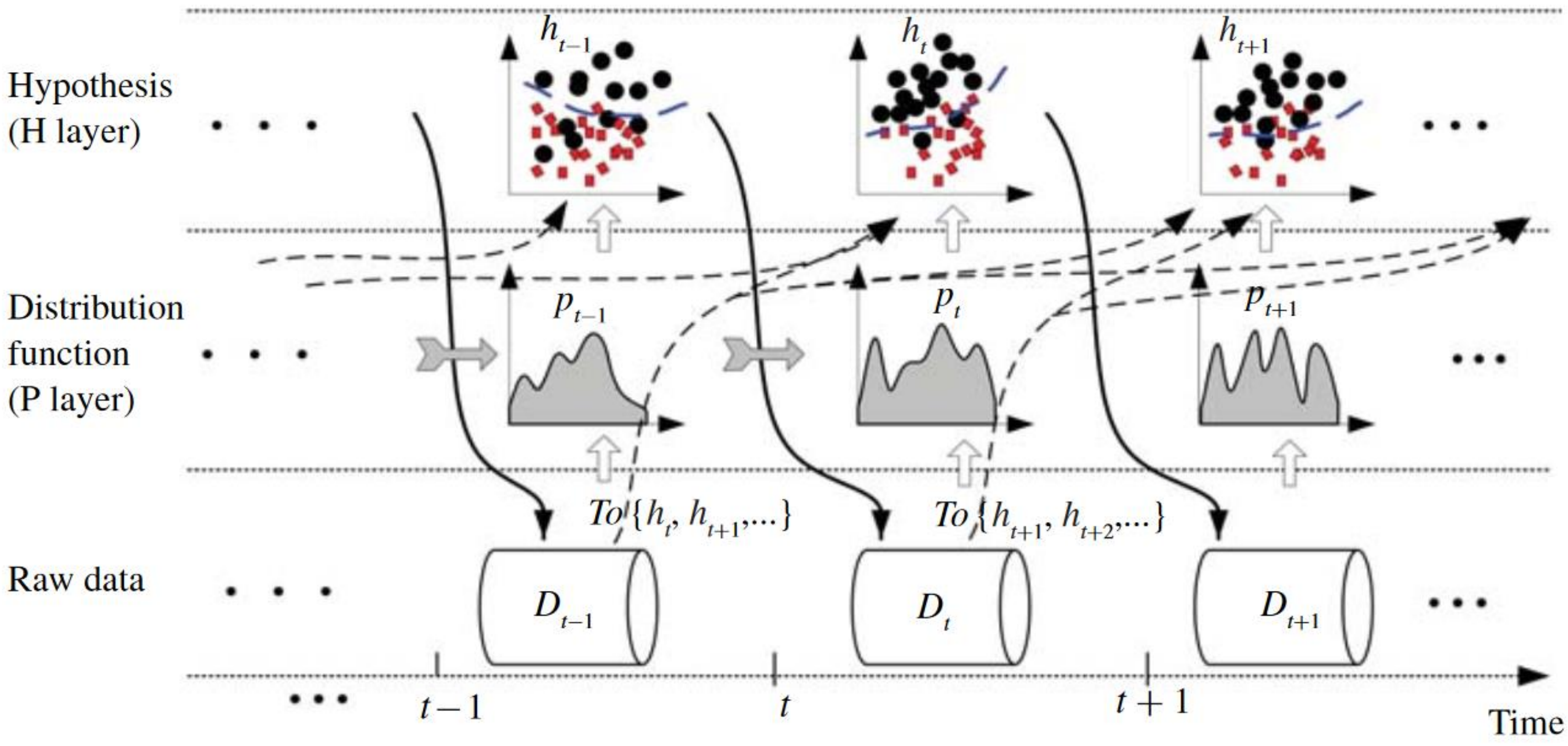





Image credit: Haibo He, Sheng Chen, Kang Li, and Xin Xu (2011)

# Future steps

- Set up a benchmark of classification results (scores) with commonly used ML models over static light curve feature data.
  - For this we are planning to use Redshift Completeness Factor (RCF) real light curve data as a training set.
  - Fit and estimate parameters of transients with appropriate tools (e.g. MOSFIT).
- Adapt the current models for running as part of an incremental learning framework that receives transient alert data as input.
- Compare the results between the original benchmark and running the models over the same training set in a streamed form.
- Integrate the model to the AMPEL system and evaluate its performance.

# References


- Andras A. Benczur, Levente Kocsis, and Robert Palovics. Online machine learning in big data streams. CoRR, abs/1802.05872, 2018.
- Leon Bottou and Yann Le Cun. On-line learning for very large data sets. Applied stochastic models in business and industry, 21(2):137–151, 2005.
- Mohamed Medhat Gaber, Arkady Zaslavsky, and Shonali Krishnaswamy.Mining data streams: A review. SIGMOD Rec., 34(2):18–26, June 2005.
- Haibo He, Sheng Chen, Kang Li, and Xin Xu. Incremental learning from stream data IEEE Transactions on Neural Networks, 22(12):1901–1914,2011.
- James Guillochon, Matt Nicholl, V. Ashley Villar, Brenna Mockler, Gautham Narayan, Kaisey S. Mandel, Edo Berger, Peter K. G. Williams. MOSFiT: Modular Open-Source Fitter for Transients. Oct 5, 2017. 15 pp.
- Anais Möller, Thibault de Boissière. SuperNNova: an open-source framework for Bayesian, Neural Network based supernova classification. Jan 2019. arXiv:1901.06384 [astro-ph.IM]