Mario Pasquato

Hosts: Prof. Michela Mapelli, Raffaele Gratton

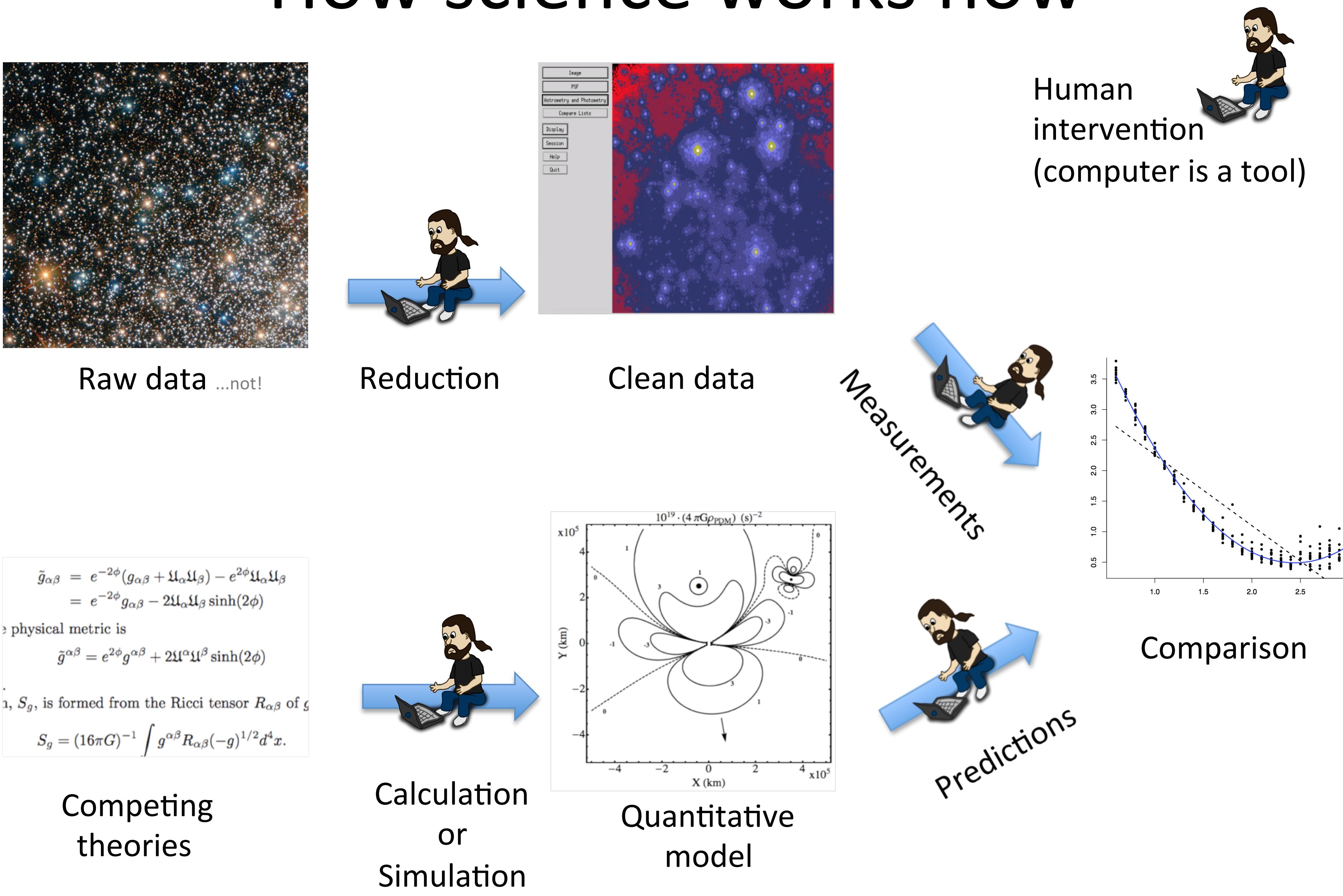
Annual Astrofit Review meeting

Rome Oct. 23-24 2018

ARTIStIC: ARTificial Intelligence Search for Intemediate mass black holes in star Clusters

Astro FIt 2

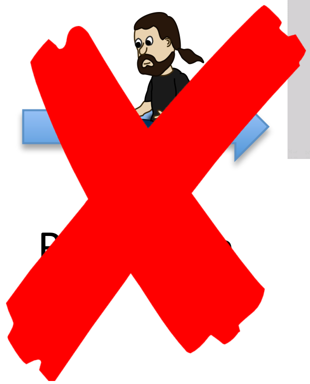
How science works now

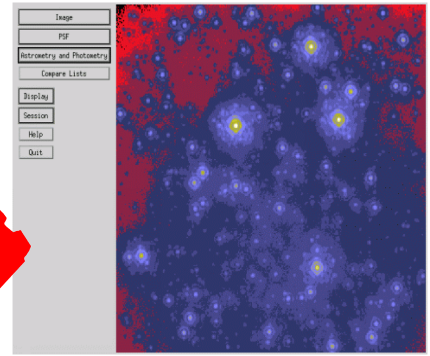


Making it more automatic

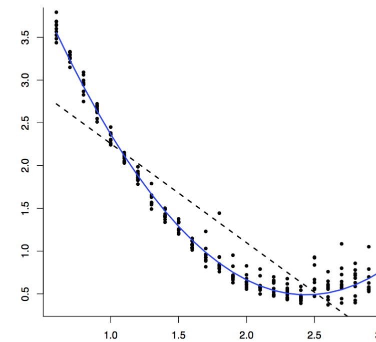
Reduce human intervention

Raw data





Clean data



Comparison

$$\tilde{g}_{\alpha\beta} = e^{-2\phi}(g_{\alpha\beta} + \mathfrak{U}_\alpha \mathfrak{U}_\beta) - e^{2\phi}\mathfrak{U}_\alpha \mathfrak{U}_\beta$$
$$= e^{-2\phi} g_{\alpha\beta} - 2\mathfrak{U}_\alpha \mathfrak{U}_\beta \sinh(2\phi)$$

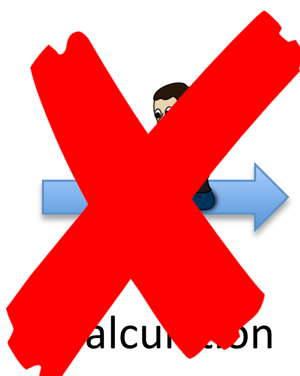
e physical metric is

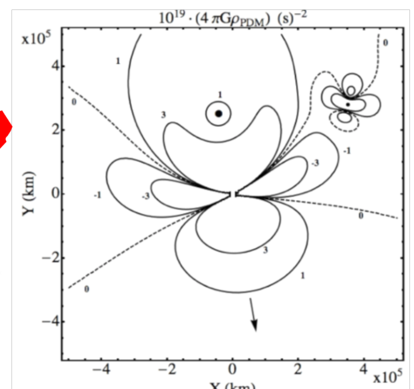
$$\tilde{g}^{\alpha\beta} = e^{2\phi} g^{\alpha\beta} + 2\mathfrak{U}^\alpha \mathfrak{U}^\beta \sinh(2\phi)$$

n, S_g, is formed from the Ricci tensor $R_{\alpha\beta}$ of g

$$S_g = (16\pi G)^{-1} \int g^{\alpha\beta} R_{\alpha\beta} (-g)^{1/2} d^4x.$$

Competing theories

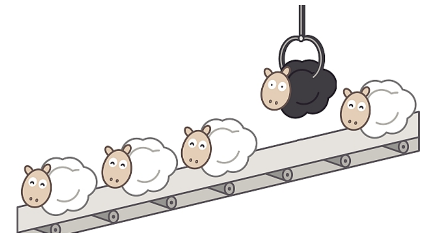




Quantitative model

How it could work

Raw data



Classification

$$\tilde{g}_{\alpha\beta} = e^{-2\phi}(g_{\alpha\beta} + \mathfrak{U}_\alpha \mathfrak{U}_\beta) - e^{2\phi}\mathfrak{U}_\alpha \mathfrak{U}_\beta$$
$$= e^{-2\phi} g_{\alpha\beta} - 2\mathfrak{U}_\alpha \mathfrak{U}_\beta \sinh(2\phi)$$

e physical metric is

$$\tilde{g}^{\alpha\beta} = e^{2\phi} g^{\alpha\beta} + 2\mathfrak{U}^\alpha \mathfrak{U}^\beta \sinh(2\phi)$$

n, S_g, is formed from the Ricci tensor $R_{\alpha\beta}$ of g

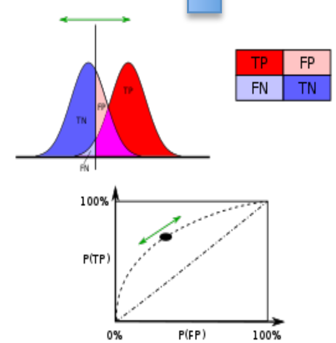
$$S_g = (16\pi G)^{-1} \int g^{\alpha\beta} R_{\alpha\beta} (-g)^{1/2} d^4x.$$

Competing theories

Simulation

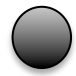
Mock data

Training

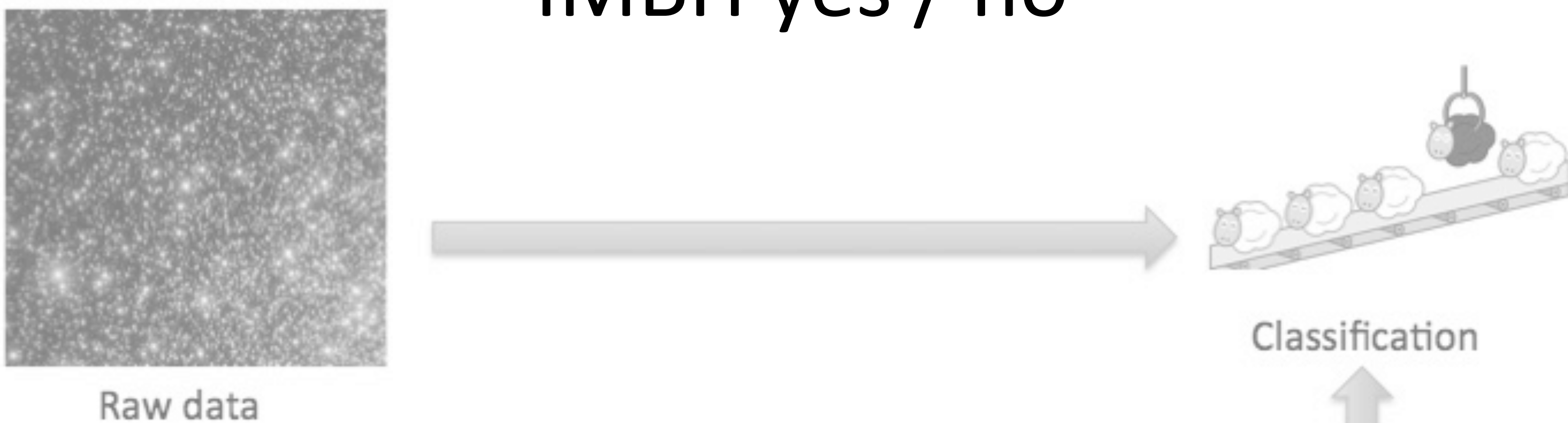


Trained machine learning model (classifier)

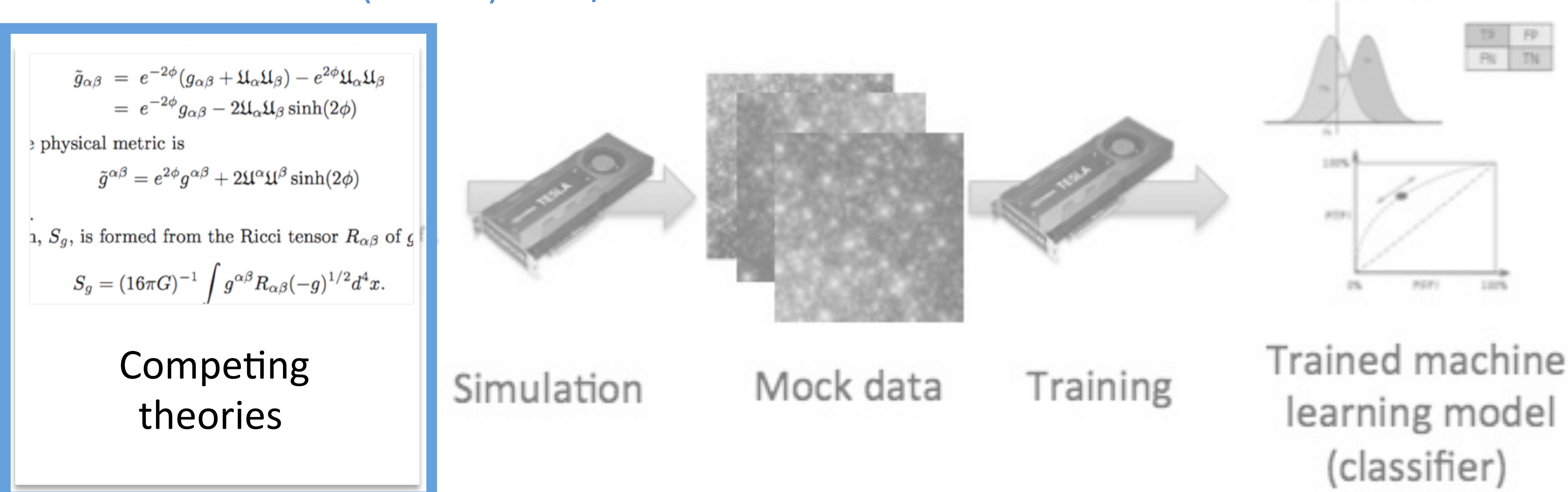
An actual example: my project

- Do Galactic star clusters contain Intermediate Mass Black Holes (IMBHs)? 
- IMBHs = BHs more massive than stellar BHs (> some 10 M_{Sun}) and less massive than Supermassive BHs (< 10^5 M_{Sun})
- Gravitational wave astronomy gave us the first firm detection of such black holes
- Electromagnetic astronomy still did not provide a firm detection

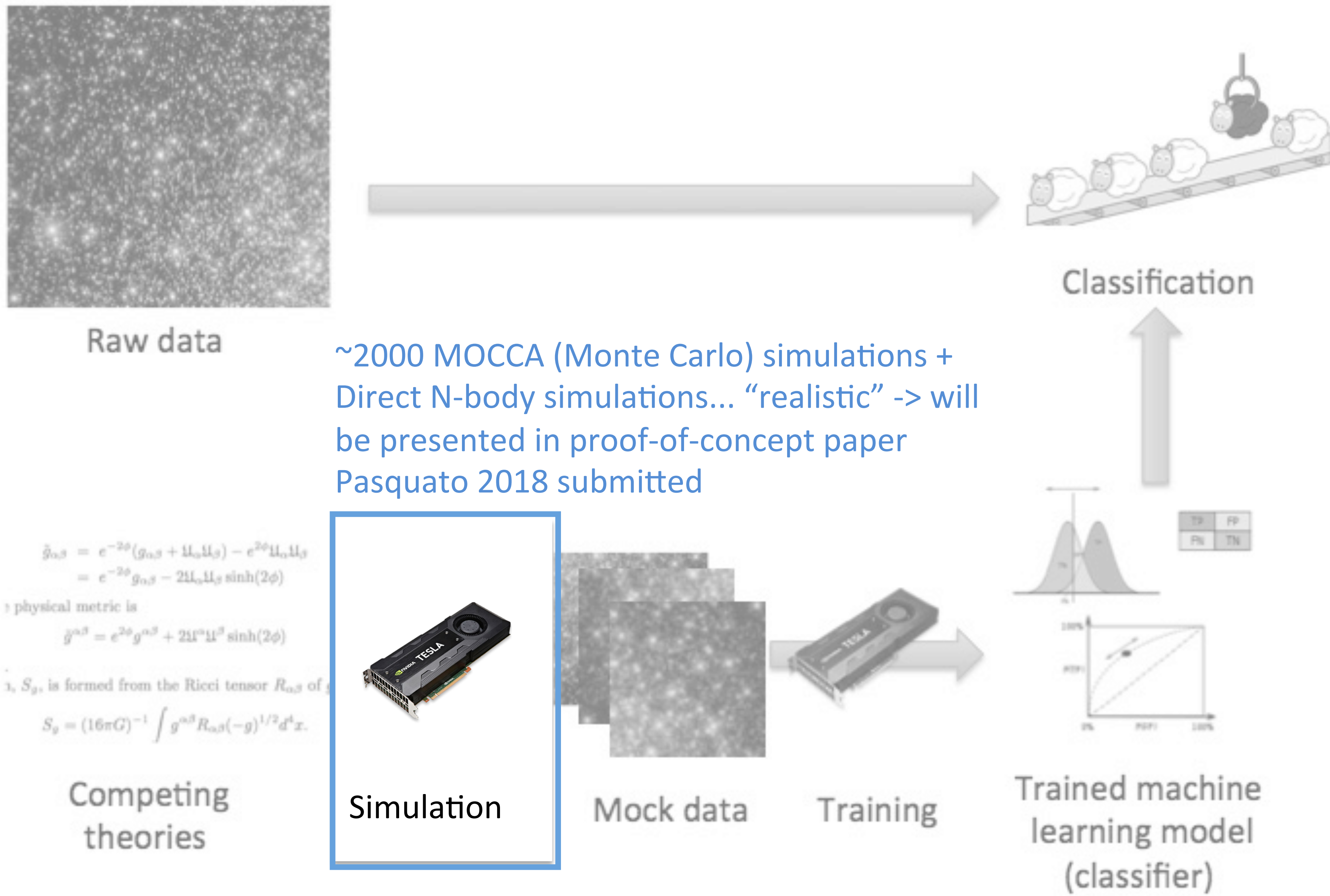
In my project: competing theories = IMBH yes / no



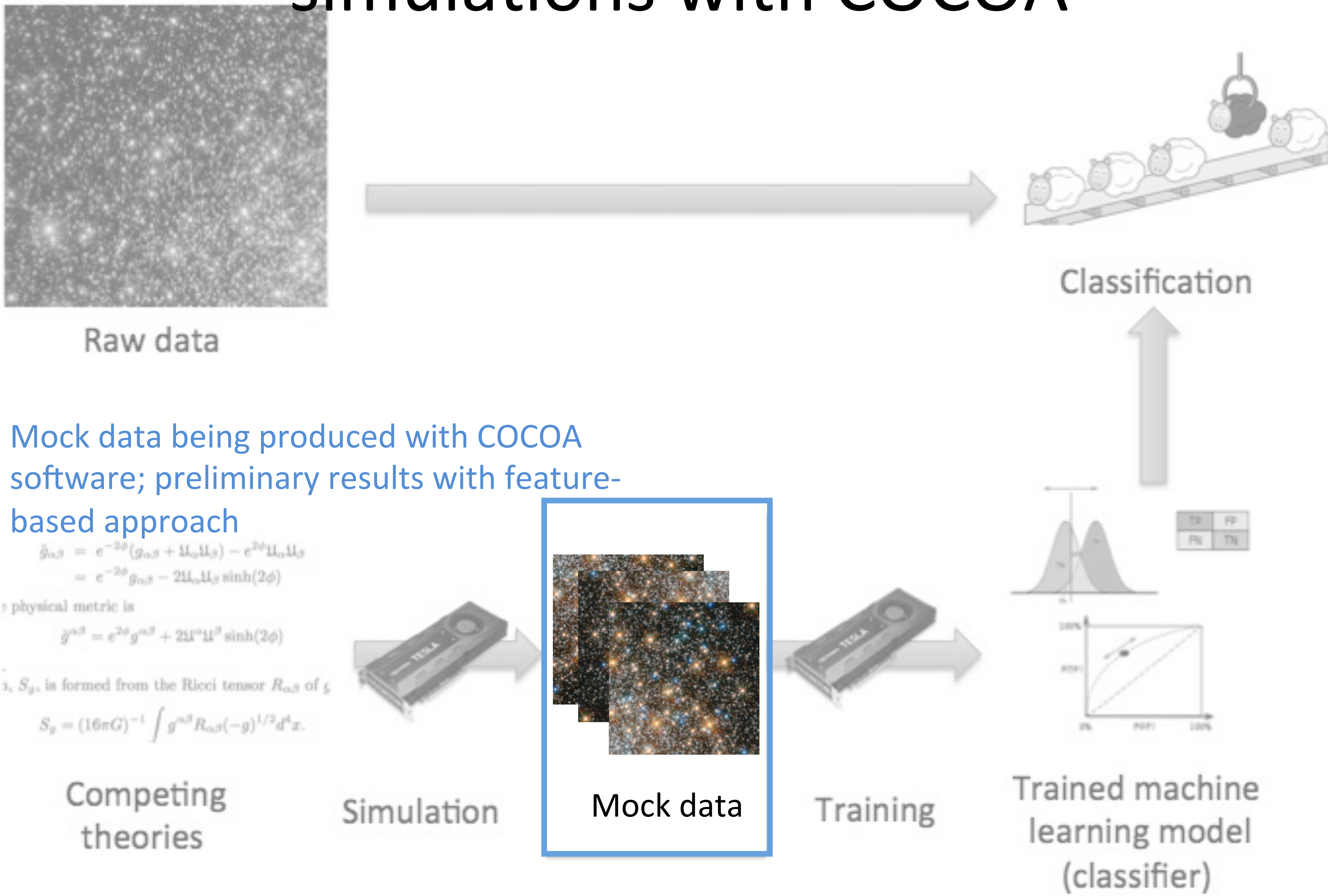
Do Galactic star clusters contain Intermediate Mass Black Holes (IMBHs)? YES/NO



In my project: simulations already fully run

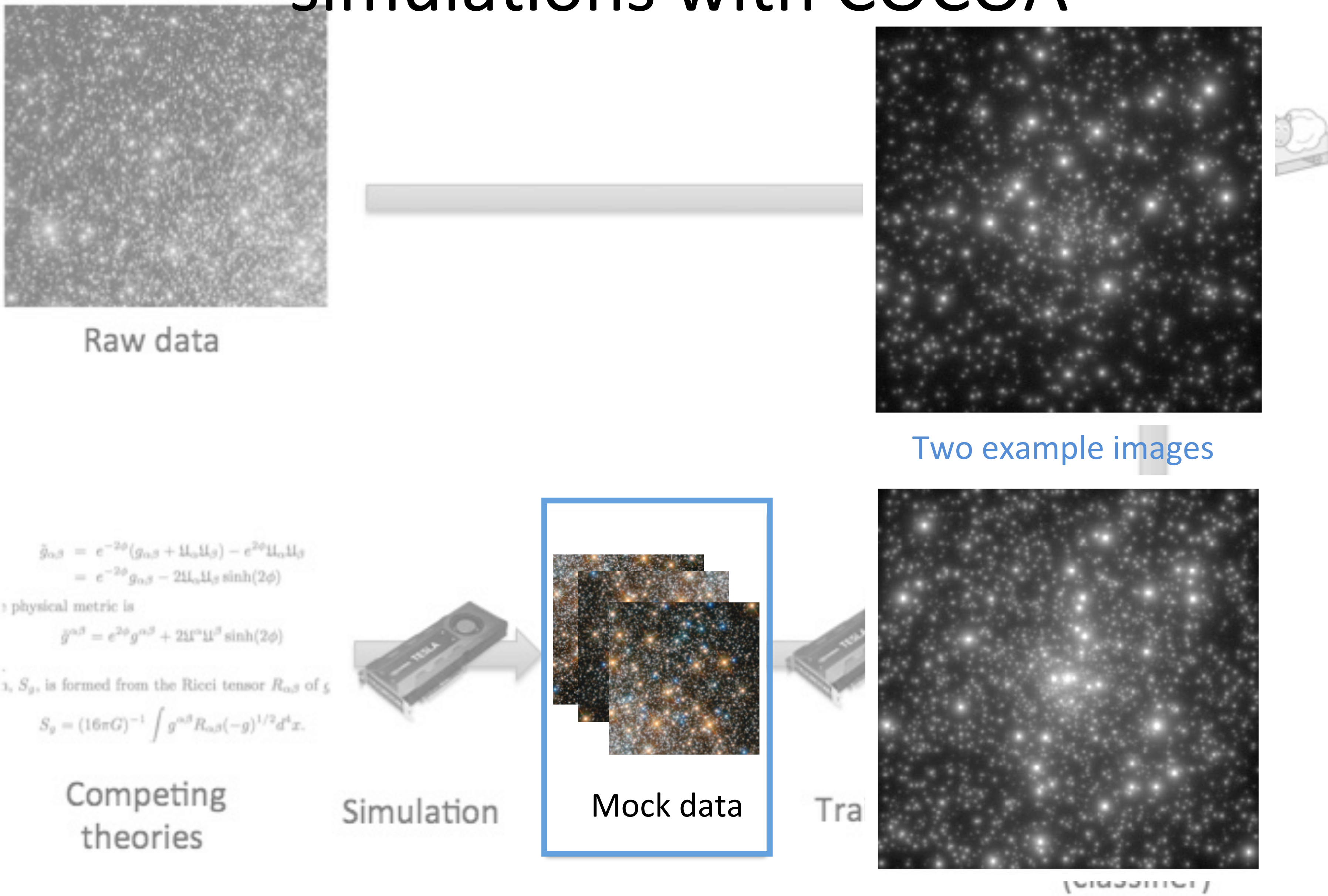


In my project: mock images from simulations with COCOA



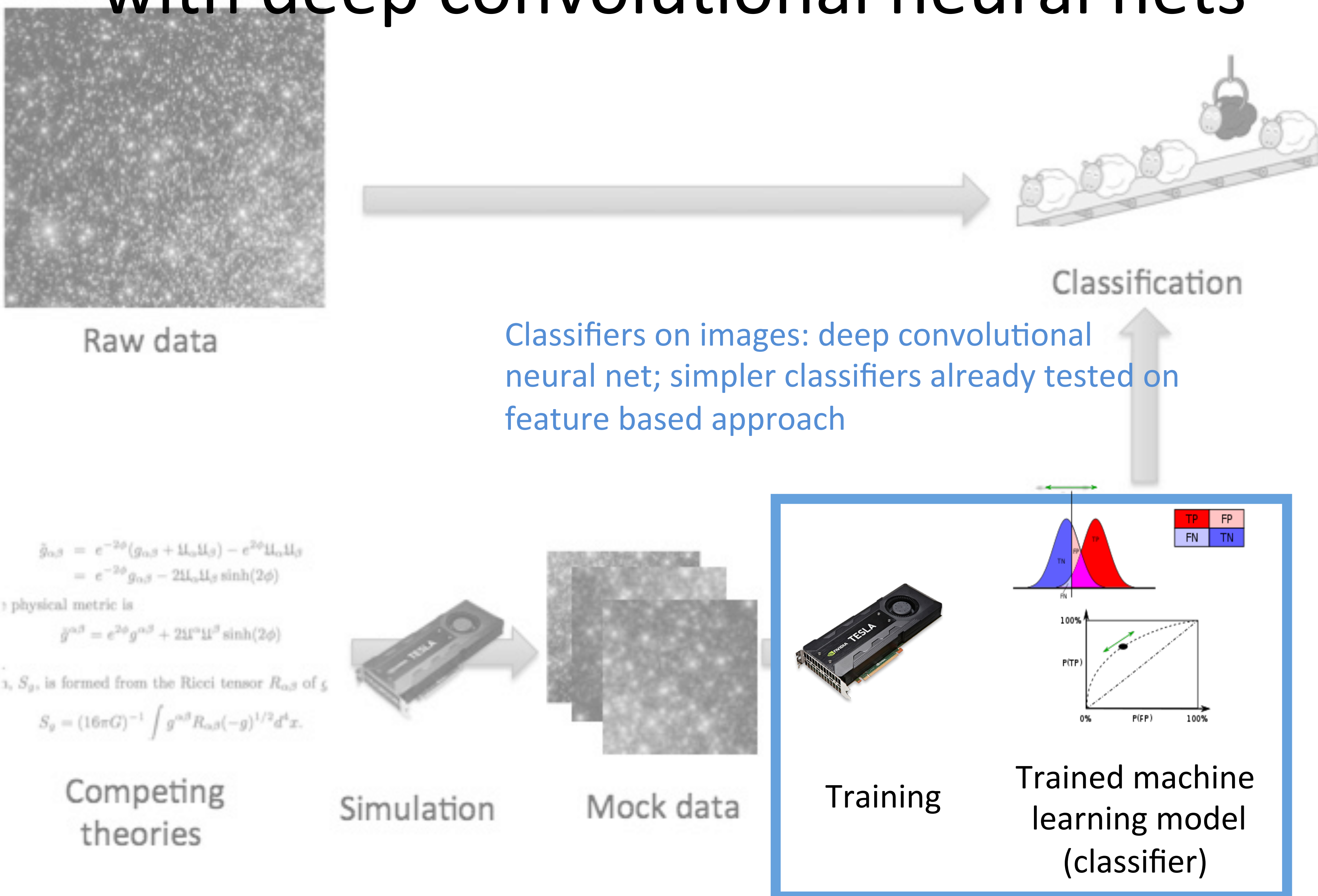
Mock data being produced with COCOA software; preliminary results with feature-based approach

In my project: mock images from simulations with COCOA



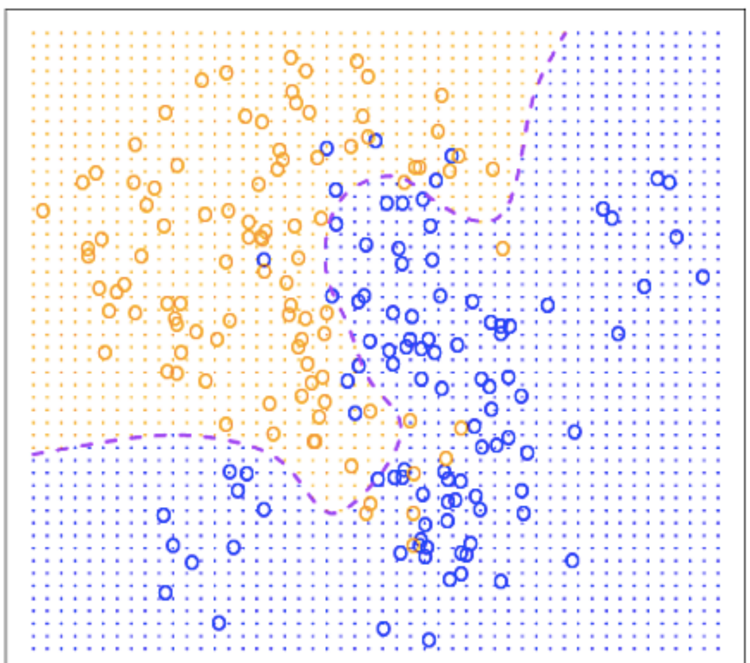
In my project: mock images classified with deep convolutional neural nets

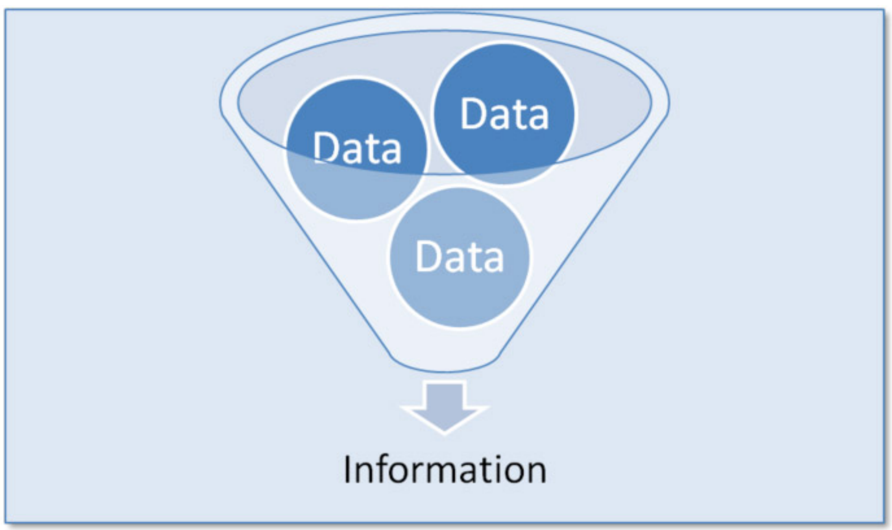
Classifiers on images: deep convolutional neural net; simpler classifiers already tested on feature based approach



By the way, what is machine learning?

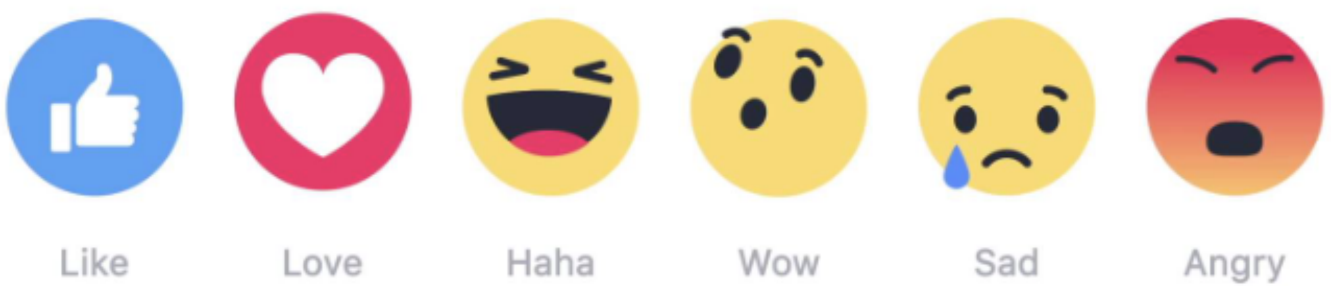
- Machine learning is **teaching computers by example instead of programming them**
- Make a classifier that eats data and spits (IMBH yes / IMBH no)
- This easily generalizes to any data, to any ~~yes/no~~ question





But… everybody is doing this already how is your project new?

- Ever heard of Galaxy Zoo: https://www.zooniverse.org/about/publications
- Wonder why facebook wants you to click on this:



You are hand-labeling data (for free) to train a classifier! (search for "sentiment analysis")

No hand labelling in our case

- We can’t hand label real observations with IMBH/no IMBH because we don’t know
- So we run simulations (where we know, by construction, if IMBH is there or not)
- Train on simulations, predict on observations
- This is what makes my project new

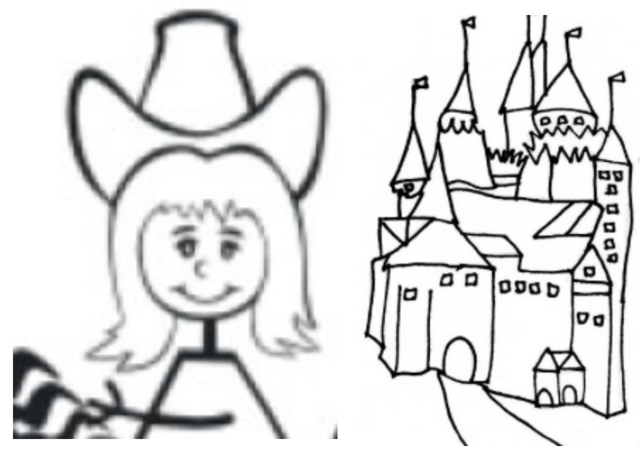
Train/deploy mismatch

- We have no real data with reliable labels (IMBH host / non-host)
- Of course: we do not know which clusters are an IMBH host in the first place
- The classifiers are trained on simulations, where we know who is an IMBH host, by construction
- But they will be deployed on real data

Fake it until you make it

- You only ever see (bad) drawings, but you have to classify real pictures!
- What to do?
- Make the drawings as similar as possible to real pictures: we need good mock observations

Real observation

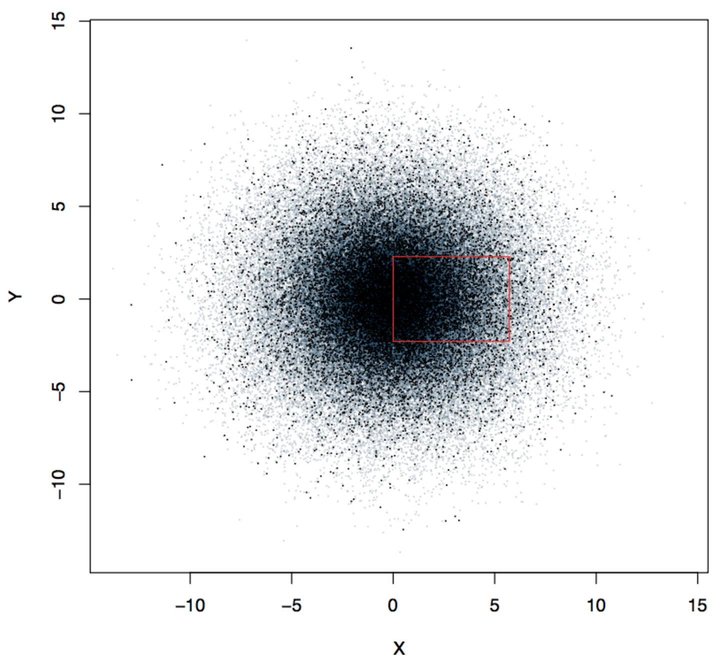


Simulation

Good mock observations

- Two criteria:
  1. A classifier cannot discriminate mocks from real observations based on the same features as the IMBH/noBH classifier
  2. an astronomer (student) cannot discriminate mocks from real observations in a blind test
- 2 is an astronomical Turing test
- I am working on this with a master's student (Mr. Piero Trevisan)

A step back: current results



2000 simulations of star clusters using MOCCA, an all-inclusive Monte Carlo* code (Hypki & Giersz 2013)

MOCCA Survey Database I (Askar et al. 2017)

~1300 simulations evolved to 12Gyr, ~40% produced an IMBH

*a Fokker-Planck code that is solved using Montecarlo techniques

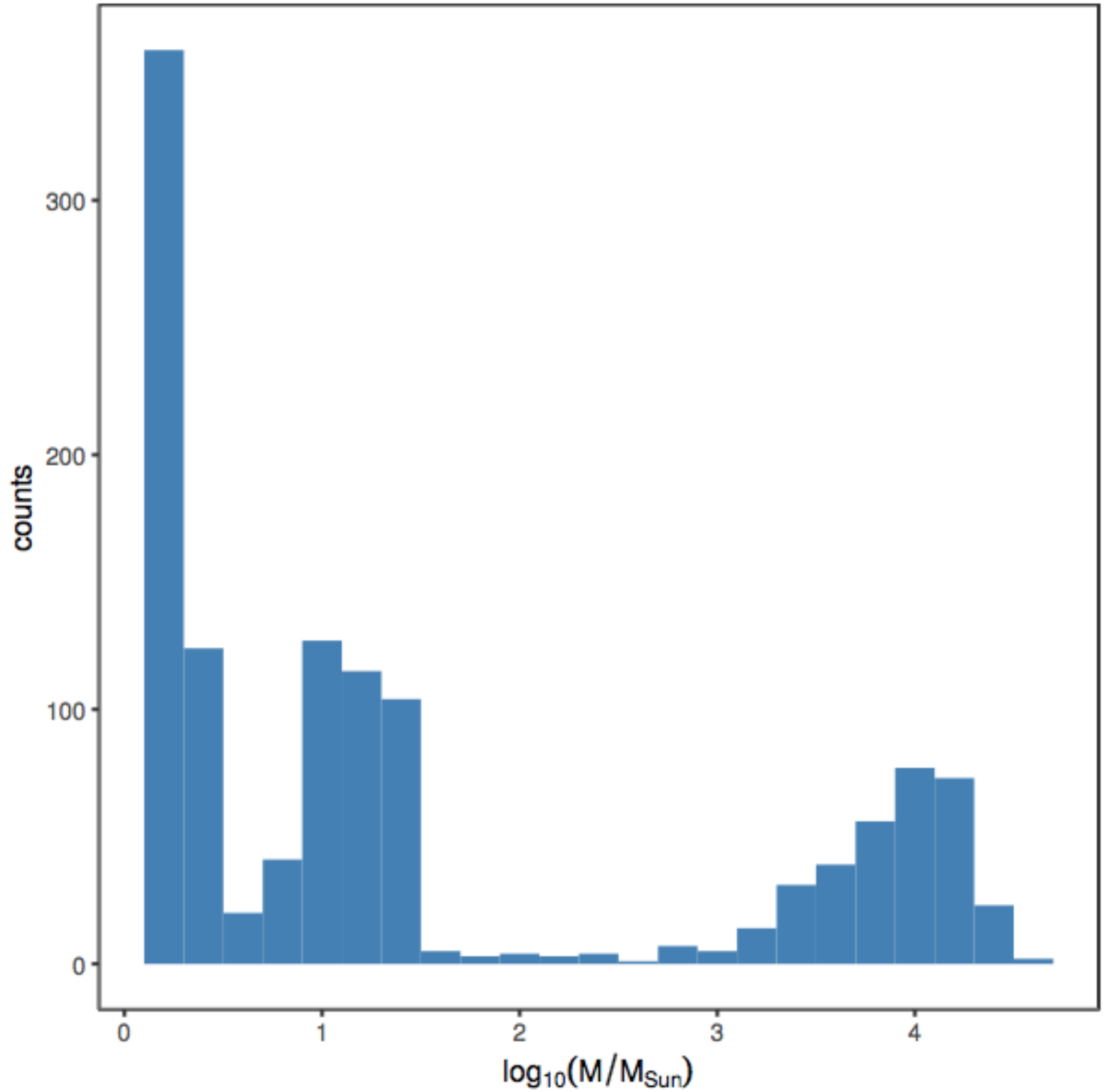


Figure 1. Histogram of the log masses of the heaviest black hole in the MOCCA Survey simulations, from snapshots taken at 12Gyr. The mass function is clearly multimodal, and stellar-mass BH range is well separated from the IMBH range, as there are few BH with mass in the 10^2-10^3 range.

- Log mass of the heaviest BH in each simulation at 12Gyr is bimodal
- IMBHs are well separated from stellar-mass BHs -> classification into IMBH host / not host is justified

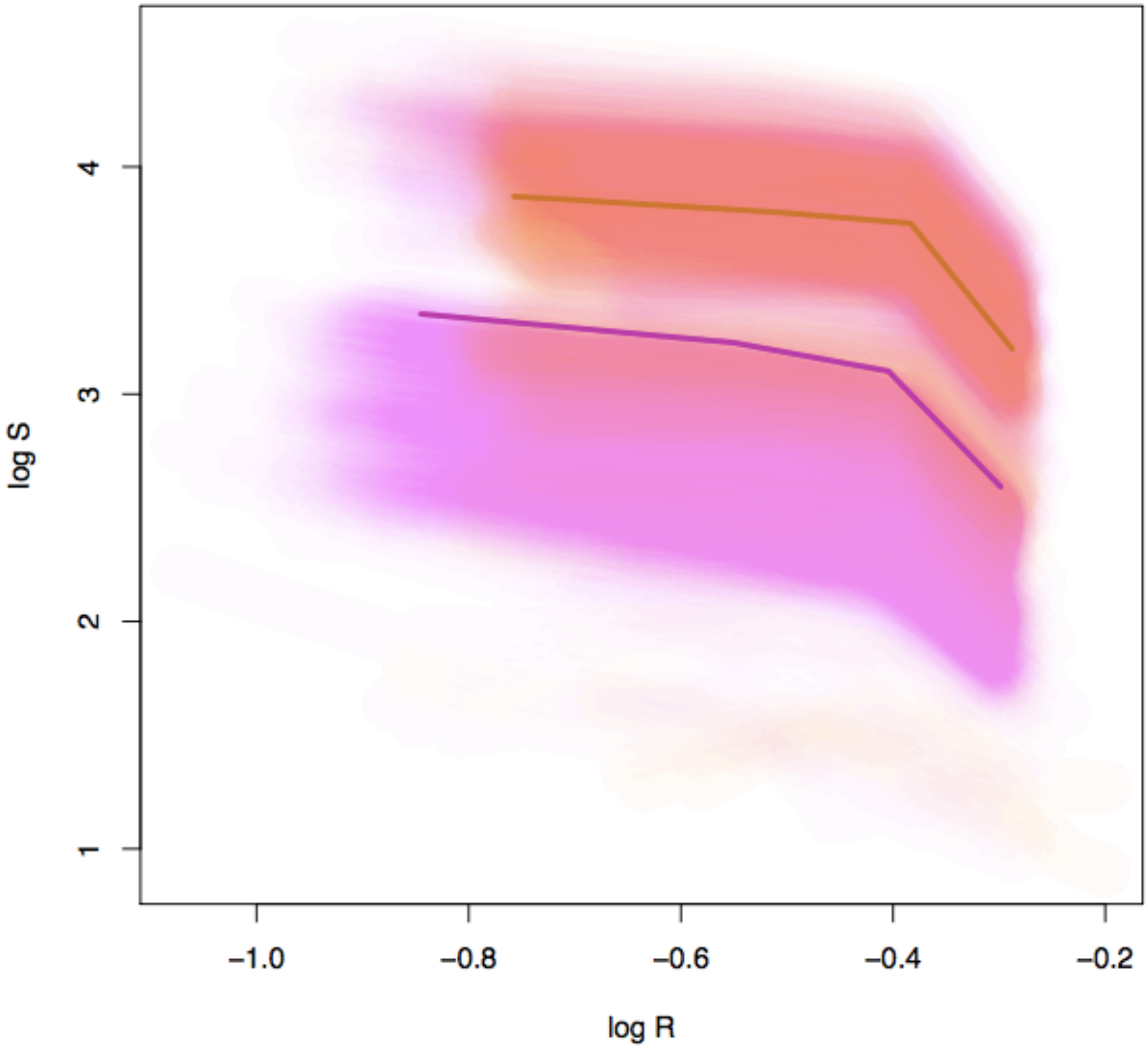
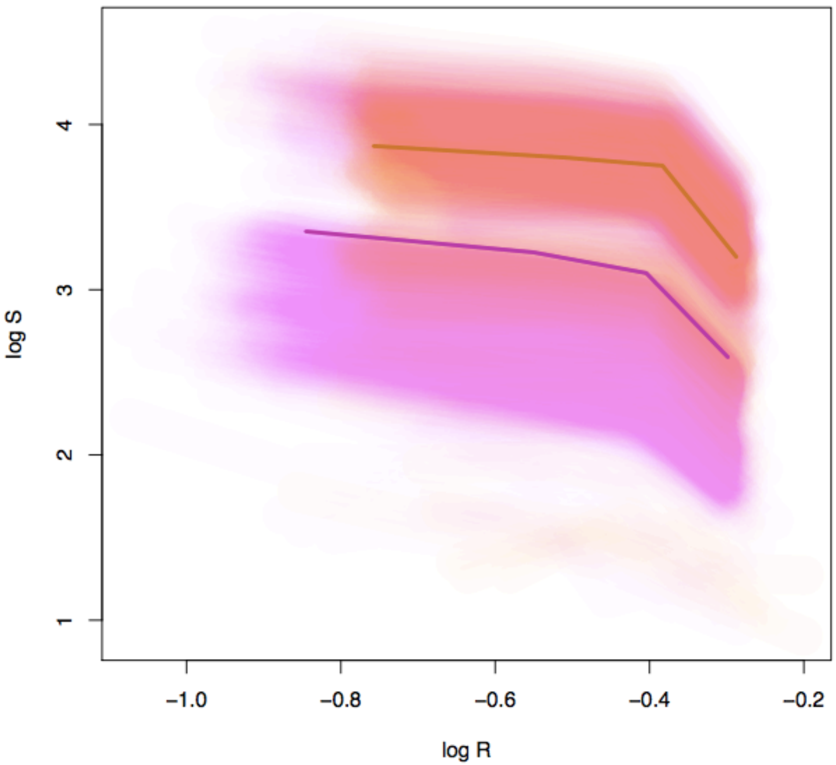


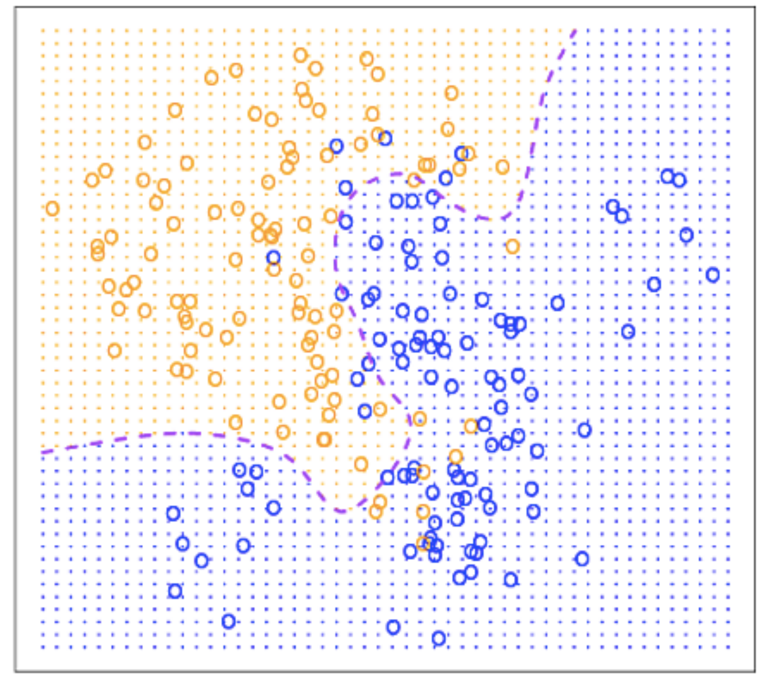
Figure 2. Number density profiles of all simulation snapshots, in log-log scale, shown as shaded areas based on kernel density estimation. IMBH hosts are shown in ochre, non-hosts in pink. Solid lines are actual profiles corresponding to the two medoids of the IMBH host and non-host group.

- Mass density profile at quantiles of the radius (within FOV, for selected stars)
- 4D feature space: 2 bin mids + 2 densities

Feature space



=



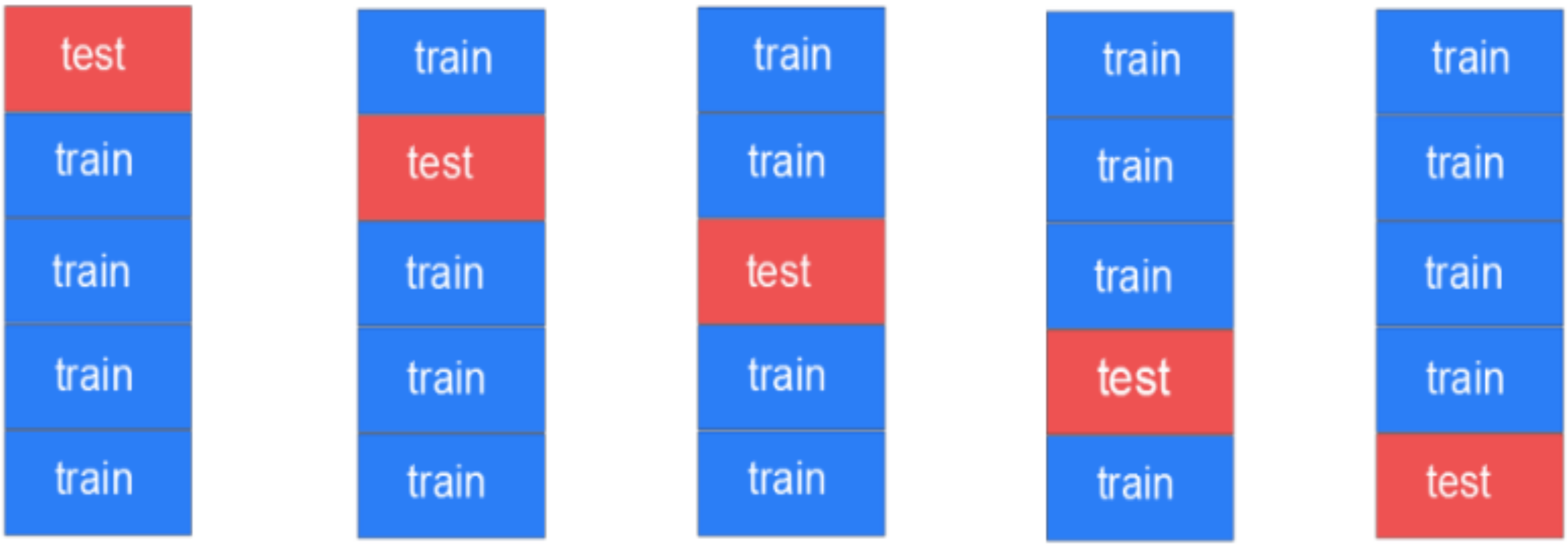
We need to find a boundary between orange (IMBH) and blue (NO IMBH) points... remember?

The simple example (right) is 2-D, the real one is 4-D

Learning how to learn

- Selected algorithms (R libraries)
  - k-nn (FNN): a point is orange if the majority of its k-nearest neighbors in feature space is orange;
  - svm (e1071): the feature space is transformed into a much-higher dimensional space, where it is easy to linearly separate the orange points from the blue; then the separating hyperplane is transformed back;
  - decision trees (party): the feature space is recursively partitioned along one of its 20 dimensions so that each split improves the division between blue and orange;
  - neural net (h2o / keras in python not in R): the hottest algorithm, aka deep learning. I run out of space to explain it.

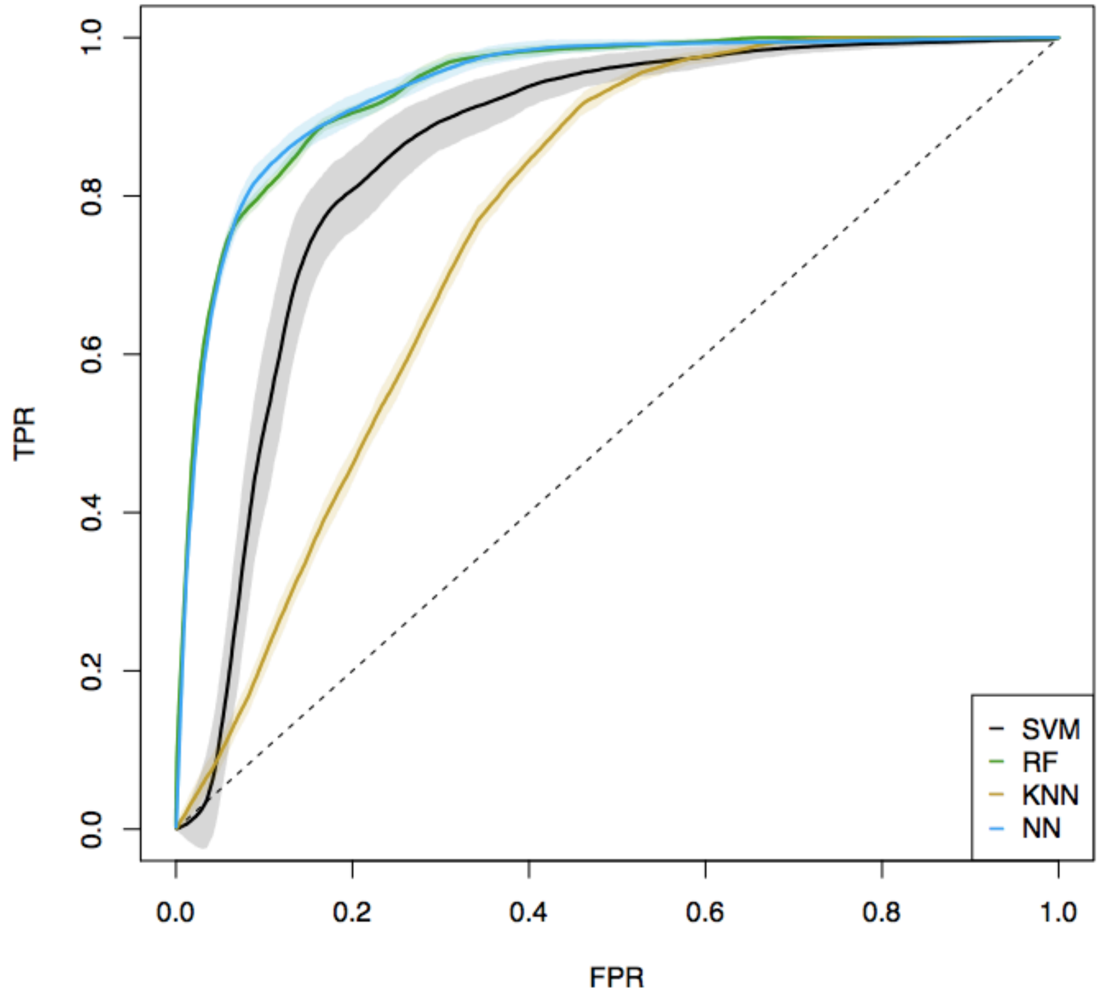
Measuring performance with cross validation



- Evaluate the trained model on unseen data, but still use all data we have (here data = mock observations)
- Split data, use subset for training and complement for testing
- Loop over data (here five times, 5-"fold" CV)

Comparing ROCs

- knn (k = 7)
- svm
- random forest
- fully connected neural net
- ROC curve from 5-fold CV, 2-sigma CA from 100 iterations
- Who performs better? Can you tell?

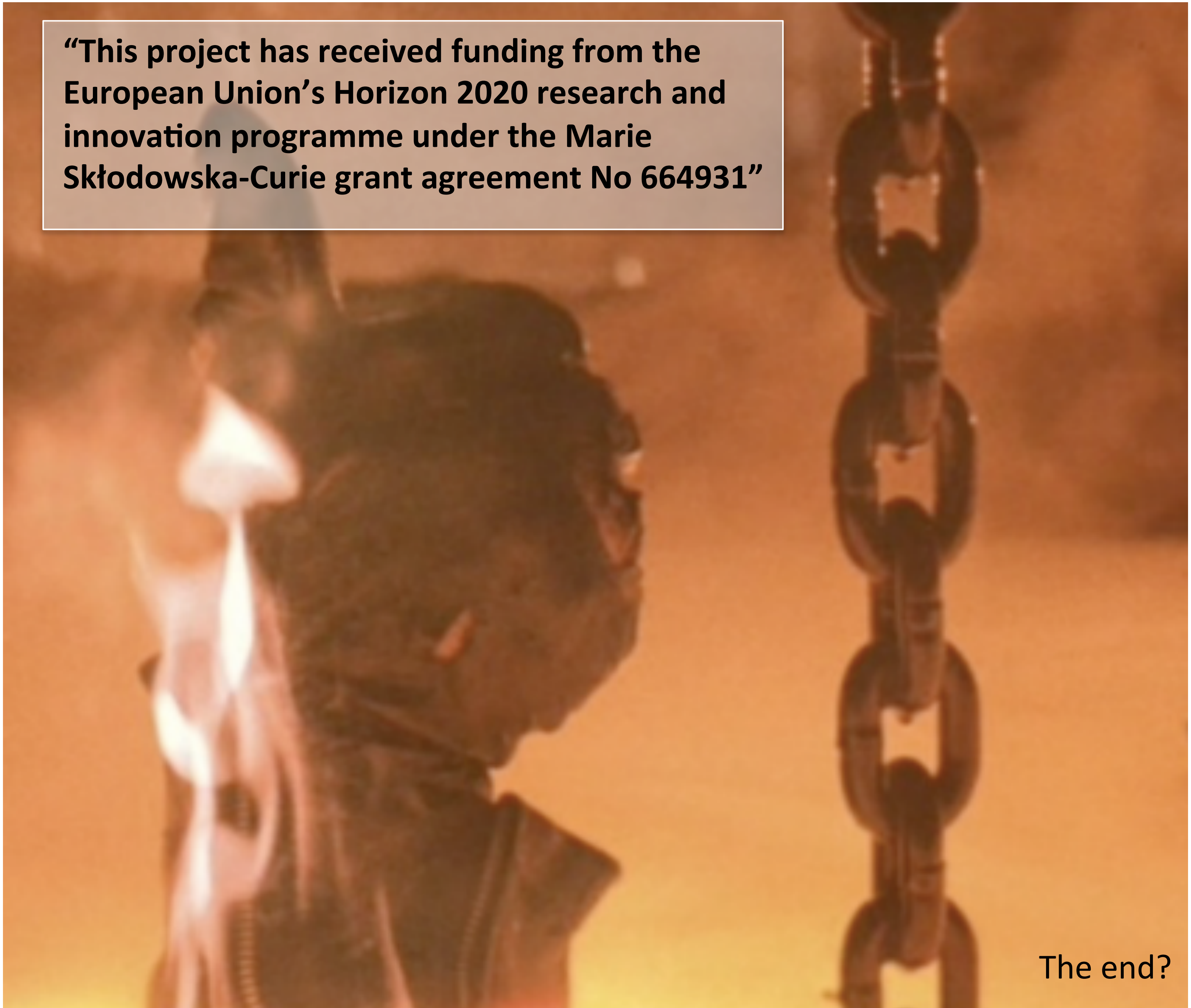


Best classifiers (neural net, random forest) achieve

> 80% true positive rate

at **10%** false positive rate

and this is **NOT SO BAD**


“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 664931”
The end?

Not the end yet

- Conferences, talks, teaching:
  - Santorini MODEST meeting (June 2018, contrib. talk)
  - Innsbruck University (Apr. 2018, invited seminar)
  - Held 12-hour course for Ph.D. students in Padova univ. on machine learning for astronomy
  - Supervising a master's student (Mr. Piero Trevisan), supervised a bachelor's student (Ms. Erica Greco)
- Achievements:
  - ISCRA C project at Cineca approved (200'000 core-hours on DAVIDE GPU cluster)
  - Submitted Pasquato et al. 2018 to MNRAS, proof of concept paper, 1st referee report
  - 3 bonus papers (unrelated projects): Ballone, Mapelli & Pasquato 2018 MNRAS 480, 4684; Pasquato, Miocchi & Yoon 2018 ApJ accepted; de Angelis et al. JHEA 19, 1

Weighing the IMBH candidate CO-0.40-0.22* in the Galactic Centre

A. Ballone,[1★] M. Mapelli[1,2,3] and M. Pasquato[1]

[1] *INAF, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[2] *Institute für Astro- und Teilchen Physik, Universität Innsbruck, Technikerstrasse 25/8, A-6020 Innsbruck, Austria*
[3] *INFN, Milano Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*

Accepted 2018 August 3. Received 2018 July 26; in original form 2018 April 27

ABSTRACT

The high velocity gradient observed in the compact cloud CO-0.40-0.22, at a projected distance of 60 pc from the centre of the Milky Way, has led its discoverers to identify the closeby mm continuum emitter, CO-0.40-0.22*, with an intermediate mass black hole (IMBH) candidate. We describe the interaction between CO-0.40-0.22 and the IMBH, by means of a simple analytical model and of hydrodynamical simulations. Through such calculation, we obtain a lower limit to the mass of CO-0.40-0.22* of few $10^4 \times M_{\odot}$. This result tends to exclude the formation of such massive black hole in the proximity of the Galactic Centre. On the other hand, CO-0.40-0.22* might have been brought to such distances in cosmological time-scales, if it was born in a dark matter halo or globular cluster around the Milky Way.

Key words: black hole physics – ISM: clouds – Galaxy: centre.

BLUE STRAGGLER BIMODALITY: A BROWNIAN MOTION MODEL

MARIO PASQUATO
INAF, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, 35122 Padova, Italy

PAOLO MIOCCHI
Istituto dei Sistemi Complessi, CNR, P.le A. Moro 2, 00185 - Roma (Italy)

SUK-JIN YOON
Department of Astronomy & Center for Galaxy Evolution Research, Yonsei University, Seoul 120-749, Republic of Korea
Draft version September 25, 2018

ABSTRACT

The shape of the radial distribution of Blue Straggler Stars (BSS), when normalized to a reference population of Horizontal Branch (HB) stars, has been found to be a powerful indicator of the dynamical evolution reached by a Globular Cluster (GC). In particular, observations suggest that the BSS distribution bimodality is modulated by the dynamical age of the host GC, with dynamically unrelaxed GCs showing a flat BSS distribution, and more relaxed GCs showing a minimum at a radius that increases for increasing dynamical age, resulting in a natural "dynamical clock". While direct N-body simulations are able to reproduce the general trend, thus supporting its dynamical origin, the migration of the minimum of the distribution appears to be noisy and not well defined. Here we show that a simple **unidimensional** model based on dynamical friction (drift) and Brownian motion (diffusion) correctly reproduces the qualitative motion of the minimum, without adjustable parameters except for the BSS to HB stars mass-ratio. Differential dynamical friction effects combine with diffusion in creating a bimodality in the BSS distribution and determining its evolution, driving the migration of the minimum to larger radii over time. The diffusion coefficient is strongly constrained by the need to reproduce the migratory behaviour of the minimum, and the radial dependence of diffusion set by fundamental physical arguments automatically satisfies this constraint. Therefore, our model appears to capture the fluctuation-dissipation dynamics that underpins the dynamical clock.

Subject headings: (Galaxy:) globular clusters: general

Detecting IMBHs with machine learning: feature-based supervised classification - I. Density profile features in MOCCA-SURVEY Database I simulations

Mario Pasquato,[1★] Michela Mapelli,[2,1] Mirek Giersz,[3] Abbas Askar[4,3]
[1] *INAF, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, 35122 Padova, Italy*
[2] *Institut für Astro- und Teilchenphysik, Universität Innsbruck, Technikerstr. 25/8, A-6020 Innsbruck, Austria*
[3] *Nicolaus Copernicus Astronomical Centre, Polish Academy of Sciences, ul. Bartycka 18, PL-00-716 Warsaw, Poland*
[4] *Lund Observatory, Department of Astronomy, and Theoretical Physics, Lund University, Box 43, SE-221 00 Lund, Sweden*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
The gap in mass between supermassive black holes and stellar black holes spans more than three orders of magnitude. Intermediate-mass black holes (IMBHs) would bridge this gap, but their very existence is still controversial. Optimizing the usage of data from electromagnetic observations to complement GW observations probing this range of masses is crucial for the development of multimessenger astronomy. However, as of now we cannot pinpoint the conditions under which a dense stellar system in the local universe is likely to be an IMBH host, due to shortcomings of the usual model-based approach to indirect IMBH detection. Here, for the first time, we apply a data-driven approach to this problem, based on Machine Learning (ML). ML techniques saw increased use in astronomy because they optimise the extraction of useful information from incomplete data, without the need for explicit interpretation based on physical models. Using an extensive sample of Monte Carlo simulations of star clusters based on the MOCCA code, the MOCCA-SURVEY Database, we show that machine learning algorithms can reliably distinguish simulations that host an IMBH from those that do not. We use a feature-based supervised classification approach. We train classifiers using different ML algorithms on a set of numeric features (independent variables) calculated from each simulation snapshot, for which the status as IMBH host is known by construction. Predictive accuracy on new simulations is measured via cross-validation. We quantify the performance of classifiers using the area under the receiver operating characteristic curve. This is the first step towards classifying actual observations, after which we plan to characterize hosts *a posteriori*.