CBPF
Centro Brasileiro de Pesquisas Físicas
UNIDADE DE PESQUISA DO MCTI

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

AI Applications in Astrophysics and Developments with the O&G Industry

Marcelo Albuquerque and **Clécio R. De Bom**

DATA SCIENCE IN (ASTRO)PARTICLE PHYSICS and COSMOLOGY: the BRIDGE to INDUSTRY

27-30 JUNHO 2022, Universidade de Coimbra, Coimbra, Portugal

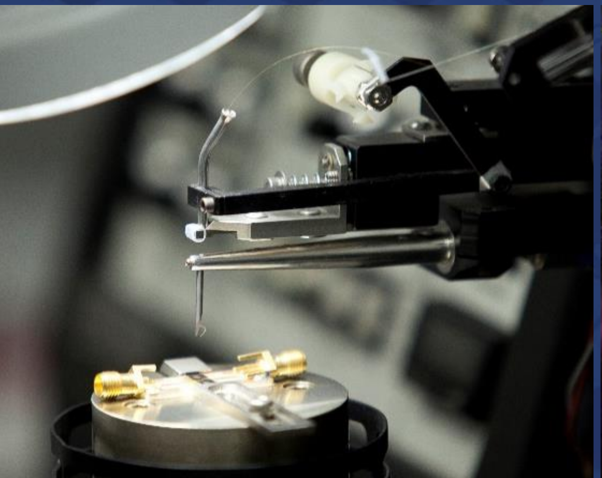
Um evento que mostra o potencial da ciência de dados na sociedade moderna e estimula sinergias entre investigação fundamental e indústria

LIP

CBPF – R&D

Experimental, Theoretical and Applied

- High Energy Physics and Astroparticles
- Materials and condensed matter
- Nanoscience and Nanotechnology
- Biophysics and Biomaterials
- Statistical Mechanics and Complex Systems
- Information and Quantum Computing
- Cosmology and Gravitation
- Signal Processing and Artificial Intelligence
- Scientific Instrumentation






Astrophysics and Deep Learning

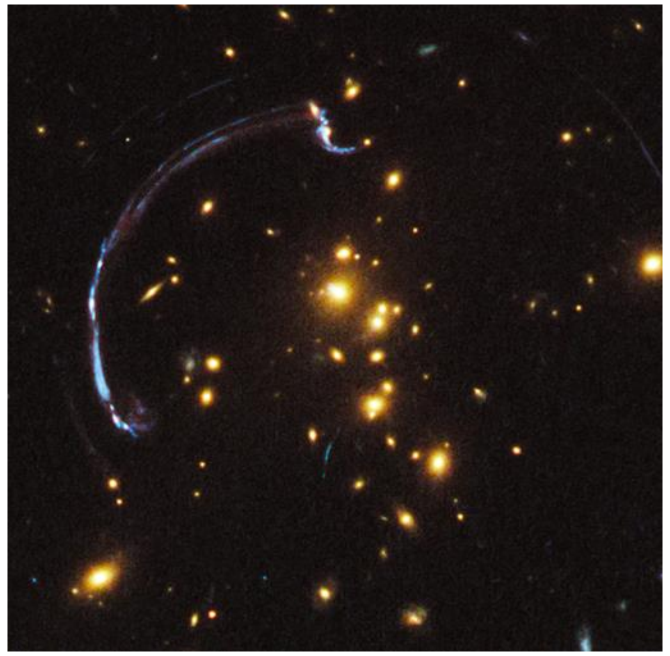
Astrophysics presents several challenges in terms of data Volume, Variety, Velocity (3 Vs in Big Data).

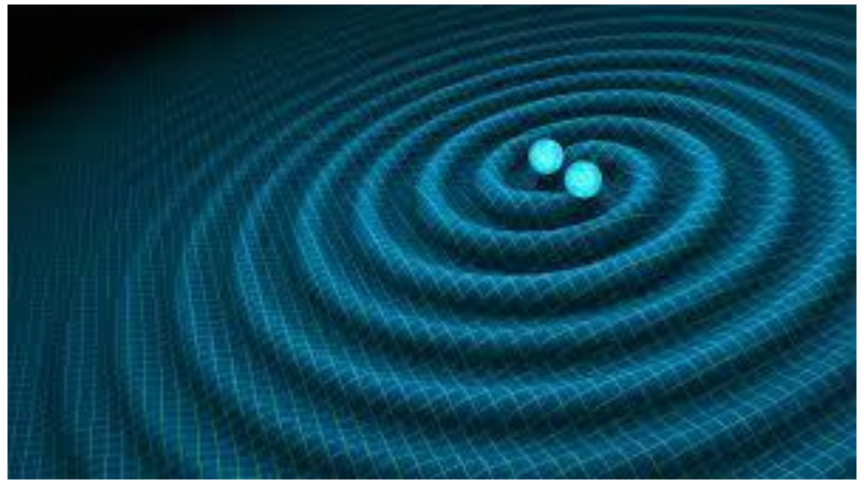
It also comprises a set of heterogenous and incomplete (complex) datasets.

Machine Learning (and in particular the expertise of our group Deep Learning) can act in several situations such as:

Search, Classification, Modelling.

Such tasks are also present in industry challenges!!!





Strong Lensing Search

The Challenge:

Classify 100k images using up to four channels (VIS, NISP J, Y and H – Euclid-like).

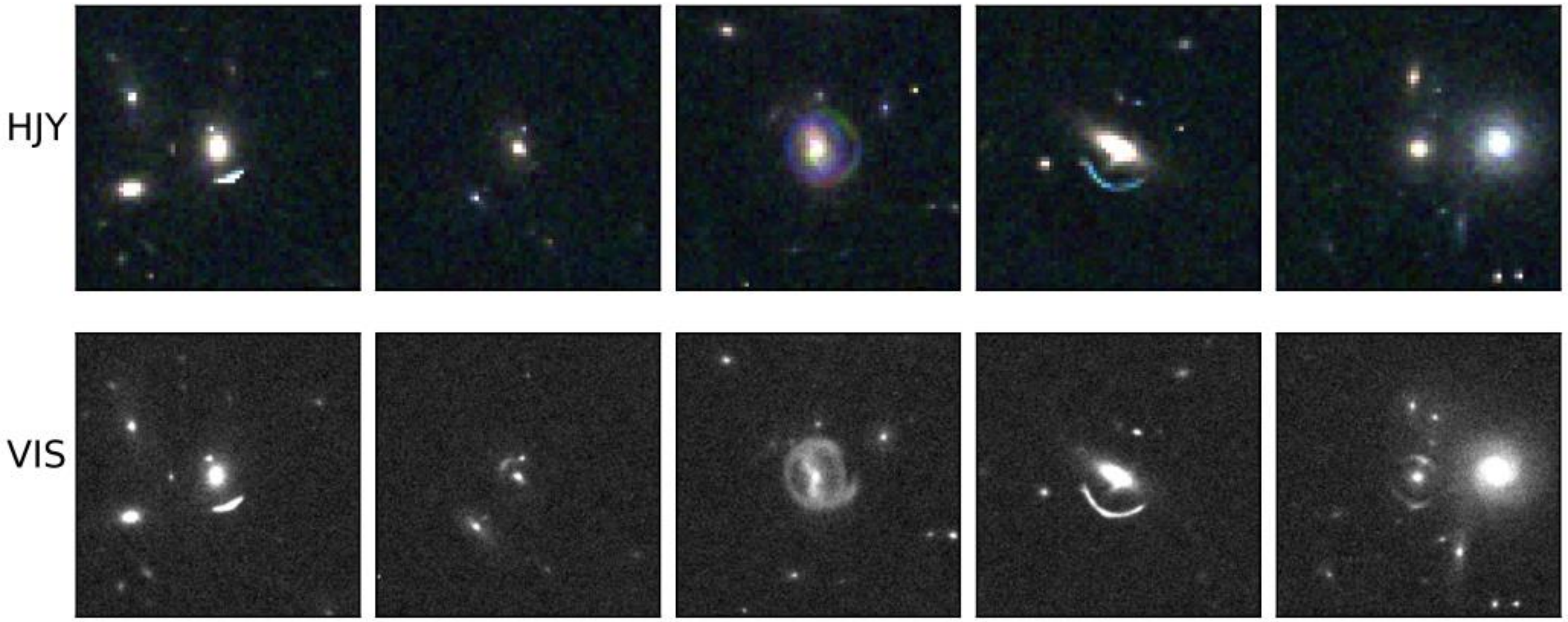
To test the algorithm we have 100k simulated images which contains all sorts of problems in the imaging system.

There was no information on how the images were simulated.

Each team developed different algorithms, mostly based in Deep Learning.

Why we need to find Strong Lenses?

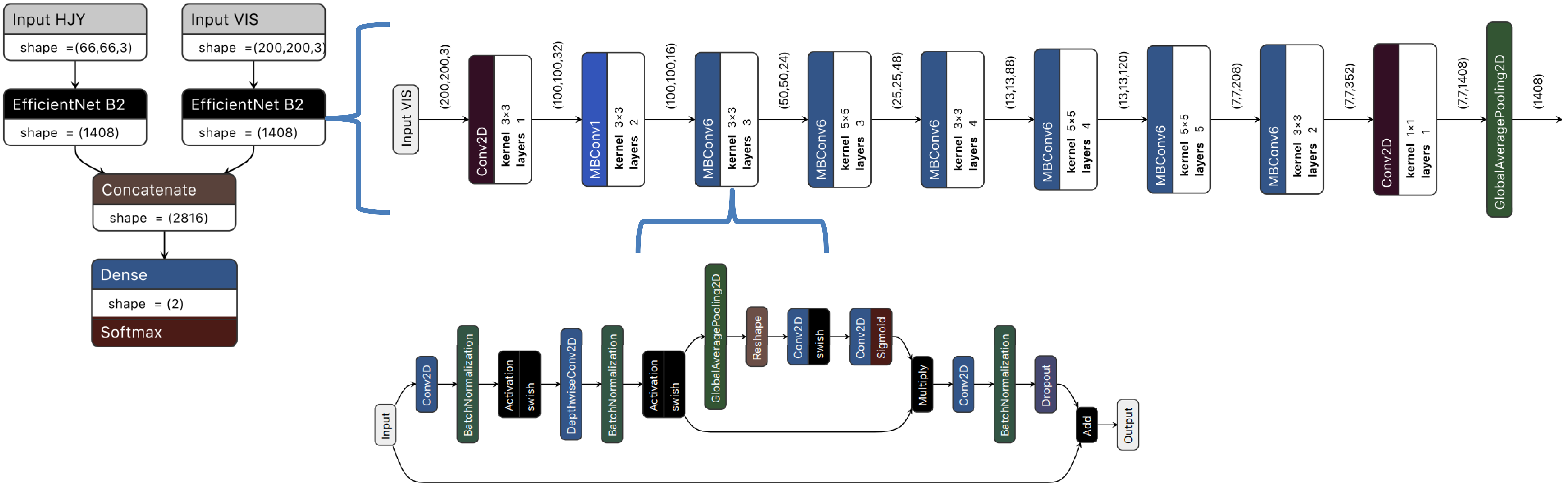
- (Dark) Matter distribution in inner cluster regions
- Gravitational telescopes to investigate faint galaxies at high redshift
- Einsten General Relativity Tests
- Cosmological probe



Strong Lensing Search

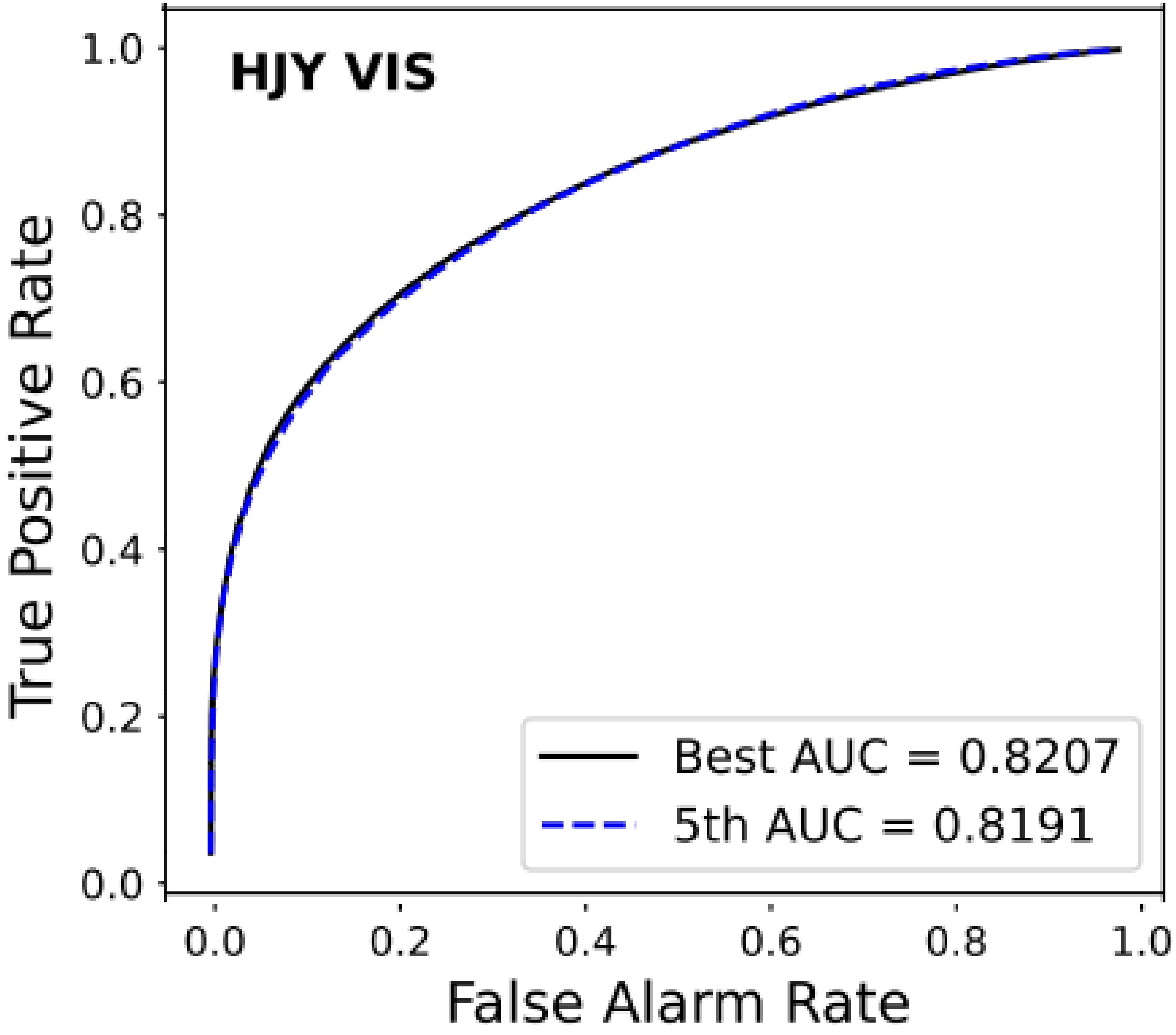
Our solution integrates images with diferente resolutions (VIS band with higher resolution and HJY with lower resolution).

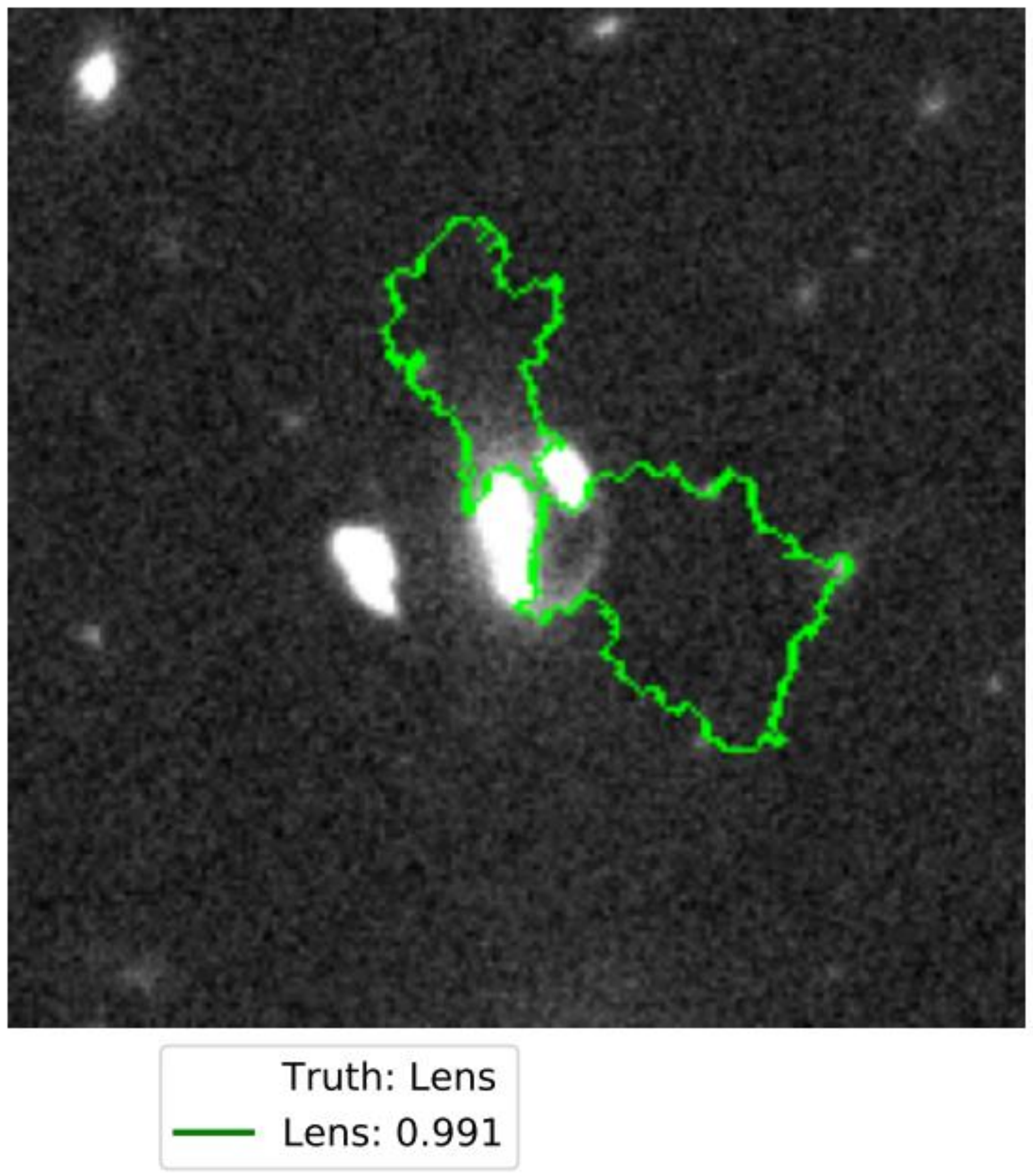
We concate them inside a state-of-art scalabe Deep Neural Network algorithm for computer vision, EfficientNet



Strong Lensing Search

Our solution integrates images with different resolutions (VIS band with higher resolution and HJY with lower resolution).
We concate them inside a state-of-art scalabe Deep Neural Network algorithm for computer vision, EfficientNet.



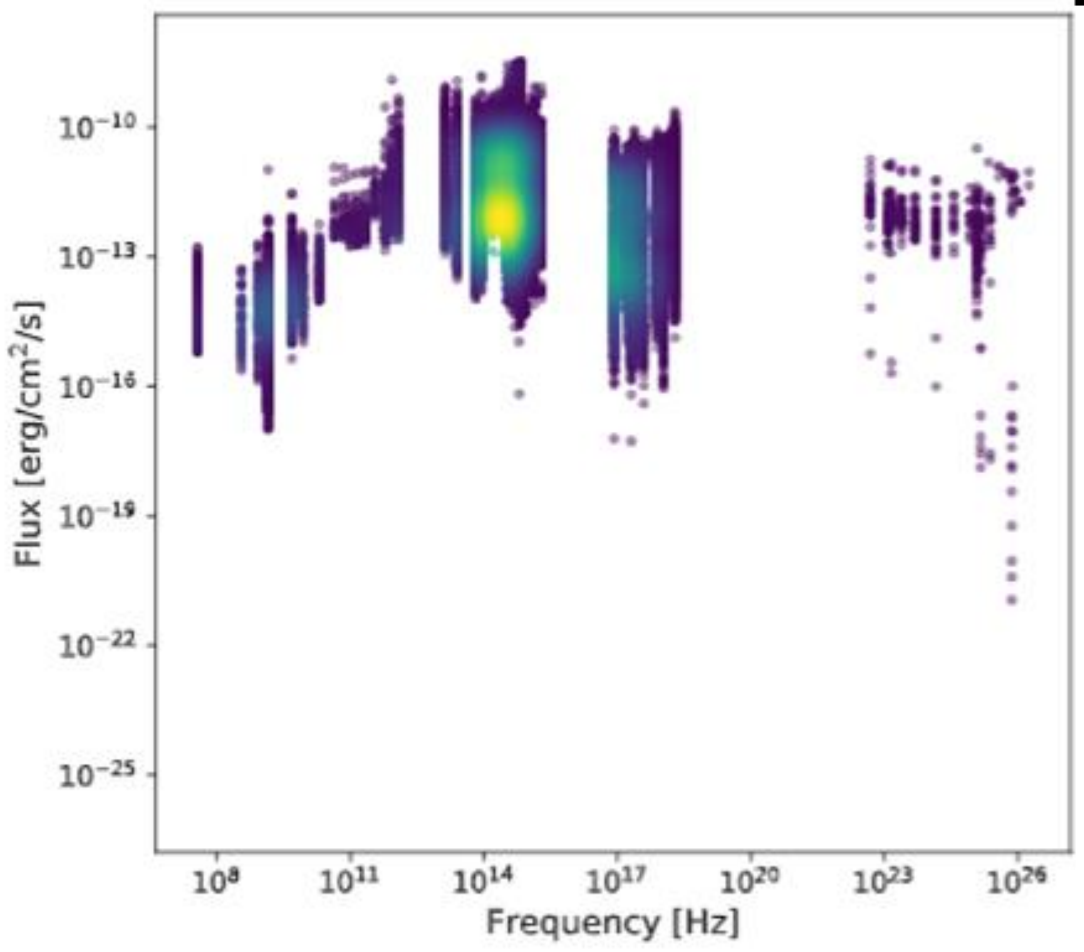


Bom et al; MNRAS; 2022

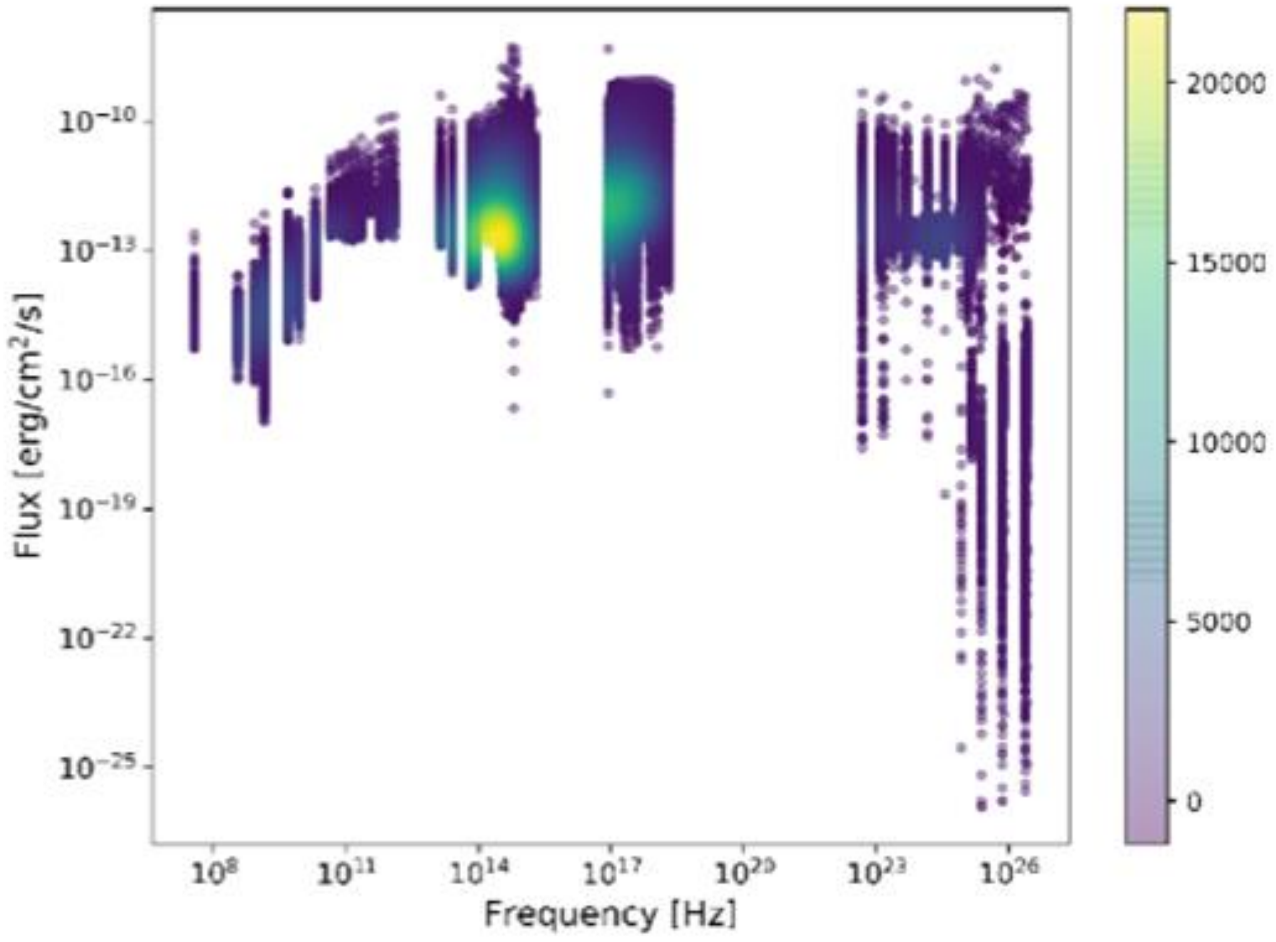
Blazar/AGN Classification

A blazar is an active galactic nucleus (AGN) with a relativistic jet directed very nearly towards an observer. Its standard identification is a manual procedure, usually depends upon heterogeneous multiwavelength coverage.
This is a sequence (curve) classification (flux x frequency), no images. From radio to gamma-rays.

Artist's impression



(a) Non-Blazars.

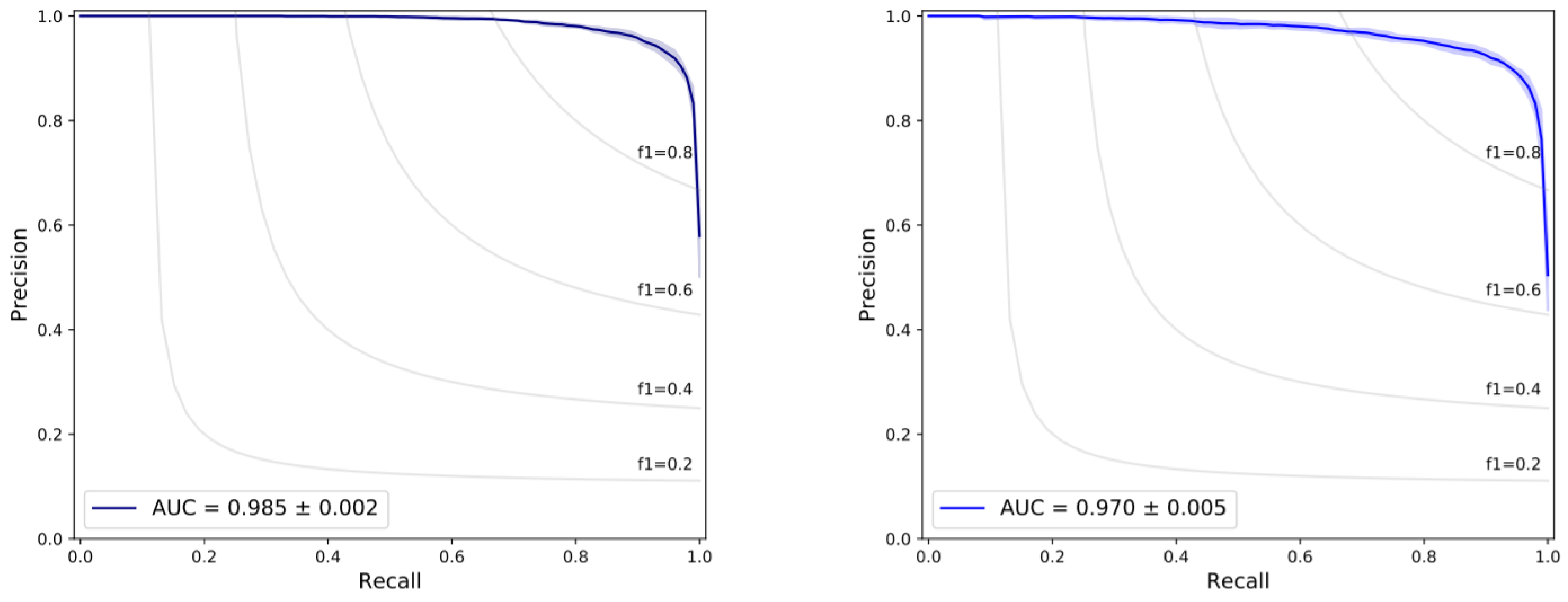


(b) Blazars

Fraga et al.; "***Deep learning Blazar classification based on multifrequency spectral energy distribution data***", MNRAS; 2021

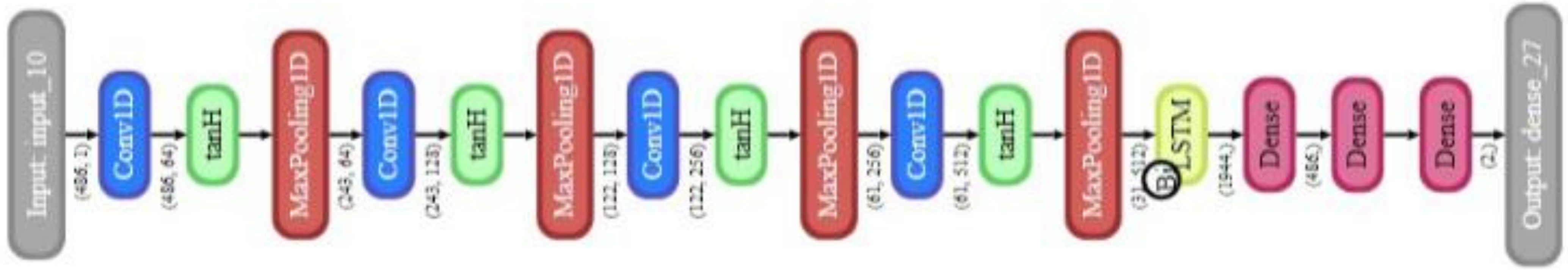
Blazar/AGN Classification

Our solution in this curve classification problem, was to use networks that works on sequences, like the ones used in text translation; a recurrent neural network. More specifically a Long Short-Term Memory (LSTM).



(a) With γ-ray data.

(b) No γ-ray data.

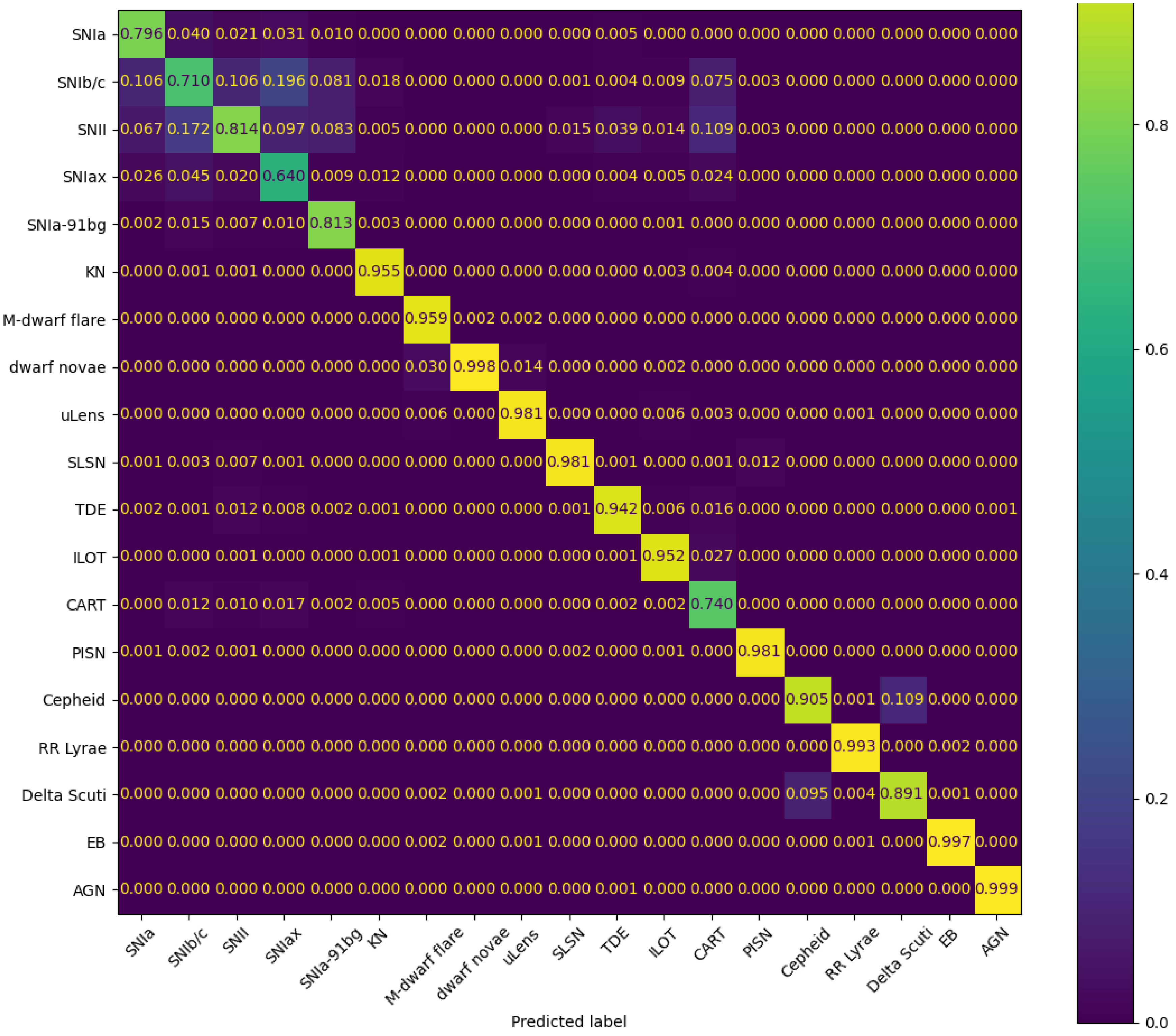


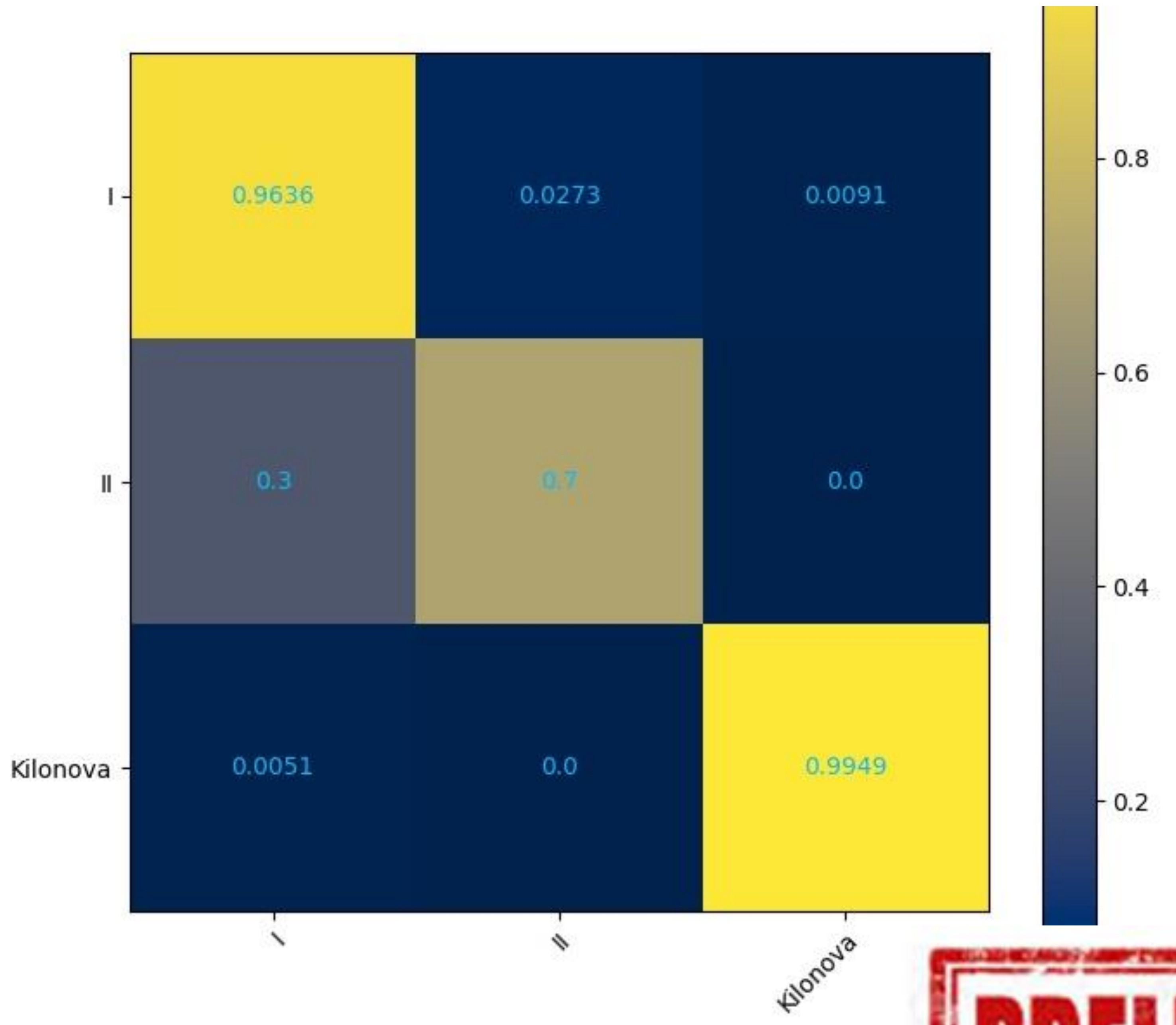
Fraga et al.; "***Deep learning Blazar classification based on multifrequency spectral energy distribution data***", MNRAS; 2021

Transient Classification from LSST

(*) Legacy Survey of Space and Time

We now are using several Deep Learning solutions to classify not SED (flux x frequency) but also optical light curves (magnitude x days) from Vera Rubin Alerts. Including AGN variability and objects that might be the sources of gravitational wave events, like Kilonovas

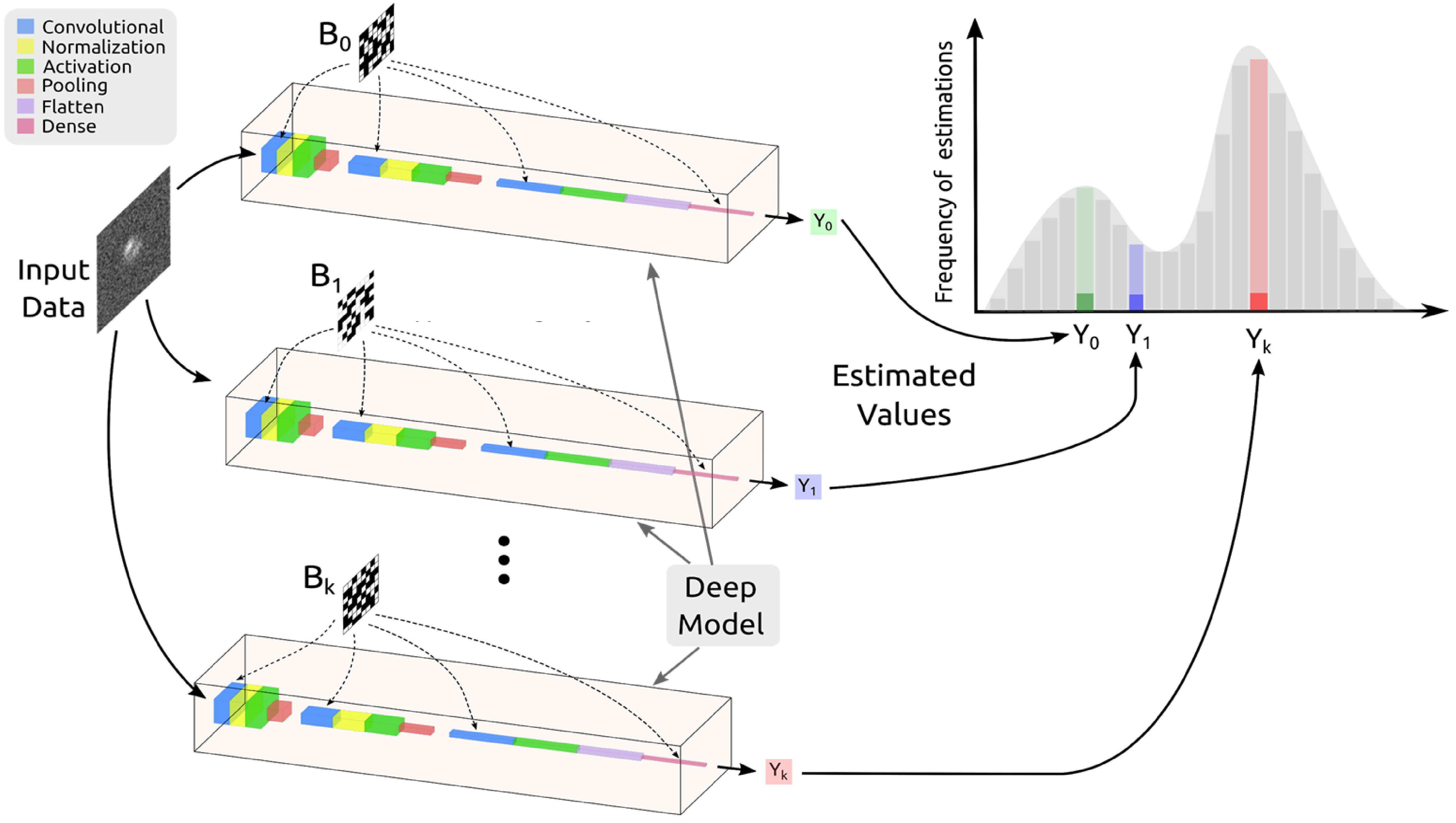


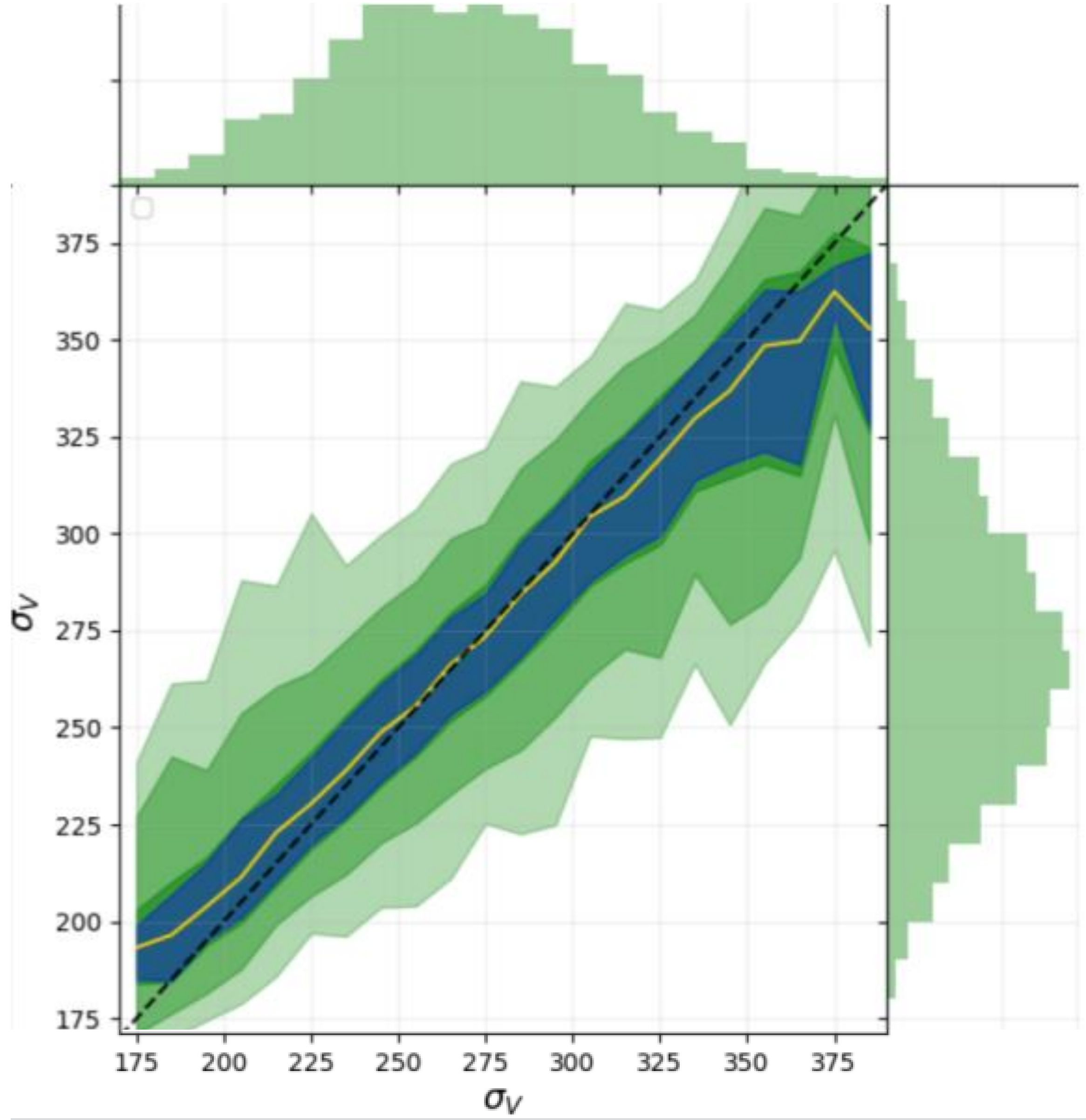


PRELIMINARY

Modelling Strong Lenses

By using Probabilistic Deep Learning (Normalizing Flows, Bayesian Deep Learning), we also model physical systems such as Strong Lenses.





1,2 and 3 sigma of velocity dispersion. A quantity directly connected with mass distribution . For a simple SL model:

$$\rho(\tilde{r}) = \frac{{\sigma_V}^2}{2\pi G\tilde{r}^2}.$$

R&D cooperation with Petrobras

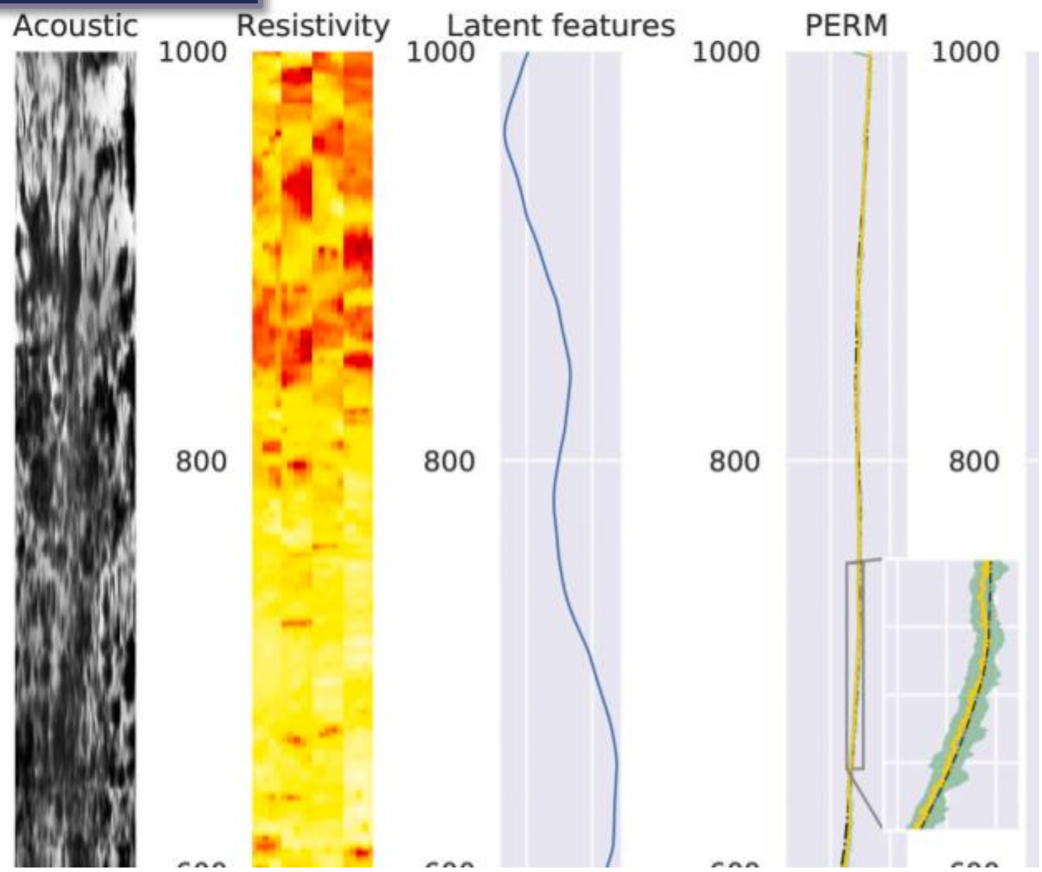
I. Artificial Intelligence in Image Profiles, Seismic and Petrographic data

II. NMR and pore structure simulation

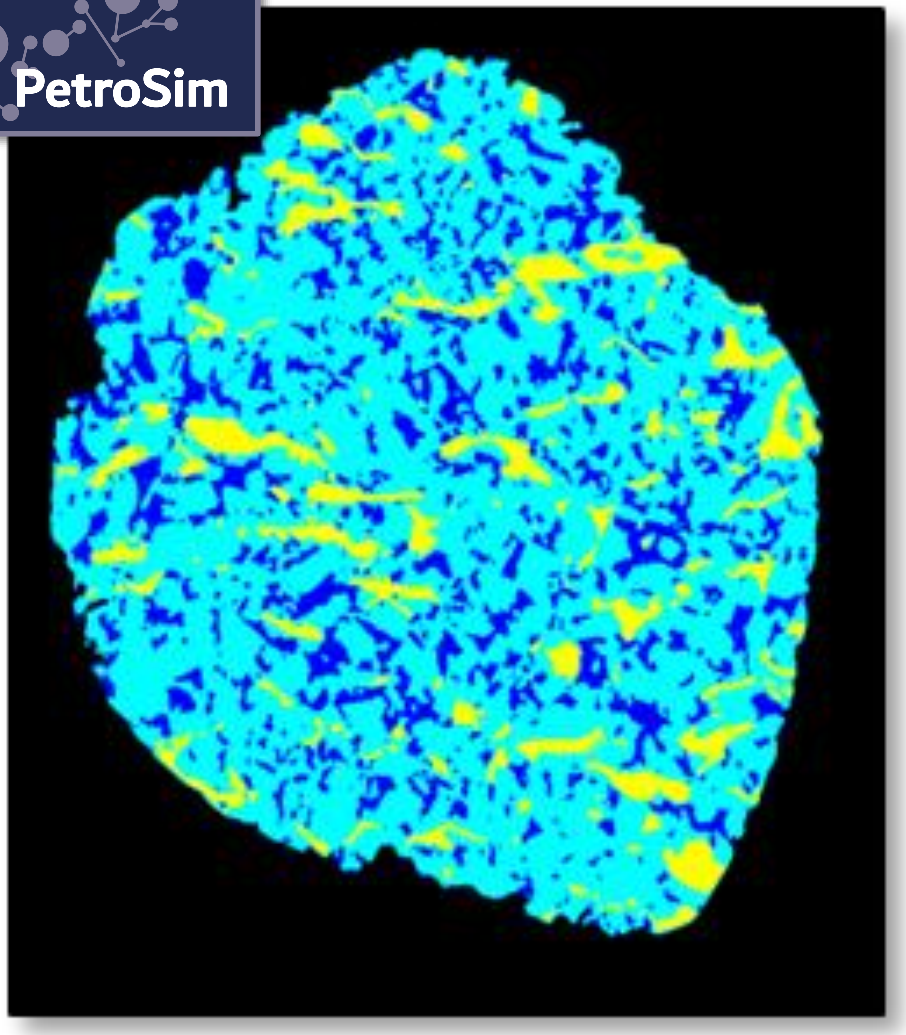
III. Instrumentation and HPC with multiGPU

Team dedicated to research and innovation at CBPF

I - PetroIA

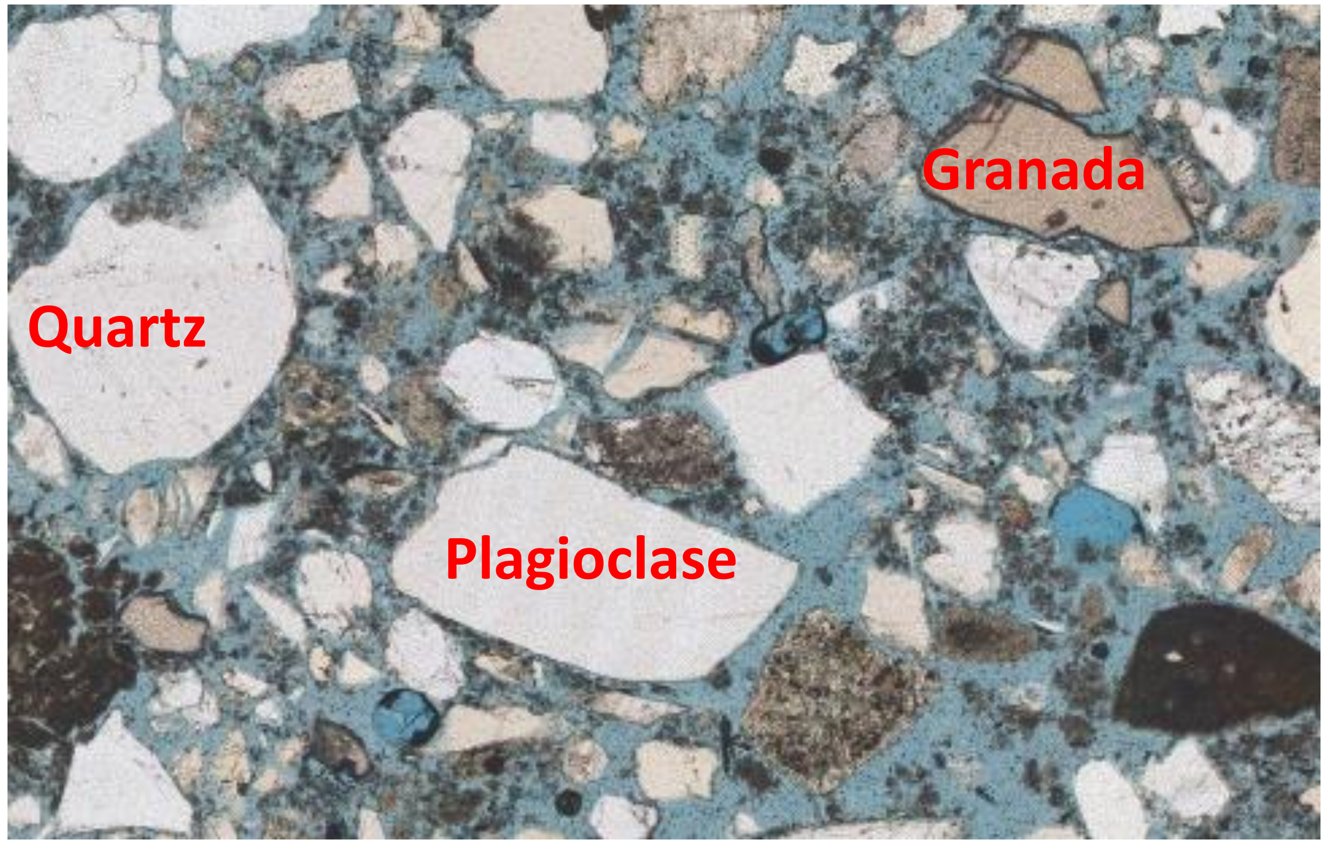


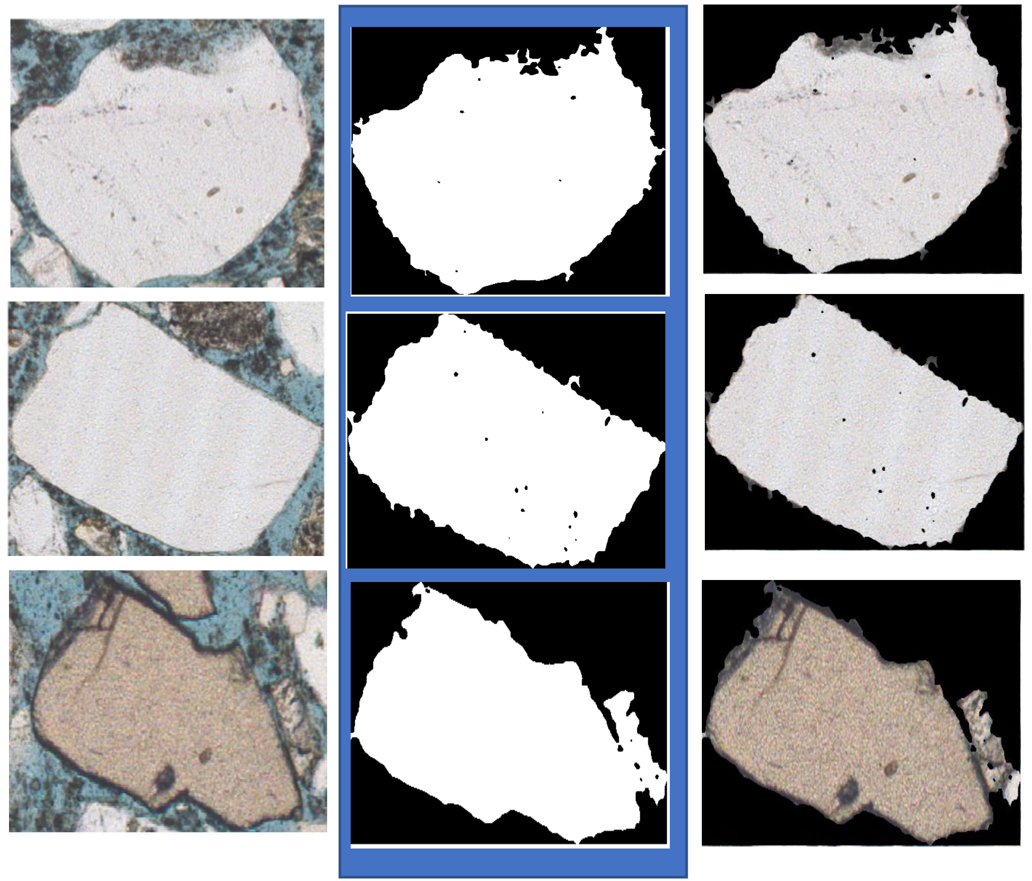
II - PetroSim

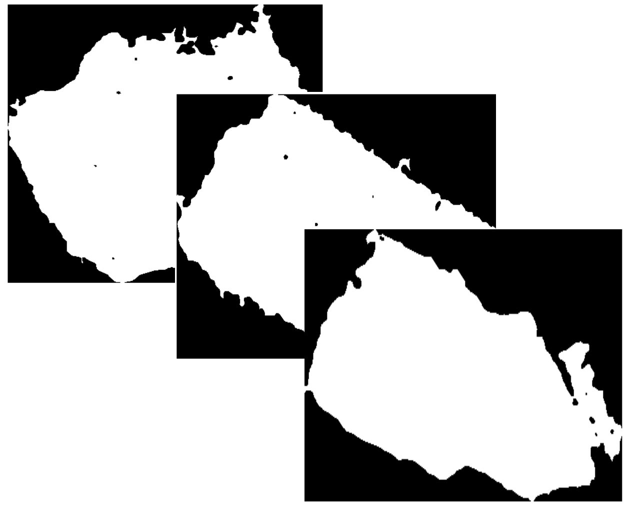


III - PetroInst

Petrographic thin section analysis & Artificial Intelligence



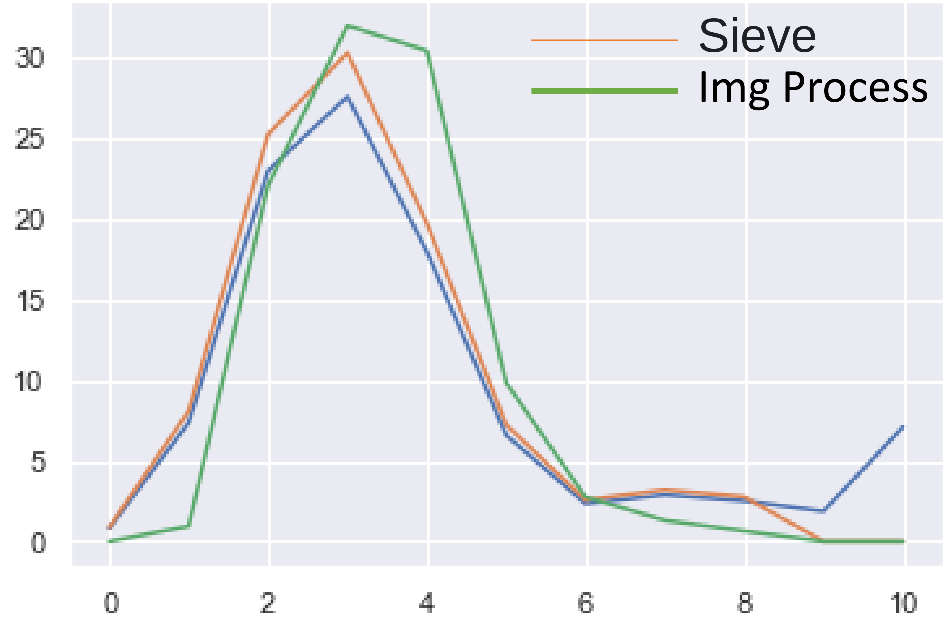




Estimation of **Granulometry** in sedimentary rock samples

Granulometry via Image Processing
- Non-destructive procedure
- Relatively fast, allowing analysis of many samples
- Maximum precision of the petrographic measures

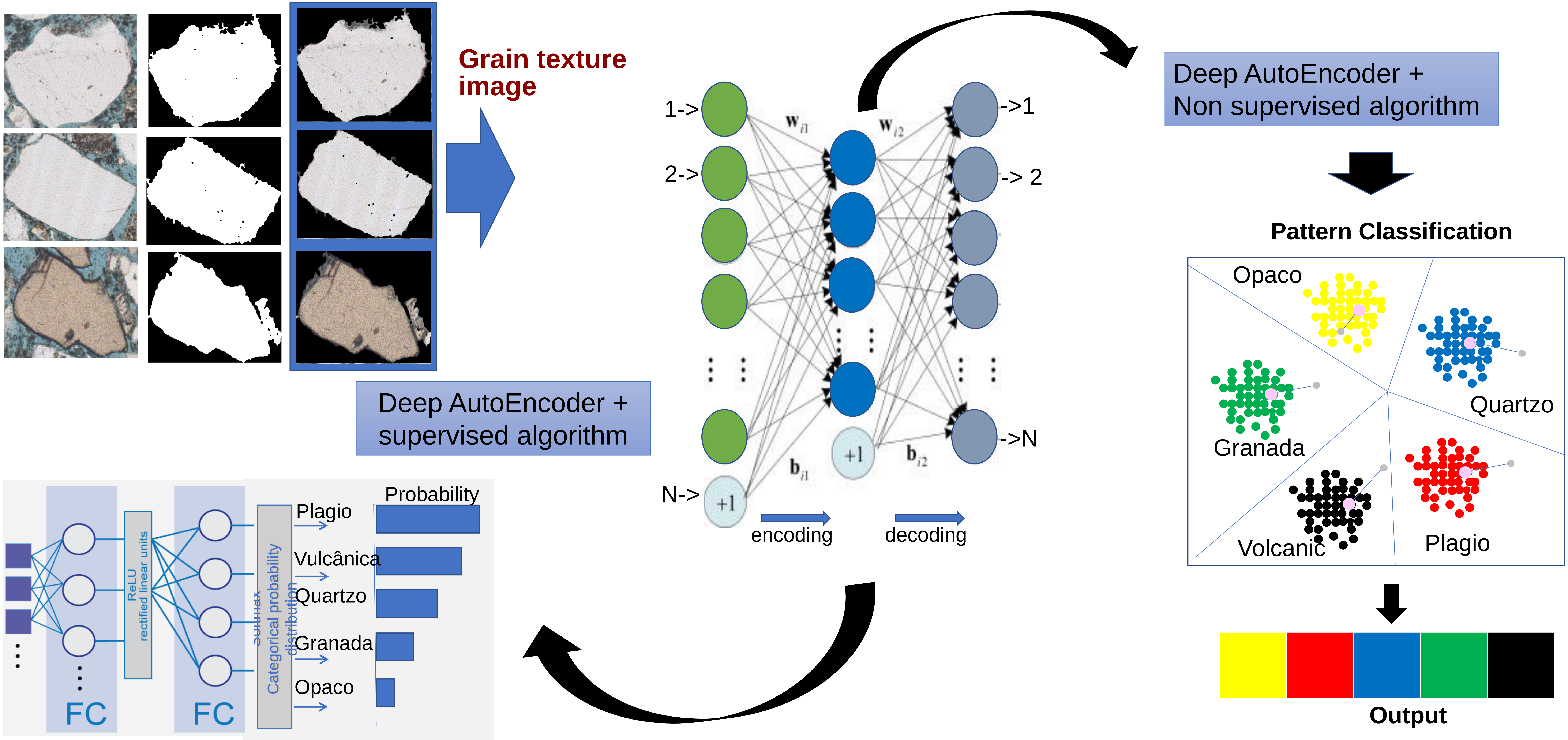
Grain binary image



	Areia					Silte					
	Cascalho	Muito Grossa	Grossa	Média	Fina	Muito Fina	Grosso	Médio	Fino	Muito Fino	Argila
Peneira	0,81	7,37	22,99	27,59	17,98	6,59	2,33	2,88	2,51	1,86	7,08
Peneira Corrigida	0,88	8,1	25,25	30,30	19,74	7,24	2,56	3,16	2,76	-	-
PDI	0,00	0,94	22,07	32,00	30,44	9,85	2,76	1,30	0,64	-	-

Petrographic thin section analysis & Artificial Intelligence

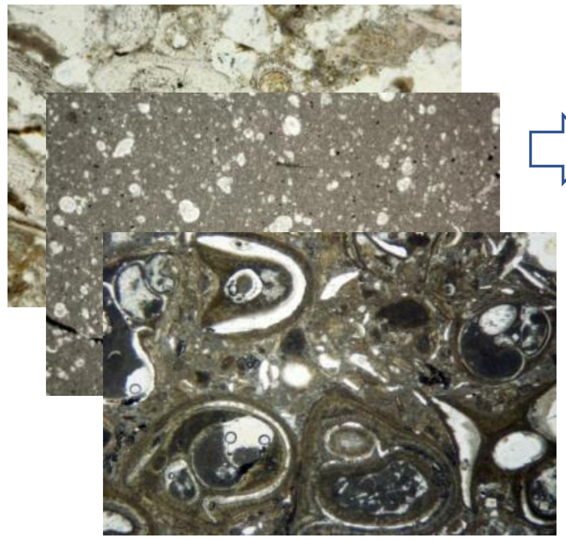
Classification or Clustering of grain images using non/supervised algorithm



Petrographic thin section analysis & Artificial Intelligence

Lithologic classification (thin section) using Stacked Deep AutoEncoder (SAE) + Optimization Algorithm

Input

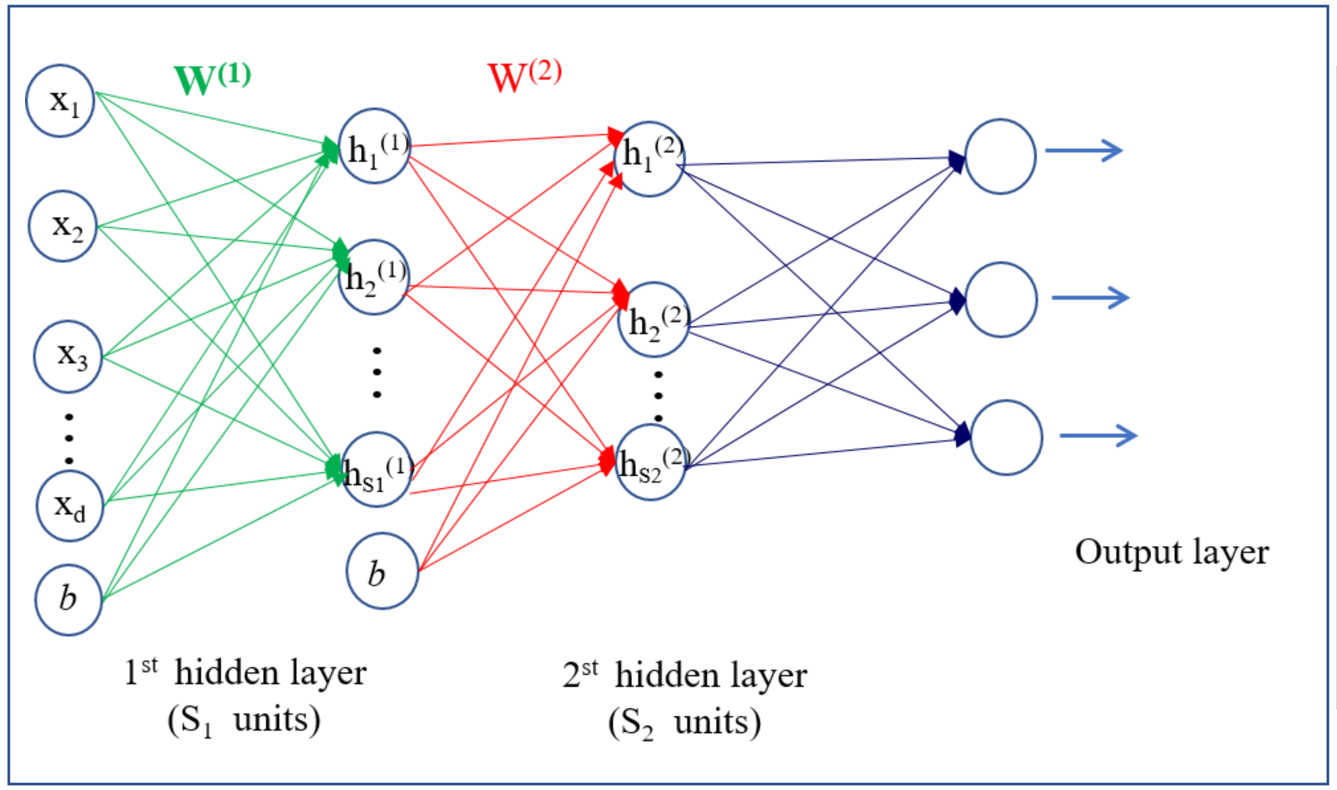


I_1, I_2, I_3, I_4, …, I_d

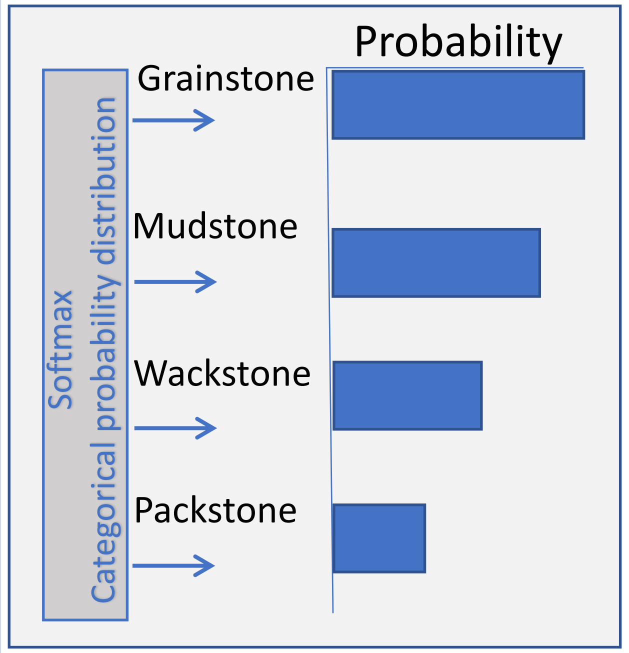
696x120 double

	1	2	3	4	5	6
1	0.2500	1.1874e+03	1.7634e+05	0.6477	1.6410	0.2182
2	0.2522	1.3162e+03	1.6246e+05	0.6299	1.5764	0.2281
3	0.2451	1.3535e+03	1.6252e+05	0.6161	1.6102	0.2256
4	0.2686	1.0063e+03	1.6698e+05	0.6839	1.6311	0.2233
5	0.2529	1.2689e+03	1.6835e+05	0.6396	1.5878	0.2262
6	0.2469	1.4258e+03	1.5642e+05	0.6069	1.5722	0.2308
7	0.2319	1.3874e+03	1.5534e+05	0.5898	1.6977	0.2145
8	0.2492	1.1904e+03	1.6968e+05	0.6417	1.6495	0.2181
9	0.2672	1.1143e+03	1.8570e+05	0.6792	1.5484	0.2288
10	0.2458	1.3666e+03	1.6464e+05	0.6155	1.5957	0.2268
11	0.2421	1.3265e+03	1.6644e+05	0.6157	1.6377	0.2205
12	0.2668	1.0633e+03	1.9006e+05	0.6846	1.5741	0.2246
13	0.2669	1.1196e+03	1.6590e+05	0.6717	1.5762	0.2280
14	0.2569	1.3480e+03	1.5820e+05	0.6306	1.5290	0.2345
15	0.2507	1.3244e+03	1.6869e+05	0.6291	1.5740	0.2281
16	0.2738	1.0458e+03	1.6453e+05	0.6859	1.5763	0.2330
17	0.2621	1.2167e+03	1.6330e+05	0.6552	1.5551	0.2302
18	0.2444	1.4229e+03	1.6562e+05	0.6057	1.5776	0.2279
19	0.2482	1.3113e+03	1.6100e+05	0.6249	1.6098	0.2245
20	0.2593	1.2327e+03	1.6308e+05	0.6495	1.5690	0.2303
21	0.2675	1.1702e+03	1.7834e+05	0.6720	1.5218	0.2341
22	0.2592	1.3316e+03	1.6725e+05	0.6392	1.5074	0.2366
23	0.2523	1.2390e+03	1.7130e+05	0.6427	1.6024	0.2235
24	0.2646	1.1326e+03	1.9099e+05	0.6741	1.5508	0.2265
25	0.2669	1.1196e+03	1.6590e+05	0.6717	1.5762	0.2280

Input layer (**Texture Parameters**)

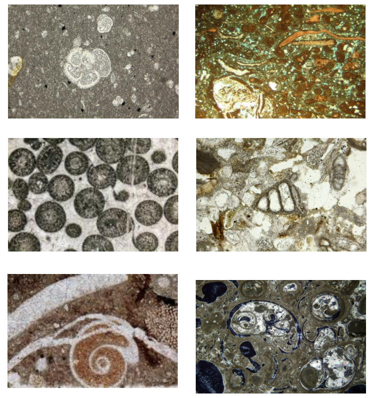


Feature representation with SAE



Softmax classifier training

Thin section



Input images for training

Auto Machine Learning

AutoKeras

TensorFlow Keras

Genetic Algorithm

- Search for the best solution to a problem
- Manipulates a population of individuals
- Individuals are possible solutions
- Individuals are combined (croosver) with each other
- The combined individuals may or may not mutate
- Populations evolve through successive generations until they find the optimal solution

HYPERPARAMETER OPTIMIZATION

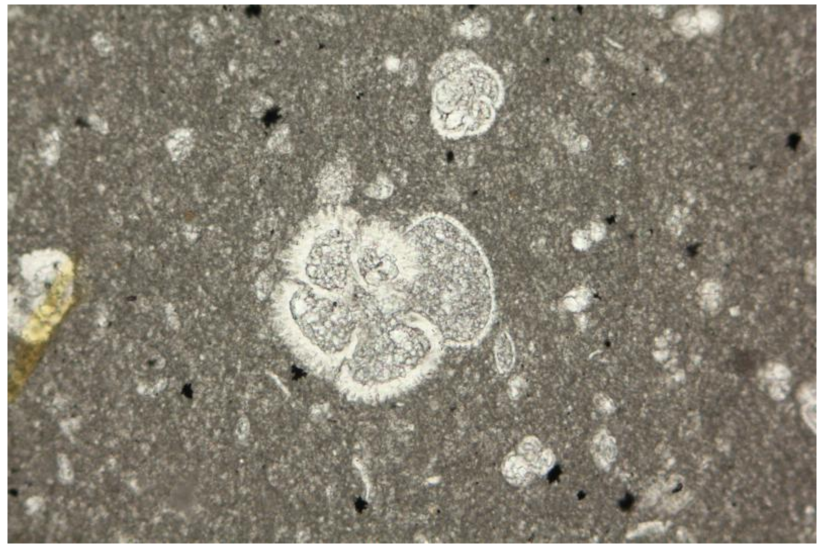
nn_param_choices = {
'**nb_neurons**': [48,64,128,256], #quantidade de neurônios das camadas encoder
'**nb_layers**': [1, 2, 3,4], #quantidade de camadas encoder
'**activation**': ['relu', 'tanh', 'sigmoid'],
'**optimizer**': ['rmsprop', 'adam', 'sgd', 'adagrad'],
'**nb_neurons_flat**': [100,150,200], #quantidade de neurônios da camada flatten

Fitness Function-> Mean Square Error validation set

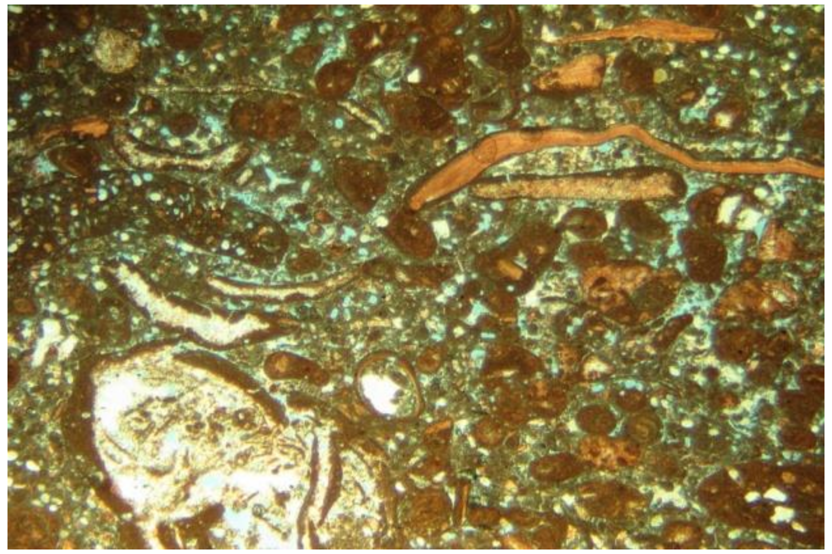
Petrographic thin section analysis & Artificial Intelligence

Lithologic classification (thin section) using Stacked Deep Autoencoder + Optimization Algorithm

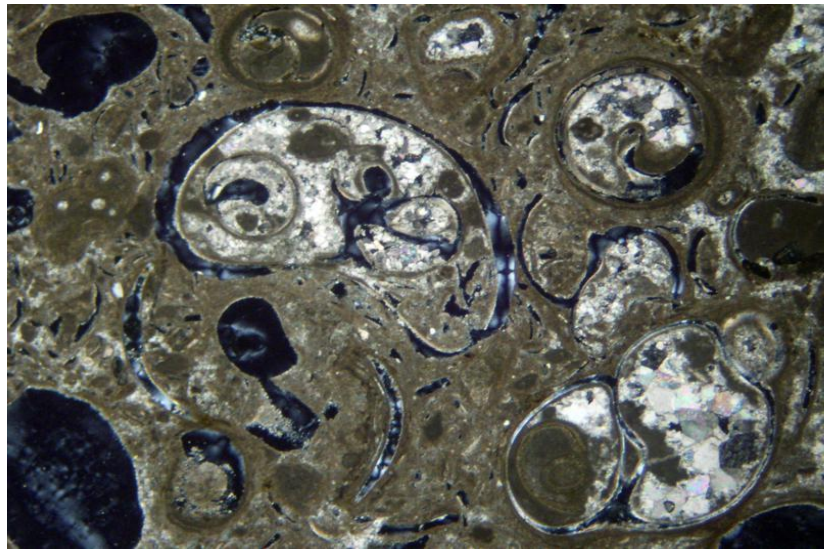
Grainstone (GR) = 228 images



Mudstone (MU) = 48 images



Packstone (PA) = 228 images



Wackstone (WA) = 180 images

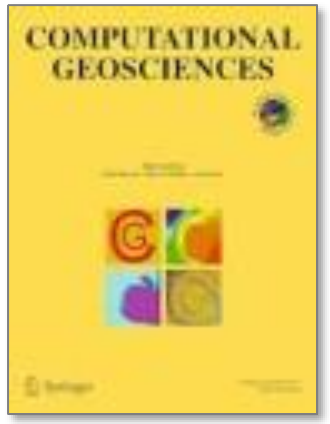
Confusion Matrix

Output Class	GR	MU	PA	WA	
1 (GR)	**224** 32.7%	**0** 0.0%	**3** 0.4%	**4** 0.6%	97.0% 3.0%
2 (MU)	**0** 0.0%	**39** 5.7%	**2** 0.3%	**0** 0.0%	95.1% 4.9%
3 (PA)	**4** 0.6%	**9** 1.3%	**220** 32.2%	**2** 0.3%	93.6% 6.4%
4 (WA)	**0** 0.0%	**0** 0.0%	**3** 0.4%	**174** 25.4%	98.3% 1.7%
	98.2% 1.8%	81.3% 18.8%	96.5% 3.5%	96.7% 3.3%	**96.1% 3.9%**

Target class

Model accuracy of 96.1%

Faria E. L. et al.; "***Lithology identification in carbonate thin section images of the Brazilian pre-salt reservoirs by the computational vision and deep learning***"; Computational Geosciences; submitted 2022.

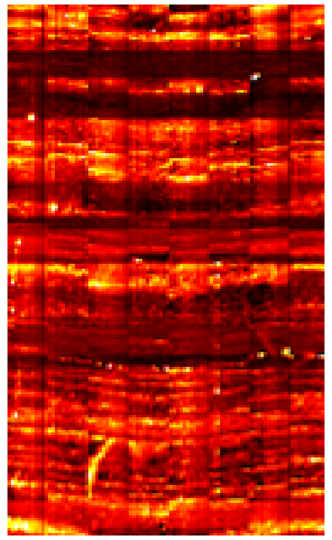


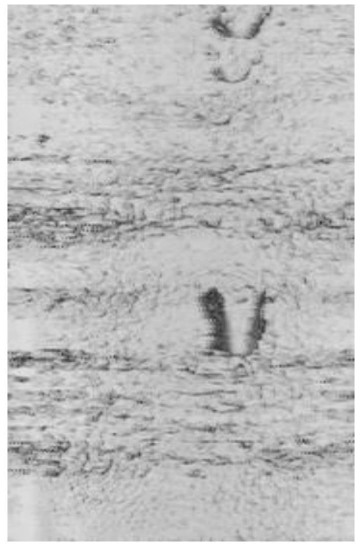
Springer

CBPF

resistivity

ultrasonic





Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

ELSEVIER

The impact score (IS) 2021 of JPSE is **4.97**

Journal of Petroleum Science and Engineering 170 (2018) 315–330

Estimation of permeability and effective porosity logs using deep autoencoders in borehole image logs from the brazilian pre-salt carbonate

Manuel Blanco Valentín[a,*], Clécio R. Bom[b], André Luiz Martins Compan[c], Maury Duarte Correia[c], Candida Menezes de Jesus[d], Anelise de Lima Souza[d], Márcio P. de Albuquerque[a], Marcelo P. de Albuquerque[a], Elisângela L. Faria[a]

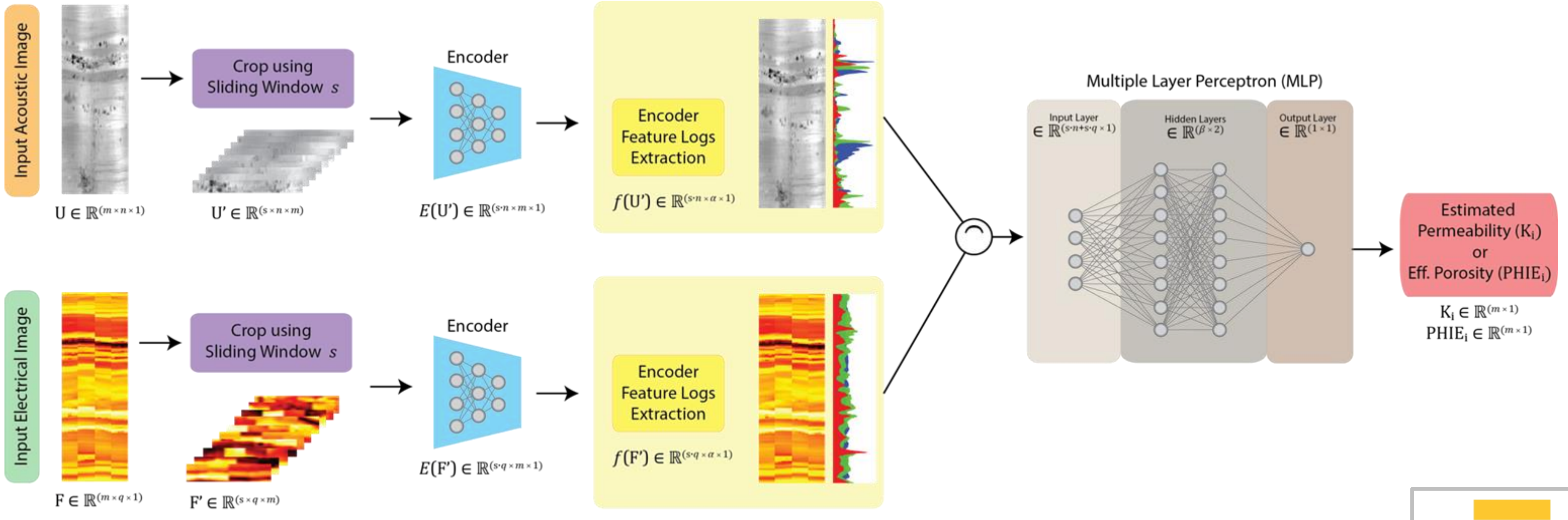
https://doi.org/10.1016/j.petrol.2018.06.038

Artificial Intelligence/Machine Learning technique that estimates permeability/porosity from ultrasound and resistivity images.

TWO STAGE APPOACH:

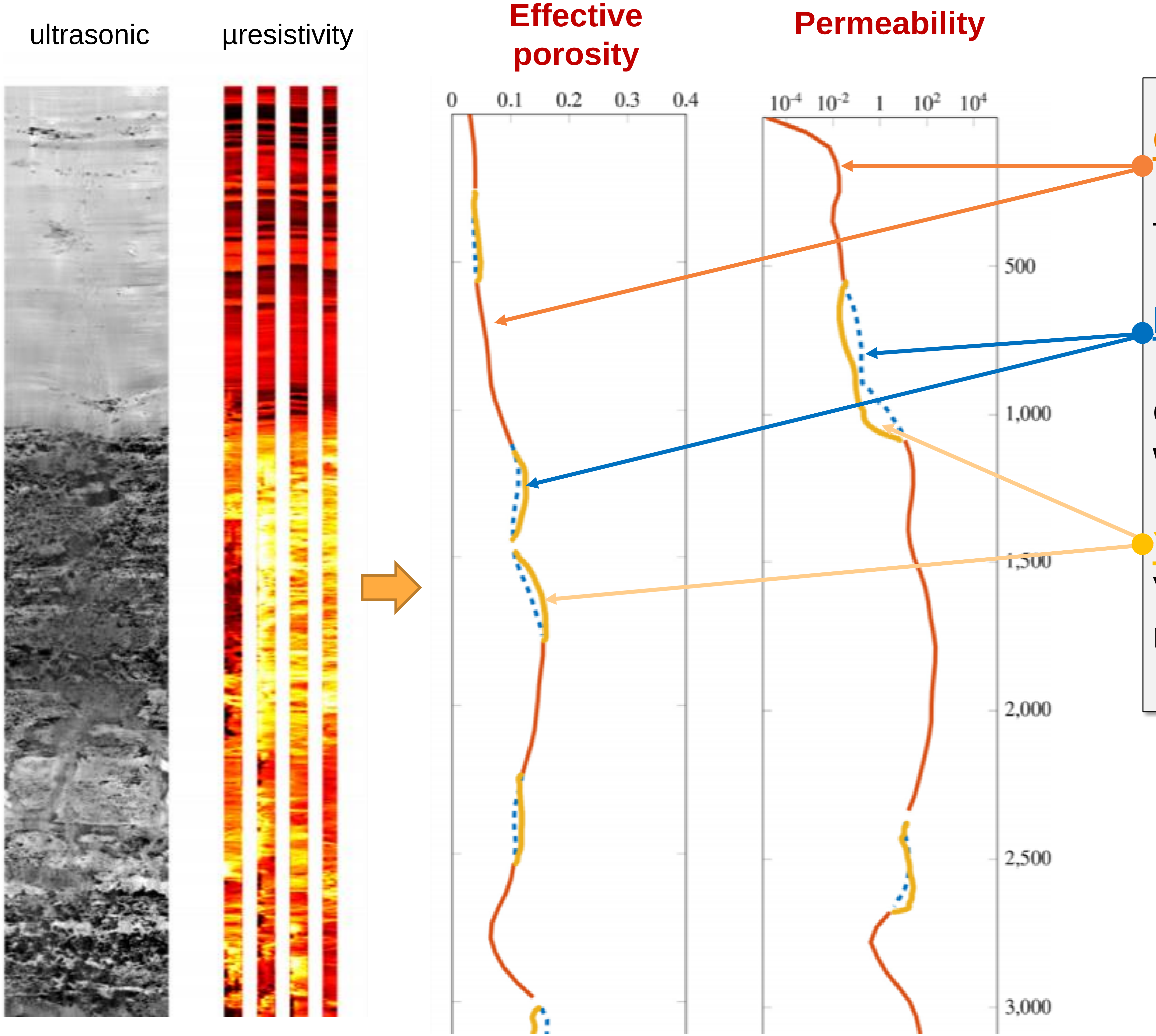
The first characterizes the image in terms of relevant attributes.

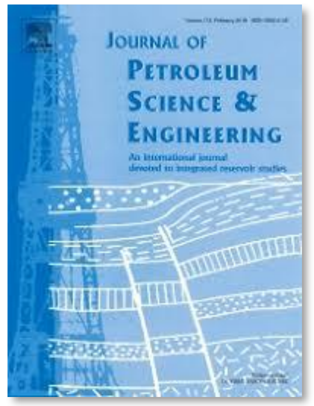
The second stage uses another Artificial Intelligence model to obtain the permeability/porosity.



Method Organization Chart

PETROBRAS


ultrasonic
μresistivity
Effective porosity
0
0.1
0.2
0.3
0.4
Permeability
10⁻⁴
10⁻²
1
10²
10⁴
500
1,000
1,500
2,000
2,500
3,000
Orange Lines:
Reference values used for training
Blue Lines:
Reference values used exclusively to compare with the network result
Yellow Lines:
Values predicted by the neural network


Journal of
Petroleum
Science &
Engineering

ELSEVIER

Journal of Petroleum Science and Engineering
Available online 16 January 2021, 108361
In Press, Journal Pre-proof

The impact score (IS) 2021 of JPSE is **4.97**

Bayesian Deep Networks for absolute permeability and porosity uncertainty prediction from image borehole logs from brazilian carbonate reservoirs

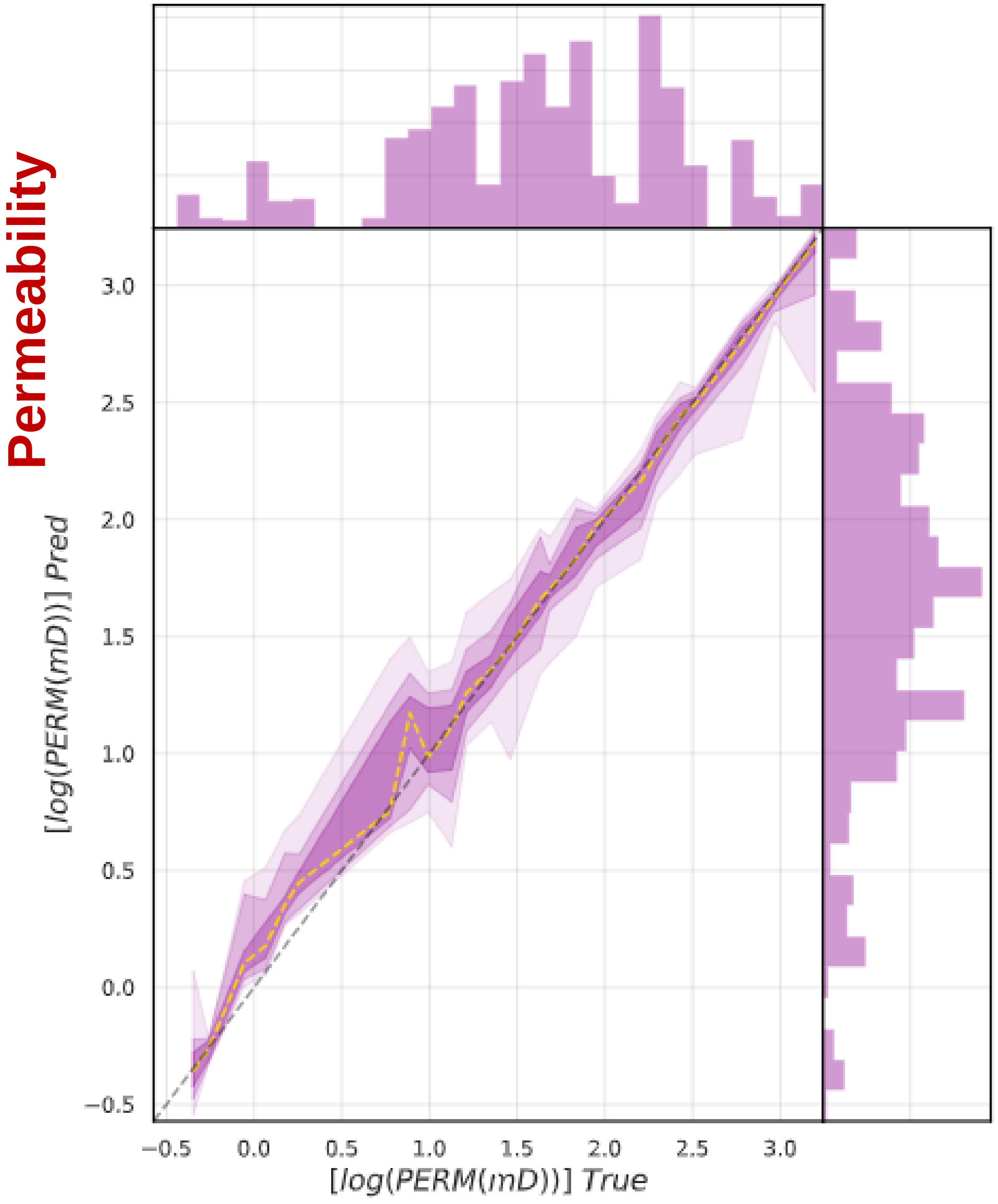
Clécio R. Bom [a, b], Manuel Blanco Valentín [a], Bernardo M.O. Fraga [a], Jorge Campos [c], Bernardo Coutinho [c], Luciana O. Dias [a], Elisangela L. Faria [a], Márcio P. de Albuquerque [a], Marcelo P. de Albuquerque [a], Maury Duarte Correia [c]

https://doi.org/10.1016/j.petrol.2021.108361

In a more recent paper, we worked on characterizing errors, using a Bayesian Artificial Intelligence approach.

In this case, we achieved a correlation greater than 99% in a pre-salt well. In addition, it was possible to estimate the error in each region of the well.

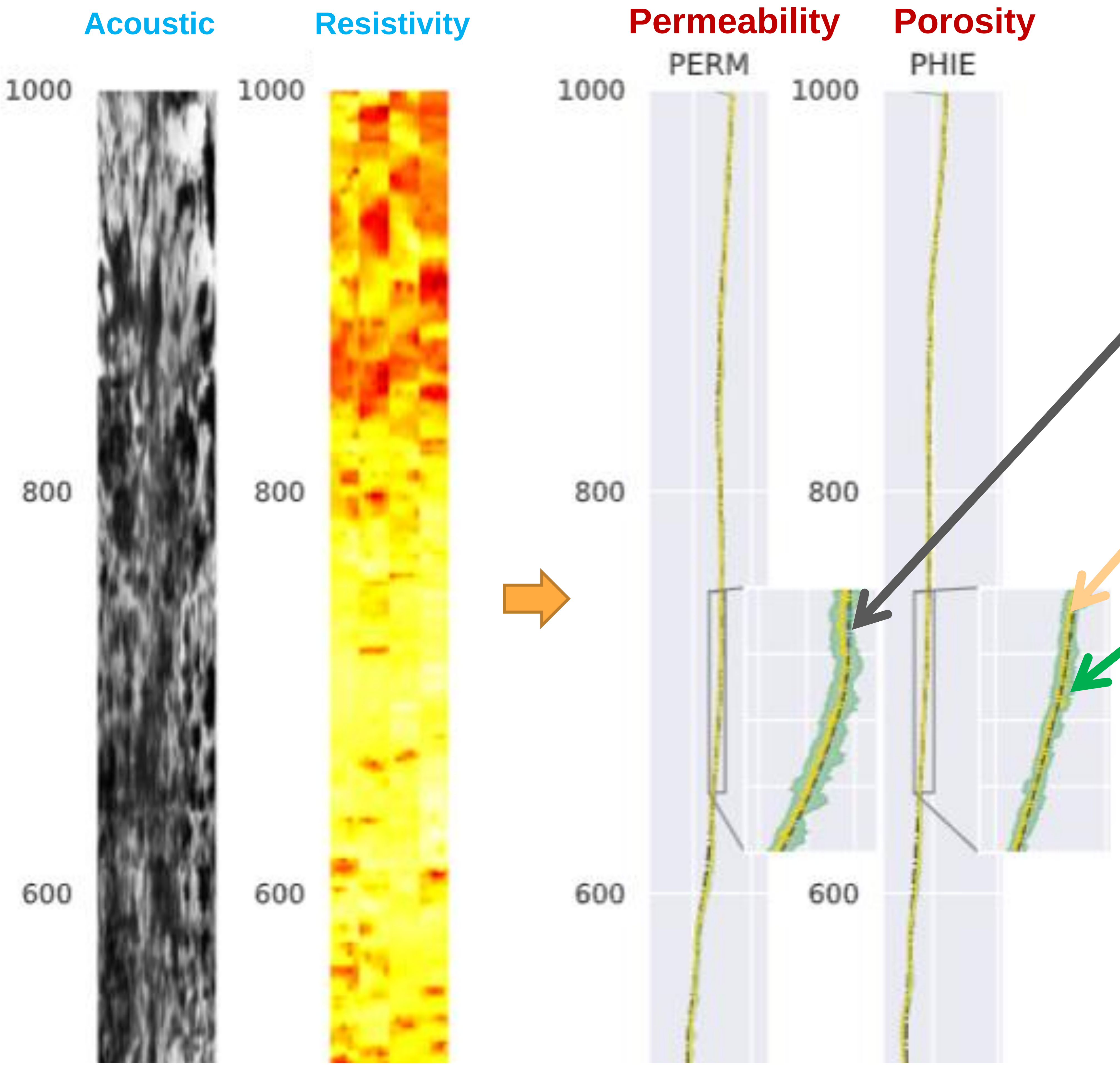
Predicted Permeability



Reference Permeability

PETROBRAS

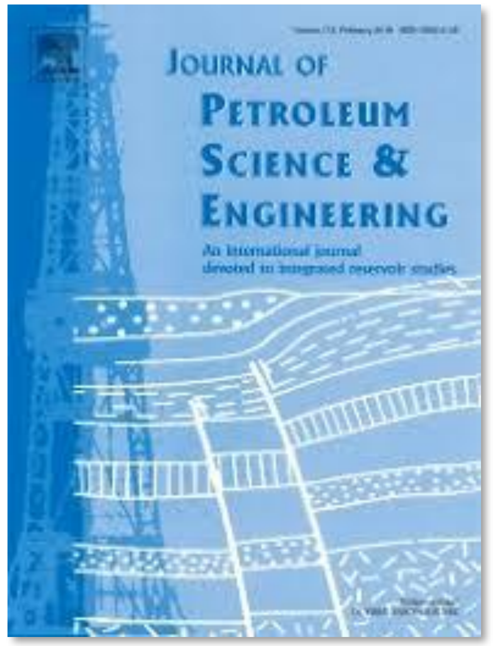
CBPF



Black Dashed Line:
Reference Values

Yellow Line:
DL predictions

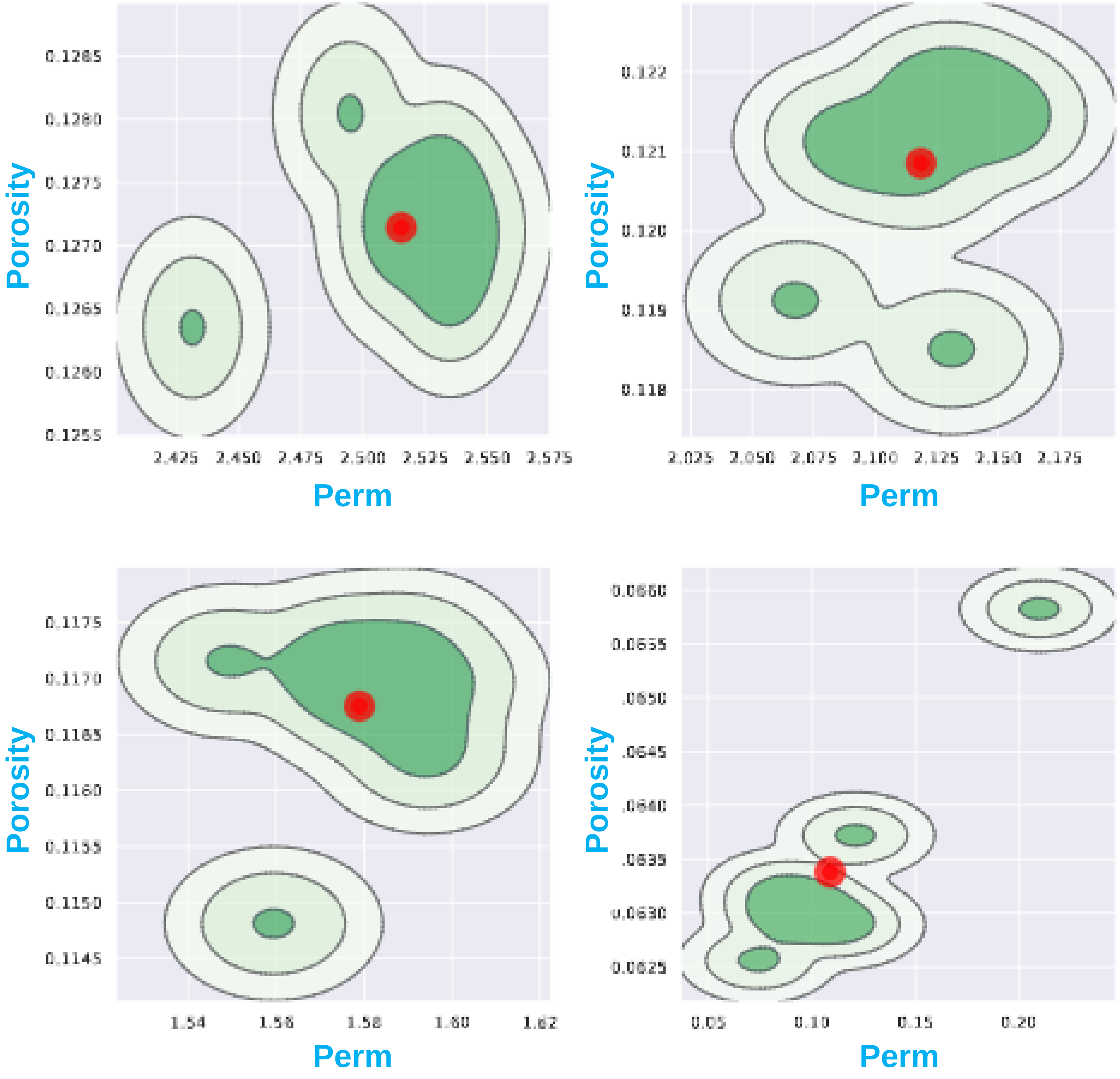
Green Shadow
Model Uncertainty



Joint Probability Constraints

Joint Probability Density Functions

● Target Value



$R^2(\%)$ Perm	$R^2(\%)$ Phie
99.63 ± 0.31	99.31 ± 0.92

Perm and Porosity correlation compared to plug,
In a Pre-salt Carbonate

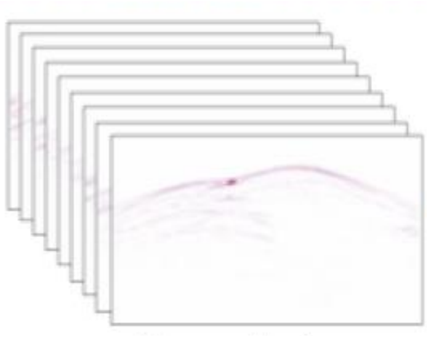
No human intervention!

Seismic analysis & Artificial Intelligence

Seismic Velocity-Model building (VM)

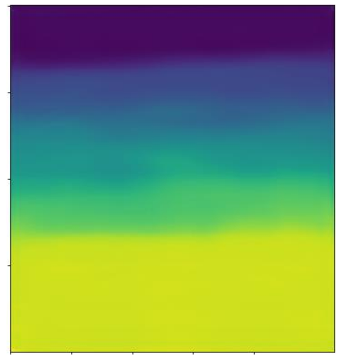
Conventional inversion method based on physical models:

Seismic Data



INVERSION METHOD BASED ON COMPLEX PHYSICAL MODELS

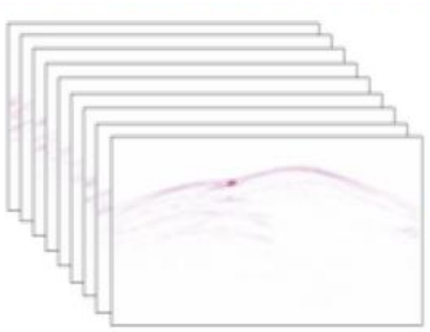
Predicted VM



- Time consuming
- Computationally expensive
- and there is always a need for human iteration

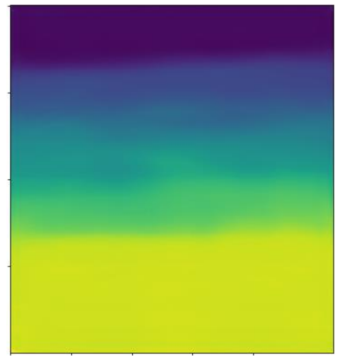
Deep Learning Inversion – Automatic seismic Velocity-Model building (VM)

Seismic Data



DEEP LEARNING INVERSION

Predicted VM



- are based on big-data training;
- in the **training** phase, the network establishes a nonlinear projection from the multi-shot seismic data to the corresponding velocity models (requires more time).
- However, during the **prediction** phase, the trained network can be used to estimate the velocity models from the new input seismic data in real time.
- No human iteration.

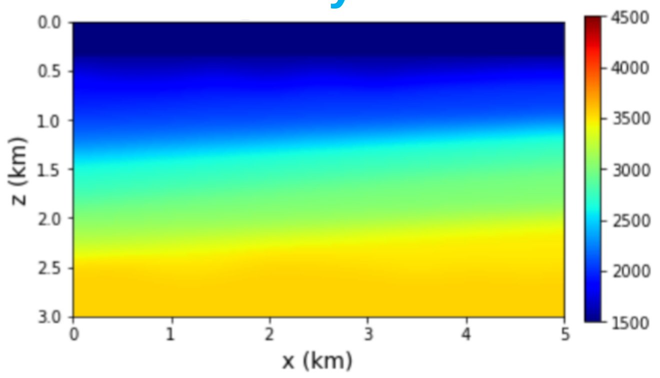
Klatt, M. et al.; "***Deep learning strategy for salt model building***", GEO-2021-0362 (Geophysics); submitted 2022;

Salt-geometries with Deep Learning

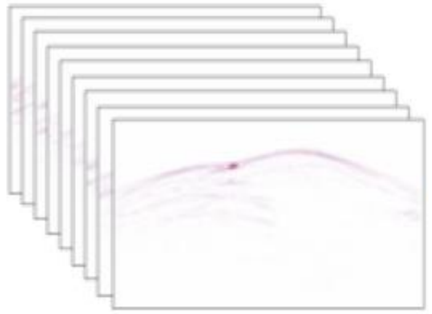
Deep Learning to predict the salt-geometries

Salt model data form SEG (Society of Exploration Geophysicists)

Velocity Model

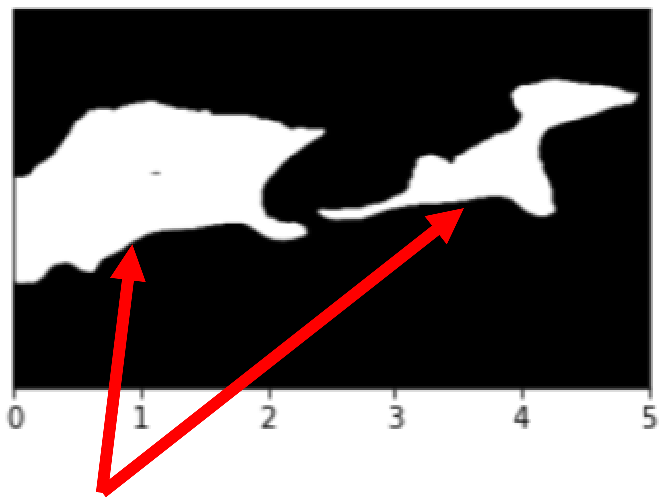


Seismic Migrated Data



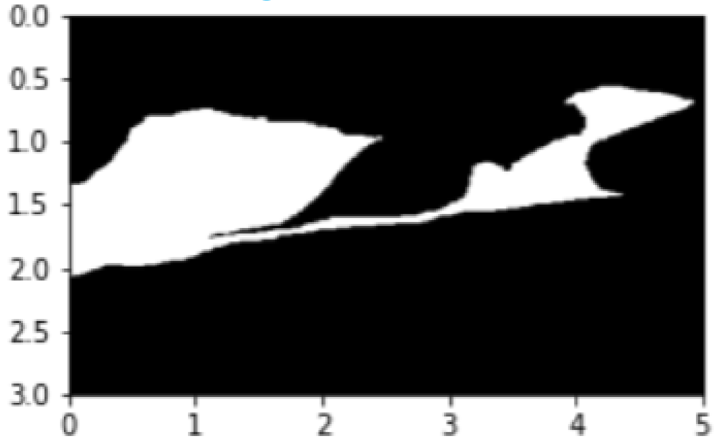
DEEP LEARNING
NN

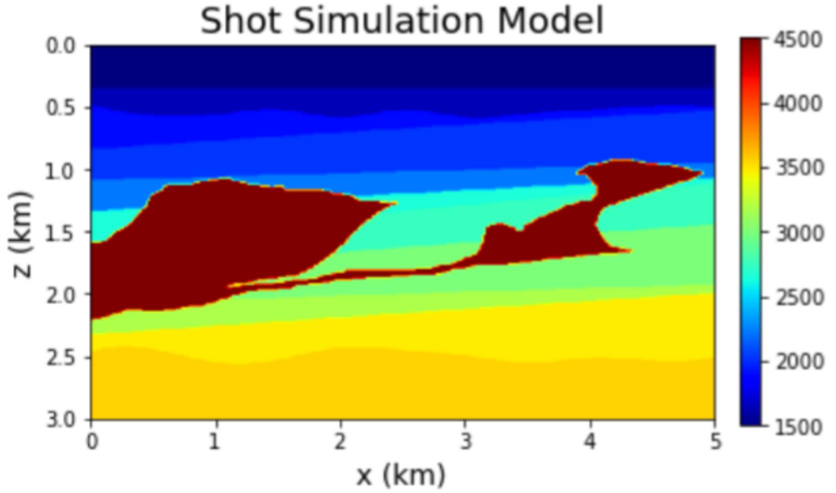
Predicted Salt Geom.



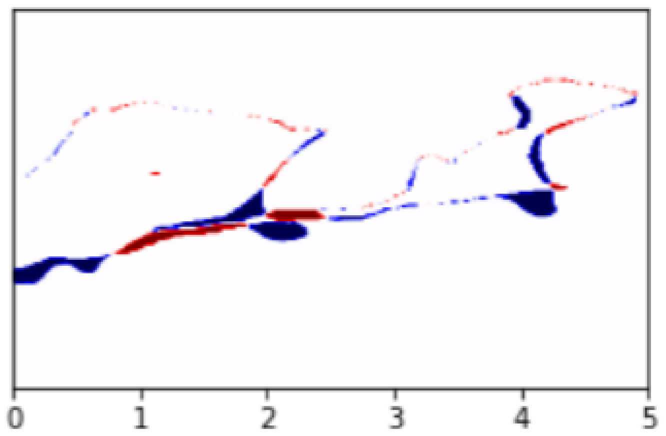
Salt Mask

Target Salt Geom.





Error



Results is currently on statistical validation on real industry cases where the salt geometry was poorly defined by traditional methods.

Submitted 2022

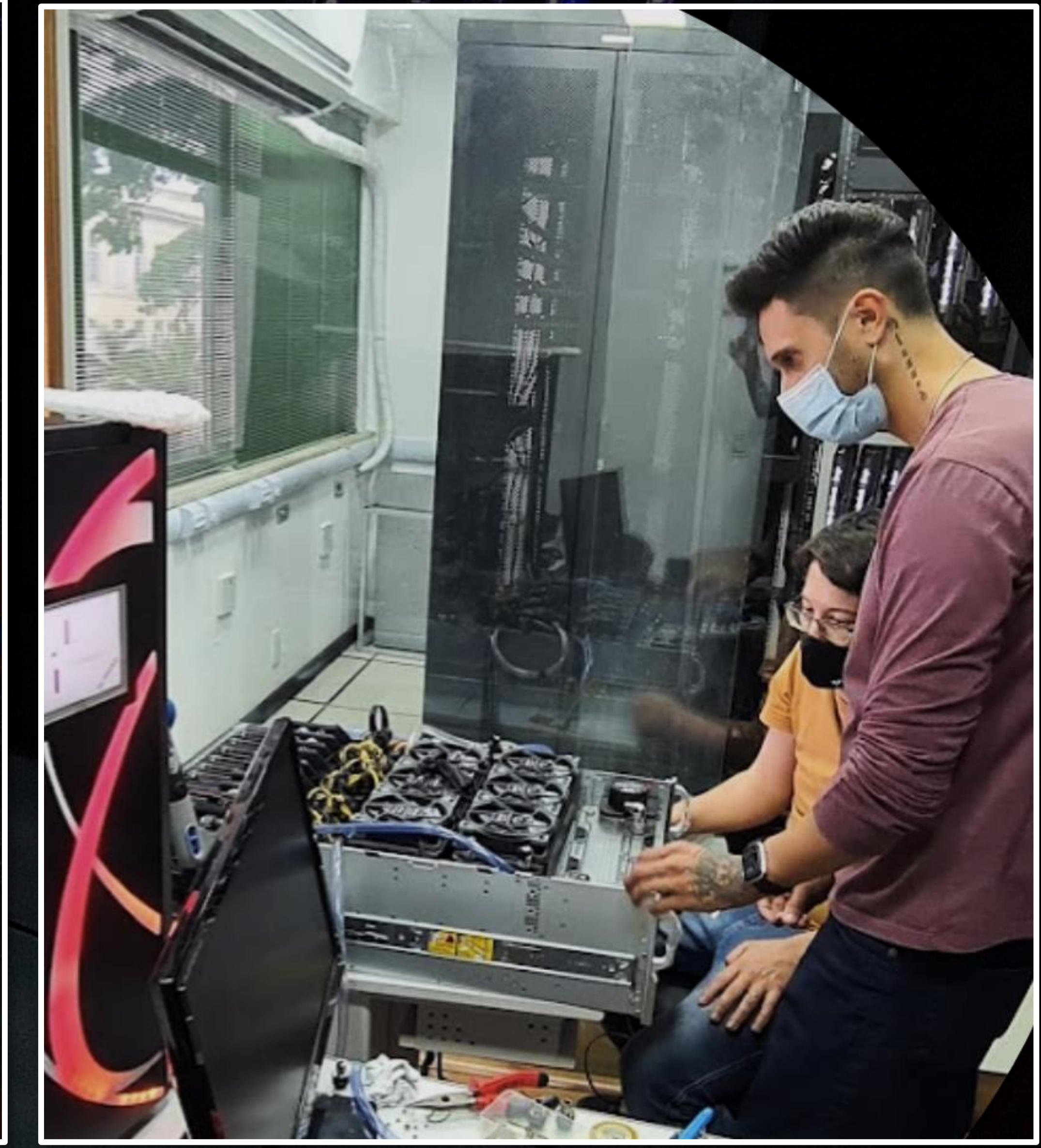
Ana Paula O. Muller, et al.; "***Complete identification of complex salt-geometries from inaccurate migrated images using Deep Learning***"; GEO-2021-0362 (Geophysics);

NEW MULTIGPUS COMPUTER
INSTRUMENTS AND SYSTEMS FOR AIDLnG

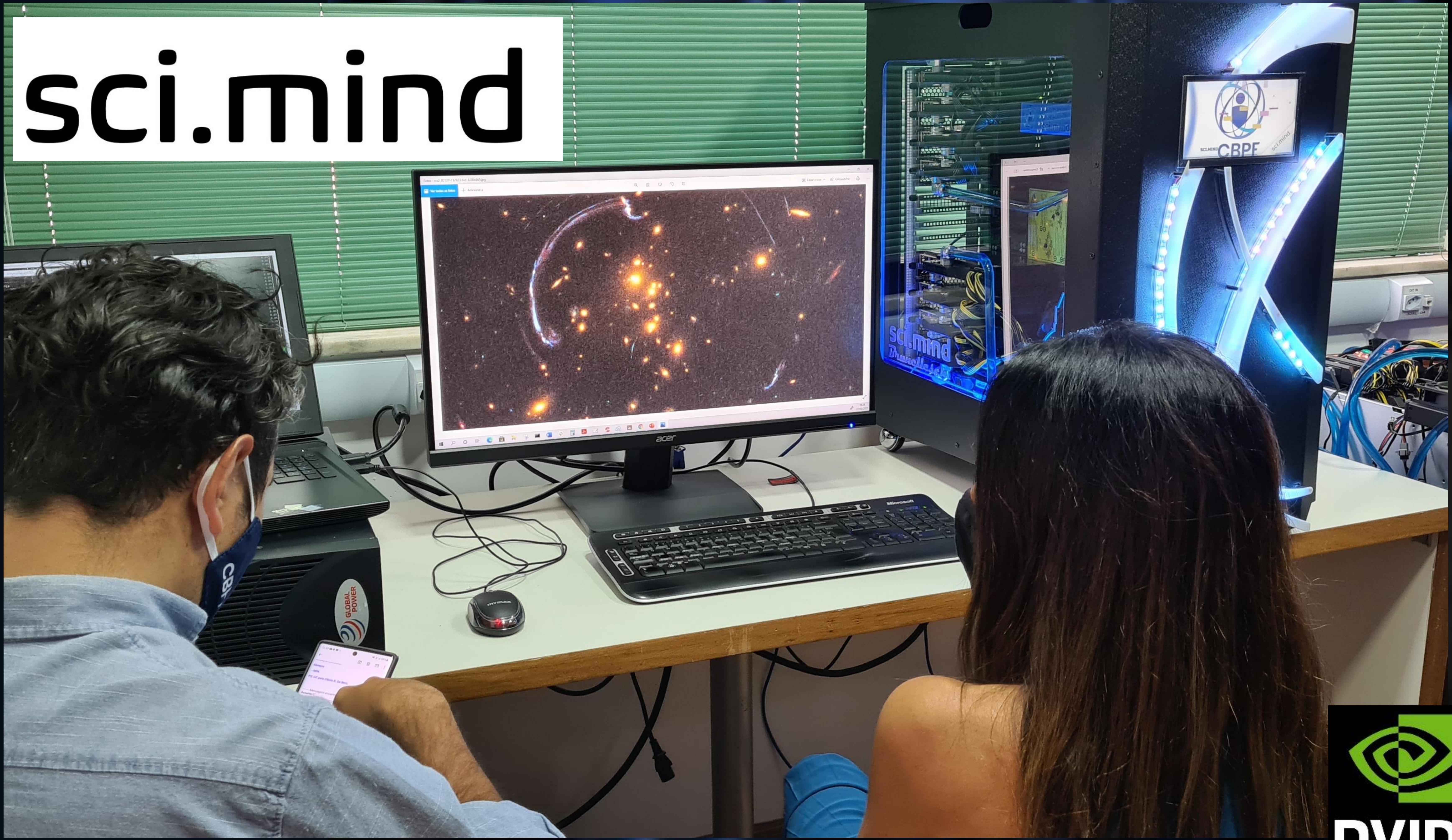
Development of multiGPUs technologies tuned to the characteristics of different data sets

These technologies contribute to the testing, fast prototyping and feasibility study of AI models

IA multiGPU Lab



➡ Sci.minds allow tunning performance parameters


sci.mind
CBPF
NVIDIA

ARTIFICIAL INTELLIGENCE
Classification, Simulations and Forecasts

CBPF

We participated and won the International Astrophysics Challenge / 2020

International challenge to identify gravitational lensing systems in simulated images.

Developing a Victorious Strategy to the Second Strong Gravitational Lensing Data Challenge

C. R. Bom[a,b], B. M. O. Fraga[a], L. O. Dias[a], P. Schubert[a,1], M. Blanco Valentin[d,2], C. Furlanetto[e], M. Makler[a,c], K. Teles[a], M. Portes de Albuquerque[a], R. Benton Metcalf[f,g]

[a]*Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, CEP 22290-180, Rio de Janeiro, RJ, Brazil*
[b]*Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rodovia Mário Covas, lote J2, quadra J, CEP 23810-000, Itaguaí, RJ, Brazil*
[c]*International Center for Advanced Studies & Instituto de Ciencias Físicas, ECyT-UNSAM & CONICET, 25 de Mayo y Francia. C.P.: 1650, San Martín, Buenos Aires, Argentina*
[d]*Electrical and Computer Engineering Department, McCormick School, Northwestern University, 633 Clark St, Evanston, IL 60208*
[e]*Universidade Federal do Rio Grande do Sul, Departamento de Física, CEP 91501-970, Porto Alegre, RS, Brazil*
[f]*Dipartimento di Fisica e Astronomia, Università di Bologna, via Gobetti 93/2, I-40129 Bologna, Italy*
[g]*INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy*

Abstract

Strong Lensing is a powerful probe of the matter distribution in galaxies and galaxy clusters and a relevant tool for cosmography. Analyses of strong gravitational lenses with Deep Learning (DL) and Convolutional Neural Networks (CNNs) have become a popular approach due to these astronomical objects' rarity and image complexity. Next-generation surveys (both ground and space-based) will provide more opportunities to derive science from these objects and an increasing data volume to be analyzed. However, finding strong lenses is challenging, as their number densities are orders of magnitude below those of galaxies. Therefore, specific Strong Lensing search algorithms are required to discover the highest number of systems possible, with high purity and low false alarm rate to minimize human intervention. The need for better algorithms has prompted the development of an open community data science competition named Strong Gravitational Lensing Challenge (SGLC) by the Bologna group. In this work, we present the Deep Learning strategies and methodology used to design the highest-scoring algorithm in the II SGLC, which was based on Euclid-like simulations. We discuss the approach used for this dataset, the choice for a suitable architecture, particularly the use of a network with two branches to work with images in different resolutions, and its optimization. We also discuss the limit of what ... tailor-made architecture in a survey in contrast ... pipeline, and discuss the best choice to easily ... ment. This work helps to take a step towards ... ks.

... networks

Submitted 2022

Monthly Notices
of the Royal Astronomical Society

The team working for the industry contributed to the scientific project.

Da esq. para dir., Patrick, Marcelo, Martín, Paulo, Clécio, Pedro, Manuel, Marcos e Luciana

(Crédito: NCS/CBPF)

https://portal.cbpf.br/pt-br/ultimas-noticias/grupo-do-cbpf-vence-desafio-internacional

EXPERTISE, TEAM & INFRASTRUCTURE

CBPF

IA multiGPU Lab

Laboratório de Computação e Inteligência Artificial do CBPF

TEAM WITH EXTENSIVE EXPERIENCE IN PD&I

- Several articles published in physics, engineering and geosciences
- Experience in filing patents with Petrobras
- Winner of the last International Astrophysics Datachallenge in 2020
- Participation in several national and international R&D projects
- International insertion in strategic themes of Physics and technological development.

CBPF COMPUTING AND ARTIFICIAL INTELLIGENCE LAB

- Infrastructure developed locally and dedicated FULLTIME to the Project:
  - 75 GPUs (522.752 #Cores; 1.3TB RAM), 110TB storage and 10Gbps connection with the internet.
- Cooperation projects with industry (Petrobras, startups, etc.)

CBPF Centro Brasileiro de Pesquisas Físicas
UNIDADE DE PESQUISA DO MCTI

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

OBRIGADO

AI Applications in Astrophysics and Developments with the O&G Industry

Marcelo Albuquerque and **Clécio R. De Bom**

marcelo@cbpf.br *debom@cbpf.br*

DATA SCIENCE IN (ASTRO)PARTICLE PHYSICS and COSMOLOGY: the BRIDGE to INDUSTRY

27-30 JUNHO 2022, Universidade de Coimbra, Coimbra, Portugal

Um evento que mostra o potencial da ciência de dados na sociedade moderna e estimula sinergias entre investigação fundamental e indústria

LIP