Ministério da Ciência, Tecnologia e Inovações

Centro Brasileiro de Pesquisas Físicas

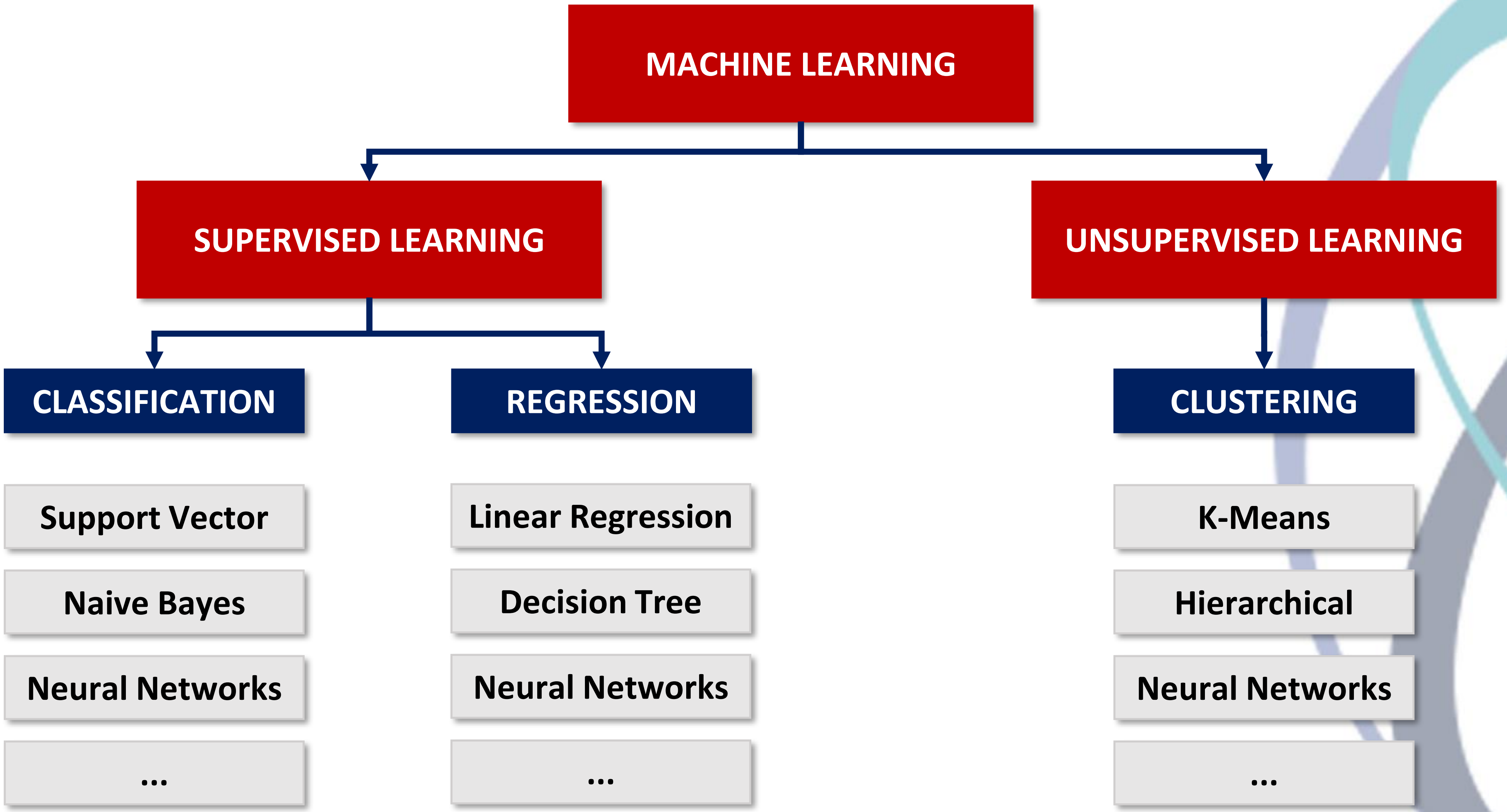
Redes Neurais profundas e aplicações Deep Learning

Clécio Roque De Bom – debom@cbpf.br

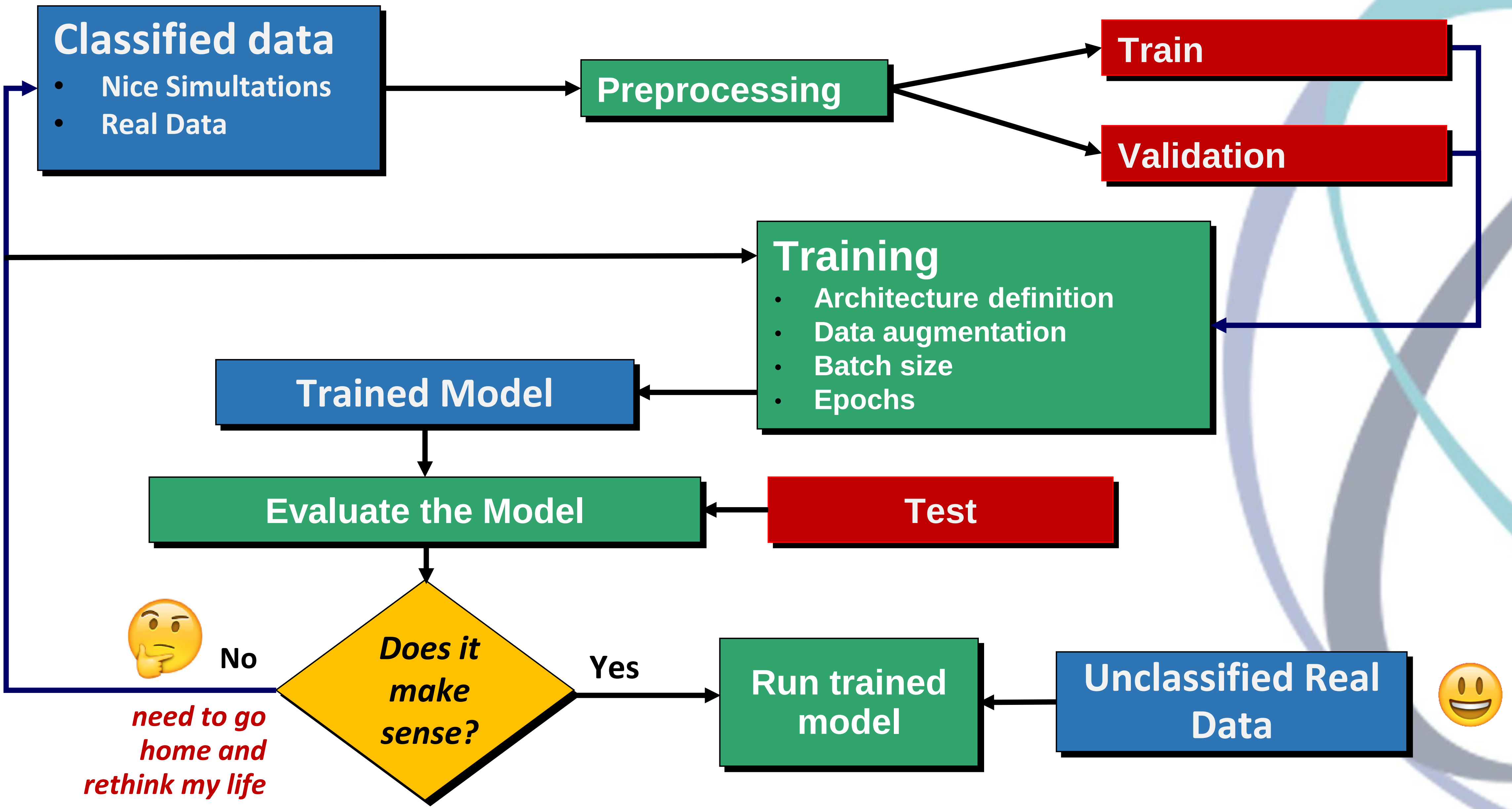
clearnightsrthebest.com

CBPF

ML: Supervised & Unsupervised

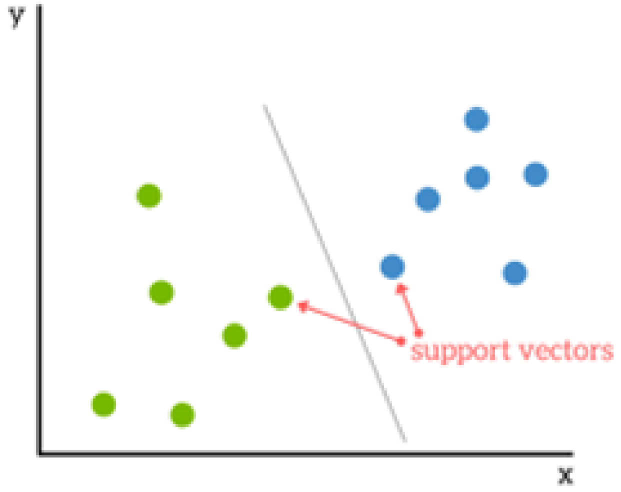


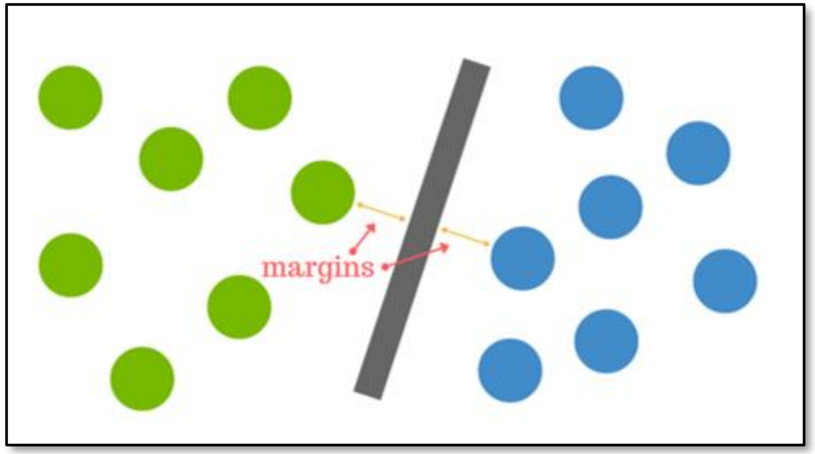
ML: Workflow Supervised

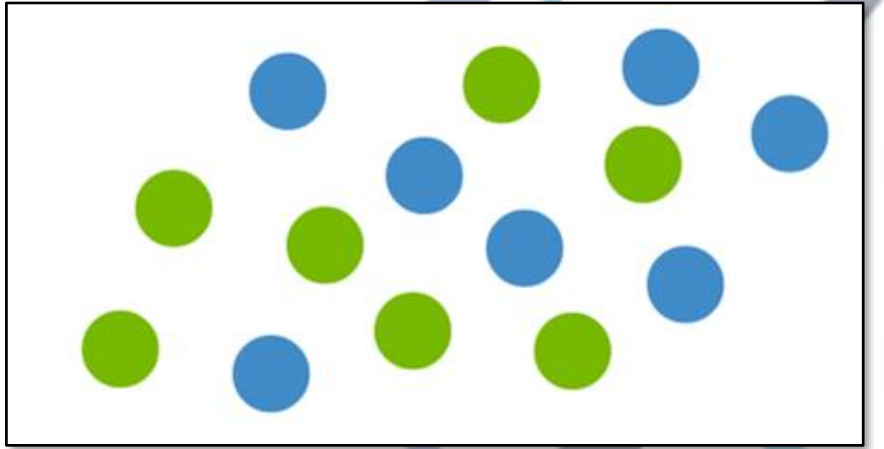


ML: Support Vector Machines - SVM

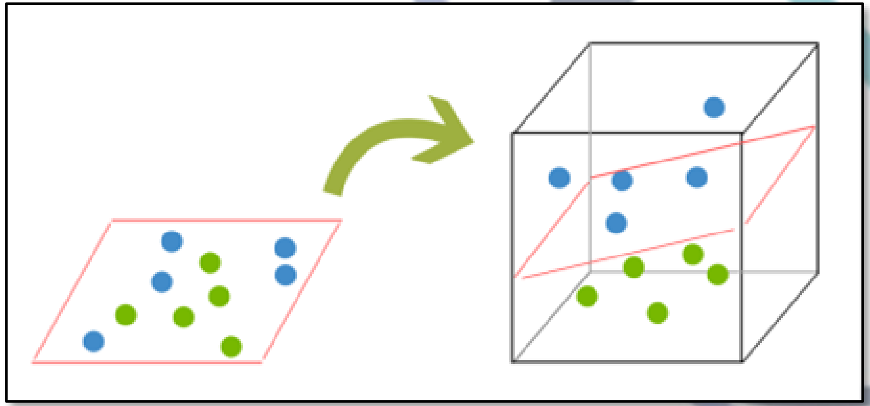
SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes.





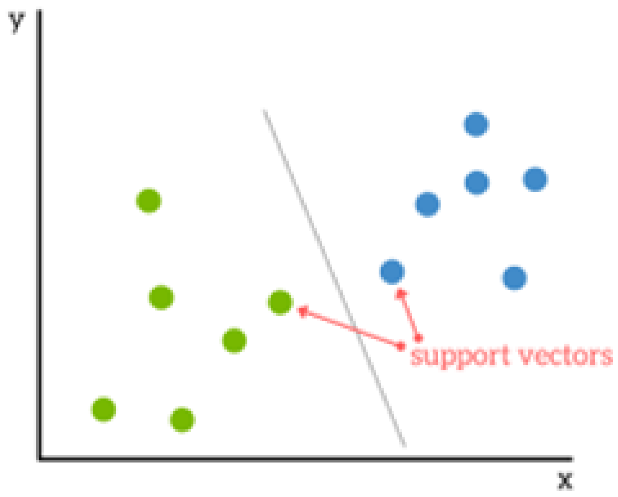


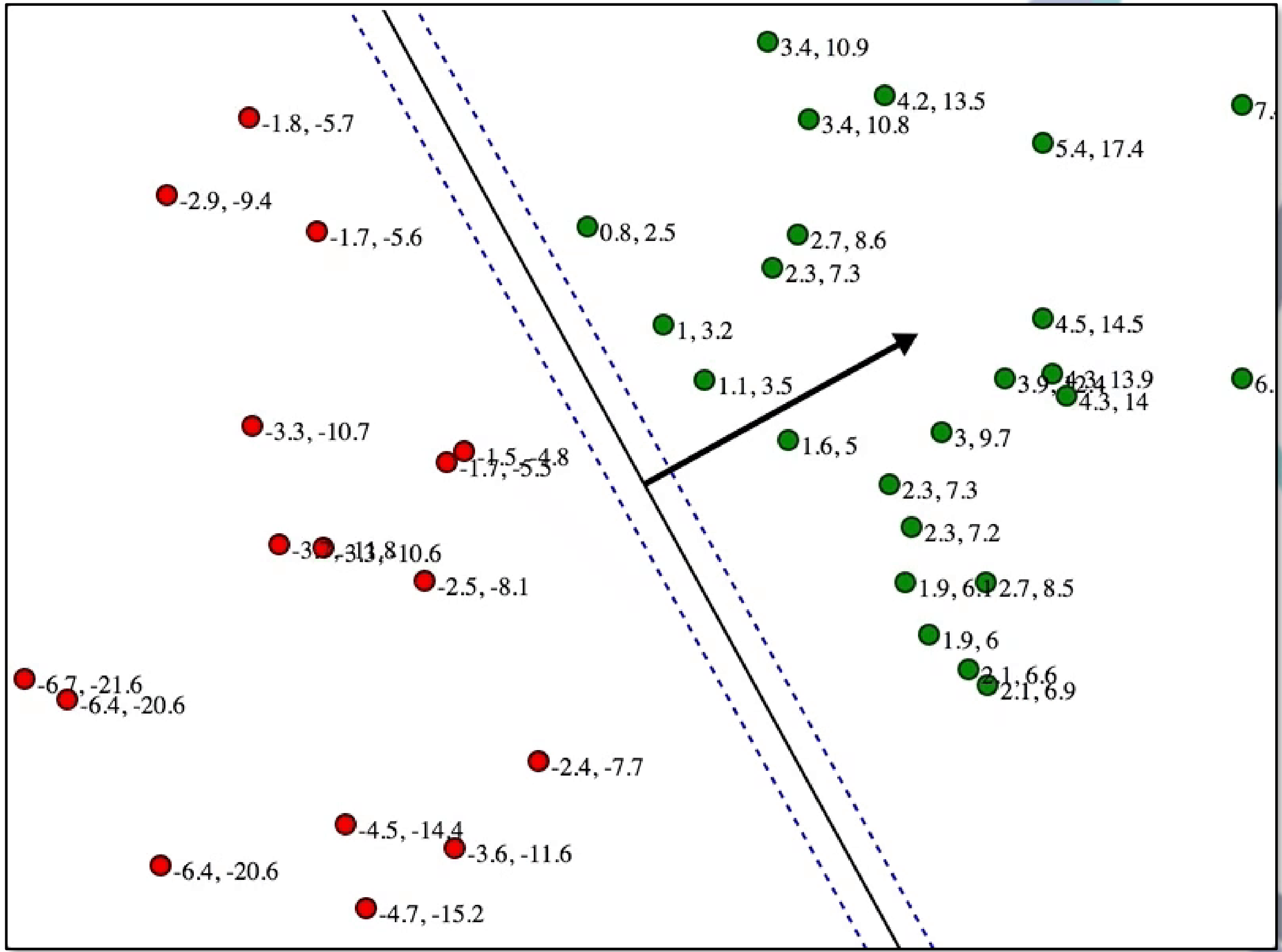
The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly.



ML: Support Vector Machines - SVM

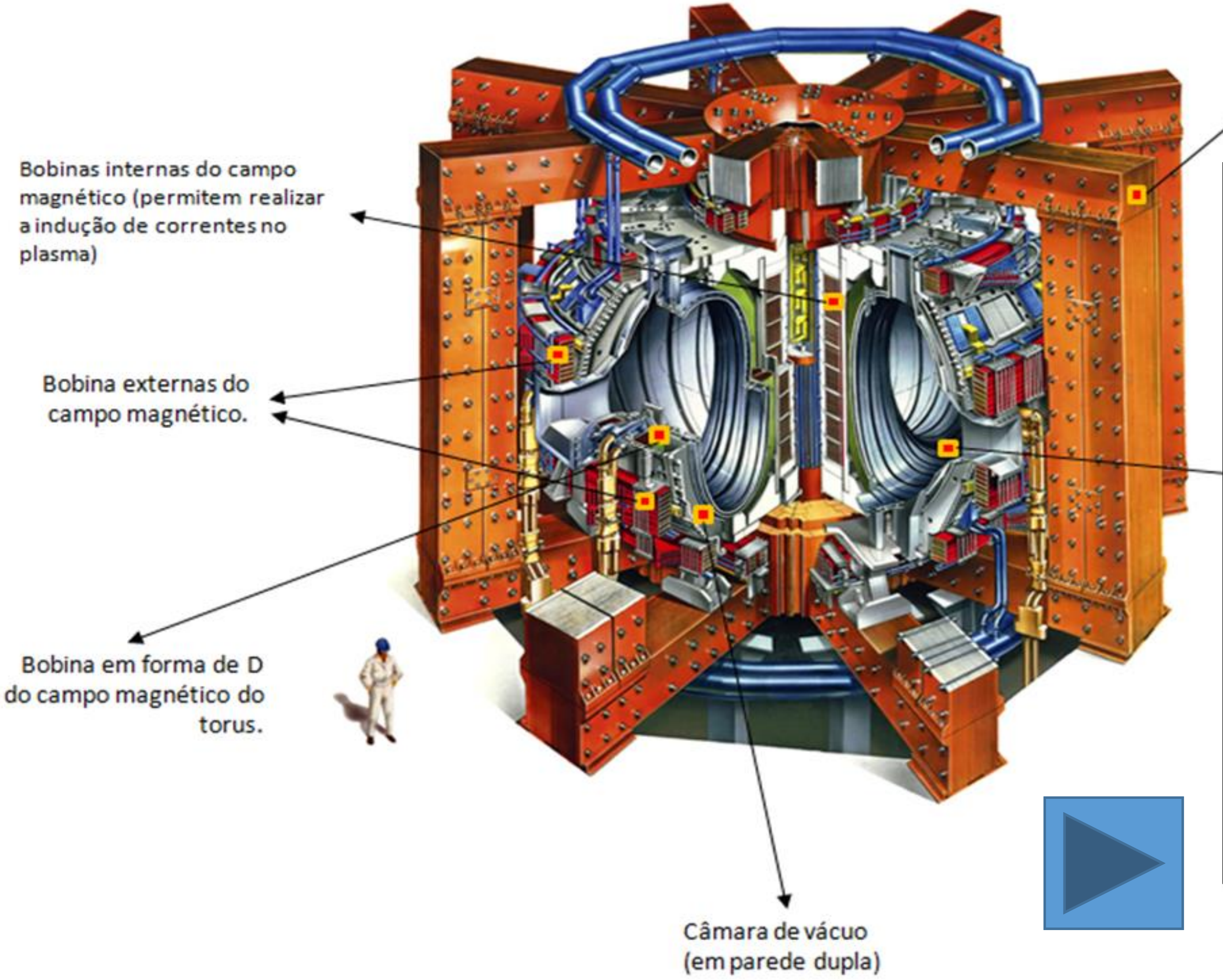
SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes.



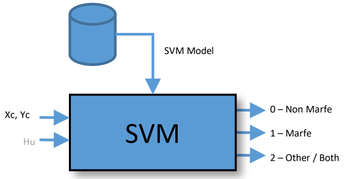


ML: Support Vector Machines – SVM (MARFE)

SVM – MARFE Classification

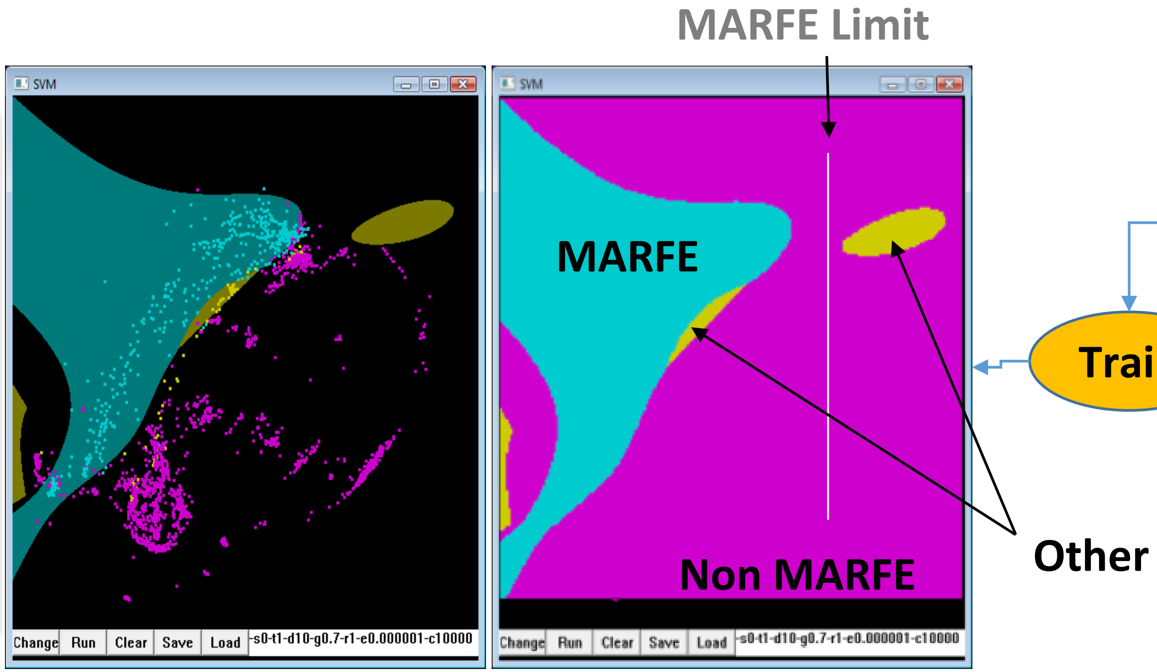


ML: Support Vector Machines – SVM (MARFE)

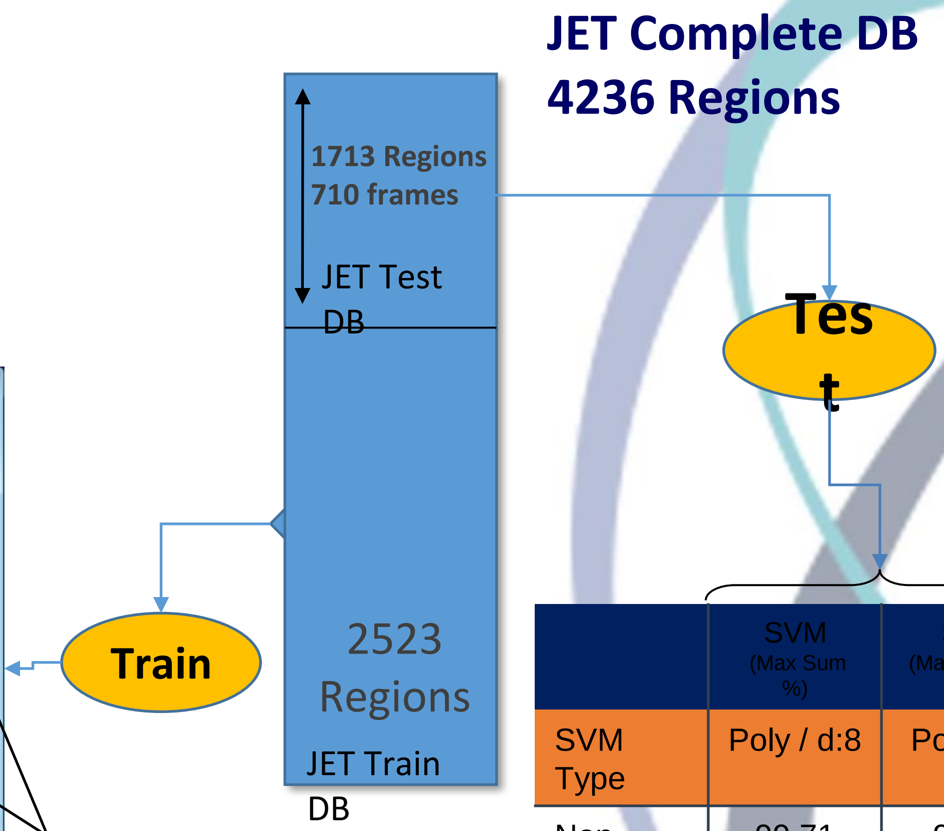


SVM Module

- **LibSVM (in C++)**



SVM decision



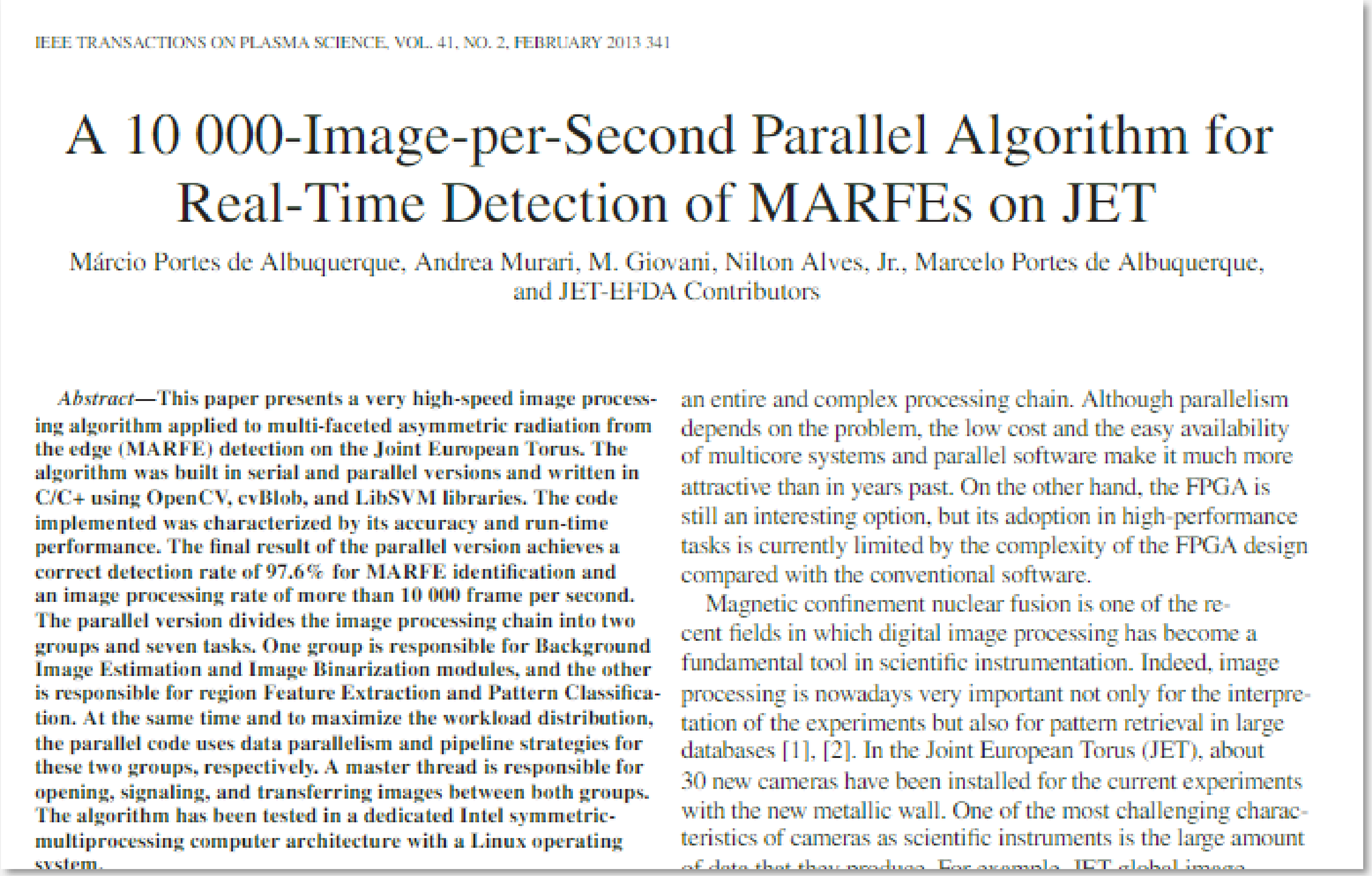
	SVM (Max Sum %)	SVM (Max MARFE %)
SVM Type	Poly / d:8	Poly /d:6
Non Marfe	99,71	93,79
Marfe	93,79	94,14
Other	25,45	14,55
Sum	96,32	96,15

Kernel type: polynomial

Evaluation test for different degree: 6, 8, 10 and 12.

(11018 SVM models were evaluated)

ML: Support Vector Machines – SVM (MARFE)


IEEE TRANSACTIONS ON PLASMA SCIENCE, VOL. 41, NO. 2, FEBRUARY 2013 341

A 10 000-Image-per-Second Parallel Algorithm for Real-Time Detection of MARFEs on JET

Márcio Portes de Albuquerque, Andrea Murari, M. Giovani, Nilton Alves, Jr., Marcelo Portes de Albuquerque, and JET-EFDA Contributors

Abstract—This paper presents a very high-speed image processing algorithm applied to multi-faceted asymmetric radiation from the edge (MARFE) detection on the Joint European Torus. The algorithm was built in serial and parallel versions and written in C/C+ using OpenCV, cvBlob, and LibSVM libraries. The code implemented was characterized by its accuracy and run-time performance. The final result of the parallel version achieves a correct detection rate of 97.6% for MARFE identification and an image processing rate of more than 10 000 frame per second. The parallel version divides the image processing chain into two groups and seven tasks. One group is responsible for Background Image Estimation and Image Binarization modules, and the other is responsible for region Feature Extraction and Pattern Classification. At the same time and to maximize the workload distribution, the parallel code uses data parallelism and pipeline strategies for these two groups, respectively. A master thread is responsible for opening, signaling, and transferring images between both groups. The algorithm has been tested in a dedicated Intel symmetric-multiprocessing computer architecture with a Linux operating system.

an entire and complex processing chain. Although parallelism depends on the problem, the low cost and the easy availability of multicore systems and parallel software make it much more attractive than in years past. On the other hand, the FPGA is still an interesting option, but its adoption in high-performance tasks is currently limited by the complexity of the FPGA design compared with the conventional software.

Magnetic confinement nuclear fusion is one of the recent fields in which digital image processing has become a fundamental tool in scientific instrumentation. Indeed, image processing is nowadays very important not only for the interpretation of the experiments but also for pattern retrieval in large databases [1], [2]. In the Joint European Torus (JET), about 30 new cameras have been installed for the current experiments with the new metallic wall. One of the most challenging characteristics of cameras as scientific instruments is the large amount

SVM (Max MARFE %)
Poly /d:6
93,79
94,14
14,55
96,15

IEEE TRANSACTIONS ON PLASMA SCIENCE, VOL. 41, NO. 2, FEBRUARY 2013 341

Evaluation test for different degree: 6, 8, 10 and 12.

(11018 SVM models were evaluated)

ML: Workflow unSupervised

Unclassified data
- Real Data

Preprocessing
- Data engineering
- Meaningful variables
- Data Cleaning
- Balancing

Training
- Architecture definition
- Data augmentation
- Batch size
- Epochs

Trained Model

Evaluate the Model
- Do we understand the pattern that emerged?
- Are we looking to the data that represents the problem?

Does it make sense?

No

need to go home and rethink my life

Yes

Extract knowledge New insights

Apply to new Data

ML: K-Means (Unsupervised)

- K-Means partition the space into K classes.
- Each point belongs to the cluster with the nearest mean
- Here “nearest” is based on some norm (e.g. Euclidean norm)

4 Classes

dx.doi.org/10.7437/NT2236-7640/2016.01.003
Notas Técnicas, v. 6, n. 1, p. 19–27, 2016

Segmentação Textural de Imagens de Rocha por Microtomografia

Segmentation of Microtomography images of rocks using texture filter

Luciana Olivia Dias*
Centro Brasileiro de Pesquisas Físicas - Rua Dr. Xavier Sigaud 150, Rio de Janeiro, RJ 22290-180, Brasil

Clécio R. De Bom†
Centro Federal de Educação, Tecnológica Celso Suckow da Fonseca, Rodovia Mário Covas, lote J2, quadra J - 23810-000 Distrito Industrial de Itaguaí, Itaguaí, RJ e Centro Brasileiro de Pesquisas Físicas - Rua Dr. Xavier Sigaud 150, Rio de Janeiro, RJ 22290-180, Brasil

Heitor Guimarães,‡ Elisângela L. Faria,§ Márcio P. de Albuquerque,¶ e Marcelo P. de Albuquerque**
Centro Brasileiro de Pesquisas Físicas - Rua Dr. Xavier Sigaud 150, Rio de Janeiro, RJ 22290-180, Brasil

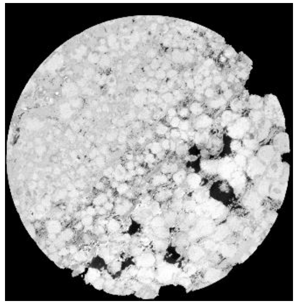
Maury D. Correia†† e Rodrigo Surmas‡‡
Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello - CENPES PETROBRAS, Av. Horácio Macedo, 950, Cidade Universitária, Rio de Janeiro, RJ - 21941-915, Brasil

Submetido: 29/09/2015 Aceito: 10/05/2106

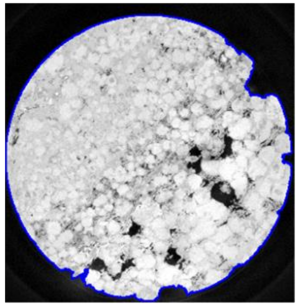
Resumo: A segmentação, realizada de maneira robusta, automatizada e eficiente, de diferentes fases em imagens de microtomografia é um fator crítico e limitador na área de Petrofísica de Rocha Digital. Abordamos a questão partindo de um algoritmo com técnicas baseadas em filtros, obtendo a maximização da Entropia Local para definir um limiar entre fundo e objeto. Validamos a qualidade da segmentação a partir de imagens de amostras de microesferas de vidro, recuperamos o raio das esferas e comparamos a técnica proposta com outros dois algoritmos de segmentação.

AttriTex

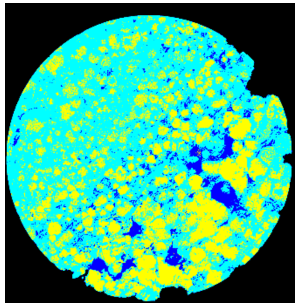
Kmeans with Automatic Contour ROI



Input Image

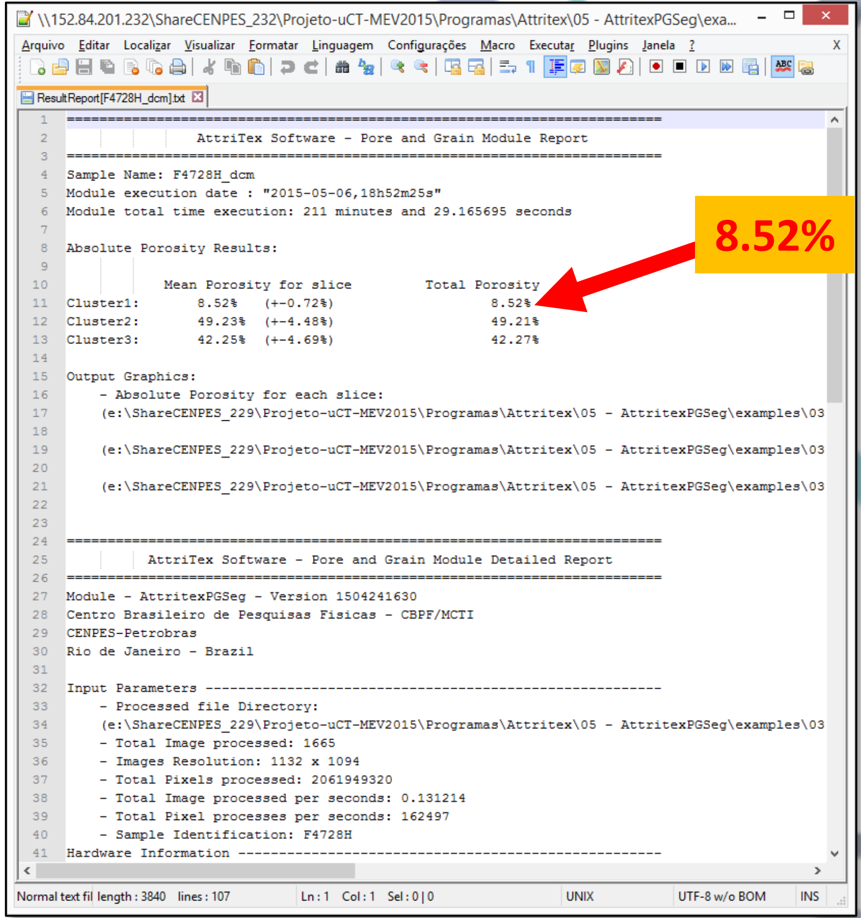


ROI Automatic Contour



3 Clusters Image

- INPUT Image: F4728H (1665 DCM Images) = 1132x1094
- Expected mean porosity by Porosimeter: 8.5%
- ROI – Automatic Contour
- Kmeans with 3 Clusters




```
=======================================================================
              AttriTex Software - Pore and Grain Module Report
=======================================================================
Sample Name: F4728H_dcm
Module execution date : "2015-05-06,18h52m25s"
Module total time execution: 211 minutes and 29.165695 seconds

Absolute Porosity Results:

             Mean Porosity for slice          Total Porosity
Cluster1:        8.52%   (+-0.72%)                 8.52%
Cluster2:       49.23%   (+-4.48%)                49.21%
Cluster3:       42.25%   (+-4.69%)                42.27%

Output Graphics:
    - Absolute Porosity for each slice:
    (e:\ShareCENPES_229\Projeto-uCT-MEV2015\Programas\Attritex\05 - AttritexPGSeg\examples\03

    (e:\ShareCENPES_229\Projeto-uCT-MEV2015\Programas\Attritex\05 - AttritexPGSeg\examples\03

    (e:\ShareCENPES_229\Projeto-uCT-MEV2015\Programas\Attritex\05 - AttritexPGSeg\examples\03


=======================================================================
        AttriTex Software - Pore and Grain Module Detailed Report
=======================================================================
Module - AttritexPGSeg - Version 1504241630
Centro Brasileiro de Pesquisas Fisicas - CBPF/MCTI
CENPES-Petrobras
Rio de Janeiro - Brazil

Input Parameters ------------------------------------------------------
    - Processed file Directory:
    (e:\ShareCENPES_229\Projeto-uCT-MEV2015\Programas\Attritex\05 - AttritexPGSeg\examples\03
    - Total Image processed: 1665
    - Images Resolution: 1132 x 1094
    - Total Pixels processed: 2061949320
    - Total Image processed per seconds: 0.131214
    - Total Pixel processes per seconds: 162497
    - Sample Identification: F4728H
Hardware Information --------------------------------------------------
```

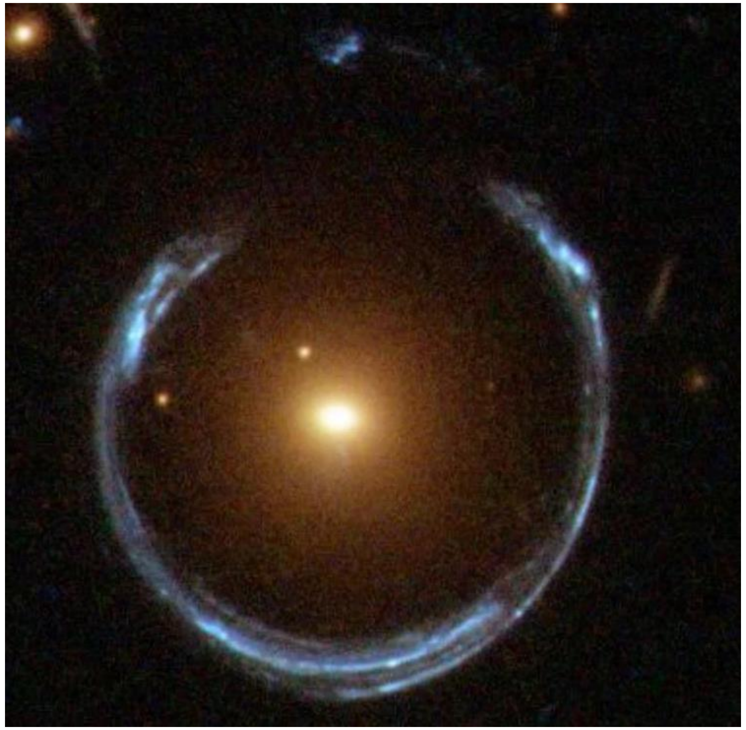

8.52%

Search, Classification and Modelling

The Search ….

Strong Lensing Applications

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



Strong Lensing Challenge 2.0

The Challenge:

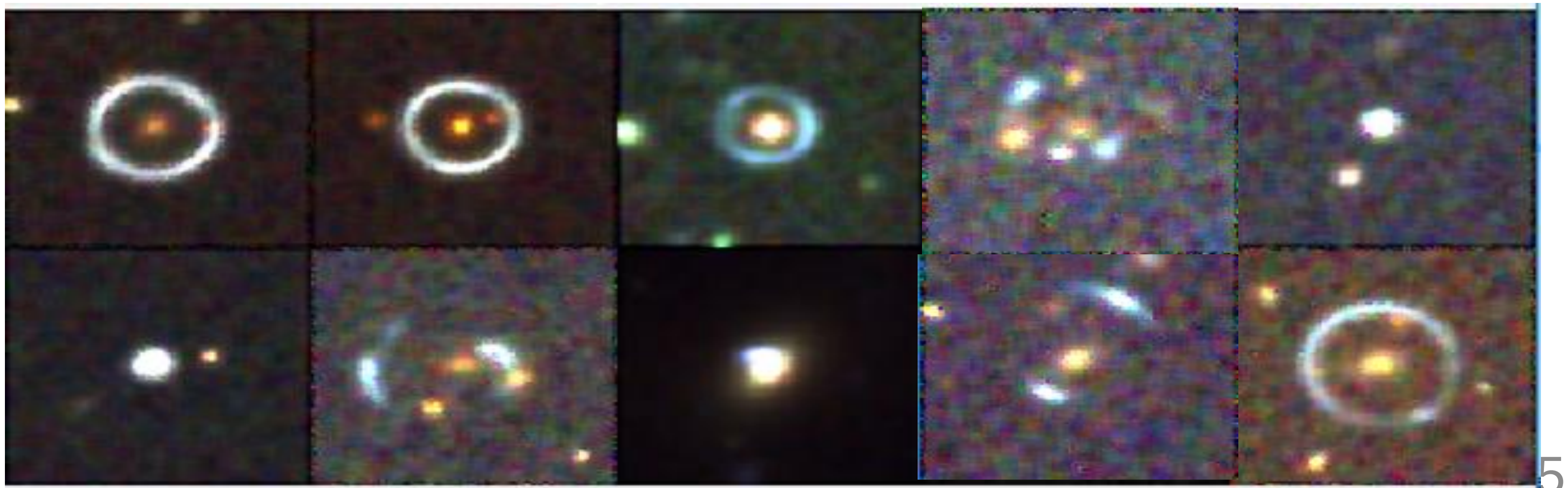
Classify 100k images using up to four channels (VIS, NISP J, Y and H – Euclid-like).

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.

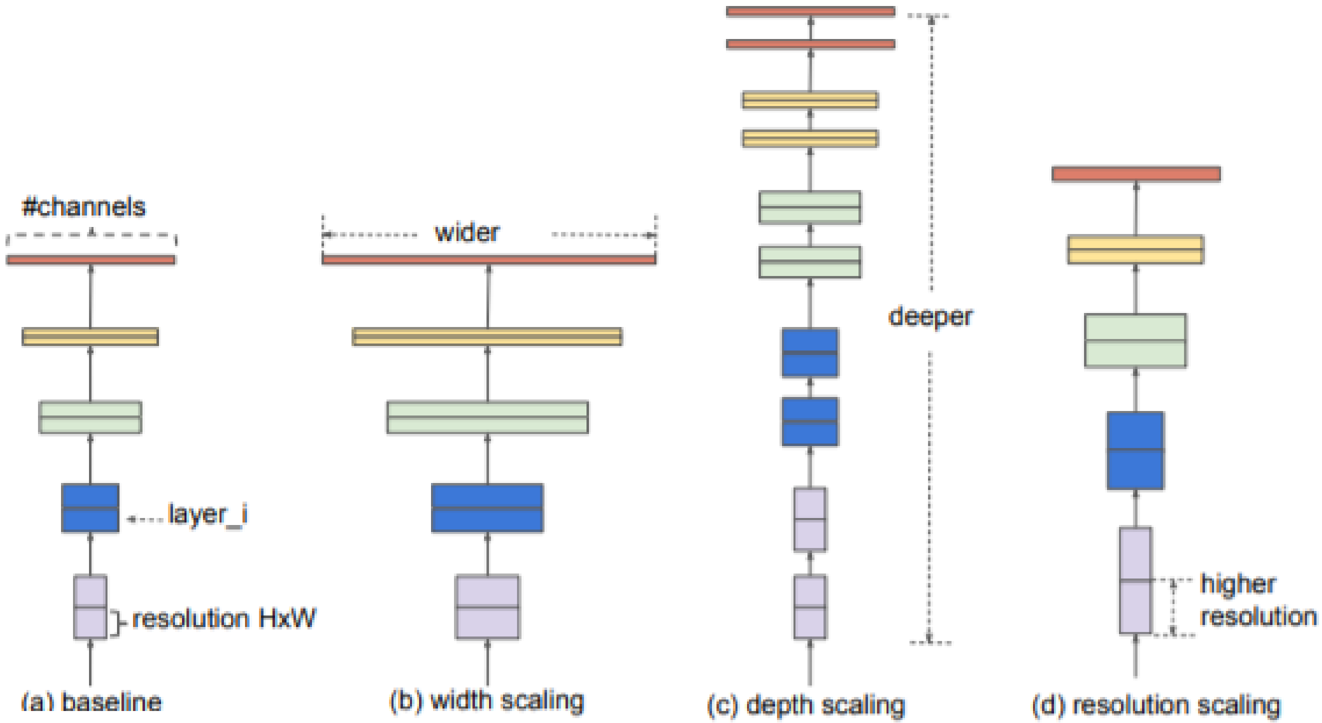
Each team developed different algorithms, mostly based in Deep Learning.

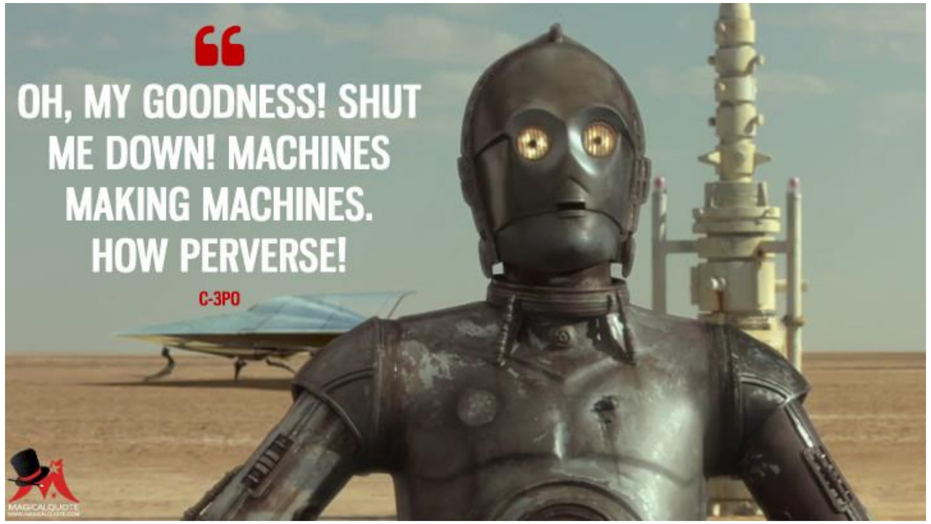
Challenge 1.0 paper : Metcalf et al. 2019 arXiv:1802.03609



EfficientNet Model

Based on compound scaling method is to perform a grid search to find the relationship between different scaling dimensions of the baseline network under a FLOPS constraint.
This determines the appropriate scaling coefficient for each of the dimensions mentioned above.



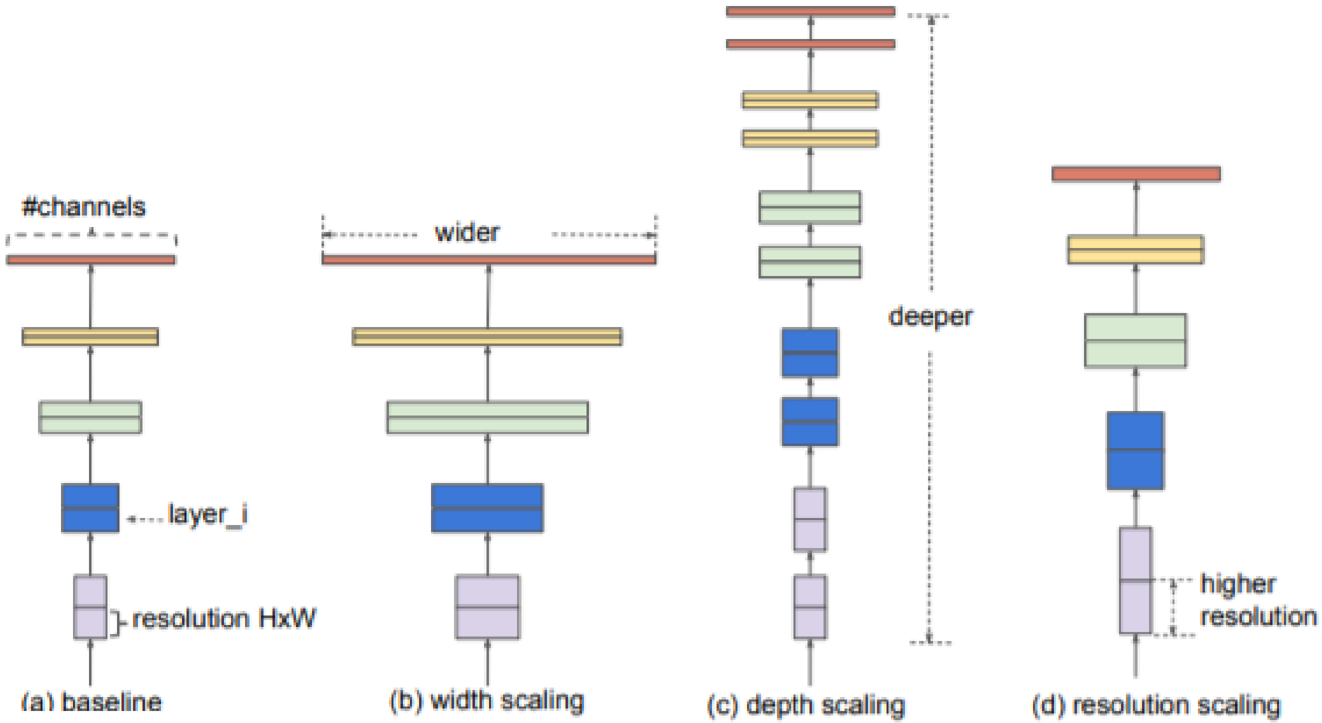


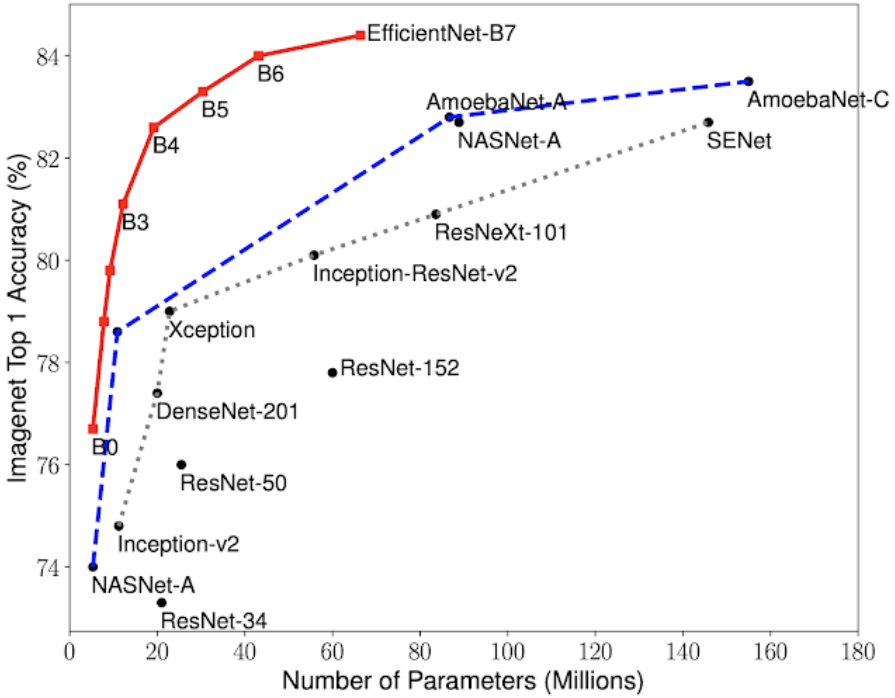
The optimization process is performed by an AutoML algorithm (MNAS).
The method search for performance with low complexity.

EfficientNet Model

Based on compound scaling method is to perform a grid search to find the relationship between different scaling dimensions of the baseline network under a FLOPS constraint.
This determines the appropriate scaling coefficient for each of the dimensions mentioned above.

$$\text{depth: } d = \alpha^{\phi}$$
$$\text{width: } w = \beta^{\phi}$$
$$\text{resolution: } r = \gamma^{\phi}$$





The optimization process is performed by an AutoML algorithm (MNAS).
The method search for performance with low complexity.

	F	Precision	Recall	names	
1	0.993468	0.997855	0.168824	Cast_efficient_auc	Clecio Bom. et al.
2	0.993456	0.996666	0.216973	CNN-napoli-groningen	Gentile F., et al.
3	0.993415	0.997041	0.197052	Cast_custom_res_auc	Clecio Bom. et al.
4	0.992514	0.997135	0.161404	Cast_efficient_f	Clecio Bom, et al.
5	0.992502	0.997300	0.156404	cnnLF	Rui Li et al.
6	0.992460	0.996276	0.188806	Cast_custom_res_f	Clecio Bom, et al.
7	0.992210	0.997031	0.155718	GAMOCLASS2	Diego Tuccillo et al.
8	0.990452	0.996228	0.133105	manchester1	Neal Jackson
9	0.989814	0.995895	0.127168	OU-VIS-JYH-200	Joshua Wilde et al.
10	0.987513	0.992573	0.148197	GAMOCLASS	Diego Tuccillo, at al.
11	0.985876	0.990229	0.167554	;-L-Finding-Challenge-2	Joshua Wilde et al.
12	0.985229	0.994180	0.089544	BarSanCNN-N1	lberto Manjón García, et al.
13	0.984689	0.991219	0.118356	OU-JYH-VIS-66	Joshua Wilde et al.
14	0.982051	0.991908	0.081559	BarSanCNN-N2	lberto Manjón García, et al.
15	0.981997	0.989571	0.103315	LASTRO_FINDER	Elodie Savary, et al.
16	0.981912	0.988461	0.117437	UIUC_ML_Lens_folks	Joshua Yao-Yu Lin, Zehao Jin
17	0.975971	0.989837	0.058916	DeepForkLens-elu	Anna Niemiec, Jonathan Vacher
18	0.975971	0.989837	0.058916	DeepForkLens-maxpool	Anna Niemiec, Jonathan Vacher
19	0.974475	0.988605	0.057726	DeepForkLens	Anna Niemiec, Jonathan Vacher
20	0.956939	0.958788	0.304440	LASTRO_Andrei_Daniel	Elodie Savary, et al.
21	0.942402	0.944213	0.300992	OU-T-SNE	Joshua Wilde et al.
22	0.561878	0.574694	0.021796	NOTCH	Nan Li, et al.
23	0.561878	0.574694	0.021796	NJUPTDeeplens	Nan Li, et al.
24	0.559905	0.559986	0.482188	IIUC_ML_lensing_group	Joshua Yao-Yu Lin, Zehao Jin
25	0.558057	0.557835	1.000000	Cast_cdropout_aug	Clecio Bom, et al.
26	0.558057	0.557835	1.000000	Cast_cdropout	Clecio Bom, et al.
27	0.558044	0.557823	1.000000	NOTCH2	Nan Li, et al.

e

	F	Precision
1	0.993468	0.997855
2	0.993456	0.996666
3	0.993415	0.997041
4	0.992514	0.997135
5	0.992502	0.997300
6	0.992460	0.996276
7	0.992210	0.997031
8	0.990452	0.996228
9	0.989814	0.995895
10	0.987513	0.992573
11	0.985876	0.990229
12	0.985229	0.994180
13	0.984689	0.991219
14	0.982051	0.991908
15	0.981997	0.989571
16	0.981912	0.988461
17	0.975971	0.989837
18	0.975971	0.989837
19	0.974475	0.988605
20	0.956939	0.958788
21	0.942402	0.944213
22	0.561878	0.574694
23	0.561878	0.574694
24	0.559905	0.559986
25	0.558057	0.557835
26	0.558057	0.557835
27	0.558044	0.557823

Lens Finding Challenge 2.0 results

Caixa de entrada ×

Ben Metcalf <RobertBenton.Metcalf@unibo.it>

para Clécio

inglês > português Traduzir mensagem

Dear Clecio,

You won! Congratulations!

I am attaching a presentation I am about to give to the
Euclid strong lensing group on it. There are a few
complications. In particular I think two of you entries where meant just for the Einstein radius contest? Also
I'm interested in how your method that trained on the
AUC does better than the one you trained on the F score.

There were other entries in the Einstein radius competition that I think are not in the right units
or are areas instead of radii. I haven't been able
to figure them out yet.

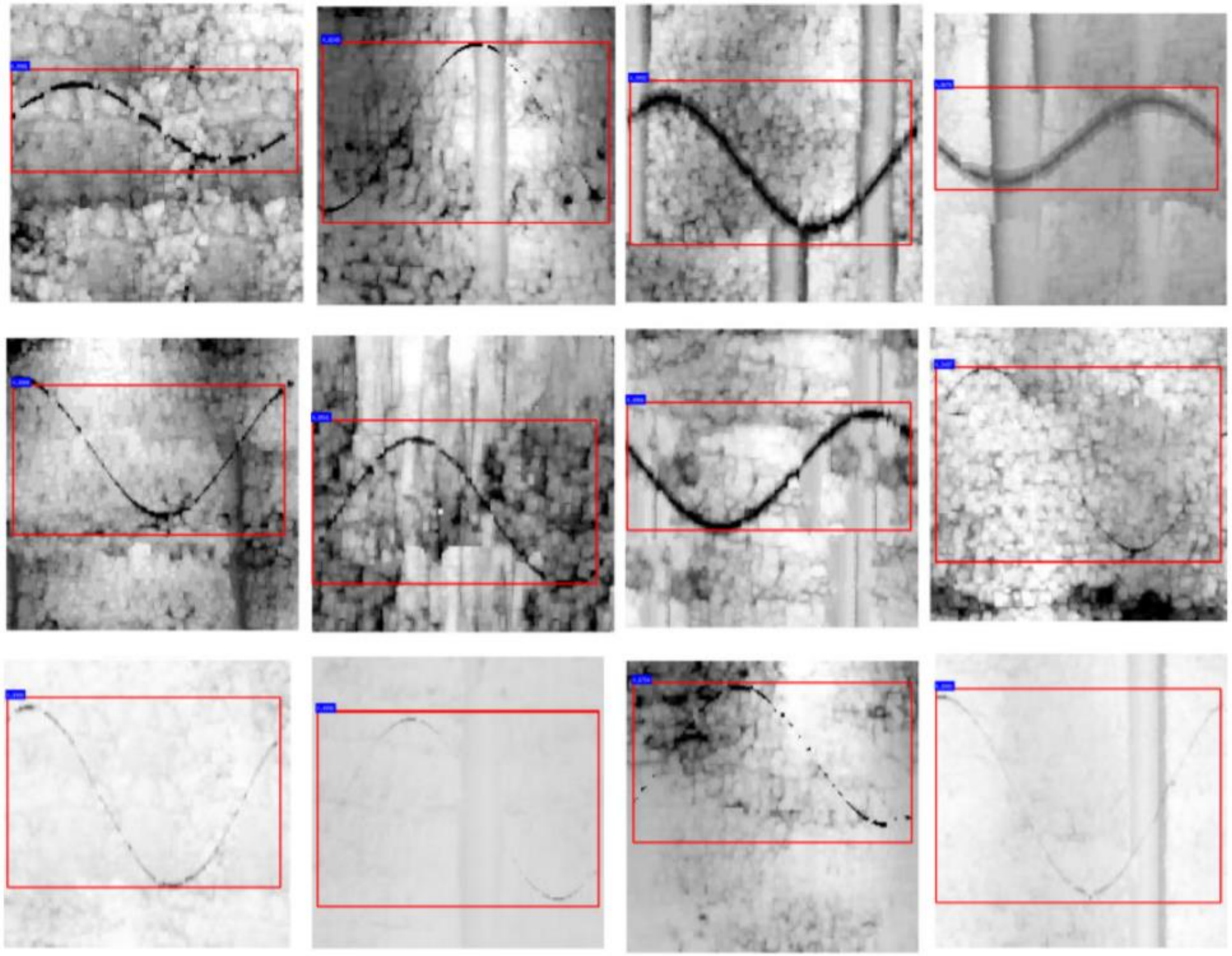
I'll be compiling a more complete report and will get in touch with you about that.

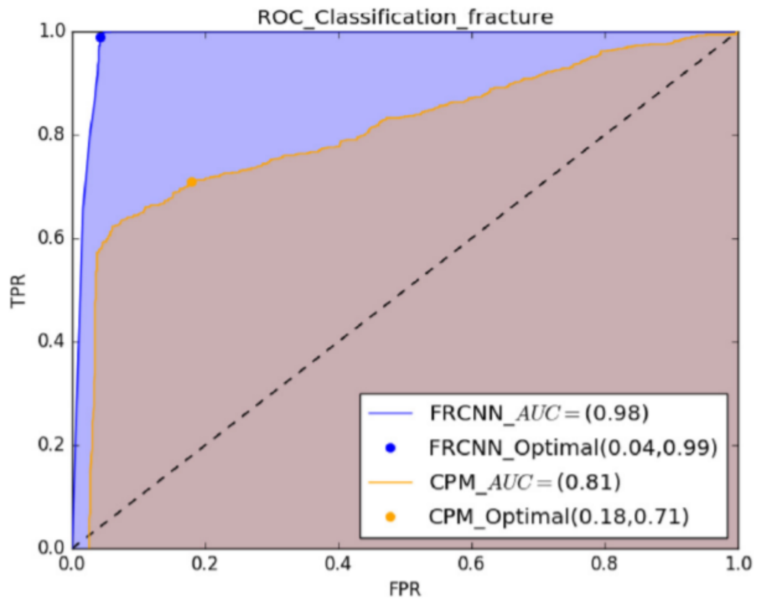
Yours
Ben

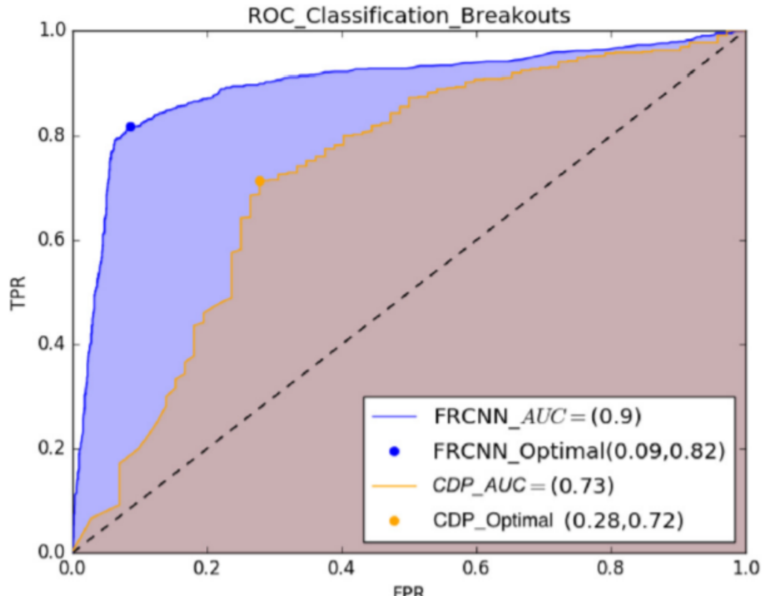
R. Ben Metcalf

DIFA, Dipartimento di Fisica e Astronomia
Alma Mater Studiorum Universita di Bologna
via Gobetti 93/2

Automatic Fractures and Breakouts detection with Deep Learning







Dias et al. 2020

The Classification ….

Galaxy Morphology Classification

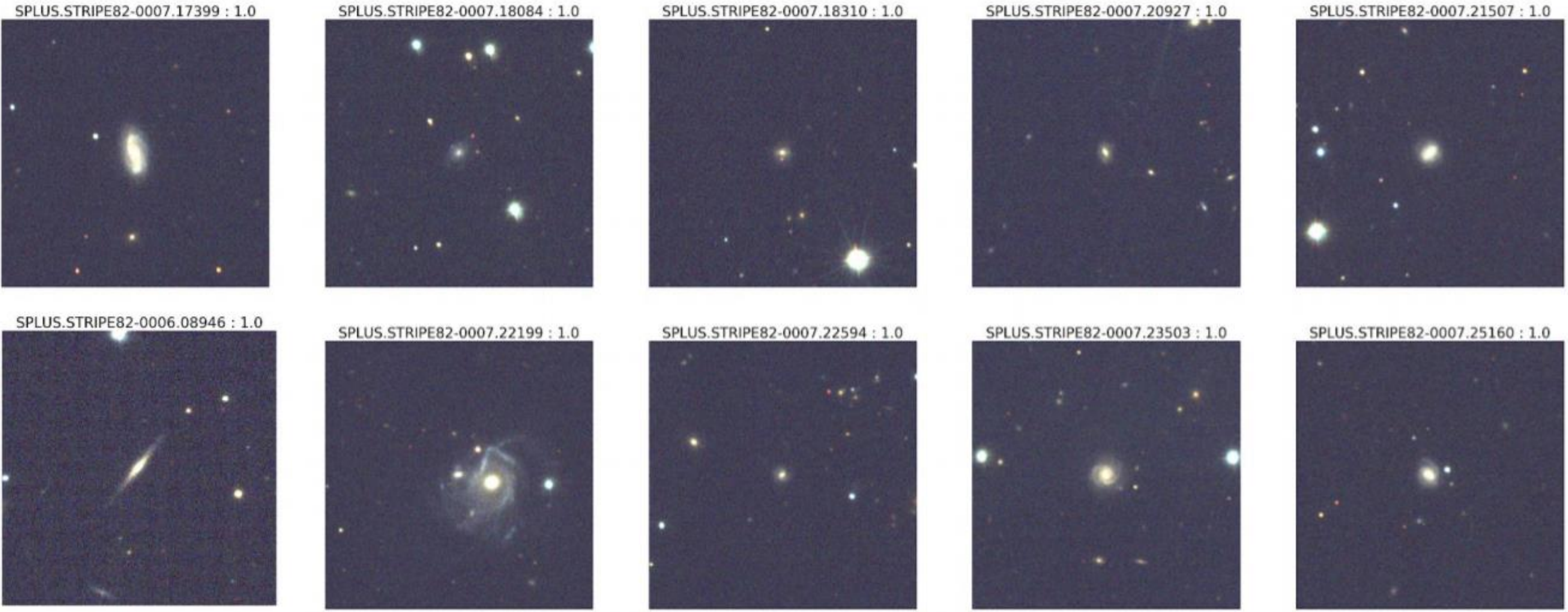
The data was later matched with Galaxy zoo database.

We combined all morphological types and subtypes in 2 major classes Elliptical and Spiral.

In DR1 we have ~8k have galaxy zoo classification (from the 14k in our original sample)

Some GalaxyZoo examples in S-Plus

Spirals



Some GalaxyZoo examples in S-Plus

Ellipticals

SPLUS.STRIPE82-0002.12730 : 1.0

SPLUS.STRIPE82-0002.14076 : 1.0

SPLUS.STRIPE82-0002.14560 : 1.0

SPLUS.STRIPE82-0002.14565 : 1.0

SPLUS.STRIPE82-0002.14682 : 1.0

SPLUS.STRIPE82-0002.15179 : 1.0

SPLUS.STRIPE82-0002.15375 : 1.0

SPLUS.STRIPE82-0002.15847 : 1.0

SPLUS.STRIPE82-0002.16038 : 1.0

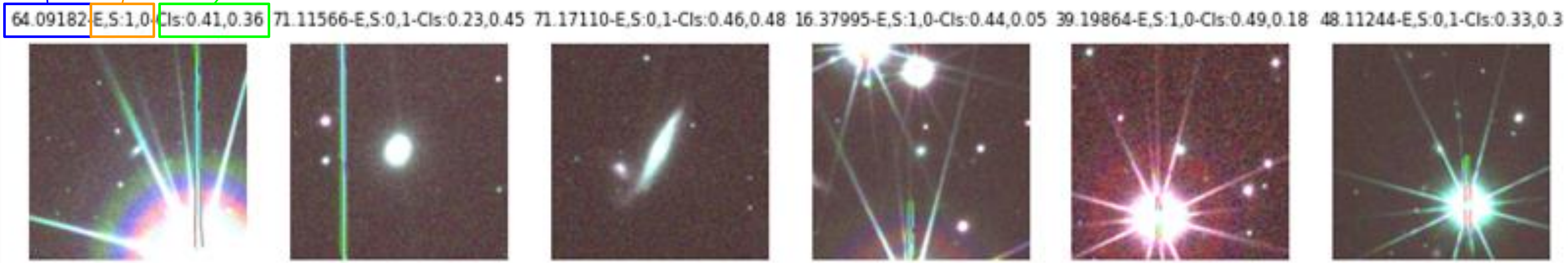
SPLUS.STRIPE82-0002.16265 : 1.0

Low probability of Elliptical AND Spiral

ID

True
E = 1
S = 0

predictions
E,S

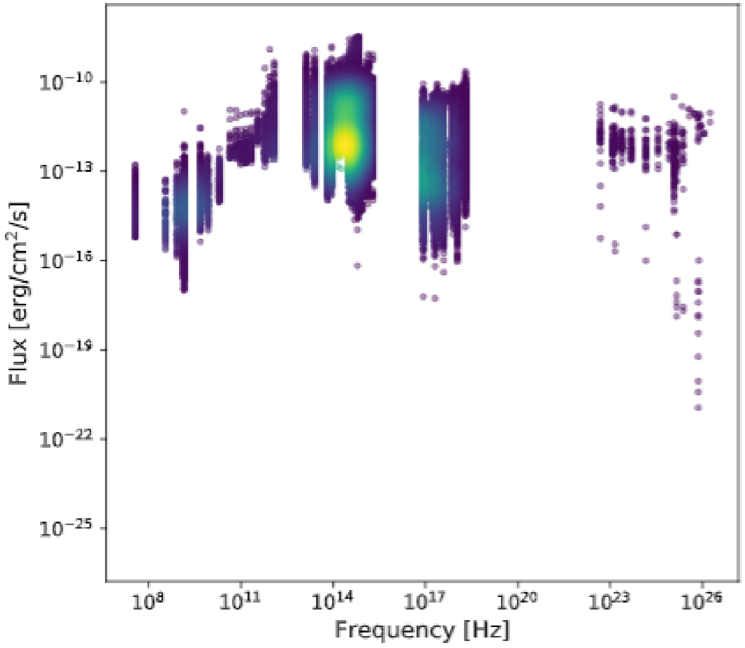


Samples Predictions:
E < 0.5 and S < 0.5

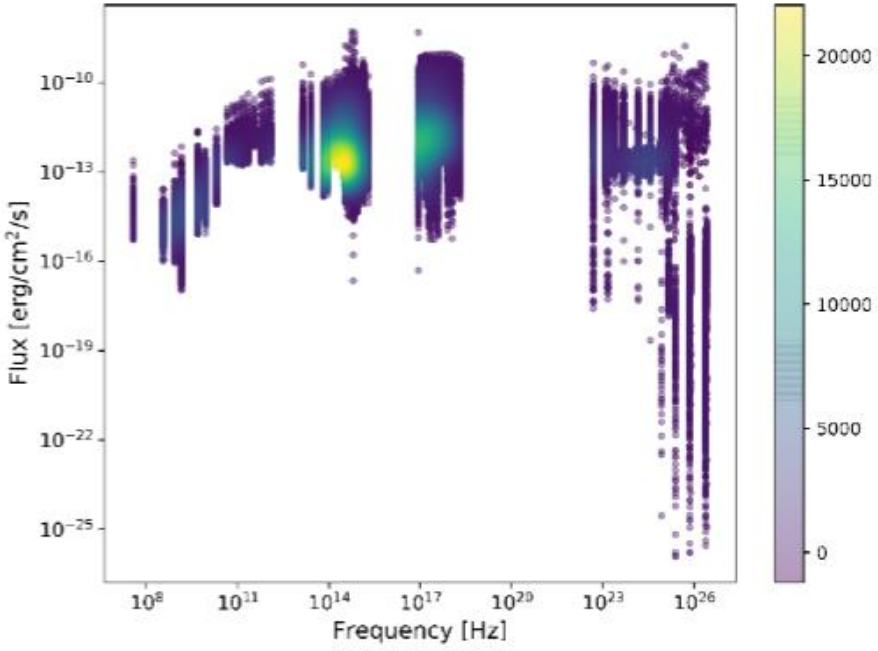
Blazars SED classification

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.

Artist's impression

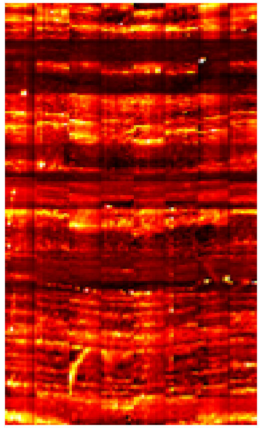


(a) Non-Blazars.

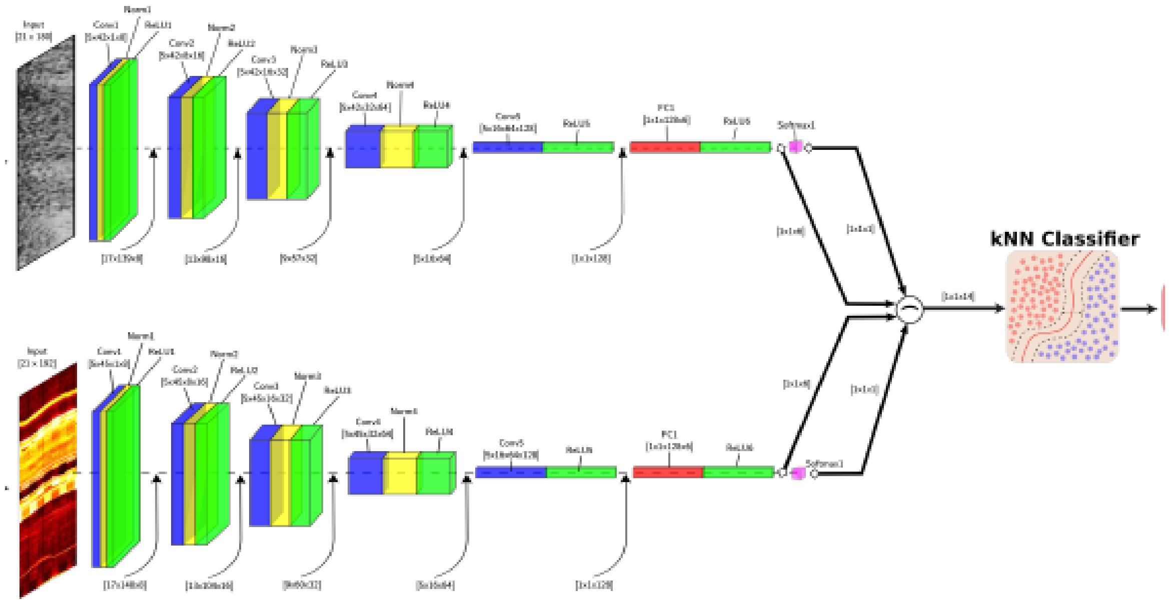


(b) Blazars

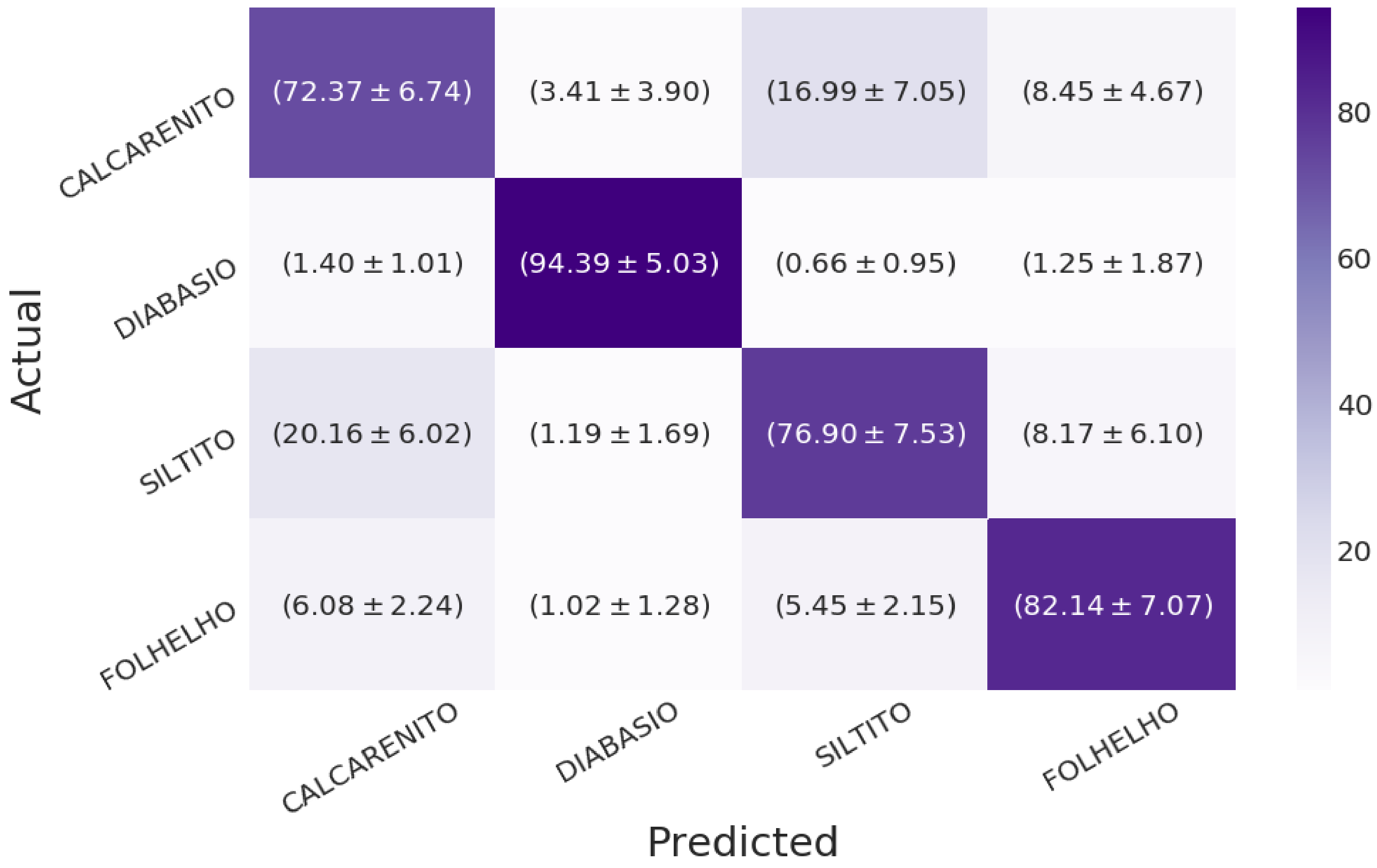
Lithology Classification in pre-salt



We proposed a Deep Learning architecture to make a fast and reliable analysis from the Ultrasound and Resistivity images that can be validated by the geologist.
Blanco-Valentin et al. 2019



Lithology Classification in pre-salt



Valentim et al. 2019

Centro Brasileiro de Pesquisas Físicas

Redes Neurais profundas e aplicações Deep Learning

Clécio Roque De Bom – debom@cbpf.br

CBPF

clearnightsrthebest.com