PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Michele Delli Veneri

Cycle: 35th

Training and Research Activities Report

Year: First

Tutor: Prof. Vincenzo Moscato

Co-Tutor: Prof. Giuseppe Longo

Date: October 21, 2020

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: 35th **Author: Michele Delli Veneri**

1. Information:

- **PhD student: Michele Delli Veneri**
- **DR number: 993895**
- **Date of birth: 06/03/1993**
- **Master Science degree: Physics University: Università di Napoli Federico II**
- **Doctoral Cycle: 35th**
- **Scholarship type: scholarship founded by *EUSTEMA S.p.A***
- **Tutor: Prof. Vincenzo Moscato**
- **Co-tutor: Prof. Giuseppe Longo**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Hardware and Software Infrastructures for Big Data – Data Science	MSc Course	96	12 CFU	08/10/19 – 01/07/2020	Prof. Antonio Picariello	Y
Data Management and Computer Networks	MSc Course	96	12 CFU	08/10/19 – 02/07/20	Prof.ssa Flora Amato	Y
Intelligenza Artificale ed Etica: La ricerca in IA alla prova delle sfide etiche	Seminar	8	1.4 CFU	06/12/19	Prof. Roberto Prevete	Y
Computational Biology: Large scale data analysis to understand the molecular basis of human diseases	Seminar	1	0.2 CFU	09/04/20	Prof. Michele Ceccarelli	Y
Matlab fundamentals	School	20	2 CFU	20/02/20 – 27/03/20	Agostino De Marco, Stefano Marrone	Y
How to Get Published with IEE, Dr. Eszter Lukacs	Seminar	2	0.4 CFU	20/04/20	Dr, Eszter Lukacs, Dr.ssa Alsessandra Scippa	Y

HyCASTLEt Cup Campania 2020 – Innnovation management, entrepreneurship and intellectual property	**Course**	**14**	**5 CFU**	**05/05/20 – 19/06/20**	**Prof. Pierluigi Rippa**	**Y**
Design and Implementation of Augmented Reality Software Systems	**Course**	**20**	**4 CFU**	**05/06/20 – 25/06/20**	**Prof. Fasolino, Prof. Amalfitano**	**N**
Large Scale Training of Deep Neural Networks	**Seminar**	**2**	**0.4 CFU**	**06/05/20**	**Dr. Giuseppe Fiameni**	**N**
Bias From The Wild	**Seminar**	**2**	**0.4 CFU**	**26/05/20**	**Prof. Nello Cristianini**	**N**
SPACE Signal Processing and Computational imagE formation	**Seminar**	**2**	**0.4 CFU**	**20/05/20**	**Saiprasad Ravishankar - IEE**	**N**
Linear regression in Pytorch and Convolutional Neural Networks	**Seminar**	**2**	**0.4 CFU**	**29/06/20**	**Dr. Giuseppe Fiameni**	**N**
Efficient Data Loading using DALI and Mixed Precision Training using Apex	**Seminar**	**1.5**	**0.3 CFU**	**01/07/20**	**Dr. Giuseppe Fiameni**	**N**
Multi-GPU Training using Horovod, Deploying Models with TensorRT and Profiling with NVTX	**Seminar**	**2**	**0.4 CFU**	**02/07/20**	**Dr. Giuseppe Fiameni**	**N**
Machine Learning	**Course**	**20**	**4 CFU**	**06/07/20 – 17/07/20**	**Marco Aiello, Anna Corazza, Carlo Sansone**	**N**
Wearable Brain-Computer Interface for Augmented Reality-based applications in industry 4.0	**Seminar**	**1**	**0.2 CFU**	**29/07/20**	**Prof. Pasquale Arpaia**	**N**
Algorithmic Accountability. Affidabilità e responsabilità degli algoritmi	**Seminar**	**2**	**0.4 CFU**	**24/09/20**	**Antonio Sassano – Fondazione Bordoni**	**N**

IBM Quantum: I primi computer quantistici per la ricerca e la didattica	**Seminar**	**1.5**	**0.3 CFU**	**09/10/20**	**CRUI, Federico Mattei**	**N**
Strategic Orientation for STEM Research & Writing	**Course**	**18**	**3.6 CFU**	**16/07/20 – 17/09/20**	**Prof. Fraser**	**Y**

1) Courses, Seminar, Doctoral School, Research, Tutorship
2) Choose: Y or N

2.1. Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	**0**	**1.4**	**0**	**0**	**1.4**
Bimonth 2	**0**	**0**	**6**	**0**	**7.4**
Bimonth 3	**2**	**0.6**	**7.4**	**0**	**10**
Bimonth 4	**33**	**2.3**	**7.4**	**0**	**42.7**
Bimonth 5	**4**	**0.2**	**7.4**	**0**	**11.6**
Bimonth 6	**3.6**	**0.7**	**7.4**	**0**	**11.7**
Total	**42.6**	**5.2**	**35.6**	**0**	**83.4**
Expected	**30 - 70**	**10 - 30**	**80 - 140**	**0 – 4.8**	

3. Research activity:

3.1. Research on Hybrid Classification Models and Semi-Supervised Classificators

In collaboration with EUSTEMA, I have studied and investigating the main approaches in the literature for hybrid classification issue. In addition, I am developing a novel hybrid classification model capable of achieving accuracies comparable with state of the art models and solving some mayor drawbacks that undermine the majority of literature models.
I began my research by reviewing the current literature on hybrid classification models, a class of algorithmic solutions aimed at combining unsupervised clustering and classification. In order to improve the classification models performances, the basic idea is that unsupervised clustering methods group instances based on their mutual relationship and thus can provide constraints suitable for a classification algorithm. In particular, after reviewing the current hybrid model scenario, it become apparent that they shared the following shortcomings: i) they are based on clustering algorithm that make assumption on the underlying data distribution resulting in a predefined number of cluster or in clusters with an imposed shape; ii) they separate clustering and classification into two independent phases without carrying the information obtained through classification to update the clustering results and iii) they do not consider the effect of noise to the obtained performances. In collaboration with EUSTEM, I developed a new hybrid algorithmic solution named

HyCASTLE (a Hybrid ClAssification System based on Typicality, Labels and Entropy) that tried to solve all three mentioned shortcomings by combining a novel clustering algorithm based on the Typicality, a non-parametric estimator introduced by Angelov et al. 2017, with a cluster aggregation/separation strategy based on both data topology and known labels. A first work concerning some preliminary results has been submitted to Knowledge Based System Journal – Elsevier, under peer review.

As future work, I am going to continue this line of research by applying HyCASTLE to both industrial and scientific use cases and by extending the current architecture to a semi-supervised one. This work could be extremely interesting, not only for the widespread interest of the community to this type of models, but also because the HyCASTLE ability to perform a completely data-driven clustering and to measure the likelihood of correctness of its prediction on unseen instances through a probabilistic measure, could be used to implement a self-supervised self-labelling method able to solve both main issues of such models (semi-supervised models), i.e.: lack of sufficient initial labelled data and inability to deal with non-spherical datasets.

3.2. Research on Signal Analysis and Compressed Sensing in Astronomical Imaging through Deep Learning

During the first year of my PhD, an important aspect was the application of Deep Learning methodology to Compressed Sensing problems in the context of Astrophysics.
The choice of this field arisen from my background studies and by considering that in the foreseeable future several large astrophysical collaborations and instruments, such as TOLIMAN, ALMA, SKA and ngVLA, will require an intensive use of machine-learning based tools. In fact, these projects are characterized by complex and mostly unresolved problems present in their data (background estimation and subtraction, detection of point-like and extended emissions, weak signal detection, image segmentation and classification). Furthermore, their focal plane instruments require efficient and automatic solutions to handle and explore the enormous amounts of produced data (hundreds of TBs per day).
Therefore, large part of my work was dedicated to the study of the current literature of compressed sensing related techniques (classical statistical methods and machine-learning) such as MMSE, MAP, MEM, deconvolution algorithms, sparse dictionary learning, Bayesian mixture models, Image compression and denoising, Convolutional Neural Networks, Autoencoders, Variational Autoencoders and, given my background, on filling my knowledge gap on Big Data handling and databases. This study was carried out either by reading the current literature and by following specific courses such as those given by Prof. Picariello and Prof. Amato.
While expanding my knowledge on the subject, I had the opportunity to work on the TOLIMAN challenge; the challenge aims to characterize the astrometric signal embedded in a time series of simulated images of the diffractive response of the TOLIMAN telescope to

the light coming from the α-Cen system. The main difficulty of the challenge lays in the amplitude of the signal that is, on average, 10^{-6} time smaller than the image pixel dimension and the presence of several noise components whose amplitude is orders of magnitudes higher than the signal itself. To solve this challenge, I developed a custom variational autoencoder with a periodic prior on the latent space that was able to recover the period of the embedded signal. This work was reviewed and accepted as a book chapter in the book *Intelligent Astrophysics* to be published in the book series *Emergence, Complexity and Computation* by Springer. I will continue to work specifically to address the increasing complexity of the problem the more noise components are injected in the simulations. Given the recent involvement of my research group (Prof. Moscato, Prof. Longo, Dr. Brescia) to the European Southern Observatory (ESO) proposal for an internal development study of the data acquirement pipeline of the Atacama Large Millimeter/submillimeter Array (ALMA), I'm currently studying the application of the aforementioned techniques to Interferometric datacubes acquired in the radio band of the ECM spectrum in order to develop a ML based method not only capable of reaching the required accuracies obtainable with state of the art statistical methods but also capable of handling the data volume and speed rate needed by the pipeline. At the beginning of the second year, I will start to apply the knowledge acquired in my first PhD year to data provided by ALMA and increase my knowledge about signal recovery and compressed sensing by following the Information Theory and Digital Signals Elaboration given by the Prof. Tulino.

4. Research products:

Delli Veneri, Michele; Picariello, Antonio; Cavuoti, Stefano; Brescia, Massimo; Sperlì, Giancarlo; Moscato, Vincenzo; Abruzzese, Roberto; Longo, Giuseppe; (2020). *HyCASTLE: a Hybrid ClAssification System based on Typicality, Labels and Entropy*. Submitted to Knowledge Based Systems, Elsevier.

Delli Veneri, Michele; Desdoigts, Louis; Schmitz, Morgan A.; Krone-Martins, Alberto; Ishida, Emille E. O.; Tuthill, Peter; de Souza, Rafael S.; Scalzo, Richard; Brescia, Massimo; Longo, Giuseppe; Picariello, Antonio; (2020). *Periodic Astrometric Signal Recovery through Convolutional Autoencoders*. To appear in the peer review Volume "Intelligent Astrophysics" of the series "Emergence, Complexity and Computation", Book eds. I. Zelinka, D. Baron, M. Brescia, Springer Nature Switzerland, ISSN: 2194-7287, eprint arXiv:2006.13905.

Carlo Donadio, Massimo Brescia, Alessia Riccardo, Giuseppe Angora, Michele Delli Veneri, Giuseppe Riccio; (2020). *New perspective analysis to classify Earth and extraterrestrial drainage networks by Deep Learning.* Submitted to Springer Nature.

5. Conferences and seminars attended

6. Activity abroad:

7. Tutorship

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: 35th **Author: Michele Delli Veneri**