

PERSONAL INFORMATION

Alessandro Quattrociochi

EDUCATION

Master of Science in Data Science

La Sapienza, University of Rome, Italy

Bachelor of Science in Electronic Engineering

La Sapienza, University of Rome, Italy

Thesis Title – *"Remote sensing systems and machine learning methods for natural hazards monitoring and land management"*

WORK EXPERIENCE

April - September 2021

Data Scientist Intern

CINECA, Casalecchio di Reno Bologna

This internship project aims at researching tumor markers by exploiting deep neural networks for cancer classification from genotypic expressions of different patients. Moreover, following an activation maximization approach of each class, different pathways have been tested and enriched. The project has been carried out under the supervision of Dr. Silvia Gioiosa and Dr. Bhaskar Agarwal.

SKILLS

Python – Proficiency with tools such as **TensorFlow** and **Pytorch** for developing for the development of machine learning models related to computer vision.
– Good knowledge of the **Spektral** tool, based on TensorFlow, for the implementation of Graph Neural Networks.
– Good knowledge of the NLP Google's tool **BERT** for context classification and sentiment analysis.
– Proficiency with **Captum**, a PyTorch based tool, for deep learning model interpretation and debugging.
– Advanced use of the **NetworkX** and **iGraph** libraries for the analysis and modelling of complex models represented by graphs, based on SIR, independent cascade, linear/non-linear preferential attachment model and configuration model.

R - Studio – Statistical modelling.
– Graph Theory with iGraph.
– Bootstrapping and Hypothesis Testing.

Data Managment – MySQL
– Neo4j

External Certifications – **Cisco Networking Academy** IT Essentials, PC Hardware and Software
– **Cisco Networking Academy** CCNA Discovery, Networking for Home and Small Businesses

PROJECTS PORTFOLIO

Bioinformatics **Biological Interpretability of a Deep Neural Network for Cancer Type Prediction**

This work presents some results obtained by exploiting different applications of neural networks for tumors classification using data from The Cancer Genome Atlas Program (**TCGA**) of the National Cancer Institute (**NIH**). In particular we have deepened the interpretability, the pruning and the reduction of a neural network through the analysis of the weights distributions. Exploiting the properties of the computational graphs, we determined the paths most used by the network. Finally, using the "Bioconductor" package i performed the enrichment of the pathways to validate the predictions of the models ([GitHub](#)).

Computer Vision **Learn To Count Everything...Better**

The proposal for this computer vision project, was to count object in images by using pre-trained models and density estimation. We first aim at reducing the gap between training and test error by introducing regularization techniques such Batch Normalization, Dropout and Data Augmentation. In addition, to enhance the behaviour of the model, we proposed to use different ImageNet pre-trained models (i.e. VGG16) as an alternative for ResNet50. As final novelty, we implemented an ensemble method by combining ResNet with YOLO to produce a model that outperforms the current state-of-the-art work ([GitHub](#)).

Bioinformatics **Comparative assessment of disease gene prediction algorithms**

In this project, I interviewed different methodologies to investigate protein-protein interactions and its' role in the disease gene association, with the aim of finding the best predictor algorithm to extract disease genes, scoring them by using different evaluation metrics such as precision, recall, F1 score and nDCG. As a final result, we investigated the most suitable disease as Autism Spectrum Disorder and the most powerful algorithm as DiaBLE with further enrichment analysis. ([GitHub](#)).

Graph Neural Network **Experimenting with continual learning**

This work shows the use of a graph convolutional network (GNN) for classification in the Continual Learning scenario. In particular, using the elastic weight consolidation (EWC) technique and the Fisher information Matrix, I split the model inference over multiple independent classes ([GitHub](#)).

Language Processing **Content Classification**

In this project we used some text processing combined with the SentenceTransformer library and the *paraphrase-distilroberta-base-v1* to classify sentences. Our project achieved an accuracy of **71.42** and was ranked sixth in the Kaggle Competition ([GitHub](#)).

PERSONAL SKILLS

Mother tongue Italian

Other languages	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	B2	B2	B2	B2	B1

In compliance with the Italian legislative Decree no. 196 dated 30/06/2003, I hereby authorize you to use and process my personal details contained in this document.