# **Brunella D'Anzi** - Data Scientist

### Personal Information


**Date of birth:** 03/03/1997 — Age 27
**Place of birth:** Matera (MT), Italy
**Nationality:** Italian
**Pronouns (Gender):** She, her, hers (Female)
**Home:** Matera (MT), Italy

**E-mail:** brunella.danzi@cern.ch
**Website:** https://bdanzi.github.io
**Other websites:**
**Mobile Phone:** (+39) 328 71 40 953
**Institutes:** University *Aldo Moro* and INFN Bari


### Summary


Curious and driven woman with a passion for data analysis and data science. I utilize statistics and (deep) machine learning techniques to derive insights from data to make them public.


### Education

**CERN, University of Bari and National Institute of Nuclear Physics - INFN (*Nov 2021–Present*)**
**Position:** Ph.D. in Experimental Particle Physics with 3-years Scholarship
*Advanced techniques for track reconstruction and search for non-resonant production of double Higgs bosons decaying into muon pairs and two b-jets in pp collisions with the CMS experiment at the LHC*

**Music Conservatory Niccolò Piccinni of Bari (*2016–21st March 2022*)**
**Position and Grade:** B.Sc. in Violin, 110/110 with honors
**Dissertation's title (Italian language):** Comparison between the genius of the physicist Albert Einstein and the genius of the musician Ludwig Van Beethoven

**University of Bari and INFN (*2019–26th October 2021* and *2016–30th September 2019*)**
**Positions and Grades:** M.Sc. in Nuclear, Subnuclear and Astroparticle Physics, and B.Sc. in Physics, 110/110 with honors
**M.Sc. Dissertation's title (English language):** Search for double Higgs events produced via a vector boson fusion mechanism in the decay channel $b\bar{b}4l$ with the CMS experiment at the LHC.
**B.Sc. Dissertation's title (Italian language):** Two quantum-states systems

### Computer Skills: Operating Systems, Programming Languages, Scientific libraries and Frameworks

OSX – WINDOWS – BASH – C/C++ – PYTHON3/2 – GOOGLE COLAB – NUMPY – PANDAS – SCIKIT-LEARN – TENSORFLOW – KERAS – MATPLOTLIB – SPARK HADOOP – GIT – OFFICE – LaTeX– MATLAB – ROOT – HTML – CSS – JAVASCRIPT

## Languages

ITALIAN (Mother tongue) – ENGLISH (Fluent) – FRENCH (Basic)

## Certificates

*"Machine Learning", Stanford University* – *"Addressing Challenges by Machine Learning", HSE University* – *2nd Terascale School of Machine Learning, DESY* – *13th CMS Heterogenues Architectures Hackathon* – *CMS Data Analysis School (CMSDAS) 2022* – CERN Webfest Annual Hackathon 2021 – *ESCAPE Summer School on Data Science for Particle Physics* – *Online Terascale Statistics School 2020 , PHYSTAT* – *"The complete SQL Bootcamp"* – *"Data Week April 2023"* – *"Boolean Coding Week 2022"* – *"Programming Foundations: JavaScript, HTML and CSS",Duke University* – *"Learning Cloud Computing: Core Concepts"* – *"Python:Programming Efficiently"* – *"Certified Peer Reviewer Course", Elsevier* – *Public speaking and scientific writing* – *"Technical Writing", Moscow Institute of Physics and Technology* – *"Interpersonal Skills"* – *"Solving Problems with Critical and Creative Talking"*.

## Professional Experience

**Research work for the CMS experiment. (*October 2021–Present*) — CERN**
Development of geometrical cuts autotuning techniques for Heterogenues Architectures (GPU,CPU)-based algorithms at the CMS tracker. Multi-Objective Particle Swarm Optimization (MOPSO) technique is currently exploited with a fake rate tracks reduction of 40% and efficiency increase of 20%. Also, I am currently perfoming datasets (size of petabytes per each year of CMS data taking) filtering, cleaning and statistial inference exploiting supervised Machine Learning algorithms to increase the sensitivity to double Higgs signal signatures, and using the state-of-the-art tagging Deep Learning tools (ParticleNet) for a better jets identification.

**Data Science project on Cluster counting particle identification technique datasets for future collider experiments at FCC. (*October 2021–Present*) — CERN**
I am developing peak finding algorithms for time series. A Deep Learning reconstruction algorithm has been presented at Norfolk (US) along with Beijing Institute of High Energy Physics.

**Data Analyst periodical task for Tracking efficiency performance evaluation using the data-driven Tag and Probe technique. (*October 2021–Present*) — CERN**
A framework has been developed to speed up the computation during fitting procedures on each physics quantities histogram bins using Apache Spark. Misalignment and miscalibration effects has been spotted in high pseudorapidity ($\eta$) muon system regions thanks to that.

**Supervisor during 2023 CERN Summer Student Program. (*June 2023–September 2023*) — CERN**
Development of python-based analysis framework within a Dask and JupyterHub environment, the INFN CMS Analysis Facility. The usage of a HEP python-based library bamboo in collaboration with UC Louvan institute, the data processing has been speed up of 10 times.

**Data Analyst for Innovation 4 Change Program: Identification of Lithium Batteries in waste streams for Iren Challenge. (*Febr–July 2023*) — Fondazione Agnelli, Turin and CERN**
Data Visualization to set-up a business startup, examination of competitors like the machine learn-

ing solution developed by Netherlands CORE Team. The idea had to increase the total collections of well-classified battieries in Italy of 30%.

**Tutor in INFN Machine Learning Hackathons. (*21–23 June 2023* and *13th–15th December 2021*)**
Hackathon open to INFN members and associates passionate in machine learning techniques, starting from hypothesis tests, supervised and unupervised use-cases. An example made public on INFN Confluence webpage can be found here.

**Lead Contact for Data Analysis tutorials at *CMS Data Analysis Schools 2023*. (*Jan 2023* and *June 2023*) — Fermilab, Chicago (US) and CERN**
CERN (Fermilab) CMS Data Analysis School are the first opportunities for European (American) MSc. and PhD students to get familiar with the CMSSW (C++) framework to perform data analysis and go for the publication procedure within the CMS Collaboration. A tutorial example can be found here.

**Tutor at the *Programming workshop: Learn Python and build your first website* in the *Django Girls Geneva Event*. (*21–22 April 2023*) — CERN**
The Django Girls workshop offers a crash course in computer programming for beginners aged 15 and over. The aim is to introduce digital technologies to the uninitiated, especially women, who are underrepresented in this field.

**INFN Scholarship: Gaseous detectors/Assembly and Test of a Fast timing Micropattern gaseous detector (FTM). (*July 2021–October 2021*) — CERN**
Diamond Like Carbon (DLC) has been introduced for my prototype, a very promising material of which the resistivity can be produced at will. After my studies, new optimized material composition have been used for further R&D on prototypes.

## Fellowships and Awards

- Selection as one of the McKinsey Forward Program participants **(*July 2024*)**
- Selection as one of the McKinsey Next Generation Women Leaders EMEA participants **(*April 2024*)**
- Member of Nova Top Talent Community **(*December 2023*)**
- Top Voice and Articles for LinkedIn Machine Learning, Statistics and Research **(*August 2023–Present*)**
- Collège des Ingeniéurs Innovation 4 Change (I4C) Program **(*1st Febrary 2023–4 July 2023*)**
- INFN CERN Doctoral Student Fellowship **(*1st March 2023–Present*)**
- 2022 CMS Data Analysis School 3rd Prize Team **(*4th–15th January 2022*)**
- Best Poster Awards at International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT) 2021 **(*29th November 2021–3rd December 2021*)**
- INFN 3-months Scholarship at CERN **(*26th July 2021–3rd November 2021*)**
- Winner of INFN "Scienza per tutti" competitions **(*14th December 2017–14th September 2021*)**

## Talks and Publications

I am a CMS and CEPC (Circular Electron Positron Collider) Collaborations author, I published many articles and conference proceedings (more than 80). Please refer to either my Orcid or website to access the complete list. I presented my work in more than 20 events around the world (China, US, France, Switzertland, Italy) in the last two years. Please refer to my website for the complete list.

**Driving Licence**

Category B

**Other Interests**

I am an official Guide for public at CERN, member of INFN Bari Local Outreach Editorial Committee, tutor in the Save The Children Education Program. I am also member of CMS Young Scientists Committee and I moderate discussion sessions during *Large Hadron Collider (LHC) Job Matching* and *Mentoring* programs. I love reading classic books and playing viola and piano beyond the violin studies.

I authorize the processing of my personal data pursuant to Legislative Decree 30 June 2003, n. 196 "Code regarding the protection of personal data".

**Date** **Signature**

September 11, 2024

**All Other Publications**

All other publications are listed here.


[1] B. D'Anzi, *Signal/background discrimination for the VBF Higgs four lepton decay channel with the CMS experiment using Machine Learning classification techniques*, Last accessed 13th November 2022, April 2021, https://confluence.infn.it/pages/viewpage.action?pageId=53906361.

[2] G. Chiarello et al., "Implementation of the Cluster Counting and Timing technique on FPGA for the reduction of transferred data and stored information", Nucl. Instrum. Meth. A **1045**, 167542 (2023), doi:10.1016/j.nima.2022.167734.

[3] C. Caputo et al., "Particle identification with the cluster counting technique for the IDEA drift chamber", in 15th Pisa Meeting on Advanced Detectors: Frontier Detectors for Frontier Physics, Under Review (Nov. 2022), arXiv:2211.04220.

[4] C. Caputo et al., *Cluster counting algorithms for particle identification at future colliders [Poster]*, Presented at ACAT2022, Oct. 2022, doi:10.13140/RG.2.2.17515.62243.

[5] B. D'Anzi et al., *CMS tracking performance in Run 2 and early Run 3 data using the Tag-and-Probe technique [Poster]*, Presented at ACAT2022, Oct. 2022, doi:10.13140/RG.2.2.30937.39525.

[6] B. D'Anzi et al., "Signal to background discrimination for the production of double Higgs boson events via vector boson fusion mechanism in the decay channel with four charged leptons and two b-jets in the final state at the LHC experiment", in ACAT2021, Accepted by the Conference Editor (Sept. 2022), arXiv:2209.11649.

[7] M. I. Abdulhamid et al., "Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method", Eur. Phys. J. C **82**, 36 (2022), doi:10.1140/epjc/s10052-022-09997-1, arXiv:2112.10465.

[8] "CMS Tracking performance in Early Run-3 data using the tag-and-probe technique", (2022), https://cds.cern.ch/record/2839918.

[9] F. Cuna et al., "Particle identification with the cluster counting technique for the IDEA drift chamber", PoS **ICHEP2022**, 335 (2022), doi:10.22323/1.414.0335.

[10] B. D'Anzi, "Search for double Higgs events produced via a vector boson fusion mechanism in the decay channel $b\bar{b}4l$ with the CMS experiment at the LHC", Presented 26 Oct 2021 (2021), https://cds.cern.ch/record/2788946.