José Guilherme de Almeida, PhD

Data scientist and ML researcher with 5+ years of experience with clinical data

they/them

jose.gcp.almeida@gmail.com
jose.almeida@research.fchampalimaud.org
github.com/josegcpa
josegcpa.github.io

Work and research positions

2022 → present — **Clinical AI Researcher and developer**

Champalimaud Foundation

Lisbon, Portugal

- **Team:** Computational Clinical Imaging Group (Nickolas Papanikolaou)
- Developed clinically-valuable deep-learning models for radiological image segmentation and classification
- Developed self-supervised learning methods to improve classification and segmentation performance using massive datasets
- Managed and processed large collections of MRI data (>10,000 patients)
- Developed and implemented clinical image generation methods
- Collaborated with over 70 researchers across different projects focused on creating value for clinicians and patients

2017 → 2022 — **PhD fellow**

EMBL-EBI + Cambridge University

Cambridge, UK

- **Advisors:** Moritz Gerstung and George S. Vassiliou
- Developed machine- and deep-learning methods to detect/characterize cells in digitised blood slides and predict disease genetics through uncover cytomorphological profiles
- Modelled longitudinal sequencing experiments using Bayesian statistics; determined genetic and non-genetic factors driving clonal expansion. Modelled phylogenetic and phylodynamic lifelong trajectories of clones using single-cell colonies

2016 → 2017 — **MSc. Researcher**

Center for Neuroscience and Cell Biology

Coimbra, Portugal

- **Advisor:** Irina Moreira
- Developed machine-learning protocols to determine hot-spots (important parts) in the binding interfaces of proteins
- Performed structural and statistical analysis of large collections of protein-protein complexes and structural characterization of complexes with no known structure

Education

2017 → 2022 — **PhD** in computational biology
University of Cambridge, UK

2015 → 2017 — **MSc** in cell and molecular biology (minor in neurosciences) | 18/20
University of Coimbra, Portugal

2012 → 2015 — **BSc** in biochemistry | 16/20
University of Coimbra, Portugal

Professional training

2024 — Data or Specimens Only Research
CITI Program (online)

2023 — Data Validation for Machine Learning
Weights and Biases (online)

2023 — Docker & Kubernetes: The Practical Guide
Academind (online)

2022 — Probability theory: foundations for data science
Colorado Boulder University (online)

2021 — Econometrics: methods and applications
Erasmus University Rotterdam (online)

2017 — Summer School in Computational Biology
University of Coimbra

Skills, tools and other competencies

Prog. languages	`Python`	◆◆◆◆◆
	`R`	◆◆◆◆◇
Machine-learning	`scikit-learn`	◆◆◆◆◇
	`caret`	◆◆◆◆◇
Deep-learning	`pytorch`, `lightning`	◆◆◆◆◆
	`MONAI`	◆◆◆◆◇
	`tensorflow`	◆◆◆◇◇
	`huggingface`	◆◆◇◇◇
	Large language models (`huggingface`, `langchain`, `ollama`)	◆◆◆◇◇
Computer-vision	`scikit-image`	◆◆◆◆◇
	`OpenCV`	◆◆◆◆◇
Data science	hypothesis testing	◆◆◆◆◇
	multivariate analyses	◆◆◆◆◇
	bayesian methods (MCMC)	◆◆◆◇◇
	Data manipulation (`pandas`, `polars`, `tidyverse`)	◆◆◆◆◇
Data viz	`ggplot2`	◆◆◆◆◇
	`D3.js` (`javascript`)	◆◇◇◇◇
Web dev	`flask`	◆◆◆◇◇
	`shiny`	◆◆◆◇◇
	`javascript`	◆◆◇◇◇
Workflow	version control (`git`)	◆◆◆◆◇
	containerisation (`Docker`)	◆◆◆◆◇
	workflow orchestration (`snakemake`)	◆◆◆◆◇

Soft skills

Teamwork: worked with international and pan-European teams on multiple projects

Leadership, project management: designed and management research projects for students and young researchers

Communication: clear and precise communication of technical and scientific results to academic and non-specialist audiences

Adaptability: quickly adapted to new fields, i.e. evolutionary biology and clinical machine-learning

Critical thinking: identified useful and valuable solutions to complicated problems and implemented them

Invited presentations

2024 Bridging imaging and genetics
28th Sociedade Portuguesa de Genética Humana Annual Meeting, Porto, Portugal

2024 How does machine-learning assist medical imaging specialists?
Data Modelling, AI, and Health: Bridging Models and Insights in Epidemiology, Biostatistics, and Medical Imaging (organized by Mathematics), Webinar

2023 ProCAncer-I: On manufacturer variability, automatic annotation and orphan data
European Multidisciplinary Congress on Urological Cancers, Marseille, France

Conference presentations

2024 Addressing the challenges of curating and segmenting large multi-centric prostate multiparametric MRI datasets with machine-learning (poster)
EuSoMII, Vienna, Austria

2024 Predicting Prostate Cancer Biochemical Recurrence After Radical Prostatectomy with Multiparametric MRI Radiomics: A Multicentric Study (poster)
EuSoMII, Vienna, Austria

2024 Giving new life to orphan data with self-supervised learning - a multi-institutional PCa MRI case study (poster)
European Congress of Radiology, Vienna, Austria

2021 The Natural History of Clonal Haematopoiesis (poster)
CRUK Cambridge Centre Early Detection Programme 6th Annual Symposium, Cambridge, UK

2020 Leveraging Automated Blood Cell Morphology for Myelodysplastic Syndrome Diagnosis and Prognosis Prediction
Quantitative BioImaging Conference,, Oxford, UK

2017 Using big-data to understand the protein interface landscape
Encontro de Jovens Investigadores de Biologia Computacional Estrutural, Coimbra, Portugal

2016 A Machine Learning Based Protein-Protein Hot-Spot Prediction Method — SpotOn
Encontro de Jovens Investigadores de Biologia Computacional Estrutural, Lisbon, Portugal

Honors and Awards

2017 EMBL PhD Fellowship (EMBL)

2017 MSc honours for outstanding academic performance (Universidade de Coimbra)

Teaching experience

2024 How to develop a Radiomics Signature
ESGAR, Webinar
Description: Teaching medical professionals about radiomic model development

2019 EMBL Lautenschlager Summer School
EMBL, Heidelberg, Germany
Description: Teaching young graduate students about practical bio-image analysis

2016 Workshops on Introductory Programming
University of Coimbra, Coimbra, Portugal
Description: Teaching undergraduate students about programming in Python and R

Other activies

2010 → present Music (Producer, composer, performer)
Description: Electronic music producer and member of different bands. Member of CIGA239, a musical association in Coimbra.

2019 EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2019 (Website design)
Description: Developed and designed the website for the 2019 eSCAMPS

2018 20th EMBL PhD Symposium (Organization, speaker contact)
Description: Contacted different high-profile researchers to invite them to present at the 20th EMBL PhD Symposium

2014 → 2017 Rádio Universidade de Coimbra (Radioshow host, programming director between 2016 and 2017)
Description: Hosted different radio shows on rock, experimental, jazz and metal music; coordinated the programming department (coordinated a department with approx. 50 people and designed a cohesive radio broadcasting schedule with other departments)

2016 Palco RUC (Curator and stage director)
Description: Curated and directed the stage for a small festival attended by thousands of people over four days in Coimbra. Contacted multiple artists and coordinated riders

2016 → 2017 Junior Enterprise on Science and Technology (Co-founder)
Description: JEST is a junior initiative I founded with a few colleagues that is dedicated to data science training among young students and services to external businesses

Student orientations

2024 Retrieval Augmented Generation for Intelligent Medical Chatbot Applications using open-source Large Language Models. **João Mata** (BSc)

2024 → 2025 Synthetic CT generation for MRI-only prostate radiotherapy. **Catarina Caldeira** (MSc thesis)

2024 → 2025 Automatic organ-at-risk, clinical target volume and lesion segmentation in prostate radiotherapy. **Martim Carneiro** (MSc thesis)

2024 → 2025 Any-to-any b-value generation for accelerated diffusion weighted imaging of the prostate. **Pedro Parracho** (MSc thesis)

2024 → 2025 A systematic study of voxel radiomics and deep-learning for lesion segmentation and detection in prostate cancer bi-parametric MRI. **Mariana Ferro** (MSc thesis)

Publications (papers, book chapters, *"proceedings"*, "other")

Almeida, J. G. d., Rodrigues, N. M., Castro Verde, A. S., Mascarenhas Gaivão, A., Bilreiro, C., Santiago, I., Ip, J., Belião, S., Matos, C., Silva, S., Tsiknakis, M., Marias, K., Regge, D., and Papanikolaou, N. (2025), **Impact of Scanner Manufacturer, Endorectal Coil Use, and Clinical Variables on Deep Learning-assisted Prostate Cancer Classification Using Multiparametric MRI**, *Radiology: Artificial Intelligence*, 1, e230555

Ma, J., Xie, R., Ayyadhury, S., Ge, C., Gupta, A., Gupta, R., Gu, S., Zhang, Y., Lee, G., Kim, J., and others (2024), **The multimodality cell segmentation challenge: toward universal solutions**, *Nature Methods*, Nature Publishing Group US New York, 1–11

Rodrigues, N. M., Almeida, J. G. de, Verde, A. S. C., Gaivão, A. M., Bilreiro, C., Santiago, I., Ip, J., Belião, S., Moreno, R., Matos, C., and others (2024), **Analysis of domain shift in whole prostate gland, zonal and lesions segmentation and detection, using multicentric retrospective data**, *Computers in Biology and Medicine*, Elsevier, 108216

Almeida, J., Castro Verde, A. S., Gaivão, A., Bilreiro, C., Santiago, I., Ip, J., Belião, S., Matos, C., Tsiknakis, M., Marias, K., and others (2024), **Self-Supervised Learning for Volumetric Imaging: A Prostate Cancer Biparametric Magnetic Resonance Imaging Case Study**, *Available at SSRN 4864797*

Rodrigues, N. M., Almeida, J. G. de, Rodrigues, A., Vanneschi, L., Matos, C., Lisitskaya, M. V., Uysal, A., Silva, S., and Papanikolaou, N. (2024), **Deep Learning Features Can Improve Radiomics-Based Prostate Cancer Aggressiveness Prediction**, *JCO Clinical Cancer Informatics*, Wolters Kluwer Health, 8, e2300180

Del Corso, G., Pachetti, E., Buongiorno, R., Rodrigues, A. C., Germanese, D., Pascali, M. A., Almeida, J., Rodrigues, N., Tsiknakis, M., Papanikolaou, N., and others (2024), *"Radiomics-Based Reliable Predictions of Side Effects After Radiotherapy for Prostate Cancer"*, in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4

Verde, A. S. C., Almeida, J. G. de, Fonseca, J., Matos, C., Conceição, R. C., and Papanikolaou, N. (2024), *"StitchPro for Computational Pathology Stitching in Patients with Prostate Cancer"*, in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4

Almeida, J. G. d., Gudgin, E., Besser, M., Dunn, W. G., Cooper, J., Haferlach, T., Vassiliou, G. S., and Gerstung, M. (2023), **Computational analysis of peripheral blood smears detects disease-associated cytomorphologies**, *Nature Communications*, Nature Publishing Group, 14, 4378. https://doi.org/10.1038/s41467-023-39676-y

Rodrigues, A. C., Almeida, J., Rodrigues, N., Moreno, R., Gaivão, A., Bilreiro, C., Santiago, I., Ip, J., Belião, S., Domingues, I., and others (2023), "Development and Prospective Validation of a Fully Automatic Bi-Parametric MRI Radiomics Signature to Predict Prostate Cancer Disease Aggressiveness: A Multi-Centric Study Using Over 4000 Patients"

Almeida, J. G. de, Rodrigues, N. M., Silva, S., and Papanikolaou, N. (2023), **Testing the Segment Anything Model on radiology data**, *arXiv preprint arXiv:2312.12880*

Rodrigues, N., Almeida, J., and Silva, S. (2023), *"Performance Analysis of Self-Supervised Strategies for Standard Genetic Programming"*, in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, pp. 627–630

Fabre, M. A., Almeida, J. G. d., Fiorillo, E., Mitchell, E., Damaskou, A., Rak, J., Orrù, V., Marongiu, M., Chapman, M. S., Vijayabaskar, M., and others (2022), **The longitudinal dynamics and natural history of clonal haematopoiesis**, *Nature*, Nature Publishing Group, 1–8

Preto, A. J., Matos-Filipe, P., Almeida, J. G. d., Mourão, J., and Moreira, I. S. (2021), "Predicting Hot Spots Using a Deep Neural Network Approach", *Artificial Neural Networks*, Springer

Preto, A. J., Barreto, C. A., Baptista, S. J., Almeida, J. G. d., Lemos, A., Melo, A., Cordeiro, M. N. D., Kurkcuoglu, Z., Melo, R., and Moreira, I. S. (2020), **Understanding the binding specificity of G-Protein coupled receptors toward G-proteins and arrestins: Application to the dopamine receptor family**, *Journal of Chemical Information and Modeling*, American Chemical Society, 60, 3969–3984

R Magalhães, P., Machuqueiro, M., Almeida, J. G. d., Melo, A., DS Cordeiro, M. N., Cabo Verde, S., H Gumus, Z., S Moreira, I., DG Correia, J., and Melo, R. (2019), **Dynamical rearrangement of human epidermal growth factor receptor 2 upon antibody binding: effects on the dimerization**, *Biomolecules*, Multidisciplinary Digital Publishing Institute, 9, 706

Lemos, A., Melo, R., Preto, A. J., Almeida, J. G., Moreira, I. S., and Dias Soeiro Cordeiro, M. N. (2018), **In silico studies targeting G-protein coupled receptors for drug research against Parkinson's disease**, *Current neuropharmacology*, Bentham Science Publishers, 16, 786–848

Preto, A. J., Almeida, J. G., Schaarschmidt, J., Xue, L. C., Moreira, I. S., and Bonvin, A. M. (2018), **Computational Tools for the Structural Characterization of Proteins and Their Complexes from Sequence-Evolutionary Data**, *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, John Wiley & Sons, Ltd, 1–19

Melo, R., Lemos, A., Preto, A. J., Almeida, J. G., Correia, J. D., Sensoy, O., and Moreira, I. S. (2018), **Computational approaches in antibody-drug conjugate optimization for targeted cancer therapy**, *Current topics in medicinal chemistry*, Bentham Science Publishers, 18, 1091–1109

Melo, R., Lemos, A., Preto, A. J., Bueschbell, B., Matos-Filipe, P., Barreto, C., Almeida, J. G., Silva, R. D., Correia, J. D., and Moreira, I. S. (2018), **An Overview of Antiretroviral Agents for Treating HIV Infection in Paediatric Population**, *Current medicinal chemistry*, Bentham Science Publishers

Moreira, I. S., Koukos, P. I., Melo, R., Almeida, J. G., Preto, A. J., Schaarschmidt, J., Trellet, M., Gumus, Z. H., Costa, J., and Bonvin, A. M. (2017), **SpotOn: high accuracy identification of protein-protein interface hot-spots**, *Scientific reports*, Nature Publishing Group, 7, 1–11

Almeida, J. G. d., Preto, A. J., Koukos, P. I., Bonvin, A. M., and Moreira, I. S. (2017), **Membrane proteins structures: A review on computational modeling tools**, *Biochimica et Biophysica Acta (BBA)-Biomembranes*, Elsevier, 1859, 2021–2039

Sensoy, O., Almeida, J. G., Shabbir, J., Moreira, I. S., and Morra, G. (2017), "Computational studies of G protein-coupled receptor complexes: Structure and dynamics", *Methods in Cell Biology*, Academic Press

Bastos, F. C., Corceiro, V. N., Lopes, S. A., Almeida, J. G. d., Matias, C. M., Dionisio, J. C., Mendes, P. J., Aidos, F. D. Sampaio dos, Quinta-Ferreira, R. M., and Quinta-Ferreira, M. E. (2017), **Effect of tolbutamide on tetraethylammonium-induced postsynaptic zinc signals at hippocampal mossy fiber-CA3 synapses**, *Canadian Journal of Physiology and Pharmacology*, NRC Research Press, 95, 1058–1063

Melo, R., Almeida, J. G., Verde, S. C., Gumus, Z., Moreira, I., and Correia, J. (2017), *"Structural mechanism of HER2-antibodies complexes by molecular dynamics studies"*, in *Proceedings of MOL2NET 2017, International Conference on Multidisciplinary Sciences, 3rd edition*, p. 5084

Almeida, J. G., Bonvin, A., and Moreira, I. (2017), *"Using big-data to understand the protein interface landscape"*, in *Proceedings of MOL2NET 2017, International Conference on Multidisciplinary Sciences, 3rd edition*

Almeida, J. G., Preto, A. J., Melo, R., Gumus, Z. H., Costa, J., Bonvin, A. M., and Moreira, I. S. (2017), *"Coevolution importance on binding Hot-Spot prediction methods"*, in *MOL2NET 2016, International Conference on Multidisciplinary Sciences, 2nd edition*