

---

# **Curriculum vitae**

*Jorge Miguel Ferreira da Silva*

---

UNIVERSIDADE DE AVEIRO

*Curriculum vitae,*

Jorge Miguel Ferreira da Silva

Apresentado no âmbito do Concurso para Investigador Auxiliar, na área científica de Informática, da Universidade de Aveiro, publicado no edital Ref.<sup>a</sup> CI-CTTI-108-SGRH/2025.

## Índice

<b>1 Dados pessoais e académicos</b>	<b>4</b>
<b>2 Sumário do meu percurso</b>	<b>4</b>
<b>3 Desempenho Científico</b>	<b>6</b>
3.1 Produção científica . . . . .	6
3.2 Participação em projetos científicos . . . . .	12
3.3 Candidatura a projetos científicos . . . . .	16
3.4 Intervenção na comunidade científica . . . . .	17
<b>4 Capacidade Pedagógica</b>	<b>20</b>
4.1 Constituição de equipas científicas e orientação de estudantes de doutoramento e bolseiros de pós-doutoramento . . . . .	20
4.2 Atividade Letiva . . . . .	21
4.2.1 Perspetivas de Evolução e Envolvimento Futuro . . . . .	21
<b>5 Atividades de extensão e de disseminação de conhecimento</b>	<b>22</b>
5.1 Propriedade Intelectual e Patentes . . . . .	22
5.2 Comissões de normalização, projetos legislativos e de normas técnicas . . . . .	24
5.3 Atividades de divulgação científica . . . . .	25
5.4 Palestras e seminários . . . . .	25
<b>6 Atividades de gestão institucional</b>	<b>26</b>

## 1 Dados pessoais e académicos

### Jorge Miguel Ferreira da Silva

Universidade de Aveiro (UA)

Instituto de Engenharia Electrónica e Informática de Aveiro (IEETA)

3810-193 Aveiro, Portugal

E-mail: jorge.miguel.ferreira.silva@ua.pt

Telefone: +351 913 776 273

### Posição atual

- Investigador doutorado, no IEETA, da Universidade de Aveiro, Aveiro, Portugal, desde 15/07/2023.

### Qualificações

- Doutor em Engenharia Informática, pela Universidade de Aveiro, Portugal, em 04/2023, no *Programa Doutoral em Engenharia Informática*.
- Mestre em Bioengenharia - Engenharia Biomédica, pela Faculdade de Engenharia da Universidade do Porto, em 04/2016.

### Experiência profissional

- Bolseiro do projecto GDI no IEETA da Universidade de Aveiro, no IEETA, da Universidade de Aveiro, Aveiro, Portugal, desde 03/2023 a 15/07/2023 (European Union's Digital Europe Programme No 101081813).
- Bolseiro, na Fundação para a Ciência e a Tecnologia (FCT), Lisboa, Portugal de 12/2018 a 01/2023 (SFRH/BD/141851/2018).
- Bolseiro no IEETA da Universidade de Aveiro, Aveiro, Portugal, de 06/2016 a 11/2018.

### Perfis online

Scopus: <https://www.scopus.com/authid/detail.uri?authorId=57194129937>

ORCID: <https://orcid.org/0000-0002-6331-6091>

Google Scholar: <https://scholar.google.pt/citations?user=N6ZzRPEAAAAJ&hl=en>

## 2 Sumário do meu percurso

Sou investigador com doutoramento em Engenharia Informática pela Universidade de Aveiro, especializado em teoria da informação e inteligência artificial, com ênfase em genómica e bioinformática.

O meu percurso na área de informática começou no meu Mestrado Integrado em Bioengenharia na Universidade do Porto, durante o qual estagiei no Departamento de Engenharia da Informação (DEI) em Pádua, Itália. Além disso, durante este período, co-fundei a Associação de Estudantes

de Bioengenharia da FEUP/ICBAS (NEB) e contribuí para a organização dos Simpósios de Bioengenharia.

Em 2016, após terminar o mestrado, juntei-me ao grupo de bioinformática do IEETA, na Universidade de Aveiro, onde trabalhei em projetos como o SCREEN-DR (FCT/CMU), para deteção de retinopatia diabética usando deep learning, e o MMIR (FCT), para classificação de imagens de ecocardiograma. O meu trabalho também incluiu a participação nos projetos europeus EHDEN e EMIF. Em 2018, fui contemplado com uma bolsa de doutoramento financiada pela FCT, durante a qual melhorei as minhas competências computacionais, focando-me em inteligência artificial, métodos de compressão de informação e engenharia de software aplicada a bioinformática.

Após concluir o meu doutoramento, obtive em 2023 uma posição de Investigador Doutorado na Universidade de Aveiro no âmbito do projeto europeu Genomic Data Infrastructure (GDI), que visa implementar uma infraestrutura federada para partilha, gestão e análise de dados genómicos a nível Europeu. O meu papel como investigador é atualmente liderar esta tarefa como responsável técnico do Nó Português em Aveiro, e, simultaneamente, desenvolver outras soluções inovadoras que possam ser adotadas pelo projeto no futuro.

Com a minha participação em projetos europeus e workshops internacionais, tive a oportunidade de conhecer parceiros europeus, o que me levou a colaborar em projetos internacionais, da comunidade ELIXIR, como o ELIXIR Science Tier, focadas em Federated Human Data, em que sou líder de tarefas.

Para além da minha participação em projetos, durante o meu percurso, desenvolvi diversas ferramentas relevantes para análise de complexidade, como por exemplo o PANTHER, que é uma ferramenta para autenticação de arte e explicação de conteúdo, e outras aplicadas à genómica, como CANVAS, Alcor e Altair. Atualmente, tenho um total de 43 artigos publicados (33 indexados no Scopus, mais 3 em processo de indexação), 11 artigos em revistas científicas (9 dos quais em revistas de ciência da computação de quartil 1), 26 artigos em conferências internacionais e 6 em conferências nacionais.

Para além das minhas contribuições de investigação, participei ativamente na apresentação de conferências, seminários e eventos de envolvimento público, como o ELIXIR AllHands, BioNexus, Forum de Gestão de Dados de Investigação da FCCN, Data Science Portugal, Students@DETI e Research Summit. Atuei também como revisor para várias revistas científicas e fui Programme Committe de diversas conferencias internacionais.

Atualmente, sou atualmente PI e Co-IP de 4 projetos computacionais financiados pelo RCCN, sou coordenador tecnico do Comité operacional Nó Português de Federated European Genome-phenome Archive (FEGA) e da comunidade de dados genómicos humanos Federados do ELIXIR em Portugal, e sou o responsável pela gestão e migração de dados no IEETA, coordenando a transferência de datasets para a plataforma dunas. Recebi ainda um prémio de melhor artigo do IEEE e venci desafios internacionais como o BioASQ.

### 3 Desempenho Científico

Na presente secção, são descritas todas as minhas contribuições científicas, até à data, organizadas conforme os critérios estabelecidos no edital do concurso. Em suma, publiquei 43 artigos científicos; estive envolvido em 9 projetos; fui revisor em diversas revistas científicas; fui membro do *Programme Committee* de conferências. Todas essas atividades estão detalhadas nesta secção dividida nas seguintes subsecções: i) produção científica; e ii) participação em projetos científicos; iii) intervenção na comunidade científica.

#### 3.1 Produção científica

Até ao momento publiquei um total de 43 artigos, sendo que 33 dos quais estão indexados no scopus (mais 3 serão indexados em breve).

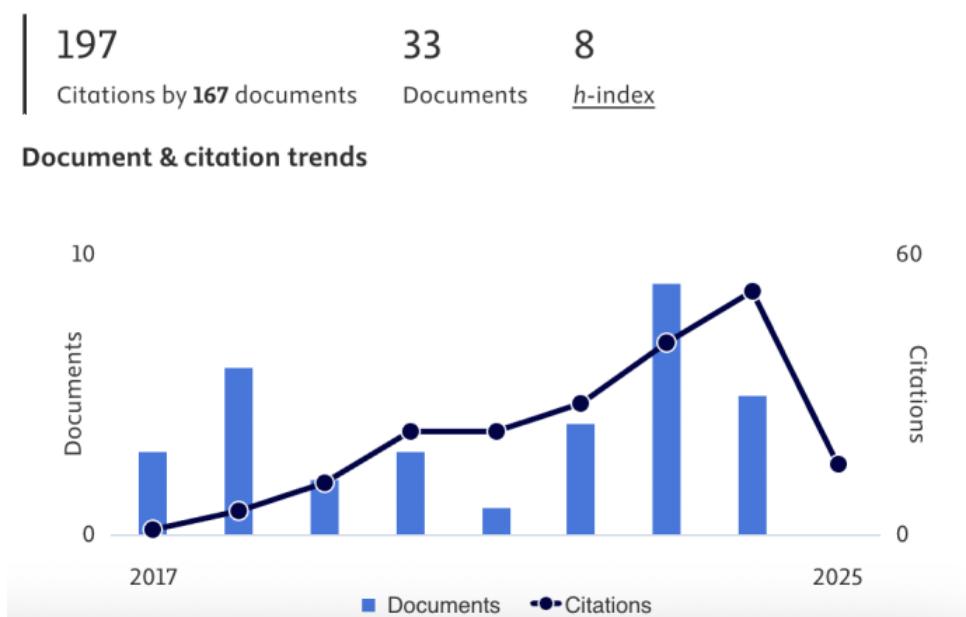


Figura 1: Current profile metrics in scopus.

O meu perfil de publicações é categorizado da seguinte forma:

- **Artigos de Jornal:** Publiquei 11 artigos em revista, sendo que 10 destes artigos são focados na área de ciência da computação e 1 artigo numa área de aplicação.
- **Artigos de Conferência Internacional:** Apresentei 26 trabalhos em várias conferências internacionais (atualmente 21 indexadas + 3 ainda em processo de indexação). Dos trabalhos atualmente indexados, 16 são dedicados à área de ciência da computação e 5 trabalhos abrangendo áreas de aplicação (os outros 3 ainda não indexados também são de ciência da computação).
- **Artigos de Conferência Nacional:** Apresentei 6 trabalhos em conferências nacionais dedicadas à deciminação de ciência da computação.

É importante mencionar que alguns destes artigos aplicam a ciência da computação para resolver problemas em outras áreas e foram indexados sob uma categoria diferente (por exemplo, informática em saúde). Esta abordagem interdisciplinar demonstra o meu compromisso em avançar na ciência da computação e enfrentar desafios do mundo real, aproveitando métodos computacionais em vários

domínios.

### **Artigos de revista**

Publiquei 11 artigos em revistas indexadas no Scopus, incluindo 9 artigos de Quartil 1 (Q1) e 2 artigos de Quartil 2 (Q2), abrangendo a área de Ciência da Computação (11 artigos) e Informática na Saúde (1 artigo). Estes incluem artigos em: Pattern Recognition (Q1, CiteScore 15.5), Artificial Intelligence in Medicine (Q1, 15.0), Bioinformática (Q1, 13.4), Gigascience (Q1, 11.6), Computer Standards and Interfaces (Q1, 8.8), Journal of Biomedical Informatics (Q1, 8.2) e International Journal of Medical Informatics (Q1, 8.0). Além disso, publiquei 2 artigos em Entropy (Q2, CiteScore 4.4). Segue abaixo o filtro de pesquisa a aplicar no Scopus e a lista completa dos artigos referidos:

```
AU-ID("Silva, Jorge Miguel" 57194129937) AND (LIMIT-T0 (DOCTYPE, "ar"))
AND (LIMIT-T0(SRCTYPE, "j"))
```

11. J.M. Silva, A.J. Pinho e D. Pratas (novembro de 2024). “AltaiR: a C toolkit for alignment-free and temporal analysis of multi-FASTA data”. Em: *GigaScience* 13, giae086. ISSN: 2047-217X. DOI: [10.1093/gigascience/giae086](https://doi.org/10.1093/gigascience/giae086).
10. J.M. Silva e J.R. Almeida (2024). “Enhancing metagenomic classification with compression-based features”. Em: *Artificial Intelligence in Medicine* 156, p. 102948. ISSN: 0933-3657. DOI: [10.1016/j.artmed.2024.102948](https://doi.org/10.1016/j.artmed.2024.102948).
9. J.M. Silva, W. Qi, A.J. Pinho e D. Pratas (dezembro de 2023). “AlcoR: alignment-free simulation, mapping, and visualization of low-complexity regions in biological data”. Em: *GigaScience* 12, giad101. ISSN: 2047-217X. DOI: [10.1093/gigascience/giad101](https://doi.org/10.1093/gigascience/giad101).
8. J.M. Silva, D. Pratas, T. Caetano e S. Matos (2022). “The complexity landscape of viral genomes”. Em: 11. DOI: [10.1093/gigascience/giac079](https://doi.org/10.1093/gigascience/giac079).
7. J.M. Silva, D. Pratas, R. Antunes, S. Matos e A.J. Pinho (2021). “Automatic analysis of artistic paintings using information-based measures”. Em: 114. DOI: [10.1016/j.patcog.2021.107864](https://doi.org/10.1016/j.patcog.2021.107864).
6. D. Pratas e J.M. Silva (2020). “Persistent minimal sequences of SARS-CoV-2”. Em: 36.21, pp. 5129–5132. DOI: [10.1093/bioinformatics/btaa686](https://doi.org/10.1093/bioinformatics/btaa686).
5. J.M. Silva, E. Pinho, S. Matos e D. Pratas (2020). “Statistical complexity analysis of turing machine tapes with fixed algorithmic complexity using the best-order markov model”. Em: 22.1, p. 105. DOI: [10.3390/e22010105](https://doi.org/10.3390/e22010105).
4. D. Pratas, M. Hosseini, J.M. Silva e A.J. Pinho (2019). “A reference-free lossless compression algorithm for DNA sequences using a competitive prediction of two classes of weighted models”. Em: 21.11. DOI: [10.3390/e2111074](https://doi.org/10.3390/e2111074).
3. J.M. Silva, T.M. Godinho, D. Silva e C. Costa (2019). “A community-driven validation service for standard medical imaging objects”. Em: 61, pp. 121–128. DOI: [10.1016/j.csi.2018.06.003](https://doi.org/10.1016/j.csi.2018.06.003).

2. M. Pedrosa, J.M. Silva, J.F. Silva, S. Matos e C. Costa (2018). “SCREEN-DR: Collaborative platform for diabetic retinopathy”. Em: 120, pp. 137–146. doi: [10.1016/j.ijmedinf.2018.10.005](https://doi.org/10.1016/j.ijmedinf.2018.10.005).
  1. J.M. Silva, E. Pinho, E. Monteiro, J.F. Silva e C. Costa (2018). “Controlled searching in reversibly de-identified medical imaging archives”. Em: 77, pp. 81–90. doi: [10.1016/j.jbi.2017.12.002](https://doi.org/10.1016/j.jbi.2017.12.002).
- Para além dos artigos de revista, aceites tenho atualmente 4 de revista em processo de revisão.
4. J.M. Silva e J.L. Oliveira (2025). “PhenoQC: An Integrated Toolkit for Quality Control of Phenotypic Data in Genomic Research”. Em: *Informatics in Medicine Unlocked*.
  3. I.B. Martins, J.M. Silva e J.R. Almeida (2025). “HYMET: A Hybrid Metagenomic Pipeline for Accurate and Efficient Taxonomic Classification”. Em: *GigaScience*.
  2. J.R. Almeida, J.M. Silva e J.L. Oliveira (2025). “Sharing Observational Health Data following FAIR Principles”. Em: *Multimedia Tools and Applications*.
  1. I.B. Martins, J.M. Silva e J.R. Almeida (2025). “A systematic review and benchmarking of modern metagenomic tools for taxonomic classification”. Em: *Computer Science Review*.

## Artigos de conferência

Publiquei 26 artigos em conferências internacionais, dos quais 21 são indexados (3 estão aceites e serão brevemente adicionados), e 6 artigos em conferências nacionais. O meu trabalho foi apresentado em várias conferências conceituadas, como IEEE CIBCB (Conference on Computational Intelligence in Bioinformatics and Computational Biology), IEEE CBMS (Computer-Based Medical Systems), IWBBIO (International Work-Conference on Bioinformatics and Biomedical Engineering), DoCEIS (Advanced Doctoral Conference on Computing, Electrical and Industrial Systems), CARS (Computer Assisted Radiology and Surgery), MIE (Medical Informatics Europe), HCIST (International Conference on Health and Social Care Information Systems and Technologies), STC (Smart Technology Conference) e IbPRIA (Iberian Conference on Pattern Recognition and Image Analysis). A minha investigação abrange vários domínios, incluindo repositórios genómicos, técnicas de compressão de dados, imagiologia médica e aplicações de machine learning em ambientes de saúde.

Abaixo está o filtro de pesquisa a ser aplicado no Scopus e a lista completa dos artigos mencionados.:

AU-ID("Silva, Jorge Miguel" 57194129937) AND (LIMIT-T0(SRCTYPE,"p"))

18. <sup>1</sup>Jorge Miguel Silva e José Luis Oliveira (2025). “VCFX: A Minimalist, Modular Toolkit for Streamlined Variant Analysis”. Em: *12th International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2025)*. Springer, pp. –. doi: [–](#).
17. <sup>1</sup>Inês Branco Martins, Jorge Miguel Silva e João Rafael Almeida (2025). “Title: A hybrid metagenomic pipeline for taxonomic classification”. Em: *12th International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2025)*. Springer, pp. –. doi: [–](#).

16. <sup>1</sup>Alexandre Cotorobai, Jorge Miguel Silva e José Luis Oliveira (2025). “A Federated Random Forest Solution for Secure Distributed Machine Learning”. Em: *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (CBMS)*. IEEE, pp. –. DOI: –.
15. Jorge Miguel Silva e José Luis Oliveira (2024). “BIT.UA at iDPP: Predictive Analytics on ALS Disease Progression Using Sensor Data with Machine Learning”. Em: vol. 3740. Cited by: 0, pp. 1401–1411.
14. João Almeida, Jorge Miguel Silva, Luís Carlos Afonso, Tiago Almeida, Rui Antunes, Richard Jonker, João António Reis et al. (2024). “HealthDBFinder: a question-answering task for health database discovery”. Em: *CBMS 2024*. DOI: [10.1109/CBMS61543.2024.00025](https://doi.org/10.1109/CBMS61543.2024.00025).
13. Inês Branco Martin, Jorge Miguel Silva e João Almeida (2024). “A comprehensive study of databases to assess the reliability of metagenomic tools”. Em: *IEEE CIBCB 2024*. DOI: [10.1109/CIBCB58642.2024.10702118](https://doi.org/10.1109/CIBCB58642.2024.10702118).
12. Diniz Cruz, Joao Rafael Almeida, Jorge Miguel Silva e Jose Luis Oliveira (2023). “SecureFASTA: Ensuring privacy and trust when sharing genomic data”. Em: vol. 2023-June, pp. 173–178. DOI: [10.1109/CBMS58004.2023.00212](https://doi.org/10.1109/CBMS58004.2023.00212).
11. Tiago Almeida, Richard A.A. Jonker, Roshan Poudel, Jorge M. Silva e Sérgio Matos (2023). “BIT.UA at BioASQ 11B: Two-Stage IR with Synthetic Training and Zero-Shot Answer Generation”. Em: vol. 3497, pp. 37–59.
10. Jorge Miguel Silva, Joao Rafael Almeida e Jose Luis Oliveira (2023). “Classifying and discovering genomic sequences in metagenomic repositories”. Em: vol. 219, pp. 1501–1508. DOI: [10.1016/j.procs.2023.01.441](https://doi.org/10.1016/j.procs.2023.01.441).
9. Tiago Almeida, Richard A.A. Jonker, Roshan Poudel, Jorge M. Silva e Sérgio Matos (2023). “BIT.UA at MedProcNer: Discovering Medical Procedures in Spanish Using Transformer Models with MCRF and Augmentation”. Em: vol. 3497, pp. 60–72.
8. Joao Rafael Almeida, Jorge Miguel Silva e Jose Luis Oliveira (2023). “A FAIR Approach to Real-World Health Data Management and Analysis”. Em: vol. 2023-June, pp. 892–897. DOI: [10.1109/CBMS58004.2023.00338](https://doi.org/10.1109/CBMS58004.2023.00338).
7. Jorge Miguel Silva e Joao Rafael Almeida (2022). “The value of compression for taxonomic identification”. Em: vol. 2022-July, pp. 276–281. DOI: [10.1109/CBMS55023.2022.00055](https://doi.org/10.1109/CBMS55023.2022.00055).
6. Rui Lebre, Eduardo Pinho, Jorge Miguel Silva e Carlos Costa (2020). “Dicoogle Framework for Medical Imaging Teaching and Research”. Em: vol. 2020-July. DOI: [10.1109/ISCC50000.2020.9219545](https://doi.org/10.1109/ISCC50000.2020.9219545).
5. Jorge Miguel Silva, António Guerra, João Figueira Silva, Eduardo Pinho e Carlos Costa (2018). “Face De-Identification Service for Neuroimaging Volumes”. Em: vol. 2018-June, pp. 141–145. DOI: [10.1109/CBMS.2018.00032](https://doi.org/10.1109/CBMS.2018.00032).
4. João Figueira Silva, Jorge Miguel Silva, António Guerra, Sérgio Matos e Carlos Costa (2018). “Ejection Fraction Classification in Transthoracic Echocardiography Using a Deep Learning

- Approach". Em: vol. 2018-June, pp. 123–128. doi: [10.1109/CBMS.2018.00029](https://doi.org/10.1109/CBMS.2018.00029).
3. Micael Pedrosa, Jorge Miguel e Carlos Costa (2018). "Reactive through services: Opinionated framework for developing reactive services". Em: vol. 2018-January, pp. 330–337. doi: [10.5220/0006661403300337](https://doi.org/10.5220/0006661403300337).
  2. João Figueira Silva, Jorge Miguel Silva, Eduardo Pinho e Carlos Costa (2017). "3D-CNN in drug resistance detection and tuberculosis classification". Em: vol. 1866.
  1. Eduardo Pinho, João Figueira Silva, Jorge Miguel Silva e Carlos Costa (2017). "Towards representation learning for biomedical concept detection in medical images: UA.PT bioinformatics in ImageCLEF 2017". Em: vol. 1866.

Os artigos estendidos para capítulos de livros, são os seguintes:

AU-ID("Silva, Jorge Miguel" 57194129937) AND (LIMIT-T0(SRCTYPE,"k"))

7. Dinis. B. Cruz, João R. Almeida, Jorge M. Silva e José L. Oliveira (2023). "A Reliable and Secure Method for Sharing Genomic Data". Em: *Studies in Health Technology and Informatics* 302, pp. 1071–1072. doi: [10.3233/SHTI230350](https://doi.org/10.3233/SHTI230350).
6. Miguel Pinheiro, Jorge Miguel Silva e José Luis Oliveira (2023). "Optimizing Variant Calling for Human Genome Analysis: A Comprehensive Pipeline Approach". Em: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 13920 LNBI, pp. 72–85. doi: [10.1007/978-3-031-34960-7\\_6](https://doi.org/10.1007/978-3-031-34960-7_6).
5. Jorge Miguel Silva, Diogo Pratas, Tânia Caetano e Sérgio Matos (2022). "Feature-Based Classification of Archaeal Sequences Using Compression-Based Methods". Em: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 13256 LNCS, pp. 309–320. doi: [10.1007/978-3-031-04881-4\\_25](https://doi.org/10.1007/978-3-031-04881-4_25).
4. Jorge Miguel Silva, Diogo Pratas e Sérgio Matos (2023). "Exploring Kolmogorov Complexity Approximations for Data Analysis: Insights and Applications". Em: *IFIP Advances in Information and Communication Technology* 678, pp. 161–174. doi: [10.1007/978-3-031-36007-7\\_12](https://doi.org/10.1007/978-3-031-36007-7_12).
3. Jorge Miguel Silva e João Rafael Almeida (2022). "Characterizing Genomics Repositories Using Feature-Based Classification". Em: *Studies in Health Technology and Informatics* 298, pp. 167–168. doi: [10.3233/SHTI220932](https://doi.org/10.3233/SHTI220932).
2. Micael Pedrosa, Jorge Miguel Silva, Sérgio Matos e Carlos Costa (2018). "SCREEN-DR: Software architecture for the diabetic retinopathy screening". Em: *Studies in Health Technology and Informatics* 247, pp. 396–400. doi: [10.3233/978-1-61499-852-5-396](https://doi.org/10.3233/978-1-61499-852-5-396).
1. Jorge Miguel Silva, Tiago Marques Godinho, David Silva e Carlos Costa (2017). "Web Validation Service for Ensuring Adherence to the DICOM Standard". Em: *Studies in Health Technology and Informatics* 235, pp. 38–42. doi: [10.3233/978-1-61499-753-5-38](https://doi.org/10.3233/978-1-61499-753-5-38).

<sup>1</sup>Artigos aceites para publicação mas ainda não disponíveis no Scopus.

Artigos de conferencias internacionais não indexados no scopus:

2. J.M. Silva, M. Pedrosa, S. Matos e C. Costa (2018). "Research PACS for diabetic retinopathy screening". Em: *International Journal of Computer Assisted Radiology and Surgery*. Part of DOI: 10.1007/s11548-018-1766-y. Berlim, Germany. doi: [10.1007/s11548-018-1766-y](https://doi.org/10.1007/s11548-018-1766-y).
1. M. Pedrosa, J.M. Silva, S. Matos e C. Costa (2018). "SCREEN-DR - Software Architecture for the Diabetic Retinopathy Screening". Em: *International Journal of Computer Assisted Radiology and Surgery*. Part of DOI: 110.3233/978-1-61499-852-5-396. Berlim, Germany. doi: [10.1007/s11548-018-1766-y](https://doi.org/10.1007/s11548-018-1766-y).

Artigos de conferencias nacionais não indexados no scopus:

5. Jorge Miguel Silva, José Luís Oliveira, Armando Pinho e Diogo Pratas (março de 2024). "An Efficient Toolkit for Large-Scale Genomic Analysis and Its Application to Multi-Resistant Bacteria". Em: *Bioinformatics Open Days*.
4. Jorge Miguel Silva, Diogo Pratas, Tania Caetano e Sérgio Matos (2021). "Archaea Taxonomic Classification". Em: *Proceedings of the 27th Portuguese Conference on Pattern Recognition, RecPad 2021*. Évora, Portugal.
3. Jorge Miguel Silva, Diogo Pratas e Sérgio Matos (2020). "Comparison and Evaluation of Information-based Measures in Images". Em: *Proceedings of the 26th Portuguese Conference on Pattern Recognition, RecPad 2020*. Évora, Portugal.
2. Jorge Miguel Silva, Diogo Pratas e Sérgio Matos (2019). "Evaluation of Statistical Complexity in Viral Genome Sequences". Em: *Proceedings of the 25th Portuguese Conference on Pattern Recognition, RecPad 2019*. Porto, Portugal.
1. Jorge Miguel Silva, Augusto Silva e Pedro Vilela (2017). "Enhanced Cerebral Vascular Segmentation with Harmonic Constraints". Em: *The 23rd Portuguese Conference on Pattern Recognition*.

## Relatórios Técnicos e White Papers

Nesta secção são listados relatórios técnicos e white papers relevantes.

1. Jorge Miguel Silva e Jorge Oliveira (2024). *Portuguese Federated Human Data ELIXIR Community White Paper*. White Paper. Published by BioData.pt. Draft 2. Lisbon, Portugal: BioData.pt / ELIXIR Portugal.

URL: [https://rgw03.tp.dsi.tecnico.ulisboa.pt/BioData-Backups/www/public/white-papers/White\\_Paper\\_FHD.pdf](https://rgw03.tp.dsi.tecnico.ulisboa.pt/BioData-Backups/www/public/white-papers/White_Paper_FHD.pdf)

### 3.2 Participação em projetos científicos

Em termos academicos tenho participado em vários projetos científicos em várias áreas, desempenhando principalmente o papel de contribuidor tecnológico e coordenador técnico. Atualmente, estou envolvido em 3 projetos internacionais relacionados com a infraestrutura de dados genómicos na Europa e coordeno varios projectos computacionais. Além disso, participei em projetos europeus relacionados com a partilha e análise de dados médicos em grande escala. Também fui contemplado com uma bolsa de doutoramento e um projeto computacional da FCT. Estas iniciativas resultaram em artigos científicos, palestras, serviços e produtos, incluindo *software open-source*. Esta seção destaca brevemente a minha participação e os resultados mais significativos de cada projeto.

#### GDI - European Genomic Data Infrastructure

O projeto tem como objetivo facilitar a colaboração e aumentar as oportunidades de pesquisa, permitindo o acesso a dados genómicos recolhidos em várias instituições em toda a Europa. Ao estabelecer uma infraestrutura federada, sustentável e segura, o projeto GDI procura promover a partilha de dados, melhorar a padronização dos dados e encorajar o desenvolvimento de ferramentas e metodologias inovadoras no campo da genómica. Isso, em última instância, contribuirá para avanços na medicina personalizada, diagnósticos e tratamentos para várias doenças. Atualmente Portugal destaca-se neste projecto por ser um nó de vanguarda, classificação atribuída aos nós que se encontram mais avançados quer em termos de maturidade da implementação, quer em termos de contribuições dadas para o projecto a nível europeu.

Neste projeto, inicialmente contribuí desenvolvendo protótipos para a troca segura de dados genómicos, e definição de pipelines de variant calling que podem ser usadas pelo nó. Atualmente, sou um dos principais contributores do projeto a nível nacional<sup>1</sup>, trabalhando como investigador, mas também como DevOps da infraestrutura nacional e developer de ferramentas para melhoramento do sistema quer a nível do nó português do GDI, quer a nível internacional.

**Posição:** Investigador Doutorado

**Referência:** European Union's Digital Europe Programme No 101081813

#### FEGA – Federated Genome-Phenome Archive

O Federated Genome-Phenome Archive (FE GA) é uma infraestrutura europeia distribuída que liga repositórios nacionais de dados sensíveis de genómica, transcriptómica e dados fenotípicos, permitindo a sua descoberta e acesso controlado para investigação biomédica, em conformidade com o GDPR.<sup>2</sup> Formalizada em 2022, a rede apoia iniciativas pan-europeias como o 1+ Million Genomes e o European Health Data Space, fornecendo mecanismos uniformes de submissão, catalogação, autorização e auditoria para dados humanos. Portugal foi um dos primeiros países a aderir, através da infraestrutura BioData.pt/ELIXIR-PT, que opera o nó nacional FEGA e disponibiliza serviços de onboarding, helpdesk, armazenamento seguro e interfaces de programação a investigadores e hospitais.

Nesta infraestrutura sou o coordenador do Comité Operacional do nó nacional. Desde que assumi o cargo, o tornamo-nos um nó operacional da rede Federated Genome-Phenome Archive

<sup>1</sup><https://gdi.onemilliongenomes.eu/about/contributors>

<sup>2</sup><https://ega-archive.org/about/projects-and-funders/federated-ega/>

(FE GA Portugal)<sup>3</sup>. Sob a minha coordenação, estabelecemos uma estrutura organizacional completa, definindo papéis claros para a equipa, incluindo membros do helpdesk e do comité ético-legal. Atualmente, estamos a desenvolver procedimentos operacionais padronizados (SOPs) para o helpdesk e a trabalhar na implementação de uma infraestrutura robusta que garanta a sustentabilidade a longo prazo do nó FEGA Portugal.

**Posição:** Investigador Doutorado e Coordenador do Comité Operacional do nó Portugues da rede FEGA

**Referência:** European Union's Digital Europe Programme No 101081813

### **FAIR-FEGA: Accelerating High-Quality FAIR Data Deposition in Federated EGA**

Este projeto, apoiado pela infraestrutura ELIXIR, tem como objetivo acelerar a submissão de dados de alta qualidade e FAIR (Findable, Accessible, Interoperable, Reusable) na rede Federated European Genome-Phenome Archive (FE GA). Este esforço visa aumentar significativamente o fluxo de dados dentro e entre os nós da FEGA, promovendo a integração de dados e fortalecendo a confiança dos investigadores nos dados armazenados na rede. O project é feito em colaboração com um consórcio estratégico composto por sete nós ELIXIR e duas comunidades ELIXIR, com foco na implementação prática e avaliação de novos processos e ferramentas para o depósito de dados.

A minha contribuição neste projeto inclui a responsabilidade por uma das tarefas centrais, focada no desenvolvimento e teste de ferramentas e desenvolvimento de SOPs (Standard Operating Procedures) a nível internacional para a submissão de dados. Esta tarefa é essencial para assegurar que os dados depositados estejam alinhados com os padrões FAIR e possam ser reutilizados de forma eficiente pela comunidade científica internacional.

**Posição:** Colaborador

**Referência:** Programa ELIXIR 2024–2028

### **Enhancing National Research Data Accessibility through Federated European Genome-phenome Archive**

Este projeto visa mover o nó nacional do Arquivo Europeu Federado de Genoma-fenoma (FE GA) para um ambiente alojado pela FCCN. O nó português do FEGA é um avanço significativo na gestão de dados de pesquisa científica nacional, visto que, como um hub centralizado, oferece armazenamento seguro de dados com criptografia avançada, sistemas de backup confiáveis e acesso eficiente aos dados.

**Posição:** Co-coordenador

**Referência:** 2023.11060.CPCA

### **Advanced Genomic Data Processing in Portuguese FEGA Node**

Este projeto visa criar um ambiente seguro para a análise de dados genómicos do Arquivo Europeu Federado de Genoma-fenoma (FE GA), coordenado pelo ELIXIR. Mais concretamente esta iniciativa melhora o nó português do FEGA, ativo desde 2023, focando-se no processamento de dados genómicos, através da implementação de pipelines de análise de qualidade dos dados, da detecção de variantes e análise do genoma.

<sup>3</sup><https://fegaportugal.biodata.pt/about/federated-ega/>

**Posição:** Coordenador

**Referência:** 2023.14342.CPCA

### **Mapping the Knowledge Landscape of Language Models**

O projeto computacional TMFinder tem como objetivo classificar, comparar e visualizar modelos de linguagem existentes com base no seu conhecimento específico de domínio, abordando uma lacuna crítica na compreensão das capacidades desses modelos. A assunção principal é que um modelo de linguagem com mais conhecimento sobre um determinado domínio ou tópico pode comprimir o texto desse domínio de forma mais eficaz. Para fazer este trabalho analisamos as distribuições de compressão normalizada (NC) em diversos conjuntos de dados textuais, como artigos de notícias, redes sociais, documentos legais e registros médicos. Assim, podemos avaliar as diferenças de conhecimento entre modelos de linguagem, permitindo que investigadores e developers escolham os modelos mais adequados para tarefas específicas.

**Posição:** Co-coordenador

**Referência:** FCT.CPCA.2022.01

### **TMFinder - Inverse Problem Approximation**

O projeto computacional TMFinder tem como objetivo desenvolver um novo método para aproximar o problema inverso usando Máquinas de Turing e Modelos de Contexto Finito. O meu papel como investigador envolve a criação de uma solução de software que equilibra métodos probabilísticos e algorítmicos, oferecendo tolerância ao ruído, flexibilidade e tempo computacional viável. O programa utiliza um algoritmo de busca guiada e um design multi-nó computacional para busca eficiente e comunicação mínima entre os nós.

**Posição:** Co-coordenador

**Referência:** 2022.27068.CPCA.A0

### **SCREEN-DR - Image Analysis and Computational Learning Platform for Innovation in Diabetic Retinopathy Screening**

O projeto tem como objetivo desenvolver uma plataforma que utilize técnicas de análise de imagem e aprendizagem computacional para melhorar o processo de triagem de retinopatia diabética. Ao automatizar a coleta e anotação de imagens, a plataforma permite uma detecção mais eficiente de sinais de retinopatia e suporta o diagnóstico e tratamento precoces. Esta abordagem inovadora tem o potencial de melhorar significativamente os resultados dos pacientes e reduzir o peso nos sistemas de saúde.

No projeto SCREEN-DR, o meu papel como investigador envolveu não só o desenvolvimento de protótipos, mas também a criação de uma plataforma para a coleta e anotação de imagens para triagem de retinopatia diabética. As imagens anotadas desta plataforma foram posteriormente usadas para realizar a anotação automática de imagens de retinopatia usando modelos de deep learning, para realizar triagem automática e detecção precoce de pacientes com retinopatia diabética.

**Posição:** Bolseiro de Investigação

**Referência:** CMUP-ERI/TIC/0028/2014

## **MMIR - Multimodal Information Retrieval in Medical Image Repositories**

O projeto tem como objetivo desenvolver soluções inovadoras para a recuperação e classificação de informações médicas em repositórios de imagens médicas. Através da utilização de técnicas de recuperação multimodal de informações, o projeto procura melhorar a eficiência e precisão do acesso a dados médicos relevantes, promovendo assim cuidados de saúde mais eficazes e tomadas de decisão mais informadas no setor da saúde.

No projeto MMIR, a minha função como investigador envolveu a prototipagem de soluções relacionadas com a recuperação e classificação de informações médicas. Além disso, colaborei com o Hospital Gaia para desenvolver e implementar estas técnicas em ambientes reais do campo médico.

**Posição:** Bolseiro de Investigação

**Referência:** PTDC/EEI-ESS/6815/2014

## **EHDEN - European Health Data Evidence Network**

O projeto EHDEN tem como objetivo construir uma rede federada de bases de dados médicas. O objetivo inicial era construir uma rede de dados composta pelas informações clínicas de 100 milhões de cidadãos europeus. Enquanto colaborador neste projeto, trabalhei a um nível mais técnico, contribuindo para protótipos, o que levou a publicações científicas em conferências.

**Posição:** Colaborador

**Referência:** EU/EFPIA Innovative Medicines Initiative 2 No 806968

## **EMIF – European Medical Information Framework**

O projeto EMIF teve como objetivo criar ferramentas e estratégias para simplificar o acesso a dados médicos de diversos estudos europeus. Neste projeto, colaborei na manutenção de uma ferramenta de gestão de estudos colaborativos e no EMIF-Catalogue.

**Posição:** Colaborador

**Referência:** EU/EFPIA Innovative Medicines Initiative Joint Undertaking No. 115372

## **GLIM – Gateway for Living Data Management**

O GLIM-BioData, coordenado pela infraestrutura BioData.pt / ELIXIR-PT, estabelece um centro nacional de competências em Gestão de Dados de Investigação (RDM) para as Ciências da Vida e da Saúde. Selecionado pela FCT-FCCN como um dos 14 Centros de RDM portugueses, o projeto fornece um ecossistema federado que apoia a partilha e reutilização de dados sensíveis e não sensíveis sob princípios FAIR e em conformidade com o GDPR, alinhando-se com iniciativas europeias como o FEGA e o 1+ Million Genomes. As suas prioridades incluem serviços de curadoria e helpdesk, roadshows nacionais, formação de \*data stewards\* e a disponibilização de ferramentas de última geração (Dataverse, eLabFTW) para a comunidade científica.

Enquanto colaborador integro o WP3 — Management of Sensitive Life and Health Data for Sharing and Reuse. Contribuo para (i) a definição de políticas e Standard Operating Procedures que regulam submissão e acesso a dados humanos, (ii) a implementação do Helpdesk FEGA Portugal, incorporando fluxos de triagem jurídica e ética, e (iii) ações de outreach junto de hospitais e biobancos para promover o uso do FEGA.

**Posição:** Colaborador **Referência:** RE-C05-i08 – Ciência Mais Digital. RI/00138/2024.

### 3.3 Candidatura a projetos científicos

Para além dos projetos científicos em estou ativamente envolvido, tenho procurado ativamente fontes de financiamento para projetos científicos. Os seguintes projetos são candidaturas que se encontram em fase de avaliação, ou não foram aprovados.

#### **BioSurfDB+: Advancing Biosurfactant Research through FAIRification and Artificial Intelligence**

Este projeto, suportado pelas infraestruturas ELIXIR e MIRRI, visa transformar o BioSurfDB num recurso FAIR baseado em grafos, enriquecido com inteligência artificial explicável para acelerar a descoberta de biosurfactantes e soluções de biorremediação. O consórcio reúne sete instituições europeias e latino-americanas de referência, cobrindo competências em curadoria biológica, interoperabilidade, inteligência artificial e sustentabilidade de serviços. A minha responsabilidade é para além da coordenação de trabalhar no desenvolvimento e validação das ferramentas de migração para uma base de dados de grafo, o desenvolvimento dos modelos preditivos e a disponibilização destes novos serviços aos utilizadores.

**Posição:** Co-coordenador **Status:** Em avaliação **Referência:** 2nd OSCARS call - ID: 02-807

#### **Federated Learning for Biomarker Discovery using Genomic and Clinical Data from Analysis in Amyotrophic Lateral Sclerosis**

Este projeto preve aplicação de aprendizagem federada a dados genómicos e clínicos de Esclerose Lateral Amiotrófica para descobrir novos biomarcadores de forma segura e interpretável; como investigador principal, irei coordenar todas as fases desde a seleção de variantes, criação de modelos preditivos e da infraestrutura federada que será pilotada no IEETA.

**Posição:** Coordenador **Status:** Em avaliação **Referência:** 2024.16754.PEX

#### **Intelligent Variant Discovery for Rare Diseases**

Este plano de investigação, propõe criar uma framework que integra teoria da informação, seleção de variantes e aprendizagem federada para identificar biomarcadores genéticos em doenças raras com garantia de privacidade.

**Posição:** Coordenador **Status:** Em avaliação **Referência:** 2024.11219.CECIND

#### **WAI - Workbench para Exploração e Análise Inteligente de Dados**

O WAI propunha-se criar uma plataforma low-code de MLOps que orquestra, em ambiente seguro e on-prem, pipelines reproduzíveis de radiologia, patologia digital e genómica, ligando ferramentas open-source, repositórios públicos e recursos HPC para que hospitais e centros de investigação usem IA sem depender de equipas técnicas. No consórcio (BMD Software, UA-IEETA, UC-FMUC/CISUC) estaria responsável pelo componente genómico: containerizar todo o fluxo de variant calling, paralelizar o alinhamento e desenvolver modelos de anotação preditiva.

**Posição:** Coordenador **Status:** Recusado **Referência:** 2024.16754.PEX

## **Federated Learning for Genomic and Clinical Data Analysis – Biomarker Discovery in Amyotrophic Lateral Sclerosis**

Este projeto exploratório propunha a aplicação de aprendizagem federada a dados genómicos e clínicos de doentes com ELA para revelar biomarcadores em modelo seguro e interpretável.

**Posição:** Colaborador **Status:** Recusado **Referência:** 2023.14011.PEX

## **Decoding the Genomic Puzzle: Uncovering Population-Specific Genomic Traits and Disease Predispositions with Machine Learning and Kolmogorov Complexity Analysis.**

Este plano de investigação, propunha-se a criar um quadro computacional que, combinando aprendizagem automática com medidas de complexidade de Kolmogorov, decifrasse a variação genética em populações europeias para revelar traços específicos e predisposições a doenças.

**Posição:** Coordenador **Status:** Recusado **Referência:** 2023.08110.CEECIND

### **3.4 Intervenção na comunidade científica**

Em termos de intervenção na comunidade científica, estive envolvido na organização de eventos como a Reunião Internacional de dados Humanos Federados do Elixir, e na organização do ELIXIR Portugal All Hands, contribuí como revisor para diversas revistas importantes, como Computers in Biology and Medicine, e participei como membro do comité de programa em conferências internacionais IEEE. Atuei como palestrante em seminários e coordenei projetos e comunidades relevantes na área de dados genómicos. Além disso, fui reconhecido com prêmios no BIOASQ Challenge 2023 e no IEEE CBMS 2018.

#### **Organização de eventos**

- *Federated Human Data International F2F Meeting* (2024), co-organizador, IST.
- *ELIXIR Portugal All Hands 3rd Edition* (13/11/2023 - 17/11/2023), co-organizador, IEETA.
- *Symposium on Bioengineering* (2014 - 2016), membro do comité organizacional, FEUP/ICBAS.

#### **Revisor em revistas**

9. Computers in Biology and Medicine
8. Journal of Biomedical Imaging
7. Artificial Intelligence in Medicine
6. Journal of Digital Imaging
5. Health Information Science and Systems
4. Briefings in Functional Genomics
3. International Journal of Medical Informatics
2. International journal of environmental research and public health
1. Life:Open Access Journal

### **Programme Committee Member**

4. IEEE 37th CBMS International Symposium on Computer-Based Medical Systems (2024), L'Aquila, Itália
3. IEEE 36th CBMS International Symposium on Computer-Based Medical Systems (2023), L'Aquila, Itália
2. IEEE 35th CBMS International Symposium on Computer-Based Medical Systems (2022), Shenzhen, China
1. HEALTHINFO, The Seventh International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (2022), Lisboa, Portugal

### **Prémios ou outras distinções**

3. *Winner of the BIOASQ Challenge 2023 Task 11B* Ethniko Kentro Ereunas Physikon Epistemon Demokritos, Greece; Google Inc, United States; Éditions Scientifiques Elsevier, France; Centro Nacional de Supercomputación, Spain
2. *Winner of the BIOASQ Challenge 2023 Task MedProcNER* Ethniko Kentro Ereunas Physikon Epistemon Demokritos, Greece; Google Inc, United States; Éditions Scientifiques Elsevier, France; Centro Nacional de Supercomputación, Spain
1. *TCCLS Best Student Paper Award IEEE* na conferência CBMS 2018, Karlstad, Suécia

### **Participação em palestras convidadas**

4. "Enhancing National Research Data Accessibility through FEGA Environment", 22 de Novembro de 2024
3. "Introduction to React", 5 de Novembro de 2024
2. "Genomic Data Infrastructure Project", 20 de Maio de 2024
1. "The Portuguese Federated Human Data Community", 13 de Maio de 2024

### **Outras atividades**

6. Responsável pela gestão e migração de dados no IEETA (2024)
5. Coordenador da Comunidade Portuguesa de dados Humanos Federados em Portugal (desde 2023)
4. Coordenador do comité tecnico operacional do Federated European Genome-phenome Archive (FE GA) Portugal (desde 2023)
3. Membro da Associação Portuguesa de Reconhecimento de Padrões (desde 2021)
2. Presidente do Concelho Fiscal do Nucleo de Estudantes de Bioengenharia da FEUP (2014-2016)
1. Co-fundador do Núcleo de Estudantes de Bioengenharia da FEUP (2014)

**Actividade de divulgação**

12. Poster "Advancing FEGA Portugal" no 11th ELIXIR All Hands Meeting (2-5 de Junho de 2025, Thessaloniki, Grécia)
11. Apresentação no ELIXIR Federated Human Data F2F Meeting (19 de Novembro de 2024)
10. Apresentação no Biodata All Hands (3 de Outubro de 2024)
9. Apresentação no Bioinformatics Open Days, Braga, Portugal (16 de Março de 2024)
8. Representante do nó nacional no Workshop conjunto do Pilar II e III do GDI (22-24 de Janeiro de 2024, Barcelona)
7. Representante do nó nacional no primeiro Workshop do Pillar II do GDI (08-10 de março de 2023, Suécia)
6. Representante do nó nacional no segundo Workshop do Pillar II do GDI (07-20 de outubro de 2023, Roma)
5. Poster no Encontro Ciência 2023 (05-07 de julho de 2023)
4. Apresentação no Research Summit 2020 (24–26 de junho de 2020)
3. Apresentação no Research Summit 2019 (3–5 de julho de 2019)
2. Apresentação no Data Science Portugal
1. Poster no Students@DETI 2019 (11 de junho de 2019)

## 4 Capacidade Pedagógica

### 4.1 Constituição de equipas científicas e orientação de estudantes de doutoramento e bolseiros de pós-doutoramento

#### Equipas Científicas

Em termos de equipas científicas sou coordenador da comunidade portuguesa de dados humanos federados<sup>4</sup> que desempenha um papel fundamental na promoção do acesso internacional a dados genómicos e biomoleculares. Esta comunidade portuguesa tem como objectivo a participação de calls internacionais de dados federados humanos, partilha das iniciativas que Portugal tem desenvolvido a nível europeu na área e divulgação trabalho científico.

#### Orientações

Atualmente, exerço o papel de orientador no mestrado de Engenharia Biomédica da Universidade de Aveiro. Seguem os detalhes deste compromisso:

- **Estudante:** Inês Branco
- **Título do Projeto:** "Enhancing Genomic Discovery through Taxonomic Identification Tools"
- **Área de Estudo:** Ciência de dados e Bioinformática
- **Publicações Resultantes:** 2 artigos em conferências internacionais (aceites IEEE CIBCB 2024 e IWBBIO 2025) e 2 artigo em revista (em processo de revisão Q1).

#### Tutoria

Em termos de tutoria, contribuí significativamente para a orientação e desenvolvimento de estudantes em projetos no campo da bioinformática e tecnologia médica. Segue uma descrição concisa desses projetos:

1. **Joaquim Rosa (2024 - Presente) - Criação de uma interface para a criação de workflows e pipelines para genómica** Estou a orientar Joaquim no desenvolvimento de uma interface de utilizador voltada para a criação de workflows e pipelines no campo da genómica. O trabalho inclui a concepção de uma experiência de utilizador intuitiva e a implementação de funcionalidades avançadas para análise genómica.
2. **Alexandre Cotorobai (2024 - Presente) - Ambientes de Pesquisa Confiáveis e Desenvolvimento de Algoritmos de Aprendizagem Federada para Dados Genómicos** Estou a orientar Alexandre no desenvolvimento de ambientes de pesquisa confiáveis (Trusted Research Environments) e no projeto de algoritmos de aprendizagem federada voltados para a análise de dados genómicos, com foco na segurança e privacidade.
3. **Alexandre Lourenço (2021) - Reference-Free Sequence Classification Tool** Orientei o Alexandre no desenvolvimento de uma ferramenta para classificação de sequências genéticas sem referência externa. Minha orientação incluiu definição técnica do projeto, garantia da precisão científica e aprimoramento da usabilidade.
4. **David Silva (2018) - Dicom Validator** Acompanhei David no desenvolvimento de uma ferramenta de validação de ficheiros DICOM. O projeto visou a conformidade e a qualidade dos ficheiros usados em diagnósticos, implementando algoritmos de validação e assegurando uma

<sup>4</sup><https://biodata.pt/communities/fhd>

interface intuitiva para os usuários. A ferramenta criada foi essencial para verificação rápida da integridade dos arquivos DICOM.

## 4.2 Atividade Letiva

Até ao momento, a minha experiência letiva formal consiste em participações pontuais e colaborações em unidades curriculares da Licenciatura em Engenharia Informática da Universidade de Aveiro. Destaco:

- **Introdução ao React (2023)** — Lecionei uma aula de 2 horas como convidado na unidade curricular *Introdução à Engenharia de Software (IES)*, onde abordei conceitos básicos do *React*, incluindo criação de componentes funcionais, manipulação de estado e integração com APIs REST. Esta experiência permitiu-me interagir diretamente com os alunos, esclarecer dúvidas e apresentar exemplos práticos de desenvolvimento de software.

Apesar de ainda não ter assumido a responsabilidade integral por uma unidade curricular, esta oportunidade ajudou-me a compreender melhor as dinâmicas de ensino e a importância de adaptar os conteúdos às necessidades dos alunos.

### 4.2.1 Perspetivas de Evolução e Envolvimento Futuro

Dada a recente conclusão do meu doutoramento e o início efetivo da minha carreira de investigação, reconheço que a minha experiência pedagógica é modesta. É, por isso, minha intenção aprofundar as minhas competências pedagógicas a curto e médio prazo, através de:

- Participação mais assídua na coorientação de projetos de final de curso e dissertações de mestrado, potenciando parcerias entre as disciplinas lecionadas e as linhas de investigação do IEETA;
- Contributo para o planeamento de aulas e avaliação das unidades de *Aprendizagem Automática* e *Técnicas Avançadas de Aprendizagem Automática*, com o objetivo de estruturar sessões teórico-práticas equilibradas entre teoria e prática;
- Criação de materiais *open-source* (notebooks, scripts, tutoriais) que facilitem a adoção de metodologias de computação federada por parte dos estudantes, incentivando a experimentação e a colaboração em pequenos *hackathons*.

## 5 Atividades de extensão e de disseminação de conhecimento

Nesta secção estão descritas as atividades mais relevantes relacionadas com extensão e de disseminação do conhecimento. Em suma, foram selecionados 18 ferramentas que resultaram de trabalhos científicos ao longo dos anos.

### 5.1 Propriedade Intelectual e Patentes

As minhas contribuições para a propriedade intelectual estão principalmente relacionadas com a criação de soluções de software de código aberto para resolver desafios em várias áreas, como genómica, metagenómica, segurança de dados e complexidade algorítmica. Desenvolvi e colaborei em vários repositórios de código aberto ao longo dos anos, resultantes do meu trabalho científico nestas áreas. Enfatizando a importância do código aberto, selecionei e partilho as ferramentas que considero mais representantes e que abrangem uma ampla gama de aplicações nos respetivos campos.

#### Repositórios públicos

20. VCFX: Toolkit de ferramentas de linha de comando C/C++ projetadas para manipulação, análise e transformação eficientes de arquivos VCF.  
Código fonte: <https://github.com/ieeta-pt/VCFX>
19. PhenoQC: Kit de ferramentas leve e eficiente para controlo de qualidade (QC) de conjuntos de dados fenotípicos; valida formatos e consistência, mapeia automaticamente termos para ontologias padrão (HPO, DO, MPO), lida com dados em falta (imputação opcional), gera relatórios e visualizações e oferece tanto CLI como GUI em Streamlit.  
Código fonte: <https://github.com/jorgeMFS/PhenoQC>
18. fed\_rf: Uma implementação em Python para aprendizagem federada com Random Forests, utilizando PySyft para permitir que diversas partes treinem modelos de forma colaborativa sem compartilhar os dados raw. Esse pacote oferece funcionalidades como treino seguro, agregação ponderada, aprendizagem incremental e avaliação distribuída dos modelos resultantes. Código fonte: [https://github.com/ieeta-pt/fed\\_rf](https://github.com/ieeta-pt/fed_rf)
17. HYMET: Pipeline metagenómico híbrido para classificação taxonómica rápida e precisa. Integra técnicas de \*sketching\* (Mash) e alinhamento (Minimap2) num fluxo totalmente automatizado escrito em Python, Perl e Bash, com instalação via Bioconda ou Docker, download automático das bases de dados de referência e geração de relatórios TSV com níveis taxonómicos e confiança.  
Código fonte: <https://github.com/ieeta-pt/HYMET>
16. BioChef: Aplicação Web que traz a suite operações genómicas para o navegador via WebAssembly, permitindo análises genómicas/proteómicas rápidas, criação de workflows \*drag-and-drop\* e execução em tempo real sem instalação.  
Código fonte: <https://github.com/JorgeMFS/gto-wasm-app>

15. AlcoR: Simulação, mapeamento e visualização sem alinhamento de regiões de baixa complexidade em dados FASTA.  
Código fonte: <https://github.com/cobilab/alcor>
14. AltaiR: Uma ferramenta de análise livre de alinhamento e espaço-temporal para dados multi-FASTA.  
Código fonte: <https://github.com/cobilab/altair>
13. Dicoogle: Um software de arquivo PACS (Picture Archiving and Communication System), extensível, independente da plataforma e de código aberto que substitui as bases de dados centralizadas tradicionais por um mecanismo de indexação e recuperação mais ágil.  
Código fonte: <https://github.com/bioinformatics-ua/dicoogle>
12. RFSC: Ferramenta de classificação de sequência sem referência que utiliza classificadores de machine learning para uma classificação eficiente em contextos metagenómicos.  
Código fonte: <https://github.com/cobilab/RFSC>
11. MedProcNER: Código da participação do grupo BITUA no desafio MedProcNer do BIOASQ 2023.  
Código fonte: <https://github.com/ieeta-pt/MedProcNER>
10. BioASQ\_11B: Código da participação do grupo BITUA no desafio BioASQ 11 B do BIOASQ 2023.  
Código fonte: [https://github.com/ieeta-pt/BioASQ\\_11B](https://github.com/ieeta-pt/BioASQ_11B)
9. xgTaxonomy: Cross-reference of Genomic Taxonomy, um método de classificação metagenómica que utiliza algoritmos de compressão de dados para melhorar a precisão da identificação do organismo.  
Código fonte: <https://github.com/ieeta-pt/xgTaxonomy>
8. SecureFASTA: O toolkit de segurança FASTA, uma ferramenta Python que permite a encriptação e desencriptação de ficheiros FASTA utilizando encriptação AES e criptografia de curva elíptica (ECC).  
Código fonte: <https://github.com/bioinformatics-ua/SecureFASTA>
7. COMPACT: Uma ferramenta que executa classificação utilizando vários compressores para realizar a classificação taxonómica.  
Código fonte: <https://github.com/bioinformatics-ua/COMPACT>
6. TMCompress: Um projeto que explora a relação entre complexidade estatística e complexidade algorítmica usando máquinas de Turing simples.  
Código fonte: <https://github.com/asilab/TMCompression>
5. CANVAS: Complexity ANalysis VirAI Sequences, uma ferramenta projetada para analisar a complexidade de genomas virais utilizando algoritmos de compressão e técnicas de aprendizagem automática.  
Código fonte: <https://github.com/jorgeMFS/canvas>
4. Archaea2: Uma ferramenta para classificação com base em características de sequências arqueais usando métodos baseados em compressão. Utiliza scripts bash e Python para

pré-processamento de conjunto de dados, criação de conjunto de dados de classificação e classificação usando métodos baseados em compressão.

Código fonte: <https://github.com/jorgeMFS/Archaea2>

3. PANThER: Uma ferramenta para medir a complexidade de pinturas artísticas usando medidas baseadas em informações. Ele fornece uma sequência para quantização, normalização, recorte, binarização e cálculo de várias medidas de complexidade.

Código fonte: <https://github.com/asilab/panther>

2. TM-Neural-Finder: Uma ferramenta que usa redes neurais para procurar as regras corretas para modificar na matriz de estados de uma Máquina de Turing (TM) para produzir a saída desejada.

Código fonte: <https://github.com/bioinformatics-ua/TM-Neural-Finder>

1. SPTTM: Uma ferramenta usada para testar e otimizar a pesquisa de Máquinas de Turing. Ele permite ao usuário modificar a matriz de estados da Máquina de Turing para alcançar uma saída específica usando várias técnicas.

Código fonte: <https://github.com/jorgeMFS/spttm>

### 5.2 Comissões de normalização, projetos legislativos e de normas técnicas

Uma das minhas funções como coordenador do comité operacional do Arquivo Europeu Federado de Genoma-fenoma (FEGA) em Portugal e investigador do Genomic Data Infrastructure (GDI) incluiu a colaboração na elaboração de normas técnicas, especificamente:

8. Como líder de tarefas do projecto internacional FAIR-FEGA estou responsável pela definição das normas de operação do FEGA a nível internacional, sendo responsável pela definição procedimentos operacionais padrão (SOPs) Europeus e de templates de SOPs para os nós do FEGA.
7. No desenvolvimento de normas técnicas e criação de SOPs para o FEGA Portugal como coordenador de operações técnicas do FEGA e pelo projecto GLIM.
6. Na elaboração do Data Processing Agreement (DPA) para o nó Portugues do FEGA.
5. Na definição dos papéis e responsabilidades dos membros do nó Portugues do FEGA Portugal. Criação de uma equipa de helpdesk, DevOps, de suporte técnico, legal e técnico.
4. Na elaboração e definição de procedimentos para a instalação do nó nacional do GDI Starter-kit.
3. Estou envolvido também na elaboração e definição de procedimentos operacionais na infraestrutura GDI, garantindo a conformidade com as regulamentações de dados biomédicos europeus e nacionais nos diferentes produtos do GDI.
2. Na contribuição para discussões técnicas e definição de políticas de dados no GDI, visando a implementação de processos unificados para a troca segura de informações entre instituições.
1. No desenvolvimento de estratégias de integração de dados para melhorar a colaboração entre os nós do GDI.

### 5.3 Atividades de divulgação científica

Entre todas as minhas atividades na área, destaco a organização dos seguintes encontros científicos e/ou tecnológicos:

4. Organizador do ELIXIR Federated Human Data International F2F Meeting (2024), Lisboa, Portugal
3. Organizador ELIXIR Portugal All Hands 3rd Edition (2023), Aveiro, Portugal
2. Organizador das Reuniões da Comunidade Portuguesa de Dados Humanos Portuguesa (Mensalmente), Online
1. Organizador do Symposium em Bioengenharia (2014-2016), FEUP, Portugal

### 5.4 Palestras e seminários

Durante o meu percurso, todas as publicações em conferências nas quais fui primeiro autor (17), fui responsável por produzir os conteúdos e realizar a apresentação dos resultados nos respetivos congressos. Para além dessas atividades de divulgação, ainda realizei as seguintes palestras:

7. "Genomic Data Infrastructure Project in Portugal"(2025), como Keynote speaker no âmbito da conferência Bionexus Conference, iBIMED, Aveiro, Portugal
6. "Enhancing National Research Data Accessibility through FEGA Environment"(2024), 11º Fórum de Gestão de Dados de Investigação, Instituto Politécnico de Viseu, Portugal
5. "Roadmap and evolution of the Federated Human Data Community in Portugal"(2024), International Federated Human Data F2F Meeting, Instituto Superior Técnico, Lisboa, Portugal
4. "Genomic Data Infrastructure Project"(2024), no âmbito do projeto GDI, palestra no Seminário - círculo de palestras do IEETA, Aveiro, Portugal
3. "The Portuguese Federated Human Data Community"(2024), no âmbito do ELIXIR Portugal, webinar, Online
2. "Analysis of Probabilistic Algorithmic Information Approximation Measures on Artistic Paintings"(2021), no âmbito do meu doutoramento, Aveiro, Portugal
1. "Algorithmic-Statistical Learning"(2019), no âmbito do meu doutoramento, DSPT Day, Porto, Portugal

## 6 Atividades de gestão institucional

Nesta secção, são detalhadas as minhas atividades de gestão científica, de infraestruturas, de comunidades, e institucionais.

8. Coordenação do Comité Operacional do nó Português da rede Federated European Genome-phenome Archive (FEGA): Responsável pelo estabelecimento da estrutura organizacional, incluindo a criação de equipas de helpdesk, DevOps, e suporte técnico e legal. Envolveu a definição de papéis, o desenvolvimento de procedimentos operacionais padronizados (SOPs) e a gestão da migração da infraestrutura para a FCCN, visando aumentar a robustez e a capacidade de processamento de dados genómicos.
7. Responsável Técnico do Nó Português da Genomic Data Infrastructure (GDI) em Aveiro: Liderança da implementação técnica e operacional do nó nacional, incluindo o desempenho de funções como DevOps da infraestrutura. Responsabilidades incluíram a elaboração e definição de procedimentos para a instalação e operação do GDI Starter-kit e a contribuição para a definição de políticas de dados e de integração a nível da rede GDI.
6. Liderança de Tarefas em Projetos Internacionais ELIXIR:
  - *FAIR-FEGA (Programa ELIXIR 2024–2028)*: Líder de tarefa responsável pela definição de normas de operação do FEGA a nível internacional, incluindo o desenvolvimento de SOPs Europeus e templates para os nós da rede FEGA.
5. Coordenação e Co-coordenação de Projetos de Investigação (FCT – Projetos de Computação Científica Avançada):
  - Coordenador do projeto "Advanced Genomic Data Processing in Portuguese FEGA Node"(Ref: 2023.14342.CPCA).
  - Co-coordenador do projeto "Enhancing National Research Data Accessibility through Federated European Genome-phenome Archive"(Ref: 2023.11060.CPCA).
  - Co-coordenador do projeto "Mapping the Knowledge Landscape of Language Models"(Ref: FCT.CPCA.2022.01).
  - Co-coordenador do projeto "TMFinder - Inverse Problem Approximation"(Ref: 2022.27068.CPCA.A0).
4. Coordenação da Comunidade Portuguesa de Dados Humanos Federados (ELIXIR Portugal, desde 2023): Responsável pela definição dos objetivos estratégicos, delineamento de metas de acessibilidade e colaboração internacional. Autor do White Paper da comunidade e organizador das reuniões mensais.
3. Responsável pela gestão e migração de dados no IEETA (2024): Coordenação da transferência de datasets para a plataforma Dunas da Universidade de Aveiro, assegurando a integridade e acessibilidade dos dados.
2. Desenvolvimento de Normas Técnicas e Estruturação de Equipas para Infraestruturas Científicas:
  - Colaboração no desenvolvimento de normas técnicas e Data Processing Agreements (DPA) para o FEGA Portugal (no âmbito do projeto GLIM e das responsabilidades no FEGA).

- Contribuição para a definição de políticas e Standard Operating Procedures (SOPs) que regulam a submissão e acesso a dados humanos no âmbito do projeto GLIM – Gateway for Living Data Management (WP3).

1. Gestão Associativa e Pedagógica (Início de Carreira):

- Co-fundador do Núcleo de Estudantes de Bioengenharia da FEUP (2014).
- Presidente do Conselho Fiscal do Núcleo de Estudantes de Bioengenharia da FEUP (2014-2016), com responsabilidades de gestão fiscal e representação.
- Membro do comité organizacional do *Symposium on Bioengineering* (2014-2016, FEUP/ICBAS), contribuindo para a gestão e execução de um evento científico-pedagógico.