

brainglobe: Improve cellfinder's classification algorithm (David Ruiz)

Personal details

- **Full name:**
David Ruiz Rodríguez
- **Email:**
davidruizrodr@gmail.com
- **GitHub username:**
DavidRZ-ai
- **Zulip username:**
David Ruiz Rodríguez
- **Location & time-zone:**
Arahal, Seville, Spain / GMT+2
- **Personal website / project portfolio:**
<https://www.linkedin.com/in/david-ruiz-rodriguez/>
- **Code contribution**
<https://github.com/brainglobe/cellfinder/pull/508>

Project proposal

Synopsis

The project aims to upgrade Cellfinder by replacing its outdated ResNet-50 classifier with a modern deep learning architecture such as EfficientNet. This update is important because the current ResNet-based model, although effective, creates a significant bottleneck when processing large datasets, leading to long processing times (e.g., a day to process ~700k candidate cells). The goals are to enhance both the speed and accuracy of cell classification, integrate the improved classifier into the existing Cellfinder pipeline, and provide a robust, well-documented solution. Ultimately, the open source community would benefit from faster, more accurate cell detection capabilities and access to a modern, efficient tool that can be adapted for a range of neuroinformatics and imaging applications.

Implementation timeline

Minimal Deliverables:

- Updated classifier integrated into Cellfinder with an EfficientNet-based model.
- Performance evaluation comparing accuracy and runtime against the existing ResNet.
- Automated tests, thorough documentation, and a detailed final report (including a blog post)

Stretch Goals (Optional):

- Experiment with alternative architectures (e.g., Vision Transformers).
- Enhance support for retraining on custom data.

12-Week Timeline (25–30 hrs/week):

- **Weeks 1–2:** Literature review, environment setup, and baseline testing.
- **Weeks 3–4:** Implement and fine-tune the EfficientNet classifier.
- **Weeks 5–6:** Evaluate model performance and refine based on feedback.
- **Weeks 7–8:** Integrate the new classifier into the Cellfinder pipeline and conduct robustness testing.
- **Weeks 9–10:** Finalize testing, clean up code, and update documentation.
- **Weeks 11–12:** Draft and publish a blog post, complete final evaluations, and submit the final report.

Communication plan:

I will maintain clear, regular communication by holding weekly mentor video calls, posting daily updates on Zulip, sharing weekly progress summaries, and regularly pushing code to GitHub. Urgent issues will be resolved with additional ad-hoc calls, ensuring transparent, timely collaboration throughout the project.

Personal statement

- **Past experience**

I have been programming with Python for six years and, as I approach the completion of my final year in Computer Engineering at the University of Seville, I have developed a robust background in machine learning, and libraries like Numpy, Pandas, Keras, PyTorch, Tensorflow, and Scikit-Learn. My academic journey and hands-on experience have given me a strong foundation in leveraging tools like Keras for image data analysis and building effective classification models. I am passionate about applying my expertise to enhance cellfinder's classification system, and I am excited by the opportunity to contribute innovative solutions to this project.

- **Motivation: why this project?**

I am driven by a genuine passion for bridging innovative computational approaches with transformative biological research. The cellfinder project offers a unique opportunity to tackle challenging problems that have real-world implications in the life sciences. I am particularly excited by the prospect of contributing to an initiative that not only pushes the boundaries of current data analysis methodologies but also promises to enhance our understanding of complex cellular processes. Being part of this project means engaging with cutting-edge techniques in a collaborative setting where every improvement can significantly impact research outcomes. This convergence of technology and scientific discovery inspires me to commit wholeheartedly to advancing the project's goals.

- **Match: why me?**

My solid foundation in computer engineering and six years of programming experience enable me to quickly grasp and address the unique challenges of refining cellfinder's classification system. I excel at converting theoretical insights into innovative, real-world solutions, particularly in machine learning and image analysis. My proven ability to adapt to new technologies and deliver efficient, high-impact results makes me an ideal candidate for this project.

- **Availability**

I have arranged a vacation from June 27 to June 29, and no other engagements are scheduled during that period.

GSoC

- **GSoC experience**

I expect to obtain hands-on experience on a real project and on the open-source community

- **Are you also applying to projects with other organisations in GSoC 2025?**

No, all three projects that I'm applying to belong to NIU. My preferences in case I get selected in more than one would be:

- 1st → Improve cellfinder's classification algorithm
- 2nd → cellfinder support for two-dimensional brain images
- 3rd → Add to BrainGlobe's data visualisation tool