# Alberto Mario Ceballos Arroyo

ceballosarroyo.a@northeastern.edu | alceballosa.github.io | github.com/alceballosa | Google Scholar

## Education

**Northeastern University** *Boston, US*

Ph.D. in Computer Science, GPA: 4.0/4.0 (provisional) *Aug 2026 (expected)*

- Developing vision-language models for video understanding and medical applications, advised by Dr. Huaizu Jiang and Dr. Byron Wallace
- Partially supported by a Fulbright Ph.D. Fellowship amounting to 85 000 USD in funding across three years
- Member of the Khoury Graduate Student Association
- **Courses:** Machine Learning, Deep Learning, Comp. Vision & Pattern Recogn., Info. Visualization, Advanced Algorithms, Intensive Comp. Systems

**Universidad Nacional de Colombia** *Medellín, Colombia*

M.Sc. in Systems Engineering, GPA: 4.8/5.0 *Thesis defended on Nov 2022*

- Developed computational methods for a system enabling the construction of genomic scaffolds from dyed DNA images, under the supervision of Professors Juan Pablo Hernández-Ortiz and David C. Schwartz
- Thesis project: "A Computational Methodology for the Generation of Genomic Maps from Fluoroscanning Images"

**Universidad de Antioquia** *Medellin, Colombia*

B.A. in English-French-Spanish Translation, GPA: 4.77/5.0 *May 2022*

- Awarded full tuition fees waiver for semesters 2018-II and 2019-II, due to outstanding academic performance
- Volunteered for the EN-SP translation of several Coursera ML courses, the NYU DL course, and the Hugging Face transformers documentation

**Universidad del Magdalena** *Santa Marta, Colombia*

B.Sc. in Systems Engineering, GPA: 460/500 *Sep 2017*

- Highest graduation GPA and entrance exam score of the Systems Engineering 2012-II cohort (1 / 103)
- Distinction mention for the B.Sc. research thesis "Exploration of techniques for improving underwater images of fish in controlled settings"

## Academic Experience

**Ministry of Information and Communication Technologies/Universidad de Antioquia** *Medellín, Colombia*

Instructor for the Programming Fundamentals course *May 2021 - Jul 2021*

- 7 hours of weekly dedication: taught Programming Fundamentals for the Misión TIC program, which aims to introduce Colombian students and professionals from various backgrounds into programming

**AI Research and Development Group, Universidad Nacional de Colombia** *Medellín, Colombia*

Co-founder and coordinator of the TecnIA (Techniques in Artificial Intelligence) student research group *Oct 2019 - Dec 2021*

- Coordinated biweekly meetings with student members for different projects within the group, as well as monthly class sessions on scientific writing and machine learning
- Members of the group have since gone on to take positions at important companies in the South American data science world, such as Mercadolibre and Rappi

**Lab. for Molecular and Computational Genomics, U. of Wisconsin - Madison** *Madison, WI, US*

Visiting Student Researcher *May 2019 - Jun 2019*

- Developed image/signal processing algorithms for the Fluoroscanning system under Prof. David Schwartz. The developed algorithms served as the base for my M.Sc. thesis, through which I helped implement a novel system for the generation of genome scaffolds based on imaging data alone, with two publications submitted and under review at this moment

**Colombia Wisconsin One Health Consortium** *Medellín, Colombia*

Bioinformatician / Software Developer (Python, C++, Bash, Nextflow, HTML/CSS/JS) *Apr 2019 - Sep 2021*

- Developed computational tools and genomics protocols for the analysis of Influenza and Dengue samples using bash and nextflow. Workflows consisting of chaining outputs from up to 6 different tools were simplified into a single command line utility, enabling biologists in the Consortium to submit a single job and eliminating data wrangling procedures that used to take them between 3 and 4 hours
- Carried out data engineering tasks and statistical analytics for preparing meteorological and epidemiological Dengue time-series datasets toward the development of an EWS for Dengue outbreaks. The resulting data collections helped the Consortium in establishing critical trends and differences in Dengue outbreaks in 6 cities across vastly different regions of Colombia

**Functional Biology Research Group, Universidad Nacional de Colombia** *Medellín, Colombia*

Researcher / Software Developer (Python, PHP, HTML/CSS/JS) *Aug 2018 - Jun 2020*

- Constructed an antimicrobial peptide database and a peptide filtering tool and maintained the group's peptide visualization and analysis tools, increasing the visibility of the group's work and enabling biologists with little programming experience to extract features from peptides into .csv files
- Developed data processing and machine learning algorithms for the prediction of antimicrobial activity in peptides, comprising the AMPClass 1.0 and PepMultiFinder 2.0 modules of PepMultiTools, resulting in the discovery of 10 peptides with antimicrobial potential and a published paper covering these findings
- Trained undergraduate students in programming for web scraping and exploratory data analysis. This training enabled the group to carry out tighter interdisciplinary collaborations, with several students carrying out Computer Science M.Sc. and Ph.D. theses as part of the lab from 2019 to 2023

**Department of Computer and Decision Sciences, Universidad Nacional de Colombia** *Medellín, Colombia*

Teaching assistant for Programming Fundamentals and Computer Vision *Feb 2018 - Mar 2020*

- Served as a TA for 6 hours/week, leading coding sessions and advising student on their course projects. The first two iterations of the Computer Vision course led to the founding of the TecnIA student research group, with the first cohort being made of students that took said course
- Created Jupyter Notebooks and slides for the Computer Vision course that have served as the base for all versions of the course since, helping students solidify their knowledge about Computer Vision and leading to an increased complexity and variety of topics tackled in course projects

**Systems and Computing Research Group, Universidad del Magdalena** *Santa Marta, Colombia*

Computer Vision Researcher (Python, MATLAB) *Aug 2016 - Jul 2019*

- Implemented video summarization, underwater image enhancement, and fish segmentation algorithms, leading to two papers published in indexed journals. The work on underwater image enhancement resulted in the implementation of prototypes that have been tested in one local fish farm, reducing the need for fish to be extracted from the water to measure their size and weight

# Industry Experience

**South Pole Carbon Asset Management SAS** *Medellín, Colombia*

Machine Learning & Computer Vision Engineer (Python) *Jul 2020 - Aug 2021*

- Set up a geo-spatial data processing pipeline based on Google Earth Engine and various Python-based tools, deployed in the Azure cloud. This enabled the company to carry out analyses on demand using cloud-based compute instead of queuing jobs in local compute resources
- Developed remote-sensing based land cover classification tools based on DeepLabV3, improving over South Pole's SVM-based pipelines by 5.7% accuracy (averaged across classes) from a baseline of around 78%

**Gotta Ingeniería** *Medellín, Colombia*

Software developer (Python & ArcGIS) *Nov 2019 - Jan 2020*

- Developed a GUI extension for the HydroSED (ArcGIS-based) tool, which integrates Python/Fortran programs for simulating water quality in hydrographic basins. The tool sped up hydrological analysis at Gotta by eliminating the time spent manually converting simulation outputs into ArcGIS-compatible formats, as well as removing the potential for human error as part of said process

**MMedical, MM Group SAS** *Colombia*

Software development intern (C#/Java, HTML/CSS/JS) *Jan 2017 - Sep 2017*

- Maintained, improved, and developed web-based plugins for the MMPACS radiology visualization system, which along contributions from senior software developers helped us launch the system for use in its first clinical center

# Publications

### Journal articles

- Duque-Salazar, G.; Mendez, E.; **Ceballos-Arroyo, A.**; and Orduz, S. (2020). "***Design of antimicrobial and cytolytic peptides by computational analysis of bacterial, algal, and invertebrate proteomes***," Amino Acids, vol. 52, pp. 1403–1412. DOI: 10.1007/s00726-020-02900-w
- **Ceballos-Arroyo, A.**; Robles, S.; and Sanchez, G. (2020) "***A morphological convolutional autoencoder for segmenting pigmented skin lesions***," Engineering Letters, vol. 28, no.3, pp. 213-224
- **Ceballos-Arroyo, A.**; Robles, S.; and Sanchez, G. (2020) "***Remoción de lluvia en imágenes por medio de una arquitectura de autoencoder***," Revista Investigación e Innovación en Ingenierías. vol. 8, no. 1, pp 140-167, 2020. https://doi.org/10.17081/invinno.8.1.3608
- Sanchez, G.; **Ceballos-Arroyo A.**; and Robles, S. (2018). "***Automatic Measurement of Fish Weight and Size by Processing Underwater Hatchery Images***," Engineering Letters, vol. 26, no. 4, pp 461-472, 2018. ISSN 1816-093X
- **Ceballos-Arroyo, A.**; Diaz-Bolaño, I.; and Sanchez, G. (2017). "***Analyzing pre-processing filters sequences for underwater-image enhancement***," Contemporary Engineering Sciences, 10, 751-771. https://doi.org/10.12988/ces.2017.7880

### Conferences

- Barrios, W.; Soldan, M.; **Ceballos-Arroyo, A.**; Caba Heilbron, F. & Ghanem, B. "***Localizing Moments in Long Video Via Multimodal Guidance***". Download link. ICCV 2023, Paris.
- Patiño, D.; **Ceballos-Arroyo, A.**; Rodriguez, J. et al., (2020) "***Melanoma detection on dermoscopic images using superpixels segmentation and shape-based features***," Proc. SPIE 11330, 15th SIPAIM; https://doi.org/10.1117/12.2545300
- Rodriguez, J.; Oliveros-Oyola, L.; Jimenez-Giraldo, J.; **Ceballos-Arroyo, A.**; and Branch, J. (2020) "***Semantic segmentation of documents using a U-Net based architecture***." 8° Encuentro Institucional y 7° Distrital de Semilleros de Investigación, Bogotá, Colombia
- **Ceballos-Arroyo, A.**; Sanchez, G.; Hernandez-Ortiz, J.; Branch, J. (2019). "***Digital Image Processing Methods for Fluoroscanning***," in V Conferencia Internacional de Innovación en Ingeniería de Sistemas, Santa Marta, Colombia

# Achievements

| | | |
|---|---|---|
| 2023 | **Google Computer Science Research Mentorship Program**, graduated from the 2023-A cohort | *US* |
| 2023 | **Latin American Meeting In Artificial Intelligence - KHIPU**, admitted to and awarded an accommodation grant to participate in the event; received a 450 USD travel award from Northeastern Uni.'s PhD network | *Uruguay* |
| 2022 | **Bancolombia Datathon 2022**, 3rd place out of 150 teams in a hackathon focused on using NLP methods to recommend news to a bank's clients based on economic sector relevance | *Colombia* |
| 2021 | **Microsoft AI4Earth Grant**, awarded 15 000 USD in Azure credits, as part of the GeoAI unit of South Pole | *Colombia* |
| 2020 | **Fulbright Ph.D. Fellowship**, funding amounting to 85 000 USD for Ph.D. studies in the US | *Colombia / US* |
| 2020 | **Data Science for All – MinTIC scholarship**, 2020-I Cohort | *Colombia* |
| 2019 | ***Jóvenes Investigadores* – COLCIENCIAS Grant-Scholarship**, funding amounting to 26 000 000 COP | *Colombia* |