Michael Álvarez

CIENCIA DE CÓMPUTOS UPRRP

- ✓ 2014 BS Electronic Engineering
- ✓ 2018 MSc Applied Mathematics
- ✓ 2024 PhD Electrical Engineering

keywords: AI & Deep Learning, Computer Vision, Computational Imaging, Applied Mathematics.

My research interests include computational imaging applications, mainly using ML and DL algorithms. Estimating properties and enhancing attributes in multidimensional signals. I am passionate studying numerical analysis and PDEs, to integrate physics simulations with real-world measurements to achieve more accurate modeling.

My lab: **Modeling Advanced Artificial Networks (MAAN) Laboratory**
https://ccom.uprrp.edu/~malvarez/index.html

Former:

- ✓ Center for the Advancement of Wearable Technologies (CAWT)
- ✓ Laboratory for Applied Remote Sensing Imaging and Photonics (LARSIP)
- ✓ Computational Optics and Imaging Laboratory (COIL)

Recent publications:

- **Alvarez, M.**, et al. Virtual stain and Phase estimation using Encoder-Decoder Networks. Conference ISICN 2025. Springer Nature. (In publication)
- **Alvarez, M.**, et al. FDTD-net+: Improving Execution Time in Light Propagation Modeling. IEEE Access, 2025. (Under revision)

Sebastián Ramírez

Under-Graduate Student

- ❖ BSc Computer Science | Expected graduation: 2027
- ❑ Web Application Development, Computer Vision & AI

S-STEM Scholars & PI Meeting 2026

https://sstemrec.aaas.org/event/2026-s-stem-scholars-pi-meeting/

Hyperspectral Classification Methods for Future PFAS Monitoring in Coastal Environments

Exploring Binary Machine Learning Classification Methods on Benchmark Datasets

Sebastián A. Ramírez-González[1], Michael A. Álvarez-Navarro[1]
[1]Department of Computer Science, University of Puerto Rico, Río Piedras

Introduction

Plastic contamination in coastal environments is an increasingly relevant global issue, receiving growing attention due to its environmental and public health impacts. Traditional monitoring approaches often rely on in situ detection, which can be limited by logistical constraints, cost, and the need for specialized training. As a result, remote sensing techniques, such as hyperspectral imaging from satellite platforms, have emerged as promising alternatives for monitoring. However, establishing reliable relationships between spectral signatures acquired remotely and laboratory-verified contaminant measurements remains challenging due to sparse labeled observations, differences in spatial resolution, and the limited availability of synchronized datasets. This work validates machine learning classification methods on benchmark hyperspectral datasets and integrates them into the interactive PR-CREWmq dashboard, establishing the foundation for future PFAS monitoring using satellite imagery and in situ measurements.

Project Objective. Perform a binary classification to distinguish between healthy and contaminated water sources.

Methodology

For the analysis of standard segmentation techniques, the Indian Pines dataset (Figure 1) was utilized primarily. The dataset consists of 145 × 145 pixels, 200 spectral bands ranging from 400 nm to 2500 nm, and a spatial resolution of 20 meters per pixel. The ground truth identifies 16 land cover classes [1].

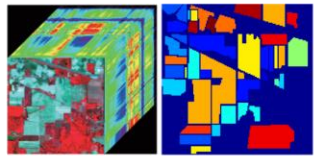


Figure 1: (a) Indian Pines hyperspectral data cube. **(b)** Indian Pines ground truth.

The data was preprocessed using the following procedure:

1. Convert the ground truth to a binary map.
2. Flatten the 145 × 145 × 200 cube into a 21,025 × 200 matrix of samples and features.
3. Normalize data using the Scikit Learn StandardScaler [2].
4. Split the data into a training set consisting of 70% of the samples and a test set with 30% of the samples.

Afterwards, three classifiers [3] were applied to both the original flattened data and the normalized data in two separate runs:

- **KMeans**: unsupervised clustering algorithm partitioning the data into two clusters based on spectral similarity.
- **Random Forest (RF)**: supervised approach using 10 decision trees using majority voting for prediction.
- **Euclidean Distance (ED)**: supervised K-Nearest Neighbors (KNN) algorithm with K = 1 and brute force neighbor calculation, corresponding to a standard Euclidean distance classifier.

Data Limitation. Majority of water quality monitoring research suffers from a lack of sufficient labeled data from remote sensing [4]. Our classification approach must achieve high accuracy with a limited number of samples.

Preliminary Results

The metrics used to evaluate the classifiers are ones present in the Scikit Learn standard: average accuracy, precision, and recall, see Tables 1 and 2.

Model	Accuracy	Precision	Recall
KMeans	0.690	0.676	0.699
RF	**0.944**	**0.953**	0.930
ED	0.935	0.927	**0.941**

Table 1: Comparison of classifications on the Indian Pines dataset using original data.

Model	Accuracy	Precision	Recall
KMeans	0.713	0.707	0.702
RF	**0.943**	**0.953**	0.930
ED	0.933	0.925	**0.939**

Table 2: Comparison of classifications on the Indian Pines dataset using scaled data.

The unsupervised KMeans algorithm yielded the lowest accuracy, precision, and recall scores (71%, 71%, and 70%, respectively). Meanwhile, the supervised Random Forest classifier obtained the highest accuracy and precision, yielding 94% and 95% respectively. The K-Nearest Neighbors algorithm with K = 1 corresponding to the Euclidean distance classifier obtained a recall score of 94% of all samples, the highest of the three models.

Figure 2 compares the predicted classification maps obtained by the three evaluated methods against the binary ground truth.

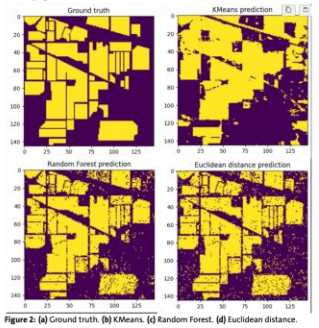


Figure 2: (a) Ground truth. **(b)** KMeans. **(c)** Random Forest. **(d)** Euclidean distance.

Acknowledgments

This work was conducted as part of the Puerto Rico Center for Research and Education in Water Monitoring and Quality (PR-CREWmq), led by Dr. Eduardo Nicolau. The authors gratefully acknowledge the support of Dr. Liz M. Díaz-Vázquez and the mentorship of Dr. Michael A. Álvarez-Navarro and Dr. Juan M. López-Encarnación.

Conference Submission (in preparation)

Analysis and Discussion

Overall, the supervised Random Forest and K-Nearest Neighbors models obtained the most accurate results. It is evident that the unsupervised KMeans classifier, while acceptable, delivered the poorest performance out of all the chosen methods. However, it is crucial to improve the accuracy of unsupervised approaches for upcoming water quality monitoring research due to the lack of accessible labeled data.

To achieve this, it may be viable to extend the data preprocessing pipeline. This experiment showed that normalizing the data presents a slight improvement on KMeans results, but a simultaneous downgrade to the supervised alternatives. Reducing the data dimensionality using principal component analysis (PCA) and introducing a spatial bias may improve the performance of unsupervised algorithms.

Future Work

The next steps for the research shall be:

1. **Extend data preprocessing**. Apply PCA and spatial bias to data to achieve higher accuracy.
2. **Apply binary classifiers to NASA data.** Apply these models tested on the Indian Pines to several hyperspectral images captured by the PACE OCI, NOAA-20 VIIRS, and Aqua MODIS sensors.
3. **Provide accessibility and validation.** Integrate the developed models into the public PR-CREWmq interactive dashboard (Figure 3), enabling researchers to analyze satellite imagery and correlate the predictions with laboratory PFAS measurements collected from field samples (Figure 4).
4. **Increase the number of classes.** Once accurate binary distinction between healthy and contaminated water is achieved in satellite data, it will be crucial to allow for more specific detection between contents of substances such as perfluorobutane sulfonate (PFBS), perfluorooctane sulfonate (PFOS), and perfluorooctanoic acid (PFOA), among others.

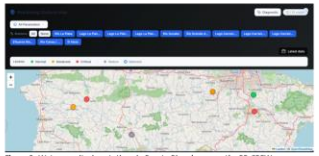


Figure 3: Water monitoring stations in Puerto Rico shown on the PR-CREWmq interactive dashboard.

Figure 4: Sample taken in the La Plata river with its PFAS content measurements.

References

[1] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015.

[2] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[4] Nathaly Wambugu, Yiping Chen, Zhenlong Xiao, Kun Tan, Mingqiang Wei, Xiaoxue Liu, and Jonathan Li. Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review. *International Journal of Applied Earth Observation and Geoinformation*, 105:102603, 2021.

Emma Moreira

Under-Graduate Student

- ❖ BSc Computer Science | Expected graduation: 2027
- ❑ Image Processing
- ❑ AI & Deep Learning

Great Minds in STEM 2026

https://gmisconference.org/

CAHSI Parameter Optimization of a Foundation Model for Cell Segmentation in Microscopy Images

Emma G. Moreira-Irizarry [1], María del C Ramos Álamo[2], Camilo Mora[3], Michael A. Alvarez-Navarro[1]

emma.moreira@upr.edu, maria.ramos21@upr.edu, camilo.mora@upr.edu, michael.alvarez2@upr.edu

[1] Department of Computer Science, University of Puerto Rico - Río Piedras
[2] Bioengineering Program, University of Puerto Rico, Mayagüez
[3] Department of Chemical Engineering, University of Puerto Rico, Mayagüez

Abstract

Accurate cell segmentation is essential for extracting quantitative information from microscopy images. This project evaluates Cellpose to investigate how parameter selection affects segmentation performance in microscopy images of fibroblasts. The generated masks will support subsequent quantitative analysis of cellular morphology.

- **Model:** Cellpose
- **Goal:** Identify suitable segmentation parameters
- **Application:** Quantitative cell morphology analysis

Introduction

Fibroblasts play an important role in tissue repair and extracellular matrix remodeling. During activation, fibroblasts can transition toward a myofibroblast phenotype, producing changes in cellular morphology such as cell size, shape, and structural organization.

- **Brightfield and phase-contrast microscopy** provide a label-free approach to monitor these changes over time[1], enabling:
  - Quantitative morphological analysis
  - Automated and accurate cell segmentation
- **Cellpose**[2] provides a deep-learning-based approach for automated cell segmentation:
  1. diameter,
  2. cell probabilities,
  3. flow threshold,
  4. nomralize and invert

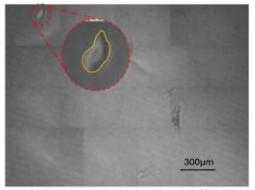


Figure 1: Example phase-contrast microscopy mosaic generated from 16 images, illustrating variation in cell size and morphology. A zoomed-in view of a cell highlights the biological complexity and visual challenges involved in manual annotation, which provides reference masks for evaluating automated segmentation.

Methodology

Brightfield and phase-contrast microscopy images (1920 × 1440 pixels) were processed using Cellpose with GPU acceleration.

- **Parameter Search.** Tested multiple Cellpose parameter combinations.
- **Mask Generation.** A segmentation mask and overlay generation of images.
- **Cross-Image Evaluation.** Compared results across images and time points.

Method 1: Parameter Optimization

Multiple Cellpose parameter combinations were systematically tested to identify promising segmentation settings.

Method 2: Multi-Image Analysis

Promising parameter configurations were applied across multiple microscopy images and experimental time points.

Preliminary Results

- Segmentation varied across parameter combinations.
- Smaller/intermediate diameter values generally detected more visible cellular regions.
- Some settings missed larger cells or produced incomplete/spurious detections.

Objective. Optimize Cellpose parameters to improve cell detection and reduce false detections.

Effect of Diameter on Segmentation

- Cell diameter strongly affected which regions were detected.
- Different diameter values changed mask completeness and cell detection.

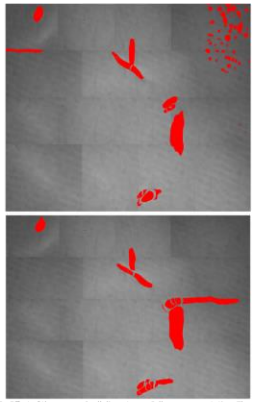


Figure 2: Effect of the assumed cell diameter on Cellpose segmentation. The same phase-contrast image was segmented using cell diameters of 200 pixels (top) and 250 pixels (bottom), demonstrating the effect of this parameter on cell detection and mask completeness.

Key Findings

- Cell detection and mask completeness were highly sensitive to the estimated cell diameter.
- Different diameter values produced substantially different masks from the same microscopy image.
- No single configuration consistently captured all visible cells without segmentation errors.
- These observations highlight the need for systematic parameter optimization across different brightfield microscopy images.

Acknowledgement

This work was partially supported by the Dr. Álvarez Seed Funds (https://ccom.uprrp.edu/~malvarez). The authors thank Maria Ramos, M.S., and Dr. Camilo Mora for providing brightfield and phase-contrast images through their work supported by the National Science Foundation (NSF) AIRFoundry under Cooperative Agreement DBI-2400135, as well as for their valuable discussions.

38th Annual Great Minds in STEM (GMiS) Conference
September 17–19, 2026

Analysis and Discussion

Segmentation performance varied with cell morphology and image characteristics. These results highlight the importance of parameter optimization for adapting Cellpose to brightfield microscopy.

- **Large and irregular cells** were more difficult to segment
- **Weak boundries** created in incomplete masks.
- Promising parameter settings **improved visual cell detection** and reduced apparent false detections.

Segmentation Challenges

Segmentation quality varied across parameter configurations. Some settings missed visually evident cells, while others introduced spurious detections, highlighting the difficulty of selecting a single configuration for all images.

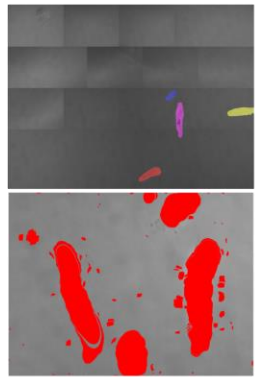


Figure 3: Examples of Cellpose segmentation challenges in phase-contrast microscopy. The top image shows missed cell detections across multiple assumed diameters, despite several cells being visually evident. The bottom image shows that a diameter of 450 pixels increases apparent cell detection but also produces spurious detections much smaller than the specified diameter.

Future Work

The next steps of this research will focus on improving the reliability and generalizability of the Cellpose segmentation workflow:

1. **Quantitative Evaluation.** Evaluate segmentation performance using quantitative metrics and validated reference masks to measure the accuracy of the generated cell masks.
2. **Parameter Generalization.** Determine whether the selected parameter combinations remain effective across different microscopy images, Z-planes, and experimental time points.

References

[1] María del C Ramos Álamo, Camilo Mora-Navarro, and Michael A Álvarez-Navarro. Morphological marker analysis for fibroblast differentiation using foundational segmentation models. In *Medical Imaging 2026: Image Processing*, volume 13925, pages 517–525. SPIE, 2026.

[2] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106, 2021.

Gladymar Colón

CIENCIA DE CÓMPUTOS UPRRP

Under-Graduate Student

- ❖ BSc Computer Science | Expected graduation: 2027
- ❑ Computer Vision
- ❑ AI & Deep Learning

1st Symposium in Computer Science Research

https://sites.google.com/upr.edu/sicc2025/

Monte Carlo Simulation of PPG Signals Using Computational Modeling

Gladymar Colon-Montanez
gladymar.colon@upr.edu
Michael A. Alvarez-Navarro
michael.alvarez2@upr.edu
Department of Computer Science, University of Puerto Rico, Río Piedras Campus

Abstract

Photoplethysmography (PPG) is a non-invasive optical technique that measures blood-volume changes in the microvascular bed of tissue. This research uses a Monte Carlo-based computational model to study light propagation in a simulated biological tissue and to generate synthetic PPG signals under different skin conditions and physiological parameters, while maintaining high computational efficiency using limited hardware resources.

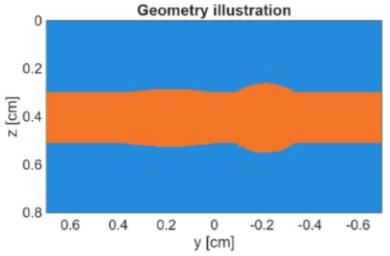
Background

The optical behavior of PPG signals can be modeled using the Radiative Transfer Equation (RTE), which describes light propagation through scattering and absorbing media. Monte Carlo (MC) methods provide a statistically robust approach for solving the RTE by simulating photon trajectories.

$$\frac{\partial L/c}{\partial t} = -\hat{s}\cdot\nabla L - \mu_t L + \mu_s \int_{4\pi} LP(\hat{s}',\hat{s})d\Omega' + S$$

This research builds upon these principles to create a computational model that simulates PPG signals across different skin optical properties.

Methodology



To recreate and customize tissue geometry and optical parameter mapping we use MC-MATLAB. This is done by integrating a 3D voxel-based light propagation model to extend the simulation into volumetric domains. This model consists of water and blood. The domain is defined as a grid of 201 x 101 x 201. Computational efficiency is achieved by using this domain size and by maintaining the simplified model with only blood and water. The simulations are currently executed on the CPU.

Methodology Cont.

Material	μ_a (cm^{-1})	μ_s (cm^{-1})	g	n
Water	0.00036	10	1.0	1.3
Blood	0.0327	75.76	0.9	1.4

The geometry function *geometryDefinition_BloodVessel* constructs this structure, media properties function *mediaPropertiesFunc* assigns the optical constants for each medium at 660 nm. These functions together define how light interacts with the different parts of the model created.

Currently we perform five trials using MC, in which it shifts the vessel position parameters *pL1* and *pL2*, which adjusts the elliptical dimensions *IDx* across a range from 0.8 [cm] to 1.25 [cm].

The equation:

$$\left(\frac{X}{a}\right)^2 + \left(\frac{Z-d}{b}\right)^2 + \left(\frac{Y-pL}{c}\right)^2 \leq 1,$$

defines the shape of the vessels, the parameters *pL1* and *pL2* shift the shape along the Y-axis, producing movement.

The equation:

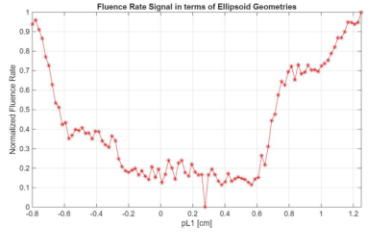
$$\left(\frac{X}{a-t}\right)^2 + \left(\frac{Z-d}{b-t}\right)^2 + \left(\frac{Y+pL}{c-t}\right)^2 \leq 1,$$

creates the smaller ellipsoids; The equation forms the inner shape that represents the blood volume which is surrounded by the water layer.

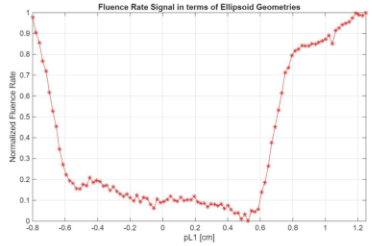
Preliminary Results

The MATLAB model successfully generated 3D representations of tissue sections and optical interactions, providing a clear basis for incorporating more complex scattering and absorption models. We approximate the PPG signal by integrating the *normalizedFluenceRate* vector, that corresponds to the relative photon density distribution within the simulated tissue. The minimum and maximum range of the fluence signal.

	$IDx: 0.33$	$IDx: 0.41$
min	2.0391	2.0518
max	2.0347	2.0516



This is the example of a signal generated using the MC solver in two media, where the ellipsoid diameter is controlled by the parameter IDx = 0.33.



Example of a signal generated using the MC solver in two media, where the ellipsoid diameter is controlled by the parameter IDx = 0.41.

Conclusion and Future Work

The MATLAB implementation enables flexible control of optical and geometric parameters to visualize light–tissue interactions. The results show that the model can reproduce basic signal changes due to variations in geometry and optical properties. Future work will focus on improving computational efficiency and incorporating more realistic physiological variables for biomedical and educational use.

Bibliography

[1] M. Alvarez, E. Arzuaga, and H. Sierra, *Using Monte Carlo simulated PPG signals to train a deep learning model to predict hemoglobin levels*, in **Emerging Topics in Artificial Intelligence (ETAI) 2023**, vol. 12655, pp. 15–22, SPIE, 2023.

[2] D. Marti, R. N. Aasbjerg, P. E. Andersen, and A. K. Hansen, *MCmatlab: An open-source, user-friendly, MATLAB-integrated three-dimensional Monte Carlo light transport solver with heat diffusion and tissue damage*, **Journal of Biomedical Optics**, vol. 23, no. 12, 2018.

[3] T. Boonya-Ananta, A. J. Rodriguez, A. Ajmal, V. N. Du Le, A. K. Hansen, J. D. Hutcheson, and J. C. Ramella-Roman, *Synthetic photoplethysmography (PPG) of the radial artery through parallelized Monte Carlo and its correlation to body mass index (BMI)*, **Scientific Reports**, vol. 11, no. 1, p. 2570, 2021.

Gabriel Torres-Morales

CIENCIA DE CÓMPUTOS UPRRP

Under-Graduate Student

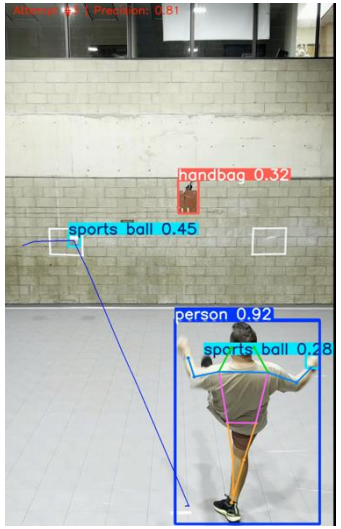
- ❖ BSc Computer Science | Jan 2023 – Present
- ❖ BB Administration | Aug 2021 – Present

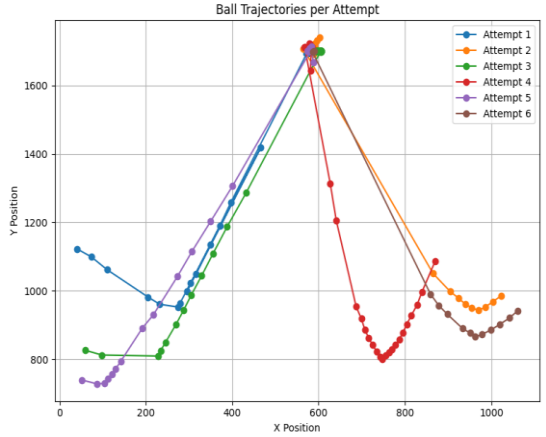
- ❑ Computer Vision
- ❑ Computer Architecture
- ❑ AI & Deep Learning

CV

GTM

Computer Vision and Artificial Intelligence in Sports Performance Analysis





Poster paper Accepted to present at CARLA 2025 (https://carlaconference.org/call-posters/). This is a high-performance computing event in the Caribbean, focused on advancing computational science, AI, and data-intensive applications.

CARLA 2025

Yadiel Mercado-Rivera

CIENCIA DE CÓMPUTOS UPRRP

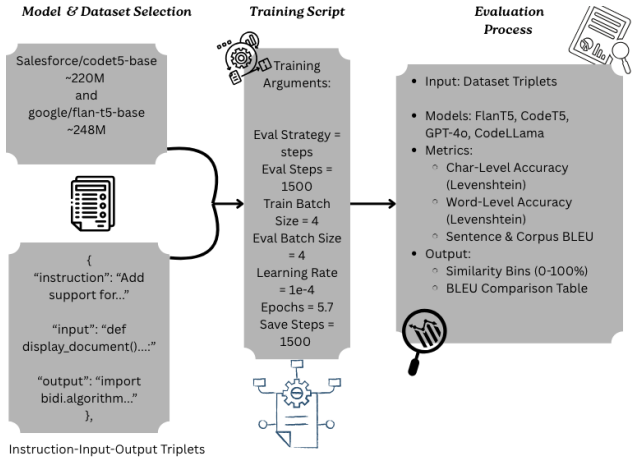
Under-Graduate Student

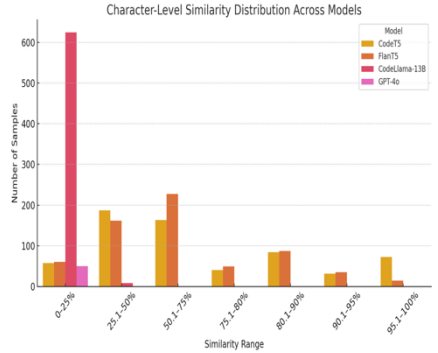
- ❖ BSc Computer Science | 2019 – Present
- ❖ Unreal Engine Developer | 2024 – Present

- ❑ Data Structures
- ❑ Cybersecurity
- ❑ Large Language Models

CV YMR

Instructional Code Editing Using Transformer Models





Poster paper Accepted to present at CARLA 2025 (https://carlaconference.org/call-posters/). This is a high-performance computing event in the Caribbean, focused on advancing computational science, AI, and data-intensive applications.

CARLA 2025