Morphological classification of radio sources in star-forming regions

Elena Díaz-Márquez[1,2,3], Gemma Busquet[2,3,4], Josep Miquel Girart[1,4], Álvaro Sánchez-Monge[1,4]

[1]Institut de Ciències de l'Espai (ICE) [2]Institut de Ciències del Cosmos (ICCUB) [3]Departament Física Quàntica i Astrofísica, Universitat de Barcelona (UB) [4]Institut d'Estudis Espacials de Catalunya (IEEC)

Introduction

Radio continuum emission at cm wavelengths from young stellar objects (YSOs) reveals insights into their **formation** and **evolution**.

The **morphology** of **protostars** is closely related to the type of radio continuum emission they emit, reflecting **underlying processes** like accretion and outflows.

Our model, using data from the Very Large Array (VLA) telescope and Stochastic Gradient Descent (SGD) aims to classify **radio source morphology** in **star-forming regions**, enhancing our understanding of star formation dynamics.

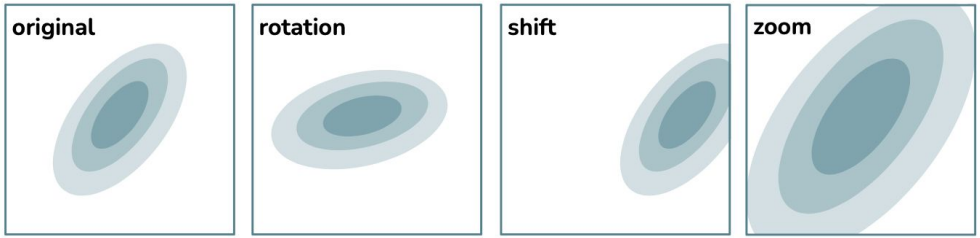
You can check the complete poster here

Dataset

Observations

Combined surveys include radio continuum observations from the **IRDC G14.22-0506** (G14.2) [1] , the **Orion Molecular Cloud 2** (OMC-2) [2], and the **Orion Nebula Cluster** (ONC) [3], and additional data from the **GLObal view on STAR formation in the Milky Way** (GLOSTAR) survey [4].

All observations were conducted using the VLA at **C-band** frequencies (~6 cm, 4–8 GHz).



Training sample

Images centered at each source with a size of 100x100 pixels

Initial dataset of **2471** sources categorized with three **morphological labels**

Since most of the detected sources are compact, in order to avoid **biased results**, we applied some modifications on the initial images, obtaining a final dataset of **4556** sources

	compact	elongated	extended
Initial sample	1307	157	1007
Aug. sample	1535	1519	1502

The modifications that we applied in order to obtain a balanced dataset were **random rotations**, **shifts** and **zooms** on the images of the less represented classes

Methods

Histogram of **O**riented **G**radients (**HOG**)

HOG measures how **local gradient orientations** correlate in images, revealing features like shapes and clumps, even in radio maps with smooth gradients.

We applied the HOG method to the images in order to reveal more features in the morphology that will be used to **train** the **classification algorithm.**

SGD Classifier

SGD optimizes the model parameters by iteratively updating them **based on gradients** computed from mini-batches of training data.

We used SGD to classify radio continuum morphologies based on learned **spatial patterns** and **characteristic features**.

Preliminary results

The classifier achieved a **mean accuracy** of 0.81

Elongated sources were classified with the highest precision and recall among the categories

However, the augmentation of the dataset **did not** substantially enhance overall **classification performance**

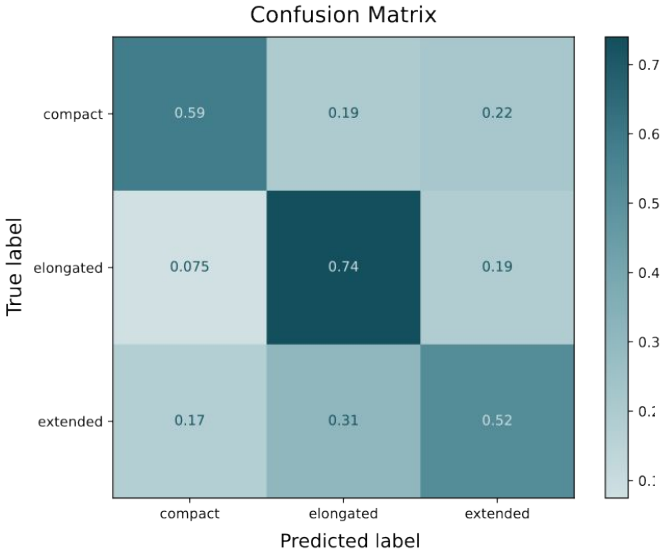
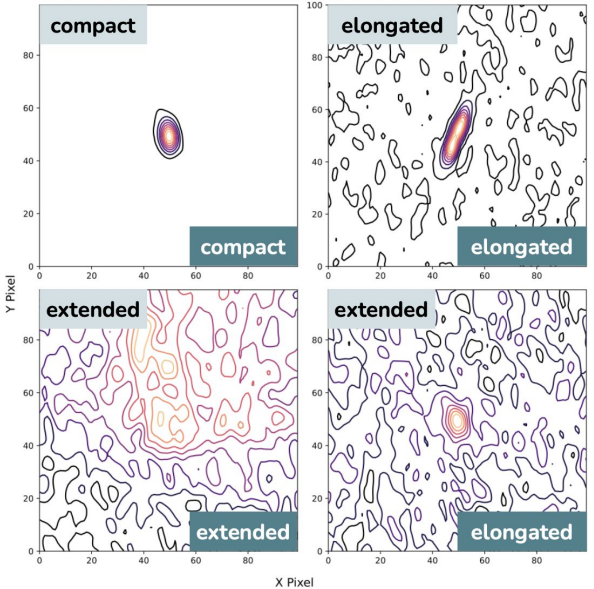


Figure 1. Confusion matrix of the morphological classes with the predicted label and the true label.

Figure 2. VLA Continuum images at C-band showing the performance of the morphological classification algorithm. **Light blue** labels indicate the true class of the sources. **Dark blue** labels indicate the predicted class.



Conclusions and future work

- Our model achieves an accuracy of ~80%, which leaves **room for improvement.**
- Traditional data augmentation techniques have not been effective, so we aim to develop **synthetic images** to enhance the dataset
- We aim to standardize **beam size** and shape to ensure consistent inputs.
- We expect to integrate more information, such as the **radio spectral index** since the classification based only on morphology can be ambiguous due to the coexistence of different radio emission mechanism.

Thank you for your attention

References:
[1] Díaz-Márquez et al. 2024 [2] Díaz-Rodríguez 2021 [3] Vargas-González et al. 2021 [4] Medina et al. 2019

Questions? Email-me at ediazm@fqa.ub.edu