



Contribution ID: 33

Type: **not specified**

## Machine Learning developments for Radio Astronomy in the SKA era

*Wednesday, 1 June 2022 09:30 (50 minutes)*

In the context of the preparation of the first SKA observations in a few years from now, more and more public surveys based on precursor and pathfinder instruments are getting released. The analysis of the resulting datasets can already be very challenging as they are getting closer to real upcoming SKA observations. Even if the type of task to perform on such datasets is often rather classical (detection, classification, denoising, etc.), they have become heavily demanding for classical approaches due to datasets size and dimensionality. It is not a surprise then, that many astronomers started to focus their work on Machine Learning approaches that demonstrated their efficiency in similar applications. However, radio-astronomical images are very different from images used to train state-of-the-art pattern recognition algorithms. Moreover, astronomers have specific expectations for the predicted results, be it in terms of robustness, reproducibility, or explainability. As a direct consequence, these methods do not always perform as well as expected when directly applied to astronomical datasets. For this reason, astronomers have started to propose in-depth modifications of widely adopted approaches and also have initiated the development of new dedicated methods.

In this talk, I will present an overview of several Machine Learning approaches that have successfully been employed for HI analysis. I will describe methods for a variety of tasks including image denoising, foreground removal, model inversion, etc. but I will put the emphasis on galaxy detection, classification, and characterization techniques, which is a necessary preliminary task for the vast majority of studies. For this, I will present a technical overview of different Machine Learning methods that have been developed by various teams that participated in the SKAO Science Data Challenge 2, which consisted of a 3D detection and characterization task inside a 1TB simulated cube of HI emission. I will discuss the main difficulties identified that are specific to this type of dataset and the various tweaks used by the different teams to mitigate them. I will also discuss what could be future technical developments for these methods in order to overcome the remaining difficulties. Finally, I will present some of the efforts being made in the application of these approaches on pathfinder and precursor instruments, along with the additional difficulties that arise, like the proper definition of learning samples.

**Presenter:** CORNU, David