



Contribution ID: 40

Type: **Oral Presentation**

## A machine learning classifier for LOFAR radio galaxy cross-matching techniques

*Wednesday, 1 June 2022 11:30 (20 minutes)*

New-generation radio telescopes like LOFAR are conducting extensive sky surveys, detecting tens of millions of radio sources. To maximise the scientific value of these surveys, radio source components must be properly associated into physical sources before being cross-matched with their optical/infrared counterparts. We use machine learning to identify those radio sources for which either source association is required or statistical cross-matching to optical/infrared catalogues is unreliable. We train a binary classifier using manual annotations from the LOFAR Two-metre Sky Survey (LoTSS) first data release. We find that, compared to a classification model based on just the radio source parameters, the addition of features of the nearest-neighbour radio sources, the potential optical host galaxy, and the radio source composition in terms of Gaussian components, all improve model performance. Our best model, a gradient boosting classifier, achieves an accuracy of 95 per cent on a balanced dataset and 96 per cent on the whole (unbalanced) sample after optimising the classification threshold. Unsurprisingly, the classifier performs best on small, unresolved radio sources, reaching almost 99 per cent accuracy for sources smaller than 15 arcseconds, but still achieves 70 per cent accuracy on resolved sources. It sends 68 per cent more sources than required to visual inspection, but this is still fewer than the manually-developed decision tree used in LoTSS, while also having a lower rate of wrongly accepted sources for statistical analysis. The results have an immediate practical application for the cross-matching of LoTSS second data release and can be generalised to other radio surveys.

**Presenter:** ALEGRE, Lara (Institute for Astronomy, University of Edinburgh)

**Session Classification:** SKA and Precursors