Classification of Radio and Multiwavelength Transients with Machine Learning

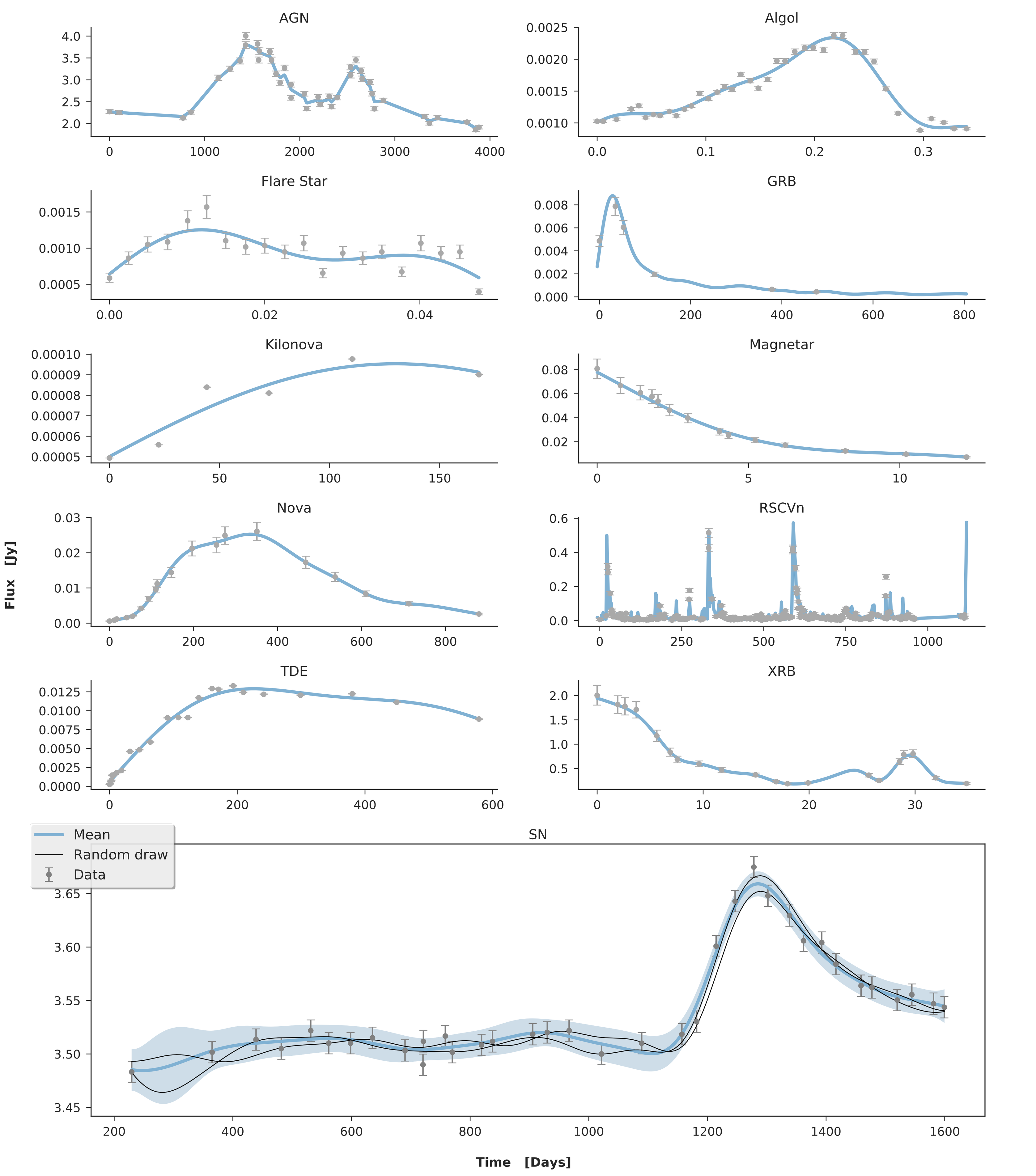
K. Sooknunan[1] · M. Lochner[2,3,5] · Bruce A. Bassett[1,2,3,4]

[1]Department of Maths and Applied Maths, University of Cape Town, Cape Town, South Africa
[2]African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg, 7945, South Africa
[3]South African Radio Astronomy Observatory, Pinelands, Cape Town 7405, South Africa
[4] South African Astronomical Observatory, Observatory, Cape Town, 7925, South Africa

In the burgeoning era of multimessenger astronomy, incorporating data from different telescopes could dramatically improve classification of events. A prime example of this is the MeerLICHT [a] telescope an optical telescope tethered to the radio telescope MeerKAT, resulting in simultaneous optical and radio observations of transients. Alert streams from telescopes such as Fermi[b] and LSST will also enable rapid coordination for multimessenger observations. Combining these data sources necessitates a new universal framework for multimessenger machine learning. We outline a method for the automatic classification of radio transients that makes use of multiwavelength data and machine learning.

[a]http://www.meerlicht.uct.ac.za/
[b]https://fermi.gsfc.nasa.gov/

Data Interpolation and Augmentation



Wavelet Decomposition

Time series data can be decomposed into a linear combination of basis functions:

$$f(x) = \sum_k a_k \phi_k(x) \tag{1}$$

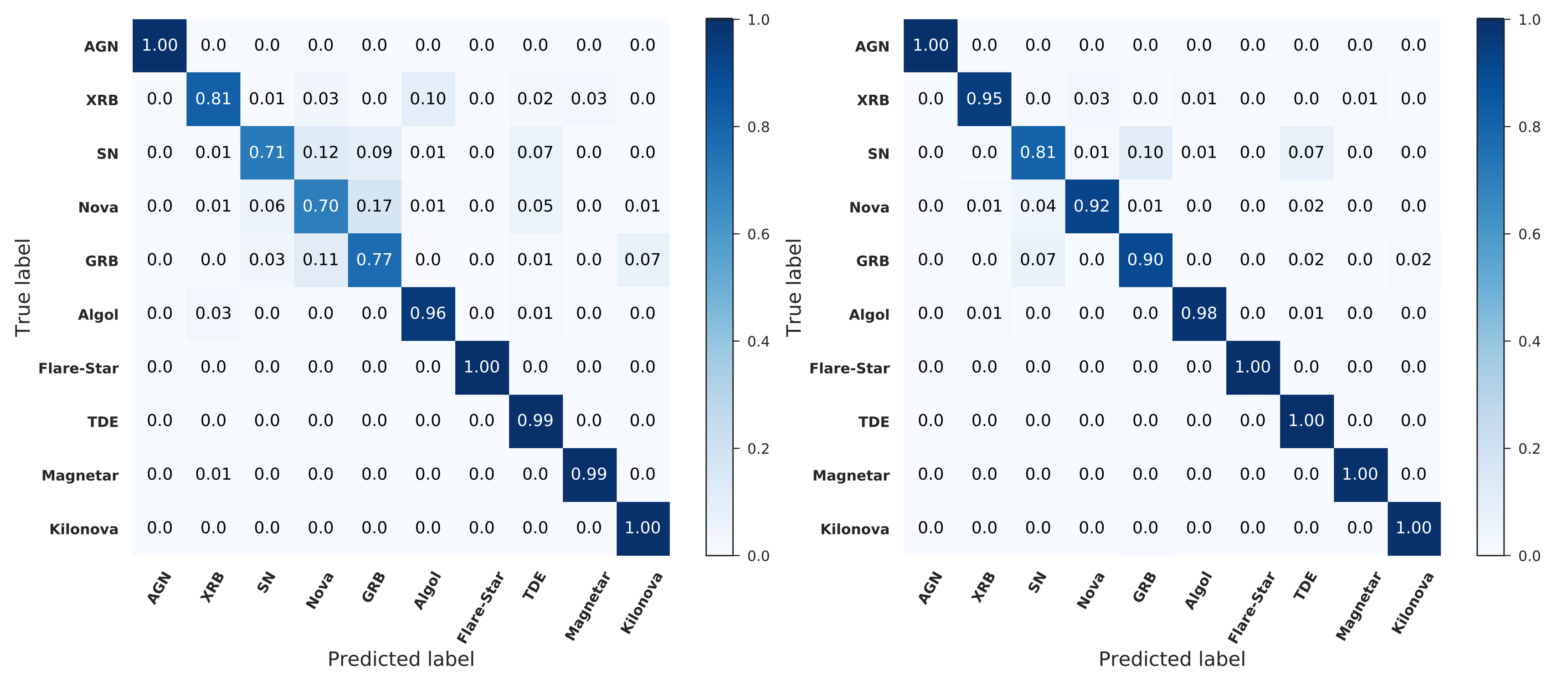
Where $\phi_k(x)$ are orthogonal basis functions and a_k are the respective coefficients. This is commonly done in the field of signal processing and can be a powerful tool for feature extraction as the set of coefficients can be used as the features for a machine learning algorithm. A critical issue with transient classification is that the transient may be observed at any point in its light curve and the algorithm must still recognise its class. Thus we require a decomposition method that is translation invariant but still sensitive to the intrinsic shape of the curve. A form of decomposition that is approximately scale- and translation-invariant is known as the stationary wavelet transform

Combining Multiple Data Sources

There are two ways information from other sources can be incorporated:

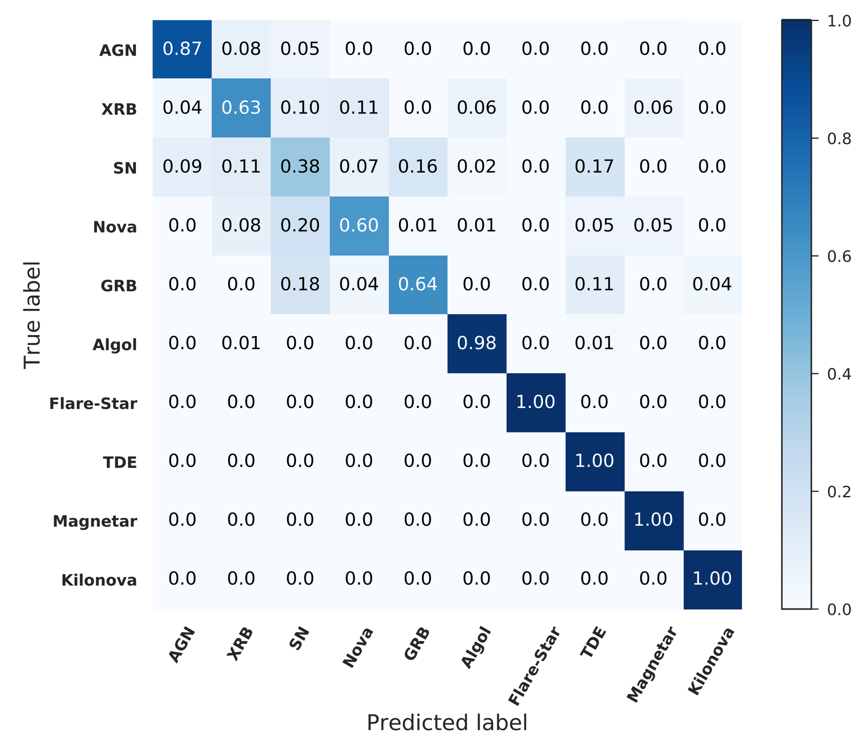
- **Probabilistic Approach:** most machine learning classification algorithms are capable of producing a score that can be interpreted as a probability of an object belonging to a particular class. To combine this with external information, such as the presence of an alert in another wavelength, we can calculate the prior probability, $P(\mathcal{C})$, of the object being in a certain class $\mathcal{C}$, given all prior information. This probability, $P(\mathcal{C})$, would then be multiplied by the probability given by the classifier to give a final probability of some object being in class $\mathcal{C}$.
- **Extra Features:** the second method is to use the information as an extra feature in the machine learning process. For example, if one has a flux measurement in any other wavelength, one could add that flux as a feature. The advantage of this approach is that correlations between the different features are learned automatically by the machine learning algorithm, potentially resulting in improved classification accuracy.

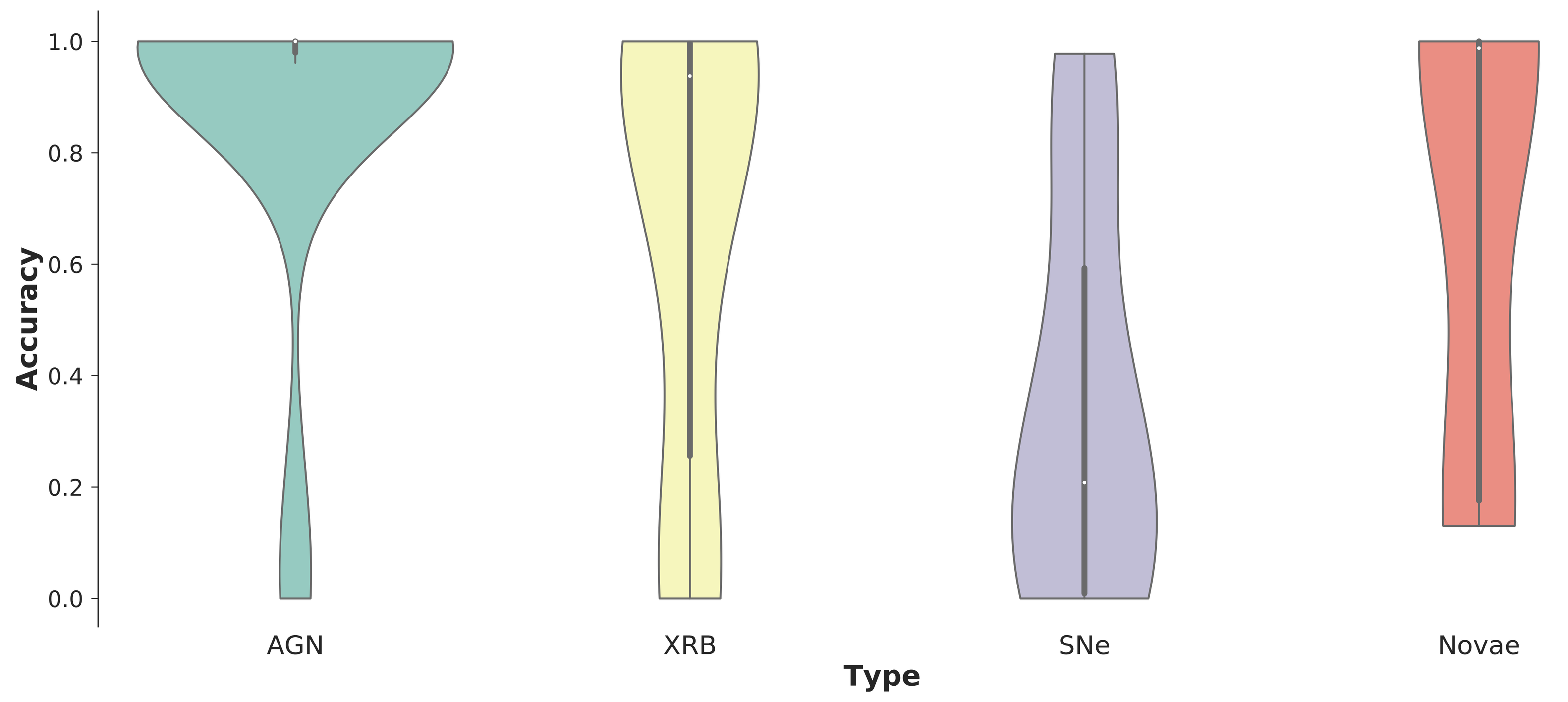
Results on Representative Training Set



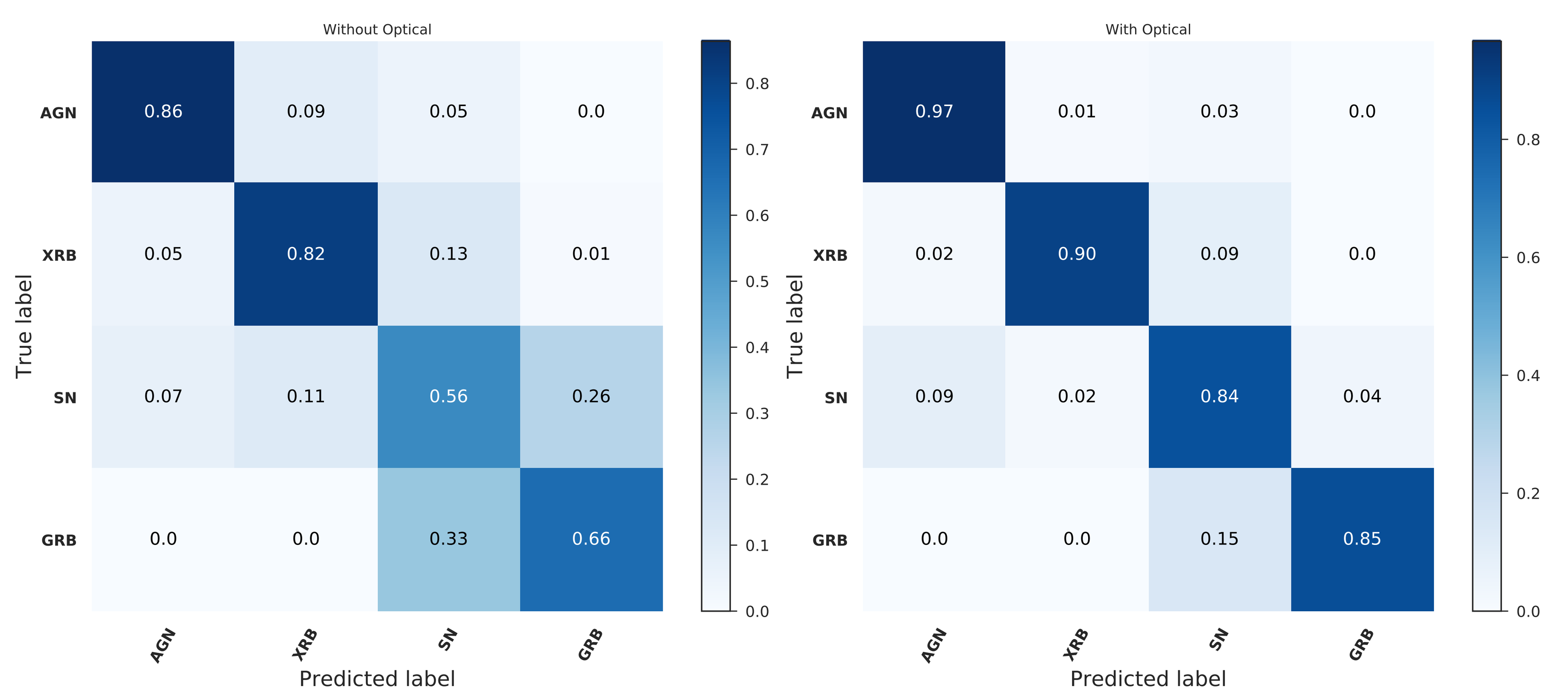
(a) Confusion Matrix without contextual information (b) Confusion matrix with contextual information

Results on Non-representative Training Set





Results on Training Set with Optical data



(a) Confusion Matrix without optical feature (b) Confusion matrix with optical feature

Acknowledgements

This work is funded by the National Astrophysics and Space Science Program South Africa as well as the SKA.