Machine learning for photometric classification of GOTO transients

Umar F. Burhanudin[1], Justyn R. Maund[1,2], and the GOTO collaboration[3]

ufburhanudin1@sheffield.ac.uk

Introduction/Motivation

The **Gravitational wave Optical Transient Observer** (GOTO) is a project that aims to identify the optical counterpart of gravitational waves (GWs) arising from binary neutron star mergers [1]. GOTO is situated at the Roque de Los Muchacos Observatory on La Palma, and conducts observations with four 40 cm diameter unit *telescopes* (UTs), for combined 18 sq. deg field of view.

During its search for GW optical counterparts, GOTO will observe many transients and variable sources not associated with GWs. The use of *machine learning* will aid in classifying different types of transients and variable sources, and also to identify GW optical counterparts and other interesting objects.

The work presented here focuses on **classifying light curves** of objects observed with GOTO into two classes: *star/variable star* and *supernova/transient*.

GOTO data

At present, GOTO conducts observations in a single wide optical bandpass (*L* filter) with 3 x 60 second exposures per pointing. With the current 4 UT configuration, GOTO achieves an average cadence of 10 days. The current 4 UT GOTO survey began in February 2019.

After data reduction is done on images, candidate sources are extracted and catalogued into the GOTO marshall. GOTO marshall sources are then cross-matched with sources on the Transient Name Server (TNS) and SIMBAD [2,3] to identify sources that have been previously classified into known classes. For this work, light curves of sources catalogued in the marshall are used. Example light curves are shown in Fig. 1.

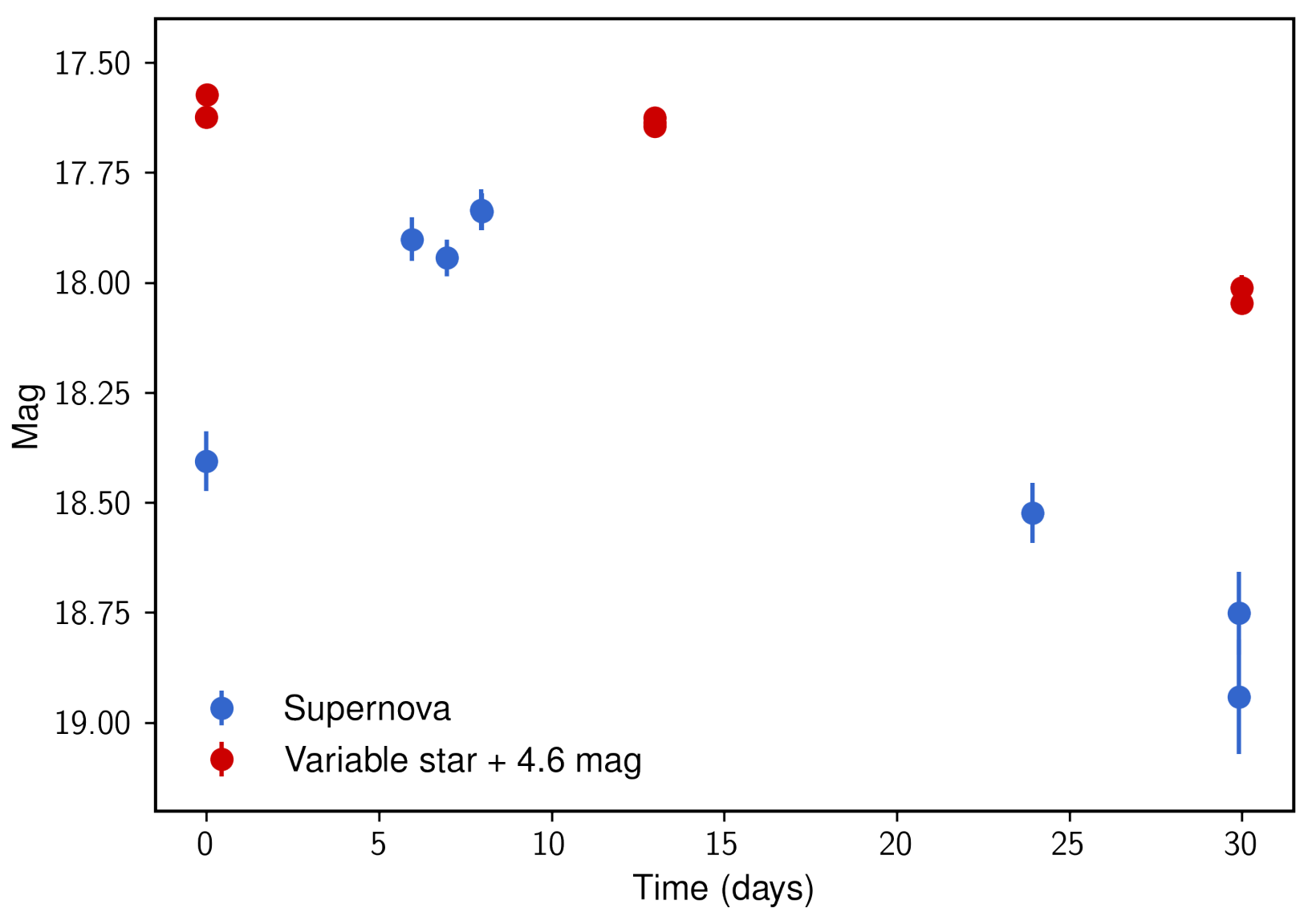


Figure 1: Lightcurves for a variable star (red) and a supernova (blue) observed with GOTO. The variable star light curve has been scaled for clarity.

Within the GOTO marshall, 68 sources had known classifications:

- 52 stars showing variability
- 9 supernovae (7 type Ia, 1 type Ia-91T-like, 1 type II)
- 7 unclassified transients

To train the classifier, GOTO light curves for supernovae, stellar flares, and variable stars were simulated to create a training and test set. The classifier was then used to attempt classification on real GOTO light curves.

GOTO currently operates with 4 unit telescopes housed in a single dome. Its scalable design allows for up to 16 unit telescopes utilising both domes in La Palma.

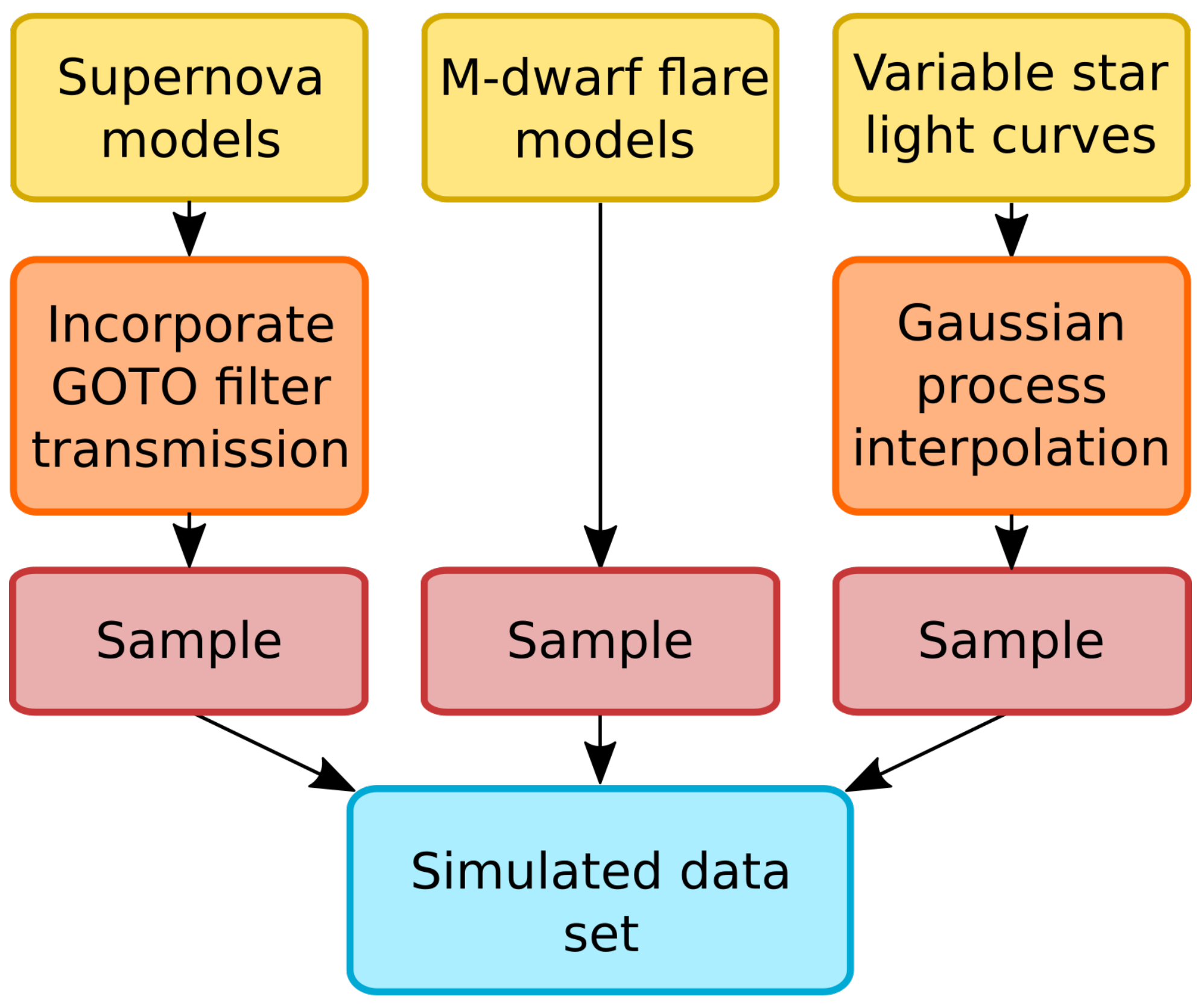


Figure 2: A flow chart outlining the process for creating a simulated data set to train and test the classifier.

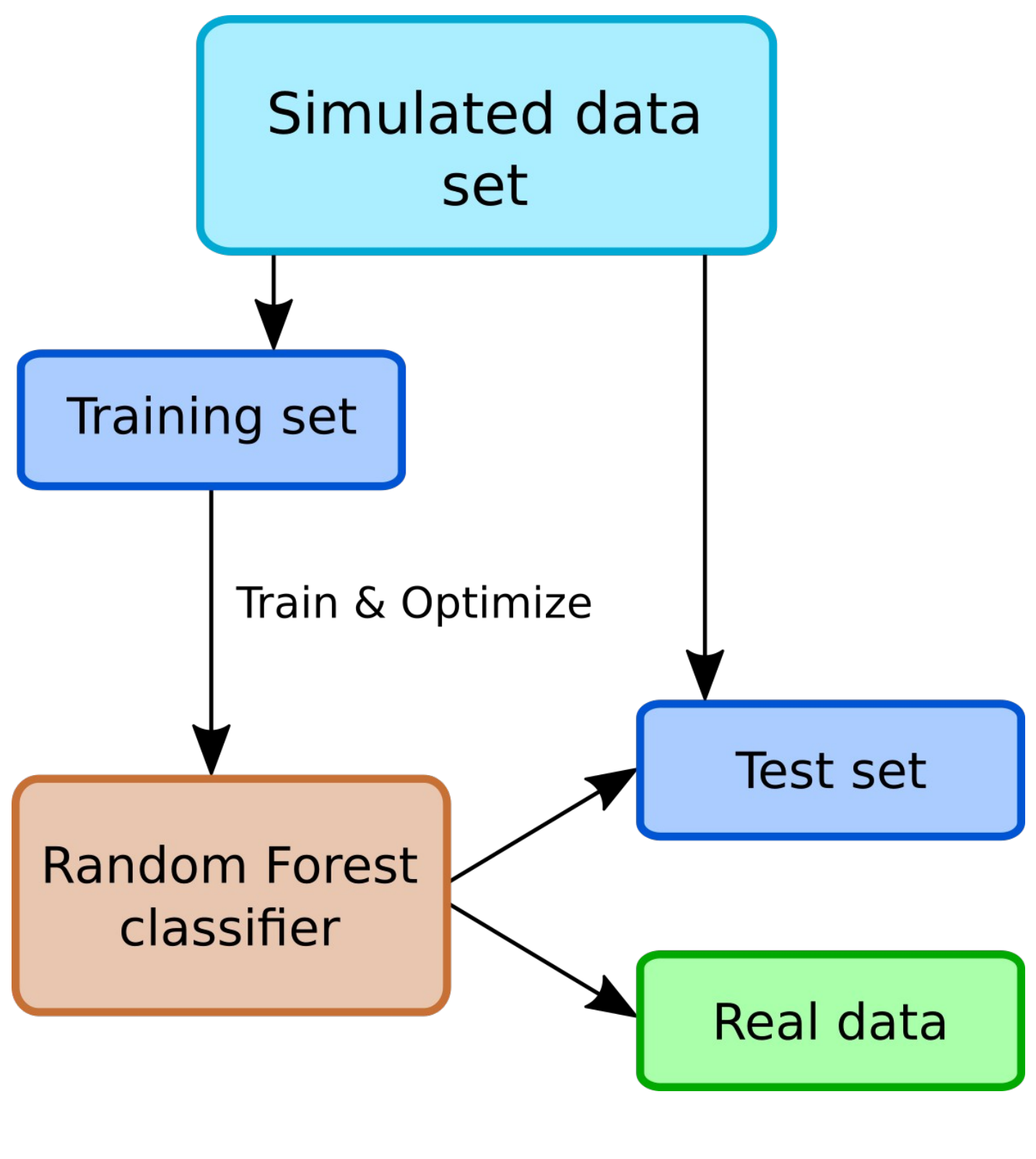


Figure 3: A flow chart outlining the machine learning workflow.

Simulating GOTO light curves

Python packages were used to simulate M-dwarf flare and supernovae light curves from models; `AltaiPony` [4] for flares and `sncosmo` [5] for supernovae. Gaussian processes were used to interpolate GOTO variable star light curves to create templates.

Model and template lightcurves were then sampled according to the GOTO cadence to create simulated light curves. Information about the GOTO *L* filter was incorporated in simulating supernovae light curves. Fig 2. outlines the steps taken to simulate GOTO light curves.

Random Forest classifier

To perform the classification task, the **random forest** (RF) algorithm was used. The RF implements an ensemble of decision trees to produce a robust classifier. For this work, the Random Forest classifier from the `scikit-learn` [6] Python package was used. Light curves are grouped into either being a *star/variable star* or *supernova/transient*.

Feature Analysis for Time Series (FATS) [7] features were used to extract features from light curves. FATS features are designed to take into account the sparse and uneven sampling of time-series data.

The simulated data were split into a *balanced training set* of 11088 samples and a test set of 44354 samples. Five-fold cross validation was performed on the training set to determine the best classifier hyperparameters, before classifying the test set. Then, classification was performed on the sample of 68 GOTO sources with known classifications. Fig. 3 outlines the machine learning workflow.

Results

Five-fold cross validation found that the classifier performed equally well for a range of hyperparameters, achieving 100% overall classification accuracy on the validation sets. A random forest classifier with *300 trees* each with a *maximum depth of 16 splits* was used to classify the test set and real data.

The classifier achieved 100% overall accuracy on the test set, and 90% overall accuracy on real data. All supernova/transient sources were correctly classified, but 7 variable stars were classified as being a supernova/transient. Fig 4. shows confusion matrices for the simulated test set and real data.

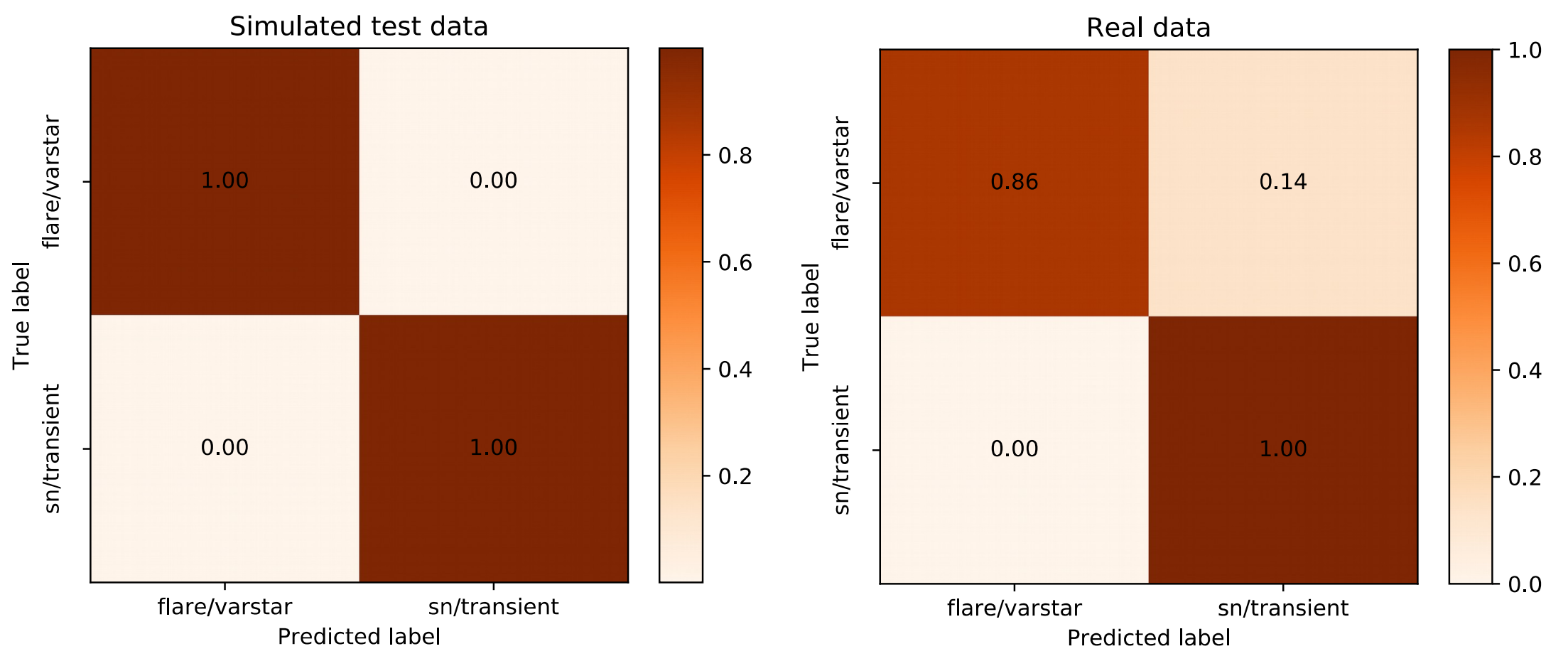


Figure 4: Confusion matrices for the test set (right) and real data (left).

Conclusions & Further Work

A summary of the work presented in this poster:

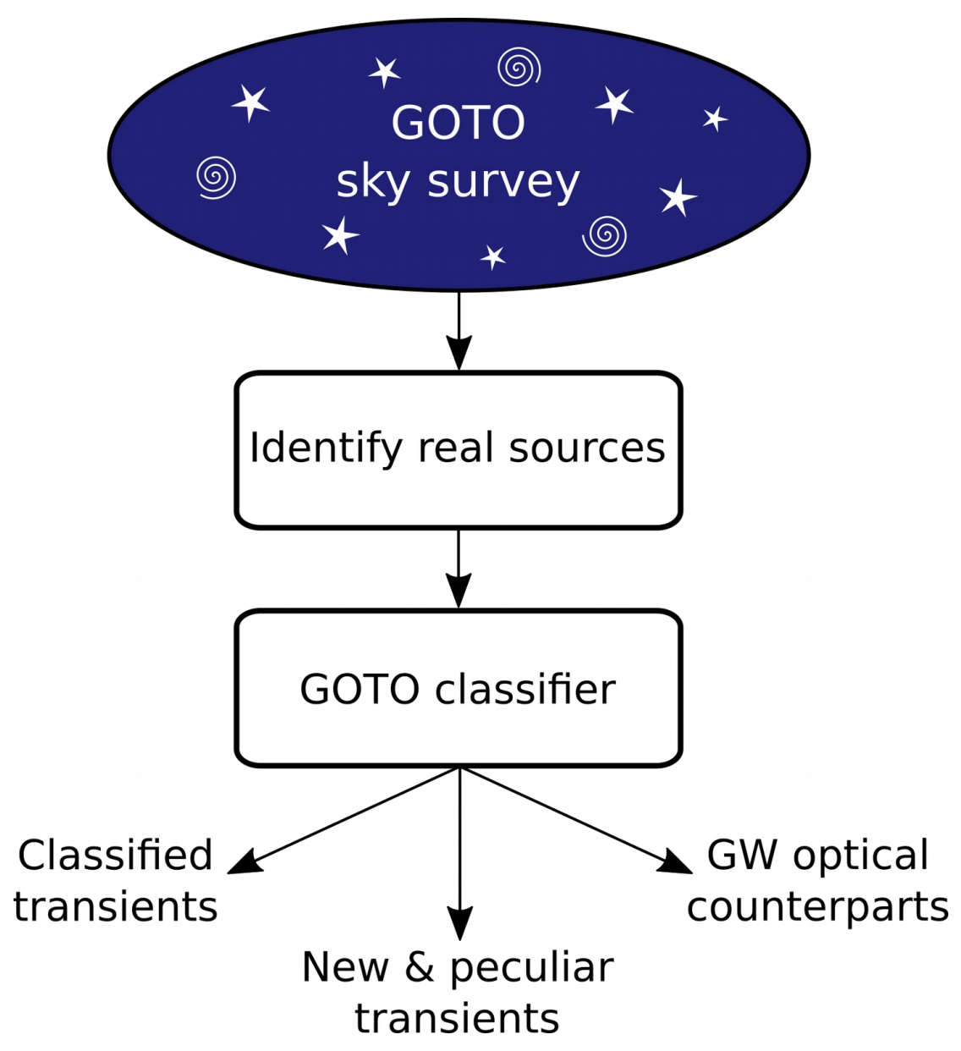
- **Aim**: classify GOTO light curves into *star/variable star* or *supernova/transient*
- Supernova and M-dwarf flare models were used to create simulated light curves
- Variable star light curves were sampled to create simulated variable star light curves
- A random forest classifier was trained and tested on a simulated data set
- The classifier was then used to classify real GOTO light curves

GOTO is still in its early phases, and so there is a limited amount of survey data that can be used. The use of a simulated data set to train and test a classifier to classify real data has shown good performance. While the work presented here shows a simple two-way classification task, it is possible to extend this model to include other types of transient events such as **kilonovae**, **superluminous supernovae** and **tidal disruption events**.

GOTO has a scalable design, allowing for additional UTs to be added. A full dome will house 8 UTs, and two full domes will allow GOTO to conduct surveys with 16 UTs. The addition of more UTs will allow for:

- A larger field of view, allowing for faster coverage of the sky and inreased cadence (avg. 7 days for 8 UTs, 3 days for 16 UTs.
- Multi-colour observations in the *R,G,B* filters.

It is also possible to perform follow-up with smaller telescopes such as the pt5m telescope operated by Sheffield and Durham [8]. The additional information gained from more UTs and follow-up observations will allow for **better characterization of transients**.



A machine learning classifier will aid GOTO in producing a variety of data products that serve multiple science goals.

---

1. University of Sheffield
2. Royal Society University Research Fellow
3. The GOTO project is a collaboration between University of Warwick, Monash University, University of Sheffield, University of Leicester, Armagh Observatory & Planetarium, National Astronomical Research Institute of Thailand, Turun Yliopisto (University of Turku), University of Manchester, University of Portsmouth, Instituto de Astrofísica de Canarias

References: 1. https://goto-observatory.org/, 2. https://wis-tns.weizmann.ac.il/, 3. Wenger et al., 2000,A&AS,143,9 , "The SIMBAD astronomical database", 4. https://github.com/ekaterinailin/AltaiPony, 5. Barbary K., et al., 2016, sncosmo/sncosmo: v1.4.0, doi:10.5281/zenodo.168220, https://doi.org/10.5281/zenodo.168220, 6. Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825, 7. Nun I. et al., 2015, ArXiv e-prints arXiv:1506.00010, 8. https://sites.google.com/site/point5metre/home

Scan this QR code to see more of my work!