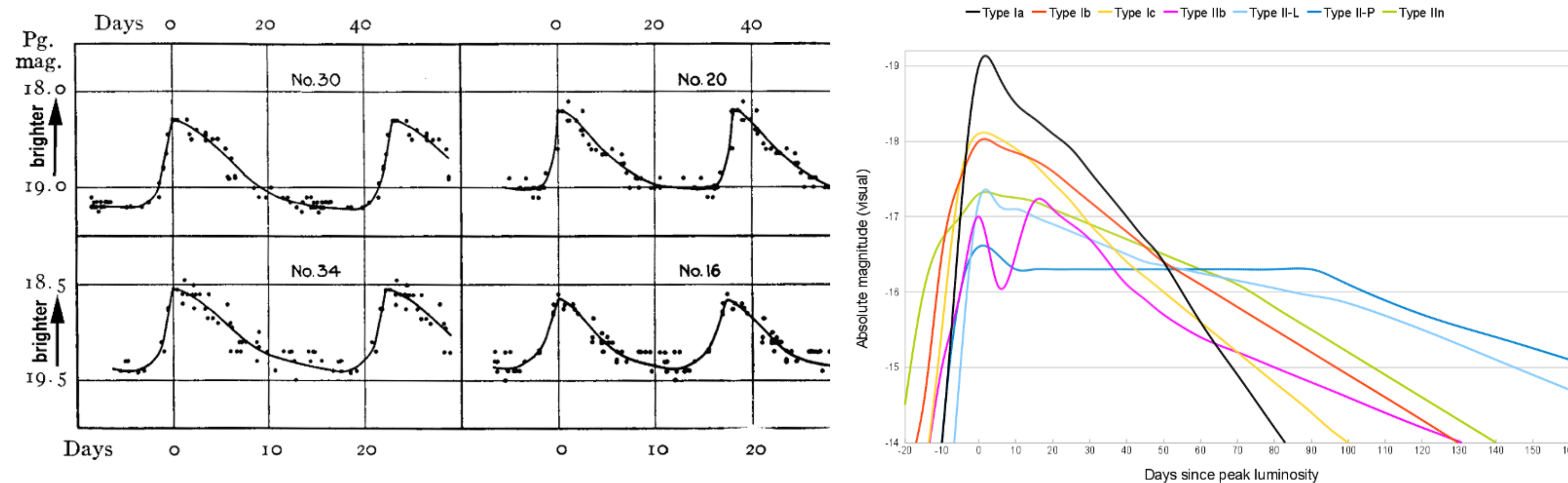


CLASSIFYING THINGS IN THE SKY

Classification of astronomical sources using simulated light curve data from the PLAsTiCC dataset

Richard Burke and Neeraj Adhikari

1. Dataset and Goal



- Photometric LSST Astronomical Time-Series Classification Challenge
- Simulation of light variable light sources in the sky
- Goal: help manage the petabytes of data from LSST
- Time-series of photometric flux in 6 different passbands
- Both intra-galactic and extra-galactic objects

2. Problems

- Gaps in the data (all sources can't be observed all the time)
- Training set non-representative of test set (sheer size difference, unseen objects)
- Lack of expertise in Astronomy
- Large uncertainty of flux observations for many objects
- Distribution of classes unequal in training and test sets

3. Preprocessing and Feature Extraction

- Cesium library for time-series handling and feature extraction
- Astropy for loading and manipulating astronomical data tables
- Descriptive features computed from training time-series
- Some of the features: median absolute deviation, variance, amplitude, skew, etc.
- Raw data: 1.4 Million individual flux samples with time and error
- Post-processed: 11 features for each of the 6 passbands, giving a $7848 * 66$ input table

4. Classification Techniques and Results

Model	Accuracy
Random Forest n=200, criterion=gini	72.26%
SVM (sigmoid) C=1.0, kernel=sigmoid	27.62%
SVM (RBF) C=1.0, kernel=RBF	29.47%
Logistic Regression regularization=L2, C=1.0	47.67%
Neural Net (MLP) hidden layers=1 with 100 neurons	58.00%
KNN (model based) k=1	64.30%