# Application of machine learning algorithms in asteroid taxonomic classification

Hanna Klimczak, Agnieszka Kryszczyńska, Dagmara Oszkiewicz, Tomasz Kwiatkowski, Emil Wilawer, Francesca DeMeo*, Wojciech Kotłowski**

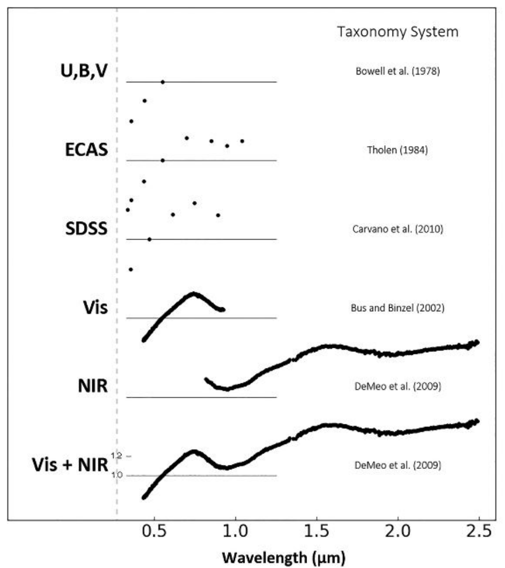
Astronomical Observatory Institute, Adam Mickiewicz University, Poznań
*Department of Earth, Atmospheric, and Planetary Sciences, MIT, USA
**Institute of Computer Science, Technical University, Poznań

*XLI Polish Astronomical Society Assembly, Toruń, 12 Sep 2023*

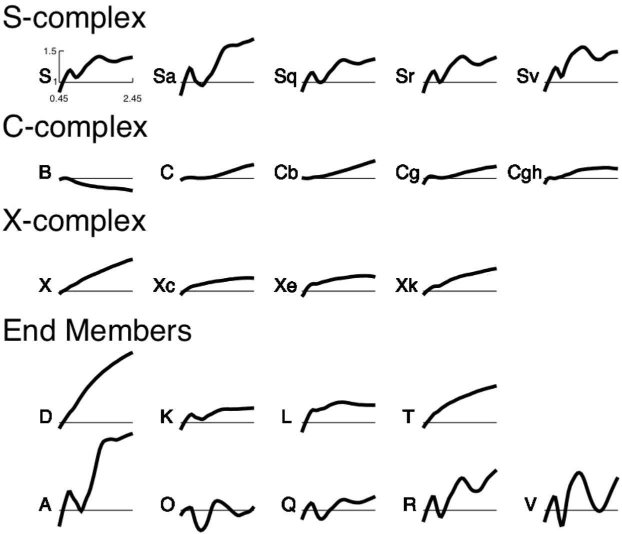
# Introduction

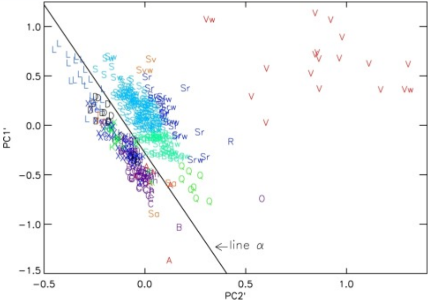
- Asteroids reflect solar radiation so their spectra are the solar spectrum modified by absorption of surface minerals.
- Dividing asteroid spectrum by the solar analogue spectrum we get a **reflectance spectrum** (often normalised to unity at 550 nm).
- Reflectace spectra are used to divide asteroids into different taxonomic types or classes.
- Best taxonomy classification (by Bus-DeMeo) is based on Vis+NIR reflectance spectra of only few thousand asteroids (out of more than 1 million observed).
- Due to the difficulty of observing spectra, Bus-DeMeo taxonomy **cannot be easily extended to new objects**.
- Multi-filter photometric sky surveys may be used for classification (colour indices instead of spectra), but **they were not optimised for asteroids**.

# Taxonomic classifications: from colour indices to spectra



# Bus-DeMeo taxonomy




Distribution of asteroids from DeMeo et al. (2009) in a reduced feature space using PCA. Each reflectance spectrum is characterised by two numbers: PC1 and PC2 (at a first step).

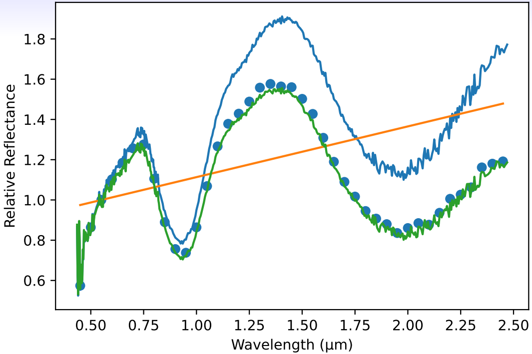
# Machine learning terminology

- **Feature**: a measurable property of a dataset to be analysed
- Features can be selected manually, based on prior knowledge of the dataset (slopes, bends, absorption bands in reflectance spectra)... or avialable wavelenghts
- ...or automatically, with one of the machine learning algorithms (the algorithms know nothing about astronomy, learn from the data)

# Aims of the work

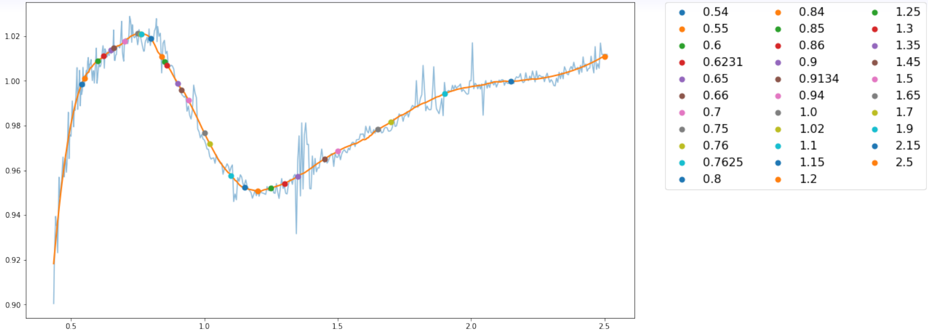
- **Check applicability of various machine learning techniques for asteroid classification** using 371 spectra from DeMeo et al. (2009) and 195 from Binzel et al. (2019).
- **Find optimal set of spectral features** (manually and automatically) for best classification.
- **Find the best set of photometric passbands** (used in existing surveys, but also new onces) for spectrophotometric classification of asteroids (spectra replaced by colour indices).

# Selection of spectral features from DeMeo spectra




The plot shows the original spectrum (blue line), the fitted slope (orange line), the spectrum after processing (green line) and final points (blue dots) of asteroid (1929) Kollaa. Spectral features (reflectance values) are based on the 0.45 $\mu m$ to 2.45 $\mu m$ spectral range, processed similarly as in DeMeo et al. (2009) with step of 0.05 $\mu m$. This resulted in a set of **41 features**.

# Selected spectral features from DeMeo spectra




The plot presents a spectrum before processing (blue line), smoothed spectrum (orange line) and **32 wavelengths** (dots) used for calculating reflectance values $f(\lambda)$ (each $f(\lambda)$ was a separate feature). They were taken from several publications.

# Reflectance differences used as features

- Reflectance values $f(\lambda)$ extracted from spectra at different wavelengths
- Reflectance differences calculated as: $f(\lambda_{begin}) - f(\lambda_{end})$

# Machine Learning Methods

- Multiclass Logistic Regression (LR)
- Naive Bayes (NB)
- Support Vector Machine (SVM)
- Gradient Boosting (GB)
- Multilayer Perceptron (MLP)

# Evaluation metrics

Accuracy of predicting asteroid taxonomy is measured with:

$$\mathrm{Acc} = \sum_{i=1}^{k} \frac{TP_i}{N},$$

and balanced prediction accuracy:

$$\mathrm{BAcc} = \frac{1}{k} \sum_{i=1}^{k} \frac{TP_i}{TP_i + FN_i},$$

where $TP_i$ is the number of correctly classified objects ("true positives") from class $i$, $k$ is the total number of classes, while $FN_i$ ("false negatives") is the number of incorrectly classified objects from class $i$, N is the total number of all samples.

# Testing our methods on asteroid types

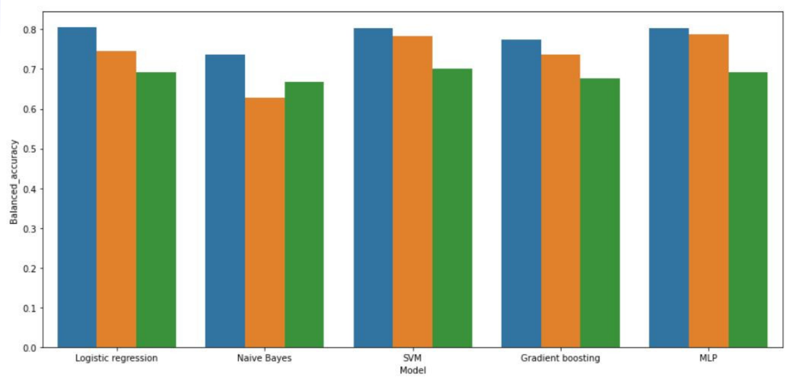
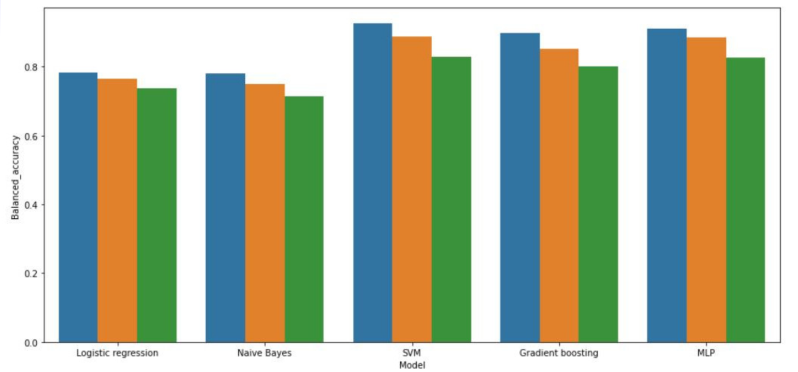



Figure: Average accuracy scores. Blue is based on reflectance differences, orange – on spectral features, green on PCA results used as features

SVM and MLP are the best results for all feature sets. The best scores per feature set: **79%** for spectral features (MLP), **80%** for reflectance differences (SVM, MLP and LR). Conclusion: our methods give reliable results best for reflectance differences.

# Testing on asteroid taxonomic complexes




SVM and MLP achieve the best results across all feature sets. The best scores per feature set are: **88%** for spectral features (MLP and SVM), **92%** for reflectance differences (SVM and MLP).

The best scores per feature set are: **88%** for spectral features (MLP and SVM), **92%** for reflectance differences (SVM and MLP).

# Feature selection - spectral features or differences

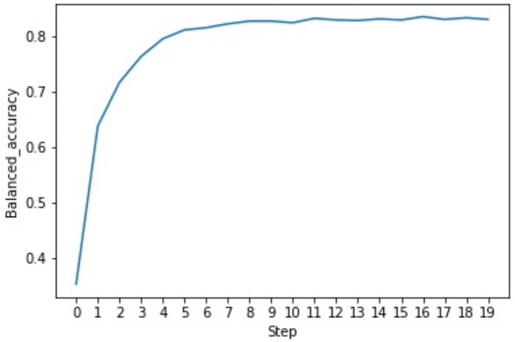
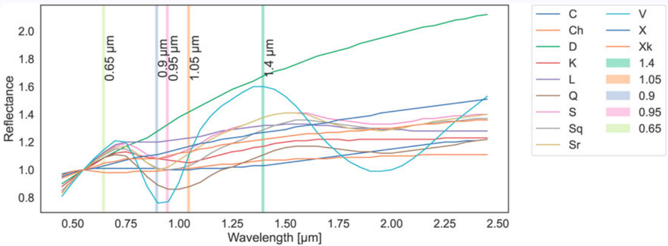
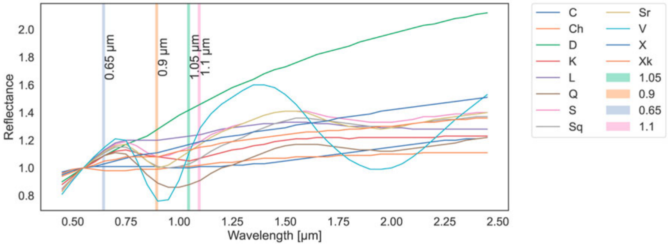



Figure: Average accuracy per step during Sequential Feature Selection

The results show that about **six/five** spectral features or reflectance differences are sufficient to obtain **81%** balanced accuracy for taxonomic types or even **93%** for complexes.

# Top spectral features for types and complexes



# Application for survey data

- Having experience in our analyses we decided to check applicability of the data from major sky surveys that already provide asteroids photometric colors or are expected in the future: **SDSS, LSST, PanSTARSS, SkyMapper, APASS, Gaia, J-PLUS, VISTA, DENIS, Euclid, 2MASS**.
- Multifilter photometry based on different sky surveys is commonly used in estimating taxonomy of individual asteroids. Colour indices are used to study the distribution of asteroids across the Solar System, distribution of families, their age relation, Solar System evolution and many others.
- Even though these surveys are optimized for other astronomical objects, they use filters that span several main spectral features of asteroids. Hence, we decided to study if they can be successfully used for taxonomic class prediction.
- Photometric colours (in magnitudes) were converted to reflectance values which were used as features.

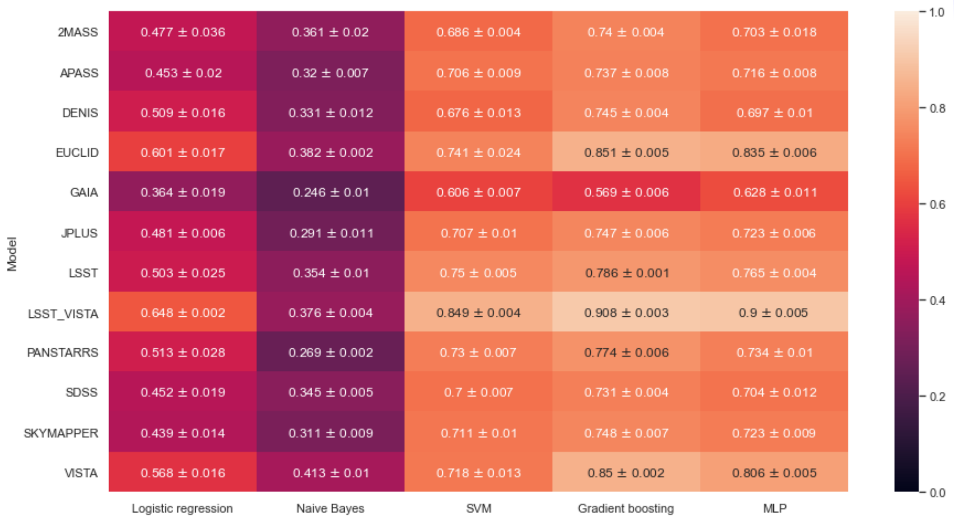
# Methods

We used the same machine learning algorithms: MLR, NB, SVM, GB, and MLP.
Due to the high imbalance of classes in our data, the main metric we used is balanced accuracy:

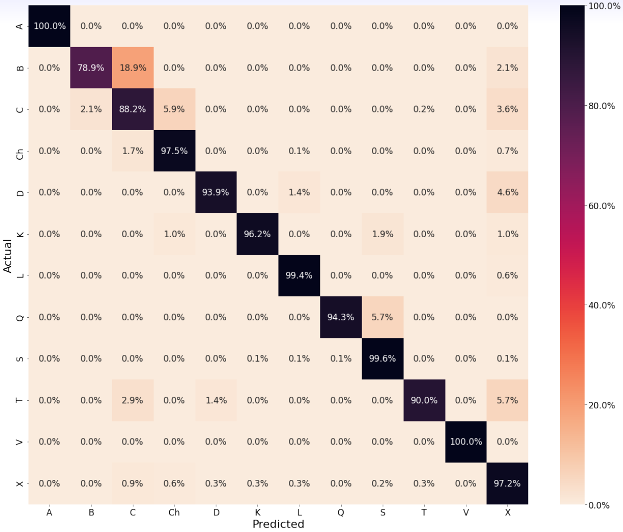
$$\mathrm{BAcc} = \frac{1}{k}\sum_{i=1}^{k}\frac{TP_i}{TP_i + FN_i}$$

where $k$ is the number of classes, $TP_i$ is the number of correctly classified objects from class $i$, and $FN_i$ the number of incorrectly classified objects from class $i$.

# Surveys comparison



# Confusion matrix for LSST+VISTA for MLP



# Results

- Using the right machine learning algorithms can improve the accuracy of asteroid taxonomic classification, the algorithm should be optimized for each sky survey individually.
- We found that best performing surveys are Euclid and LSST+VISTA reaching 85 and 91% of balanced accuracy.
- Those surveys cover the pyroxene and olivine IR absorption bands. Moreover we found that selecting the right machine learning algorithm can improve the accuracy by a factor of two in the most extreme cases.
- Among the studied methods multi-layer perceptron and gradient boosting resulted in the highest balanced accuracy.
- Modern wide-field surveys can record thousands of asteroids.

# Appendix

How to optimize future multi-filter surveys towards asteroid characterisation?

- Find a set of photometric passbands which will give optimal results for spectrophotometric classification of asteroids' into taxonomic types and classes.
- To determine the taxonomic complexes with a balanced accuracy of 85%, a set of five spectrophotometric bands is required.
- For taxonomy type determination with the balanced accuracy of 80% a set of eight bands is necessary.
- Furthermore, only a 3 band system is enough for distinguishing the C complex asteroids with 92% balanced accuracy.
- These results can be used for designing future asteroid multifilter sky surveys.
  - B1: $\lambda_0 = 0.55$ μm, range $[0.52, 0.58]$ μm
  - B2: $\lambda_0 = 1.00$ μm, range $[0.94, 1.06]$ μm
  - B3: $\lambda_0 = 1.20$ μm, range $[1.14, 1.26]$ μm
  - B4: $\lambda_0 = 1.50$ μm, range $[1.44, 1.56]$ μm
  - B5: $\lambda_0 = 1.65$ μm, range $[1.59, 1.71]$ μm

Thank you!