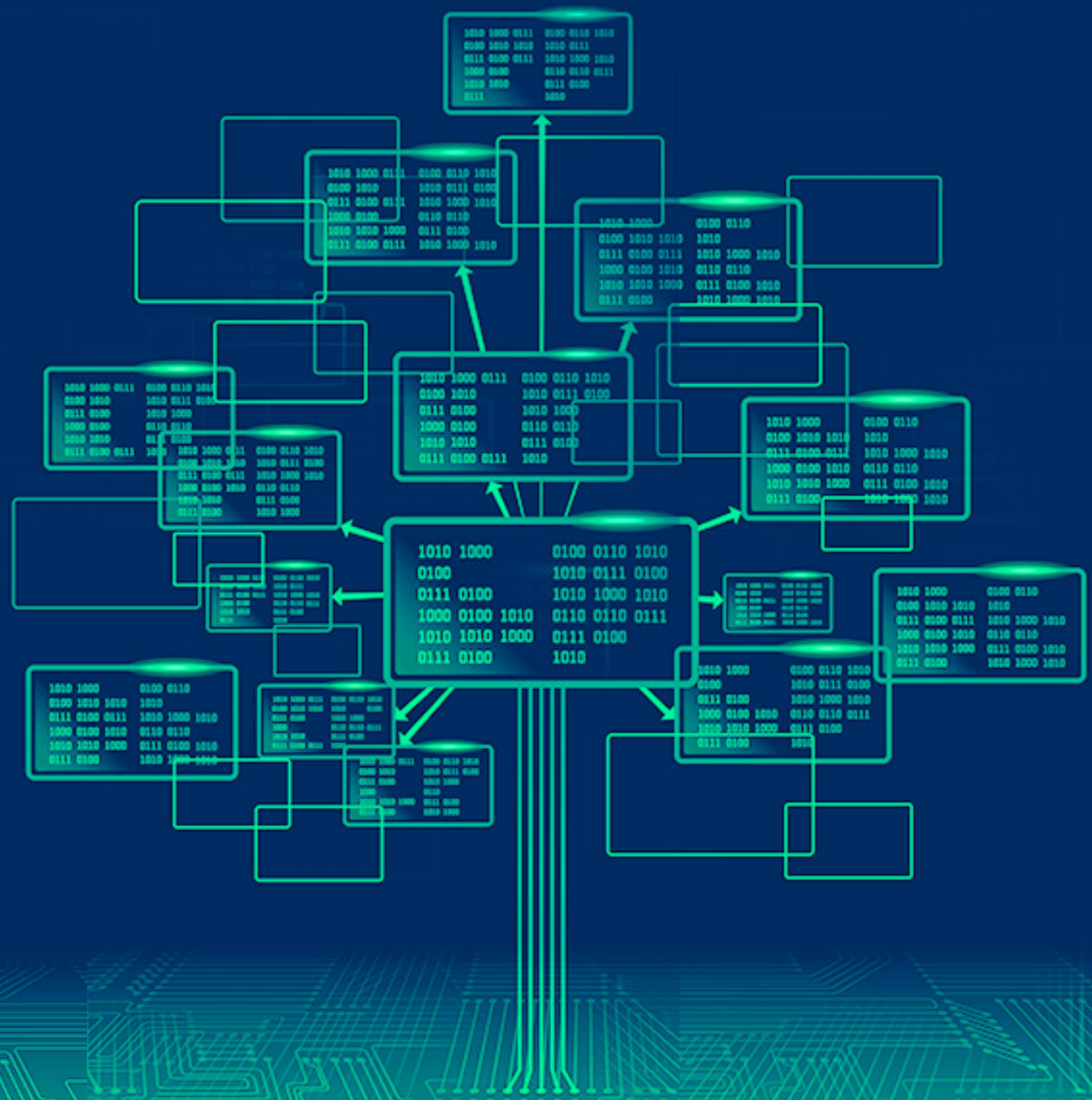
MACHINE LEARNING para clasificación automática de estrellas binarias de rayos X: preparación para la Misión espacial THESEUS

Elisa Nespoli, Adrián Pérez-Suay, Pere Blay, Julia Suso, Juan Fabregat, José Luis Gasent-Blesa, Javier Navarro-González

ESENCIA
Centro de Estudios en Ciencia de Datos e Inteligencia Artificial

viu Universidad Internacional de Valencia



theseus
transient high energy sky and early universe surveyor

1er Workshop de Investigación Centro ESenCIA - 30/01/2025

THESEUS/M7: contribution from Valencia

within the Spanish consortium

- THESEUS Scientific Contribution
- XGIS Simulator
- XGIS Imaging System Design
- Machine learning techniques applied to scientific data related to the THESEUS goals

WHY X-RAY BINARIES?

- ☑ High variability across timescales from msec to years in periodic, quasi-periodic, or aperiodic domains.
- ☑ Diverse types exist, classified from phenomenology and nature of the counterpart.

The phenomenology of X-ray emission from XRBs is broad and varied, making their classification a complex goal.

WHY X-RAY BINARIES?

- ☑ High variability across timescales from msec to years in periodic, quasi-periodic, or aperiodic domains.
- ☑ Diverse types exist, classified from phenomenology and nature of the counterpart.

The phenomenology of X-ray emission from XRBs is broad and varied, making their classification a complex goal.

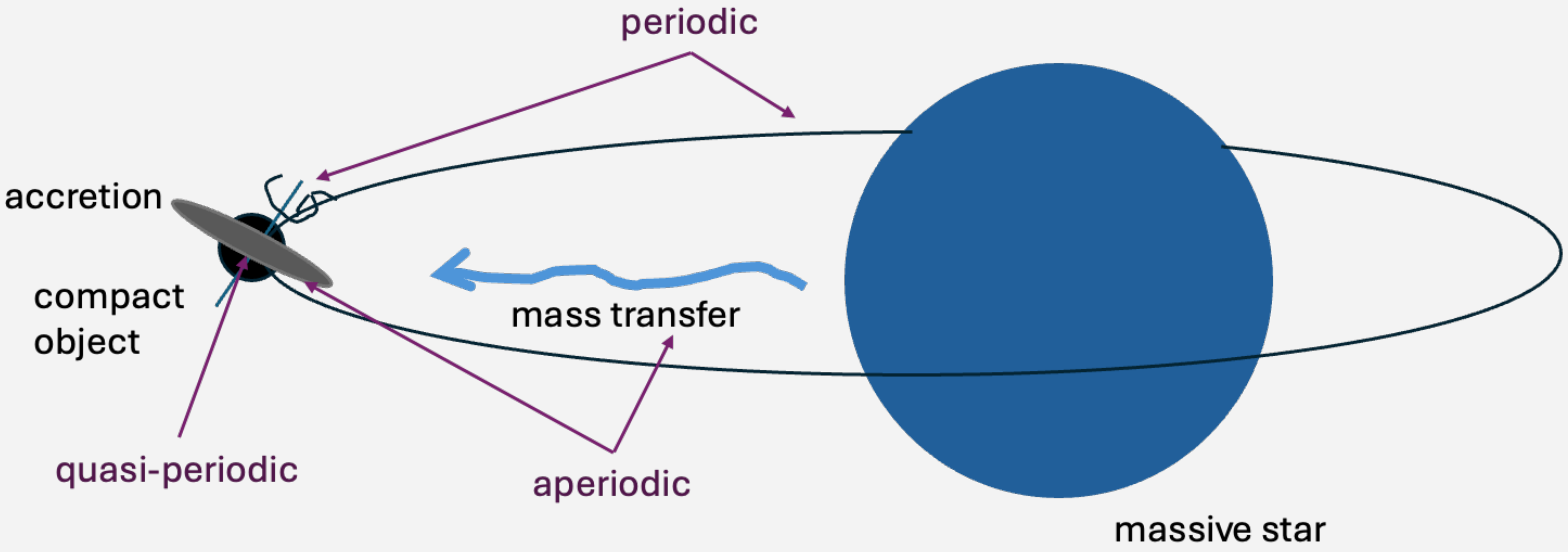
X-ray binaries: compact object (NS, BH, WD) +

- Low mass X-ray binaries: solar mass or less massive stars as counterpart.
- High mass X-ray binaries: > 8 solar masses counterpart.
- Intermediate systems > systems that may not fit in the previous categories.

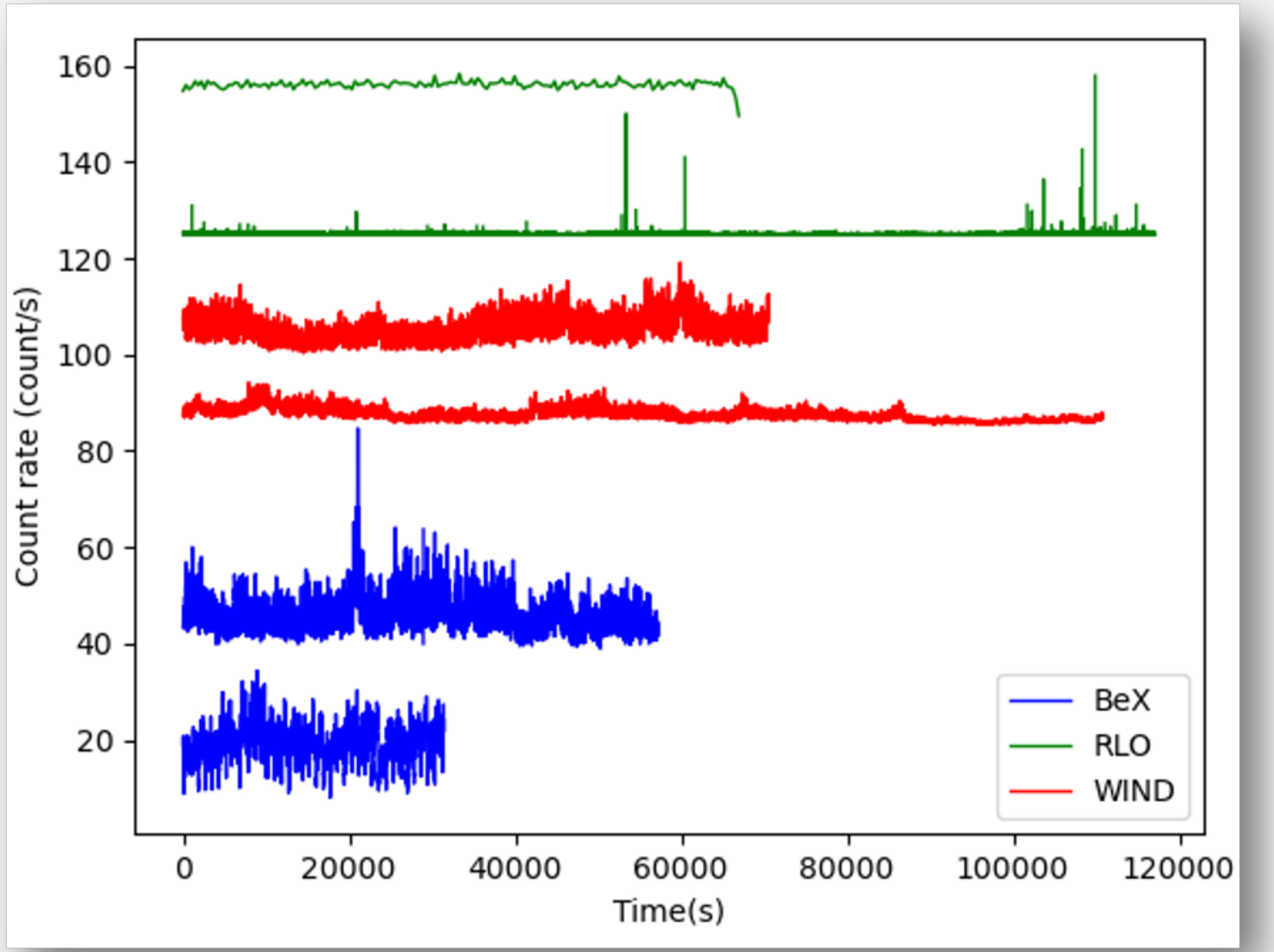
This work, as first approach, is focused on High Mass X-Ray Binaries.

HIGH-MASS X-RAY BINARIES

Variability (periodic, QPO or aperiodic) yields lots of information about the system and the physical processes taking place.



HIGH-MASS X-RAY BINARIES



Examples of light curves (XMM/EPIC) for High-Mass X-Ray Binaries with three types of mass transfer:

- Roche-lobe overflow
- Stellar wind accretion
- Accretion from the circumstellar disk of a Be star

The challenge is to classify HMXBs according to the type of mass transfer.

WHY MACHINE LEARNING?

Automatic classification of X-ray binaries is ideal to be addressed with ML techniques for several reasons:

- Large amount of data: huge datasets are ideal for training ML algorithms, as the more information available, the better the models can learn.
- Data complexity: ML techniques are capable of handling this complexity and learning patterns even in noisy or difficult-to-interpret data.
- Non-linear features: The relationships between the features of X-ray binaries and their classes can be non-linear or difficult to model using traditional analytical methods. ML algorithms, such as neural networks, can learn these non-linear relationships and capture complex patterns in the data.
- Ability to adapt: ML algorithms are flexible and can adapt to different types of data (observation duration, sampling frequency, instrument sensitivity, etc.).
- Feature discovery: In many cases, it can be difficult to determine which specific features of the light curves in the X-ray range are important for classification.

PREVIOUS RESULTS IN THE LITERATURE

The application of Machine Learning techniques, still a very recent tool of analysis in astrophysical applications, has proven effective in the classification of X-ray sources, even with a limited number of labelled sources and without additional information from optical and infrared catalogues.

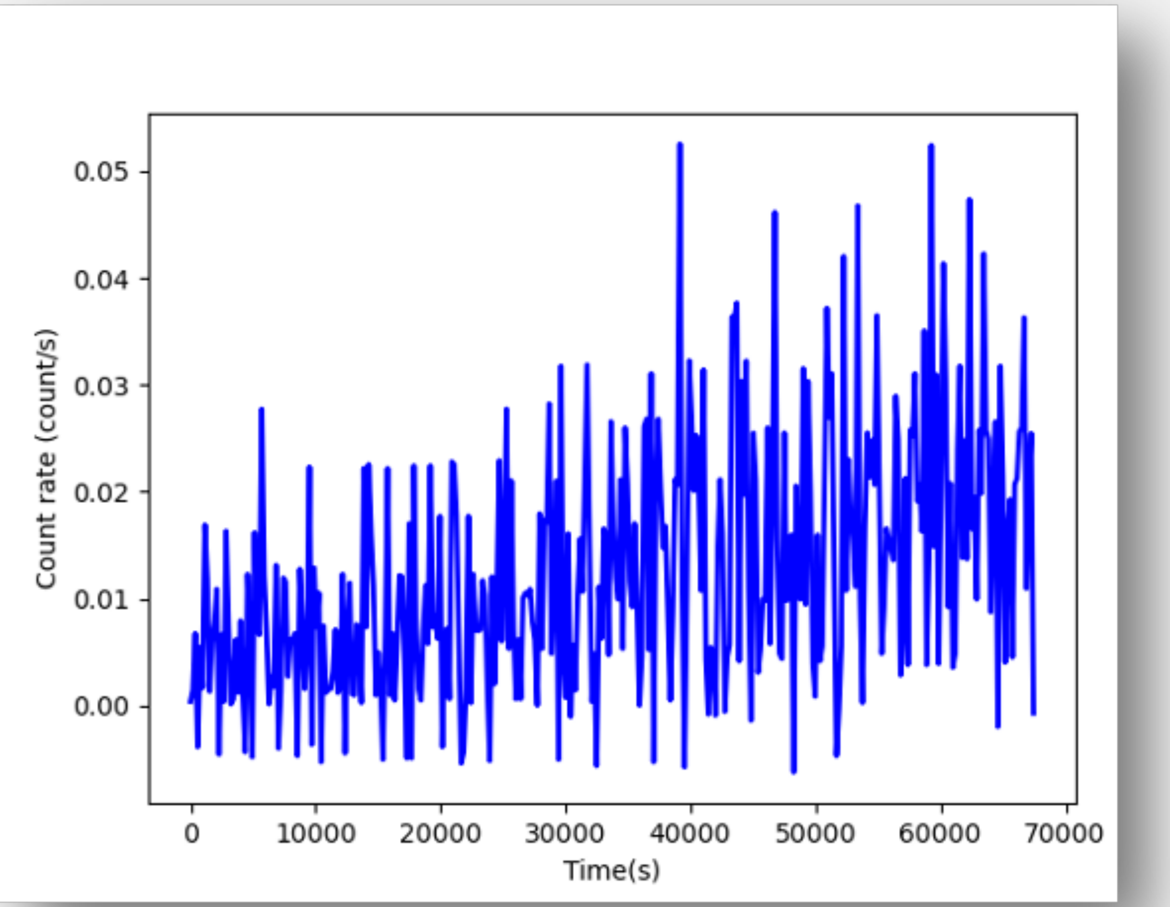
Nevertheless, a general lack of a systematic analysis of variability emerges from previous works, which opens to a large area for further studies and improvements

See, for instance:

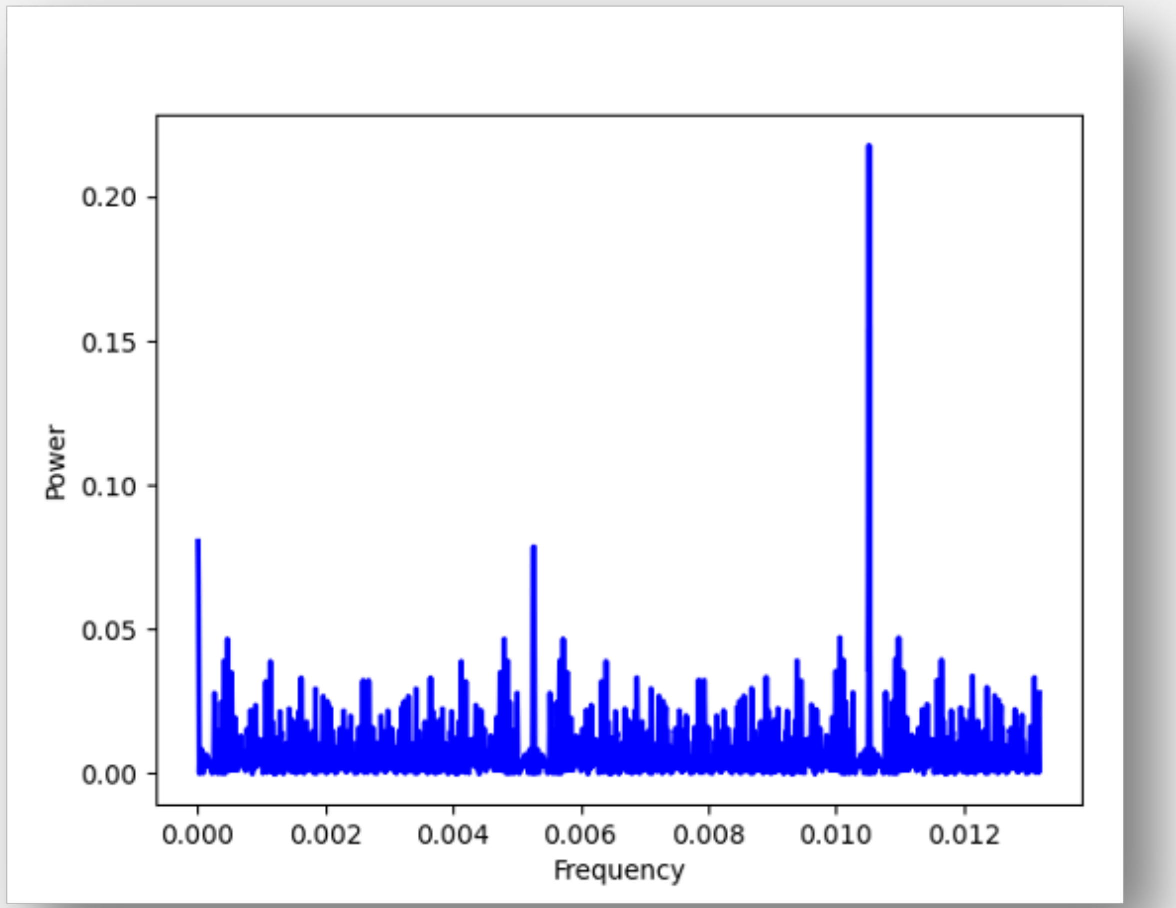
- Pérez-Díaz, V., et al., MNRAS 528, 4852–4871 (2024)
- De Beurs, Z., et al., ApJ 933-116 (2022)
- De Luca, A., et al., A&A 650, A167 (2021)

...

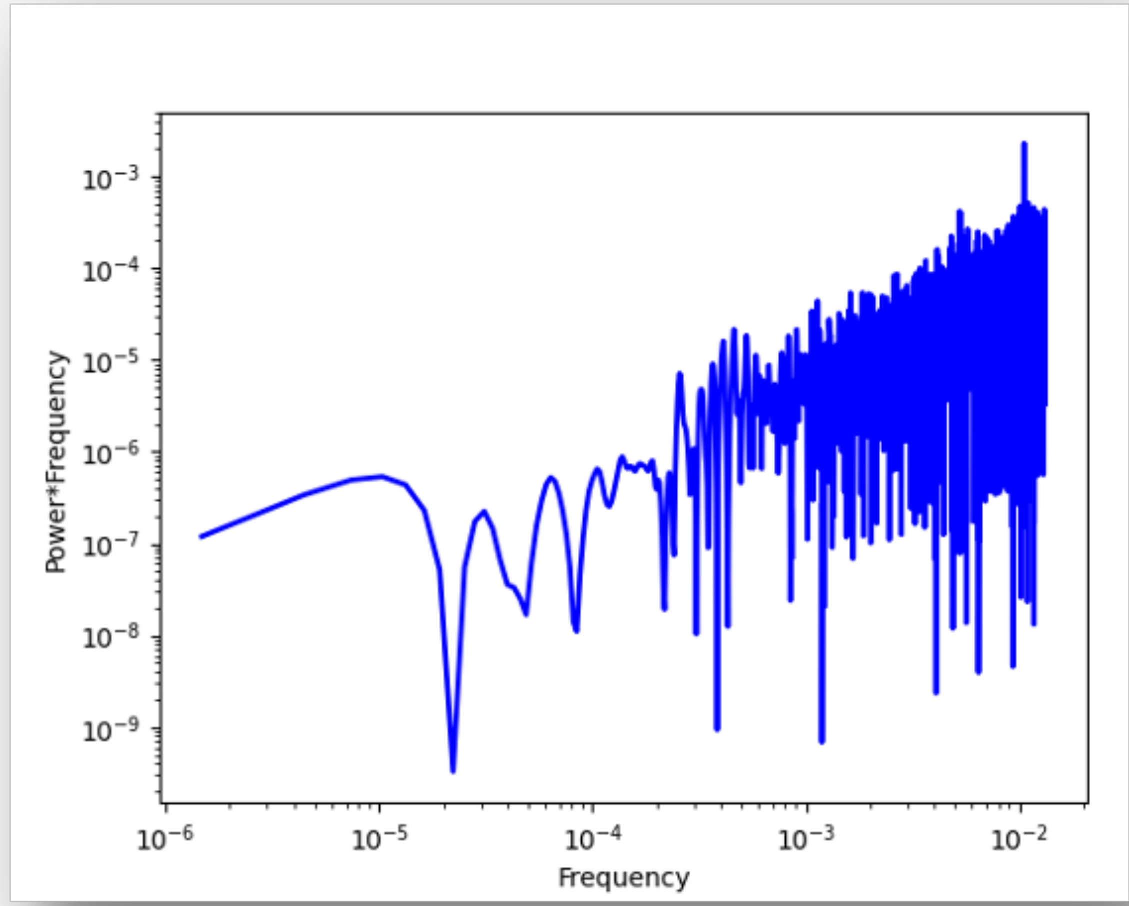
DATASETS



light curves



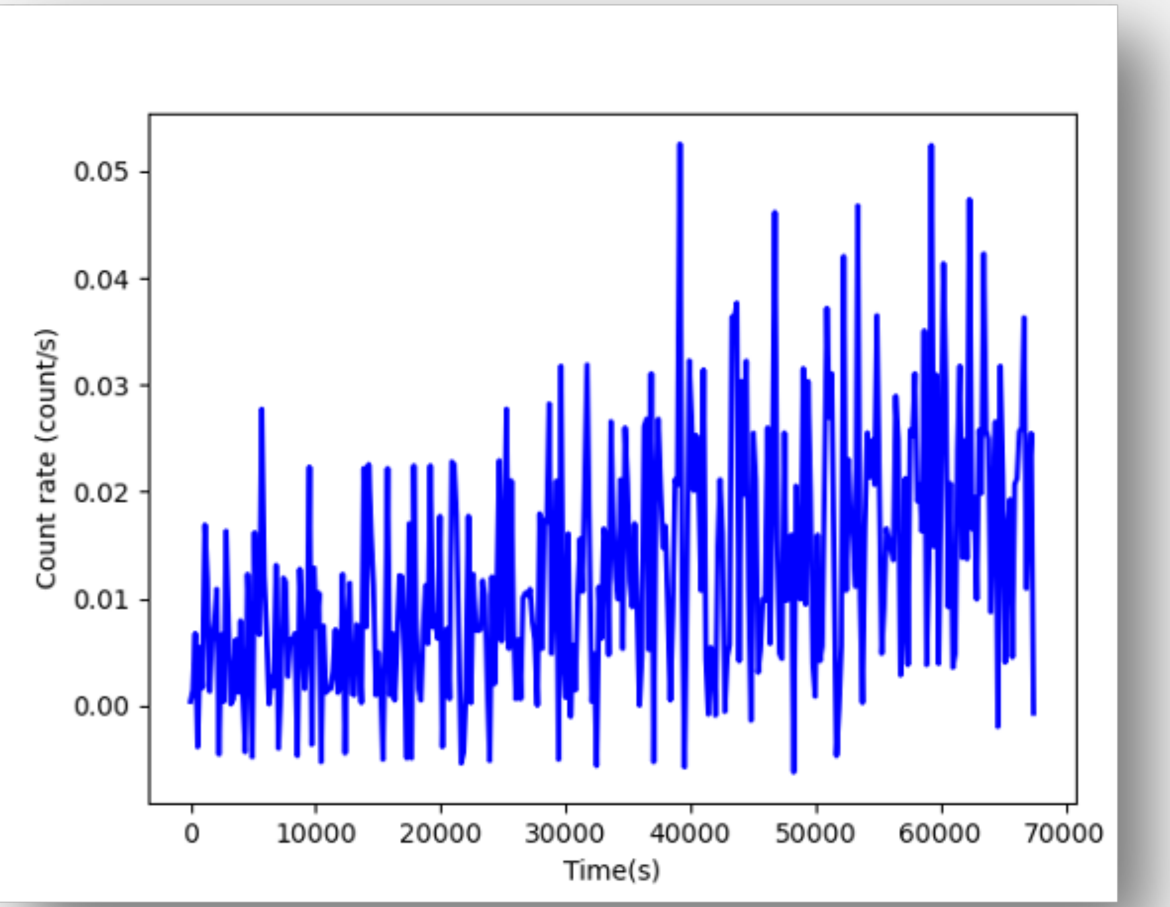
Lomb-Scargle Periodograms



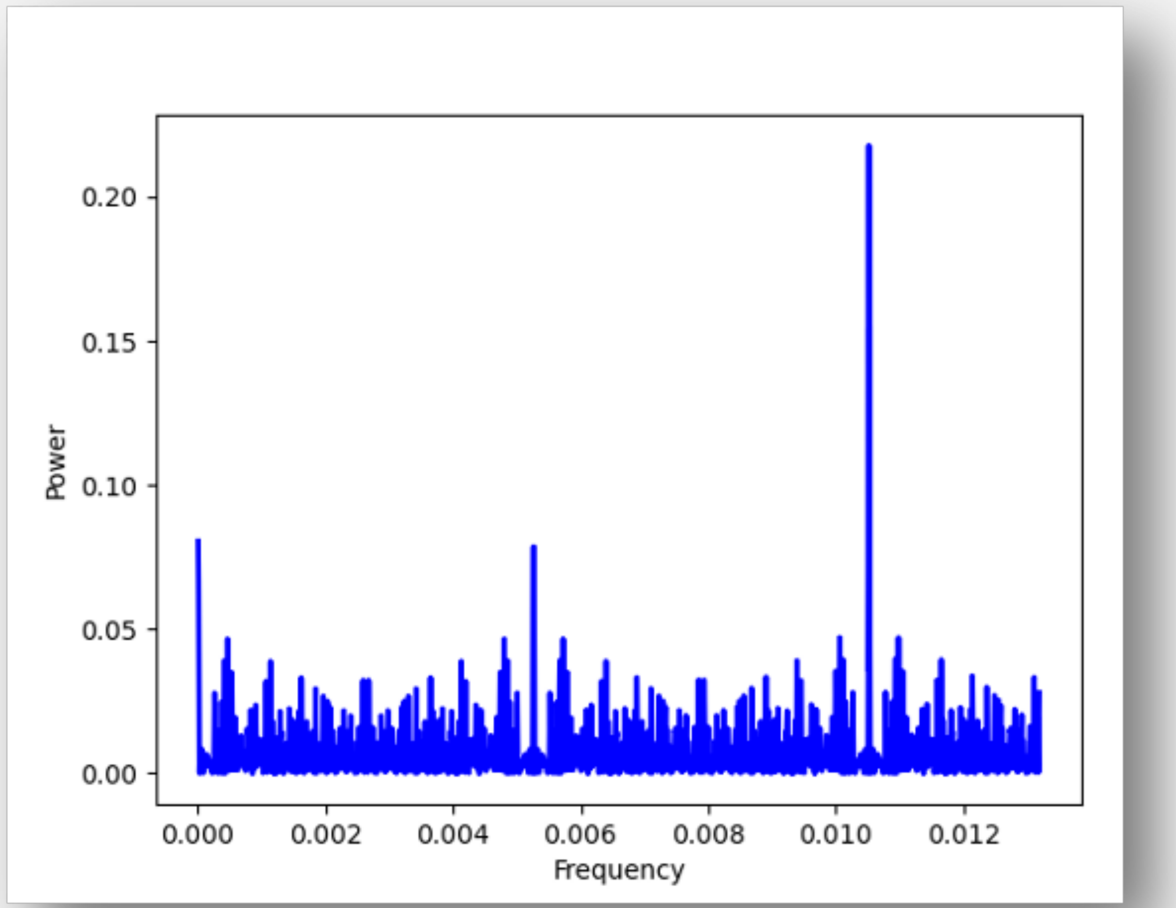
Power spectral density (PSD)

Our final goal is to combine information from all of them to classify the targets, by using different ML approaches.

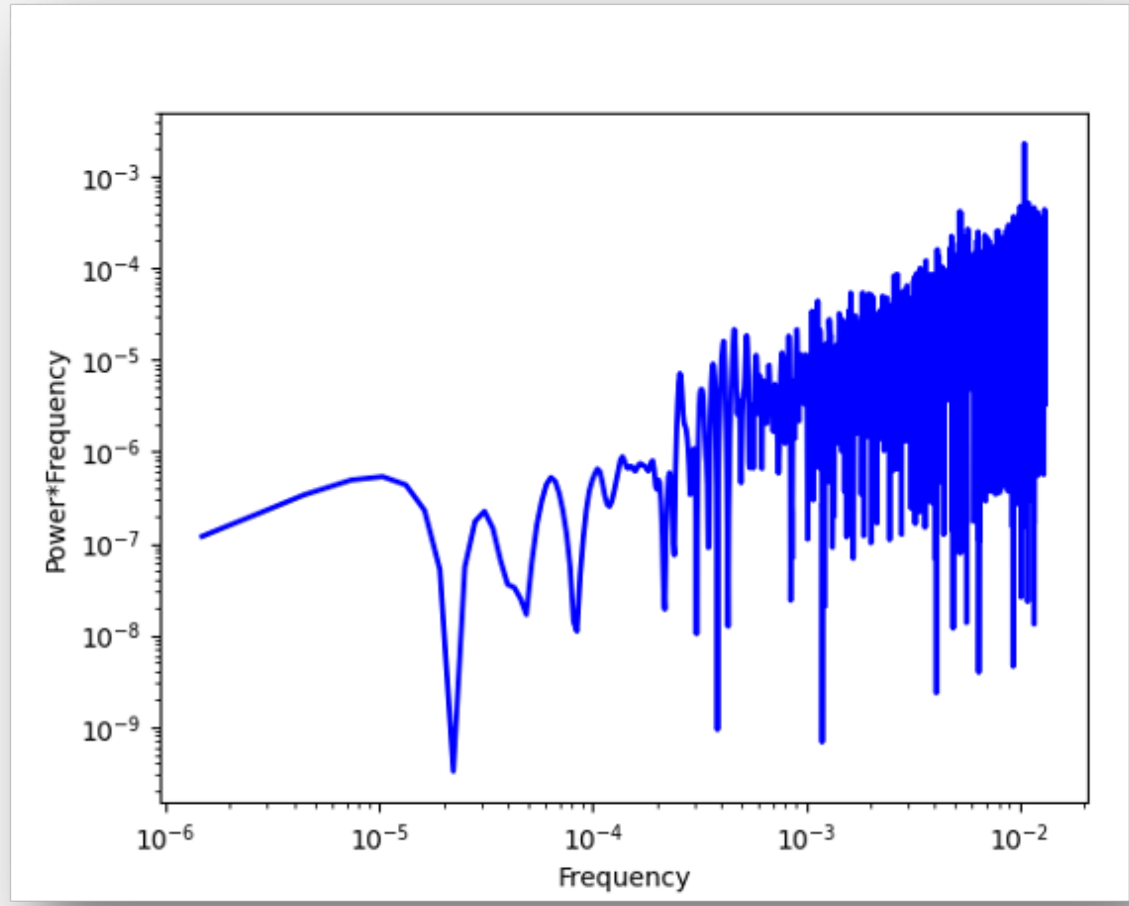
DATASETS



light curves



Lomb-Scargle Periodograms



Power spectral density (PSD)

Our final goal is to combine information from all of them to classify the targets, by using different ML approaches.

To include the largest amount of data, we used a multi-mission approach (similar energy range 2-12 keV): XMM-Newton, MAXI, Swift (in progress).

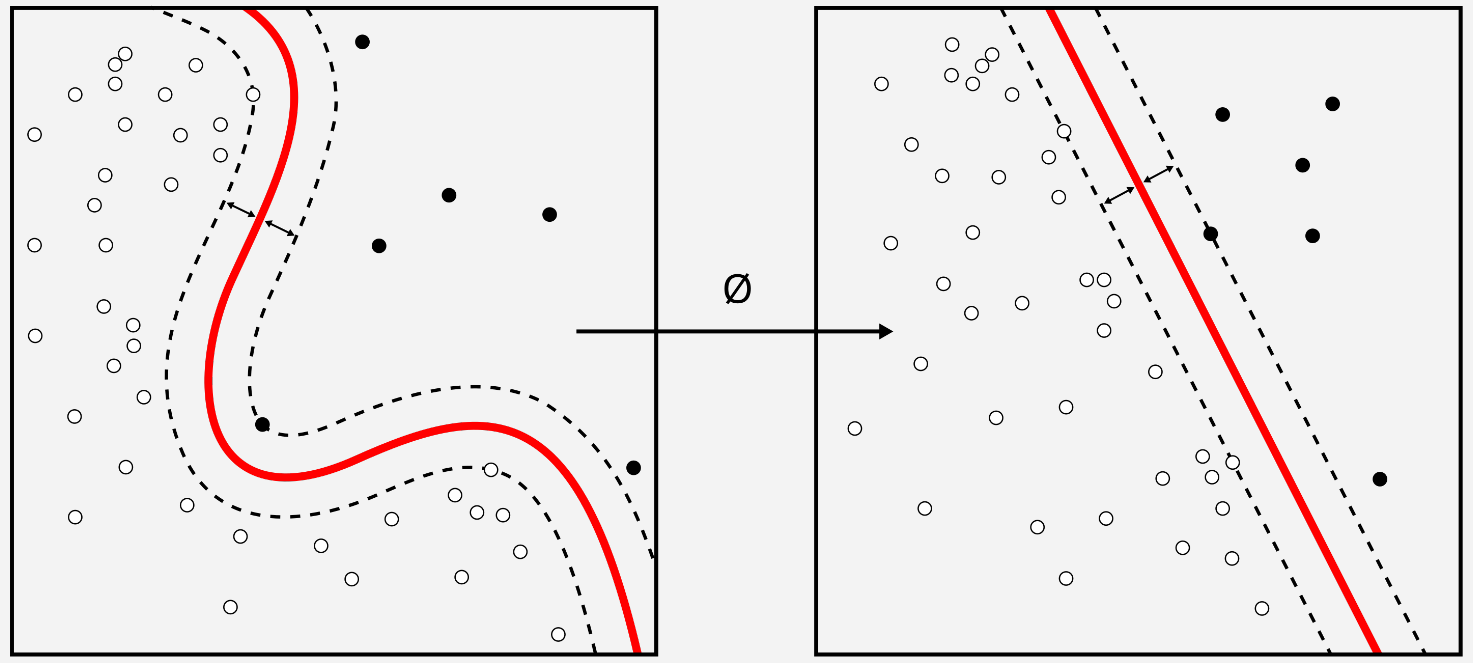
METHODOLOGY

Data are divided into two sets:

- 66% for training
- 33% for testing

Imbalanced learning: number of elements in each class is not equal (15, 93, 129 members for each class for XMM).

First tests with supervised classification methods:



non-linear methods
(KNN, Kernel SVM)

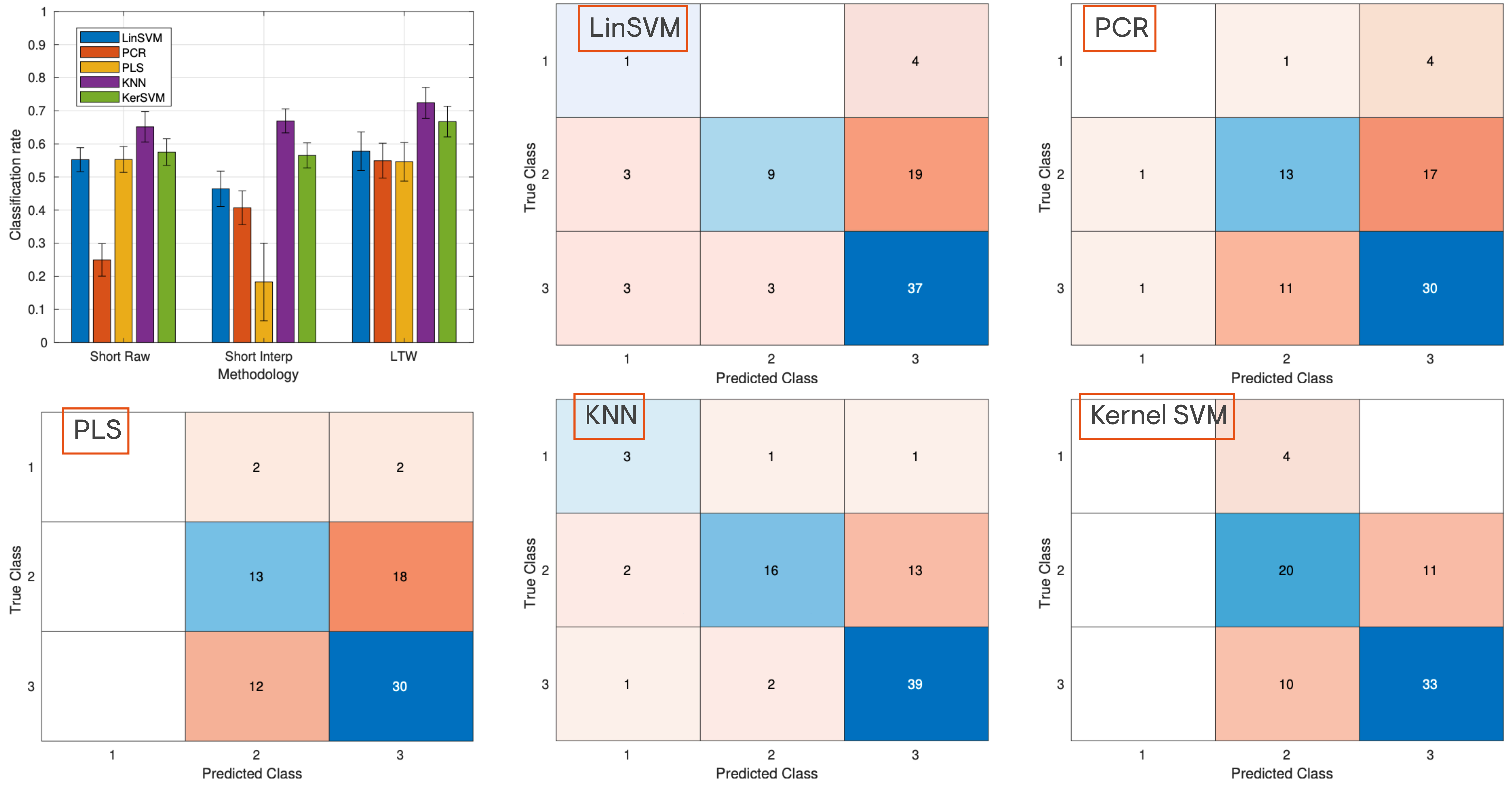
linear methods
(linear SVM, PLS, PCR)

RESULTS

XMM
Light curves

CLASES:

1. Roche-lobe overflow
2. Stellar wind accretion
3. BeX

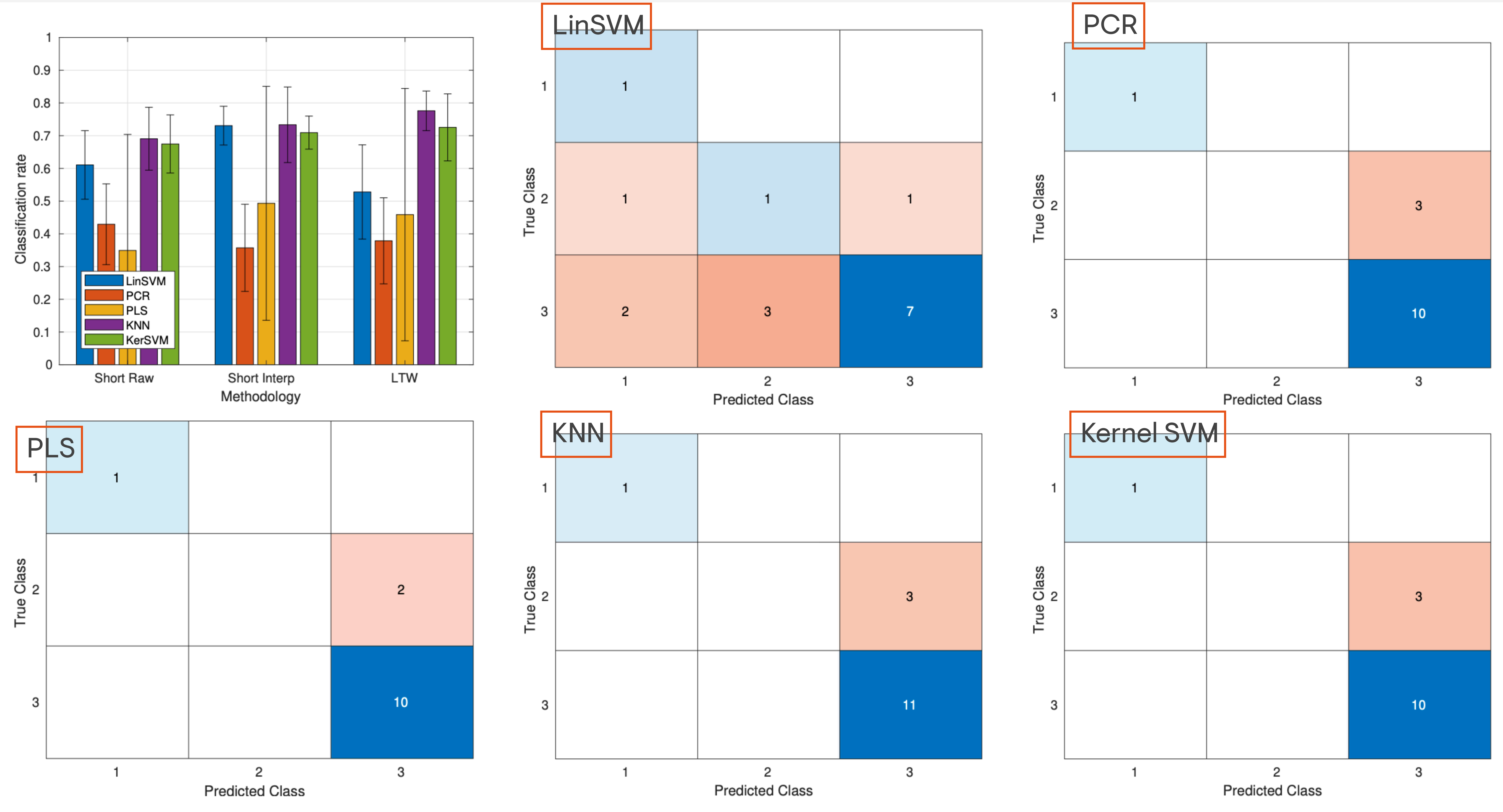


RESULTS

MAXI
Light curves

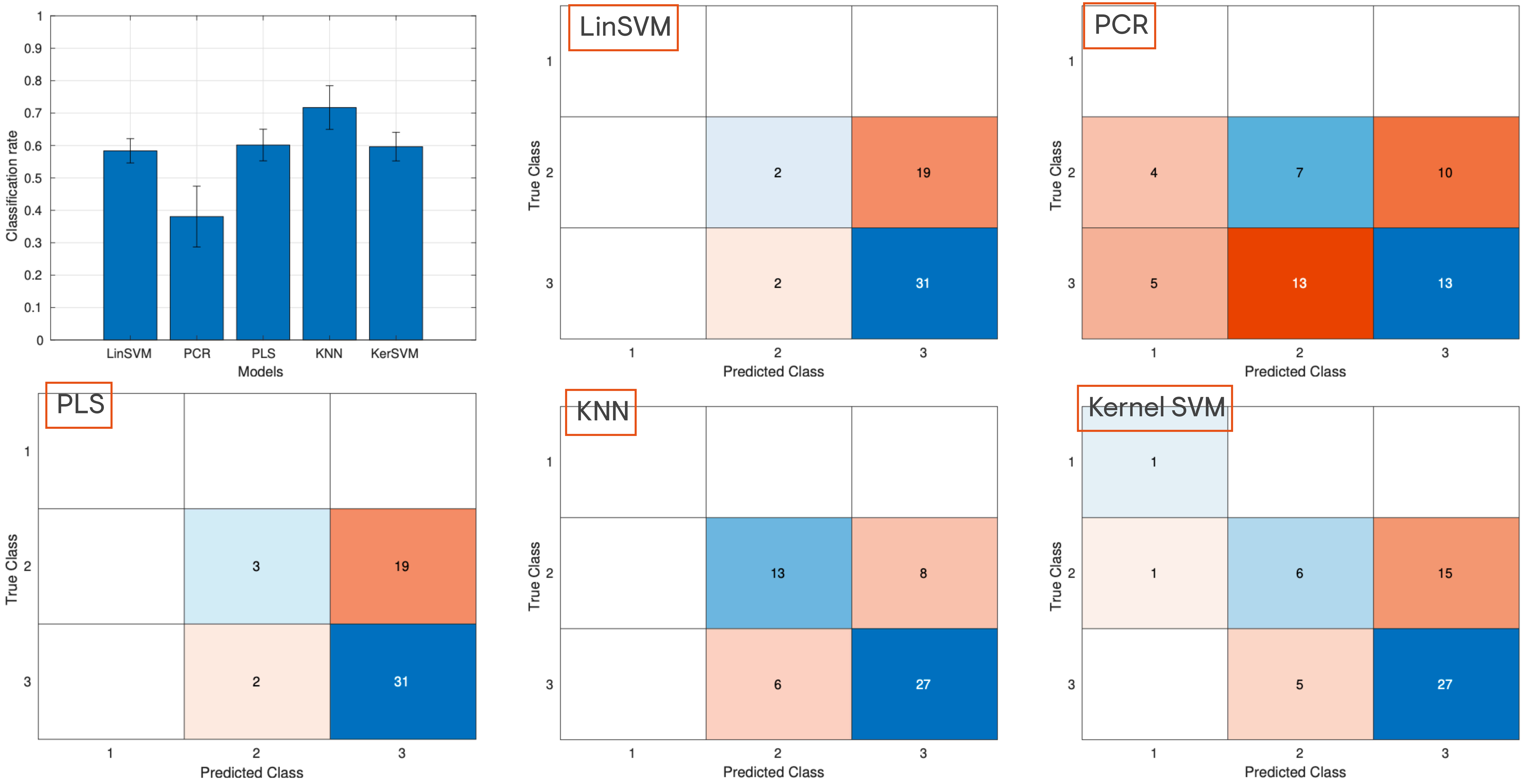
CLASES:

1. Roche-lobe overflow
2. Stellar wind accretion
3. BeX



RESULTS

XMM PSDs



CLASES:

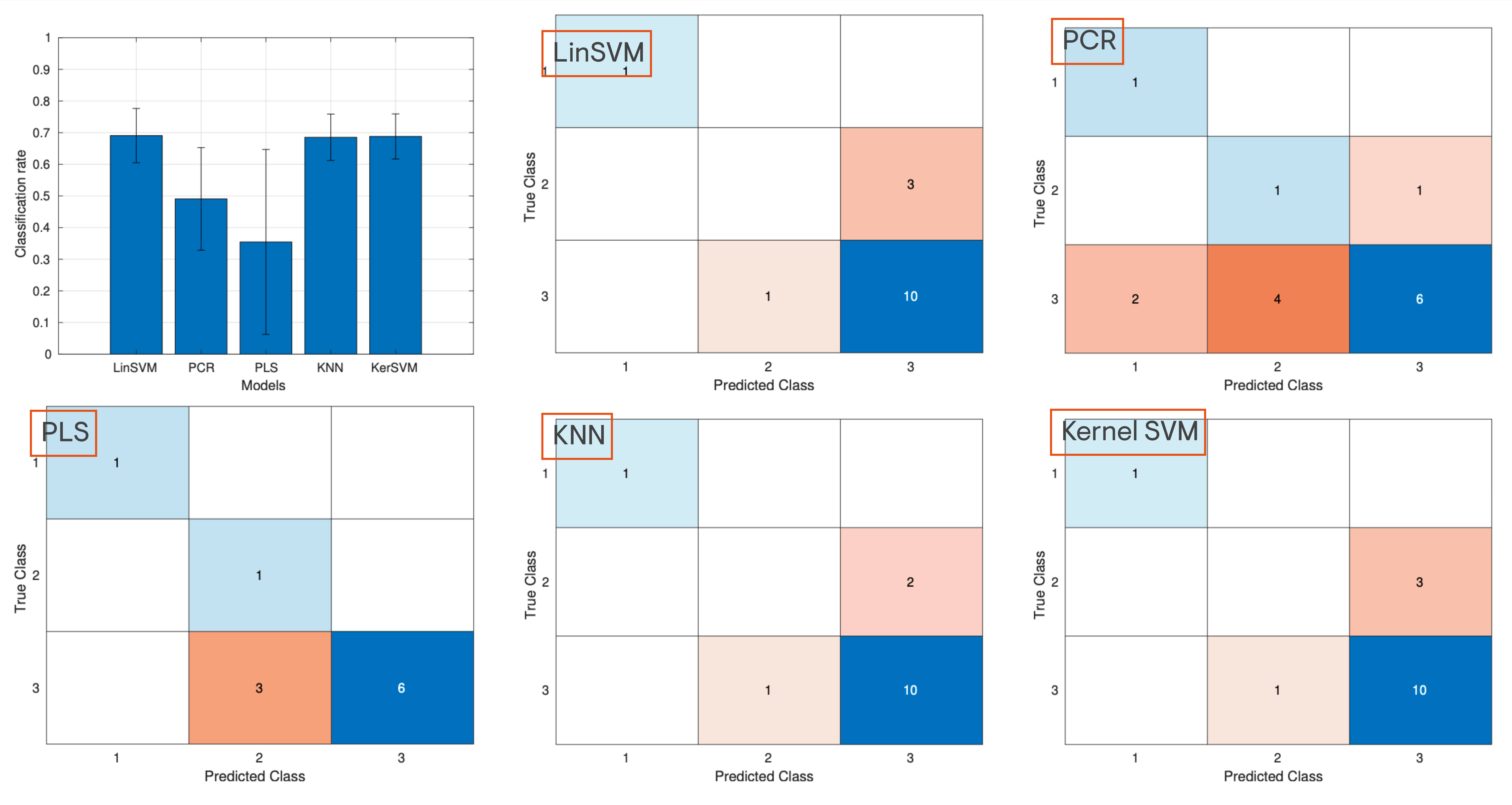
1. Roche-lobe overflow
2. Stellar wind accretion
3. BeX

RESULTS

MAXI
PSDs

CLASES:

1. Roche-lobe overflow
2. Stellar wind accretion
3. BeX



ADDITIONAL WORK

Data

- Including L-S periodograms (ongoing)
- Including data from additional missions: SWIFT (ongoing), RXTE/ASM, INTEGRAL/JEM-X, GINGA, GRO
- Extending the work to all XRBs
- Employing simulated THESEUS data

Models

- Improving current models in order to maximize the accuracy
- Applying neural network models
- Exploring the feasibility of real-time identification of targets as soon as they are observed by THESEUS.

CONCLUSIONS

- Within the THESEUS science, X-ray binaries are a key topic.
- The use of ML techniques provides advantages to identify and classify Galactic X-ray binaries. These techniques are well suited to problems where large amounts of data have to be analysed.
- This preliminary work is focused on High Mass X-Ray Binaries.
- Five ML techniques have been applied to light curves and PSDs from XMM-Newton and MAXI .
- Preliminary results are promising:
  - ☑ light curves allow for a more accurate classification than PSDs
  - ☑ KNN provides the best results… to be confirmed…