<u>Aim:</u> To categorize the states of Black Hole X-ray Binaries within synthetic data employing the unsupervised DBSCAN algorithm

INTRODUCTION:

A binary star system is a two star system which are gravitationally bound to each other. An X-ray binary systems are those binary systems which mainly emit X-rays. In an X-ray binary system one star is a ‘compact star’ (i.e. either a Black Hole or a Neutron Star or a White Dwarf) and another star called ‘companion of the compact star’ is a normal star (either like a bright star which we see in the sky everyday with naked eyes or like a faint star which can be seen with a telescope).

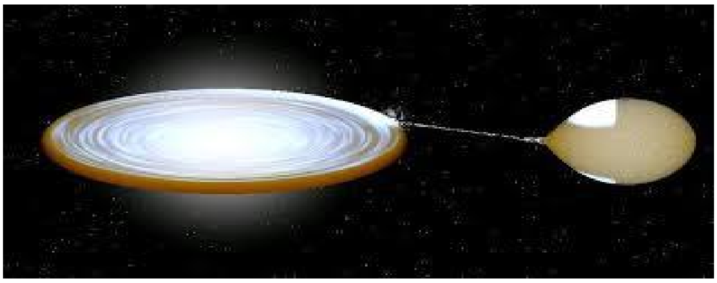


Figure 1: A cartoon view of a X-ray binary system

So, in a Black Hole X-ray Binary (BHXRB) system one star is a Black Hole and another star is a normal star. The Black Hole, because of it’s intense gravity accretes (pulls) matter from its companion star, and as the matter falls into the gravitational well of the Black Hole, it emits X-rays. The accretion process in BHXBs can exhibit different states based on the observed X-ray emission characteristics. The two main states are a) hard state (also called low/hard state) b) soft state (also known as high/soft state). In hard state the binary system emits more high energy X-ray photons compared to low energy photons. Photons below 4 keV are usually considered low (soft) energy photons and photons with energies more than 4 keV are the high (hard) energy photons. The characteristics of these two states are given in the following Table.

Characteristics	hard state (HS)	soft state (SS)
X-ray Emission	X-ray spectrum is characterized by a hard powerlaw component. The X-rays produced in this state are of higher energy.	X-ray spectrum is dominated by a soft (thermal) component. The X-rays produced in this state are of lower energy.
Luminosity	The luminosity of the system is relatively low. For this reoson this state is also called low/hard state.	The luminosity is higher compared to the hard state. The system is brighter in the soft X-ray band. Because of this reoson this state is also known as high/soft state.
Variability	Often show strong variability in their X-ray emission.	The variability is generally less pronounced in the soft state compared to the hard state.

Often BHXRBs transits from hard to soft state and vice versa. This transiting state is known as intermediate state (IMS), which shares characteristics of both the hard and soft states. Further IMS are of two types, hard intermediate state (**HIMS**) and soft intermediate state (**SIMS**).

In **HIMS** X-ray spectrum still contains a significant hard powerlaw component, but slightly less harder than the hard state. In **SIMS**, the X-ray spectrum is softer compared to the hard state but not as soft as the soft state. There is a decrease in the prominence of the hard powerlaw component, the X-ray emission in the soft energy band becomes more pronounced compared to the HIMS. The exact details can vary between different systems, and our understanding of these states continues to evolve with ongoing observations and research.

The study of these state transitions provides valuable insights into the accretion processes around Black Holes in BHXRB systems and helps us understand the complex dynamics of matter falling into the extreme gravity around this exotic object.

METHOD:

Everyday the X-ray detectors onboard space missions receive X-ray photons. The electronic system in the detector detects photon energies, their time of arrival, number of photon counts per second along with their direction in space from which they coming. Following the preprocessing of these data, timing and spectral analyses are conducted. Upon a meticulous examination of the timing and spectral parameters, the state of a BHXRB system is determined for a specific day or duration. However, manually identifying the system's state can be a laborious task, especially when dealing with large datasets.

Hence, with the aim to automate this process of identification of the BH state in a BHXRB system, I have generated synthetic data with four states (HS, HIMS, SIMS, SS). This data includes Modern Julian Day (MJD), Hardness Ratio (HR), powerlaw photon index (P.I.), inner accretion disk temperature (T_{in}), low energy (0.5-4 keV) flux (lowF) and high energy (4-10 keV) flux (highF) as features. Where MJD is the time of detection of the photons, HR is the ratio between 4-10 keV photon flux and 0.5-4 keV photon flux. The flux is obtained in ergs/sec/cm^2. The features are stored in the file **ip_feature_shuffle_1.csv**. The data is unlabelled i.e. no state infomation is given in this file. As the Black Hole transits from HS to SS through HIMS ans SIMS, the HR decreases gradually. P.I., Tin, lowF increases and highF decreases. I have generated the data from MJD 57445 to MJD 58447 with 250 datapoints for each state and hence 1000 datapoints in total. I have first generated the dats in ascending order of MJD from HS to SS through HIMS ans SIMS and then shuffled it keeping each row same as it is created. I am using this shuffled dataset (**ip_feature_shuffle_1.csv**) as the input for classifying the states using **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise) algorithm.

The code is given in **DBSCAN_with_synthetic_data_GB_2.py**. The code creates an output file with the above six features and the labels (in the last column) for the Black Hole's states. I then compared this classification with standard state classification. The details of each step used in the code is given inside the code itself. The results are described below.

RESULTS:

- Silhouette Score for the state classification: **0.63**

- The labels for the Black Hole’s states are given in the last column (‘cluster_label’) of the output file: **op_feature_labels_GB_1.csv**

- The classification accuracy is checked by comparing the label column of a standard file (**ip_feature_label_shuffle_1.csv**) and the above output file

- The comparison gives 100% accuracy

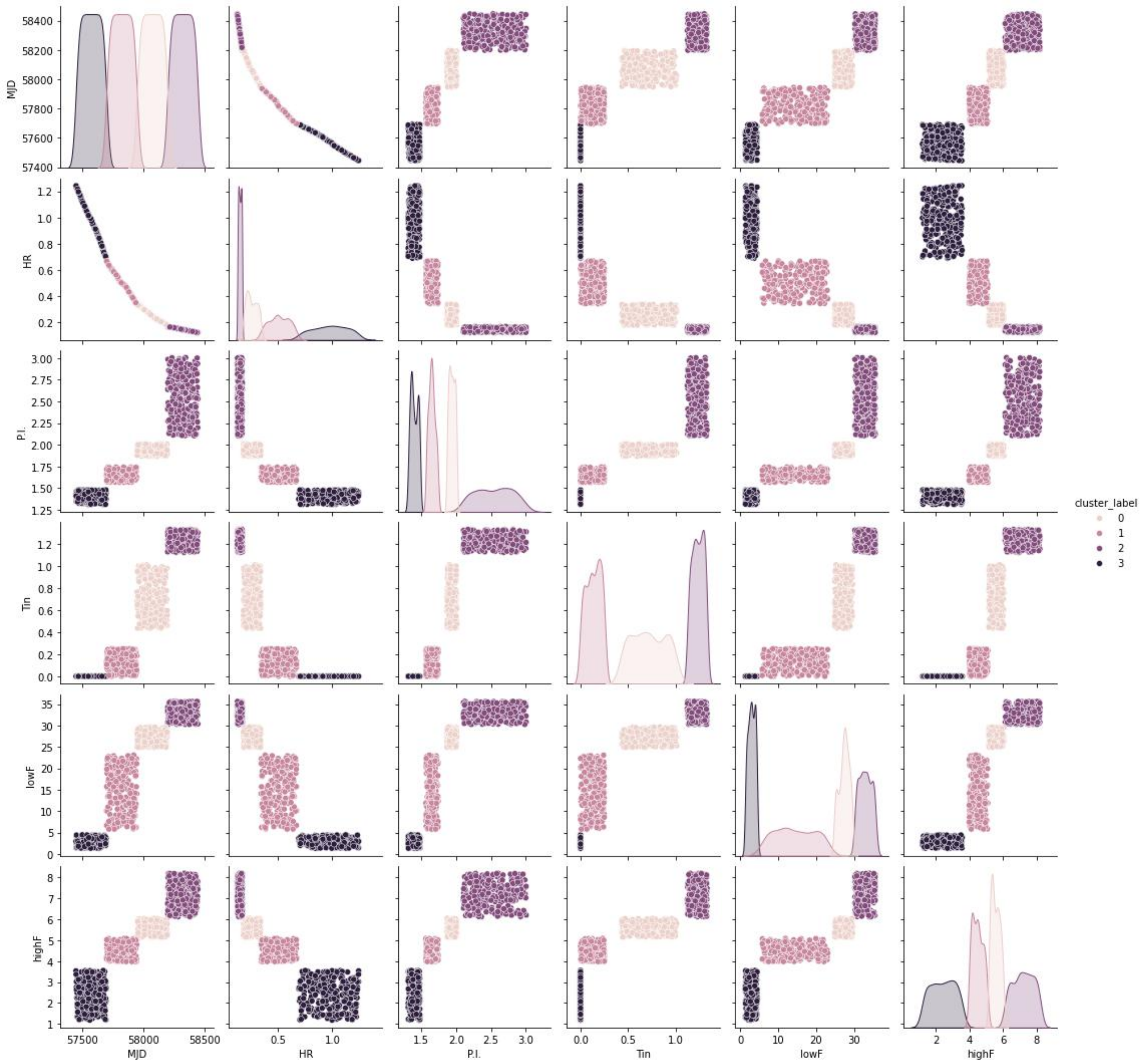
- To evaluate the performance of the model, the metric **Precision, Recall, F1 score and Accuracy** are calculated from the **Confusion Matrix** by comparing the true label (given in the last column of the standard file ip_feature_label_shuffle_1.csv) and the predicted label (given in the last column of op_feature_labels_GB_1.csv) obtained by the code. These are given below.

Confusion Matrix:

```
[[250   0   0   0]
 [  0 250   0   0]
 [  0   0 250   0]
 [  0   0   0 250]]
```

Precision: 1.0
Recall: 1.0
F1 Score: 1.0
Accuracy: 1.0

- The plot for each pair of features for the four Black Hole classes are given below.



MJD
HR
P.I.
Tin
lowF
highF
cluster_label
0
1
2
3