Classification of Transient Astronomical Object Light Curves Using LSTM Neural Networks

Fernandes, G.G.D.[1,3]; Barroca, M.A.[1]; dos Santos, M.[1]; Oliveira, R.S.[2]

[1]Centro Brasileiro de Pesquisas Físicas

[2]Universidade Federal de Minas Gerais

[3]Instituto Superior Técnico

Abstract

This study presents a bidirectional Long Short-Term Memory (LSTM) neural network for classifying transient astronomical object light curves from the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC) dataset. The original fourteen object classes were reorganized into five generalized categories (S-Like, Fast, Long, Periodic, and Non-Periodic) to address class imbalance. After preprocessing with padding, temporal rescaling, and flux normalization, a bidirectional LSTM network with masking layers was trained and evaluated on a test set of 19,920 objects. The model achieved strong performance for S-Like and Periodic classes, with ROC area under the curve (AUC) values of 0.95 and 0.99, and Precision-Recall AUC values of 0.98 and 0.89, respectively. However, performance was significantly lower for Fast and Long classes (ROC AUC of 0.68 for Long class), and the model exhibited difficulty distinguishing between Periodic and Non-Periodic objects. Evaluation on partial light curve data (5, 10, and 20 days from detection) revealed substantial performance degradation, with increased misclassification toward the S-Like class. These findings indicate that class imbalance and limited temporal information are primary limitations, suggesting that class balancing strategies and preprocessing techniques focusing on detection moments could improve performance.

1   Introduction

Artificial intelligence (AI) has become increasingly important in scientific research, particularly for analyzing large-scale datasets that exceed human processing capabilities. In astronomy, the exponential growth of observational data from modern telescopes requires automated classification systems to identify and categorize celestial objects efficiently.[1] Machine learning approaches, particularly deep neural networks, have shown great promise in handling such classification tasks.

Recently, the Large Synoptic Survey Telescope (LSST) proposed a challenge on Kaggle, a Google subsidiary website,

called The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC).[2] This challenge is aimed at classifying light curves, data of the observed brightness of celestial objects as a function of time, which were simulated in preparation for LSST observations. These curves are capable of revealing the presence of some phenomena, such as supernovae[3]. The LSST will revolutionize our understanding of the sky, however, this type of study is hindered by the large volume of data obtained by the telescope, making automatic analysis procedures indispensable to differentiate and classify them. In this challenge, the question was raised: How well can we classify objects in the sky that vary in brightness from simulated LSST time-series data? The PLAsTiCC dataset and challenge were created to help classify these astronomical sources. [4]

Before the use of AI, this type of classification still depended on manual analysis. Common statistical methods such as template fitting[5] were used, but these are not scalable for a large volume of data. Usually, a group of experts manually eliminated obvious cases of false positives, something that by itself can take several days. Of the remaining data, each case should be reviewed by at least three experts, which can lead to disagreements about a particular case, since experts might not have the same definition for classification. For these reasons, a reliable system is necessary that repeatedly selects the most important candidates that will be manually reviewed for confirmation at a later stage.

2 Theoretical Framework

2.1 Long Short-Term Memory Neural Networks

Since light curves are a function that varies with time, one possible way to solve the PLAsTiCC problem is to use a Long Short-Term Memory (LSTM) Neural Network. The LSTM is a recurrent neural network (RNN) architecture, developed by Hochreiter and Schmidhuber in 1997. Recurrent networks have the characteristic of making some information persist in the network through a loop at arbitrary intervals. The RNN by definition has a simpler repetition structure than LSTMs, as there is only a single tanh layer, as shown in figure 1, while LSTMs have a chain structure that contains four neural network layers and different memory blocks called cells as shown in figure 2. Thus, LSTMs are well suited for classifying, processing, and predicting time series with time intervals of unknown duration. [6]

3 Materials and Methods

The light curve data were obtained through LSST PLAsTiCC via the Kaggle challenge[2]. To complete the project within the scheduled time of EAFEXP, a subset of the total data was used. In total, two datasets were used: the first had 7848 objects, of which 10% were used to form a validation set and the remainder formed the training set; the second dataset had 19920 objects and was used as the test set.

The data are in tabular format and consist of flux measurement information, measurement error, the measurement time
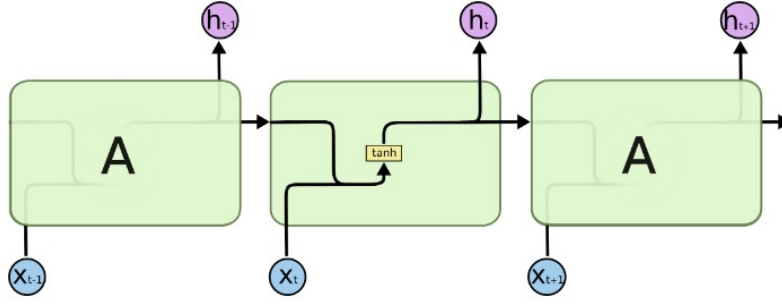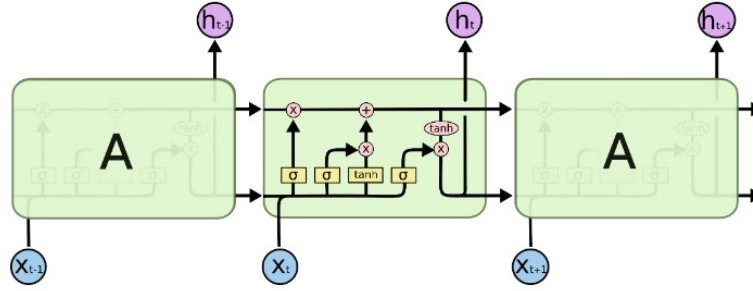
Figure 1: Functioning of an RNN network. [7]

Figure 2: Functioning of an LSTM network. [7]

in Modified Julian Date (mjd), the filter used (the object's emission spectrum, ugriZY, represented by a numerical value from zero to five), a boolean variable (0 or 1) that indicates detection or non-detection, the object class, and a unique identifier. An example can be seen in Table 1.

| Flux	Error	mjd	Filter	Detection	Class	Id
-544.810	3.623	59750.4	2	1	92	615
-816.434	5.553	59750.4	1	1	88	713
-471.386	3.802	59750.4	3	1	42	730
...	...	...	...	...	...	...
-388.985	11.395	59750.4	4	1	90	745
-2.940	1.771	59798.3	3	0	90	116720
-12.810	5.380	59798.4	5	0	92	117016
...	...	...	...	...	...	...

Table 1: Table exemplifying the dataset.

3.1 Preprocessing

Initially, the provided data had fourteen categories, however, due to the discrepancy in the amount of data per category,

the data were reorganized into five categories according to Table 2. Examples of each of the classes can be seen in figure 8. The distribution by original classes and generalized classes can be seen in figure 3.

Id	Original Class	Generalized Class	New Id
6	Single micro-lens	Fast	1
15	TDE	Long	2
16	Eclipsing Binary	Periodic	3
42	SNII	S-Like	0
52	SNIax	S-Like	0
53	Mira	Periodic	3
62	SNIbc	S-Like	0
64	Kilonova	Fast	1
65	M-dwarf	Fast	1
67	SNIa-91bg	S-Like	0
88	AGN	Non-Periodic	4
90	SNIa	S-Like	0
92	RR Lyrae	Periodic	3
95	SLSN-I	Long	2

Table 2: Table with the fourteen original classes and their generalized classes.
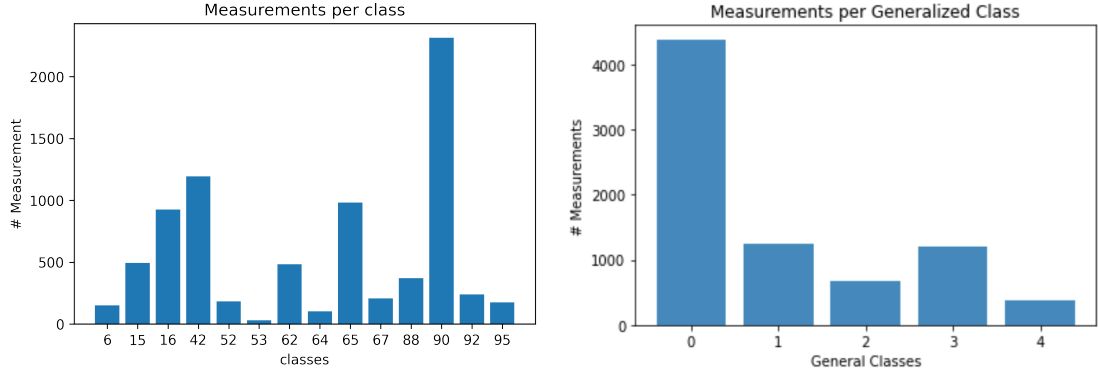
(a) Distribution of measurements by original classes.

(b) Distribution of measurements by generalized classes.

Figure 3

To train the network, it is necessary to ensure that each object has the same number of measurements, since this will be the input dimension. This is not true for this dataset, so the data with smaller dimensions were filled with arbitrary values in a process known as padding. As shown in figure 4, after padding, all objects have the same dimension.

In addition, the temporal measurements were rescaled so that the first measurement of each object was set to zero on the temporal scale. Finally, the flux data were normalized for each object separately using min-max normalization, scaling values to the range [0,1], since the amplitude of measurements varies significantly between objects.

Finally, this project is interested in verifying the ability to classify objects with only partial light curve data. For this purpose, two new test sets were selected, where the data were advanced in time by ten and twenty days of observation from the moment each object is detected, and subsequently the network was evaluated with sets of five, ten, and twenty days into the future.

3.2   Neural Network

Since the provided data are time series, an LSTM network was chosen for its ability to capture temporal dependencies. The network is bidirectional to capture both forward and backward temporal patterns, as both previous and subsequent measurements are relevant for classification. Each column in the dataset (Table 1) was treated as a feature, resulting in five input features: flux, error, modified Julian date, filter, and detection.

Due to the padding procedure, all objects have 352 measurements and 5 features. The network input dimensions are therefore 352×5, and the output is a probability distribution over the five generalized classes.

A masking layer was implemented to identify and ignore padded values, preventing them from being treated as noise. A GlobalMaxPooling layer was used to reduce the temporal dimensions by extracting the maximum value across the sequence, which helps capture the most significant features regardless of their temporal position. Finally, a dense layer with softmax activation performs the final classification. The complete network archi-
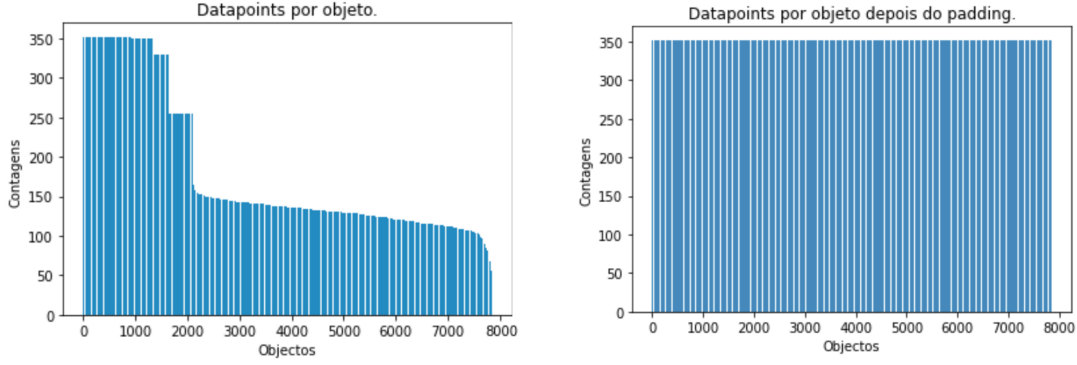
(a) Distribution of the number of measurements per object.

(b) Distribution of the number of measurements per object after padding.

Figure 4

tecture can be visualized in figure 5.

The network was trained using the Adam optimizer with categorical cross-entropy as the loss function. Training was performed with early stopping based on validation loss to prevent overfitting. The model was trained for a maximum of 50 epochs with a batch size of 32.

All code used and the trained model to reproduce these results are available in a public repository on GitHub at the link `https://github.com/MarcoBarroca/VI_EAFEXP_Proj3`.

4   Results and Discussion

4.1   Performance on Complete Light Curves

The network was evaluated using ROC curves, Precision-Recall curves, and confusion matrices, all constructed from the test set. The ROC curves showed excellent performance for the S-Like and Periodic classes, achieving area under the curve (AUC) values of 0.95 and 0.99, respectively. However, the Long class achieved a significantly lower AUC of 0.68, indicating poor performance for this category.

Similarly, the Precision-Recall curves for the S-Like and Periodic classes achieved high AUC values of 0.98 and 0.89, respectively. The Non-Periodic class showed the worst performance with an AUC of 0.40. All curves can be visualized in figure 6.

Analysis of the confusion matrix (figure 7 (a)) reveals several important patterns. The classification of S-Like curves achieves high accuracy, and the model successfully distinguishes Periodic and Non-Periodic classes from the other three categories. However, the model struggles significantly with Long and Fast type curves, showing poor classification performance. Additionally, the model has difficulty differentiating between Periodic and Non-Periodic objects, often confusing one for the other.

4.2   Performance on Partial Light Curves

To assess the model's ability to classify objects early in their light curve evolution, the network was evaluated on test sets with temporal advances of 5, 10,
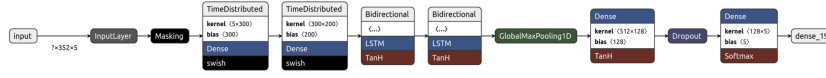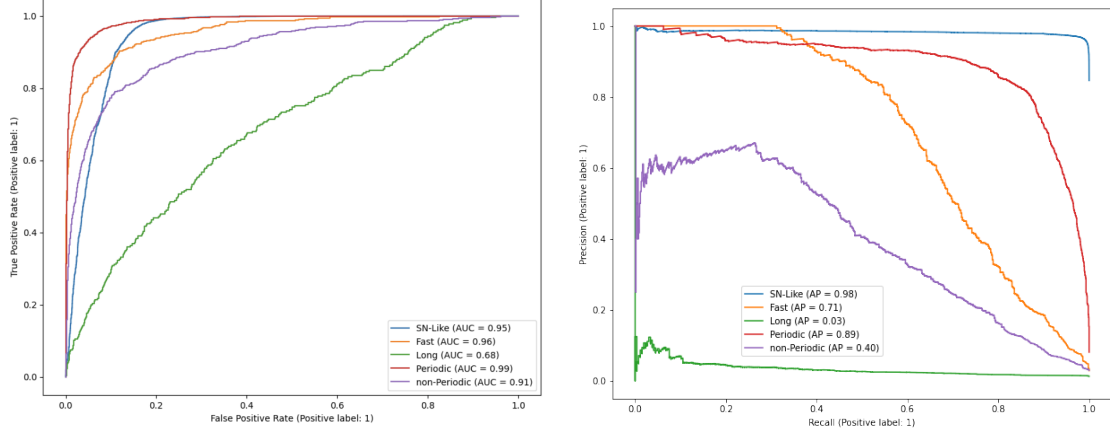
Figure 5: Neural network used in the classification process. The LSTM network is bidirectional because equal importance must be given to past and future data; the Mask layer serves to attenuate the noise effects that padding can cause

(a) ROC curve for our classification model. The Periodic and S-Like classes show satisfactory results, however the same cannot be said for the Fast and Long classes.

(b) Precision-Recall curve for our classification model. The Periodic and S-Like classes show satisfactory results, however we notice that the model does not seem to correctly classify Non-Periodic curves.

Figure 6

and 20 days from the detection moment. The results show a significant degradation in performance as the temporal advance increases. Even with only 5 days of advance, the network fails to achieve performance comparable to the complete dataset and begins to misclassify most objects as S-Like. This degradation becomes more pronounced with longer temporal advances, as shown in the confusion matrices in figure 7 (b), (c), and (d). The ROC and Precision-Recall curves for these partial datasets are provided in the Appendix (figures 9, 10, and 11).

5 Conclusion

This work presents a bidirectional LSTM neural network for classifying transient astronomical object light curves from the PLAsTiCC dataset. The model achieved strong performance for S-Like and Periodic classes, with ROC AUC values of 0.95 and 0.99, respectively, demonstrating the effectiveness of LSTM networks for time-series classification in astronomy.

However, several limitations were identified. The model struggled significantly with Fast and Long classes, likely due to their characteristic short-duration peaks that occur only at the beginning or end of measurements, making them difficult to capture with the current ap-
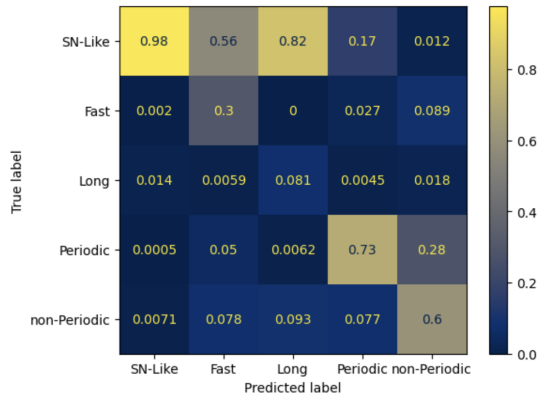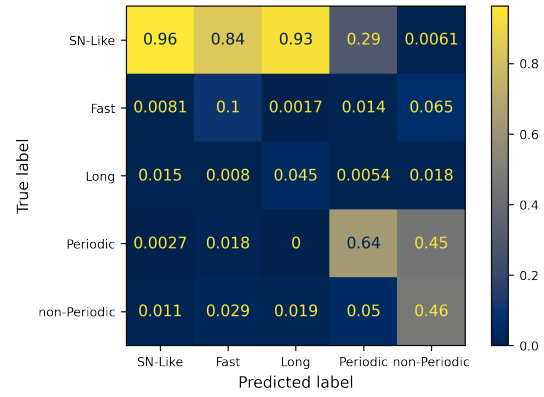
(a) Confusion matrix for the complete test set.
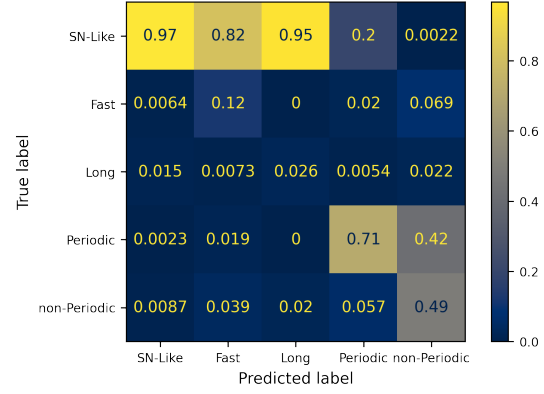
(b) Confusion matrix for the test set with five days ahead. The results worsen compared to the complete test set.
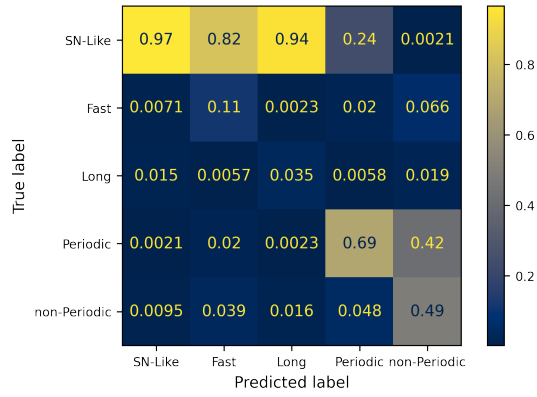
(c) Confusion matrix for the test set with data ten days ahead. The results worsen compared to the complete set and the set with five days.

(d) Confusion matrix for the test set with data twenty days ahead. The network now confuses more classes with S-Like type objects.

Figure 7

proach. Additionally, the model had difficulty distinguishing between Periodic and Non-Periodic objects, often misclassifying Non-Periodic as Periodic, though the reverse error was rare. This suggests that both classes exhibit similar peak patterns, leading the model to incorrectly attribute periodicity to non-periodic signals.

The model's performance degraded substantially when evaluated on partial light curve data, with even 5-day advances causing significant misclassification, primarily toward the S-Like class.

This degradation is likely due to the class imbalance in the dataset, where S-Like objects dominate the training distribution.

Several approaches could address these limitations. First, class balancing through oversampling underrepresented classes or using more sophisticated class weighting schemes could improve performance on minority classes. Second, preprocessing strategies that focus on detection moments (using only data where the detection variable equals 1) could help capture the characteristic features of Fast

and Long objects while reducing background noise. Third, alternative architectures such as attention mechanisms or transformer networks might better capture long-range dependencies and distinguish periodic from non-periodic patterns. During this work, class weighting was attempted but did not yield significant improvements, suggesting that more substantial architectural or preprocessing changes may be necessary.

6   Acknowledgements

References

[1] DJORGOVSKI, S. G. et al. **The Large Synoptic Survey Telescope: From Science Drivers to Reference Design**. arXiv preprint arXiv:0805.2366, 2008.

[2] KAGGLE. **The Photometric LSST Astronomical Time-series Classification Challenge**. Available at: `https://www.kaggle.com/c/PLAsTiCC-2018`

[3] RUSSELL, H. N. **On the determination of the orbital elements of eclipsing variable stars**. Astrophysical Journal, v. 35, p. 315-340, 1912.

[4] THE PLAsTiCC TEAM. **The Photometric LSST Astronomical Time-series Classification Challenge**. arXiv preprint arXiv:1810.00001, 2018.

[5] TANVIR, N. R. et al. **A gamma-ray burst at a redshift of z=8.2**. Nature, v. 461, n. 7268, p. 1254-1257, 2009.

[6] HOCHREITER, S.; SCHMIDHUBER, J. **Long short-term memory**. Neural computation, v. 9, n. 8, p. 1735-1780, 1997.

[7] COLAH. **Understanding LSTM Networks**. Available at: `https://colah.github.io/posts/2015-08-Understanding-LSTMs/`

A   Appendix

If it is of interest to the reader, the ROC and Precision-Recall curves are shown in figures 9, 10, and 11 for the test sets of 5, 10, and 20 days.
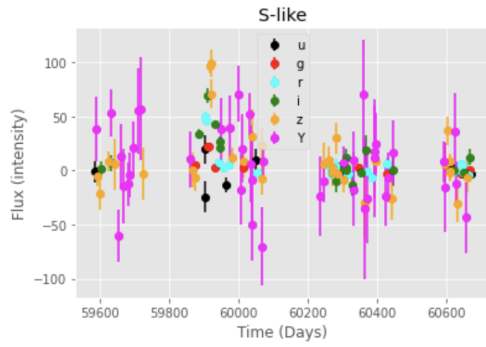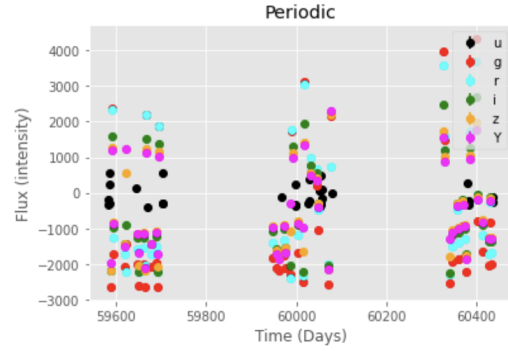
(a) Example of a light curve of an object from the S-Like class. The values in the legend indicate the filter used and the emission spectrum.
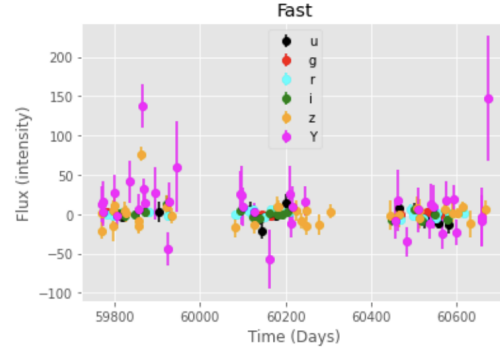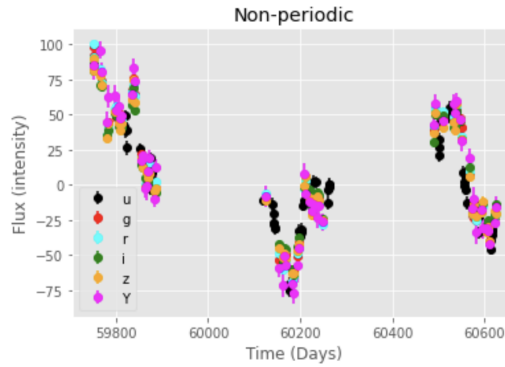
(b) Example of a light curve of an object from the Fast class. The values in the legend indicate the filter used and the emission spectrum.
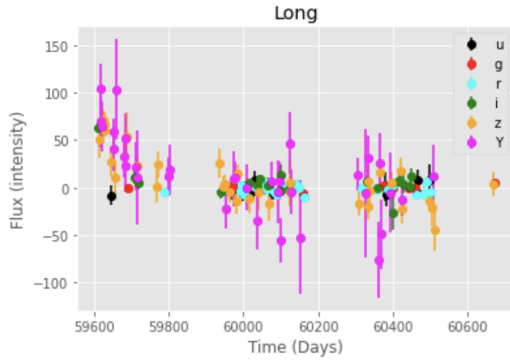
(c) Example of a light curve of an object from the Long class. The values in the legend indicate the filter used and the emission spectrum.
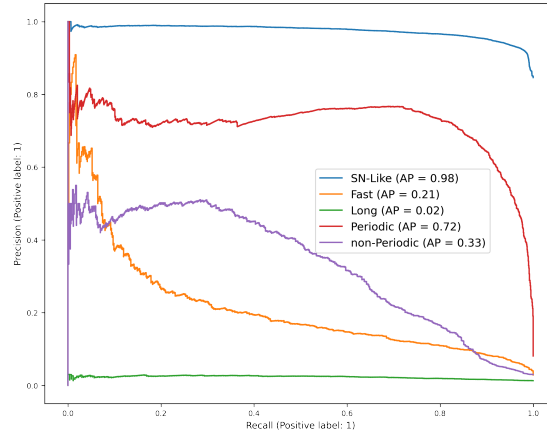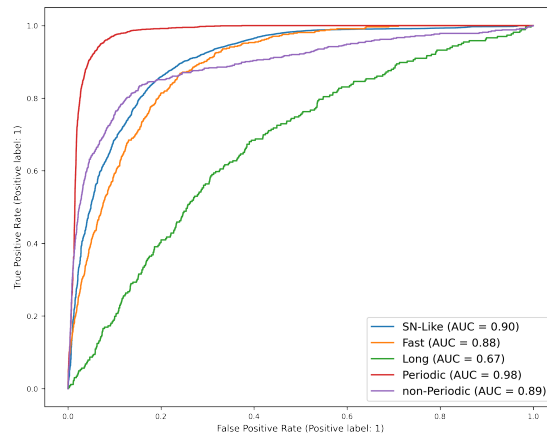
(d) Example of a light curve of an object from the Periodic class. The values in the legend indicate the filter used and the emission spectrum.

(e) Example of a light curve of an object from the Non-Periodic class. The values in the legend indicate the filter used and the emission spectrum.
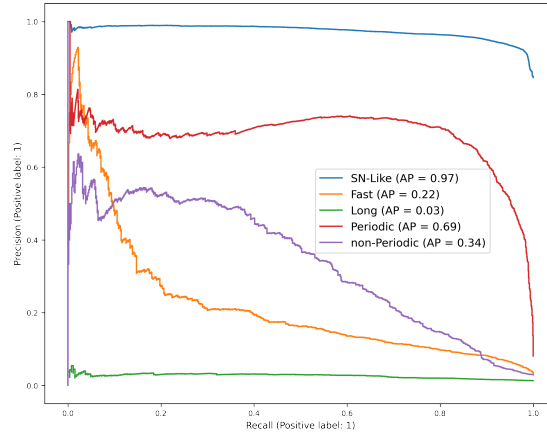
Figure 8

(a) Precision-Recall curve for the dataset with a temporal advance of 5 days.
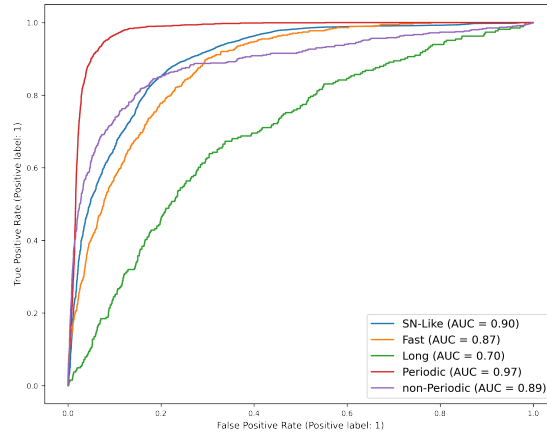

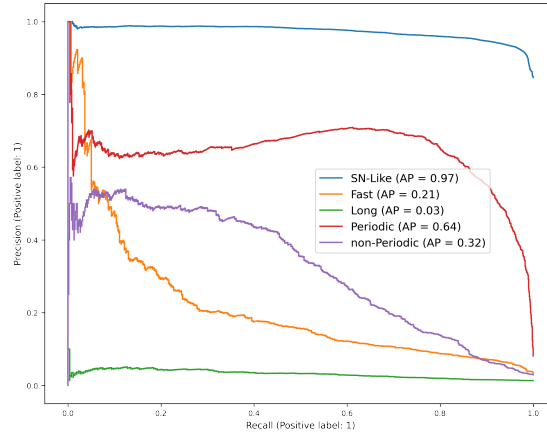(b) ROC curve for the dataset with a temporal advance of 5 days.

Figure 9

(a) Precision-Recall curve for the dataset with a temporal advance of 10 days.
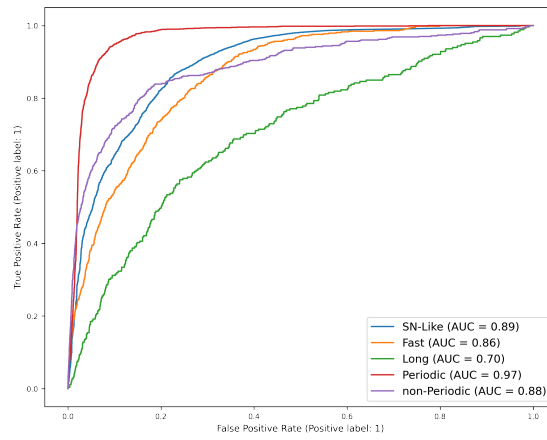

(b) ROC curve for the dataset with a temporal advance of 10 days.

Figure 10

(a) Precision-Recall curve for the dataset with a temporal advance of 20 days.

(b) ROC curve for the dataset with a temporal advance of 20 days.

Figure 11