# Identifying Variable AGB Stars in the Central Part of the M33 Galaxy via Machine Learning Algorithms

**Sogol Sanjaripour, Yousef Jamshidi, Amirreza Bahrami Moghadas**

## Abstract

We have conducted a near-infrared monitoring campaign at the UK Infrared Telescope (UKIRT), of the Local Group spiral galaxy M33 (Triangulum). The main aim was to identify stars in the very final stage of their evolution, and for which the luminosity is more directly related to the birth mass than the more numerous less-evolved giant stars that continue to increase in luminosity. The most extensive data set was obtained in the *K* band with the UIST instrument for the central 4 × 4 arcmin2 (1 kpc2) – this contains the nuclear star cluster and inner disc. These data, taken during the period 2003–2007, were complemented by *J*- and *H*-band images. Photometry was obtained for 18 398 stars in this region; of these, 812 stars were found to be variable, most of which are asymptotic giant branch (AGB) stars. Our data were matched to optical catalogues of variable stars and carbon stars and to mid-infrared photometry from the *Spitzer Space Telescope*. In this first of a series of papers, we present the methodology of the variability survey and the photometric catalogue – which is made publicly available at the Centre de Donnees astronomiques de Strasbourg – and discuss the properties of the variable stars. The dustiest AGB stars had not been previously identified in optical variability surveys, and our survey is also more complete for these types of stars than the *Spitzer* survey.

## 1. Introduction

Messier 33 is one of three stereotypical spiral galaxies that inhabit the Local Group. Located in the constellation of Triangulum, it spans about a degree on the sky. Its favourable inclination angle of 56◦ makes it a prime subject for the study of the detailed structure and stellar content of a spiral galaxy like our own. The distance to M33 has been determined *via* several methods as follows. Freedman, Wilson & Madore (1991) determined a distance of 850 kpc by using Cepheid-type variables, i.e. a distance modulus $\mu$ = 24.65 mag, but Scowcroft et al. (2009) revised this to $\mu$ = 24.53 ± 0.11 mag. By using RR Lyrae-type variables, Sarajedini et al. (2000) found $\mu$ = 24.84 mag.

## 2. Observation

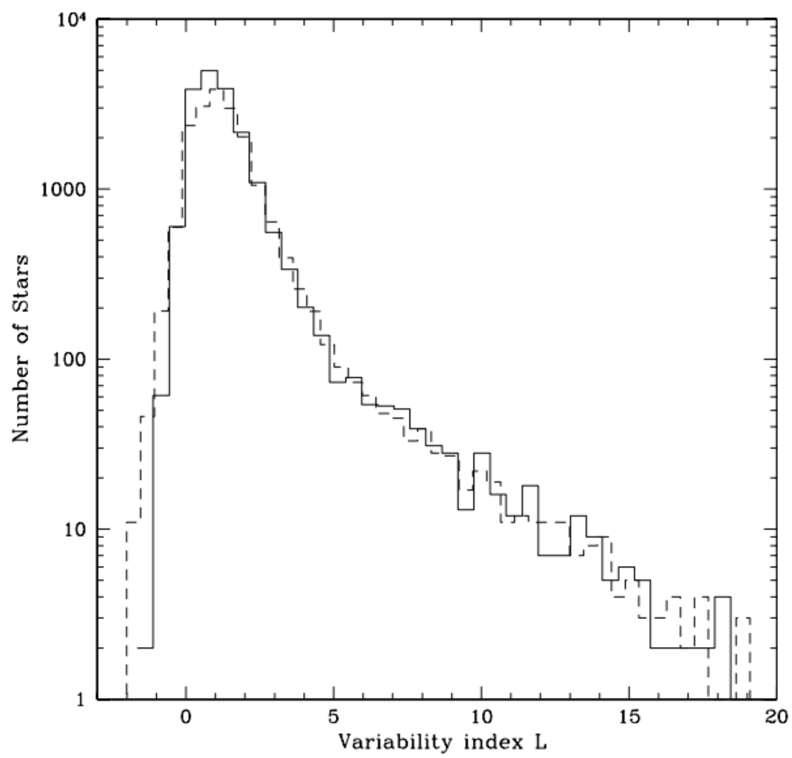
Observations were made with three of UK Infrared Telescope (UKIRT)'s imagers: UIST, UFTI and WFCAM. The WFCAM observations cover a much larger part of M33 and are discussed in Paper IV of this series.

## 3. Description of the catalogue

The photometric catalogue including all variable and non-variable stars is made publicly available at the Centre de Donnees as tronomiques de Strasbourg (CDS).

## 4. Data analysis

Via Stetson algorithms and analyzing the observational data by complicated procedures that take a lot of time I plotted the histogram of variability index L using all measured magnitudes for all 18398 observed stars. The final result was incredible, for a galaxy that its stars are all non-variable, the histogram is Gaussian; on the contrary, for M33 galaxy the histogram is not completely Gaussian and for L >4 we can see the distribution of variable stars.



Therefor, we decided to use Machine Learning Algorithms to identify variable stars of M33 galaxy.

During the second phase of the project, we tested KNN, Random Forest, SVM, Naive Bayes algorithms with high accuracy (roughly 99.7%).

KNN: We divided stars into 2 classes, variable and non-variable; then we transferred our data to a 12-dimensional vector space and used KNN algorithm to classify entries according to nearest neighbors in this vector space.

Random Forest: In order to find best depth of decision tree we tried to find ideal decision parameters for lowering loss by minimizing “gini” and “entropy” criterion.

SVM: This method is useful for high dimension data. In order to classify our data we try to separate classes using a line with best safety margin; for this purpose we transferred our data to higher dimensions using “phi” function.

Naive Bayes: In this method we assume all random variables are independent so we can classify our data by calculating the probability of each random variable.

In the third phase, we tested Neural Network algorithms on our data set to identify variable stars.

For the third phase we used Keras library with back-end Tensorflow.

From 4750 stars (750 variable stars) we dedicated 80% for training, 20% for validation and 20% for testing.

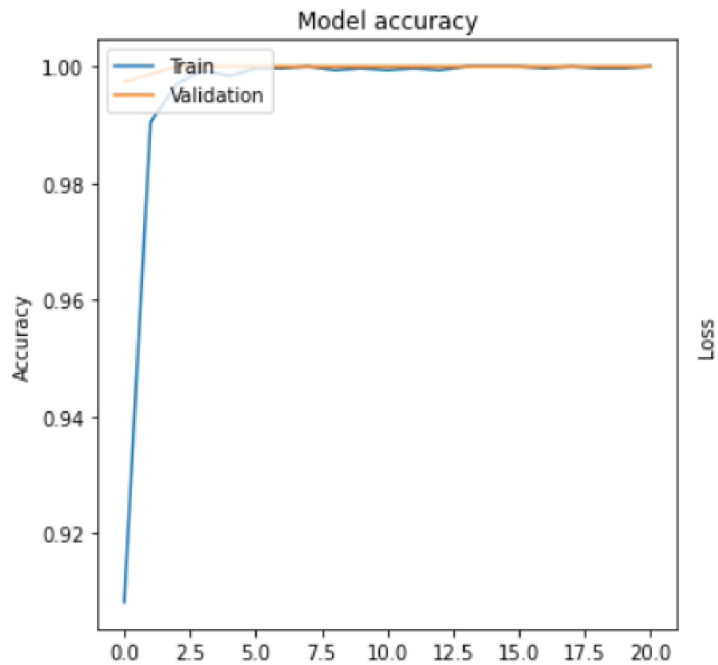
We made use of Split method to have a contribution with the same random seed as the previous phase.

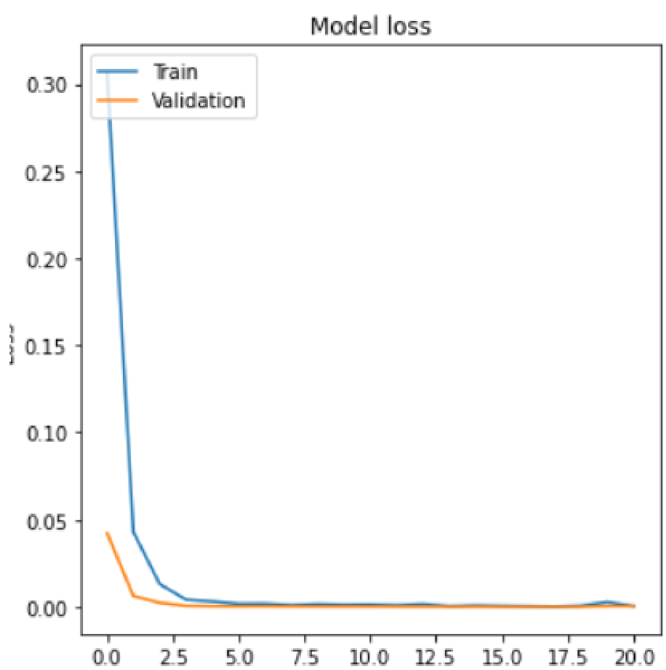
We had two fully connected layers respectively with 64 nodes followed by a dropout with the size of 0.5 and 64 nodes in the second layer; again a dropout with the size of 0.5 and the procedure ends with final layers having 1 node (the output). The reason for which we used dropout was to impede the model from over-fitting.

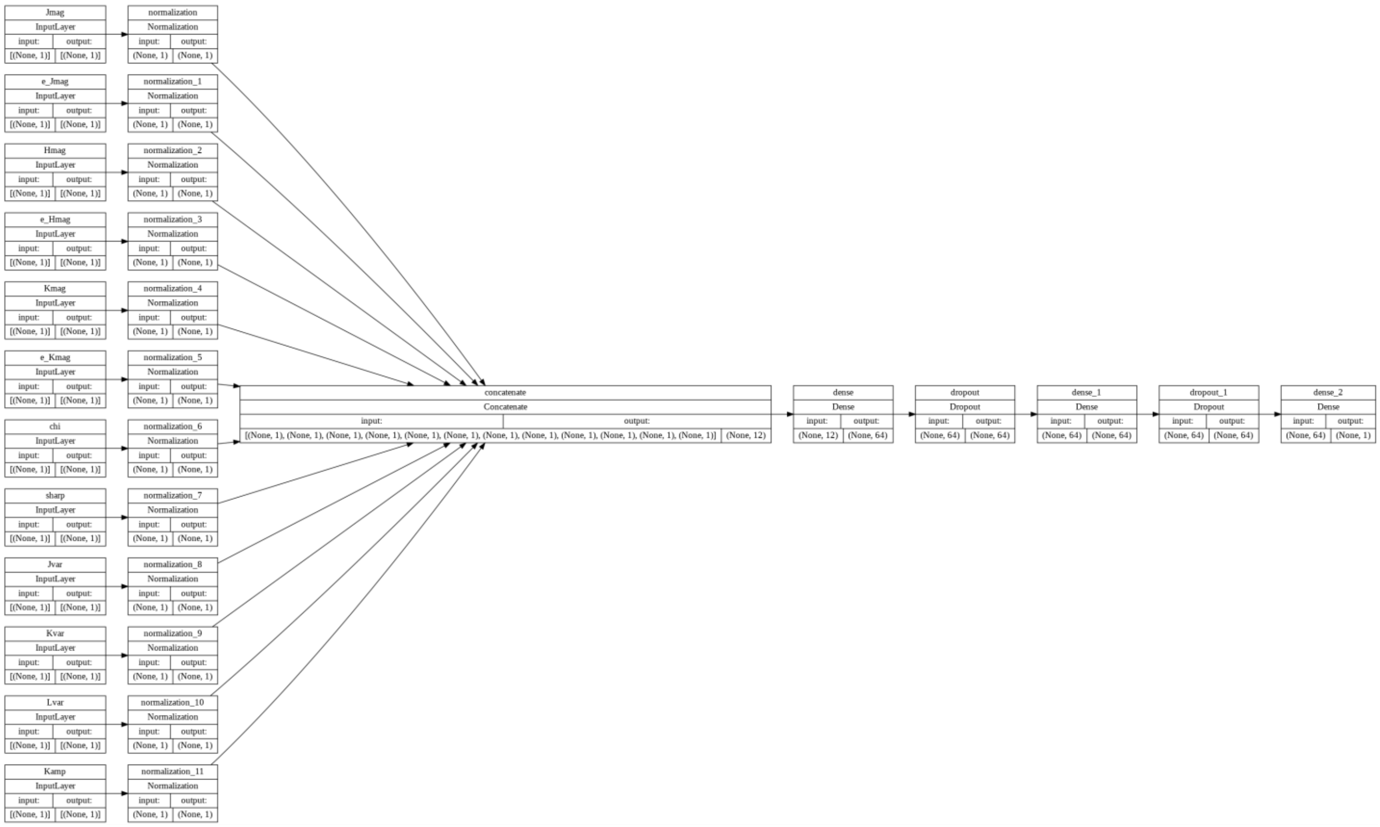
The activation function in the model was “relu” which does not cause the vanishing gradient. In the following steps, we used Adam optimizer and cross entropy loss function.

In the training part, we made use of two prime functions named “early stopping” and “check point”. In the whole, we let the function to reach 50 epochs to minimize the val-loss. “Early stopping” function stops the procedure after 2 epochs of increasing consecutively from being continued. In the end, “check point” function calculates the best result.

The results regarding the loss and accuracy of our model is illustrated:






## 5. Conclusion

We carried out research on stars of M33 galaxy to identify variable ones. In this process we had to identify the variable stars via Stetson algorithms and through methods that takes us considerable amount of time with a lot of effort for analyzing the data on the histogram manually. Taking advantage of Machine learning algorithms let us do our survey in less time with less effort. All methods that we used including KNN, Random Forest, Neural Network were well suited to our model and we got reliable results. The results were amazing and our machine learning methods had perfect accuracy in comparison to other methods.