Photometric LSST Astronomical Time-series Classification (PLAsTiC) Using Gradient Boosting Techniques

Author

Asad Mansoor Khan

00000204850

Supervisor

Dr. Sajid Gul Khawaja

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

FEBRUARY 2019

Photometric LSST Astronomical Time-series Classification (PLAsTiC) Using Gradient Boosting Techniques

Author

Asad Mansoor Khan

00000204850

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor

Dr. Sajid Gul Khawaja

Thesis Supervisor's Signature: ____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

FEBRUARY 2019

Declaration

I certify that this research work titled "*Photometric LSST Astronomical Time-series Classification (PLAsTiC) Using Gradient Boosting Techniques"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Asad Mansoor Khan

00000204850

Language Correctness Certificate

This thesis has been read by an English expert and is free of most typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student
Asad Mansoor Khan
00000204850

Signature of Supervisor
Dr. Sajid Gul Khawaja

Copyright Statement

Acknowledgements

By the grace of the lord above, my Master's thesis is finally complete. It was quite an arduous exercise that could not have been completed without the help of God Almighty and the strength that he bestowed upon me.

I am also eternally grateful to all those, especially my friends and family, who helped me with the thesis every step of the way. Their support was what made me accomplish this herculean task. Without them, it would not have been possible.

Lastly, I would like to offer my sincere thanks to my incredible supervisors and the entire committee: Sir Usman Akram, Sir Ali Saeed, Sir Sajid Gul Khawaja, Sir Farhan Hussain and Sir Arsalan Shaukat. Without their support – in every sense of the word – and constant guidance, this thesis would never have been conceived or completed (and definitely not in time). I cannot thank them enough for their role in the completion of this thesis and report.

Not Worth Dedicating to Anyone

Abstract

The light curve analysis of the heavenly bodies is an indispensable tool for understanding the physical phenomena that govern them. Doing so not only leads to new discoveries but also enhances our understanding of the universe. Large telescopes like the Large Synoptic Survey Telescope (LSST) will produce an excess of data that will necessitate the need for automated methods to sift through it quickly and efficiently, as doing so manually can be truly laborious. Furthermore, such a method should be able classify the observed astronomical objects accurately. Keeping this in view, this research presents an automated classification method using the simulated, photometric light curves provided in the Kaggle Challenge PLAsTiCC hosted by the LSST Team, in to 14 different classes. The classification model has been built around extracting several features and employing three different classifiers: Random Forest, eXtreme Gradient Boosting and Light GBM into an ensemble rounded off by a 5-layer Multilayer Perceptron (MLP). The training dataset containing 7848 samples has been used to train all three classifiers with different subsets of features sorted on the bases of their importance to the classifier. The MLP has then been trained on the concatenated probabilities of the three classifiers to predict the probabilities for 14 classes. The proposed methodology performs reasonably well for most of the classes achieving around an accuracy of 85% on the 3.5 million testing samples present in the test dataset. As the proposed methodology relies on features extracted from photometric light curves, therefore it can be adapted and extended for use in other fields that rely on similar light curves.

Key Words: *Large Synoptic Survey Telescope, Light Curves, Random Forests, eXtreme Gradient Boosting, Light GBM, Multilayer Perceptron, Deep Learning, Computer Aided Classification System, Astronomy*

Table of Contents

List of Figures

List of Tables

Chapter 1 : INTRODUCTION

Since time immemorial, man has looked upwards towards the heavens in an endeavor to decipher its mysteries, or at the very least, understand his place in the cosmos. Centuries of advancements have transformed the science of astronomy from its humble beginnings in ancient time [1] to a field whose refinement of numerous ideas and techniques, when applied to other aspects of modern life, greatly benefit them. The never-ending quest for knowledge of the skies and the ensuing cycle of advent of better and modern tools have helped improve life here on Earth in health, transportation, public safety and computer technology segments [2]. Technological breakthroughs have paved the way for today's massive telescopes (both optical and radio) [3], [4], [5] that peer into the darkest and oldest corners of the universe and have helped develop a better understanding of the inner workings of the universe and different forces that govern it.

A new ground based telescope called Large Synoptic Survey Telescope (LSST) [6] is being added to astronomy's arsenal that will allow us to look farther and deeper than we have done before [7]. Owing to its massive size, the amount of data that LSST will generate will be unparalleled and so it necessitates an automated method for swiftly and correctly identifying different astronomical phenomena so that the interesting astronomical objects can be studied in detail. This research presents an automated classification model for 14 different types of astronomical objects.

1.1 Motivation

Due to new, powerful telescopes becoming operational all around the globe, the amount and the availability of astronomical data has skyrocketed. This has led to several new discoveries which have helped progress our understanding of the cosmos. Understanding the world around us is a basic human instinct similar to space exploration and space travel. In addition, recent advances in computation have allowed us to manipulate this abundance of data with ease. The matrimony of this data and computational prowess can lead to better, speedier and efficient solutions of characterizing the said data. Accurate and swift classification of astronomical objects from their light curves can significantly cut back on the manual effort required by the researchers and can really help the diverse scientific community to achieve equally diverse scientific goals.

In order to comprehend how a system works and what laws and rules make it turn, we have to appreciate how its components work. Therefore, identifying and studying far-flung stars and galaxies can grow our knowledge base and can help us make sense of the universe.

While interstellar travel might still be a distant dream, the pursuit of understanding the heavens by looking up from our pale blue dot is as old as time itself and playing a small portion in this very human endeavor is the motivation for taking on this research work.

1.2 Problem Statement

Swift and accurate prediction of astronomical entities from the data gathered from LSST (and similar telescopes) can speed up the process of identifying and studying old and new cosmological phenomena. However, using light curves for identifying astrophysical phenomenon poses a serious challenge as they can include a large amount of noise that results in difficulty in identifying the classes of these objects.

The purpose of this research is to explore different features that can help separate objects into their respective class based on their light curves along with building an ensemble of different classifiers including deep learning techniques.

1.3 Aims and Objectives

Major objectives of the research are as follow:

- To reform raw, simulated astronomical data to be used for classification
- Explore gradient boosting techniques for analysis of astronomical data
- Devise an algorithm for classification of astronomical bodies into existing and new classes
- To find out best features for proper representation of data

1.4 Structure of Thesis

This work is structured as follows:

Chapter 2 covers the basics of LSST including the advances that make this telescope possible in addition with characteristics of the data captured by it.

Chapter 3 gives review of the literature and the significant work done by researchers in past few years for classification of heavenly entities using light curves.

Chapter 4 consists of the proposed methodology in detail. It includes the details about the features extracted and subsequently used along with the ensemble of different classifiers.

Chapter 5 includes all the experimental results accompanied by relevant figures.

Chapter 6 concludes the thesis and reveals future scope of this research.

Chapter 2 : LARGE SYNOPTIC SURVEY TELESCOPE

Telescopes have revolutionized the way we observe the universe and, consequently, our understanding of it. Since the first telescope in the 1600s [8], they have improved leaps and bounds, with the ability to “see” more than the naked human could ever see. This chapter will briefly cover the Large Synoptic Survey Telescope, its observational and imaging capabilities, the database management system that it will employ, and, finally, the form of data that will be made available to scientists courtesy of this telescope.

During the last couple of decades, the scientific community has greatly benefited from several large-scale surveys of the sky ranging from Sloan Digital Sky Survey [9] to Galaxy Evolution Explorer [10]. Owing to the success of these surveys in the form of a slew of new scientific discoveries, the genesis of another large-scale survey of the entire Southern sky [11] was just a natural progression of such scientific ventures. The Large Synoptic Survey Telescope (LSST) is at the heart of this ambitious undertaking.

As the name implies, LSST aims to photograph and study a large portion of the sky for a course of 10 years [6] once it becomes fully operational in 2022 [11]. Located at Cerro Pachon, Chile [11], this telescope will observe the flux changes of millions upon millions of different astronomical entities every night and will catalogue the changes observed for each of these objects.

LSST has been conceived with a number of diverse science drivers in mind that also dictate the physical dimensions and other restrictions of the design. Primarily, the data gained from the LSST will be used by the scientific community for:

- Better understanding of the dark energy and dark matter, how they shape the universe and how the dark energy’s behavior is influenced over time [6], [12]
- Better understanding and charting of the millions of small to medium sized bodies that populate the Solar System along with keeping track of potentially hazardous asteroids that are larger than 140 meters [6], [12]
- Better understanding of the fast transients exhibited by different astronomical objects by repeatedly photographing the night sky in addition to providing timely notifications of such phenomenon [6], [12]
- Better understanding and charting of our Milky Way Galaxy with high quality, accurate data than ever before [6]

The current LSST design and dimensions ensure that the above-mentioned goals are met.

2.1 LSST Site and Facility

The LSST is under construction in north-central Chile with the Andes Mountain range as the backdrop on a site owned by the Association of Universities for Research in Astronomy [13]. Building such a telescope necessitates that the housing of the telescope is adequately equipped to alleviate the environmental issues that are bound to plague a setup of this proportion. The testing of the site in the form of geotechnical studies, computational fluid dynamics modelling and weather simulations was carried out and it was determined that this location provides the most stable platform for installing the LSST [14]. Figure 2.1 shows the cutaway diagram of the various sections of the entire facility.

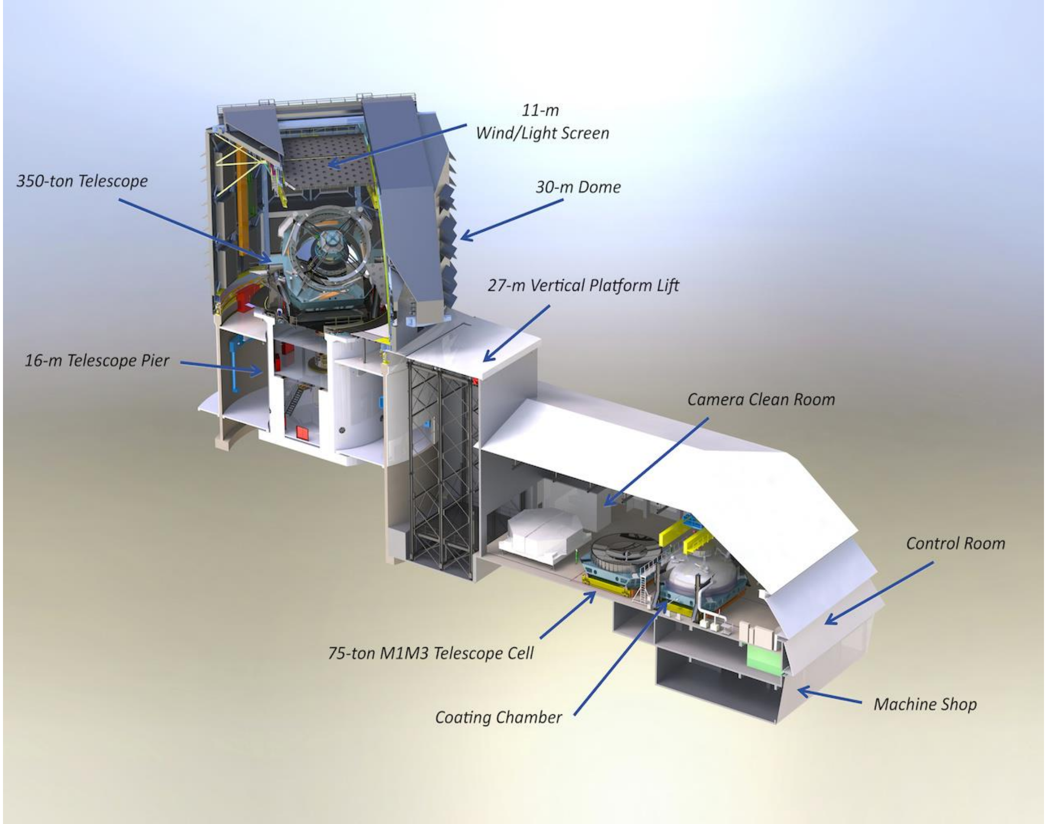


Figure 2.1 Cutaway View of the LSST Facility [14]

The telescope pier that will harbor the optical system is built on the highest point of the peak and rises to 16 meters culminating in a 30-meter wide dome that is capable of rotating thus allowing the whole telescope to point at the desired patch of the sky. An attached 3000 m^2 building will act as the services and operations hub where not only the mirrors of the LSST will be coated before initial installation but will also be serviced during their entire lifetime. [14]. For maintenance or resuming operations, an 80-ton platform will carry the telescope components to and from the pier. The entire facility will temperature-controlled and different sections of the housing will be kept at different temperatures [14]. Once the LSST facility is up and running, it will be highly automated requiring only limited number of workers to operate it [16].

2.2 LSST Design and Dimensions

LSST has been designed from the ground up by keeping in view the diverse goals that were to be met. This has resulted in radical new designs and improved subsystems that push the boundaries of the design and functionality of such ground based telescopes.

The LSST optical design has incorporated a number of innovative concepts that set it apart from telescopes of similar size. While LSST's primary mirror has a diameter of 8.4 meter, its three-mirror design results in a remarkable improvement in its Field of View (FoV). The primary mirror M1 has a diameter of 8.4 meter but its effective diameter comes out to be at 6.5 meter; M2, the secondary convex mirror has a diameter of 3.5 meter that reflects the light into a tertiary, 5-meter mirror M3 [6]. Figure 2.2 shows mirror assembly the LSST and how all mirrors reflect the light.

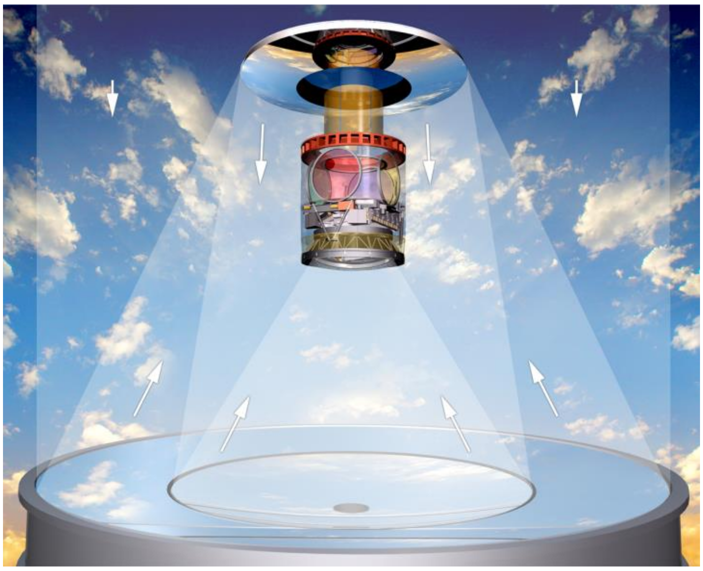


Figure 2.2 Rendering of LSST at 45 Degrees

LSST packs the above three-mirror assembly along with the camera in a compact 6.4-meter length from the vertex of M2 to the vertex of M3 [17]. Because of this design, LSST has a FoV of 9.62 degree^2 (a radius of 1.75 degrees) which is a massive improvement compared to other telescopes in the same 8 meter category. This enables the LSST to capture an area of the sky in which seven full moons can fit end-to-end. In comparison, the Hubble Space Telescope is only able to view a small portion of the moon itself [6]. Figure 2.3 illustrates the difference between the FoV of a typical telescope of similar size primary mirror and that of the LSST.

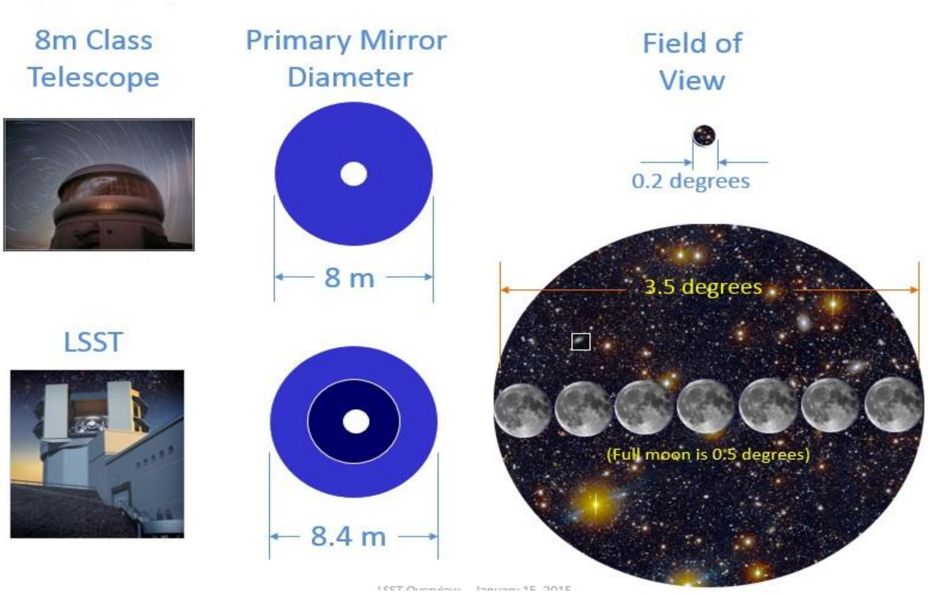


Figure 2.3 LSST's Field of View Comparison to a Similar Telescope [13]

Owing to its incredible optical capabilities, the LSST will be able to resolve details that are even smaller than the width of a human hair that is held at arm’s length [14]. All these improvements will enable the scientists to observe the sky in unprecedented detail.

The incredible optical system of the LSST is complemented by an equally impressive camera system. LSST’s camera takes the crown of the largest digital camera ever constructed at 1.65 meter by 3 meter and is nearly as tall as an average human adult is while weighing mammoth 2.8 metric tons [20]. Figure 2.4 shows the cutaway diagram of the camera.

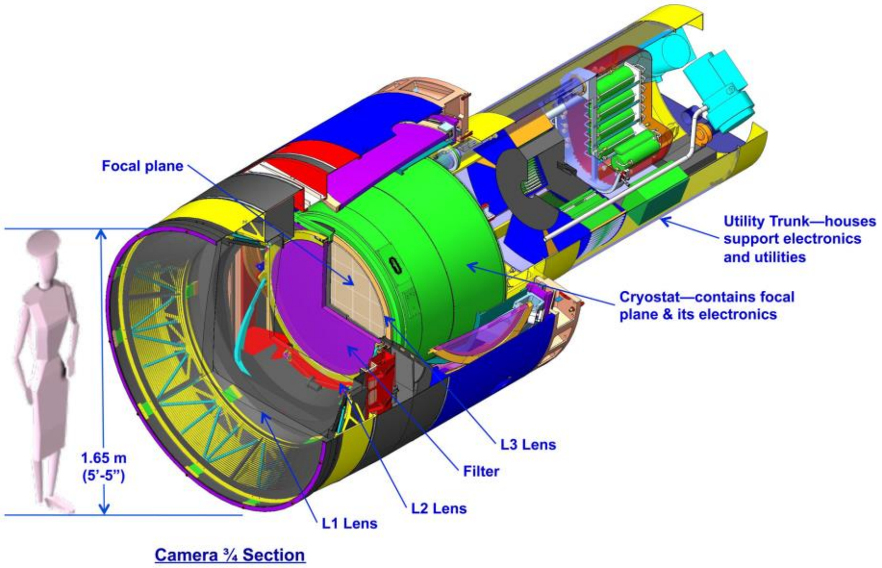

Focal plane
Utility Trunk—houses support electronics and utilities
Cryostat—contains focal plane & its electronics
1.65 m
(5’-5”)
L3 Lens
Filter
L2 Lens
L1 Lens
Camera ¾ Section

Figure 2.4 Cross Section of LSST Camera [16]

LSST’s camera houses three lenses; the first lens L1 is also the largest one measuring 1.55 meter in diameter, the second lens L2 has a diameter of 1.1 meter and the third lens L3 is relatively small at 0.72 meter. In addition to focusing light as part of the three-lens system, L3 also acts as a vacuum barrier for the pixel array of the camera that lies after it. The different filters that allow to LSST to observe the universe in several wavelengths is present between the L2 and L3 [17].

To build such a massive camera that lies at the heart of the LSST, 9 charge-coupled device (CCD) imaging sensors of 4096x4096 pixels each are arranged in a 3x3 grid known as a “raft”. Each raft measures about 16.8 Megapixel. 21 such rafts are then combined to form the entire imaging plane of the LSST’s camera that has a diameter of 64 centimeters [6], [20]. Each CCD sensor is capable of delivering 16 outputs, so, consequently, each raft has 144 outputs adding up to 3024 channels for the entire camera. In order to carry the data from the sensors to the on-facility data storage and processing equipment, the rafts are mounted on a tower-like structure that holds the electronics to convert the signals into a digital format and then transmit it onwards. As the camera will be operational for long periods, therefore, it is necessary that they have adequate cooling built in to prevent overheating. This is accomplished by adding cooling elements to the tower that help keep the temperature of the sensors to -100 degree Centigrade. The benefit of using such tower like structures is that each tower acts like an individual camera that can then be interconnected to form the entire array [22]. Figure 2.5 shows how each sensor is packaged, joined with other sensors in a raft and how each raft is mounter on a tower.

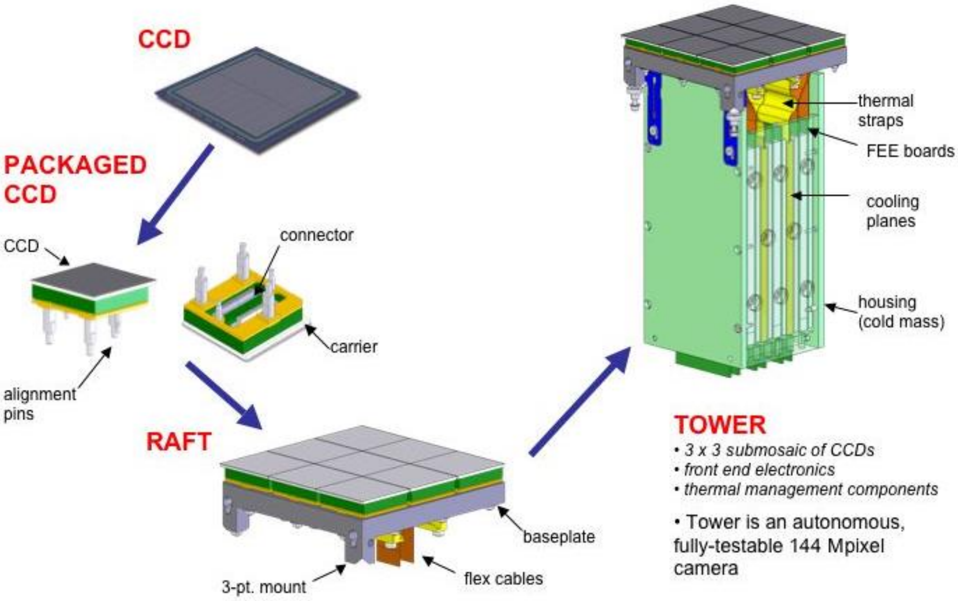


Figure 2.5 Raft and Tower Assembly [22]

Each pixel in the camera has a significant physical size of 10 micrometers with a dynamic range of 18 bits/pixel and the entire array is capable of imaging the entire sky in just 2 seconds. In addition to the imaging pixels, different types of sensors (wave front and guide sensors) are also present on the imaging plane. The focal plane and its accompanying electronics are contained in a cryostat chamber. [6], [20]. Figure 2.6 shows the intricate structure of the imaging focal plane.

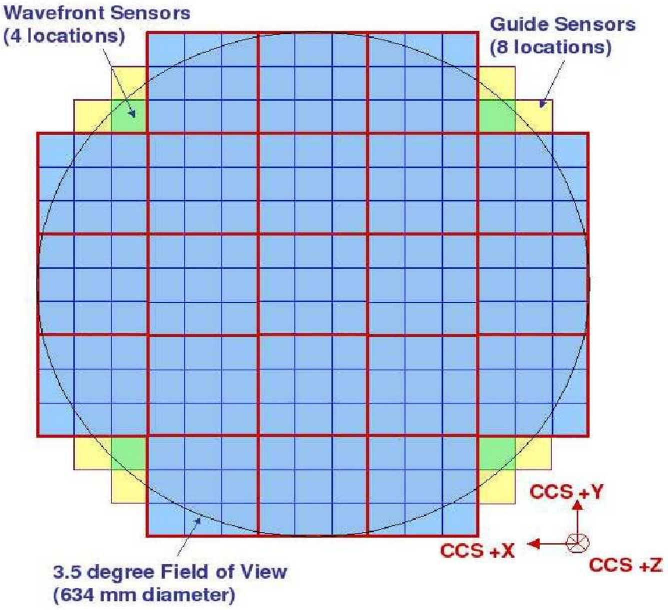


Figure 2.6 LSST Pixel Array measuring a mammoth 3.2 Gigapixels [6]

The camera system of the LSST has been coupled with a series of different filters that allow the telescope to detect different wavelengths of light and thus view different types of astronomical entities in different light.

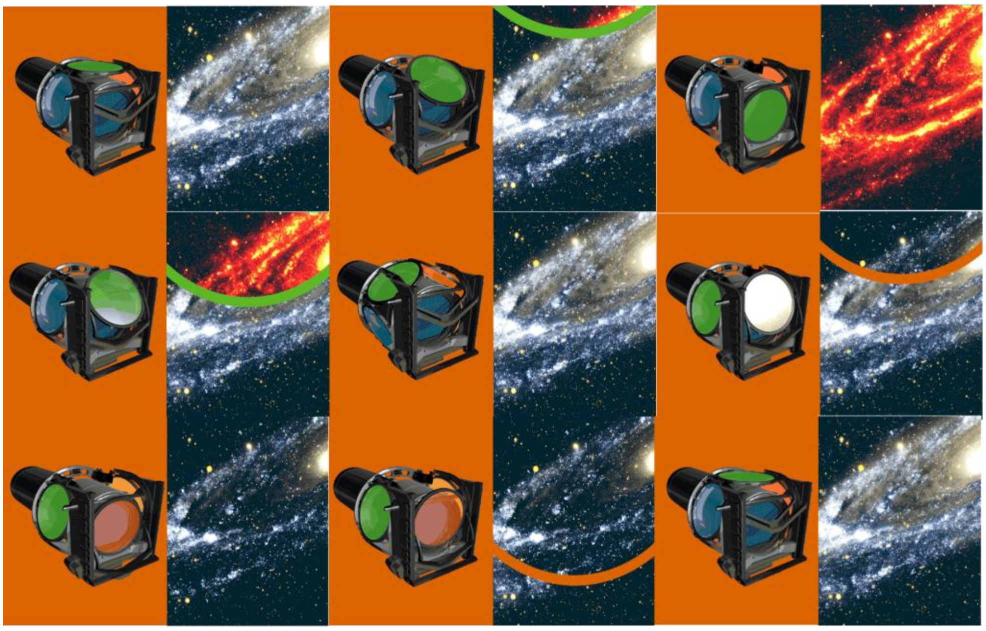


Figure 2.7 Filter Changing Mechanism [20]

The filters reside above the cryostat chamber in a circular pattern and are pulled down in front of the L3 by a rail-type mechanical arm within two minutes. This location of the filters ensures that they are in a temperature-controlled environment thus significantly reducing light distortions. Figure 2.7 shows how the lenses will be switched and the effect that it will have on the images captured by the LSST.

The mechanical shutter of the camera has been designed to reduce the wear and tear of the components over the lifetime of the telescope by making the shutter blinds close alternatively [20]. Figure 2.8 (a, b and c) depict the action of the shutter blinds.

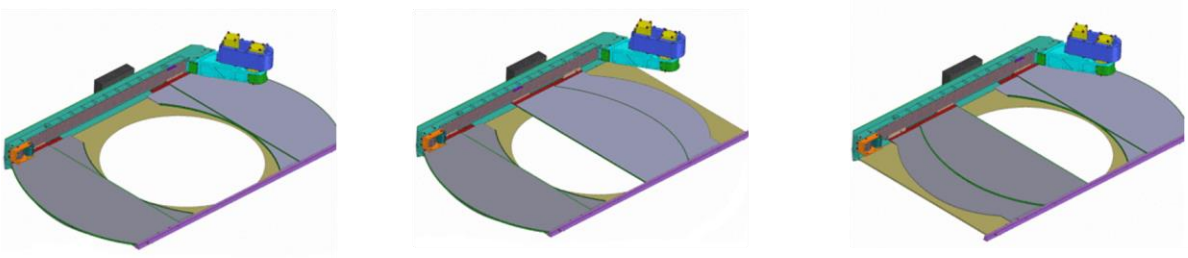


Figure 2.8 (a) Shutter Open, (b) Right Shutter Blind Closing, (c) Left Shutter Bind Closing

2.3 LSST Passbands

Astronomical entities exhibit different behavior in different bands of electromagnetic spectrum and observing them in several wavelengths can reveal more about these bodies than observing them in a single wavelength would ever accomplish [6]. For example, infrared wavelength can pass through dust clouds enabling to survey the astronomical entities that are hidden from direct view [17]. The data gathered through such techniques has resulted in important discoveries ranging from the determining the redshift values of the galaxies [21] to the selection of quasars based on their photometric values [22]. LSST filter bank has been inspired Sloan Digital Sky Survey [9].

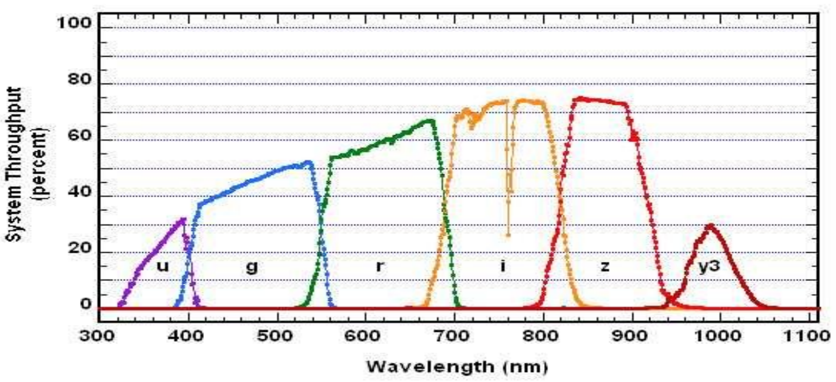


Figure 2.9 LSST Band passes [6]

The filters have been designed to detect bands of light from ultraviolet to near infrared and have been symbolized as u, g, r, i, z and y. u band accommodates the wavelengths between 300 to 400 nanometers, g band houses the wavelengths between 400 to 600, r band lies in 500 to

700 nanometer range, 650 to 850 nanometer wavelength are present in i band, z band houses 800 to 950 mm and y band holds 950 to 1050 nm wavelengths [11]. Figure 2.9 depicts the design of the filters of the LSST.

If the light coming from an astral entity is directly incident on the center of the focal plane (or within 0.7 degrees on either side), the LSST optical system is able to capture 63% of this light. The percentage steadily drops to 57% at the end of the FoV of the camera. Figure 2.10 shows how the throughput of the system drops when moving away from the center of the FoV [17].

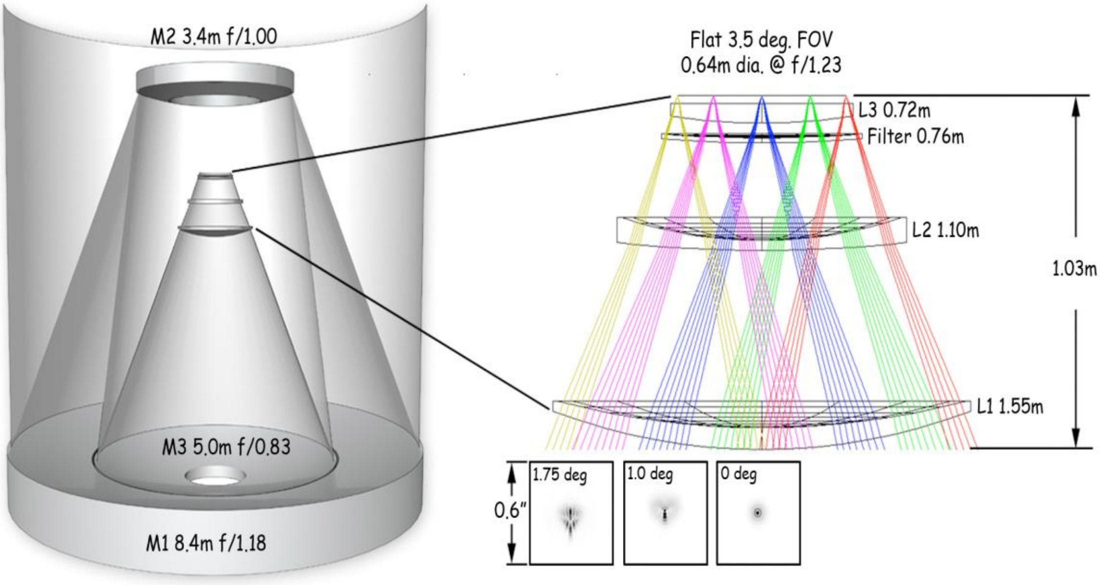


Figure 2.10 Effect on Throughput when Moving Away from the Centre of the Focal Plane [23]

2.4 LSST Data Management

Owing to the scale of the LSST, it will generate data at an unprecedented rate totaling more than 20 terabytes of raw data every night. Over the course of its 10-year life, LSST will gather 60 Petabytes of raw data that will result in an enormous catalogue of astronomical bodies amounting to 15 Petabytes [6]. LSST’s source catalogue will have approximately 7 trillion rows while the object catalog will comprise of 37 billion rows containing more than 200 attributes each [27]. Such an extraordinary amount of data will require significant computation prowess; 150 trillion floating point operations per Second (TFLOPS) will process the data creating several hundred Petabytes of data. As the data collected by the telescope will increase with every year, so will the computational power required to process it. By the last year of LSST’s lifetime, 950 TFLOPS will be required to process the entirety of the data [28].

In order to deal with this onslaught of data, a data management system had to be developed from scratch capable of handling the storage and processing requirements. Not only that, LSST’s data management system will also have to ensure data integrity i.e. the data is not damaged when being converted from raw image capture of the LSST's camera to such a format that truly represent the unfolding of the universe. In order to accomplish this herculean task, LSST's data management system consists of three main layers. The bottommost infrastructure layer that handles the bulk of the tasks such as storage and computation along with networking and software stack that links them all together. The middle layer is mainly responsible for distribution of computation among the nodes and an interface for users to access data. The top application layer centered on the data products that will be produced by LSST [28]. All the data gathered by the LSST will be transmitted to several, remote data centers for access and archival purposes [6].

In order to streamline the data collection and cataloging, the LSST team has come up with an inventive sphere that divides the sky visible from the Southern hemisphere into different quadrants. Figure 2.11 illustrates this concept.

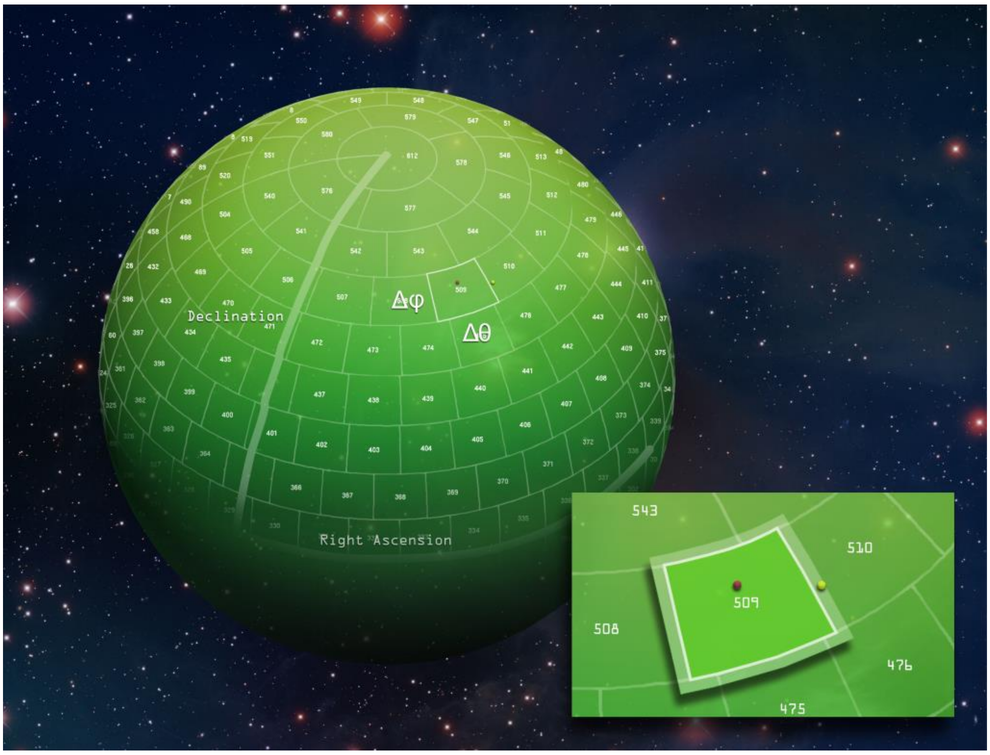


Figure 2.11 LSST Data Mining Sphere [25]

As a testament to the capability of data management system, interesting transient events detected by the subtraction of two consecutive nightly images will be relayed to the scientific community within 60 seconds of such an event being detected [28].

The data gathered by the difference of images from a reference image will result in flux value measurements for each heavenly body in the FoV of LSST. This flux values will be used to generate time-series light curves that will help identify different types of astrophysical phenomenon visible from this vantage point [11].

The innovations in the subsystems of the Large Synoptic Survey Telescope make it one of the most powerful telescope of the next decade. The data that the LSST will capture will undoubtedly result in new discoveries that will broaden our current understanding of the universe.

Chapter 3 : LITERATURE REVIEW

The recent years have seen an upsurge in the data generated by the telescopes – both ground based and orbital – and that has led to a need for automated methods for detection of different astronomical entities by finding patterns in the data that require minimal to no human intervention. For this reason, researchers have applied machine-learning techniques for solving different problems. This chapter summarizes some of the valuable contributions that have been made to this field.

3.1 Traditional Machine Learning Techniques

Richards et al. [30] were one of the first ones to use machine learning techniques specifically Random classifier. The scientists computed both statistical and periodic features for classifying 25-class dataset of various stars. They fitted different periods to the light curves in order to estimate a true representation of the actual period. In addition, the researches employed hierarchal classification for certain star classes while using the feature importance calculated by the Random Forrest classifier as a heuristic as to use which features and drop others. Using this procedure, they were able to achieve an overall improvement of 24% over the previous best solution while for a couple of classes, the accuracy was over 98 percent representing a huge improvement.

Jenkins et al. [31] proposed a method for reducing the number of Threshold Crossing Events for vetting by Kepler Threshold Crossing Event Review Team that relied on a classifier based on Random Forest. This facilitated the researchers by bringing down the initial number of the interesting objects by a wide margin (along with the time requirements of accomplishing this task) thus providing a smaller list that could be analyzed in-depth. By classifying the planet candidates as planet candidate, astrophysical false positive or a non-transiting phenomenon, the researchers wanted to eliminate as much possible candidates as possible so that most interesting cases could be focused on first.

Tackling the ever-persistent problem of gigantic amount of data that could not be classified in a reasonable amount of time by human effort, Coughlin et al. [32] proposed a set of algorithms for automatic vetting of the threshold crossing events from the Kepler pipeline that tried to mimic the decision making process of humans vetting the same data. These algorithms named "Robovetters" tried to determine whether the shape of the signal is transit like or not and whether another transit like event exists in the same signal indicating an eclipsing binary system. In addition, they also used the pixel level data captured by the telescope in conjunction with the light curves. The researchers used Q1-Q16 catalogues of Kepler Data Pipeline as benchmarks. Their methodology performed almost on par with the human expertise. For Q17, the results of only Robovetter were presented.

Thompson et al. [33] made improvements to the "Robovetter" algorithms. In order to improve the performance of their system, the researchers used light curves that have been detreneded. Each light curve was folded in bins totaling N which represented the number of initial dimensions. Then leveraging dimensionality reduction the number of features were further reduced based on Locality Preserving Projections.

These "improved" samples were used for k-nearest neighbor algorithm for removing non-transient events. The researchers' solution was able to successfully detect 99% of already known planet candidates while removing 90% of non-transiting events.

Mullally et al. [34] also proposed a method to eliminate false positives further improving the Robovetter. The researchers modelled actual transiting planets and other artifacts that mimic such behavior. Then, threshold crossing events were compared with each of the model. Bayesian Information Criterion was used to determine which model fits the profile of an individual event more closely. They were able to detect successfully false positives approximately 60 to 70 percent of the time while retaining the correct classification for 95 percent of true transits. This approach can be used for data of other missions as well as it relies on modelling both the true transits and the false alarms and then using them for comparison on sample-to-sample basis.

For classification of different supernovae using photometric data, Lochner et al. [35] tested different combinations of machine learning methods and handcrafted features obtained from different techniques. Starting with template fitting the researchers moved on to wavelet based decomposition after interpolating the data with Gaussian Process Regression Technique. Principal Component Analysis was further applied to the output of Wavelet decomposition to reduce the dimensionality. The researchers tried a number of robust machine learning techniques ranging from naïve Bayes to k-nearest neighbor, from Support Vector Machines to Multilayer Perceptron, finally settling on Boosted Decision Trees. The team was able to achieve a score of 0.98 on Area under the Curve metric commonly used for comparing the performance of different machine learning algorithms.

Hartley et al. [36] used Support Vector Machines in order to detect gravitational lensing systems i.e. the gravity of a massive object distorting the image of another object in its background to bigger proportions. Galaxies that exhibit such behavior usually have elliptical shape due to strong gravitational lensing. Detection of arc like and ring like objects in raw images while rejecting false positive given by ring galaxies was accomplished by the use of Gabor Filters that not only detected the desired shape but also made use of the information of the colour in the image. Feature elimination was done by using brute force. Their results showed that the algorithm developed by the scientists outperformed detection by visual inspections in majority of the cases.

Mislis et al. [37] realized that better predictions could be made if time-series data was cleaned of undesirable noise even after it had undergone detrending. To accomplish this, they employed a new clustering technique called TSARDI which itself was based on DBSCAN. DBSCAN is a clustering algorithm that divides a set of points into separate clusters, which have a minimum number of points, separated not more than a maximum distance defined by a certain distance criterion.

TSARDI creates "de-noised" light curves by implementing DBSCAN algorithm four times consecutively. However, at each step the maximum distance between the points, the distance criterion and the minimum number of points in each cluster is varied. The output of one step of TSARDI is becomes the input of the next step. With this method, their results showed that they were able to improve the results by almost 11 percent compared to the simple technique of sigma clipping. TSARDI is built from the ground up to work on any time series data.

3.2 Deep Learning Techniques

In addition to conventional machine learning techniques, some researchers have explored deep learning techniques for the classification of astronomical entities or phenomenon using photometric or photographic data. Although, deep learning itself and its application to such problems is not new, it has seen a definite uptick in the recent years due to computation becoming cheap and accessible. Therefore, researchers have applied deep learning methods such as neural networks and convolutional neural networks (CNN) for classification.

[38] applied a simple 3 layer perceptron with 5 input neurons, a hidden layer of the same size and one neuron on the output layer with the goal to correctly identify microlensing event against other variable events such as eclipsing variable stars, eruptive stars, cataclysmic stars and pulsating stars. As the microlensing event is an extremely rare event in the universe therefore it can be easily confused by any of the astronomical entities mentioned here. Instead of feeding the raw light curves, five features were computed and that feature set was fed to the neural network. The results showed that the network was able to classify the microlensing events with great accuracy, misclassifying only 3 instances out 800.

In order to improve the detection of microlensing events, Belokurov et al. [39] proposed a combination of networks connected one after another. The first network was trained to detect whether an event could be a microlensing event instead of any other types of the variable stars. The second network in the series was trained to eliminate supernovae masquerading as microlensing event. A separate algorithm is then used on this information to compute the microlensing rate. In order for an event to be classified as a microlensing event, it must pass through both the neural networks. Both the networks used variable number of neurons and hidden layers.

[40] applied multiple Multilayer Perceptron (MLP) and 1 dimensional Convolutional Neural Network to time-series data from the Kepler pipeline for the detection of exoplanets. Exoplanets are Earth like planets (similar in size, atmosphere etc.) that orbit distant starts in that star's habitable zone and can, theoretically, support life. MLPs had 4 layers each with 64, 32, 8 and one neuron in each successive layer while the input features were approximately 180. For one of the MLPs, they used the Wavelet transform on the light curves and fed the detail coefficients as features at the input layer.

For the CNN approach, the team used a set of 4 filters with length 6. After convolution, the outputs were concatenated together and treated as input of an MLP with similar architecture described above. While all the neural networks outperformed the conventional machine learning algorithms like SVM and Wavelet transformations, the Wavelet MLP showed the best performance.

Kipping et al. [41] trained a MLP for prediction of transiting planets that might be in the same region as other transiting planets missed by a combination of sensitivity of the telescope (Transiting Exoplanet Survey Satellite, in this case) and noise in the data. The researchers used hand crafted features that were then fed to the MLP for the prediction of additional planets orbiting the host star. The approach used by the scientists improved the possible planet candidates by a factor of two.

Shallue et al. [42] also used a 1D CNN for identification of exoplanets. The CNN architecture that performed the best 20 layers deep at its maximum depth. The researchers fed their CNN two different input presenting different "views" of the light curve namely Global view and Local View. These views were obtained by different sized binning of the light curve. The Global View had a size of 1x2001 while the Local View was limited to 1x201. The Global View passed 10 Convolutional layers while the Local View was subjected to only 4 such layers with intermittent Max Pooling layers. After both the views have traversed through the

Convolutional Layers, they were concatenated together and fed to 4 Fully Connected layers connected in a series before the final output layer.

Their approach was able to correctly identify planet candidates 98.8% of the time reducing greatly the overall number of false positives. This led to the statistical validation of two new Kepler Planet Candidates one of which was part of a 5-planet system orbiting Kepler-80 while the other one was member of an 8-planet system around Kepler-90. This discovery made Kepler-90 to be the only know star with 8 planets like our Sun.

For classifying supernovae based on their photometric data, [43] utilized a specific type of Recurrent Neural Networks termed as Long Short Time Memory (LSTM). Recurrent Neural Networks and by extension LSTM perform better on sequential data such has the time-series data of the light curves. The investigator improved upon the design of LSTM by integrating the time of observation as an input. Furthermore, as deep learning techniques require a huge amount of data, so data augmentation was also performed by using Gaussian noise for generating new data points. Another technique that was used to increase the amount of data was early truncation of artificially generated light curves. The team was able to achieve an accuracy of approximately 93.2% on classification of type Ia supernovae classification against the rest.

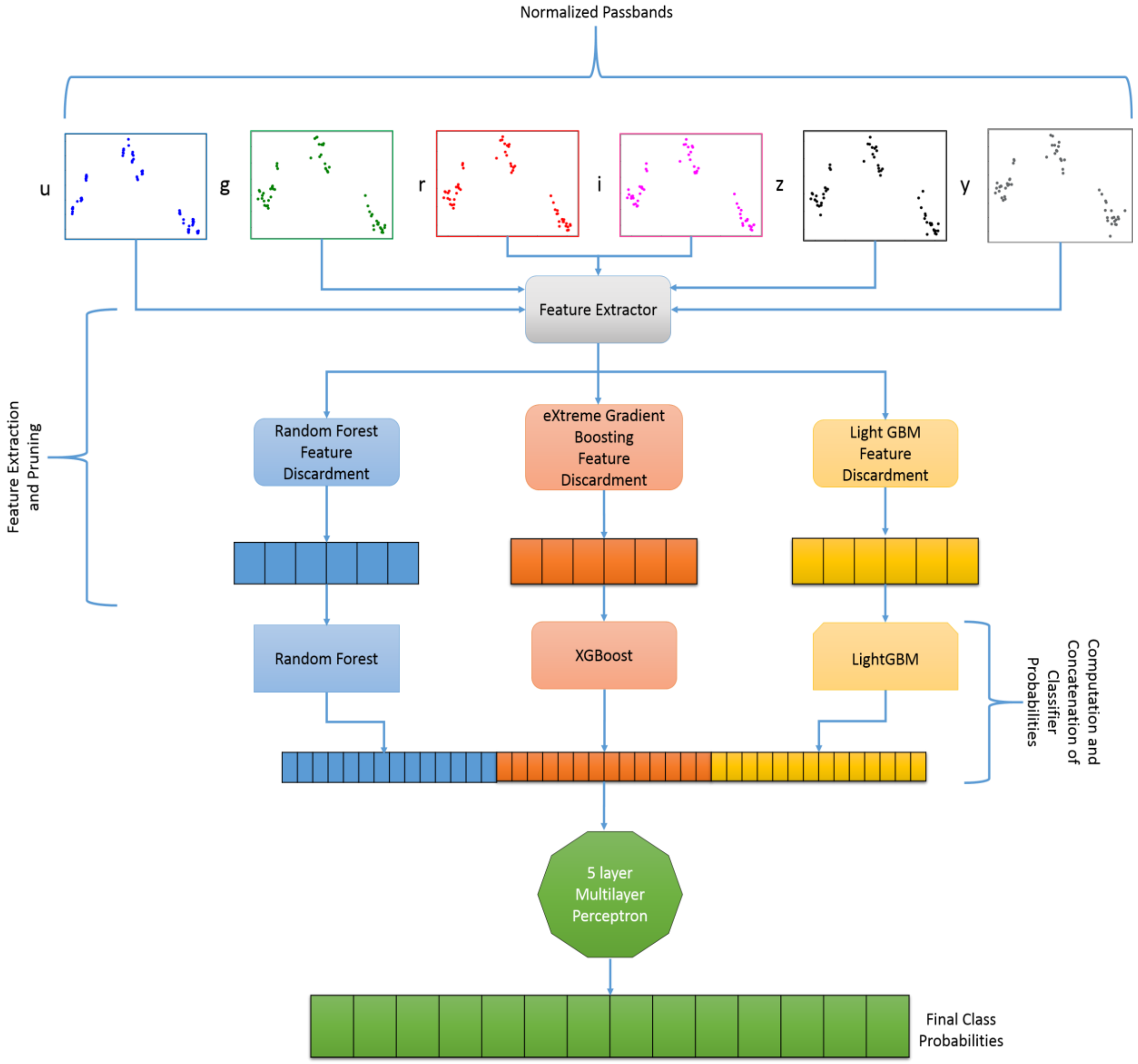
The literature review shows that the automatic classification of light curves (regardless of whether they are light curves of transiting planets or other astronomical entities) has been accomplished using several machine-learning techniques to varying degrees of success and accuracy. The algorithms that perform the best are those that mimic the human decision making process. Therefore, algorithms like bagging (Random Forests) and boosting perform really well for this kind of data. In addition, modelling the source of the light curve along with template matching can give good results. Another important thing to note is that detrending the raw flux values using different clustering or noise removing techniques can have a large effect on the performance of the subsequent classifiers used.

In addition to traditional machine learning techniques, the use of deep learning for light curve classification is on the rise as it eliminates the need for excessive preprocessing and feature engineering. Using multilayer perceptrons and 1D convolutional neural networks, researchers have achieved reasonable accuracy for different datasets. While deep learning eliminates the need for feature engineering and extraction for the most part, feeding the neural networks hand crafted features can improve the accuracy even further. In fact, the combination of these two outperforms most of the traditional machine learning techniques. One shortcoming of using deep nets is that there needs to be a large amount of data for the training stage. For this reason, for some of the cases, the data has be augmented.

Chapter 4 : METHODOLOGY

This thesis presents a method for automatic classification of light curves to identify different astronomical objects observed by the LSST. The proposed methodology consists of two main phases i.e. feature extraction and classification. However, before the feature extraction phase, some pre-processing is applied to the raw data. Then the first phase extracts a number of features for an accurate presentation of light curves that eventually supports the classification of these objects.

The classification phase builds on the extracted features and uses an ensemble comprising of Random Forest (RF) [44], eXtreme Gradient Boosting (XGB) [45], Light GBM (LGBM) [46, 46] and a Multilayer Perceptron (MLP) at the end to identify the astronomical entities. Figure 4.1 shows the constituent steps of the proposed methodology.



Normalized Passbands
u
g
r
i
z
y
Feature Extractor
Feature Extraction and Pruning
Random Forest Feature Discardment
eXtreme Gradient Boosting Feature Discardment
Light GBM Feature Discardment
Random Forest
XGBoost
LightGBM
Computation and Concatenation of Classifier Probabilities
5 layer Multilayer Perceptron
Final Class Probabilities

Figure 4.1 Proposed Methodology

4.1 Dataset Description

The dataset, its attributes and the distribution of classes are explored in detail in Chapter 5. However, it is necessary to briefly explain the type of data from which the features have been extracted.

For each object, we have an observed value of the flux (brightness) along with an error value against a timestamp. Furthermore, there is also a binary label “detected” that indicates whether a certain flux value is at least bigger than ±3σ compared to the rest of the signal.

A collection of these flux values forms a light curve. Each object in the dataset has been observed in six different bands so each object has six corresponding light curves of varying lengths. Due to the design of LSST, the light curve in a passband have certain gaps in them as the LSST is observing that object in another passband at that particular time or is aimed at another patch of sky altogether. There is another important value associated with the object called Photometric Redshift provided with the dataset. These values are 0 for galactic objects i.e. found in our own Milky Way galaxy and has some value for extragalactic objects. Generally, the farther the object, the greater this value [11]. Figure 4.2 shows the light curve in a certain passband of a randomly selected object from the dataset.

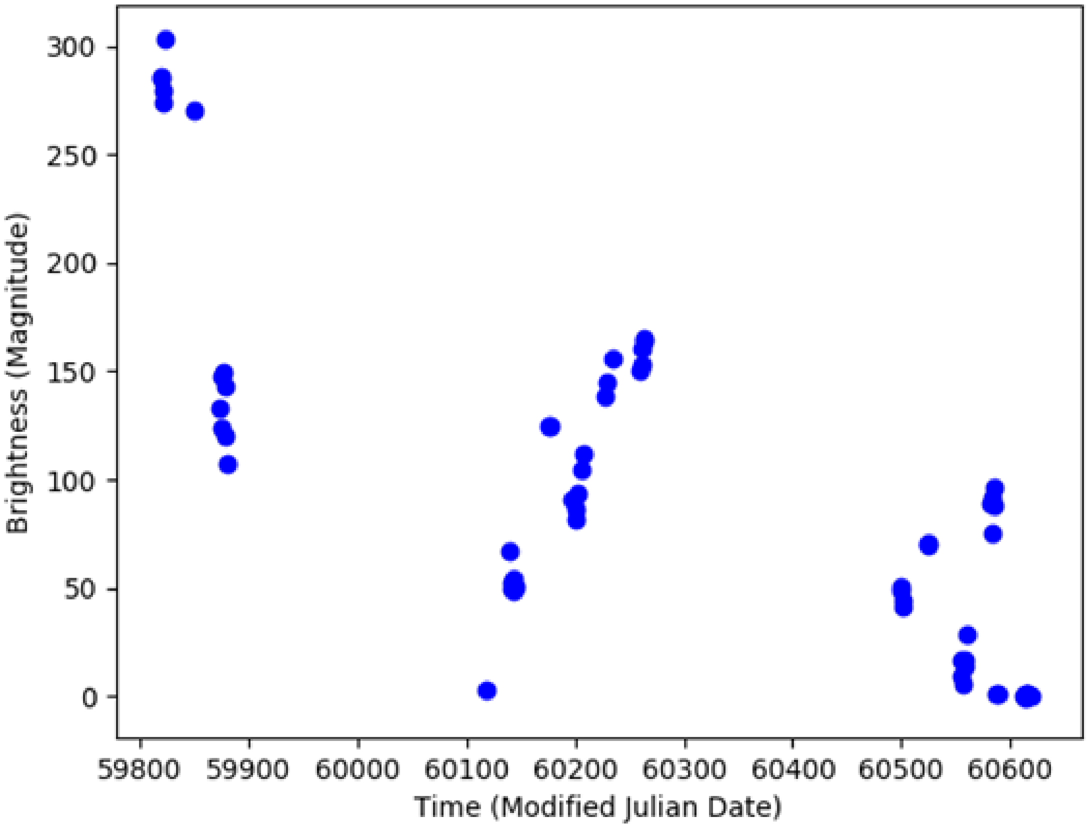


Figure 4.2 Flux Values for an Object in a Certain Band

4.2 Data Preprocessing

In order to prepare data for the feature extraction step, it is preprocessed. Preprocessing can be divided into the following steps:

4.2.1 Normalization

The first step is to remove most of the noise so that it does not adversely affect the extracted features. This is accomplished by normalizing the data by using the following equation:

$$\frac{x_i\sigma^2 + {x_{err}}^2\mu}{\sigma^2 + {x_{err}}^2} \tag{4.1}$$

where x_i is the raw flux value, σ^2 is the square of the standard deviation of the signal, ${x_{err}}^2$ is the square of error value for x_i and μ is the mean of the light curve. Normalizing in this way brings the values close to 0 closer still and consequently, the remaining values stand out more. Figure 4.3 shows the effect of this normalization on multiple light curves.

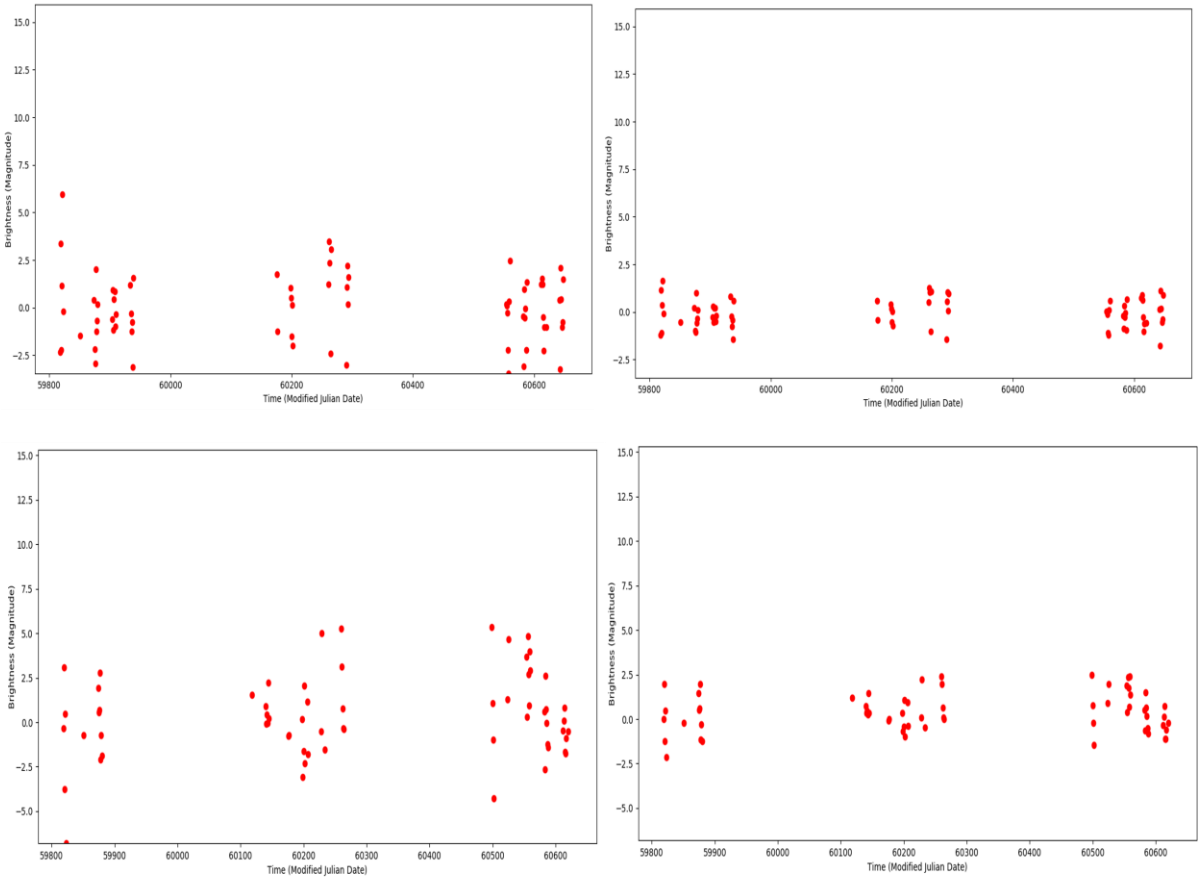
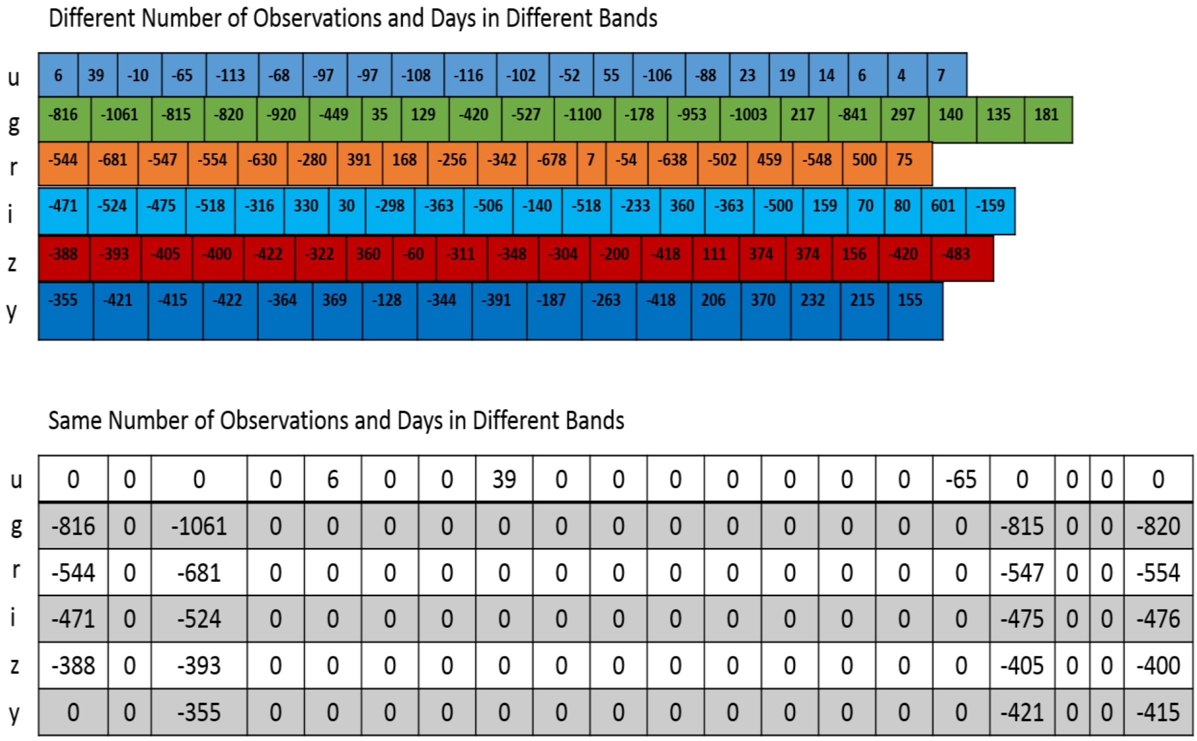


Figure 4.3 Effects of Normalization on Flux Values

4.2.2 Equating the Number of Days

Up until this point, the light curves in different channels can have varying length because of varying number of observations on different days. To remedy this, we “stretch out” the flux observations to span all the days that are present in any of the bands such that the resultant light curve (in each band) has normalized flux values with intermittent zeros. These zeros represent the days when an observation was not made for a certain passband. This step ensures that each passband has same length after this process starting from the minimum MJD to the maximum MJD. Figure 4.4 depicts the changes in size of the light curves after this process.



Different Number of Observations and Days in Different Bands

u	6	39	-10	-65	-113	-68	-97	-97	-108	-116	-102	-52	55	-106	-88	23	19	14	6	4	7
g	-816	-1061	-815	-820	-920	-449	35	129	-420	-527	-1100	-178	-953	-1003	217	-841	297	140	135	181	
r	-544	-681	-547	-554	-630	-280	391	168	-256	-342	-678	7	-54	-638	-502	459	-548	500	75		
i	-471	-524	-475	-518	-316	330	30	-298	-363	-506	-140	-518	-233	360	-363	-500	159	70	80	601	-159
z	-388	-393	-405	-400	-422	-322	360	-60	-311	-348	-304	-200	-418	111	374	374	156	-420	-483		
y	-355	-421	-415	-422	-364	369	-128	-344	-391	-187	-263	-418	206	370	232	215	155				

Same Number of Observations and Days in Different Bands

u	0	0	0	0	6	0	0	39	0	0	0	0	0	0	0	-65	0	0	0	0
g	-816	0	-1061	0	0	0	0	0	0	0	0	0	0	0	0	0	-815	0	0	-820
r	-544	0	-681	0	0	0	0	0	0	0	0	0	0	0	0	0	-547	0	0	-554
i	-471	0	-524	0	0	0	0	0	0	0	0	0	0	0	0	0	-475	0	0	-476
z	-388	0	-393	0	0	0	0	0	0	0	0	0	0	0	0	0	-405	0	0	-400
y	0	0	-355	0	0	0	0	0	0	0	0	0	0	0	0	0	-421	0	0	-415

Figure 4.4 Effect of Making the Number of Days Constant over All the Bands

4.2.3 Scaling of Flux Values

The raw flux values are not used as is, instead they are scaled by multiplying each flux value with the square of photometric redshift value of that object if the object is extragalactic. If the object is galactic (Photometric Redshift Value of 0) then the flux values are not changed.

The benefit of scaling the values in this way is that the extragalactic values and thus objects are easy to distinguish from the galactic values due to the scaling factor. Furthermore, as the scaling is non-linear, the effects of it can also be seen within the extragalactic classes.

4.3 Feature Extraction

While examining the light curves of the astronomical objects manually can shed some light on the type of the object, doing so with large amounts of data is just not feasible. Therefore, a summary of the curve can be created by using a set of features extracted from the curve itself. As objects of different classes exhibit different behavior, therefore, the values of different features can be used to differentiate objects of one class from another. This feature vector can then be fed to a classifier (or more) for training and testing.

We start by extracting a number of features from each band that can be grouped together in five distinct categories namely statistical features, features obtained from wavelet decomposition, shape characteristics of the signal, ratios of certain attributes of the signal and miscellaneous features. In addition to computing the feature for each of the bands, some features are also computed for an object in two ways; one, by adding up all the values of the passbands to create an object level signal or "profile", and, two, by concatenating each passband one after another to make a similar object level "profile".

Once a superset of the features has been created, the features are then pruned based on the importance of a feature for classification for a particular classifier. This results is three different subsets of features for the three classifiers.

4.3.1 Statistical Features

The following statistical features have been extracted from each of the bands and from the two overall profiles of the object:

- **Mean**: The mean value of the signal.
- **Weighted Average**: Weighted average of the signal. The weights are error values provided against each flux observation.
- **Standard Deviation**: Standard deviation of the signal.
- **Percent Beyond 1 STD**: The percentage of values that lie beyond 1 STD from the weighted average.
- **Skewness**: The measure of how much the signal is different from a normal distribution
- **Kurtosis**: The measure of the sharpness of peak of the signal.
- **Maximum Value**: Maximum value of the signal.
- **Minimum Value**: Minimum value of the signal.
- **Median**: Median value of the signal. It should be noted that in case of added object profile, the median is calculated by ignoring zeros.

- **Median Absolute Deviation**: Median of absolute differences of values of the signal from the median of the signal. This feature is not calculated for the two profiles of the object.
- **First Quartile:** The median of the values between the minimum value of the signal and the median of the signal. It should be noted that in case of added object profile, the first quartile is calculated by ignoring zeros.
- **Third Quartile:** The median of the values between the median of the signal and the maximum value of the signal. It should be noted that in case of added object profile, the third quartile is calculated by ignoring zeros.

4.3.2 Features Obtained from Wavelet Decomposition

Wavelet decomposition has been performed to decompose each light curve into 3 levels as a means of calculating energy of the decomposed components. DB1 wavelet has been used for the decomposition. The decomposition has been performed not only for the bands of each object but also for the complete object profile (both added and concatenated). Energy of the decomposed signal components is calculated as follows:

$$E = \sum_{i=0}^{n} {x_i}^2 \tag{4.2}$$

where n is the number of samples in the decomposed component**.** The energy of the resulting decomposed components have formed 4 features. Wavelet decomposition has been performed to observe the energy of the signal with varying frequencies. The four features that have been computed are as follows:

- Energy of Wavelet Decomposition (Detail Level 1)
- Energy of Wavelet Decomposition (Detail Level 2)
- Energy of Wavelet Decomposition (Detail Level 3)
- Energy of Wavelet Decomposition (Approximation Level 3)

The energy of the decomposed components is quite useful as astronomical entities with higher energy in a particular passband can be distinguished from other astronomical entities with comparatively lower energy in another passband. Wavelet decomposition of the signal before the computation of energy allows to observe whether the higher frequency bands have more energy or whether the energy lies in the lower range of the energy spectrum. Furthermore, the energy makeup of the overall object may be different from the energy makeup of a single passband.

4.3.3 Shape Based Features

The light curve of an object of one class differ visually from the object of another class and the astronomers have relied on these differences for manual vetting for years. Figure 4.5 shows how the typical light curves of objects from different classes can differ from one another.

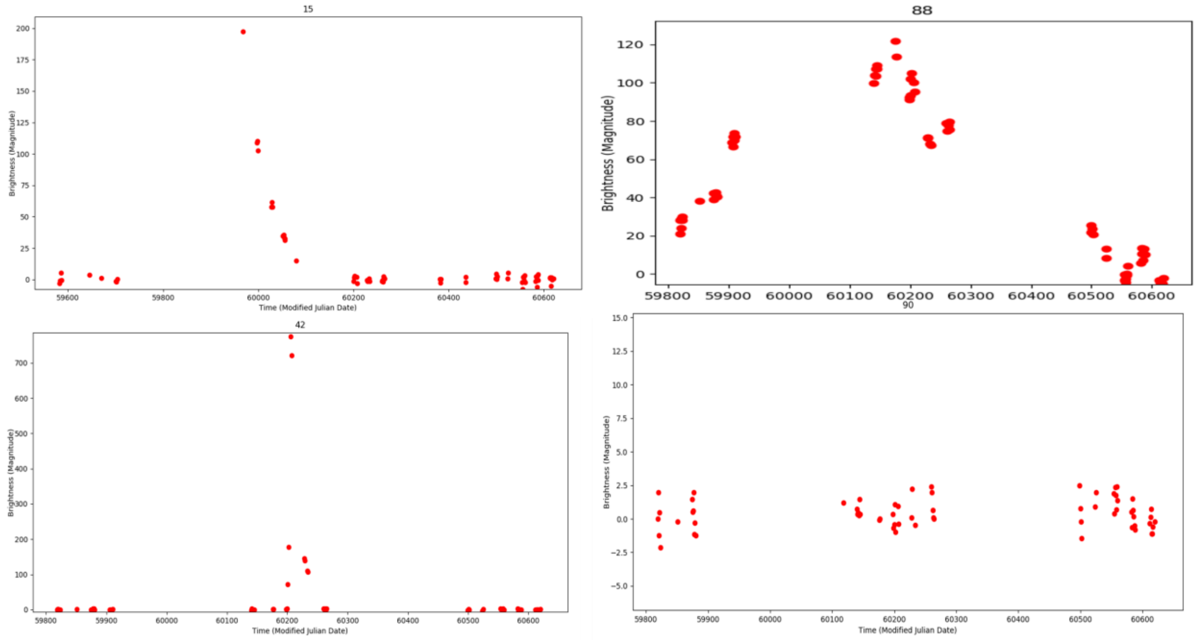


Figure 4.5 Light Curves of Objects from Different Classes

These differences in the shapes of the light curves have be extracted for better identification of the objects. As for other categories, shape based features have also been calculated for both the six bands and overall profiles of the object. 46 such features have been used.

- **Maximum Slope:** The largest, absolute rate of change in the values of the signal. Figure 4.6 shows how the maximum slope can differ for different types of objects. This feature is not calculated for the two profiles of the object.

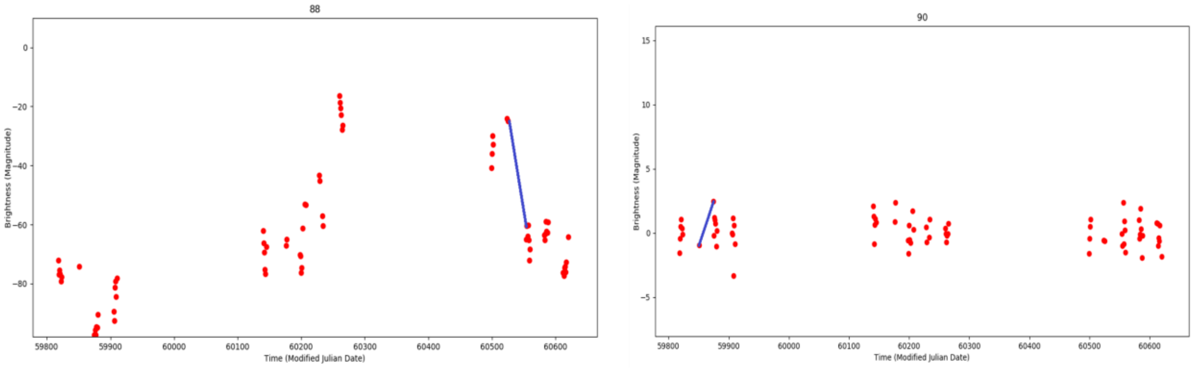


Figure 4.6 Different Slope Values for Different Classes

- **Amplitude:** Half of the difference of the maximum and the minimum value of the signal. This feature is not calculated for the two profiles of the object.
- **Average Rate of Change:** The average of rate of change (difference between two consecutive values) for the entire signal. This feature is not calculated for the two profiles of the object.
- **Peaks above Average:** Count of peaks in the signal that are above the weighted average of the signal.
- **Average Distance between Peaks above Average:** Average of the distances between the middle of the peaks that lie above average in the signal. This feature is not calculated for the two profiles of the object.
- **Average Peak Width:** Average of the width of the peaks detected that lie above average of the signal.

Figure 4.7 shows how the peaks above average, the average distance between them and their average width is calculated for the light curves.

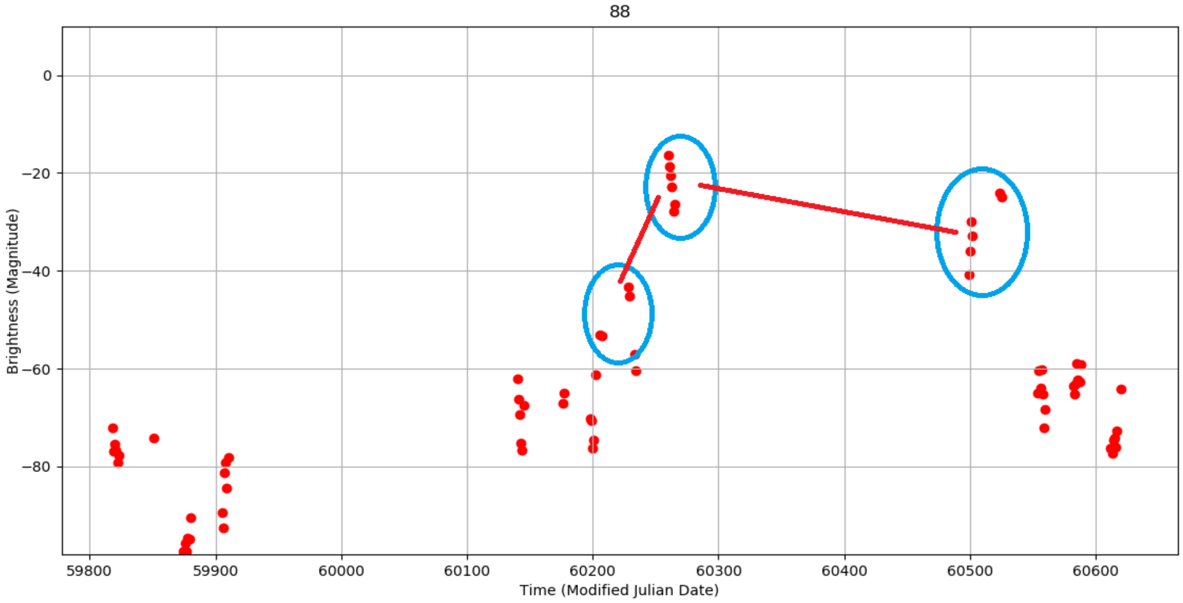


Figure 4.7 The Line Denote the Distance between the Encircled Peaks above Average in the Signal

- **Maximum Peak Prominence:** Maximum of all the values of peak prominence in the signal. Peak prominence is the measure of how much a peak stands out from its surroundings. The point of comparison is the midpoint of a horizontal line from the peak in consideration where that line crosses the signal [47].
- **Minimum Peak Prominence:** Minimum of all the values of peak prominence in the signal.

- **Average Time to Brighten:** Average of the times (in terms of samples) each peak above average in the signal takes to brighten. It is the time (in terms of samples) from the left intersection point where the horizontal line representing the average of the signal intersects the signal to the middle of the peak.
- **Average Time to Fade:** Average of the time (in terms of samples) each peak above average in the signal takes to fade. It is the time (in terms of samples) from the middle of the peak to the right intersection point where the horizontal line representing the average of the signal intersects the signal.
- **Time from max to 75% of max:** The time taken in terms of samples for a signal to drop from its maximum value to the first instance of the values that is equal to 75 percent of the maximum value. Intermittent zeros are ignored and a shift is added to make all negative values greater and equal to zero while calculating this feature.
- **Time from max to 50% of max:** The time taken in terms of samples for a signal to drop from its maximum value to the first instance of the values that is equal to 50 percent of the maximum value. Intermittent zeros are ignored and a shift is added to make all negative values greater and equal to zero while calculating this feature.
- **Time from max to 25% of max:** The time taken in terms of samples for a signal to drop from its maximum value to the first instance of the values that is equal to 25 percent of the maximum value. Intermittent zeros are ignored and a shift is added to make all negative values greater and equal to zero while calculating this feature.
- **Modified Julian Date (MJD) Difference:** Difference in terms of days between flux observations having the detected flag set to 1. Medium distance between such observations shows that the object has a cyclic behavior of brightening and fading while a large or small value shows that the brightening or fading happens only once during the observation period. If no two close values are found then this feature value is set to a really high value e.g. 10000. This feature is not calculated for the two profiles of the object.

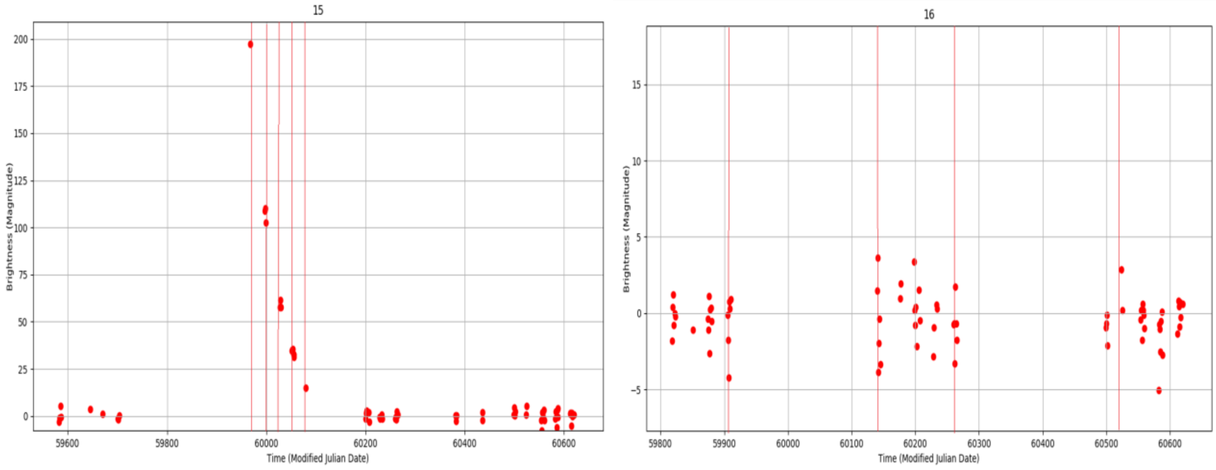


Figure 4.8 Object with Decreasing Flux Value vs. Object with Cyclic Behavior

Figure 4.8 shows the MJD Difference of a cyclic object and that of an object that has a declining or increasing light curve. The MJD Difference of the cyclic object is considerably greater.

In addition to the above features that focus on the shape of the light curve, some more features are also calculated by overlaying a grid over the signal that divides the signal along the flux axis. These features take into account how light curves of different classes exhibit different behavior in different ranges of the flux values. Figure 4.9 illustrates the effect of dividing a light curve into different ranges and then using those specific ranges for different features.

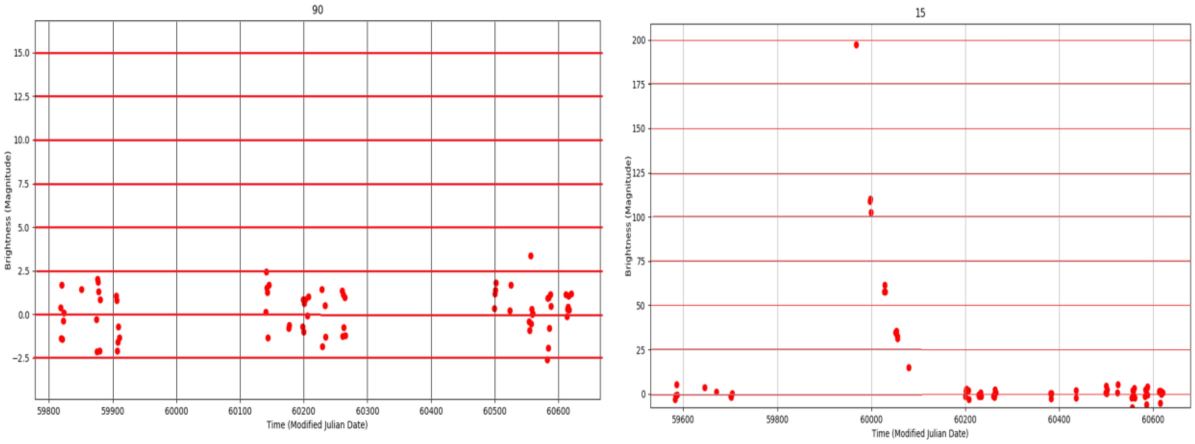


Figure 4.9 Values of Observed Flux for Two Different Samples of Different Classes from the Dataset

These features can be divided into four distinct categories: number of peaks between certain positive ranges, number of peaks between certain negative ranges, sum of the flux values in certain positive ranges and finally, sum of the flux values in certain negative ranges. For number of peaks between certain negative ranges, -1 is multiplied with the signal first to flip it around the axis and then the values are computed. It is worth noting that number of peaks in positive and negative ranges are not calculated for the two profiles of the object. Below is a list of the certain ranges that are covered by these four features mentioned here.

- **Positive Peaks Between 0 to 1:** Number of the peaks that lie in 0 to 1 range.
- **Positive Peaks Between 1 to 2:** Number of the peaks that lie in 1 to 2 range.
- **Positive Peaks Between 2 to 4:** Number of the peaks that lie in 2 to 4 range.
- **Positive Peaks Between 4 to 8:** Number of the peaks that lie in 4 to 8 range.
- **Positive Peaks Between 8 to 16:** Number of the peaks that lie in 8 to 16 range.
- **Positive Peaks Between 16 to 32:** Number of the peaks that lie in 16 to 32 range.
- **Positive Peaks Between 32 to 64:** Number of the peaks that lie in 32 to 64 range.
- **Positive Peaks Between 64 to 128:** Number of the peaks that lie in 64 to 128 range.
- **Positive Peaks Between 128 to 256:** Number of the peaks in 128 to 256 range.
- **Positive Peaks Beyond 256 :** Number of the peaks that lie in 256 and above range.
- **Negative Peaks Between 0 to 1:** Number of the peaks that lie in 0 to 1 range.
- **Negative Peaks Between 1 to 2:** Number of the peaks that lie in 1 to 2 range.
- **Negative Peaks Between 2 to 4:** Number of the peaks that lie in 2 to 4 range.
- **Negative Peaks Between 4 to 8:** Number of the peaks that lie in 4 to 8 range.
- **Negative Peaks Between 8 to 16:** Number of the peaks that lie in 8 to 16 range.
- **Negative Peaks Between 16 to 32:** Number of the peaks that lie in 16 to 32 range.
- **Negative Peaks Between 32 to 64:** Number of the peaks that lie in 32 to 64 range.
- **Negative Peaks Between 64 to 128:** Number of the peaks that lie in 64 to 128 range.
- **Negative Peaks Between 128 to 256:** Number of the peaks in 128 to 256 range.
- **Negative Peaks Beyond 256:** Number of the peaks that lie in 256 and beyond.
- **Sum of Flux Value Between 0 to +20:** Sum of flux values in this particular range.
- **Sum of Flux Value Between +20 to +40:** Sum of flux values in this particular range.
- **Sum of Flux Value Between +40 to +60:** Sum of flux values in this particular range.
- **Sum of Flux Value Between +60 to +80:** Sum of flux values in this particular range.
- **Sum of Flux Value Between +80 to +100:** Sum of flux values in this particular range.
- **Sum of Flux Value Between +100 and beyond:** Sum of flux values in this particular range.
- **Sum of Flux Value Between 0 to -20:** Sum of flux values in this particular range.
- **Sum of Flux Value Between -20 to -40:** Sum of flux values in this particular range.
- **Sum of Flux Value Between -40 to -60:** Sum of flux values in this particular range.
- **Sum of Flux Value Between -60 to -80:** Sum of flux values in this particular range.
- **Sum of Flux Value Between -80 to -100:** Sum of flux values in this particular range.

- **Sum of Flux Value Between -100 and beyond:** Sum of flux values in this particular range.

4.3.4 Ratio Based Features

While the features extracted from individual light curve bands can tell a lot about the category of a heavenly body, more information can be gathered by comparing a particular feature of one band to the same feature of another band. These ratios can greatly help in the classification of an object. The ratio based features that have been used are given below.

- **Standard Deviation of Band u to Band i:** The ratio of the standard deviation of band u to band i.
- **Standard Deviation of Band u to Band y:** The ratio of the standard deviation of band u to band y.
- **Standard Deviation of Band i to Band y:** The ratio of the standard deviation of band i to band y.
- **Skewness of Band z to Band y:** The ratio of the value of skewness of band z to the skewness of band y.
- **Percent Beyond 1 STD of Band r to Band i:** The ratio of the percentage of values beyond 1 STD of band r to the percentage of same values of band i.
- **Percent Beyond 1 STD of Band r to Band y:** The ratio of the percentage of values beyond 1 STD of band r to the percentage of same values of band y.
- **Percent Beyond 1 STD of Band i to Band y:** The ratio of the percentage of values beyond 1 STD of band i to the percentage of same values of band y.
- **Standard Deviation of Change in Flux to Standard Deviation of Band:** The ratio of the standard deviation of changes in flux to the standard deviation of the band. This feature value is calculated for all six passbands.
- **Average of Absolute Flux to Flux Error of Band:** Average of the absolute of the values of flux to flux error. This is calculated for all six passbands.

Using the observed flux values, the color of astronomical entities can also be calculated which sheds light on how hot or cold (relatively) the body is as the color is directly related to its temperature. Numerically, the color is just the ratio of observed flux values between two bands [48]. Equation 4.3 is used to compute different colours.

$$C = \frac{\sum w_{k\_i} x_{k\_i}}{\sum w_k} - \frac{\sum w_{j\_i} x_{j\_i}}{\sum w_j} \tag{4.3}$$

where j and k are the two passbands. Using the weighted averages of different bands, the following colour values are computed:

- **U-G Colour:** Colour calculated by using u and g band.
- **U-I Colour:** Colour calculated by using u and i band.
- **U-Z Colour:** Colour calculated by using u and z band.
- **U-Y Colour:** Colour calculated by using u and y band.
- **G-R Colour:** Colour calculated by using g and r band.
- **G-I Colour:** Colour calculated by using g and i band.
- **G-Z Colour:** Colour calculated by using g and z band.
- **G-Y Colour:** Colour calculated by using g and y band.
- **R-I Colour:** Colour calculated by using r and i band.
- **R-Z Colour:** Colour calculated by using r and z band.
- **R-Y Colour:** Colour calculated by using r and y band.
- **I-Z Colour:** Colour calculated by using i and z band.
- **I-Y Colour:** Colour calculated by using i and y band.
- **Z-Y Colour:** Colour calculated by using z and y band.

4.3.5 Miscellaneous Features

In addition to the features mentioned above, there are 6 more features that are computed or added to the feature vector for each object. These features are:

- **Maximum Detected Flux:** The maximum value of flux that has been detected in any of the six bands. For this feature, we use the flux values without the intermittent zeros i.e. the provided flux data.
- **Passband that Holds the Maximum Detected Flux:** The number of the band in which the maximum detected flux lies. This value can be one of the following: 0, 1,2,3,4 or 5.
- **Minimum Detected Flux:** The minimum value of flux that has been detected in any of the six bands. For this feature, we use the flux values without the intermittent zeros i.e. the provided flux data.
- **Passband that Holds the Minimum Detected Flux:** The number of the band in which the minimum detected flux lies. This value can be one of the following: 0, 1,2,3,4 or 5.
- **Photometric Redshift:** The value of the photometric redshift for the object.

- **Redshift Flag:** A binary value indicating whether the object is galactic or extragalactic. This flag is generated based on the value of photometric redshift.

All the features are then concatenated to form a feature vector of length 488. Table 4.1 summarizes the number of features that have been extracted using the six different bands and the over profile of the object.

Table 4.1 Breakdown of Different Categories of All Extracted Features

Type of Features	No. of Features Calculated per Band	No. of Features from All 6 Bands	No. of Features calculated from Added Object "Profile"	No. of Features calculated from Concatenated Object "Profile"	Category Total
Statistical	12	72	11	11	94
Wavelet Decomposition	4	24	4	4	32
Shape Based	46	276	22	22	322
Ratio Based	34				34
Miscellaneous	6				6
Total Features					488

4.4 Feature Pruning

In order to reduce the number of the extracted features from 488 (100%) down to 160 (~ 33%) we prune the less useful features. The effect of this strategy is two-fold: one, this not only helps with reducing the training time but also makes the testing stage a lot faster, and, two, removing weak features (with greater noise) can generally improve the performance of a classifier by reducing overfitting [49].

Pruning of features is done based on the feature importance values [50] calculated by the classifier itself that are used internally to split the training data so that it can be classified accurately. Top 160 features are selected based on the feature importance values by each classifier. Tables 4.2 lists the features used for the final model of Random Forest.

Examining the above selected features for the three classifiers reveals that while there is quite a bit of overlap in the final features of each classifier (redshift flag and photometric redshift being the prominent examples), a fair number of features are different. More importantly, even if the features are same, they have different importance value for different all three classifiers.

A feature having the same importance value for two or more classifiers is a rarity. The Venn diagram in Figure 4.10 shows the breakdown.

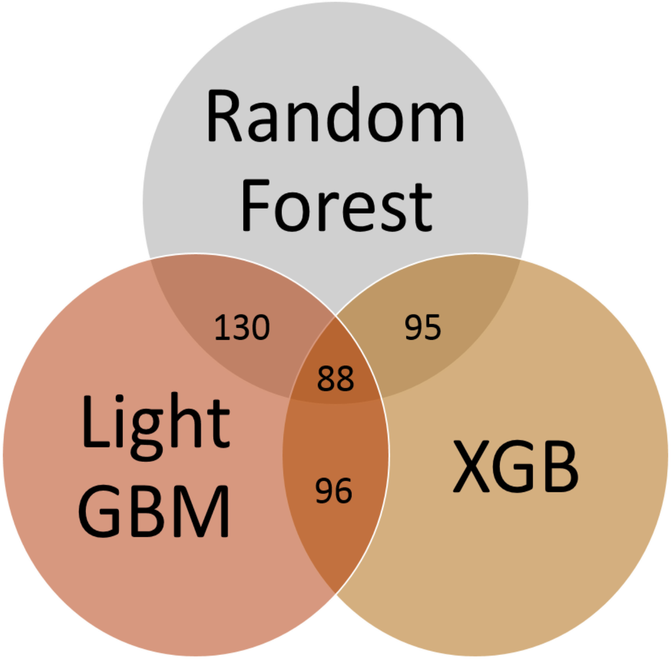


Figure 4.10 Overlap of Features between the Classifiers

4.5 Classification

4.5.1 Random Forest

Random Forests (RF) leverage the fact that combining several weak classifiers can result in a final, strong classifier. RFs work by creating an ensemble of multiple, independent decision trees by using a randomly selected subsample of the feature vector. For each tree in the ensemble, the features are different. This combination of several supplementing decision trees results in improved performance by increasing accuracy and provides better generalization [44]. The final classification is obtained by combining the output of all the classifiers using a particular method; either the outputs can be averaged (in essence, the averaging of models generated) [51] by using a certain discriminant function or voting can used with the class having the most votes being declared the output class [52]. Figure 4.11 shows the typical working of a Random Forest Classifier.

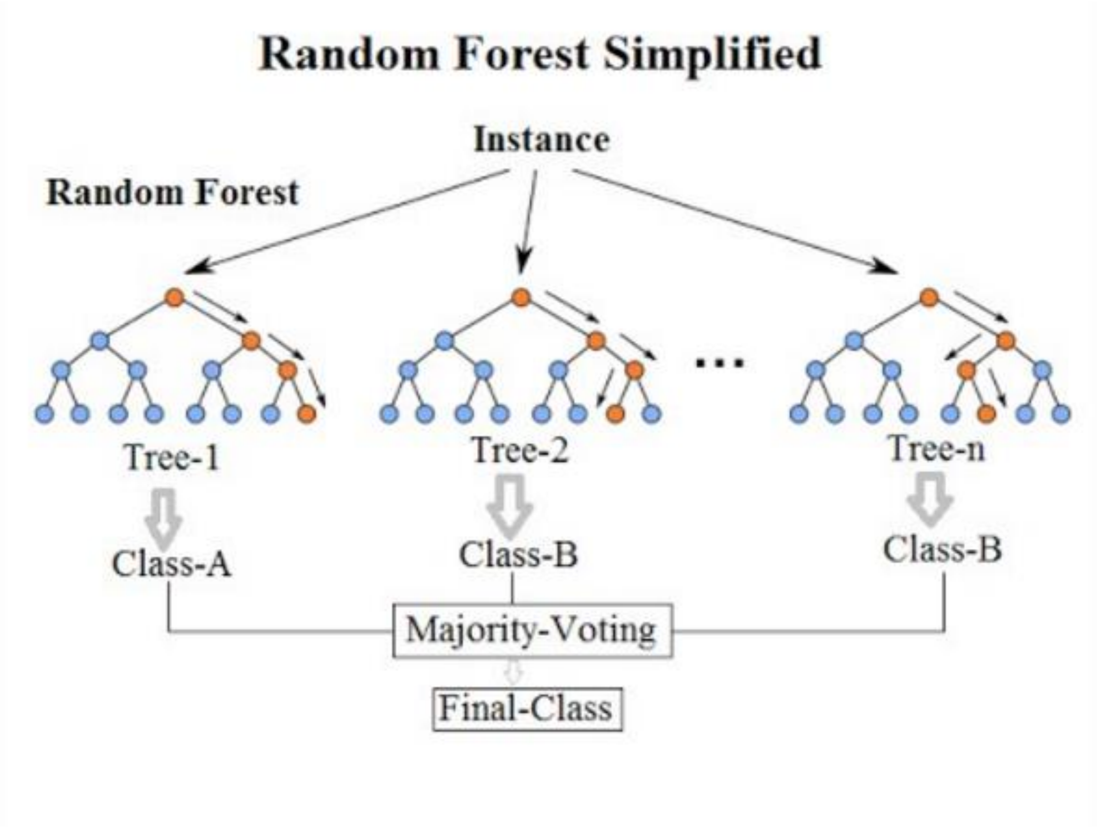


Figure 4.11 Random Forest Working [53]

Random forests can be tuned to work better for a specific problem by tuning some of the hyper-parameters. We have used a grid based searching method for hyper-parameters tuning. When the number of tree in the forest is set to 5550 with max tree depth restricted to 10 and using all features for each estimator, we get the best results by this model. In addition to the above parameters, attribute selection measure is set to Gini Matrix. The features that are most important for the RF are then used to training. The output for each sample is a 14 class probability.

4.5.2 eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGB), as the name implies, is a gradient boosting technique relying on decision trees. Boosting techniques are similar to bagging techniques as they also employ a number of weak classifiers to form a strong classifier but have the added benefit of learning from the last weak classifier's mistake. Simply put, XGB builds a strong classifier by adding on weak classifiers that optimize a learning objective function. XGB has several benefits the most prominent of which is that it is highly scalable [45].

Just like Random Forests, the hyper-parameters of XGB can also be tuned for better performance. We have used a grid-based approach to find the best hyper-parameters for this problem. The number of trees has been set to 2650 while the tree depth has been constrained to just 4. In addition, each tree in the ensemble uses only 70% of the features. In order to stop over-fitting, Alpha regularization has been used with a value of 0.01. XGB is trained on the shortlisted 160 features by their importance. XGB also produces a 14 class probability vector for each test sample.

4.5.3 Light Gradient Boosting Machine

Light Gradient Boosting Machine is another boosting technique that speeds up the process of training by making use of two novel techniques. One of them is Gradient-based One Side Sampling that dramatically reduces the number features that are used to calculate the information gain by only considering the features with large gradients. The other technique is Exclusive Feature Bundling that allows for combining mutually exclusive features. This reduces the overall number of features to consider and results in improved speed. Furthermore, the trees in LGBM are grown leaf wise instead of being grown level wise as in other bagging or boosting techniques. The tree with the most entropy is selected for this purpose [46]. Figure 4.12 demonstrates the process of a tree being grown using LGBM.

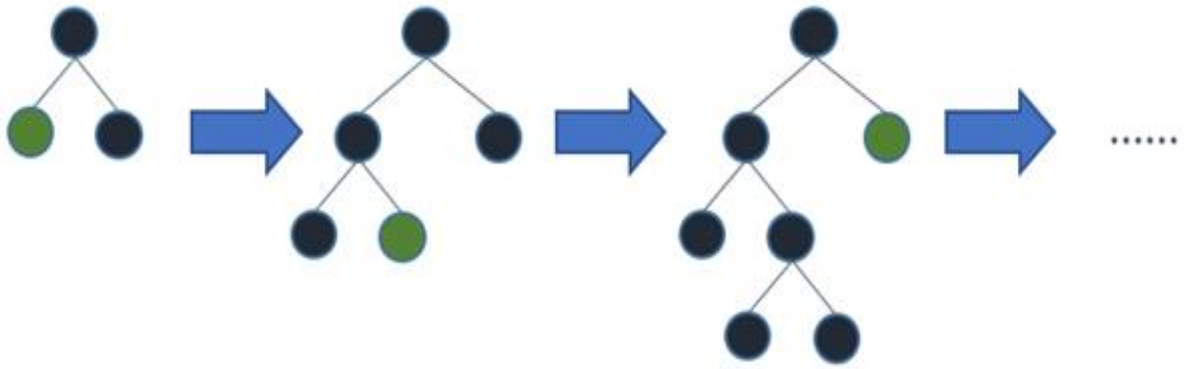


Figure 4.12 Leaf Wise Tree Growth in LGBM [54]

Using the grid-based search methods for hyper-parameters of LGBM, the number of trees is set to 50 while the maximum tree depth is capped at 5. Like XGB, the 70% of the features are used for each tree and the value of alpha regularization is set to 0.01. LGBM also produces a 14-class probability vector against each test sample.

4.5.4 5-Layer MLP

A multilayer perceptron with 5 hidden layers is used atop of the previous classifiers to combine the advantages that the different classifiers have with regard to different classes. Neural nets get their name from the biological neural nets as their architecture is quite similar consisting of computational nodes that perform non-linear operations on the input. These nodes in each consecutive layers are connected by certain weights and learning these weights is the goal of the training process through repeated passes of forward and backward propagation [55]. Figure 4.13 illustrates the structure of the MLP that has been used in our methodology.

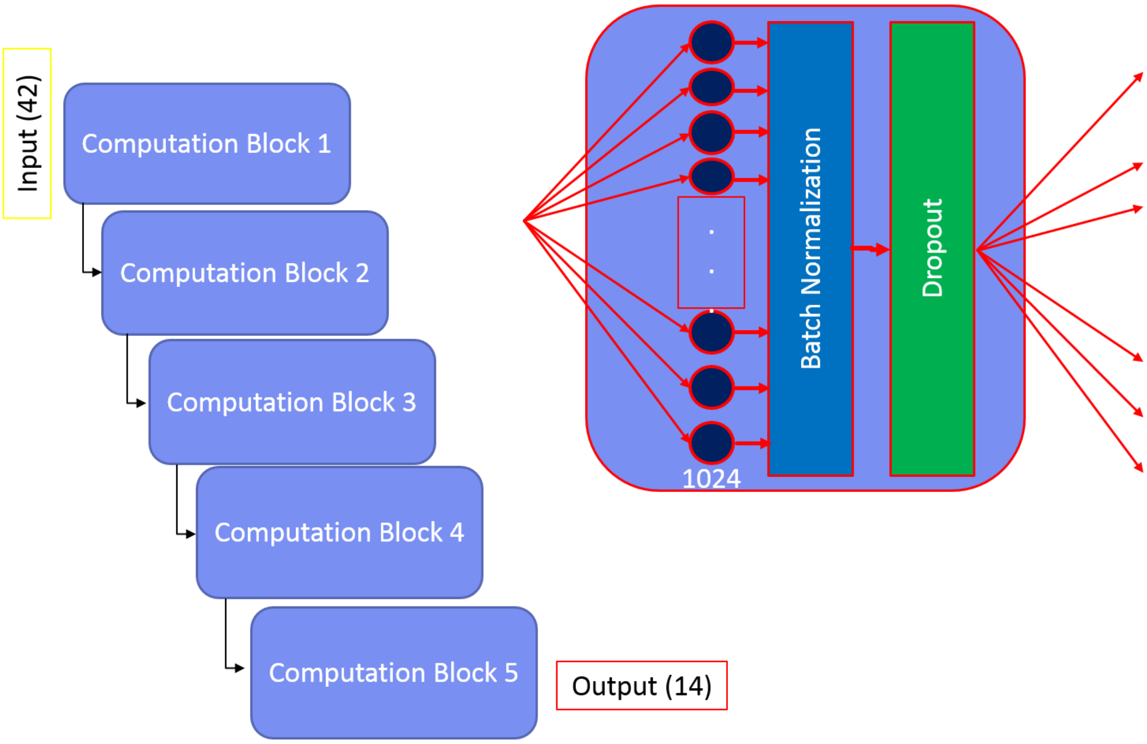


Figure 4.13 MLP Architecture

The concatenated vector of 14 class probabilities is fed to the MLP as an input. Each layer of 1024 nodes is followed by a batch normalization [56] layer which itself is followed by a dropout layer. Random dropout of 0.2 is applied at the dropout layer. Both the normalization layer and the dropout layer help to reduce overfitting. The output of the MLP is a 14-class probability vector which is treated as the final probability vector.

The MLP is trained on the probabilities of all three classifiers obtained using multiple runs at the validation stage using random shuffling in each run.

The main benefit of employing this methodology is while one classifier might perform well for one class and terribly for another, combining the classes in which all the classifiers perform well will result in an overall increased accuracy. Feature selection also allows to speed up the process of testing of incoming samples which can be an important feature if the amount of data is humongous.

Chapter 5 : EXPERIMENTAL RESULTS

5.1 Dataset

LSST dataset consists of simulated light curves of 14 different astronomical occurrences covering a wide range of objects. These light curves have been created using models presented by collaborators from different institutes all around the globe [51]. The 14 classes themselves can be divided into two main groups, namely galactic and the extragalactic objects. The Galactic objects are found in our galaxy Milky Way while extragalactic objects lie beyond. Table 5.1 contains the class labels along with the model that they represent.

Table 5.1 Dataset Classes and Models

Class Label	Astronomical Model
Class 6	μ-lens from single lens
Class 15	Tidal Disruption Event
Class 16	Eclipsing Binary stars
Class 42	Core Collapse, Type II SN
Class 52	Peculiar SNIax
Class 53	Pulsating variable stars
Class 62	Core Collapse, Type Ibc SN
Class 64	Kilonova (NS-NS merger)
Class 65	M-dwarf stellar flare
Class 67	Peculiar type Ia: 91bg
Class 88	Active Galactic Nuclei
Class 90	WD detonation, Type Ia SN
Class 92	RR lyrae
Class 95	Super-Lum. SN (magnetar)

The dataset is divided into two subsets; the training data and the testing data. The training data consists of 7848 samples while the testing data consist of 3492890 (approximately 3.5 million) samples. Both the training and the testing datasets are biased and the training dataset does not truly represent the testing dataset. In addition to the 14 classes in the testing dataset, it also contains a very small percentage of outliers i.e. samples which do not belong to any of the classes in the training dataset [11]. Figure 5.1 shows the distribution of the classes in both the training data and the testing data while Figure 5.2 shows the distribution of the samples according to the galactic and extragalactic subsets.

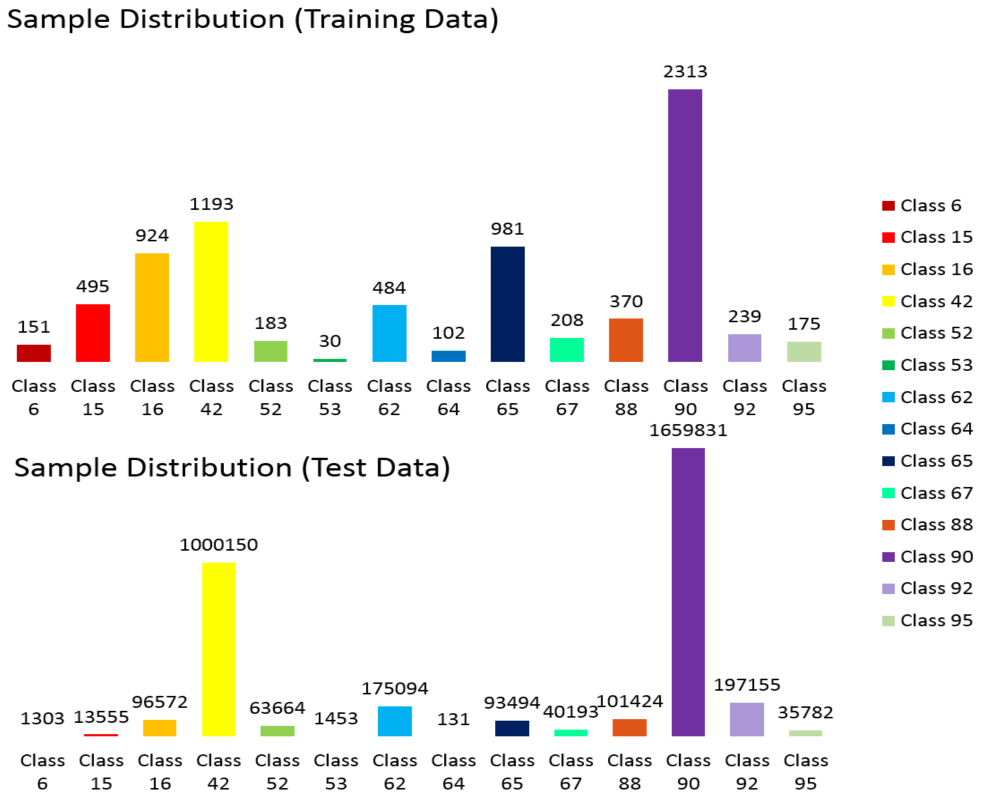


Figure 5.1 Distribution of Classes in Training and Testing Data

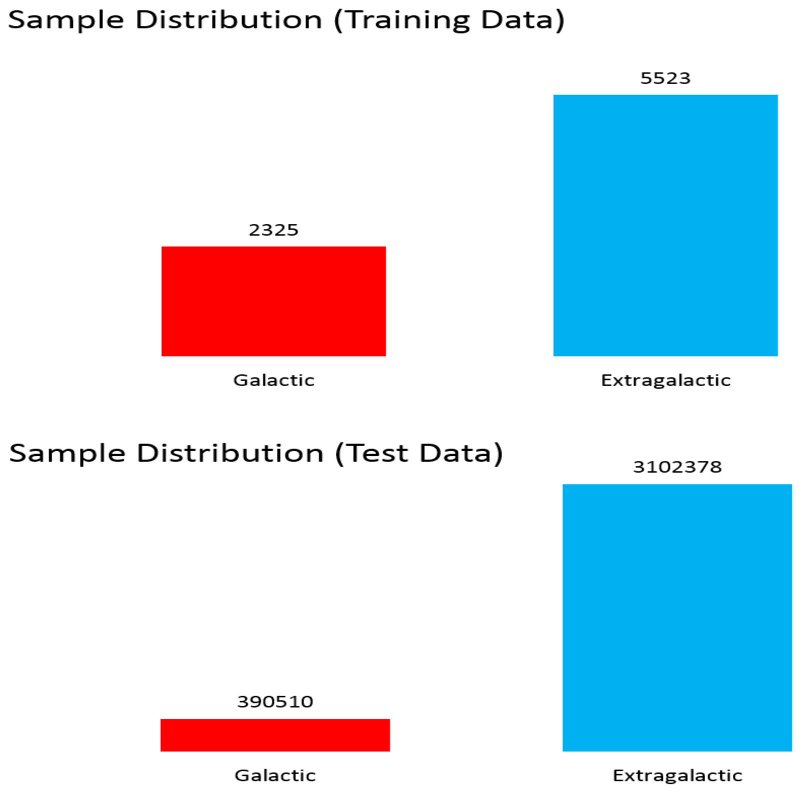


Figure 5.2 Distribution of Galactic and Extragalactic Objects

The dataset is provided with metadata for each object consisting of a number of fields the most important of which is the value of photometric redshift. In addition to the value of photometric redshift, the value of spectroscopic redshift is also given. Spectroscopic redshift is a far more accurate version of the photometric redshift. While the value of spectroscopic redshift is given for all training samples, only a small portion of testing samples have this value [11].

5.2 Performance Measures

Two metrics have been used for the evaluation of the results; the first one is the accuracy of each class while the second metric is the PLAsTiCC Metric Score [52]. It has been conceived due to the fact that it can be meaningfully interpreted for several diverse fields. The two parameters are defined as:

$$ACC = \frac{TP}{TP + FP} \tag{5.1}$$

$$L = \frac{\sum_{j=1}^{M} \left(w_j \cdot \sum_{i=1}^{N} [\![\frac{I}{N_j} \tau_{i,j} l\, n\left(P_{i,j}\right) \right)]\!]}{\sum_{j=1}^{M} w_j} \tag{5.2}$$

In case of accuracy, TP (True Positive) is the number of samples of a class that have been identified correctly while FP (False Positive) are the samples that have been classified as belonging to another class.

The PLAsTiCC Metric Score is a weighted form of cross entropy where $(i,)$ = 1, if i is from j, 0 otherwise, N_j: Number of objects in class j and w_j: weight of class j [52]. A higher weight is assigned to the class that has more samples than the rest.

5.3 Results

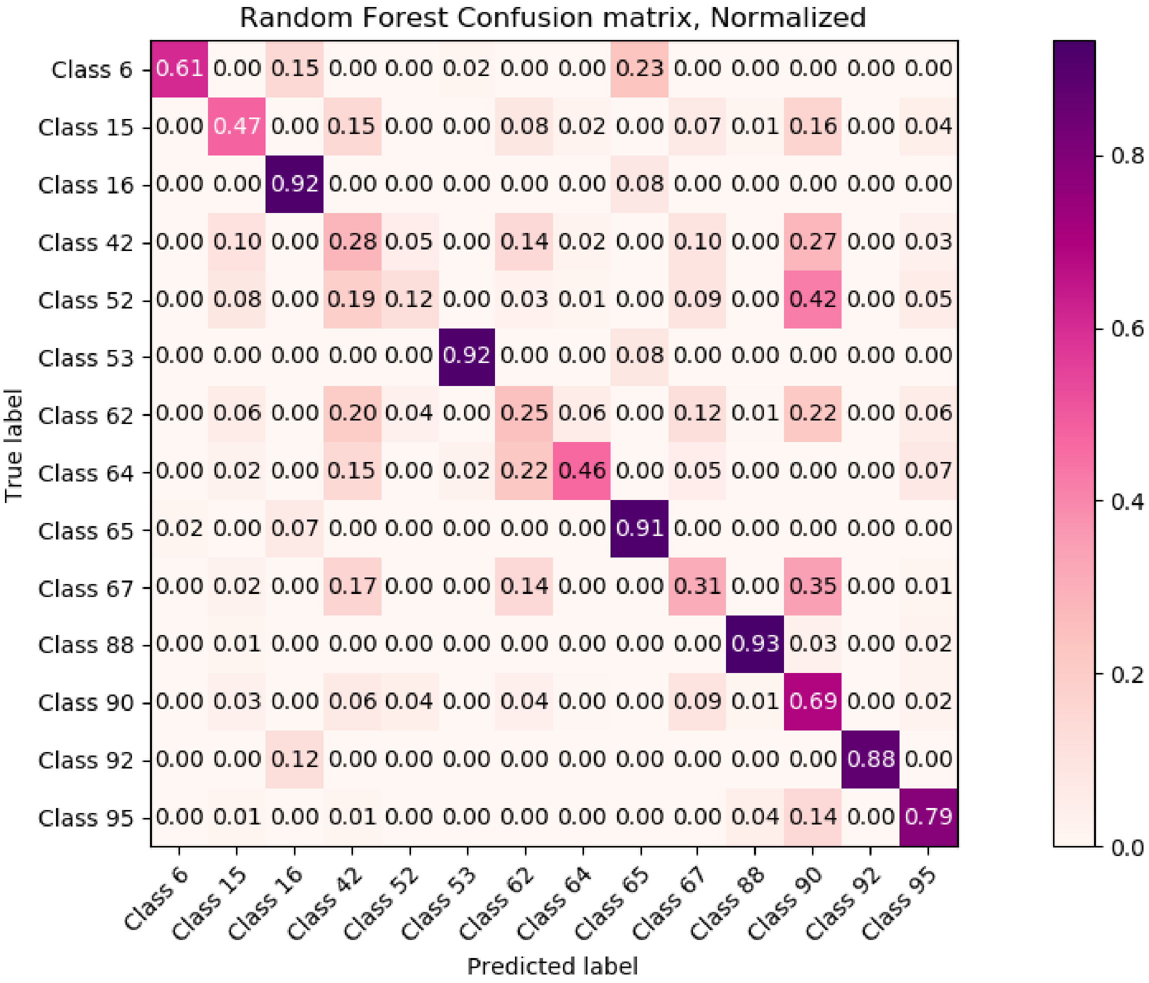
Results were computed for both the training dataset and the testing dataset and are given in the sections below.

5.3.1 Cross Validation

For Cross Validation, the training data had a 60:40 split. It is worth nothing that different classifiers performed differently for the classes in the dataset.

Random Forest (RF)

The confusion matrix of Random Forest is given in Figure 5.3. RF outperformed the other two classifiers for the following classes: Class 6, Class 15, Class 16, Class 53, Class 62, Class 64, Class 67, Class 88 and Class 95.



	Class 6	Class 15	Class 16	Class 42	Class 52	Class 53	Class 62	Class 64	Class 65	Class 67	Class 88	Class 90	Class 92	Class 95
Class 6	0.61	0.00	0.15	0.00	0.00	0.02	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.00
Class 15	0.00	0.47	0.00	0.15	0.00	0.00	0.08	0.02	0.00	0.07	0.01	0.16	0.00	0.04
Class 16	0.00	0.00	0.92	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00
Class 42	0.00	0.10	0.00	0.28	0.05	0.00	0.14	0.02	0.00	0.10	0.00	0.27	0.00	0.03
Class 52	0.00	0.08	0.00	0.19	0.12	0.00	0.03	0.01	0.00	0.09	0.00	0.42	0.00	0.05
Class 53	0.00	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00
Class 62	0.00	0.06	0.00	0.20	0.04	0.00	0.25	0.06	0.00	0.12	0.01	0.22	0.00	0.06
Class 64	0.00	0.02	0.00	0.15	0.00	0.02	0.22	0.46	0.00	0.05	0.00	0.00	0.00	0.07
Class 65	0.02	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00	0.00
Class 67	0.00	0.02	0.00	0.17	0.00	0.00	0.14	0.00	0.00	0.31	0.00	0.35	0.00	0.01
Class 88	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.03	0.00	0.02
Class 90	0.00	0.03	0.00	0.06	0.04	0.00	0.04	0.00	0.00	0.09	0.01	0.69	0.00	0.02
Class 92	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.00
Class 95	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.14	0.00	0.79

Figure 5.3 Confusion Matrix for Random Forest Classifier

The random forest achieved an accuracy of 63.33% while a log loss score of 1.2339.

eXtreme Gradient Boosting (XGB)

The confusion matrix of eXtreme Gradient Boosting is given in Figure 5.4. XGB outperformed the other two classifiers for the following classes: Class 42, Class 65 and Class 92. The XGB achieved an accuracy of 69.7% while a log loss score of 1.6194. The higher accuracy is the result of performing quite well for Class 42 and Class 65 as they contain a reasonably large number of samples in the training dataset.

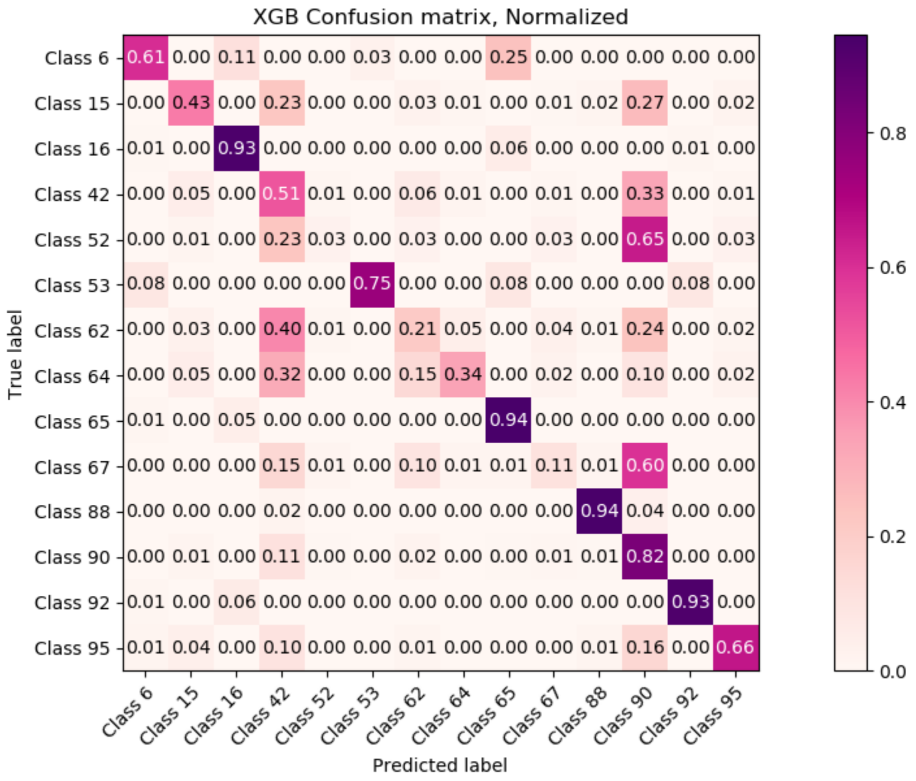


Figure 5.4 Confusion Matrix for eXtreme Gradient Boosting

Light Gradient Boosting Machine (LGBM)

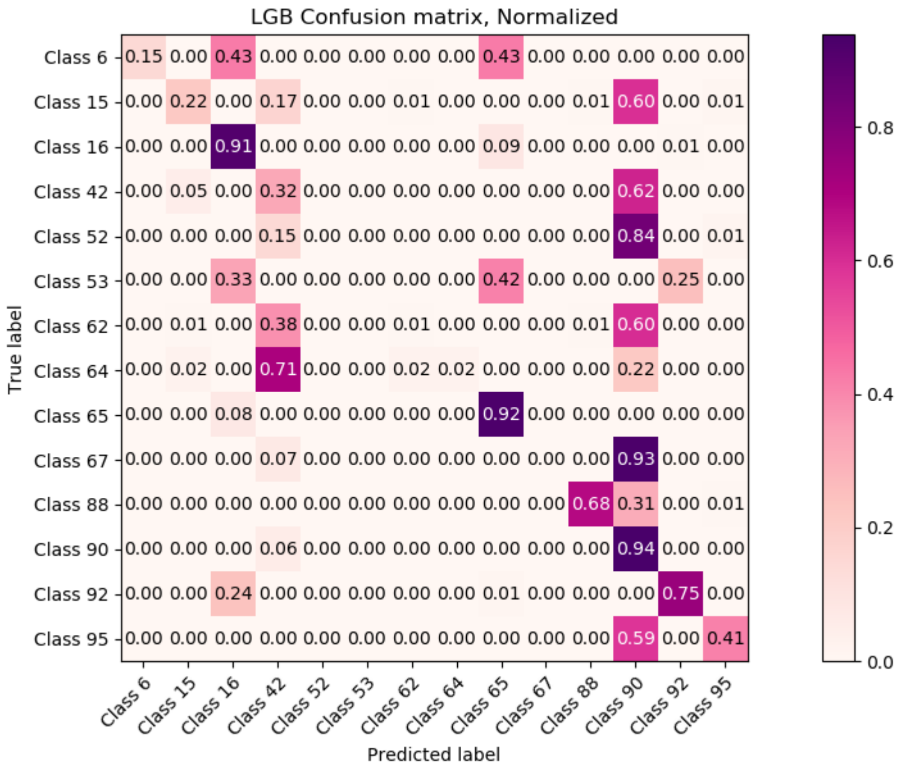


Figure 5.5 Confusion Matrix for Light Gradient Boosting Machine

The confusion matrix of Light Gradient Boosting Machine is given in Figure 5.5. LGBM outperformed the other two classifiers for only one class that is Class 90. LGBM achieved an accuracy of 62.895% while a log loss score of 2.1541. The performance of LGB was considerably worse for all other classes.

Multilayer Perceptron

The confusion matrix of the Multilayer Perceptron is given in Figure 5.6. MLP combines the output of all the classifiers in a way that the accuracy of all the classes is improved.

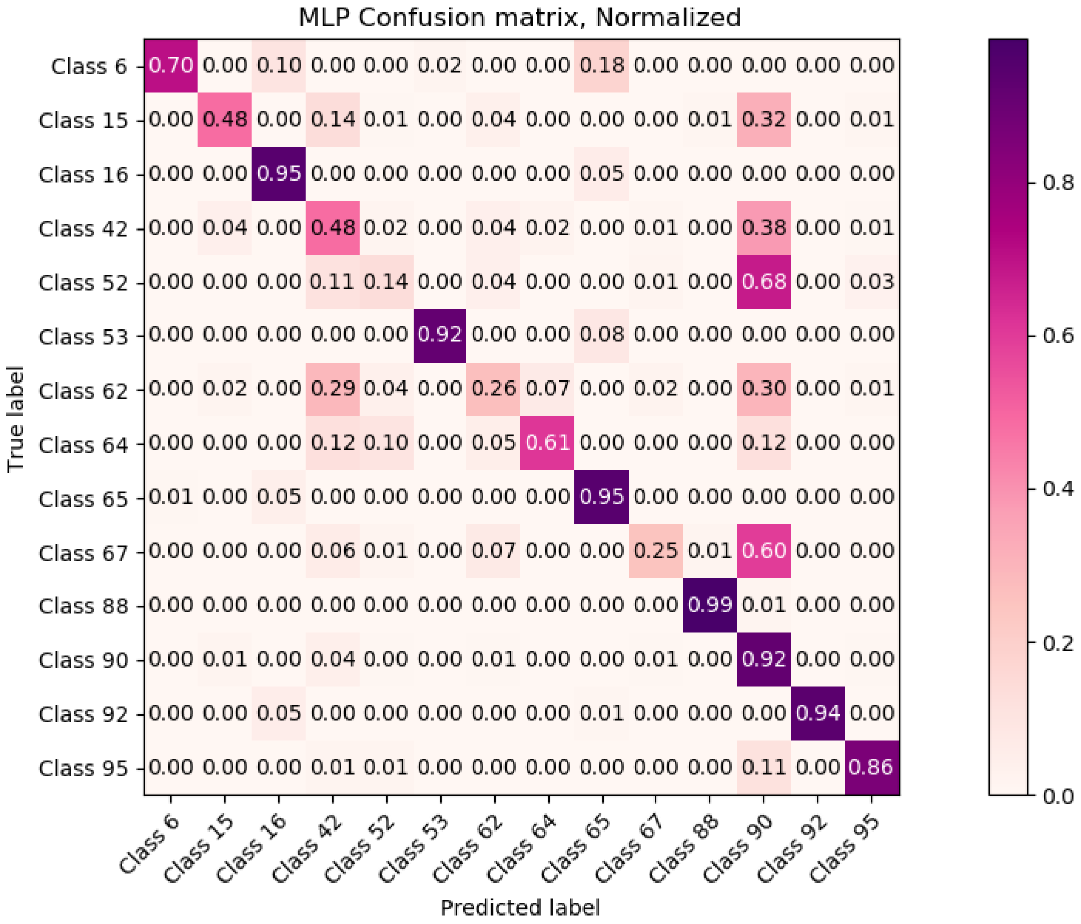


Figure 5.6 Confusion Matrix of Multilayer Perceptron

The multilayer perceptron achieved an accuracy of 74.9% while a log loss score of 0.92175. Table 5.2 summarizes all the accuracies and the log losses.

Table 5.2 Summary of Classifiers' Accuracies and Log Losses

Classifier	Accuracy	Log Loss
RF	63.33	1.2339

XGB	69.7	1.619
LGBM	62.895	2.1541
MLP	74.9	0.91275

5.3.2 Testing

As mentioned earlier, the testing dataset contains approximately 3.5 million samples. The data has been cleaned up slightly by removing the outliers present in the data. As these samples account only for a small percentage (less than 0.39%), therefore removing them does not affect the testing data very much.

Random Forest (RF)

The confusion matrix of Random Forest for testing dataset is given in Figure 5.7. The random forest achieved an accuracy of 61.792% while a log loss score of 1.455.

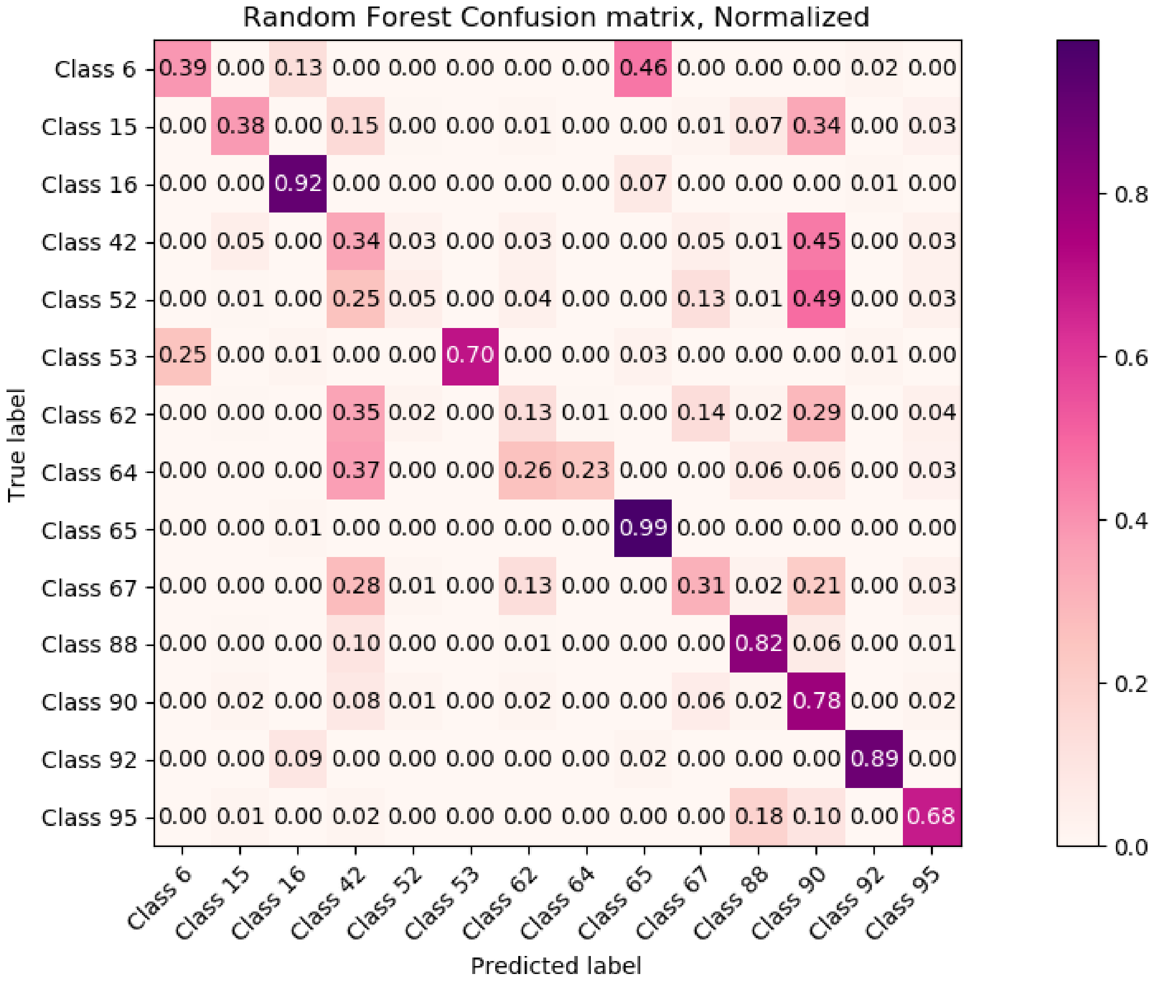


Figure 5.7 Random Forest Confusion Forest for Testing Data

eXtreme Gradient Boosting (XGB)

Figure 5.8 shows the confusion matrix of testing data. XGB achieved an accuracy of 60.63% while a log loss score of 1.65248 on the test data performing similarly for each of the classes as in cross validation.

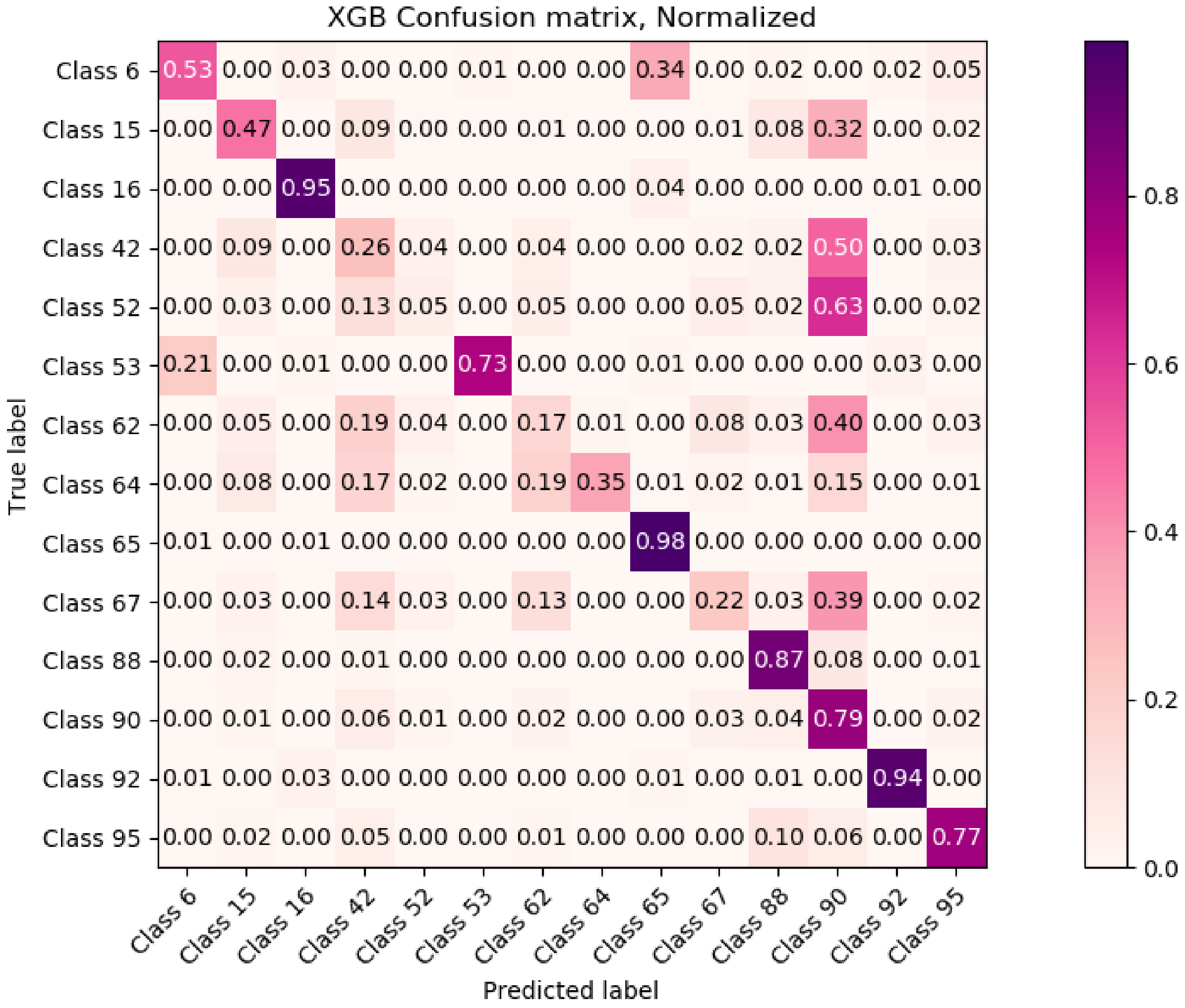
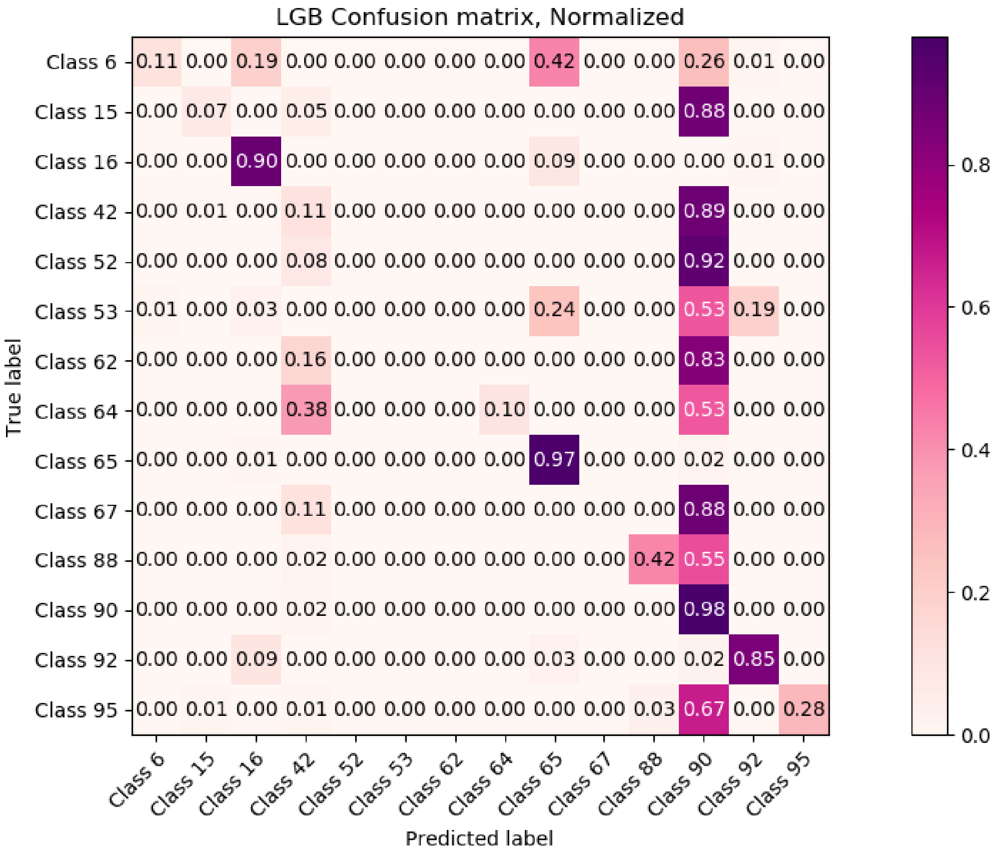


Figure 5.8 Confusion Matrix for eXtreme Gradient Boosting for Testing Data

Light Gradient Boosting Machine (LGBM)

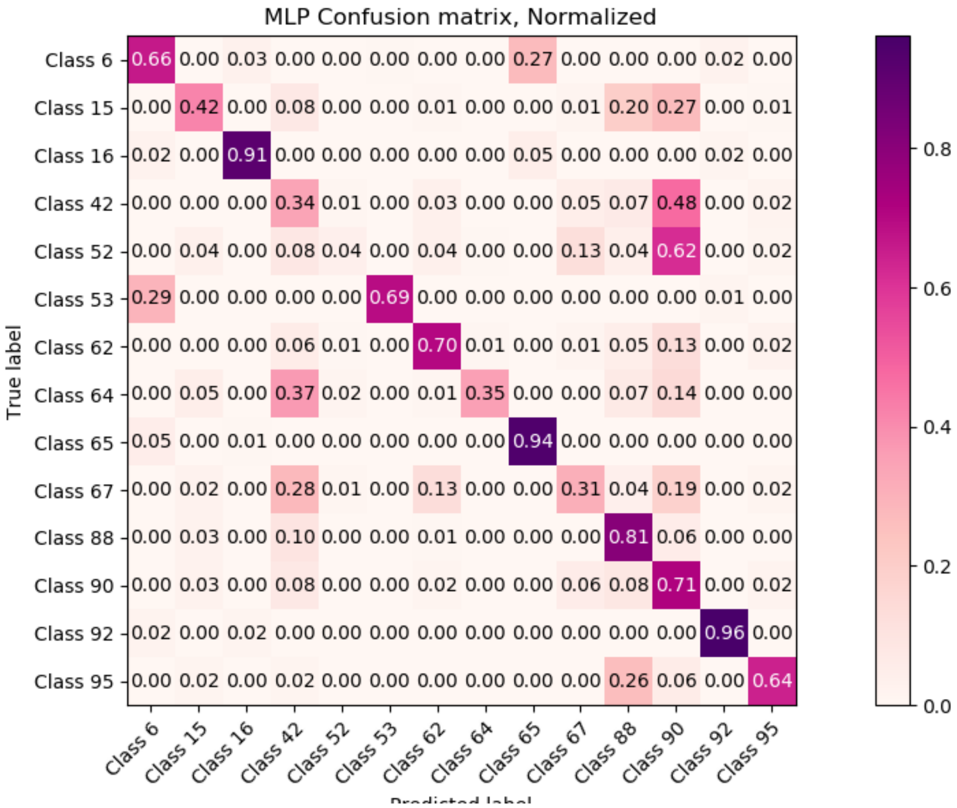
As in the cross validation, Light Gradient Boosting Machine outperformed the other two classifiers for Class 90 in testing data also. LGBM achieved an accuracy of 61.23% while a log loss score of 2.180. For all other classes, the performance of LGBM was pretty abysmal. The confusion matrix of Light Gradient Boosting Machine is given in Figure 5.9


LGB Confusion matrix, Normalized

	Class 6	Class 15	Class 16	Class 42	Class 52	Class 53	Class 62	Class 64	Class 65	Class 67	Class 88	Class 90	Class 92	Class 95
Class 6	0.11	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00	0.26	0.01	0.00
Class 15	0.00	0.07	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.00
Class 16	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.01	0.00
Class 42	0.00	0.01	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.00	0.00
Class 52	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00	0.00
Class 53	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.53	0.19	0.00
Class 62	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00
Class 64	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.53	0.00	0.00
Class 65	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.02	0.00	0.00
Class 67	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.00
Class 88	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.55	0.00	0.00
Class 90	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.00	0.00
Class 92	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.02	0.85	0.00
Class 95	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.67	0.00	0.28

Figure 5.9 Confusion Matrix of LGB using Test Data

Multilayer Perceptron


MLP Confusion matrix, Normalized

	Class 6	Class 15	Class 16	Class 42	Class 52	Class 53	Class 62	Class 64	Class 65	Class 67	Class 88	Class 90	Class 92	Class 95
Class 6	0.66	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.02	0.00
Class 15	0.00	0.42	0.00	0.08	0.00	0.00	0.01	0.00	0.00	0.01	0.20	0.27	0.00	0.01
Class 16	0.02	0.00	0.91	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.02	0.00
Class 42	0.00	0.00	0.00	0.34	0.01	0.00	0.03	0.00	0.00	0.05	0.07	0.48	0.00	0.02
Class 52	0.00	0.04	0.00	0.08	0.04	0.00	0.04	0.00	0.00	0.13	0.04	0.62	0.00	0.02
Class 53	0.29	0.00	0.00	0.00	0.00	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Class 62	0.00	0.00	0.00	0.06	0.01	0.00	0.70	0.01	0.00	0.01	0.05	0.13	0.00	0.02
Class 64	0.00	0.05	0.00	0.37	0.02	0.00	0.01	0.35	0.00	0.00	0.07	0.14	0.00	0.00
Class 65	0.05	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00
Class 67	0.00	0.02	0.00	0.28	0.01	0.00	0.13	0.00	0.00	0.31	0.04	0.19	0.00	0.02
Class 88	0.00	0.03	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00	0.81	0.06	0.00	0.00
Class 90	0.00	0.03	0.00	0.08	0.00	0.00	0.02	0.00	0.00	0.06	0.08	0.71	0.00	0.02
Class 92	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00
Class 95	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.06	0.00	0.64

Figure 5.10 Confusion Matrix of Multilayer Perceptron for Testing Data

The ensemble achieves an accuracy of 61.7009% for the test data and a log loss score of 1.43. The degradation in performance is due to the extremely large number of samples as compared to the training dataset.

Table 5.3 summarizes all the accuracies and the log losses for the 3.5 million test samples.

Table 5.3 Summary of Classifiers' Accuracies and Log Losses for Testing Data

Classifier	Accuracy	Log Loss
RF	61.792	1.455
XGB	60.63	1.65
LGBM	61.23	2.180
MLP	61.7009	1.43

Chapter 6 CONCLUSION & FUTURE WORK

6.1 Conclusion

In the age of LSST and similar telescopes, automatic prediction of astronomical entities in a timely manner is the need of the hour. Keeping that in view, the results presented here are encouraging. An aggregate of diverse features can form a clear picture of an object for classification instead of relying on just a handful of features. The benefit of using multiple features is that a subset of features that might work for one classifier might not give good results for another. In addition, using the bagging and boosting classifiers in an ensemble, the overall accuracy can be increased to some extent. The benefit of using such an ensemble is that classifiers like Random Forest and Light GBM are quite fast to train so we can cut down on the training time if we just use a deep learning technique. An ensemble also curbs overfitting.

6.2 Contribution

- Fully automated system for classification of astronomical entities using light curves from photometric data
- Engineering and extraction of useful features from the light curves to help in the classification
- Reformatting of raw flux values and saving them with relevant information in an easy to access file format

6.3 Future Work

While the ensemble improves the accuracy of the overall model, increasing the accuracy of the individual classifier can greatly benefit the ensemble. In addition, the classification at the bagging and boosting classifiers can be done in a hierarchal manner to further increase the accuracy. Moreover, light curves from other sources can also be subjected to this approach and this model can be expanded for classification of other astronomical entities.

REFERENCES

[1] A. Aaboe, "Scientific astronomy in antiquity.," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 276, no. 1257,* pp. 21-24, 1974.

[2] "NASA Spinoff Technologies," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/NASA_spinoff_technologies#Mistakenly_attributed_NASA_spinoffs. [Accessed 21 January 2019].

[3] NASA, "Hubble Space Telescope," NASA, [Online]. Available: https://www.nasa.gov/mission_pages/hubble/story/index.html. [Accessed 25 January 2019].

[4] NASA, "Kepler and K2," NASA, [Online]. Available: https://www.nasa.gov/mission_pages/kepler/main/index.html. [Accessed 25 01 2019].

[5] NRAO, "Very Large Array," [Online]. Available: https://public.nrao.edu/telescopes/vla/. [Accessed 25 01 2019].

[6] Z. Ivezi¶c, T. Axelrod, W. N. Brandt, D. L. Burke, C. F. Claver, A. Connolly, K. H. Cook and e. al., "Large Synoptic Survey Telescope: From science drivers to reference design," *Serb.Astron,* no. 176, pp. 1-13, 2008.

[7] S. M. Kahn, N. Kurita, K. Gilmore, M. Nordby, P. O'Connor, R. Schindler, J. Oliver and e. al., "Design and development of the 3.2 gigapixel camera for the Large Synoptic Survey Telescope," in *Ground-based and airborne instrumentation for astronomy III, vol. 7735, p. 77350J. International Society for Optics and Photonics*, San Diego, 2010.

[8] H. C. King, The History of the Telescope, Courier Corporation, 2003.

[9] D. G. York, J. Adelman, J. E. Anderson Jr. and e. al., "The Sloan Digital Sky Survey: Technical Summary," *The Astronomical Journal,* vol. 120, no. 3, pp. 1579-1587, 2000.

[10] D. C. Martin, J. Fanson and D. Schiminovich, "The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission," *The Astrophysical Journal Letters,* vol. 619, no. 1, pp. 1-6, 2005.

[11] A. Jr, Tarek, A. Bahmanyar, R. Biswas, M. Dai, L. Galbany, R. Hložek, E. EO Ishida and e. al., "The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Data set," arXiv preprint arXiv:1810.00001., 2018.

[12] LSST Org, "LSST Science Goals," LSST Org, [Online]. Available: https://www.lsst.org/science. [Accessed 14 December 2018].

[13] LSST Org, "Home," Association of Universities for Research in Astronomy, [Online]. Available: https://www.aura-astronomy.org/. [Accessed 15 12 2018].

[14] LSST Org, "Summit," LSST ORG, [Online]. Available: https://www.lsst.org/about/tel-site/summit. [Accessed 15 12 2018].

[15] *LSST Facility and Operations Building Cutaway View.* [Art]. LSST Org, 2018.

[16] LSST Org, "LSST Telescope & Site," LSST Org, October 2018. [Online]. Available: https://www.lsst.org/about/tel-site. [Accessed 15 December 2018].

[17] LSST Org, "LSST Optical Design," LSST Corporation, 2018. [Online]. Available: https://www.lsst.org/about/tel-site/optical_design. [Accessed 19 11 2018].

[18] *LSST Field of View Comparison.* [Art]. LSST Project Office, 2018.

[19] G. Narayan, "The PLAsTiCC Astronomy "Starter Kit"," 31 08 2018. [Online]. Available: https://www.kaggle.com/michaelapers/the-plasticc-astronomy-starter-kit. [Accessed 02 09 2018].

[20] LSST Org, "Camera," LSST Org, 2018. [Online]. Available: https://www.lsst.org/about/camera. [Accessed 19 11 2018].

[21] *Camera Layout Full.* [Art]. LSST/DOE, 2018.

[22] LSST Org, "Detector Design - Rafts and Towers," LSST Org, September 2018. [Online]. Available: https://www.lsst.org/about/camera/rafttower. [Accessed 16 December 2018].

[23] *Front End Electronics.* [Art]. LSST/DOE, 2008.

[24] T. Budavari, A. J. Connolly, A. S. Szalay, I. Szapudi, I. Csabai, R. Scranton, N. A. Bahcall and e. al, "Angular Clustering with Photometric Redshifts in the Sloan Digital Sky Survey: Bimodality in the Clustering Properties of Galaxies," *The Astrophysical Journal,* vol. 595, no. 1, pp. 59-70, 2003.

[25] G. T. Richards, X. Fan, H. J. Newberg, M. A. Strauss, D. E. Vanden Berk, D. P. Schneider, B. Yanny and e. al, "American Astronomical Society logo," *The Astronomical Journal,* vol. 123, no. 6, pp. 2945-2975, 2002.

[26] L. Corporation, Artist, *FPA.* [Art]. LSST Org, 2018.

[27] LSST Org, "Technology Innovation," [Online]. Available: https://www.lsst.org/about/dm/technology. [Accessed 25 December 2018].

[28] LSST Org, "Data Management," LSST Org, [Online]. Available: https://www.lsst.org/about/dm. [Accessed 12 December 2018].

[29] LSST, Artist, *Data Mining Sphere.* [Art]. LSST Org, 2018.

[30] J. W. Richards, D. L. Starr, N. R. Butler, J. S. Bloom, J. M. Brewer, A. Crellin-Quick, J. Higgins, R. Kennedy and R. Maxime, "On machine-learned classification of variable stars with sparse and noisy time-series data.," *he Astrophysical Journal,* vol. 733, no. 1, p. 10, 2011.

[31] J. M. Jenkins, S. D. McCauliff, J. Catanzarite, J. D. Twicken, C. J. Burke, J. Campbell and S. Seader, *Likely Planet Candidates Identified by Machine Learning Applied to Four Years of Kepler Data,* In American Astronomical Society Meeting Abstracts# 223, vol. 223. 2014, 2014.

[32] J. Coughlin, F. Mullally, S. E. Thompson and K. Team, *Humans Need Not Apply: Robotization of Kepler Planet Candidate Vetting.,* In American Astronomical Society Meeting Abstracts# 225, vol. 225., 2015.

[33] T. Susan E., F. Mullally, J. Coughlin, J. L. Christiansen, C. E. Henze, M. R. Haas and C. J. Burke, "A MACHINE LEARNING TECHNIQUE TO IDENTIFY TRANSIT SHAPED SIGNALS," *The Astrophysical Journal,* vol. 812, no. 1, p. 46, 2015.

[34] F. Mullally, J. L. Coughlin, S. E. Thompson, J. Christiansen, C. Burke, B. D. Clarke and M. R. Haas, "Identifying False Alarms in the Kepler Planet Candidate Catalog," *Publications of the Astronomical Society of the Pacific ,* vol. 128, no. 965, 2016.

[35] M. Lochner, J. D. McEwen, H. V. Peiris, O. Lahav and M. K. Winter, "Photometric supernova classification with machine learning," *The Astrophysical Journal Supplement Series,* vol. 225, no. 2, p. 31, 2016.

[36] P. Hartley, R. Flamary, N. Jackson, A. S. Tagore and R. B. Metcalf, "Support vector machine classification of strong gravitational lenses," *Monthly Notices of the Royal Astronomical Society,* vol. 471, no. 3, pp. 3378-3397, 2017.

[37] D. Mislis, S. Pyrzas and K. A. Alsubai, "TSARDI: a Machine Learning data rejection algorithm for transiting exoplanet light curves," *Monthly Notices of the Royal Astronomical Society,* vol. 481, no. 2, pp. 1624-1630, 2018.

[38] V. Belokurov, N. W. Evans and Y. L. Du, "Light-curve classification in massive variability surveys—I. Microlensing," *Monthly Notices of the Royal Astronomical Society,* vol. 341, no. 4, pp. 1373-1384, 2003.

[39] V. Belokurov, N. W. Evans and Y. L. Du, "Light-curve classification in massive variability surveys–II. Transients towards the Large Magellanic Cloud," *Monthly Notices of the Royal Astronomical Society,* vol. 352, no. 1, pp. 233-242., 2004.

[40] K. A. Pearson, L. Palafox and C. A. Griffith, "Searching for exoplanets using artificial intelligence," *Monthly Notices of the Royal Astronomical Society,* vol. 474, no. 1, pp. 478-491, 2017.

[41] D. M. Kipping and C. Lam, "Transit clairvoyance: enhancing TESS follow-up using artificial neural networks," *Monthly Notices of the Royal Astronomical Society,* vol. 465, no. 3, pp. 3495-3505, 2016.

[42] C. J. Shallue and A. Vanderburg, "Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90," *The Astronomical Journal,* vol. 155, no. 2, p. 94, 2018.

[43] A. Moss, *Improved Photometric Classification of Supernovae using Deep Learning,* arXiv preprint, 2018.

[44] T. K. Ho, "Random decision forests," in *3rd international conference on document analysis and recognition*, 1995.

[45] T. Chen and G. Carlos, "Xgboost: A scalable tree boosting system," in *22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.

[46] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *In Advances in Neural Information Processing Systems*, 2017.

[47] MathWorks, "Prominence," MathWorks, 2018. [Online]. Available: https://www.mathworks.com/help/signal/ug/prominence.html. [Accessed 16 December 2018].

[48] Wikipedia, "Color Index," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Color_index. [Accessed 24 December 2018].

[49] yellowbrick, "Feature Importances," 2016. [Online]. Available: https://www.scikit-yb.org/en/latest/api/features/importances.html. [Accessed 14 Januray 2019].

[50] scikit-learn.org, "3.2.4.3.1. sklearn.ensemble.RandomForestClassifier," Scikit-Learn, 2007. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier.feature_importances_. [Accessed 30 November 2018].

[51] J. VanderPlas, Python data science handbook: essential tools for working with data., O'Reilly Media, Inc., 2016.

[52] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001.

[53] W. Koehrsen, Artist, *Random Forest Simplified.* [Art]. Medium, 2017.

[54] M. Corporation, Artist, *Leaf Wise.* [Art]. Microsoft Corporation , 2017.

[55] R. P. Lippmann, "Anintroduction to computing with neural nets," *IEEE Assp magazine ,* vol. 4, no. 2, pp. 4-22, 1987.

[56] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,* arXiv, 2015.

[57] PLASTICC, Team; PLASTICC, Modelers;, "Unblinded Data for PLAsTiCC Classification Challenge," 21 January 2019. [Online]. Available: https://zenodo.org/record/2539456#.XJfCi1UzbIU. [Accessed 24 02 2019].

[58] A. Malz, R. Hložek, T. Allam Jr and e. al., *The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Selection of a performance metric for classification probabilities balancing diverse science goals,* 2018.

[59] T. Mason, Artist, *LSST Optical Design.* [Art]. Mason Productions Inc., 2018.