Università degli Studi di Napoli “Federico II”

Scuola Politecnica e delle Scienze di Base

Corso di Laurea in Fisica

Anno Accademico 2018/2019

Tesi Sperimentale di Laurea Magistrale

Machine learning based classification of transients in astronomical synoptic surveys

Relatori:
Prof. Giuseppe Longo
Dr. Massimo Brescia
Dr. Stefano Cavuoti

Candidato: Marco Vicedomini
Matricola: N94000305

Abstract

Astronomy has entered the big data era and Machine Learning based methods have found widespread use in a large variety of astronomical applications. The exploitation of present and future synoptic (multiband and multi-epoch) surveys, like LSST, requires an extensive use of automatic methods for data processing and data interpretation. With data volumes already in the terabyte and petabyte domain, the discrimination of time-critical information has already exceeded the capabilities of human operators and also crowds of citizen scientists cannot match the task. This thesis is focused on an analysis of several critical aspects related to the approach, based on Machine Learning and parameter space optimization, to variable and transient sky sources classification, with special care to the various types of Supernovae, one of the most important subjects of Time Domain Astronomy, due to their fundamental role in Cosmology.

The work is based on a test campaign, with incremental complexity, carried out to first classify the various astrophysical transients present in the LSST simulation catalogue (PLAsTiCC dataset) and subsequently the various types of Supernova. Another simulation catalogue (SNPhotCC) has been also explored to fine tune a specific time series classification model (LSTM), recently introduced in literature. The classification was carried out by comparing the performances among several Machine Learning algorithms (LSTM, Random Forest, Nadam, RMSProp, Adadelta), either on light curve data and their statistical parameters. The analysis of results makes in evidence some critical aspects related to the data quality and their parameter space characterization, propaedeutic to the preparation of processing machinery for the real data exploitation in the incoming decade.

Contents

Introduction

The scientific topics covered in this thesis falls within what is called TDA (Time Domain Astronomy). TDA is the study of variable sources, i.e. astronomical objects whose light changes with time. Although the taxonomy of such sources is extremely rich, there are two main kinds of objects, respectively, transients and variables. The first change their nature during the event, while the second present just a brightness variation. The study of these phenomena is fundamental to identify either the mechanisms causing light variations and the progenitors of the various classes of objects.

Since ancient times the phenomenon of Supernovae has fascinated human beings, but only recently we understood, in most cases, why and how this explosion happens. Obviously there are still many open questions, but knowledge about the type of galaxy in which every kind of Supernovae happens and at which rate they take place, could help us to better understand this phenomenon and many other properties of the Universe connected to the Supernovae. In order to understand and push ourselves further and further into the universe, ever more powerful incoming observing instruments, like LSST (Large Synoptic Survey Telescope), will be able to deliver impressive amounts of data, for which astronomers are obliged to make an intensive use of automatic analysis systems. Methods that fall under the heading Data Mining and Machine Learning have now become commonplace and indispensable to the work of scientists. But then, where the human work is still needed? For sure in terms of final analysis and validation of the results. This thesis work is therefore based on this virtuous combination, by exploiting data science methodology and models, such as Random Forest, Nadam, RMSProp, Adadelta and LSTM, to perform a deep investigation on time domain astronomy, by focusing the attention on Supernovae classification, performed on realistic sky simulations.

In the first chapter we will explore all the bounds between Supernovae and the modern Astrophysics, showing the importance of a good classification for such fundamental objects; in the second chapter we will describe all scientific aspects of time domain sources and methods used in this work. The third chapter is dedicated to introduce the two data simulations selected for experiments. In the fourth chapter we report all series of experiments performed and finally in the fifth we discuss the results. After the conclusion section, four appendices are present, dedicated, respectively, to ΦLAB related tables, confusion matrices of the experiments, model parameter setup and data distribution histograms for train and test sets used for model learning and validation.

Chapter 1

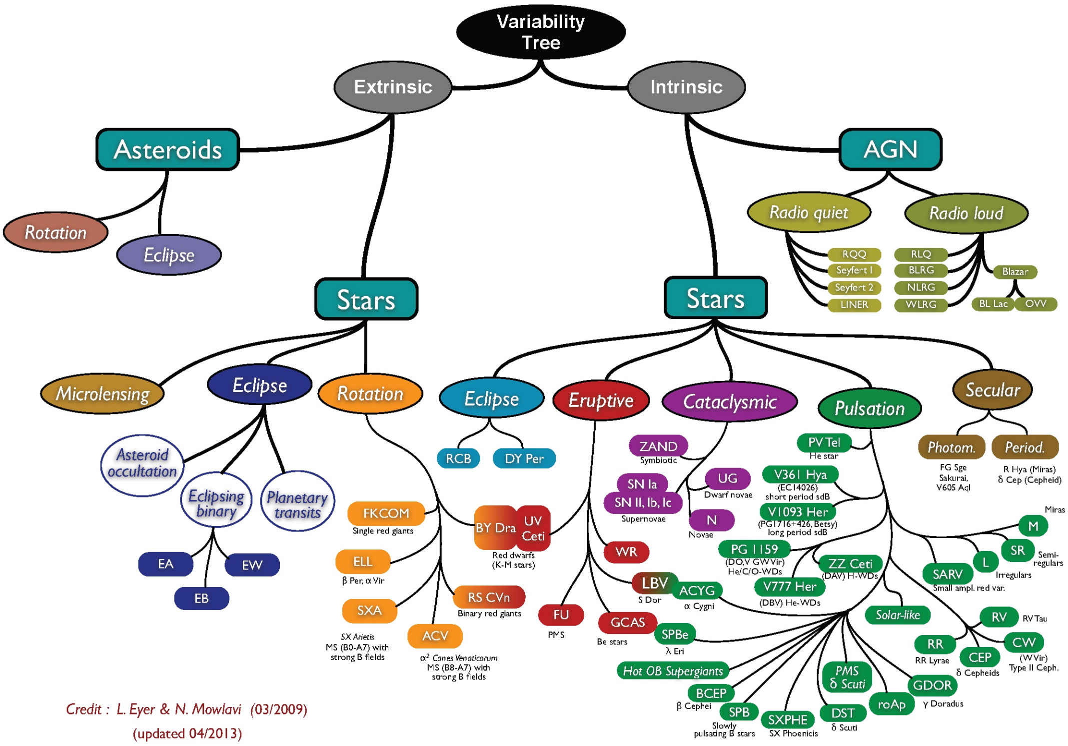
The importance of Supernovae in Astrophysics and Cosmology

Main subject of this work are transients, in particular the Supernovae (SNe).

This unique phenomenon has a great importance in Astrophysics as well as in Cosmology [3], as being referred to Neutron Stars, Black Holes, Neutrinos, gravitational waves, Interstellar Medium, early and latter Universe star formation, accelerated cosmic expansion, Hubble constant and distance measurements, gravitational lensing, discovery of dwarf galaxies and intracluster populations. In the next sections we provide a general overview on how SNe are related to them.

1.1 Neutron stars and Black Holes

Both Neutron stars and Black Holes (BHs) are the compact objects that remains after the explosion of a Supernova, in function of the progenitor stellar core mass. At the time of formation, the Neutron star mass depends also by the amount of ejecta fallback and could increase its mass if a donor star remains in the hypothetical existing binary system. The mass range of these objects is approximately



Credit : L. Eyer & N. Mowlavi (03/2009)
(updated 04/2013)

Figure 1.1: Taxonomy of Time Domain Astronomy

1.3 $\div$ 2 solar masses and with a mean radius of 10 km, their matter, following the equation of state, is almost stiff. Regarding the angular momentum, we use the parameter of Kerr or Spin, which has a value between 0 and 1, even if stars close to breakup have a value of about 0.7. Although the Neutron stars, barely visible after the explosion, rotate with low Spin (~ 0.02), it is assumed that the speed at their formation is very high, caused by the hydrodynamic processes during the explosion of the SN and that it then decreases to feed the SN itself. When SN remnants are sufficiently sparse, if the Neutron star belongs to a binary system, it could accumulate mass from the companion, and hence increase or decrease its rotation speed. Another important factor linked to rotation is the presence of a magnetic field. The greater the rotation of the star, the greater the magnetic field generated by the dynamic effect or by hydro-magnetic instabilities. Finally, even the position of Neutron star is perturbed by the possible asymmetry of the core collapsed SN, which pushes the star out of the galactic disk towards the halo. Until now, we do not know the exact mass limit between Neutron stars and Black Holes; a reasonable hypothesis is that, since the maximum mass of a neutron star is around 2 solar masses, this represents the minimum for the Black Hole. Anyway, the estimated minimum mass of a Black Hole is about 5 solar masses; so this gap may be due to an error in the mass estimate, a selection error or a flawed theory. For sure, in this mass gap we will find a better understanding of the explosion of a SN and the role played by Neutrinos. The relation between the mass of a compact object and the Zero Age Main Sequence (ZAMS) mass of the progenitor is unclear and complex, because there are many affecting factors like metallicity, stellar winds and the eventually binary evolution of the system. Moreover, all Black Holes candidates are within binaries system, thus revealing the presence of a selection effect. Furthermore, Black Holes seem formed by the mass of the inner core, or by stars that have lost their envelope becoming a BH, but this needs very massive stars with a core of more than 15 solar masses, or by stars that have lost its envelope, transferred to its companion, or by a core collapse of a red giant very massive, which has lost its envelope. Comparing BHs with Neutron stars, it is possible that the rotating BHs are very fast (Spin ≥ 0.7) and this velocity is such that no any accretion disk or companion should have sufficient strength or time to modify it, so the Black Hole should be produced in SNe with jets. In positional terms, the BHs move like Neutron stars towards the halo, but at lower speed.

1.2 Neutrinos and Gravitational Waves

During a core collapse event, Gravitational Waves and 10^{58} Neutrinos emerge from the deep layers of star, carrying with them lot of information about the core-collapse mechanism. The neutrino emission begins few seconds after the beginning of the collapse, while photons escape after hours, so the neutrinos arrive first and their detection announces the incoming photonic event. An anomalous interruption of the neutrino flow could lead to the hypothesis of a BH formation. The upper limit of the cosmic diffuse neutrino background produced by SNe is at most ten times the predicted flux and the neutrino emission should increase with the shock revival time, and with this understand a basic parameter of the core collapse mechanism.

As regards Gravitational Waves emission predictions from core collapse, this waves could be due to the breaking of spherical symmetry during the collapse and the subsequent rebound and the hydro-dynamical instabilities during the post rebound. Such instabilities, together with many theoretic explosion mechanisms could become detectable in the next future by Gravitational Waves. Furthermore, the SNe Gravitational Waves emission contributes to the Gravitational Wave background, by mixing it with the predicted contribution of standard inflationary models, and it is unclear if it will be possible to distinguish them and obtain clues about the early universe.

1.3 Interstellar Medium and star formation in early and latter universe

The information on the nature and distribution of dust in galaxies comes from studies on interstellar extinction of SNe, because the SN spectra superimposed to narrow absorption lines provide useful information on interstellar gas. SNe of type Ib-Ic have strong diffuse interstellar bands (DIB) which vary on short timescales, and seems to be associated with mass loss from the progenitor star. Therefore, more investigations about the inhomogeneities on small scales of the ISM can be performed whenever the time-varying forces of interstellar absorption lines interfere with the expanding photosphere of the SN.

In the early universe, the III stars population begins the chemical enrichment of interstellar and intergalactic medium. Today we cannot observe them at $z \geq 10$, but we hope in future to observe light of their SNe, maybe IIn type or supermassive thermonuclear explosions.

When a star reaches the end of its journey and explodes as a SN to change its nature, it emits an extremely energetic shock wave, containing most of its mass which heats and presses all the surrounding ISM. This combination helps the nucleosynthesis and the star formation and is in perfect equilibrium with cooling processes which stretch their timescales with respect the dynamical one [17]. Afterwards, the process is controlled by gas density and interaction strengths. To understand in which way the hot SN gas interacts with the cool one and how these cumulative interactions shape the galactic disk, is a fundamental step for understanding the evolution of galaxies. A dynamical approach to the ISM and gas disks try to reveal the disk-halo connection, thus many studies prove a low energy transfer efficiency, but some models show the gas compression and the vertically expulsion of the gas from the disk. It is proven that a SN feedback to the star formation rate will improve it by a factor of two, getting a better porosity of ISM, increasing the gas velocity dispersion and avoiding the formation of smaller structures. SNe and stellar wind, by sweeping and condensing the surrounding gas, can trigger star formation in a positive feedback, but this mechanism is not generally confirmed by observations. For the condensation through the fragmentation timescale, there are many phenomena influencing it, like the galactic differential rotation, the value of the sound speed and of the total energy provided by all young star formed in that region. These fragments form disks of various thickness and this is lead by the pressure in a inversely proportional way. Many numerical simulations have been performed to understand the heating efficiency of SNe, but they are too simplistic and poorly spatially resolved to obtain quantitative results. Anyway, the best star formers are SNe type II, caused by their short lifetimes and therefore have maintained their central position in the stars forming process.

1.4 Gravitational lensing, time dilation, intra-cluster populations and the discovery of Dwarf galaxies

The observed luminosity dispersion of SNe [3] is observed through inhomogeneities in the weak lensing and this is an upper limit on the cosmic matter power spectrum. Massive cosmological objects like galaxies and cluster of galaxies can magnify many times the flux of events like SNe that would be too faint to detect and bring them into our analysis picture. Studies on lensed SNe type Ia by cluster of galaxies may be used to probe the distribution of dark matter on them.

Time delay between the multiple images of lensed SNe could provide a good estimates of its high redshift. Furthermore there are two factors that makes SNe better than quasars in measuring time delay [18]: *(i)* if the Supernovae is taken before the peak, the measurements are easier and on short timescale compared to the quasar; *(ii)* the SNe light fade away with time, so we can measuring the lens stellar kinematics and the dynamics lens mass modeling.

In the next decade, the LSST will play a key role in the discovery of new lensed

SNe Ia. LSST will help to find apparently host-less SNe of every type, and this may help to study dwarf galaxies with a mass range of $10^4 \div 10^6$ solar masses. These galaxies, indeed, play a key role in large scale structure models, and despite their very big predicted population, over 1 Mpc we cannot see them, until now. Same story for theorized intracluster population of stars stripped from their galaxies, which could be seen through the SNe host-less events.

1.5 Cosmological aspects (Hubble constant, distance measurements and the accelerated cosmic expansion)

Cosmology is based on two axioms [23]: At cosmological distances the dominant interaction is gravity, and the cosmological principle is a good approximation to the universe. For cosmological principle we mean the assumption that the whole Universe, seen on a large scale, is homogeneous and isotropic. Homogeneous implies that at a given instant the Universe seems everywhere the same, while isotropic implies that locally the Universe seems the same in every direction for an observer who moves together with matter. So we consider that the metric of Universe at zeroth order is well described by the cosmological principle and every inhomogeneities are treated as perturbations of the background. From a mathematical point of view, introducing a pseudo-Riemann manifold M with a metric $g_{\mu\nu}$, we could interpret homogeneity as a one-parameter family of space-like hyper-surfaces in which the whole manifold is divided. All these slices are homogeneous. So far, at any time t, it exists a diffeomorphism of space-time that carries the point p on the point q on the same slice, leaving the metric invariant. Regarding the isotropy, it must be taken in mind that the Universe is not isotropic for all observers, so we must consider the lines of the observers. If the slices are homogeneous, the Universe line intersects them perpendicularly. Let p be one of the points where they intersect and let u^μ be the vector tangent to the Universe line at that point. Moreover v^μ is a vector of type space orthogonal to it. Isotropy then means that there exists a diffeomorphism of space-time with fixed p and u^μ that carries v_1^μ into v_2^μ and leaves the metric $g_{\mu\nu}$ invariant. A homogeneous and isotropic manifold around a point has the maximum symmetry, and a slice with this property constitutes a three-dimensional space with constant curvature. It can be shown that in such a time space the following coordinates exist:

$$x = (x^0, x^1, x^2, x^3) = (ct, \chi, \theta, \varphi) \qquad ,$$

such that the metric $g_{\mu\nu}$ takes the form of the Robertson-Walker metric whose line element is:

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu = -c^2dt^2 + R^2(t)\gamma_{ij}(\chi, \theta, \varphi)dx^i dx^j \tag{1.1}$$

with:

$$\gamma_{ij}(\chi, \theta, \varphi)dx^i dx^j = \frac{d\chi^2}{1 - K\chi^2} + \chi^2(d\theta^2 + sin^2(\theta)d\varphi^2)$$

where γ_{ij} is the three dimensional space of constant curvature, K is the constant that specifies the curvature of space, t is the cosmic time, R(t) is the cosmological radius and (χ, θ, φ) are spatial spherical co-moving coordinates. The range for these coordinates are:

$$0 \leq \chi < \begin{cases} \infty, & K = 0, -1 \\ 1, & K = 1 \end{cases} \quad , \quad 0 \leq \theta < \pi \quad , \quad 0 \leq \varphi < 2\pi$$

$K = 0$ implies flat space, $K = 1$ implies positively curved space and $K = -1$ negatively curved. Furthermore, the dimensionless scale factor *a(t)* must be entered

$$a(t) \equiv \frac{R(t)}{R(t_0)} \tag{1.2}$$

where t_0 is, generally, the present time. With the coordinate transformation $r = R(t_0)\tilde{r} = R_0\tilde{r}$, Eq. (1.1) becomes

$$ds^2 = -c^2dt^2 + a^2(t)\left[dr^2 + R_0^2 f_K^2(r/R_0)(d\theta^2 + sin^2(\theta)d\varphi^2)\right] \tag{1.3}$$

In this equation, *a(t)*, θ and φ are dimensionless, and r has the dimension of a length.

An observer in $x = 0$, and in any other point with fixed co-moving coordinates such that he sees the entire isotropic universe due to the rotation symmetry of γ_{ij}, is a co-moving observer. For this observer, the equation (1.3) become

$$ds^2 = -c^2dt^2 \tag{1.4}$$

and since the proper time of a co-moving observer is linked to the line element by

$$ds^2 = -c^2d\tau_p^2 \tag{1.5}$$

the cosmic time t in the Robertson-Walker metric for a co-moving observer is just its proper time τ_p. The proper distance $D_p(t)$, instead, between two points on the slice t is defined like the minimal distance on the slice between them. With homogeneity and isotropy, the problem can be simplified and put $x_0 = (0, 0, 0)$ and $x_1 = (r, 0, 0)$. Parameterizing $x(\lambda) = (\lambda, 0, 0)$, we have from equation (1.3)

$$D_p(t) = a(t)\int_0^r d\lambda = a(t)r \tag{1.6}$$

This tell us that r is the standard radial coordinate and that the proper distance of a co-moving observer depends from the time dependent scale factor. Instead, we can define the co-moving distance $D(t)$ that, for co-moving observers, is always constant

$$D(t) = \frac{D_p(t)}{a(t)} \tag{1.7}$$

The importance of the factor scale emerges when, after solving the Einstein field equations, we come to the Friedman's equation

$$\begin{cases} \dot{H} - H^2 = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\varepsilon + 3p) \\ H^2 = \frac{\dot{a}^2}{a^2} + \frac{kc^2}{a^2} = \frac{8\pi G}{3c^2}\varepsilon \end{cases} \tag{1.8}$$

where $\varepsilon = \rho c^2$, $H(t)$ is the Hubble parameter which, if considered at time zero, or present, is called the Hubble constant H_0; the dots on the letters imply the first or second derivative of time. In the first equation is present the acceleration which is related to both energy and the spatial part through pressure, so this equation describes the system evolution. The second one is the energy equation and represents a constraint for the system. But these two equations are not enough; a third equation for the evolution of matter, namely equation of continuity, and a constitutive equation, that is a state equation giving the relation between ρ and p, are required. Starting from the identity of Bianchi for the impulse energy tensor, using a bit of algebra we arrive at the cosmological relationship that regulates the evolution of matter

$$\dot{\varepsilon} + 3\frac{\dot{a}}{a}(\varepsilon + p) = 0 \tag{1.9}$$

For the last equation, instead, in the hypothesis that the relationship between pressure and density remains constant over time, we obtain

$$\rho(t) = \rho_0 a(t)^{-3(1+\gamma)} \tag{1.10}$$

which is the state equation used to solve the cosmological equations previously obtained, once γ and K have been fixed. Setting $\gamma = -1$ in the relation (1.10), we obtain a constant density over time, and since, by solving the Einstein equations, it results that the Universe decelerates, it is necessary to introduce a cosmological constant Λ in the Einstein equations to respect the observational data

$$G_{\mu\nu} \ \rightarrow \ G_{\mu\nu} + g_{\mu\nu}\Lambda \tag{1.11}$$

so

$$\rho_0 = \frac{\Lambda c^2}{8\pi G} \tag{1.12}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{R_0^2 a^2} + \frac{\Lambda c^2}{3} \tag{1.13}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.14}$$

and the solution of the second equation (1.8) is

$$a(t) = a(t_*)e^{H_0(t-t_*)} \tag{1.15}$$

where t_* is an arbitrary fixed time and $H_0 = H(t_*) = \sqrt{8\pi G\rho(t_*)/3} \equiv const.$ This equation shows how the scale factor of the Universe changes with time, fixed the Hubble constant.

Photons are our fundamental way of information and we must understand the phenomenon by which the expansion of the Universe modifies them. This phenomenon is called *Cosmological Redshift.* Modifying the Robertson-Walker metric we can write

$$dr = \frac{c}{H(z)}dz \quad , \quad H(z) = H_0\sqrt{\sum_f \Omega_f(1+z)^{3(1+\gamma_f)} + \Omega_K(1+z)^2} \tag{1.16}$$

where z is the cosmological redshift, Ω_f is the density parameter of the sum of baryonic matter, dark matter and dark energy and Ω_K is the curvature density parameter. This passage is true because between co-moving objects $D = r$, there's a two-way correspondence between the scale factor and the emission epoch, and between the scale factor and redshift. And all this guide us to the co-moving distance to an extragalactic source with redshift z

$$D(z) = c \int_0^z \frac{dz}{H(z)} \tag{1.17}$$

The direct validation test about such relation is coming from the analysis of SNe Ia, thanks to their strong constant luminosity. This leads to the direct observation of the Universe acceleration, allowing to set better constrain on cosmological model. Therefore, with SNe Ia data and with data of Cosmic Microwave Background (CMB) and Baryonic Acoustic Oscillation (BAO), the density parameter were set

$$\Omega_K \simeq 0 \quad , \quad \Omega_\Lambda \simeq 0.73 \quad , \quad \Omega_{DM} \simeq 0.23 \quad , \quad \Omega_B \simeq 0.04 \tag{1.18}$$

A few years ago, SNe Ia were found [27] to be weaker at the peak, faster in descent and with lower photosphere speeds. This has put the standard candles system in trouble, since the relationship between peak weakness and distance was no longer unambiguous. However, since most of these weaker SNe were found in large z [1], the Ia remained good distance indicators for relatively small z. In any case, the uncertainties on the distance measurements have been reduced thanks to the development of better photometric calibrations and to the better understanding of how these measures are biased by the selection effects [21]. We have also seen that a difference in brightness of the SNe Ia is due to the greater or smaller presence of metals [15]; the smaller its quantity, the greater the absolute brightness. An integration with the gravitational waves, created by the generation of black holes to compensate for the Ia SNe uncertainty, has been proposed. The lack of LIGO data currently limits the effectiveness of the methodology. Alternative routes have been sought and one of these uses type IIn SNe. The dense shell method [1], is a mix of the expanding photosphere method, the spectral fitting expanding atmosphere technique and the expanding shock front method. SNe type IIn are used and the propagation of a thin envelope layer is used to determine its brightness. Since the whole measurement is based on the photosphere, this makes it independent of the processes that led to the explosion. By measuring the speed with which the photosphere expands, the temperature and the radiative flux, it is possible to trace the SN distance. The constancy of the speed of the photosphere is guaranteed by the great density that characterizes the photosphere of the SNe of type IIn.

In 2020 various experiments with SNe Ia are planned to improve the estimates of cosmological parameters. Let's see some approaches (see [30] and references therein for details).

The first approach is by fitting the cosmological models on the Hubble Diagram (redshift, number of SNe). We try to distinguish between dark energy models, increasing the statistics and the redshift range. The range is from 0.1 to ~ 2.3 and the number of SNe will increase of $\sim 5x$ at low z and $\sim 2x$ to mid z. When, in future, LSST and WFIRST will come in play, the multipliers will become 300 for mid z and 1000 at high z.

A second approach involves the calculation of the intercept of the Hubble parameter. Using Cepheids and SNe at a similar distance, the calibration set is created. The more distant SNe instead create the Hubble flow set. The average difference in brightness between these two sets allows the intercept to be calculated. For this approach, sources with z < 0.5 are used.

Another one is the measure of the peculiar velocities of galaxies that deviate from the homogeneous expansion of the Universe. There are deviations due to the gravitational attraction from the large-scale structure in the Universe and this allows to probe the total distribution of matter, including dark matter, as well as to measure on large scales if there are deviations from general relativity. These peculiar velocities are measured by comparing the Hubble residues of the distances of the SNe Ia with low z.

A further approach uses the lensing of the SNe by the structure of the Universe, which modifies their brightness as a function of the density of the regions crossed by the photons, as we said previously. It is possible to measure cosmological parameters either by correlating the magnitudes of lensed SNe with the density observed along the line of sight, or by using amplification to estimate the properties of the halo of dark matter. For this approach will be used SNe in redshift range by $1 < z < 2$.

The last approach is the time-delayed cosmography, which measures the time delay between multiple images of a strongly lensed transient. Combining this method with a model for the lensing potential, we obtain the ratio of cosmological distances between source and lens. This distance ratio is inversely proportional to H_0. This approach will be used for both type Ia and type II SNe with $z > 1$.

1.6 Test campaign

As we have seen, SNe are very important elements for understanding our Universe, and their clear classification allows us to use only the types of SNe required. The best classification is undoubtedly the spectroscopic one, but today the spectroscopic techniques are too time consuming, so we had to work with photometry. As you go towards tools like LSST that can discover almost 2000 SNe per night, it is essential to be able to classify them in a short time and with as few points available as possible. Data mining can help, as being an indispensable tool for prompt data analysis.

In pursuing the goal of photometric SN classification, a test campaign was identified, hierarchically diversifying and based on the different scientific complexity of the various experiments:

1. Generic distinctions
    - Periodic Vs No Periodic: We started with this experiment, as being quite simple and therefore suitable as a preliminary approach.

- SNe Vs All: Among the various transients available, it was decided to focus on SNe, due to their extreme importance at a cosmological level. In fact, the uniformity of type Ia light curves allows them to be used as standard candles for distance measurements as well as for local clocks. To improve the estimates, large samples of SNe are needed, unreachable only spectroscopically. That is why research is pushing towards photometric classification.

2. Analysis of performance related to supernovae

    - SNe Ia Vs SNe II: In theory these two classes should be the most easily classifiable among the various classes of SNe.
        (a) PLAsTiCC dataset
        (b) SNPhotCC dataset
        (c) Long-Short Time Memory model optimization
    - Superluminous SNe Vs SNe Ia mixed: This experiment is important because in recent years the study has proceeded not only to understand the mechanisms that lead to the explosion of such massive stars, but also to understand their contribution to the chemical evolution of the Universe and to its reionization.
    - Six class problem: This is the most complex experiment, due to the widest dimension of the parameter space and to the presence of relatively similar objects, which further complicates the classification work.

Chapter 2

Data mining

Data mining is the ensemble of methodologies and techniques for mining information from massive data with automatic or semi-automatic methods to discovery new patterns and correlations.

Data mining is not a modern thing [41], in fact, it was born many decades ago and its basis comes from three century of statistical and scientific studies. In 1763, the Bayes theorem was published; it explains the conditional probability of two events. In 1805, Legendre and Gauss introduced a key functional case for data mining: the Regression. After more then a century, in 1936, Alan Turing, in his paper "On Computable Numbers", introduced the first prototype at the base of future computers. Another milestone was set in 1943 with McCulloch and Pitts, through their paper "A logical calculus of the ideas immanent in nervous activity" that poses the basis of Neural

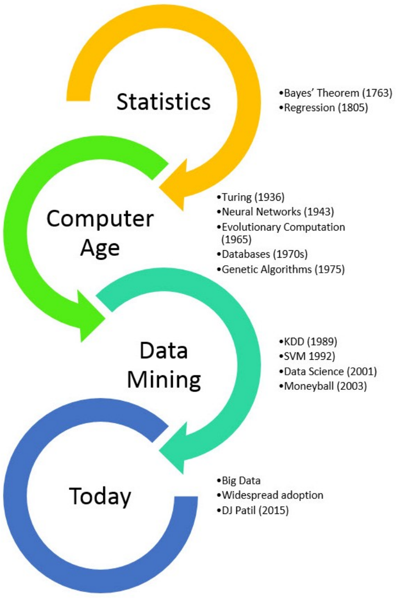


Figure 2.1: Data mining timeline

Networks. After these, there have been many other progresses, from databases (1970) to Knowledge Discovery in Database(KDD) in 1989. In the 1990s, for the first time data mining was used in the financial field for analysis of trends. From that, the computer science sector has developed up to what we know today.

2.1 Data

Data are described by a number of attributes featuring an object [31]. An attribute (or feature) is a characteristic of an object that may vary in time or from one object to another. To compare more objects with the same attribute, we must use a measurement scale that is a function associating a numerical value with an attribute of an object.

We can define four kind of attributes:

1. Nominal;
2. Ordinal;
3. Interval;
4. Ratio.

The first two are qualitative attributes, while the last two are quantitative.

The Nominal attribute includes the properties of *distinctness*; it is only for the distinction of the objects.

The Ordinal attribute is strictly related to the properties of *order*, in fact, this kind of attributes gives only ordering information.

The Interval and Ratio attributes are related, respectively, to the properties of *addition* and *multiplication*.

Another property of attributes is the *asymmetry.* When in a distribution the only important value of an attribute is the non zero value, we talk about asymmetric binary attribute.

Before moving on to data quality, we spend a few words on the characteristics of the datasets. A dataset is a structured ensemble of data and the three most important features are: *Dimensionality, Sparsity* and *Resolution.* The first is the number of features that an object possesses. In general, the more attributes are present, the better the analysis is, but too much attributes could make us fall into the curse of dimensionality (see section about pre-processing). Sparsity is related to asymmetric attributes. In fact, in a dataset with few non zero elements, there is the possibility to save time and storage, because only the non zero values need to be stored. The third one, Resolution, is important because such phenomena are visible in a given resolution and may disappear in a different one.

2.1.1 Data quality

The most common element of a measure that reduce the quality of a data is the noise. It is a random component of measurement error and its removal could be very hard.

Another type of ambiguous *noise*, is the outliers. An outlier, often, is a false noise, consisting frequently in a peculiar object that comes out from the distribution and could be interesting to study; in other cases may be an object with too much noise. So one must be careful to evaluate them.

The quality of a data is measured in terms of *precision* and *bias.* The first is defined like the closeness between two repeated measures of the same quantity. Often, the precision is measured by standard deviation of a data distribution. The bias is a systematic variation of measure from the correct value; in fact, bias is measured by subtracting the mean of measurements from the measure. Obviously, this kind of quantity can be determined only if the correct result is a priori known.

A more general term for the degree of measurement error is the *accuracy*, which is defined like the closeness to the true value of the measured quantity.

It does not always happen that an attribute has a value for all objects. In this case we talk about of *missing value*. There are many kinds of solutions to face this problem; the most common are:

- Eliminate objects or attributes subject of missing entries;
- Estimate and replace missing values (through techniques called imputation);
- Ignore missing values. In Astrophysics this option should be avoided, due to the uncertain source of such missing entries (instrumental troubles or too low S/N in the observed flux).

In some cases it is preferable to try to guess any estimate of any missing value, like the case of smooth time series; the value can be calculated by interpolating the existing close values or by replacing it with any statistical measure of other values assumed by the attribute (e.g. mean, median, standard deviation, etc.). In other cases, when the attribute is qualitative, the value may be estimated by setting the most representative one for that attribute.

Another way is to ignore the missing values. But this should be avoided, especially in case of machine learning, since they would dramatically impact on the learning performance and generalization capability of the trained model.

2.1.2 Data preprocessing

Propaedeutic to any approach based on data mining or machine learning on a dataset, is the initial setup and choice of data and their features (parameter space) to be used. This phase is crucial for the good outcome of any experiment, both in terms of computational costs and of qualitative performance. There are many techniques suitable for data preprocessing:

- Aggregation;
- Sampling;
- Dimensionality reduction (i.e. compression);
- Feature Selection;
- Feature extraction;
- Discretization and Binarization;
- Variable transformation or normalization.

Aggregation is an interesting technique, allowing the reduction of the data volume, thus reducing time and costs of data analysis. Another advantage is the potential discovery of unexpected patterns or correlations among features. Finally, it is statistically proved that aggregated data are more stable than single ones, just by considering that the average or the total sum of a given attribute is less variable than the single object. A weak side of this technique is that patterns appearing evident with single data, disappear by aggregating them.

Sampling, like aggregation, reduces the size of a dataset, as long as the subset is representative of the entire dataset. Representative means that it maintains main properties of the original dataset. There are many techniques, among which the most simple is the *random sampling* that have two variants: sample with/without replacement. Without replacement, sampling has a different percentage after each step, but every object may be raffled only one time. With replacement, instead, the percentage is always the same, but there may be duplications.

If there are many groups of objects into a dataset, the simple sampling may fail, because groups less populated have less percentage to be draw. In this case we can use the *stratificated sampling*, in which we could draw the same number of objects from every group or in a more complex case, the same percentage from every group. Since it is not simple to know the right size of the subset, *progressive sampling* is sometimes used, increasing the model's accuracy. The size grows in the same way, until a break point, where the accuracy stops to increase and the right size is reached.

Dimensionality reduction is an ensemble of techniques that reduce the dimensionality of a dataset, by creating new features that are combination of the old ones. This kind of technique have many benefits. A reduction of dimensionality can lead to a more understandable model, because the model involves fewer features. Moreover, less dimensions involve a better data visualization, because is more easy to find a significant subset of features that reveal some correlation. Another benefit is by saving memory and time of algorithms with the reduction of features. The most important benefit is that with fewer dimensions, data mining algorithms work better. First, because the reduction erases the irrelevant features, thus reducing their intrinsic noise; second, because it solves the problem of the "curse of dimensionality". It refers to the classical problem for data mining algorithms, occurring in case of high dimensional data. The more features included within the parameter space, the more sparse will result it, due to the increase scatter (distance) among data. As result, classification may have less accuracy and clustering poorer quality.

The most common technique for dimensional reduction is the linear algebra technique known as *Principal Component Analysis* (PCA), which finds eigenvectors of the covariance matrix of original feature space, thus obtaining the maximum amount of variation in the data and reduction of the final parameter space.

Another way to optimize the parameter space of data is to identify and remove the *Irrelevant* and *Redundant features.* This latter category of features includes attributes with an informative content already provided by other features of the space, while the irrelevant ones contain useless information for the problem under examination. Such kinds of features are related to the difficult problem to select the best subset of attributes within a dataset sample, able to improve the solution of any real problem connected by those data.

Features Selection is the ensemble of techniques aimed at identifying a subset of more representative features within a given parameter space. There are methods able to extract the best minimal subset of relevant features for a given problem or the so-called "all-relevant" feature set, including best and weak relevant features [5, 12]. The most simple way to perform a feature selection would be to analyze all possible combinations of subsets, but due to the exponential cost, alternative approaches are required. These are usually divided into three categories, *Embedded, Filter* and *Wrapper.*

Embedded feature selection is a method in which the selection is made by the algorithm itself, like decision tree, used to solve the data related problem.

Filter based selection makes use of an additional dedicated method before to apply the problem solving algorithm.

The Wrapper method, instead, ignores the algorithm type and offers many kinds of random subsets to test the best one.

Instead of select a subset, we could make a *feature generation*, by finding new features from the original parameter space. Obviously the target is to obtain a new set of features smaller than the original one. The most common methodologies are *Feature Extraction*, *Mapping the data to a new space* and *Feature Construction.*

Feature extraction is the generation of a feature set from original raw data. By changing the viewpoint of data may reveal hidden features, like in time series analysis, where, for example, the Fourier transform could help to identify the oscillation period of a periodic signal over the noise level.

Finally, whenever the given dataset contains the right information, but not the right shape to apply any solving algorithm, then any technique of feature construction can be used to adjust it.

Discretization implies the transformation of continue attributes into categorical ones; some classification algorithms need categorical features to work.

Binarization, instead, transforms continue and discrete features into one or more binary attributes; this is needed for some associative algorithms for example. For binarization the best approach is to create a number of features equal to the amount of categorical attributes and then to apply the asymmetric property. The discretization of a continuous attribute is simple. The feature's value is sorted and then divided into n intervals, by setting n-1 split points. Afterwards, every value in an interval switches to a categorical value for that interval. Obviously, main problem is to choose the number of split points and where to put them.

Discretization for classification could be supervised or unsupervised and it depends on the availability of a target label. For unsupervised cases the simple equal width of intervals could be heavily affected by outliers, so it is better a frequency approach, namely same number of objects in each interval. An approach for supervised discretization is to partition a continuous attribute by bisecting the initial values, so that the resulting two intervals give minimum entropy. This technique considers each value as a possible split point, because each interval contains an ordered set of values. The splitting process is repeated by choosing the interval with the highest entropy, until a stopping criterion is satisfied. Entropy of an interval is defined as a measure of purity of that interval. In formulas, the entropy of a generic interval i is:

$$e_i = \sum_{i=1}^{k} p_{ij} \log_2 p_{ij}$$

where $p_{ij} = m_{ij}/m_i$ is the probability of class j in the i^{th} interval; k is the number of different class labels, m_i is the number of values in the i^{th} interval of a partition, and m_{ij} is the number of values of class j in the interval i. The total entropy is defined as the weighted average of entropy:

$$e = \sum_{i=1}^{n} w_i e_i$$

where $w_i = m_i/m$ is the fraction of values in the i^{th} interval, m is the number of values and n is the number of intervals. It is evident that the greater the entropy e_i, the lower the purity in the i^{th} interval, so the goal is always to minimize entropy.

The last kind of technique for data preprocessing is the variable transformation. For simple transformations, the most common functions are $log(x)$, e^x, $\sqrt{x}$, etc. An example is the transformation of flux into magnitude with log function.

Another type of common trasformation is the *Standardization* or *Normalization.* If a dataset has a peculiar property, than we could use it for standardization purposes. An example is the z-score in which a set of values, with mean $\bar{x}$ and standard deviation σ, is transformed in a set of values with null mean and unitary standard deviation with this formula: $x' = (x - \bar{x})/\sigma$.

2.2 Classification and Regression tasks

The data science models are usually applied in optimization tasks, involving mainly two kinds of functionalities, respectively, classification and regression, for which the supervised learning paradigm is generally chosen.

The Regression is based on predictive techniques, where the target label is a continuous number, resulting from an a priori unknown analytical correlation of input features. Its goal is to find the target function that fits the data with the minimum error. The error function for regression can be expressed by the sum of absolute or squared errors:

$$Abs\ err = \sum_{i=1} |y_i - f(x_i)|$$

$$Sq\ err = \sum_{i=1} (y_i - f(x_i))^2$$

where y_i is the target label and x_i is the set of features that can be either discrete or continuous.

The Classification, instead, is a task where the target label may assume discrete numbers or categorical values and the output function f maps each feature set x to one of the predefined class labels y. In this work we focused the attention to the classification of time domain sources.

2.2.1 General approach

A classifier can be used as a descriptive model to distinguish among objects of different classes, and as a predictive model to predict the class label of input patterns. Classification techniques work better for predicting or describing data sets with binary or nominal categories. Each technique uses a different learning algorithm to find a model that fits the relationship between the feature set and class labels of the input data. The goal of the learning algorithm is to build models with good generalization capability. The typical approach of machine learning models is to randomly shuffle and split the given input dataset with known assigned class labels into three subsets: training, validation and blind test sets. The validation set can be used to validate the learning process, while the test set is used blindly to verify the trained model performance and generalization capabilities. The performance of a classification model is based on some statistical estimators, extracted from a matrix known as *confusion matrix* shown below.

		Predicted	
		P=0	**N=1**
Target	**p=0**	True Positive	False Negative
	n=1	False Positive	True Negative

This is a confusion matrix for a binary classification. Each entry a_{ij} in this table is the number of records from class i predicted to be of class j. The numbers a_{00} and a_{11} show correct classified records. The a_{01} records named *False Positive* indicate wrong records classified in class 0, when their correct classification was class 1; instead, a_{10} named *False Negative* show the records classified in class 1 but belonging to class 0. The total number of correct predictions is $a_{11} + a_{00}$, and the total number of wrong ones is $a_{10} + a_{01}$. For a better comparison between different models, summarizing the results through a confusion matrix is the common way. We can do this using a *performance metric*, such as *accuracy*, defined as follows:

$$Accuracy = \frac{a_{00} + a_{11}}{a_{00} + a_{11} + a_{01} + a_{10}}$$

Another way to show the performance of a model can be in terms of its error rate:

$$Error\ rate = \frac{a_{01} + a_{10}}{a_{00} + a_{11} + a_{01} + a_{10}}$$

Highest accuracy or equivalently lowest error rate, is the target of every classifier.

Other important statistical estimators, for a better understanding of the results for each class, are:

$$Purity = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Completeness = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$Contamination = 1 - Purity = \frac{FalsePositive}{TruePositive + FalsePositive}$$

$$F1_{Score} = \frac{2}{(Purity)^{-1} + Completeness^{-1}}$$

Purity of a class is the percentage of correctly classified objects in that class, divided by the total classified objects in that class. Also named as precision of a class.

Completeness of a class is the percentage of the correctly classified objects in that class divided by the total amount of objects belonging to that class. Also named as recall of a class.

Contamination of a class is dual of purity.

F1-Score of a class is the harmonic mean between purity and completeness of that class and it is a measure of the test accuracy.

2.2.2 Photometric Features

For any statistical approach it is necessary to create a set of features representing the peculiar characteristics of the astrophysical objects. Within this work we used the following features [10], resulting from a preliminary mapping of variable object light curves into a statistical parameter space:

- Amplitude (ampl);

  The arithmetic average between the maximum and the minimum magnitude:

$$ampl = \frac{mag_{max} - mag_{min}}{2} \tag{2.1}$$

- Beyond1std (b1std);

  The fraction of photometric points above or under one standard deviation from the weighted average:

$$b1std = P(|mag - \overline{mag}| > \sigma) \tag{2.2}$$

- Flux Percentage Ratio (fpr);

  The ratio between two flux percentiles $F_{n,m}$. The flux percentile is defined as the difference between the flux value at percentiles n and m, respectively. For this work, the following fpr values have been used:

$$\begin{gathered} fpr20 = F_{40,60}/F_{5,95} \\ fpr35 = F_{32,5,67,5}/F_{5,95} \\ fpr50 = F_{25,75}/F_{5,95} \\ fpr65 = F_{17,5,82,5}/F_{5,95} \\ fpr80 = F_{10,90}/F_{5,95} \end{gathered}$$

- Lomb-Scargle Periodogram (ls);

  the period obtained by the peak frequency of the Lomb-Scargle periodogram.

- Linear Trend (lt);

  the slope a of the light curve in the linear fit:

$$\begin{gathered} mag = a * t + b \\ lt = a \end{gathered} \tag{2.3}$$

- Median Absolute Deviation (mad);

  the median of the deviation of fluxes from the median flux:

$$mad = median_i(|x_i - median_j(x_j)|) \tag{2.4}$$

- Median Buffer Range Percentage (mbrp);

  the fraction of data points which are within 10% of the median flux:

$$mbrp = P(|x_i - median_j(x_j)| < 0.1 * median_j(x_j)) \tag{2.5}$$

- Magnitude Ratio (mr);

  An index to see if the majority of data points are above or below the median of the magnitudes:

$$mr = P(mag > median(mag)) \tag{2.6}$$

- Maximum Slope (ms);

  the maximum difference obtained measuring magnitudes at successive epochs:

$$ms = max(|\frac{(mag_{i+1} - mag_i)}{(t_{i+1} - t_i)}|) = \frac{\Delta mag}{\Delta t} \tag{2.7}$$

- Percent Difference Flux Percentile (pdfp);

  the difference between the fifth and the 95th percentile flux, converted in magnitudes, on median flux:

$$pdfp = \frac{(mag_{95} - mag_5)}{median(mag)} \tag{2.8}$$

- Pair Slope Trend (pst);

  the percentage of the last 30 couples of consecutive measures of fluxes that show a positive slope:

$$pst = P(x_{i+1} - x_i > 0, i = n - 30, ..., n) \tag{2.9}$$

- R Cor Bor (rcb);

  the fraction of magnitudes that is above 1.5 magnitudes with respect to the median:

$$rcb = P(mag > (median(mag) + 1.5)) \tag{2.10}$$

- Small Kurtosis (sk);

  the ratio between the 4th order momentum and the square of the variance. For small kurtosis it is intended the kurtosis on a small number of epochs:

$$sk = \frac{\mu_4}{\sigma^2} \tag{2.11}$$

- Skew (skew);

  the ratio between the 3rd order momentum and the variance to the third power:

$$skew = \frac{\mu_3}{\sigma^3} \tag{2.12}$$

- Standard deviation (std);

  The standard deviation of the flux.

2.2.3 Optimization

High accuracy usually implies that the classification model has been optimized. The methodology for finding the maximum or minimum value of a function is called *Optimization.* There are two kinds of optimization: *Unconstrained* and *Constrained.*

Unconstrained Optimization

By setting $f(x_i)$ as a multivariate continuous function with continuous first and second order derivatives, the goal of this kind of optimization is to find the value of x_i^*, named *stationary point*, that minimizes or maximizes the function. This solution is found by deriving the function from the first order and setting the result equal to zero:

$$\left.\frac{\partial f}{\partial x_i}\right|_{x_i=x_i^*} = 0, \ \forall i = 1, 2, \ldots n$$

To understand if the stationary point is a maximum or a minimum is difficult because we need to verify the second derivative sign for all of them. They are enclosed in a matrix, for instance the *Hessian matrix*:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

A Hessian matrix is positive defined if and only if $x^T Hx > 0$ for any non-zero vector x; while it is negative defined if and only if $x^T Hx < 0$ for any non-zero vector x and is indefinite if $x^T Hx$ is positive for some values of x and negative for others.

If $H(x^*)$ is positive defined, then x^* is a minimum stationary point; if $H(x^*)$ is negative defined, then x^* is a maximum stationary point and if indefinite, then it is a saddle point.

In many case, an analytical way is not feasible, so far we must find the stationary point numerically. We will see three numerical methods, respectively, *Golden Search*, *Newton's Method* and *Gradient Descent Method.*

- Golden Search

Let us consider a continuous function $f(x)$ with a minimum enclosed between points a and b. Iterating, we must find a minimum interval to approximate the stationary point. Adding a smaller interval $c{-}d$ at the minimum, such that $a{-}c{-}d$ width is equal to $c - d - b$, we could say:

$$\begin{cases} c - a = b - d = \beta * (b - a) \\ d - c = \alpha * (b - a) \end{cases}$$

so,

$$\begin{cases} 1 = \frac{(b-d)+(d-c)+(c-a)}{b-a} = \alpha + \beta + \alpha \\ \frac{d-c}{b-c} = \frac{c-a}{b-a} \end{cases}$$

and then

$$\begin{cases} \beta = 1 - 2\alpha \\ \frac{\beta}{1-\alpha} = \alpha \end{cases}$$

This equations system leads to $\alpha = 0,382$ and $\beta = 0,236$.

By comparing $f(c)$ against $f(d)$, it is possible to detect whether the minimum value occurs in that interval. The interval that contains the minimum value is recursively partitioned, until the interval width is small enough to approximate the minimum value.

- Newton's Method

This method starts from Taylor's series expansion of the function $f(x)$:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(x_0)$$

Taking the derivative of $f(x)$ with respect to x and setting it to zero, we have:

$$x = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

This equation will be updated until x converges to the location of the minimum value.

In the multivariate way we can use the gradient operator $\nabla f(x)$ and the Hessian matrix, instead of the first and second order derivatives. The final equation is:

$$x = x_0 - H^{-1} \nabla f(x)$$

- Gradient Descent Method

Newton's method can be generalized like:

$$x = x_0 - \lambda g(x)$$

The gradient descent method sets $g(x) = \nabla f(x)$ where $f(x)$ is a continuous and derivable function. The location of x is updated to the steepest descent, which means that x moves to the lower value of f. The equation is:

$$x = x_0 - \lambda \nabla f(x)$$

Constrained Optimization

There may be two types of constraints, equal or unequal.

- Equality Constrains

To our function $f(x_1, x_2, \ldots x_n)$ is placed a constrain of the form:

$$g_i(x) = 0, \; i = 1, 2, \ldots m$$

The Lagrange multipliers can solve this problem. Define a Lagrangian:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

where λ_i are Lagrange multipliers.

Then we set the first derivative of the Lagrangian with respect to x and λ equal to zero. Then, by solving this system of $m+n$ equations, we can find the minimum x^* and the λ values.

- Inequality Constrains

The same above problem, but the constrains are:

$$h_i(x) \leq 0, \; i = 1, 2, \ldots m$$

With the same Lagrangian we can solve the problem, but the conditions, named Karush-Kuhn-Tucker (KKT), are different:

$$\begin{cases} \frac{\partial L}{\partial x_i} = 0, \; i = 1, 2, \ldots n \\ h_i(x) \leq 0, \; i = 1, 2, \ldots m \\ \lambda_i \geq 0, \; i = 1, 2, \ldots m \\ \lambda_i h_i(x) = 0, \; i = 1, 2, \ldots m. \end{cases}$$

2.2.4 Parameter Handling Investigation Laboratory (ΦLAB)

The choice of an optimal set of features is connected to the concept of feature importance, based on the measure of a feature's relevance [12]. Formally, the importance of a feature is its percentage of informative contribution to a learning system. We approach the feature selection task by the all-relevant feature selection, able to extract the most complete parameter space, i.e. all features considered relevant for the solution to the problem. This is appropriate for problems with highly correlated features, as these features will contain nearly the same information. With a minimal-optimal feature selection, choosing any one of them (which could happen at random if they are perfectly correlated) means that the rest will never be selected. The method ΦLAB, includes properties of both embedded and wrappers categories of feature selection to optimize the parameter space, by solving the all-relevant feature selection problem, thus indirectly improving the physical knowledge about the problem domain.

ΦLAB is based on the combination of two components: shadow features and Naive LASSO statistics. Given a data set of N samples, represented through a D-dimensional parameter space, we double the parameter space by introducing a shadow feature for each real one, by randomly shuffling its values among the N samples. Shadow features are random versions of the real ones and their importance percentage can be used as a threshold for the information noise. This threshold is important since feature selection methods only provide a ranking of the features. The second component of ΦLAB is based on the Naive LASSO statistics. The LASSO (Least Absolute Shrinkage and Selection) performs both a variable selection and a regularization of a ridge regression, enhancing the prediction accuracy of the statistical model. The regularization is a process based on the addition of a functional term to a loss function. LASSO performs the L_1 regularization, based on the L_1 norm, which has the effect of sparsifying the weights of the features, turning off the least informative features. We included two Naive LASSO techniques in ΦLAB. One is the Alternate-LASSO, able to find all weakly relevant features that could be removed from the standard LASSO solution. This method calculates a list of features alternate to those selected by the standard LASSO, each one associated with a calculated score, reflecting the performance degradation from the optimal solution. In ΦLAB, we select only the alternate features that achieve the lowest score difference from the best features, trying to reach the best trade-off between feature selection performance and flexibility in the analysis of the parameter space. These alternate features smoothly degrade the solution score, but relax the intrinsic stiffness of the best solution system. The second version of the standard LASSO is Enumerate-LASSO, which enumerates a series of different feature subsets, considered as solutions with a decreasing level of approximation. By enumerating a variety of potential solutions, there is a chance to obtain better solutions for the problem domain task. The shadow features and Naive LASSO are then combined by selecting the candidate weak relevant features through the shadow feature noise threshold and by extracting the final set of weak relevant features, based on the A-LASSO and confirmed by E-LASSO. To summarize, we find the list of candidate features through the shadow features technique and then we use the LASSO operator to explore the parameter space and verify the effective contribution carried by those features considered as weak relevant to the solution of the problem. The loss function based on L_1 regularization is crucial to quantify the degradation of performance when other features subsets are replacing the best one, identifying the redundancy of features that the shadow features technique is unable to disentangle. The pseudo-code of the features selection method can be summarized by the following steps:

1. Let the set $x_1, x_2, ..., x_D$ be the initial complete parameter space composed by D real features;

2. Apply the shadow feature selection and produce the following items:

    I. $SF = x_{s_1}, ...x_{s_D}$, the list of shadow features, obtained by randomly shuffling the values of real features;

    II. max(IMP[parameter space, SF]) $\forall x \in$ parameter space $\&\forall xs \in SF$, the importance list of all 2D features, original and shadows.

    III. st: noise threshold, defined as the max{IMP[SF], $\forall xs \in SF$}.

    IV. BR = $x \in$ parameter space with $IMP[x] \geq st$, the set of best relevant real features;

    V. RF = $x \in$ parameter space, rejected by the SFS, the set of excluded real features, i.e. not relevant;

    VI. WR = $x \in$ parameter space with $IMP[x] < st$, the set of weak relevant real features.

3. At this stage, the complete parameter space is now split into BR, WR, and RF. Now we consider the reduced parameter space, space red $= BR + WR$, obtained by excluding the rejected features. In principle, it may correspond to the original parameter space if there is no rejections by the SFS:

    I. If $RF == \emptyset \&\& WR == \emptyset$, the SFS method confirmed all real features as high relevant, therefore return ALL-RELEVANT (parameter space), i.e. the full parameter space as the optimized parameter space and EXIT.

    II. If $RF = \emptyset \&\& WR == \emptyset$, the SFS method rejected some features and confirmed others as high relevant, therefore return ALL-RELEVANT (BR) as the optimized parameter space and EXIT.

    III. If $WR = \emptyset$, regardless some rejections, SFS confirmed the presence of some weak relevant features that must be evaluated by LASSO methods, therefore go to step (4).

4. Apply E-LASSO method on the space red $= BR + WR$ producing:
    I. EL S: a list of M subsets of features, considered as possible solutions, ordered by decreasing score;
    II. If $WR \subseteq ELS$, then all weak relevant features are possible solutions, therefore return ALL-RELEVANT (BR + WR) as the optimised parameter space and EXIT.
    III. Else go to step (5);

5. Apply A-LASSO method on the space red $= BR + WR$ (set of candidate features) producing:
    I. AL S, a set of T features, each one with a corresponding list of features List(t) considered as alternate solutions with a certain score;
    II. If AL $S == \emptyset$, then no alternate solutions exist, therefore:
        i. If EL $S == \emptyset$, then return ALL-RELEVANT(BR) as the optimized parameter space and EXIT.
        ii. Else if EL $S = \emptyset$, then return ALL-RELEVANT(BR + EL S) as the optimized parameter space and EXIT.
    III. Else extract $\forall t \in T$ the alternate solution with Score(as) = min{Score(y), $\forall y \in$ List(t)};
    IV. Go to step (6).

6. For each x ∈WR:

  I. If x is alternate solution of at least one feature $t \in T$, with $[t \in BR||t \in ELS]$, then retain x within WR set;
  II. Else reject x (by removing x from WR);

7. Return ALL-RELEVANT(BR + WR) as the final optimized parameter space and EXIT.

2.2.5 Random Forest

A Random Forest [4] is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, ...\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x.

Now we characterize the accuracy of Random Forest (RF). First we can say that RF does not overfit adding trees to the forest, but it sets a limiting value of the generalization error PE*

$$PE^* = P_{X,Y}(mr(X, Y) < 0)$$

where X are inputs, Y are targets and $mr(X, Y)$ is the margin function. The margin measures the difference between the average number of votes for the right class and the average number of votes for other classes.

The margin function is:

$$mr(X, Y) = P_\Theta(h(X, \Theta) = Y) - max_{j \neq Y} P_\Theta(h(X, \Theta) = j)$$

and the value which converges PE^* is:

$$P_{X,Y}((P_\Theta(h(X, \Theta) = Y) - max_{j \neq Y} P_\Theta(h(X, \Theta) = j)) < 0)$$

So for a larger margin, more confidence have the classification.

Now we demonstrate that an upper bound for generalization error is given by:

$$PE^* \leq \overline{\rho}(1-s^2)/s^2 \tag{2.13}$$

We set the strength of a set of classifiers $\{h(x,\Theta)\}$ like:

$$s = E_{X,Y} mr(X,Y) \geq 0 \tag{2.14}$$

and the raw margin function as:

$$rmg(\Theta,X,Y) = I(h(X,\Theta)=Y) - I(h(X,\Theta)=\widehat{j}(X,Y))$$

where

$$\widehat{j}(X,Y) = \arg\, max_{j \neq Y} P_{\Theta}(h(X,\Theta)=j).$$

Because $s \geq 0$, the Chebychev's inequality gives:

$$PE^* \leq var(mr)/s^2 \tag{2.15}$$

var(mr) is the variance of margin function and we can write it as:

$$var(mr) = \overline{\rho}(E_{\Theta} sd(\Theta))^2 \leq \overline{\rho} E_{\Theta} var(\Theta) \tag{2.16}$$

where $\overline{\rho}$ is the mean value of the correlation between margin function and raw margin function and sd is the standard deviation of raw margin function. We can write:

$$E_{\Theta} var(\Theta) \leq E_{\Theta}(E_{X,Y} rmg(\Theta,X,Y))^2 - s^2 \leq 1 - s^2 \tag{2.17}$$

and by putting the last three equations together we get the result.

For more than two classes, the measure of strength depends on the forest and on single trees, because the forest determines $\widehat{j}(X, Y)$. Another approach could be:

$$\begin{aligned} PE^* &= P_{X,Y}(P_\Theta(h(X,\Theta) = Y) - max_{j \neq Y} P_\Theta(h(X,\Theta) = j) \leq 0) \\ &\leq \sum_j P_{X,Y}(P_\Theta(h(X,\Theta) = Y) - P_\Theta(h(X,\Theta) = j) \leq 0) \end{aligned}$$

if the strength of the set of classifiers relative to class j is:

$$s_j = E_{X,Y}(P_\Theta(h(X,\Theta) = Y - P_\Theta(h(Y,\Theta) = j)) \tag{2.18}$$

Using Chebyshev's inequality we have:

$$PE^* \leq= \sum_j var(P_\Theta(h(X,\Theta) = Y) - P_\Theta(h(X,\Theta) = j))/s_j^2 \tag{2.19}$$

and using mathematical steps similar to previous proof, we can express variance in terms of average correlations. It is important to note that this definition of strength does not depend on the forest.

We can say that the generalization error for this algorithm depends on the strength of single trees and from their correlations through the raw margin functions. The upper bound, instead, tell us that smaller the ratio of those quantities is, better the RF performance are.

How to improve accuracy by keeping trees strength? We must decrease the correlation between them, and a way is to use bagging with a random selection of features.

Bagging or Bootstrap Aggregating, is an algorithm designed to improve the stability and accuracy of machine learning algorithms. It also reduces variance and helps to avoid overfitting. Given a training set of size n, bagging generates m new training sets each of size p, by sampling from the original one uniformly and with replacement. This kind of sampling is known as a bootstrap sample. The m models are fitted using the m bootstrap samples and combined by averaging the output (for regression) or voting (for classification).

Bagging is useful because, in addition to improving accuracy when using random features, it provides an estimate of the generalized error of the set of trees and the strength and correlation of trees. The estimation is done out-of-bag. Out-of-bag means that the error estimate of each pair (x,y) is made on all those bagging datasets that do not contain that given pair.

2.2.6 Long Short Term Memory (LSTM)

The LSTM [39, 40] is a particular type of Recurrent Neural Network (RNN).

A RNN is a neural network composed by multiple copies of the same network, each passing the information to the next as in Figure 2.2.

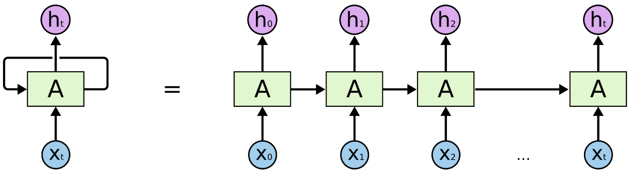


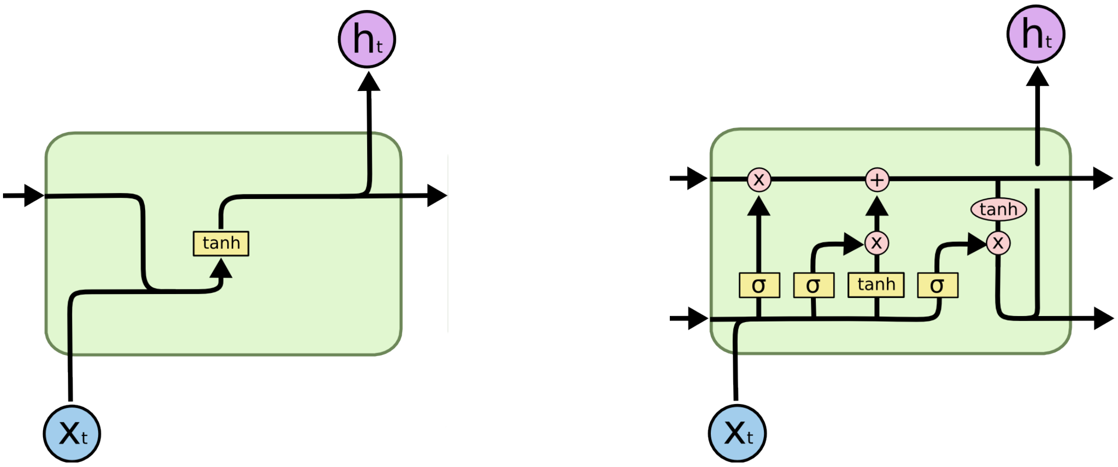
Figure 2.2: Recurrent Neural Network example

The limit of a generic RNN is the long temporal memory of information. The LSTM goes beyond with a particular composition of its cells. While in the RNN the cell contains only a tanh (hyperbolic tangent) gate, the LSTM contains multiple gates with sigmoids and tanh that, as we shall see, allow a long storage of information.

The most important innovation is the top channel linking all cells. On this line the information runs, it undergoes the modifications linked to the various gates and is then passed to the next cell. Now we look step by step at the gates.

The first is the Forget gate composed by a sigmoid, which varies between 0 and 1, that decides how much previous information must continue to be remembered.

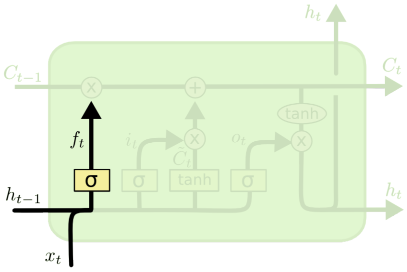
f_t is the function of this gate, with σ, the sigmoid function, which have as



(a) RNN cell

(b) LSTM cell

Figure 2.3: Cell examples of a LSTM



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

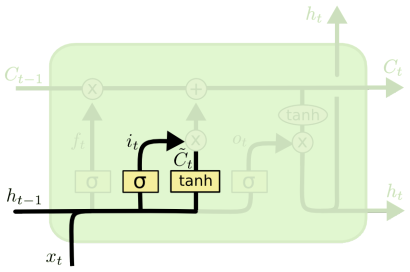
Figure 2.4: LSTM - Forget gate

arguments W_f the weights, h_{t-1} the hidden state at time $t-1$, x_t the input at time t and b_f the bias. More this function is close to 1, more information will be kept.

The second step is the updater of cell state, formed by two gates. The first is the input gate that decides with a sigmoid which value will be updated, while the second gate, with a tanh, creates a vector state $\widetilde{C}_t$ to add to the cell state C_{t-1}.

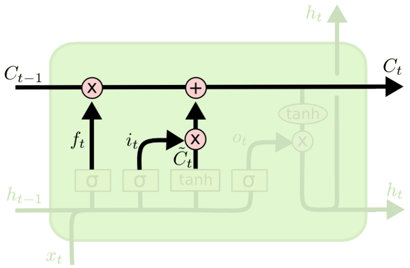
With this two steps we can update the old cell state and create the new cell state

The $*$ symbol is the Hadamard product, in which the matrix product is made element by element.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

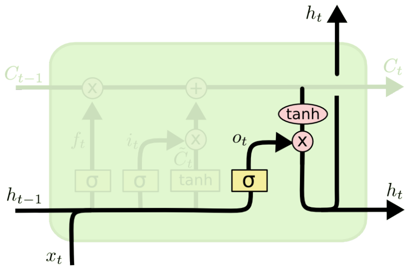
Figure 2.5: LSTM - Updater gate



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Figure 2.6: LSTM - New cell state

The last step is the update of the hidden state. The old one passes through a sigmoid gate, while the new cell state is the argument of a tanh. These two results are, then, multiplied.



$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 2.7: LSTM - Output state

In this work we used many of LSTM composed by two layers with a linear layer and softmax at the end, like done in the paper of Charnock&Moss [8].

2.2.7 Nadam, RMSProp and Adadelta

We will analyze the Nadam algorithm first, then the RMSProp and finally the Adadelta.

The simplest optimization algorithm is the *Gradient Descent*, in which the gradient of the function to be minimized is calculated. This depends on the parameter θ_{t-1}. Only a portion of the gradient is used to update the parameters; this portion is given by the parameter η.

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ \theta_t \longleftarrow \theta_{t-1} - \eta g_t \end{cases}$$

The next step is the *Classical Momentum* algorithm, in which, instead of the gradient, the parameters are updated by the vector momentum m, generated by the gradient and the decaying sum of the previous gradients with a μ decay constant. This method increases the speed with which the gradient decreases in the directions in which the gradient tends to remain constant, while reducing it in those where the gradient tends to oscillate.

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ m_t \longleftarrow \mu m_{t-1} + g_t \\ \theta_t \longleftarrow \theta_{t-1} - \eta m_t \end{cases}$$

We can see that in the momentum definition, the terms m_{t-1} and g_t are independent, so we can improve the algorithm in the *Nesterov's accelerated gradient* (NAG).

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1} - \eta\mu m_{t-1}) \\ m_t \longleftarrow \mu m_{t-1} + g_t \\ \theta_t \longleftarrow \theta_{t-1} - \eta m_t \end{cases}$$

To obtain Nadam, we must introduce the Adam algorithm, based on the combination between the momentum implementation and another type, which is based on the L_2 normalization. This type of normalization changes the η member, dividing it by the L_2 norm of all previous gradients. Indeed the Adam algorithm is:

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ m_t \longleftarrow \mu m_{t-1} + (1-\mu) g_t \\ \hat{m}_t \longleftarrow \frac{m_t}{1-\mu^t} \\ n_t \longleftarrow \nu n_{t-1} + (1-\nu) g_t^2 \\ \hat{n}_t \longleftarrow \frac{n_t}{1-\nu^t} \\ \theta_t \longleftarrow \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{n}_t}+\epsilon} \end{cases}$$

Now, revisiting the NAG and merging it with the Adam, we get Nadam:

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ \hat{g} \longleftarrow \frac{g_t}{1-\prod_{i=1}^{t} \mu_i} \\ m_t \longleftarrow \mu m_{t-1} + (1-\mu) g_t \\ \hat{m}_t \longleftarrow \frac{m_t}{1-\prod_{i=1}^{t+1} \mu_i} \\ n_t \longleftarrow \nu n_{t-1} + (1-\nu) g_t^2 \\ \hat{n}_t \longleftarrow \frac{n_t}{1-\nu^t} \\ \overline{m}_t \longleftarrow (1-\mu_t)\hat{g}_t + \mu_{t+1}\hat{m}_t \\ \theta_t \longleftarrow \theta_{t-1} - \eta \frac{\overline{m}_t}{\sqrt{\hat{n}_t}+\epsilon} \end{cases}$$

RMSProp is a L_2 normalization based algorithm and to reach it we must start from Adagrad algorithm that is a L_2 normalization based too. In the Adagrad algorithm η is divided for every step by the L_2 norm of all previous gradients. This has the advantage of compensating for the speeds along the different dimensions by stabilizing the model on common features and allowing the rare ones to emerge.

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ n_t \longleftarrow n_{t-1} + g_t^2 \\ \theta_t \longleftarrow \theta_{t-1} - \eta \frac{g_t}{\sqrt{n_t}+\epsilon} \end{cases}$$

On other hand, a great problem of this algorithm comes from the norm vector that could becomes so large to stop the training, preventing the model from reach the local minimum. This problem is resolved by RMSProp by replacing the sum of n_t with a decaying mean parameterized by a costant value ν. This allows the model to no stop the learning.

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ n_t \longleftarrow \nu n_{t-1} + (1-\nu) g_t^2 \\ \theta_t \longleftarrow \theta_{t-1} - \eta \frac{g_t}{\sqrt{n_t}+\epsilon} \end{cases}$$

The Adadelta algorithm [37] goes a step further by modifying the parameter update. Indeed, in terms of units of measurement, the current relationship is incorrect. To rectify it was introduced a diagonal Hessian. Since the RMS of the previous gradients is already represented in the denominator, a measure of the $\Delta\theta$ quantity will be in the numerator. An approximation of $\Delta\theta_t$ was computed by the exponentially decaying RMS over a window of size ω of previous $\Delta\theta$, assuming the locally curvature smoothly.

$$\begin{cases} g_t \longleftarrow \nabla_{\theta_{t-1}} f(\theta_{t-1}) \\ n_t \longleftarrow \nu n_{t-1} + (1-\nu) g_t^2 \\ \Delta\theta \propto H^{-1} g \propto \frac{\partial f/\partial\theta}{\partial^2 f/\partial\theta^2} \propto units\ of\ \theta \\ \theta_t \longleftarrow \theta_{t-1} - \frac{\sqrt{[\Delta\theta]_{t-1}+\epsilon}}{\sqrt{n_t}+\epsilon} g_t \end{cases}$$

Chapter 3

Data

In this work two datasets were used; the Supernova Photometric Classification Challenge (SNPhotCC) [22] and the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) [33, 34]. The first was used for tuning and testing the LSTM classifier, while the second for testing all other classifiers.

3.1 SNPhotCC

This challenge was performed in 2010 and consists of a mixed set of simulated SN types, respectively, Ia (5088), Ibc (2801) and II (13430), selected respecting the relative rate. The volumetric rate was found by Dilday et al.(2008) [13] as $r_v = \alpha(1+z)^\beta$, where for SN Ia parameters we have $\alpha_{Ia} = 2.6 \times 10^{-5} Mpc^{-3} h_{70}^3$ yr^{-1}, $\beta_{Ia} = 1.5$ and $h_{70} = H_0/(70\ kms^{-1}Mpc^{-1})$. H_0 is the present value of the Hubble parameter. For non Ia SNe the parameters come from Bazin et al.(2009) [2] and are $\alpha_{NonIa} = 6.8 \times 10^{-5} Mpc^{-3} h_{70}^3\ yr^{-1}$ and $\beta_{NonIa} = 3.6$. The simulation is based on four bands, *griz*, with cosmological parameters $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $\omega = -1$, where Ω_M is the density of barionic and dark matter, Ω_Λ is the density of dark energy and ω is the cosmological constant. Moreover, the pointspread function, atmospheric transparency and sky-noise were measured in each filter and epoch using the one-year chronology.

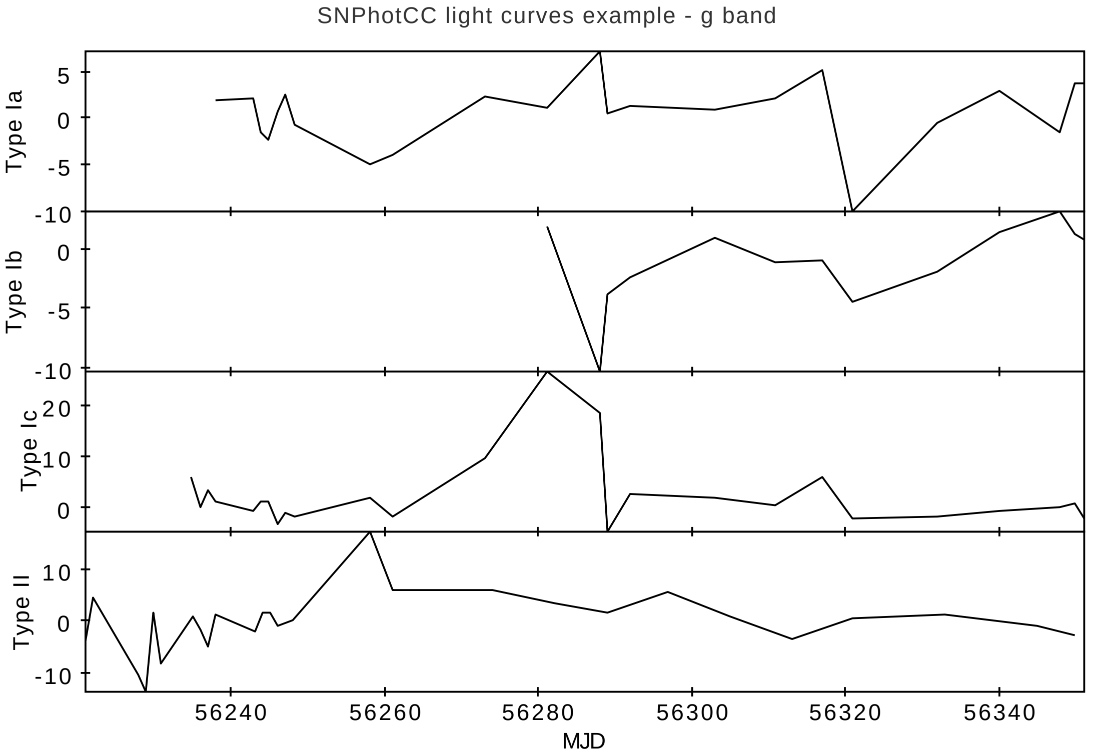


Figure 3.1: Examples of SNPhotCC light curves in g band. From the top to the bottom: SN004923(Ia), SN000760(Ib), SN003475(Ic), SN001986(II).

The dataset sources are based on two variants, respectively, with or without the host-galaxy photometric redshift. For this work only the samples without redshift information were used.

Every simulated light curve has at least one observation, in two or more bands, with signal-to-noise ratio > 5 and five observations after the explosion. A spectroscopically confirmed training subset was provided; it was based on observations from a 4m class telescope with a limiting r-band magnitude of 21.5 and on observations from an 8m class telescope with a limiting i-band magnitude of 23.5.

Types	Bands	Sampling	%	Amount
SNIa	g,r,i,z	uneven	23,86	5088
SNIbc	g,r,i,z	uneven	13,14	2801
SNII	g,r,i,z	uneven	63	13430

Table 3.1: SNPhotCC dataset information.

3.2 PLAsTiCC

This catalogue arises from a challenge focused on the future use of the Large Synoptic Survey Telescope (LSST), [20], by simulating the possible objects on which science will be based. In particular, all these objects are transients.

LSST will be the largest telescope specialized for the Time Domain Astronomy, whose first light is foreseen in 2020. Its field of view will be $\sim$ 3.5 degrees (the diameter will be about seven full moons side by side), with a 6.5 meter effective aperture, a focal ratio of 1.23 and a camera of 3.2 Gigapixel.

Every four nights it will observe the whole sky visible from the Chile (southern emisphere). Therefore, it will find an unprecented amount of new transients: Supernovae Ia, Ia-91bg, Iax, II, Ibc, SuperLuminous Supernovae, Tidal Disruption Events, Kilonova, Active Galactic Nuclei, RR Lyrae, M-dwarf stellar flare, Eclipsing Binary stars, Pulsating variable stars, μ-lens from single lens, μ-lens from binary lens, Intermediate Luminosity Optical Transient, Calcium Rich Transient and Pair Instability Supernovae.

LSST data will be used for studying stars in our Galaxy, understanding how solar systems and galaxies formed and the role played by massive stars in galaxy chemistry and measuring the amount of matter in the Universe. PLAsTiCC includes light curves with realistic time-sampling, noise properties and realistic astrophysical sources [42]. Each object has observations in six bands: u($300 \div 400$

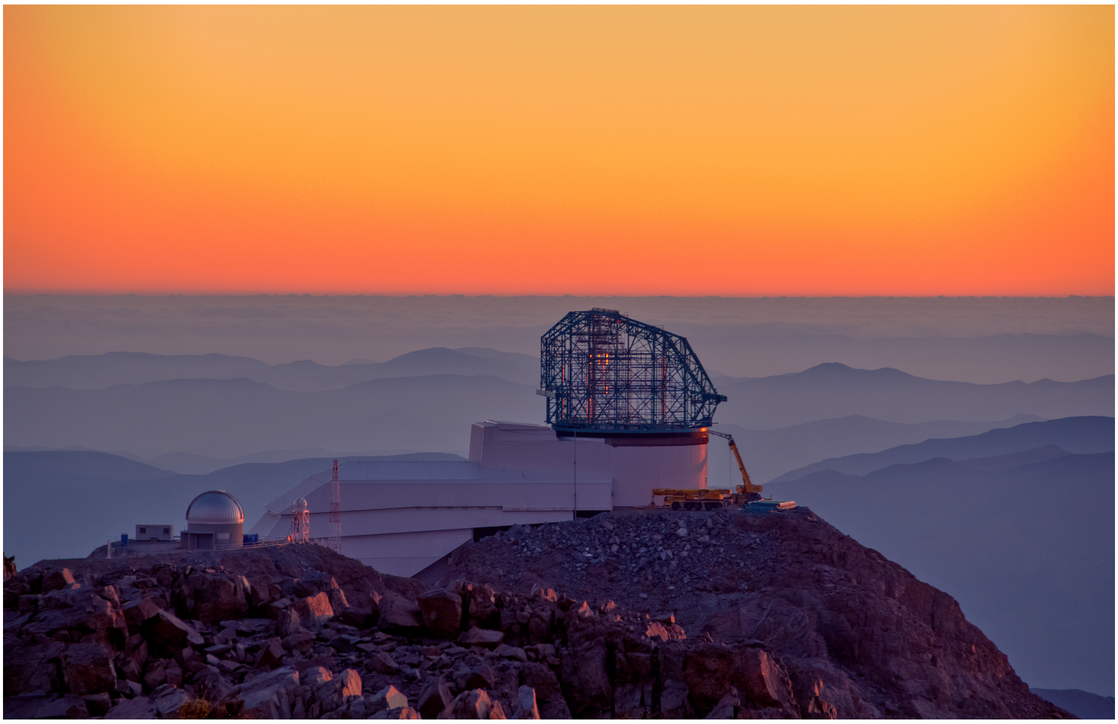


Figure 3.2: LSST under construction.

nm), g($400 \div 600$ nm), r($500 \div 700$ nm), i($650 \div 850$ nm), z($800 \div 950$ nm), and y($950 \div 1050$ nm). The training set is a mixture of what we can expect to have before LSST, so it is a quite homogeneous ensemble of ~ 8000 objects; the test set, instead, is based on what we expect to have after 3 years of LSST operations and it is formed by $\sim 3,5$ million of objects. The observations are limited in magnitude in single band to 24.5 in the r band and to 27.8 r stacked band.

By combining training and test, we have the following objects per class: SNIa (1662144), SNIa-91bg (40401), SNIax (63847), SNII (1001343), SNIbc (175578), SLSN-I (35957), TDE (14050), KN (233), AGN (101794), RRL (197394), M-dwarf (94475), EB (97496), Mirae (1483), μLens-Single (1454). In the test set there are four more classes: μLens-Binary (533), ILOT (1702), CaRT (9680), PISN (1172).

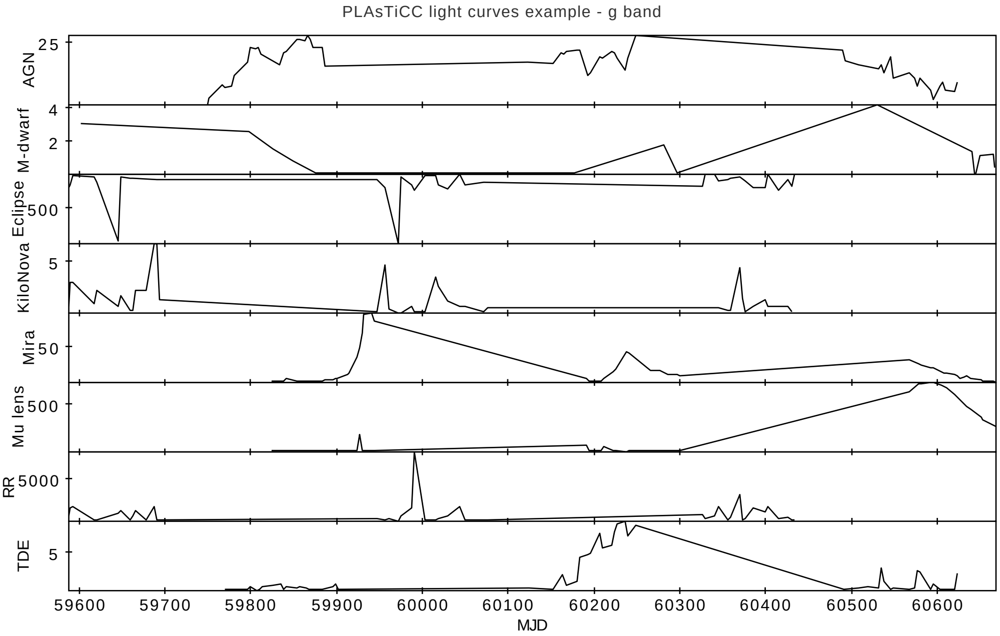


Figure 3.3: Examples of PLAsTiCC light curves in g band. From the top to the bottom: 2198(AGN), 2157270(M-Dwarf), 22574(Eclipsing Binary), 139362(Kilonova), 80421(Mirae), 45395(μ-lens), 184176(RR lyrae), 9197(TDE).

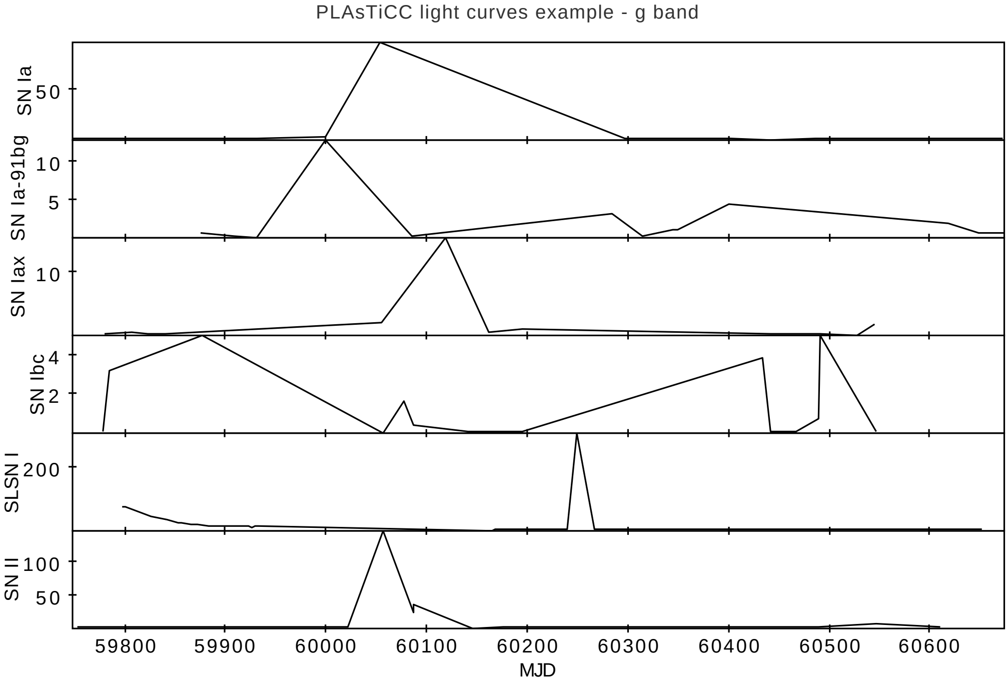


Figure 3.4: Examples of PLAsTiCC light curves in g band. From the top to the bottom: 15461391(SNIa), 1143209(SNIa-91bglike), 1019556(SNIax), 1076072(SNIbc), 73610(SLSN I), 1028853(SNII).

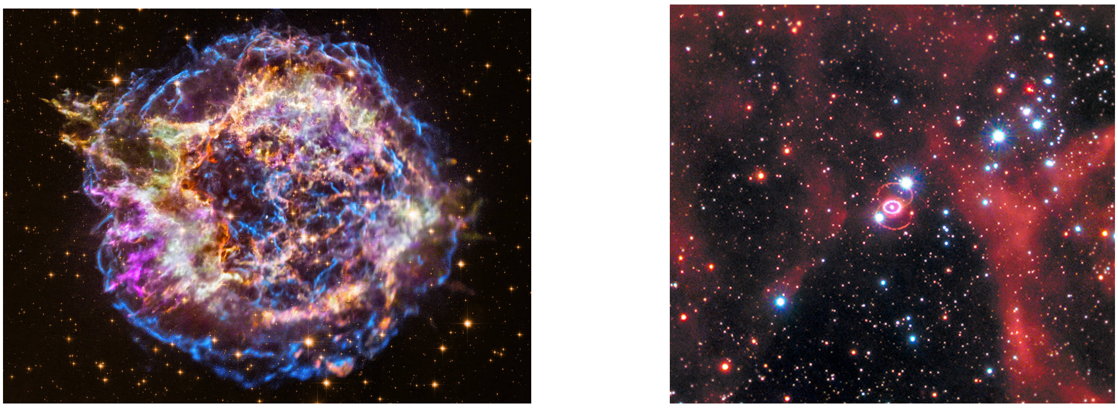
Types	Training	Test	Bands	Sampling	%	Amount
SNIa	2313	1659831	u,g,r,i,z,y	uneven	47.57	1662144
SNIax	183	63664	u,g,r,i,z,y	uneven	1.81	63847
SNIa 91bglike	208	40193	u,g,r,i,z,y	uneven	1.15	40401
SNIbc	484	175094	u,g,r,i,z,y	uneven	5.00	175578
SNII	1193	1000150	u,g,r,i,z,y	uneven	28.65	1001343
SLSN I	175	35782	u,g,r,i,z,y	uneven	1.02	35957
AGN	370	101424	u,g,r,i,z,y	uneven	2.89	101794
M-Dwarf	981	93494	u,g,r,i,z,y	uneven	2.68	94475
RR Lyrae	239	197155	u,g,r,i,z,y	uneven	5.63	197394
Mirae	30	1453	u,g,r,i,z,y	uneven	0.04	1483
Eclipse	924	96572	u,g,r,i,z,y	uneven	2.77	97496
KN	100	131	u,g,r,i,z,y	uneven	0.01	231
TDE	495	13555	u,g,r,i,z,y	uneven	0.38	14050
μ Lens	151	1303	u,g,r,i,z,y	uneven	0.04	1454
Other	0	13087	u,g,r,i,z,y	uneven	0.36	13087

Table 3.2: PLAsTiCC dataset information.

3.3 Types of objects

3.3.1 Supernovae

A SN [3, 9, 7] is an explosive phenomenon that leads the star towards the final stage of its life. A SN may arise in three ways. The first is when the core of a star has a high electron degeneracy, becoming unstable and causing a supersonic shock wave, called detonation, or a subsonic combustion, called deflagration (i.e., it expels the outer layers of star). If the original star has a mass lower than 8



(a) Cassiopea A

(b) SN1987a

Figure 3.5: Supernovae Remnants

solar masses, it becomes a white dwarf in the final state of its life; if the white dwarf belongs to a binary system, its companion could transfer part of its envelope to white dwarf, exploding when it reaches critical conditions of temperature and density. Being an old star, it has not hydrogen when it explodes, so this kind of exploding mechanism is suggested for type Ia. The second way is the core collapse supernovae. When the original star exceed the 8 solar masses, in final stages of its life, instead forming white dwarf, its core implodes because endergonic reactions in iron nucleus remove energy and pressure from star, leading to collapse. The 99% of the energy generated by collapse is carried out by neutrinos, while only 1% of the energy is transferred to the star for the explosion. For stars between 8 and 12 solar masses, the final destiny is a neutron star; more massive stars generate a black hole. The third way is the most catastrophic. For stars of over 100 solar mass, during the helium burning, the core is formed by oxygen, but the temperature is so high that pairs of electron positron are spontaneously generated. These brings the oxygen core to a strong thermonuclear explosion, destroying the entire star. This may have been the fate of the primitive stars after the cosmological dark age. Spectral classification of supernovae is based on optical spectra for historical motivation, however, it can be extended into near UV and near IR bands.

Type Ia - Iax - Ia 91bg-like

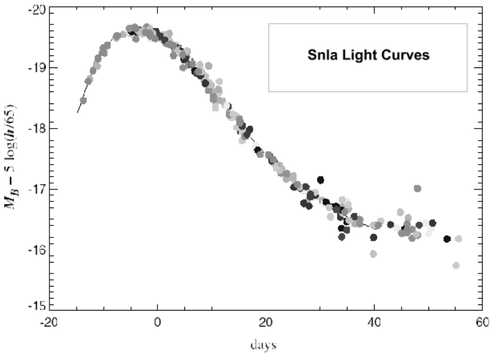


Figure 3.6: SN Ia light curve example

These types of SN are found in every type of galaxy. Taking into account that a white dwarf is accreting matter from its companion star, it will become a SN when its mass approach to Chandrasekhar limit of 1.4 solar masses. The conditions at the final stage of the mass accretion are a radius of $\sim 10^8 cm$, a central density of $\sim 10^9\ gcm^{-3}$ and a central temperature of $\sim 10^8 K^\circ$. In 1 second the nuclear energy brings the temperature to $10^9 K$, synthesizing heavier elements and starting the nearly adiabatic expansion. The conditions of the photosphere after the explosion are a velocity of 10000 km s^{-1}, a radius of 10^15 cm and a Temperature of 10^4 K. At the maximum luminosity (20 day after) the SN radiates 10^{43} erg s^{-1} and have a Mean Bolometric Magnitude value of $\sim -19.3 \pm 0.3$. Light curve's shape follows the density trend of the expelled mass. For first 200 days after the explosion, all gamma rays escape without interactions, but for hundreds more days the shape is linked to positrons. If there is a strong magnetic field, positrons are trapped and deploy their energy by collision with electrons generating gamma rays. In this case the slope of bolometric light curve follows the ^{56}Co decay rate; if the magnetic field is weak, positrons can escape before deploying energy and the light curve declines faster than ^{56}Co decay rate. Spectra of such kind of SNe do not have Hydrogen, but an absorption near 6100A produced by Silicon (Si II). Months after the maximum, spectra shows forbidden lines of iron.

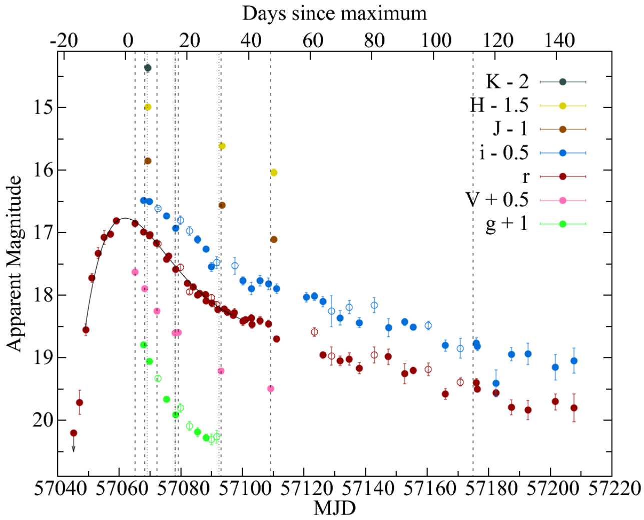


Figure 3.7: SN Iax light curve example

Type Iax [19] is a particular SN Ia class with a range of expansion velocity of $2000 - 7000$ km s^{-1}, a slower decline in redder bands and a lower luminosity at maximum. Moreover type Iax in the UV band start bluer than general type Ia, but few days after the maximum, it rapidly comes redder then a classical Ia. This type of SN does not have a fully nebular phase in which forbidden lines are predominant into the optical spectrum. The distribution of SNIax is concentrated into young galaxies with strong star forming region but with no predilection for high or low metallicity. By now, the best model for this type of SN is that a white dwarf with a Carbon/Oxygen core, or with the adding of Neon, increases its mass of Helium from a He-star companion. After reached the Chandrasekhar mass, it explodes with a deflagration that in many cases does not destroy the star. All these statements require further studies to confirm their accuracy.

Supernova 91bg-like [32] is less luminous than its cousin Ia, in fact, its absolute magnitude is maximum $-16.7 \div -17.7$; These luminosity differences are smaller by going to the reddest bands. The bolometric peak is used to calculate how much ^{56}Ni was produced into the explosion; The quantity is important because the end of light curve is sensitive to the ^{56}Ni mass, but also to gamma rays and positrons captured.

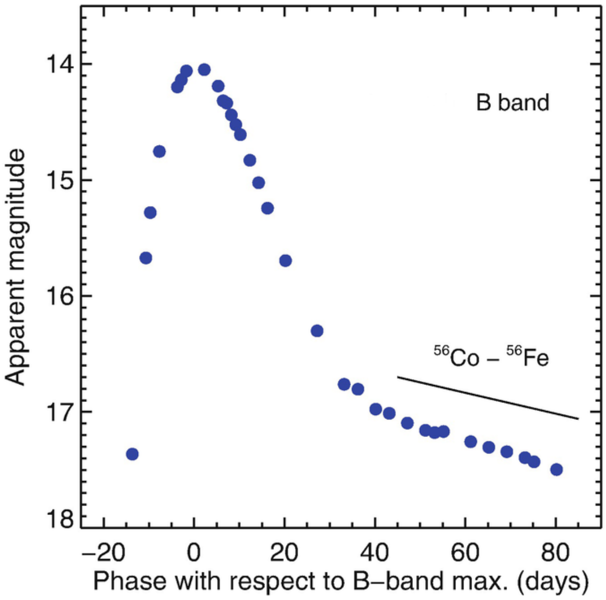


Figure 3.8: SN Ia91bg-like light curve example

Earlier than type Ia, optical spectrum after maximum is dominated by Fe and Co, but in general the spectra of this two types of SNe are very similar. Some lines particularly relevant for this type are O I, Ti II and Si II; C I in the NIR. This kind of SN is observed especially in massive elliptical or S0 galaxies. The fraction of SNe 91bg-like, with respect to all Ia types, oscillates between 6% and 15%; this value has a great range because there is the Malmquist bias that acts as an upper limit on these sub-luminous SNe.

Type Ib - Ic

These SNe are found exclusively in spiral and irregular galaxies. In spiral galaxies they are principally in the arms, in fact, they are associated with stars in recent star formation region. Type Ib occur in galaxies with less metal-rich region compared to type II region, and in galaxies with lower star formation rate compared to type Ic ones.

These stars are more massive and do not generate a white dwarf in final stage, so the type of explosion is different from type Ia. Another topic that pushes for the core collapse mechanism for this type of SN, is the similarity of nebular spectra between type Ib/c and type IIP.

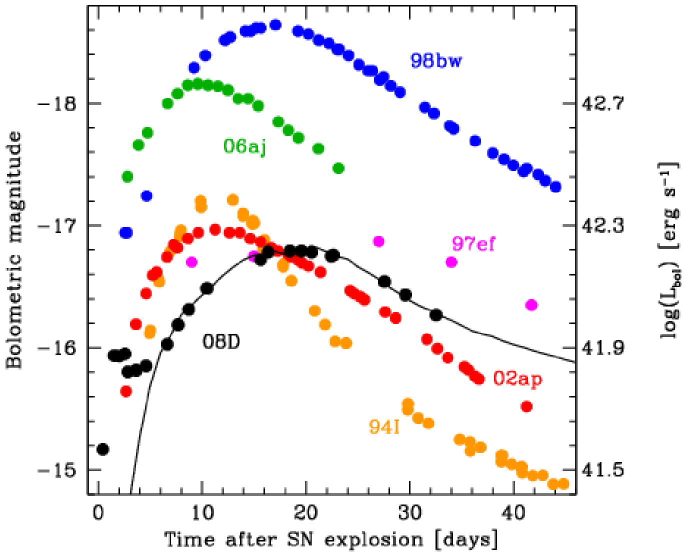


Figure 3.9: SN Ibc light curve examples

Type Ib have not Hydrogen lines, have weak or absent Silicon (Si II) absorption lines and have strong absorption lines of Helium (He I), instead type Ic have weak absorption line of Helium(He I) and strong absorption lines of Oxygen (O I). After the maximum luminosity, both reveal lines of semi forbidden Magnesium (Mg I) and forbidden Oxygen (O I) and Calcium (Ca II) Both types are referred like STRIPPED-ENVELOPE Supernova because most of their hydrogen envelope was lost before the explosion. The Mean Bolometric Magnitude values are (b) -17.4 ± 0.3 and (c)-17.3 ± 0.3. Light curves of these two types of SN tend to be heterogeneous, but generally type Ic has a faster light curve rise time than type Ib; the decline close to the peak is slower than the rise for all kind. For type Ic, the light curve ends follow an exponential decline for hundred days faster than ^{56}Co decay. Most type Ib show light curve ends steeper than type II.

Type IIP - IIL - IIn

These SNe are exclusively in spiral and irregular galaxies. In the spiral galaxy they are mainly in the arms.

The Mean Bolometric Magnitude values are (p)-16.7 ± 0.3 (l)-17.9 ± 0.3 (n)-18.7 ± 0.3.

Type IIP is a SN showing Balmer lines and a great plateau phase in its light curve that lasts 100 days. This SN type forms the most part of core collapse in the Universe. The light curve plateau is due to great radius of red super giant progenitor of about 100 ÷ 1000 solar radius and a envelope mass of at least 10 solar masses. The shock from core takes from hours to days to reach the photosphere and produces a flash into UV and X bands. The moment in which the shockwave reaches the photosphere has not observed yet. After few weeks, the outer layers cool down 10000 K and the external hydrogen starts to recombine; during this phase the external recombined hydrogen has low opacity in reverse of not recombined ones below the recombination front, so the photosphere follows the recombination front. This leads to plateau because, while the ejecta expands, the photosphere decreases, maintaining constant radius and temperature of about 5500 K and consequently the luminosity. This phase of constant luminosity is related to envelope mass and in general has a duration of 100 days. When the entire hydrogen envelope is recombined, the recombination front takes the metal rich material ejected from core and this causes a fall of luminosity; this is the plateau-tail phase.

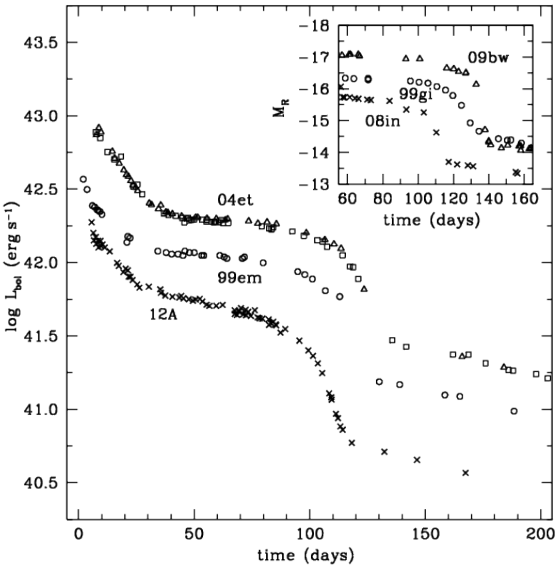


Figure 3.10: SN IIP light curve examples

A SN type IIL shows hydrogen lines in its spectra and instead of a plateau, in light curve it has a linear decline in magnitude from the maximum, although the physical nature of this linearity still remains not well defined. Most SNe of this type are more luminous than type IIP and with longer rise time to maximum. After the maximum, the slope of the light curve is halfway from ^{56}Co decay and plateau-tail phase.

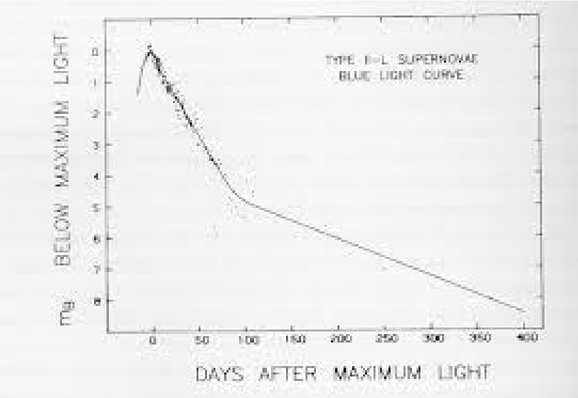


Figure 3.11: SN IIL light curve example

The nebular spectra of this type show a hydrogen emission, Mg I, O I and Ca II. Some IIL do not have this last three lines. Type IIL may represent a peculiar class of IIP, together IIn, in which there are very little plateaus or absent plateaus. Spectra also sometimes overlap these types of SNe.

Type IIn is featured by narrow spectral lines of hydrogen in the optical range; this kind of SNe is not well studied yet, but in spiral galaxies they are not strictly confined in the arms or correlate to star formation regions. Its light curve has not steep slope and this may represent some energy source like pulsar. The radiant energy is modest and this implies that phenomenology of SNe IIn include collisions with circumstellar medium that may involve multiple ejections and shell's collisions. Both a front wave shock and a reverse shock wave are present, last one propagating itself into the ejecta. In some conditions this shock may clear the broad forbidden lines that charac-

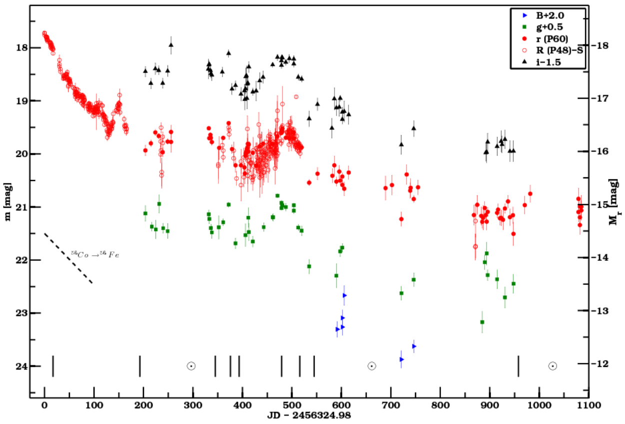


Figure 3.12: SN IIn light curve examples

terize the core collapse spectra. Type IIn shows an IR excess that may have two contributors: dust just formed after shocks and dust ejected by original star.

Superluminous SNe

Every SN brighter at peak than -21 absolute magnitude is defined Superluminous. The rate of this type of SN is 10^{-4} the core collapse rate.

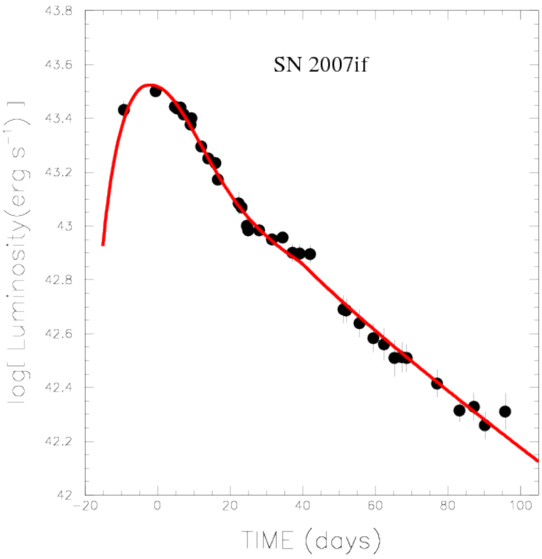


Figure 3.13: SL SN I light curve examples

Like classical SNe, if there are strong hydrogen lines into the spectra, the SN will be classified SLSNII; if not or only during the last phase, it will be classified like SLSNI. One of the mechanisms proposed for high luminosity of SLSNII is the conversion of kinetic energy into radiation by interaction with matter; another one is the interaction between more shells ejected due to pulsation of pair instability, or interaction between gamma ray burst and stellar envelope, or a magnetar or a super energetic core collapse explosion and many others. SLSNI happen in metal poor extreme emission line galaxies, which have low mass but high star formation rate and hard ionizing radiation field. This implies that origin stars have short live in a range of mass between $25 \div 120$ and the intense radiation field may cause strong mass loss and the lack of hydrogen. Its light curve has a fast rise time of some weeks and the decline is more steep than ^{5}6Co decay. Generally its spectrum has a blue continuum and many O II lines but its shape and emission lines may depend from the phase, if the shock is still propagating in the CSM or is emerged from. The best model, for now, to explain SLSNI phenomenology, is the magnetar because it is easy that a magnetar was the core of the origin star from which hydrogen and helium was stripped away. Moreover, this model predicts that an intense pulsar radiation sweeps away shuck of matter and a hot cavity within the ejecta, characterized by a flux of hard X-rays, that lead the ionization front through the ejecta.

3.3.2 RR Lyrae variables

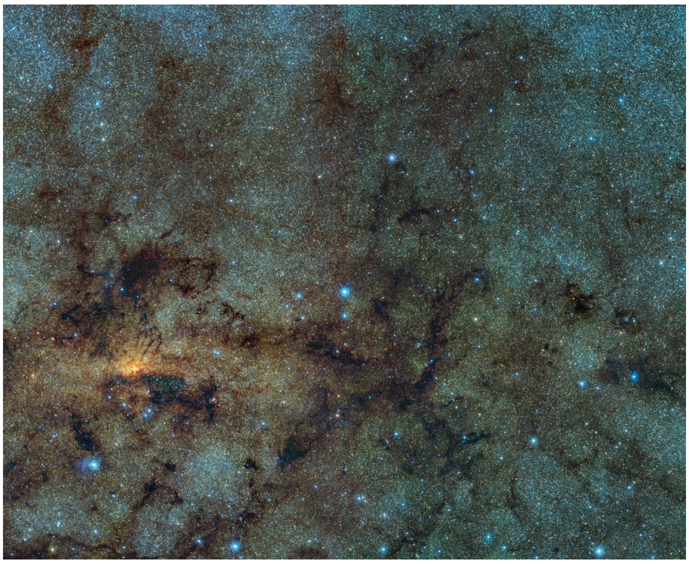


Figure 3.14: RR lyrae variable

This kind of variable stars [38] burns Helium in the core to produce Carbon and Oxygen. These stars have a radial pulsation with period lower than a day that change its luminosity. The pulsation, caused by k-mechanism, is linked to ionized helium opacity that varies with temperature. When a layer is compressed, the temperature grows up; in this case also the opacity and the pressure that expand the star grow up. By expanding the density, also the temperature is reduced, with the consequent opacity; and the star returns to standard dimension. Bailey introduced period-amplitude diagram for better studying them and this allowed to note small difference among subgroups having different periods or someone which have double pulsation.

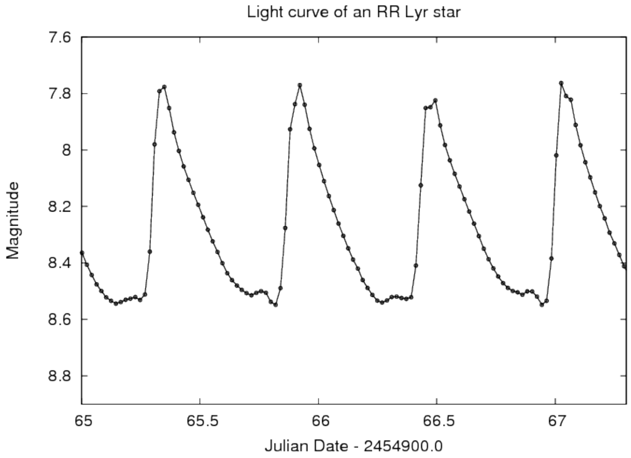
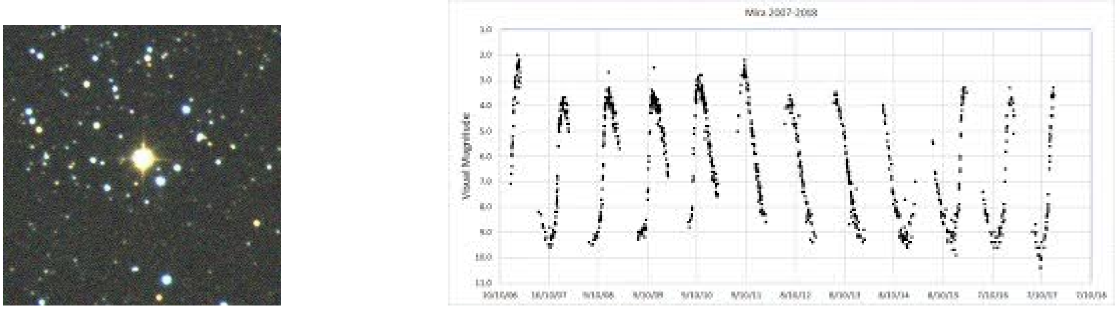


Figure 3.15: RR lyrae light curve example

The period-luminosity relation in the infrared K band is stable and known. Therefore, these stars are used as standard candles for extragalactic distances. Principally they lived into globular clusters but also into the alone and, in the Hertzsprung-Russell diagram, are located into the instability strip on the horizontal branch. The mean absolute magnitude is +0.75 and they are stars with mass lower than the sun of spectral class between A and F.



(a) Mirae variable

(b) Mirae light curve example

Figure 3.16: Mira, star and light curve

3.3.3 Mira variables

Like RR Lyrae, Mira [35] variables are stars with radial pulsation. They are very red stars with pulsation period of over 100 days and variation of even a magnitude. Most of them are red giants in the last phases of their life on asymptotic giant branch.

For their condition of red giants, they can be thousands of times brighter than the sun, but their mass do not exceed the two solar masses. A subgroup of them shows a variable time pulsation with period which can change of even a factor three in hundreds of years. These phenomena may depend by thermal pulsation linked to an hydrogen mass close to the core, which starts an independent nuclear fusion. The adaptation to this new pulsation causes the variability of the period.

3.3.4 Eclipsing binaries

Every binary system with two stars, gravitationally linked to the rotation plane close to the line of sight, is defined as eclipsing or photometric binaries [38].

There are many kinds of such systems, all periodic, with different periods from lower than a day to more than a year, with different stars from a young star to a white dwarf or neutron star.

Figure 3.17: Artistic Eclipsing Binaries

The light curve is characterized by two holes of brigthness in correspondence of the two eclipses. There are more differences between the two stars, in particular the two minima are different, related to temperature, radii, distances and inclination angle. For example, the rotation plane on a line of sight and a distances between stars small, but such that the

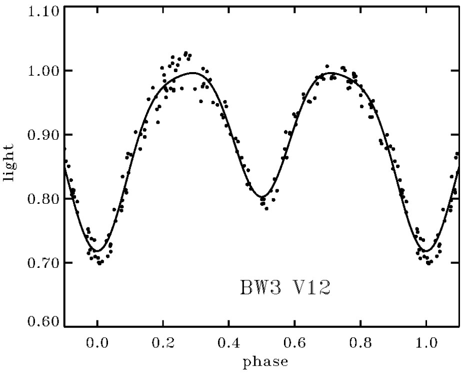


Figure 3.18: Eclipsing Binaries light curve

relative atmospheres do not exceed own Roche
lobe. Moreover, the couple formed by a white dwarf and a red giant. In this case we will have an anular eclipse and the greater minimum comes when white dwarf is eclipsed by red giant, because part of red giant is visible and contributes to total luminosity.

3.3.5 Kilonovae

When two neutron stars or two black holes merge, electromagnetic waves on all spectrum are producted [29]. After the fusion, the material, which fall on the new black hole, generates a strong beam, the gamma ray burst. Another event generated by mass onto the black hole is the kilonova, which has lower energy and life of a classical supernova; most probably this shock wave is supported by radioactive decay of heavy r-process elements.

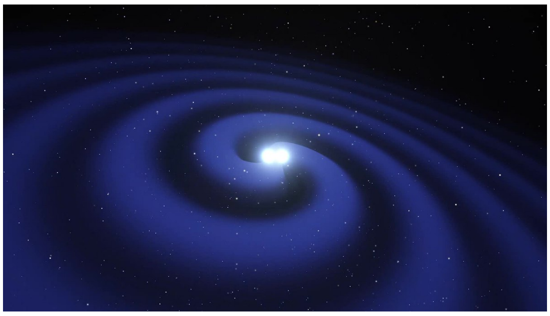


Figure 3.19: Artistic neutron stars merge

This process may increase opacity, reducing significantly the observable properties. Moreover, depending on which elements were formed during the merge, there will be a blue and a red component related to lanthanides poorness or richness of ejecta.

This was necessary, because models with one component or fixed opacity or light curve, which follows ^{56}Ni decay, fails following observations.

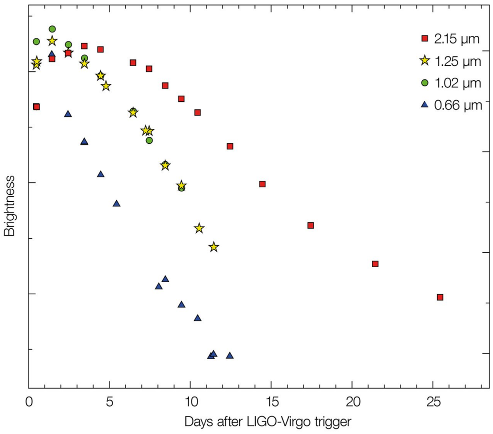


Figure 3.20: Kilonova light curve

3.3.6 Tidal Disruption Event

When a star suffers by tidal forces [24] of a supermassive black hole and the distance between them is lower than tidal radius, the difference of gravitational force on star is so strong that it is disrupted. When that happens, a flare of electromagnetic radiation is produced with maximum in UV and soft X bands and matter ends up in the accretion disk. When this happens to a compact star, together to electromagnetic radiation, also gravitational waves are generated.

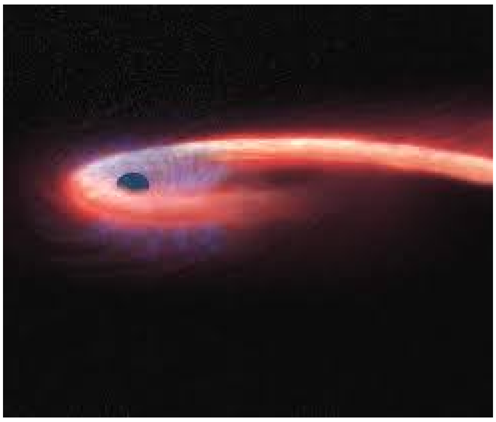
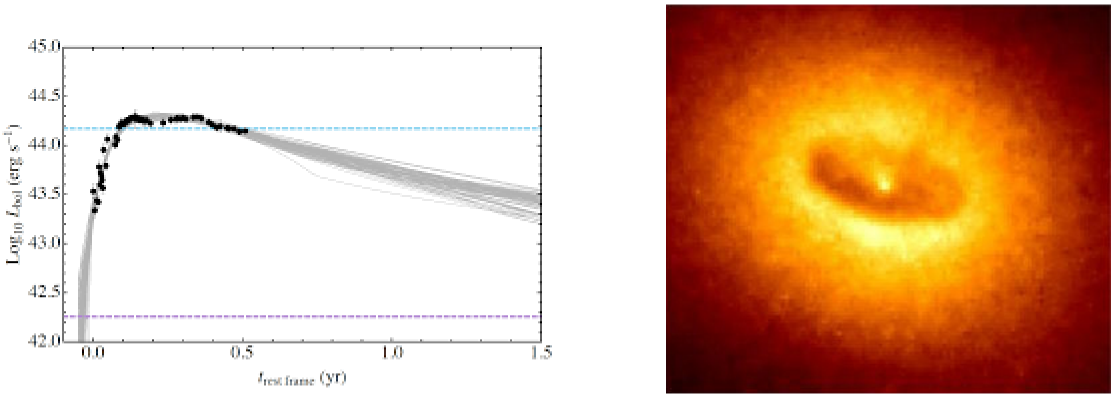


Figure 3.21: Tidal Disruption Event

This kind of phenomena allows to study many characteristics of black holes and relativistic matter, like effects in Kerr metric, physic near Eddington's luminosity, distribution of many kind of black holes and their spins, and many others.



(a)

(b)

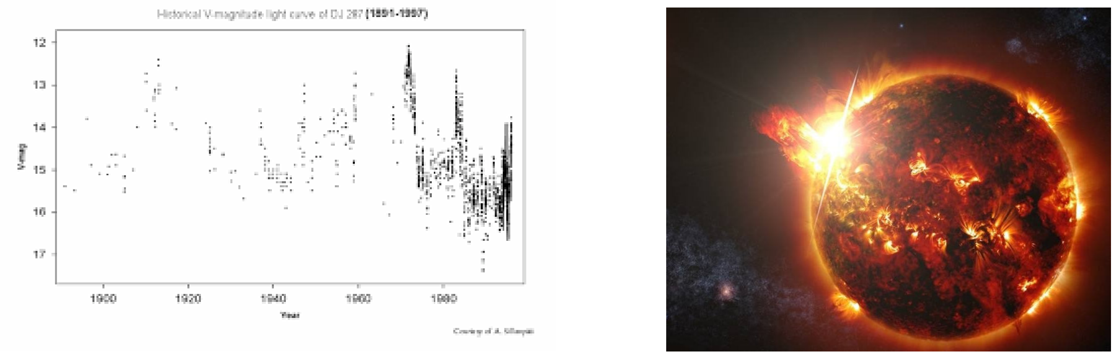
Figure 3.22: Tidal Disruption Event light curve (a) and Core of galaxy NGC 4261 (b)

3.3.7 Active Galactic Nuclei

When a galaxy has a very luminous central zone, more than a standard galaxy over the most of spectrum, and if it is clear that is not a stellar emission, we called it Active Galactic Nuclei [25].

There were many kinds of AGN, like Seyfert I, with broad lines of transition allowed and narrow lines of those forbidden (O III) with great emission in the soft X band; Seyfert II, with narrow lines both forbidden and allowed and no emission in the X soft band; Quasar, that are great radio sources, with jets and lobes containing compact objects and plasma with superluminar motions. This kind are named radio-loud, instead of radio-quiet with no jets and lobes.

At low luminosity it is difficult to distinguish quasar from Seyfert I. Another kind of AGNs is the Blazar, which emits from radio to gamma rays and the most important of this type are BL Lac, with polarized emission in optical and radio bands, with no emission lines in its continuum spectrum. All these types can be considered as the same objects, but seen in a different position: Seyfert I, if the galactic plane is on the line of sight and the disk on hot gas visual; Seyfert II, if the galactic plane is almost on the line of sight, so the broad line region is obscured by torus, although the narrow line region is visible. Quasar are more luminous Seyfert and BL Lac are AGNs with jets pointed on this direction. Only a supermassive black hole can justify this kind of power, and the accretion disk is given by all stars disrupted by tidal forces, because in the center there is a high stellar density. The spectrum appears like a power law; the jet dominates in the radio band the torus, surrounding the broad line region produced in the infrared band, while the accretion disk emits in the optical and UV bands, and, moreover, for reverse Compton with electrons, X and gamma rays by disk are emitted.



(a)

(b)

Figure 3.23: AGN oj287 light curve (a) and Artistic stellar flare (b)

3.3.8 M-Dwarf stellar flare

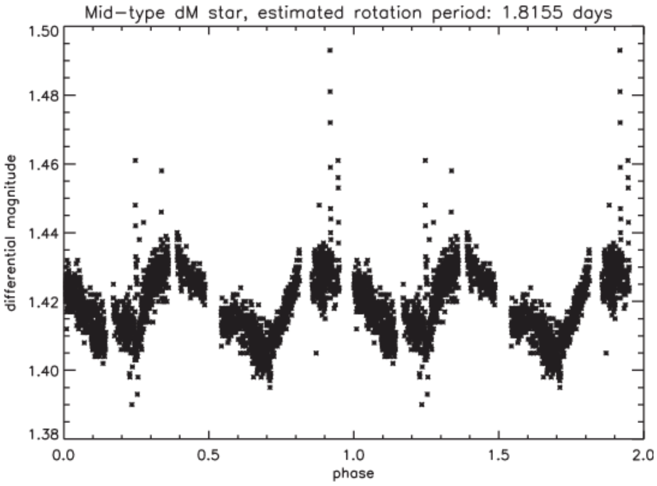
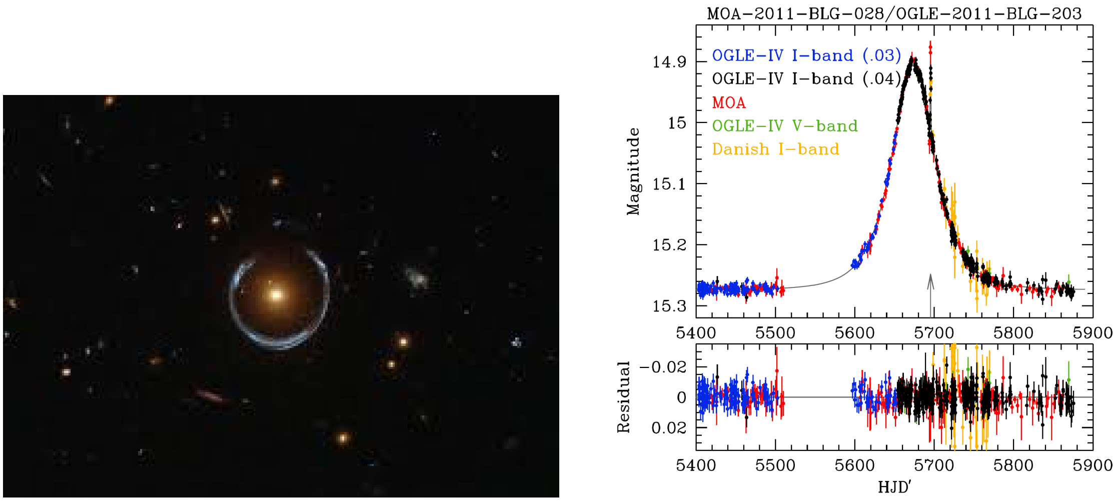


Figure 3.24: M-dwarf stellar flare light curve example

A very common phenomena for active stars is the flare [36], a great energy outburst from stellar surface that can last for hours. Its nature comes from the re-connection of the magnetic field from the inner to the outer atmosphere of the star. The spectral extension of these bursts goes from radio to gamma rays bands. There have been observed M-dwarf's flare with energy from 10^{32} to 10^{35} erg instead of $10^{29} \div 10^{32}$ erg like our sun. On M-dwarf stars, these super-flares are very common and the convective nature of this kind of stars makes stronger and more stable poloidal magnetic fields, generated by a inner turbulent dynamo effect.



(a) (b)

Figure 3.25: Gravitational lens effect (a) and Gravitational microlensing light curve example (b)

3.3.9 μ-lens from single lens

This kind of phenomena, named Gravitational Microlensing from single lens [28], occurs when an object passes in front of a background star. Therefore, when a massive object passes between observer and source, its gravitational field deflect photons coming from the source, generating distorted images of the source on the lens plane for the observer. This concentration of photons generates a magnification of the brightness of the source linked to the lens position on lens plane. It is a powerful instrument for studying stellar atmosphere, massive compact objects in the milky way halo, the structure of our galaxy, extrasolar planets and many other astrophysical phenomena.

Chapter 4

Experiments

In order to pursue main goal of the present work, related to a deep analysis of SNe in terms of their classification and characterization of the parameter space required to recognize different types, we relied on two simulation datasets, one in particular developed and specialized within the LSST project (PLAsTiCC). We preferred a statistical approach, by mapping object light curves into a set of statistical features, as done in other contexts [10]. For completeness we also tried the direct approach (i.e. by using light curves as input data) with a LSTM-based algorithm, obtaining low performances on PLAsTiCC and better results on the SNPhotCC dataset. The classification with statistical data have been performed through the comparison of different types of classifiers, respectively, Nadam, RMSProp, Adadelta and Random Forest.

A data pre-processing phase was carried out on the PLAsTiCC dataset, based on a pruning on the flux and related error, in order to reduce the amount of negative fluxes, which could in principle affect the behavior of the machine learning models. On the SNPhotCC dataset, both the errors in the flux and the quantity of negative fluxes were such that it was not deemed necessary to perform the pruning. The curves in the PLAsTiCC dataset were selected in successive steps so as to minimize the presence of negative fluxes, reaching, where possible, a subset of 35,000 light curves. In the SNPhotCC dataset, on the other hand, all the given 5088 SN Ia curves were selected and the type II curves were reduced so as to balance the classes; the other types of SNe have been discarded, due to their negligible amount available.

The sequence of classification experiments followed an incremental complexity, starting from the most simple exercise on the PLAsTiCC dataset, i.e. the separation between periodic and non-periodic objects, expected to be well classified due to their very different features within any parameter space. In terms of initial minimization of negative fluxes, it was decided to apply the following replacement: for each class of objects, the observations related to the same day were grouped, by taking the least positive flux value. This value has been replaced to all the negative fluxes of that day.

As expected, the classifiers revealed a high capability to disentangle periodic from non-periodic objects. Therefore, in all further experiments we excluded periodic sources, by focusing the exclusive attention to variable objects, increasing the complexity of classification by considering different sub-classes of transients and evaluating the performances of the selected machine learning classifiers.

The next step was to recognize the SNe from all the other non-periodic objects available in the dataset. But before we tested different methods for replacing the negative fluxes. In addition to the first method (minimum positive flux extracted from same day observations) already applied, a second method was chosen, in which negative fluxes were replaced by the constant number 0.001, considered as the absolute minimum flux emitted by the sources; we tried also a third method, in which the negative fluxes were simply excluded from the input dataset, without any replacement. In theory, such third method was considered the worst case, since it would cause a drastic reduction of the light curve sampling. As we will show, the second method (the constant minimum flux value) obtained best classification performances for all models. Therefore it was applied as reference for all further classification experiments.

The subsequent classification use cases concern some fine classifications of most interesting SNe types, starting from the classic case of SNIa Vs SNII, proceeding through a mix of SNIa Vs Superluminous SNe I, up to the most complex case, based on the multi-class experiment, in which we tried to classify all six different types of SNe in a single shot.

Besides the negative flux replacement, we investigated also the feature selection problem, in order to identify the most significant parameter space able to recognize different types of SNe. After the selection process we verified that such reduced amount of data dimensions could maintain sufficiently high the classification performances. Trying to standardize the number of features between the different use cases while respecting their statistical importance, a different feature selection was drawn up, even smaller than the first and common to almost all use cases.

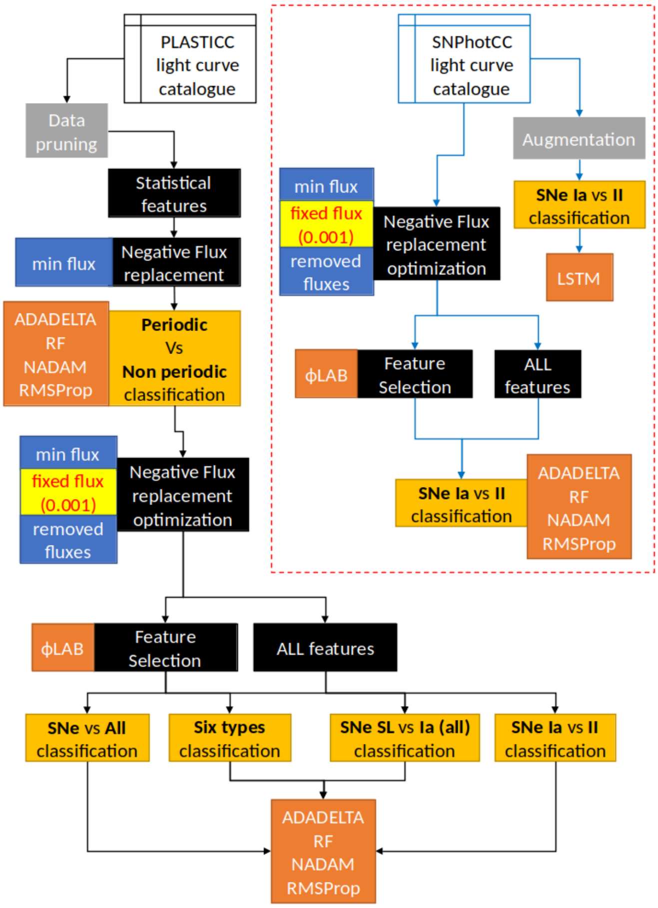


Figure 4.1: Summary of the procedure designed and followed for the experiments.

To evaluate these deductions, tests were carried out using the best setting of the parameters deriving from the experiments in the various use cases in the case with all the features. For each use case, a series of experiments was performed also with the spaces optimized to evaluate the goodness of the selection made by the algorithm.

The SNIa Vs SNII use case was also performed on the SNPhotCC dataset, this dataset being composed almost exclusively of these two types of SNe. The results were then compared with those performed on the PLAsTiCC dataset deprived of the U and Y bands, so as to obtain a fair comparison.

As shown above, in this work, five series of experiments were performed on the PLAsTiCC dataset and, only one, on the SNPhotCC dataset. They were chosen hierarchically and considering the most important goal for us: the fine classification of SNe. An overview of the different sections of this chapter is visible in figure 4.1, all ΦLAB feature importance histograms are in the Appendix A and all confusion matrices are in the Appendix B. All model parameter details are in the Appendix C, while all histograms of data distribution are in the Appendix D.

4.1 Preprocessing

For these experiments was used the PLAsTiCC dataset after the challenge, so every object has the target flag. From the whole dataset a maximum of 200,000 objects per class were randomly taken, where possible. For every class a pruning in flux and its error was done. No pruning was done on the SNPhotCC dataset.

Table 4.1 shows the limits set for pruning.

Object	Band	Flux	Flux Er.	Object	Band	Flux	Flux Er.
AGN	u	> -50	<160	M-Dwarf	u	>-60	<300
	g	>-50	<160		g	>-60	<100
	r	>-50	<160		r	>-60	<100
	i	>-50	<160		i	>-60	<80

Continued on next page

	z	>-50	<160		z	>-60	<80
	y	>-50	<160		y	>-60	<180
E. Binary	u	>-200	<800	Kilonova	u	>-10	<60
	g	>-800	<800		g	>-10	<20
	r	>-900	<800		r	>-10	<20
	i	>-800	<800		i	>-10	<25
	z	>-1100	<800		z	>-20	<40
	y	>-800	<650		y	>-30	<70
Mirae	u	>-30	<2500	μ Lens	u	>-40	<1700
	g	>-20	<800		g	>-20	<250
	r	>-50	<900		r	>-30	<400
	i	>-1200	<1700		i	>-40	<300
	z	>-8000	<3000		z	>-60	<400
	y	>-11000	<3300		y	>-90	<500
RR Lyrae	u	>-1300	<1500	SN Ia	u	>-50	<1350
	g	>-6000	<1500		g	>-20	<500
	r	>-6000	<1500		r	>-20	<400
	i	>-4500	<1500		i	>-40	<170
	z	>-4500	<1200		z	>-60	<200
	y	>-5500	<1200		y	>-100	<300
SN Iax	u	>-30	<550	SN Ia91bg	u	>-30	<800
	g	>-10	<150		g	>-20	<200
	r	>-20	<150		r	>-20	<200
	i	>-30	<100		i	>-30	<150
	z	>-50	<125		z	>-40	<150
	y	>-90	<200		y	>-90	<325
SN Ibc	u	>-50	<800	SN II	u	>-40	<200
	g	>-20	<200		g	>-20	<100
	r	>-20	<150		r	>-20	<100
	i	>-30	<100		i	>-30	<100
	z	>-60	<125		z	>-60	<100
	y	>-110	<350		y	>-110	<150
SL SN I	u	>-30	<1000	TDE	u	>-20	<200
	g	>-10	<150		g	>-10	<50
	r	>-15	<125		r	>-10	<50
	i	>-20	<100		i	>-20	<50
	z	>-40	<100		z	>-30	<75
	y	>-70	<175		y	>-60	<150

Table 4.1: Table of values preserved from pruning on the classes of PLAsTiCC.

After this first skimming, the number of objects of the various classes was reduced to a maximum of 35,000 curves. The selection for classes with more than 35K objects was driven by the choice of the curves with the least number of observations with negative fluxes and with at least 6 observations per band.

Negative fluxes have remained a problem to be addressed, so in the first instance, it was decided to try the following substitution method to prove its validity.

We have considered all the curves of a class and we checked all the observations of a given day; if in that day there was a negative or zero flux, then it was replaced with the lowest positive flux present. If in that day only negative fluxes were present, they were replaced with the lowest positive flux of the previous day. This replacement has been applied to every day, for all curves and for all classes.

An example of the replacing method is shown in table 4.2.

ID	MJD	Flux	
		Before	After
1	59820.0015	−25.154862	0.284215
2	59820.0238	15.458932	15.458932
3	59820.1234	−5.848961	0.284215
4	59820.4451	−20.548951	0.284215
5	59820.8251	0.284215	0.284215
6	59820.0234	−9.542318	0.284215
7	59820.6234	10.854215	10.854215

Table 4.2: Replacing method example.

19 features have been chosen for our statistical approach. Since PLAsTiCC has 6 bands, there are a total of 114 features.

After the statistical datasets creations, some curves showed some NaN (Not-a-Number), so they have been excluded. The total number of curves per class is reported in table 4.3.

Dataset	Object	Curves	Object	Curves
PLAsTiCC	AGN	34666	E. Binary	34484
	Kilonova	232	M-Dwarf	34849
	Mirae	1154	μ Lens	1187
	RR Lyrae	32698	SN Ia	34953
	SN Iax	34977	SN Ia 91bg	34923
	SN Ibc	34932	SN II	34828
	SL SN I	34959	TDE	14023
	Total objects		361711	
SNPhotCC	SNIa	5088	SNII	12027
	Total objects		17115	

Table 4.3: Summary table of the curves belonging to the datasets.

4.2 Periodic Vs Non Periodic

This was the first series of experiments, performed only on PLAsTiCC. Having no need, at this level, to make comparisons in terms of the treatment of negative fluxes, we have only used the method previously illustrated. We had RR lyrae, Mirae variables and Eclipsing Binaries in the periodic class (P) and all the others in the non periodic (NP) class. To balance the classes we excluded some objects in the second class, as shown in table 4.4. The percentage between training and test was set at 80-20%.

Object	Number of curves		Object	Number of curves	
	Training	Test		Training	Test
RR Lyrae	26158	6540	Kilonova	187	46
E. Binary	27587	6897	M-Dwarf	6001	1501
Mirae	923	231	μ Lens	950	238
AGN	6001	1501	SN Ia	6001	1501
SN Iax	6001	1501	SN Ia 91bg	6001	1501
SN Ibc	6001	1501	SN II	6001	1501
SL SN I	6001	1501	TDE	6001	1501
Total P Training		54668	Total NP Training		55146
Total P Test		13668	Total NP Test		13793

Table 4.4: Summary table of the curves belonging to the PLAsTiCC dataset in the P vs NP use case divided in training (80%) and test (20%) sets.

	type	RF	Nadam	RMSProp	Adadelta
Accuracy (%)	-	99	97	98	96
Purity	NP	0.99	0.97	0.99	0.95
	P	0.99	0.98	0.98	0.97
Completeness	NP	0.99	0.98	0.98	0.97
	P	0.99	0.97	0.99	0.95
F1 Score	NP	0.99	0.98	0.98	0.96
	P	0.99	0.97	0.98	0.96

Table 4.5: Summary table of the best result for the 4 algorithms. Nadam, RMSProp and Adadelta have the decay value set at 10^{-5} and the learning rate at 0.0005.

4.2.1 Classifiers

We performed, with the 4 algorithms, a series of experiment optimizing the decay value and pruning the learning rate. The best results follows in the table 4.5, while the confusion matrices are shown in the Appendix B.

This series of experiments, as expected, being the simplest given the intrinsic difference of the objects involved, did not reveal any surprises. Accuracy was very high for all algorithms: Random Forest (99%), Nadam (97%), RMSProp (98%) and Adadelta (96%). This shows a great reliability of these algorithms in recognizing periodic objects from the variable ones.

4.3 Replacement methods for negative fluxes

In both dataset we had to face, premilinary, the negative fluxes problem since some features require fluxes converted into magnitude. Therefore it was decided to approach this question in three way. The first was to replace their value as done in the previous use case. The second approach was to replace the negative fluxes with the costant value of 0.001. The third way was the complete cancellation of the negative fluxes from the dataset without any substitution. For both the Plasticc dataset with the SN Vs All use case and for the SNPhotCC dataset with the SNIa Vs SNII use case, a comparison was performed on all three methods. So all the classes of objects have been treated with the three types of substitution and this has given rise to different numbers of objects per class. The entire composition of the datasets for the three methods is shown in table 4.6, instead the composition of the classes of SN, All, SNIa and SNII are shown in table 4.7. The difference of objects between the first and second method is very peculiar, being in fact present only for the objects of the classes SNIa and SNII of PLAsTiCC dataset.

	Object	Number of curves			Object	Number of curves		
		1° method	2° method	3° method		1° method	2° method	3° method
PLAsTiCC	AGN	34666	34666	34082	Kilonova	232	232	229
	μ Lens	1187	1187	1144	M-Dwarf	34849	34849	34191
	SN Ia	34953	34891	34423	SL SN I	34959	34959	34750
	SN Iax	34977	34977	34680	SN Ia 91bg	34923	34923	34559
	SN Ibc	34932	34932	34437	SN II	34828	34771	34393
	TDE	14023	14023	13985				
Total objects 1° method		294529	Total objects 2° method		294410	Total objects 3° method		290873
SNPhotCC	SNIa	5088	5088	5086	SNII	5088	5088	5077
Total objects 1° method		10176	Total objects 2° method		10176	Total objects 3° method		10163

Table 4.6: Summary table of the light curves belonging to the datasets for each replacing method.

		1° method		2° method		3° method	
	Object	Training	Test	Training	Test	Training	Test
PLAsTiCC	AGN	27732	6934	27732	6934	27266	6816
	Kilonova	186	46	186	46	183	46
	μ Lens	949	238	949	238	915	229
	M-Dwarf	27879	6970	27879	6970	27353	6838
	SN Ia	12001	3001	11975	2994	11802	2954
	SL SN I	12001	3001	12001	3001	11935	2979
	SN Iax	12001	3001	12001	3001	11900	2976
	SN Ia 91bg	12001	3001	12001	3001	11866	2975
	SN Ibc	12001	3001	12001	3001	11828	2951
	SN II	12001	3001	11983	2992	11835	2970
	TDE	11218	2805	11218	2805	11188	2797
	Total SN	72006	18006	71962	17990	71166	17805
	Total All	67964	16993	67964	16993	66905	16726
SNPhotCC	SNIa	4071	1017	4071	1017	4062	1016
	SNII	4071	1017	4071	1017	4070	1015

Table 4.7: Summary table of the light curves belonging to the dataset divided into training and test sets for each replacing method.

An experiment was performed by algorithm, both as regards the PLAsTiCC dataset and the SNPhotCC dataset, with all the features. For Nadam, RMSProp and Adadelta was chosen a decay value of 10^{-5}. The results are shown in tables 4.8 and 4.9, where P,C and F1 stands for Purity, Completeness and F1-score.

Analyzing the results we see that on average, as regards the PLAsTiCC Dataset, the second method is better than the others. As for the SNPhotCC dataset, on the other hand, the second and third methods are on average very similar, however, since we intend to compare the two datasets, we have chosen to prefer the second method also in this dataset.

Dataset	Use case	Algorithm	Class	Parameter	M1	M2	M3
PLAsTiCC	SN Vs All	RF	SN	P	0.86	0.91	0.85
				C	0.94	0.93	0.91
				F1	0.90	0.92	0.88
			All	P	0.93	0.92	0.90
				C	0.83	0.90	0.83
				F1	0.88	0.91	0.86
		Nadam	SN	P	0.77	0.84	0.83
				C	0.82	0.78	0.85
				F1	0.79	0.81	0.84
			All	P	0.80	0.78	0.84
				C	0.73	0.85	0.82
				F1	0.76	0.81	0.83
		RMSProp	SN	P	0.85	0.89	0.87
				C	0.83	0.89	0.91
				F1	0.84	0.89	0.89
			All	P	0.83	0.88	0.90
				C	0.85	0.89	0.86
				F1	0.84	0.88	0.88
		Adadelta	SN	P	0.80	0.85	0.85
				C	0.84	0.86	0.87
				F1	0.82	0.86	0.86
			All	P	0.82	0.85	0.86
				C	0.78	0.84	0.84
				F1	0.80	0.85	0.85

Table 4.8: Summary table of the comparison between the three methods of replacing negative fluxes on the PLAsTiCC dataset. For Nadam, RMSProp and Adadelta a learning rate of 0.001 and a decay value of 10^{-5} was set.

Dataset	Use case	Algorithm	Class	Parameter	M1	M2	M3
SNPhotCC	SNIa Vs SNII	RF	SNIa	P	0.91	0.95	0.91
				C	0.94	0.97	0.93
				F1	0.93	0.96	0.92
			SNII	P	0.94	0.97	0.93
				C	0.91	0.95	0.91
				F1	0.92	0.96	0.92
		Nadam	SNIa	P	0.86	0.91	0.92
				C	0.92	0.92	0.94
				F1	0.89	0.92	0.93
			SNII	P	0.91	0.92	0.94
				C	0.86	0.91	0.92
				F1	0.88	0.91	0.93
		RMSProp	SNIa	P	0.91	0.92	0.93
				C	0.93	0.96	0.94
				F1	0.92	0.94	0.94
			SNII	P	0.93	0.96	0.94
				C	0.91	0.92	0.93
				F1	0.92	0.94	0.94
		Adadelta	SNIa	P	0.89	0.86	0.92
				C	0.92	0.88	0.92
				F1	0.91	0.87	0.92
			SNII	P	0.92	0.88	0.92
				C	0.89	0.85	0.92
				F1	0.90	0.87	0.92

Table 4.9: Summary table of the comparison between the three methods of replacing negative fluxes on the SNPhotCC dataset. For Nadam, RMSProp and Adadelta a learning rate of 0.001 and a decay value of 10^{-5} was set.

4.4 Parameter space analysis

After choosing how to deal with the issue of negative fluxes, we wanted to analyze the parameter space as the use cases change. To do this, we applied the ΦLAB algorithm to the datasets created for the different use cases (except Periodic Vs Non Periodic) and we obtained an optimized parameter space for each of them. The summary tables of this feature selection are shown in appendix A, while the optimized subsets are shown in table 4.10. The feature selection of the SNIa Vs SNII use case of the PLAsTiCC dataset, deprived of the U and Y bands, will also be used for comparison with the SNPhotCC dataset, since if a feature is irrelevant in a 6-band dataset, it will also be irrelevant in one with 4 bands.

Feature	Band	1	2	3	4	5	Feature	Band	1	2	3	4	5
Ampl	u	X	X	-	X	X	B1std	u	X		-		
	g	X	X	X	X	X		g			X		
	r	X	X	X	X	X		r	X	X	X	X	X
	i	X	X	X	X	X		i	X	X	X	X	X
	z	X	X	X	X	X		z	X		X	X	X
	y	X	X	-	X	X		y	X		-	X	X
Fpr20	u	X		-	X	X	Fpr35	u	X	X	-	X	X
	g	X	X	X	X	X		g	X	X	X	X	X
	r	X	X	X	X	X		r	X	X	X	X	X
	i	X	X	X	X	X		i	X	X	X	X	X
	z	X	X	X	X	X		z	X	X	X	X	X
	y	X	X	-	X	X		y	X	X	-	X	X
Fpr50	u	X	X	-	X	X	Fpr65	u	X	X	-	X	X
	g	X	X	X	X	X		g	X	X	X	X	X
	r	X	X	X	X	X		r	X	X	X	X	X
	i	X	X	X	X	X		i	X	X	X	X	X
	z	X	X	X	X	X		z	X	X	X	X	X
	y	X	X	-	X	X		y	X	X	-	X	X
Fpr80	u	X	X	-	X	X	Kurt	u	X	X	-	X	X
	g	X	X	X	X	X		g	X	X	X	X	X
	r	X	X	X	X	X		r	X	X	X	X	X
	i	X	X	X	X	X		i	X	X	X	X	X
	z	X	X	X	X	X		z	X	X	X	X	X
	y	X	X	-	X	X		y	X	X	-	X	X
Ls	u	X		-		X	Lt	u	X	X	-	X	X
	g	X	X	X	X	X		g	X	X	X	X	X
	r	X	X	X	X	X		r	X	X	X	X	X
	i	X	X	X	X	X		i	X	X	X	X	X
	z	X	X	X	X	X		z	X	X	X	X	X

Continued on next page

	y	X	X	-	X	X		y	X	X	-	X	X
MAD	u	X	X	-	X	X	Mbrp	u			-		
	g	X	X	X	X	X		g			X		
	r	X	X	X	X	X		r					
	i	X	X	X	X	X		i			X		
	z	X	X	X	X	X		z			X		
	y	X	X	-	X	X		y			-	X	
Mr	u			-			Ms	u	X	X	-	X	X
	g							g	X	X	X	X	X
	r							r	X	X	X	X	X
	i				X			i	X	X	X	X	X
	z				X			z	X	X	X	X	X
	y			-	X	X		y	X	X	-	X	X
Pdfp	u	X	X	-	X	X	Pst	u			-	X	
	g	X	X	X	X	X		g					
	r	X	X	X	X	X		r	X	X	X	X	X
	i	X	X	X	X	X		i	X		X	X	X
	z	X	X	X	X	X		z	X		X	X	X
	y	X	X	-	X	X		y			-	X	X
Rcb	u			-		X	Sk	u	X	X	-	X	X
	g			X				g	X	X	X	X	X
	r	X		X				r	X	X	X	X	X
	i	X		X	X			i	X	X	X	X	X
	z	X		X	X	X		z	X	X	X	X	X
	y			-	X	X		y	X	X	-	X	X
Std	u	X	X	-	X	X							
	g	X	X	X	X	X							
	r	X	X	X	X	X							
	i	X	X	X	X	X							
	z	X	X	X	X	X							
	y	X	X	-	X	X							

Table 4.10: Summary table of the features subsets in the different experiment series; SN Vs All(1), SN Ia Vs SN II (PLAsTiCC)(2), SN Ia Vs SN II (SNPhotCC)(3), SLSN I Vs SN Ia mixed(4) and Six Class Problem(5).

For each feature selection performed, the cumulative sum of the importance of the features deemed relevant and its normalization compared to the cumulative importance of the last relevant features were calculated. For our analysis we have considered a neighborhood of the quartiles of normalization, so as to be able to divide our analysis into different steps. A summary table of the 4 use cases of the PLAsTiCC dataset has been compiled to see the total number of occurrences for each features on all use cases. A comparative table between SNPhotCC and PLAsTiCC datasets on SNIa Vs SNII use case of common features in the different quartiles has been compiled. All these tables are shown in the appendix A.

The best compromise between the same number of features and common statistics in various cases is shown in table 4.11 and corresponds to 78 features for UGRIZY cases and 52 for GRIZ cases. An anomaly occurred in the six-class problem and this can be explained by the fact that in that case there are 6 classes instead of 2 as in the other use cases.

Feature	1	2	3	4	5	Feature	1	2	3	4	5
Ampl$_X$	X	X	X	X	X	Skew$_X$	X	X	X	X	X
Pdfp$_X$	X	X	X		X	FprYY$_X$	X	X	X	X	X
Ms$_X$				X		Kurt$_X$	X	X	X	X	X
MAD$_X$	X	X	X	X	X	Ls$_X$	X	X	X	X	X
Std$_X$	X	X	X	X	X	Lt$_X$	X	X	X	X	X

Table 4.11: Summary table of the features chosen for the feature selection analysis. Each feature is intended to include all available bands. The use cases are: SN Vs All (PLAsTiCC)(UGRIZY)(1),SNIa Vs SNII (PLAsTiCC)(UGRIZY)(2), SLSNI Vs SNIa mixed (PLAsTiCC)(UGRIZY)(3), Six class problem (PLAsTiCC)(UGRIZY)(4), SNIa Vs SNII (SNPhotCC)(GRIZ)(5).

This choice was obtained keeping in mind that on the one hand it was important to minimize any percentage of discards, on the other to maximize the presence of features that had occurrences starting from 25% up to 100%; without ignoring the percentages of overall importance that were lost by excluding the different features. This type of analysis showed uniformity of features between the two datasets, so we can unify the discourse of light curve analysis without distinction between datasets.

Extremely important is the presence of some features within the first quartile and that this presence is common to most use cases; in particular the Amplitude (Ampl), which shows 50% of the total occurrences, reveals that the role is crucial for the analysis of these use cases, and in particular it can be said for the classification of SNe. Moving into the second quartile, much importance is assumed by the Standard Deviation (Std), which reaches 79.2% of occurrences. While reaching the third quartile all the other features emerge that we have decided to introduce in our selection. Equally important is the result for which the Median Buffer Range

Percentage (Mbrp), the Magnitude Ratio (Mr) and the R Cor Bor (Rcb) have been rejected with high percentages, 95.8%, 83.3% and 58.3% respectively.

In all datasets the average value of the Mbrp which is the percentage of points in an interval of 10% of the median flux from the median flux, is very high and close to the unit with standard deviation, generally, of a smaller order of magnitude. This show that most of the light curves are relatively contained in flux extension. The Mr feature representing the percentage of points above the median magnitude has values, for all use cases, greater than 40% with a standard deviation of a lower order of magnitude; except in the case of the problem with six classes in which the standard deviation is comparable with the Mr value. This show that most of the light curves are basically symmetrical in magnitude. The Rcb has an average value of about 30% with a standard deviation comparable to that value. So its value ranges over the whole spectrum of possible values without any class distinction.

The Ampl is the most important feature for these use cases and this is related to the different distribution in the classes. In the use case SN Vs All, the class of SNe shows a bimodal distribution, while the class All shows a distribution sometimes bimodal, sometimes unimodal, with different peaks from the SNe distributions. In the SL Vs SNIa mixed use case, the SNIa have a bimodal distribution, unlike the SL which instead is unimodal. The use case with the six class problem, show that the SNeIa have a different peak from the Iax, the Iabg91, the SL SN and the SN Ibc , which instead have a similar peak value. The SNe II instead have the peak similar the SNe Ia, and this should explains why this feature is present in smaller quantities in the first quartile of the SNIa Vs SNII use case compared to the others. From a physical point of view this feature shows the half-amplitude, in magnitude, of the light curves.

The Std, deviation from the mean flux, has the same trend as the Ampl, with bimodal and unimodal distributions with peaks at different values.

The Fpr, Flux Percentage Ratio, shows in the SN Vs All case and in the GRIZ bands with the major percentile intervals, that there are two distributions with distinguishable peaks. In the six class use case, the RIZ bands with the wider flux ratios contribute to solving the envelope of the 6 classes. In the SL Vs SNIa mixed use case, the different distributions can be identified particularly in the RIZY bands, again in the broader flux ratios such as 50, 65 and 80. Finally, in the SNIa Vs SNII use case the distinction is more complex and only in few RIZ band reports it is possible to see the two distributions. This feature represents the relationship between two differences; in the denominator the two extreme percentile fluxes (fifth percentile and 95-th percentile) and in the numerator the two percentile fluxes representing a range of light curve values. This feature is related to the sampling of the light curve which assumes its importance with the higher flux values, indeed the greater relevance features are the last 3 that embrace a wider range of values.

In the other relevant features we do not infer distinct distributions in the various use cases, but only different fluctuations around the same distribution. This means that all the curves of all the classes have more or less the same distribution as regards the flatness of the curve (Kurt), the symmetry of the curve (Skew), the slope deriving from the linear fit (Lt), the period obtained from the peak frequency of the Lomb Scargle Periodogram (Ls), the ratio between the difference in percentile magnitude (95-5) and the median (Pdfp) and finally the median of deviations from the median (MAD). Since these features have proved to be the most important, this implies that those fluctuations in class distributions contribute substantially to the classifier analysis. As far as the six-class use case is concerned, another feature has importance, and it is the maximum difference in magnitude between two successive epochs (Ms), which slightly in the U band and in a more consistent way in the Y one provides fluctuations relevant to the resolution of the problem.

To verify the validity of our analysis, we performed a test by use case with the decay value and learning rate that best performs using all the features. The results are shown in the tables of the respective use cases.

4.5 SNe Vs All

In this use case we had SNe type Ia, Iax, Ia 91bg-like, Ibc, II and SL SNe I in the SNe class and all the other objects, minus the periodic ones, in the All class. We performed the experiments with the 4 algorithms using all the features and the 95 best ones that optimize the parameter space. The objects were balanced as shown in table 4.12. The percentage between training and test was sets 80% and 20%, respectively.

Object	Training	Test
SN Ia	11975	2994
SN Iax	12001	3001
SN Ia91bg	12001	3001
SN Ibc	12001	3001
SN II	11983	2992
SL SN I	12001	3001
Kilonova	186	46
M-Dwarf	27879	6970
μ Lens	949	238
TDE	11218	2805
AGN	27732	6934
Total SN	71962	17990
Total All	67964	16993

Table 4.12: Summary table of the curves belonging to the dataset divided in training (80%) and test (20%) sets.

4.5.1 Classifiers

We performed, with the 4 algorithms, a series of experiment optimizing the decay value and pruning the learning rate for all the features and the 95 best ones. Between Nadam, RMSProp and Adadelta, the best performances were obtained with the RMSProp both in the case with all the features and with the optimized space. Therefore the results, which follow in table 4.13, for these three algorithms are reported using the best parameters for the RMSProp. The confusion matrices are shown in the Appendix B.

		Random Forest			Nadam			RMSProp			Adadelta		
	type	All	95	78	All	95	78	All	95	78	All	95	78
Accuracy (%)	-	92	92	92	84	84	84	89	89	89	84	84	84
Purity	SN	0.91	0.91	0.91	0.83	0.84	0.84	0.89	0.90	0.89	0.84	0.83	0.83
	All	0.92	0.92	0.92	0.86	0.84	0.84	0.88	0.87	0.88	0.85	0.85	0.85
Completeness	SN	0.93	0.93	0.93	0.87	0.85	0.85	0.89	0.88	0.89	0.86	0.87	0.87
	All	0.90	0.90	0.90	0.82	0.83	0.83	0.89	0.90	0.88	0.83	0.81	0.81
F1 Score	SN	0.92	0.92	0.92	0.85	0.85	0.85	0.89	0.89	0.89	0.85	0.85	0.85
	All	0.91	0.91	0.91	0.84	0.84	0.84	0.88	0.89	0.88	0.84	0.83	0.83

Table 4.13: Summary table of the best result for the 4 algorithms with all the features, the 95 best and the 78 best ones. RMSProp has 10^{-5} and 0.0005 as best values for the decay and the learning rate, respectively.

The best classifier for this use case is the Random Forest, across all 3 parameter spaces. For each classifier, the performances on the 3 different parameter spaces are equivalent.

4.5.2 Contamination level of the SN class and redshift distributions of SNe and AGNs

In this section, the error percentages of all object classes referring to the best experiment with all the features of the Random Forest are shown; moreover for the same experiment we show the SNe redshift distribution for every class in relation to the wrong classified one. The same we have done for the AGNs, due to their interesting cosmological aspects. The classes contamination are shown in table 4.14. In figures from 4.2 to 4.8 the redshift distributions of AGN and SNe are shown. In these histograms the test object distribution and the subset of wrong classified ones were superimposed.

Class	Total	Correct	Wrong	% Wrong
AGN	6934	6907	27	≈0
M-Dwarf	6970	5837	1133	16
KN	46	45	1	2
μ Lens	238	215	23	10
TDE	2805	2352	453	16
SN Ia	2994	2990	4	≈0
SN Iax	3001	2638	363	12
SN Ia91bg	3001	2648	353	12
SN Ibc	3001	2720	281	9
SN II	2992	2989	3	≈0
SL SN I	3001	2712	289	10

Table 4.14: Summary table of the contamination during the Random Forest classification best experiment.

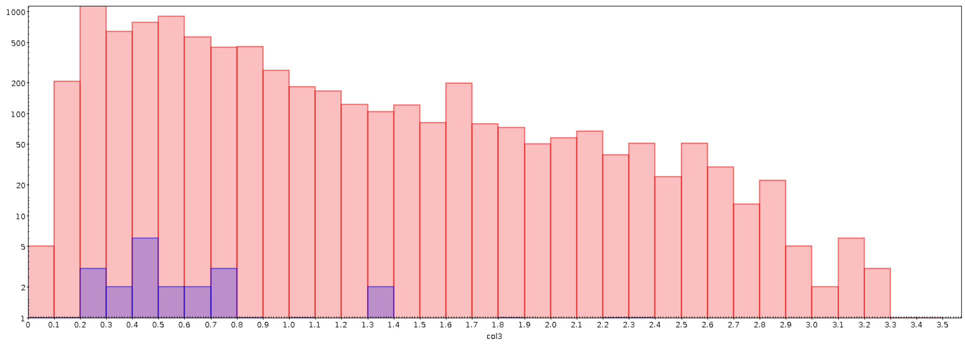


Figure 4.2: AGN redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

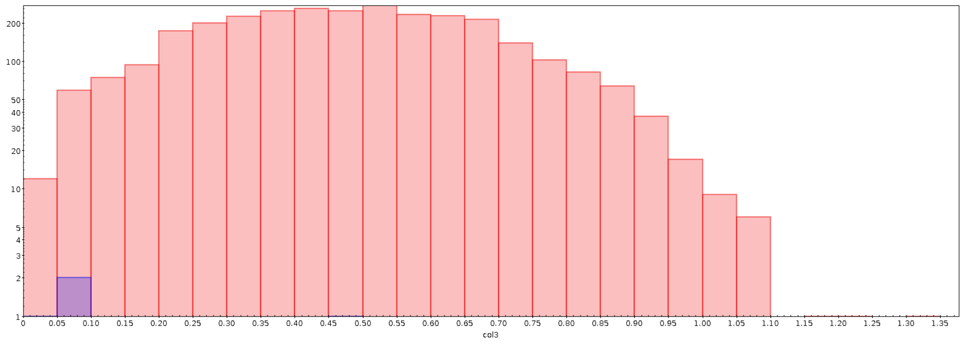


Figure 4.3: SNIa redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

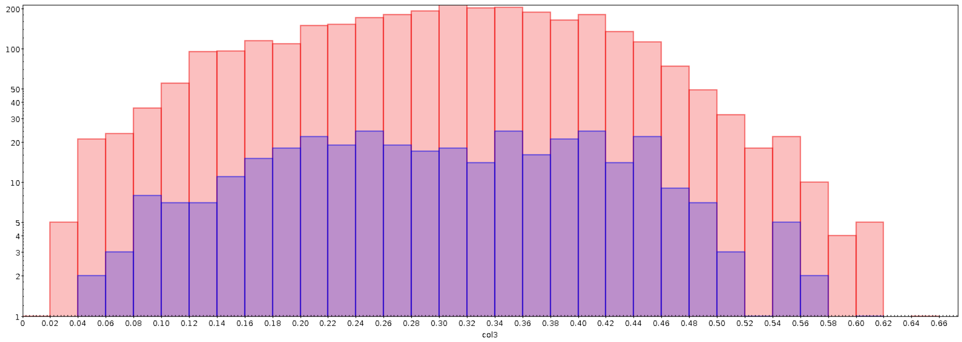


Figure 4.4: SNIabg redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

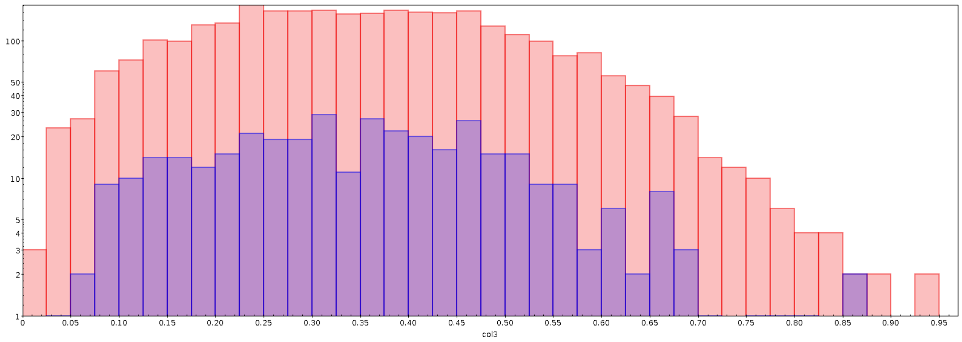


Figure 4.5: SNIax redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

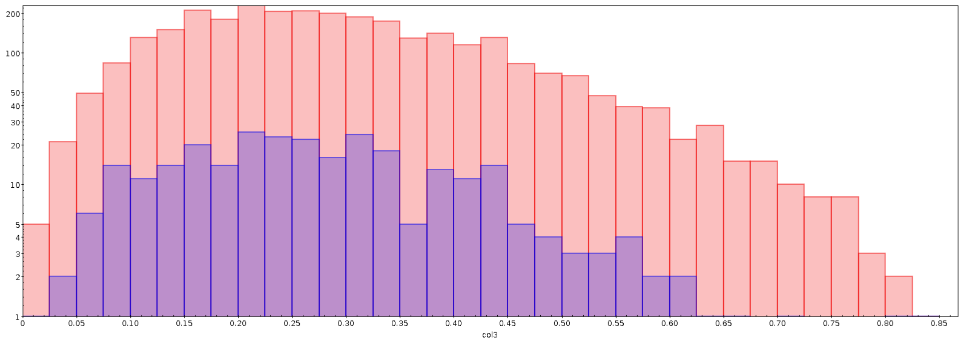


Figure 4.6: SNIbc redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

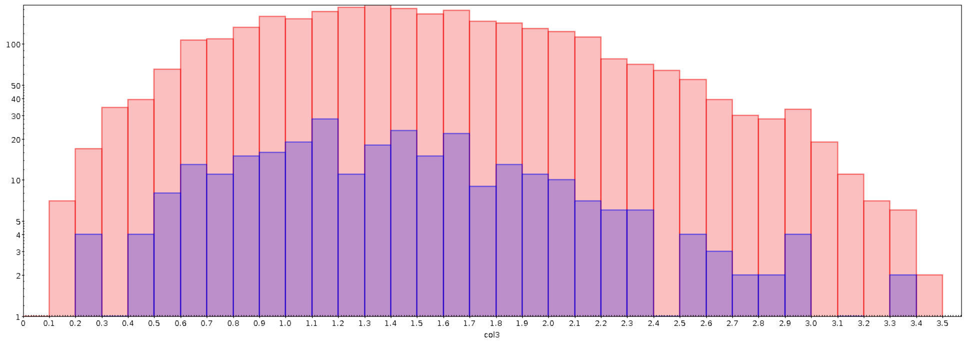


Figure 4.7: SL SN I redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

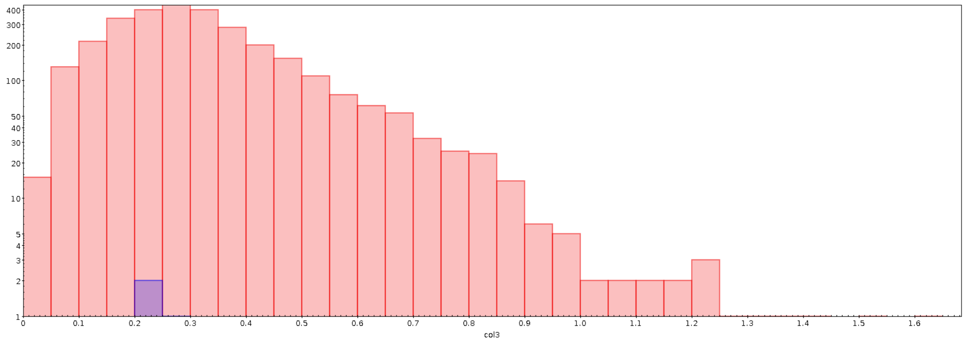


Figure 4.8: SNII redshift distribution, with test set (red) and wrong classified (blue) superimposed. The y axis is on a logarithmic scale.

The worst ranked classes are the M-Dwarf and the TDE in the All class and the SNIa91bg and SNIax in the SN class, while the SNIa, SNII and AGN objects are classified exceptionally well. All SNe and AGNs badly classified, as seen from the histograms, show no relation with the redshift despite a correlation of these two parameters was expected.

4.6 SNe Ia Vs SNe II

In this series of experiments we considered only SNe type Ia and II. The balance of the objects is shown in table 4.15, with a percentage between training and test set, imposed at 80% - 20%, respectively. In this use case we performed the experiments with the 4 algorithms using all the features and the 85 best ones that optimize the parameter space. A comparison between the SNPhotCC and PLAsTiCC datasets was performed, using all the features and their optimized parameter spaces on GRIZ bands.

Dataset	Class	Number of curves	
		Training	Test
PLAsTiCC	SN Ia	27964	6990
	SN II	27983	6966
	Total	55947	13956
SNPhotCC	SN Ia	4071	1017
	SN II	4071	1017
	Total	8142	2034

Table 4.15: Summary table of the curves belonging to the datasets divided in training (80%) and test (20%) sets.

4.6.1 Classifiers

A series of experiment, optimizing the decay value and pruning the learning rate for all the features and the 85 best ones, was performed with the 4 algorithms. Between Nadam, RMSProp and Adadelta, the best performances were obtained with the RMSProp both in the case with all the features and with the optimized space. Therefore the results, which follow in table 4.16, for these three algorithms

		Random Forest			Nadam			RMSProp			Adadelta		
	type	All	85	78	All	85	78	All	85	78	All	85	78
Accuracy (%)	-	78	78	79	71	70	71	75	75	74	72	73	73
Purity	Ia	0.76	0.76	0.76	0.71	0.70	0.70	0.73	0.73	0.72	0.71	0.72	0.73
	II	0.81	0.81	0.81	0.72	0.71	0.73	0.77	0.76	0.76	0.73	0.74	0.74
Completeness	Ia	0.82	0.82	0.83	0.72	0.71	0.74	0.79	0.77	0.78	0.75	0.76	0.75
	II	0.74	0.74	0.74	0.71	0.70	0.68	0.71	0.72	0.70	0.69	0.71	0.72
F1 Score	Ia	0.79	0.79	0.79	0.72	0.71	0.72	0.76	0.75	0.75	0.73	0.74	0.74
	II	0.77	0.77	0.78	0.71	0.70	0.70	0.74	0.74	0.73	0.71	0.72	0.73

Table 4.16: Summary table of the best result for the 4 algorithms with all the features, the 85 and the 78 best ones. RMSProp has a different pair of best values for the decay and the learning rate depending on the parameter space used; 10^{-5} and 0.0005 for the case with all and 78 features; 10^{-5} and 0.001 for the case with the 85 best features.

are reported using the best parameters for the RMSProp. The confusion matrices are shown in the Appendix B.

The Random Forest was the best classifier with all 3 parameter spaces in this use case. All four classifiers do not have a difference in performance that exceeds the percentage point between the 3 different parameter spaces.

4.6.2 SNPhotCC

In this dataset, 5088 light curves for SNIa and SNII were used, with a $80 - 20$ proportion between training and test set. The data are reported in table 4.15. We performed the experiments on GRIZ bands with the 4 algorithms using all the features and the 70 best ones that optimize the parameter space. We then compared the best results with the best of the PLAsTiCC dataset of the same use case in which only the GRIZ bands were used. A first comparison is on all the features, a second is on the distinct optimized parameter spaces, while a third is on the same subset of features coming from our statistical analysis.

The results on SNPhotCC dataset are reported in the table 4.17, instead the comparative results are in table 4.18.

		Random Forest			Nadam			RMSProp			Adadelta		
	type	All	70	52	All	70	52	All	70	52	All	70	52
Accuracy (%)	-	96	96	96	91	92	94	94	94	95	87	89	92
Purity	Ia	0.95	0.95	0.95	0.89	0.91	0.93	0.92	0.92	0.94	0.86	0.87	0.91
	II	0.97	0.97	0.97	0.92	0.93	0.94	0.95	0.96	0.96	0.89	0.91	0.94
Completeness	Ia	0.97	0.97	0.97	0.93	0.93	0.94	0.95	0.96	0.96	0.90	0.92	0.94
	II	0.95	0.95	0.95	0.89	0.91	0.93	0.92	0.91	0.93	0.85	0.86	0.91
F1 Score	Ia	0.96	0.96	0.96	0.91	0.92	0.94	0.94	0.94	0.95	0.88	0.89	0.92
	II	0.96	0.96	0.96	0.90	0.92	0.94	0.93	0.93	0.95	0.87	0.89	0.92

Table 4.17: Summary table of the best result for the 4 algorithms with all the features, the 70 and the 52 best ones. RMSProp has a different pair of best values for the decay and the learning rate depending on the parameter space used; 10^{-5} and 0.001 for the case with all and 52 features; 10^{-7} and 0.001 for the case with the 70 best features.

Number of Features		type	RF		Nadam		RMSProp		Adadelta	
			PL	SN	PL	SN	PL	SN	PL	SN
All	Accuracy (%)	-	78	96	70	91	75	94	73	87
	Purity	Ia	0.75	0.95	0.68	0.89	0.73	0.92	0.72	0.86
		II	0.80	0.97	0.73	0.92	0.77	0.95	0.74	0.89
	Completeness	Ia	0.82	0.97	0.76	0.93	0.79	0.95	0.76	0.90
		II	0.73	0.95	0.65	0.89	0.70	0.92	0.70	0.85
	F1 Score	Ia	0.79	0.96	0.72	0.91	0.76	0.94	0.74	0.88
		II	0.77	0.96	0.69	0.90	0.73	0.93	0.72	0.87
Optimized space	Accuracy (%)	-	78	96	70	92	74	94	73	89
	Purity	Ia	0.76	0.95	0.69	0.91	0.73	0.92	0.72	0.87
		II	0.80	0.97	0.71	0.93	0.76	0.96	0.74	0.91
	Completeness	Ia	0.82	0.97	0.73	0.93	0.78	0.96	0.75	0.92
		II	0.73	0.95	0.67	0.91	0.71	0.91	0.71	0.86
	F1 Score	Ia	0.79	0.96	0.71	0.92	0.75	0.94	0.74	0.89
		II	0.77	0.96	0.69	0.92	0.73	0.93	0.72	0.89
52	Accuracy (%)	-	78	96	70	94	74	95	73	92
	Purity	Ia	0.76	0.95	0.71	0.93	0.71	0.94	0.72	0.91
		II	0.80	0.97	0.69	0.94	0.78	0.96	0.74	0.94
	Completeness	Ia	0.82	0.97	0.68	0.94	0.81	0.96	0.76	0.94
		II	0.74	0.95	0.72	0.93	0.68	0.93	0.70	0.91
	F1 Score	Ia	0.79	0.96	0.69	0.94	0.76	0.95	0.74	0.92
		II	0.77	0.96	0.71	0.94	0.72	0.95	0.72	0.92

Table 4.18: Comparative table of the results of the 4 algorithms on the PLAsTiCC and SNPhotCC datasets on three different parameter spaces; PLAsTiCC (76-59-52) and SNPhotCC (76-70-52).

The best performing algorithm is the Random Forest. The results on SNPhotCC dataset compared with the ones on PLAsTiCC dataset are better, both on 4-band PLAsTiCC and on 6-band PLAsTiCC.

4.7 Superluminous SNe I vs Ia mixed

For this series of experiments all three classes of SNe Ia have been mixed in the same percentage and then classified against Superluminous SNe I. We performed the experiments with the 4 algorithms using all the features and the 99 best ones that optimize the parameter space. The balanced classes are shown in table 4.19. The percentage between training and test sets was imposed at 80% and 20%, respectively.

Class	Number of curves	
	Training	Test
SN Ia	9323	2331
SN Iax	9323	2331
SN Ia91bg	9323	2331
SLSN I	27967	6992
Total Ia	27969	6993
Total SL	27967	6992

Table 4.19: Summary table of the curves belonging to the dataset divided in training (80%) and test (20%) sets.

4.7.1 Classifiers

With the 4 algorithms, a series of experiment optimizing the decay value and pruning the learning rate was performed, for all the features and the 99 best ones. Between Nadam, RMSProp and Adadelta, the best performances were obtained with the RMSProp both in the case with all the features and with the optimized space. The best results follows in the table 4.20, while the confusion matrices are shown in the Appendix B.

		Random Forest			Nadam			RMSProp			Adadelta		
	type	All	99	78	All	99	78	All	99	78	All	99	78
Accuracy (%)	-	87	87	85	77	76	78	83	85	80	70	77	71
Purity	SL SN I	0.83	0.82	0.80	0.71	0.71	0.73	0.78	0.81	0.74	0.71	0.73	0.71
	SN Ia mix	0.93	0.93	0.93	0.87	0.85	0.85	0.90	0.90	0.89	0.70	0.82	0.71
Completeness	SL SN I	0.94	0.94	0.95	0.91	0.89	0.88	0.92	0.91	0.91	0.68	0.85	0.71
	SN Ia mix	0.80	0.80	0.76	0.63	0.64	0.67	0.74	0.79	0.68	0.73	0.68	0.71
F1 Score	SL SN I	0.88	0.88	0.87	0.80	0.79	0.80	0.84	0.86	0.82	0.70	0.78	0.71
	SN Ia mix	0.86	0.86	0.84	0.73	0.73	0.75	0.81	0.84	0.77	0.71	0.74	0.71

Table 4.20: Summary table of the best result for the 4 algorithms with all the features, the 99 and the 78 best ones. RMSProp has a different pair of best values for the decay and the learning rate depending on the parameter space used; 10^{-4} and 0.0005 for the case with all and 78 features; 10^{-5} and 0.0005 for the case with the 99 best features.

Random Forest is confirmed, even in this use case, as the best performing classifier. In this use case, for all classifiers with the exception of Nadam, the parameter space deduced from our analysis proved to be poorer than the other two used.

4.7.2 Contamination level of the SNIa mixed class

In this section, the error percentages of all SNe Ia types are shown in the best experiment with all the features of the Random Forest. The contamination classes are shown in table 4.21.

Class	Total	Correct	Wrong	% Wrong
SN Ia	2331	2328	3	≈0
SN Iax	2331	1508	823	35
SN Ia91bg	2331	1779	552	24

Table 4.21: Summary table of the contamination during the Random Forest classification experiment.

The worst ranked classes are SNIa91bg and SNIax, like in the SN Vs All use case. The SNIa is classified exceptionally well.

4.8 Six class problem

The final series of experiments is the more complex because we try to classify the six classes of SNe on the PLAsTiCC dataset. The experiments with the 4 algorithms are performed using all the features and the 96 best ones that optimize the parameter space. The balanced classes are shown in table 4.22. The percentage between training and test sets are 80% and 20%, respectively.

Class	Number of curves	
	Training	Test
SN Ia	27912	6979
SN Ia91bg	27938	6985
SN Iax	27981	6996
SN II	27816	6955
SN Ibc	27945	6987
SL SN I	27967	6992

Table 4.22: Summary table of the curves belonging to the dataset divided in training (80%) and test (20%) sets.

4.8.1 Classifiers

A series of experiment was performed, with the 4 algorithms, optimizing the decay value and pruning the learning rate for all the features and the 96 best ones. Between Nadam, RMSProp and Adadelta, the best performances were obtained with the RMSProp both in the case with all the features and with the optimized space. The best results follows in the table 4.23, while the confusion matrices are shown in the Appendix B.

		Random Forest			Nadam			RMSProp			Adadelta		
	type	All	96	78	All	96	78	All	96	78	All	96	78
Accuracy (%)	-	61	62	62	47	48	48	56	55	56	47	46	45
Purity	SN Ia	0.79	0.79	0.79	0.61	0.63	0.69	0.73	0.74	0.75	0.59	0.58	0.57
	SN Ia 91bg	0.82	0.79	0.78	0.50	0.50	0.71	0.77	0.74	0.82	0.46	0.49	0.57
	SN Iax	0.58	0.58	0.57	0.39	0.38	0.51	0.54	0.52	0.50	0.36	0.39	0.36
	SN II	0.74	0.74	0.75	0.56	0.57	0.55	0.66	0.65	0.69	0.53	0.52	0.53
	SN Ibc	0.40	0.40	0.42	0.31	0.31	0.30	0.35	0.34	0.35	0.32	0.31	0.31
	SL SN I	0.62	0.61	0.59	0.47	0.53	0.43	0.54	0.54	0.56	0.50	0.45	0.45
Completeness	SN Ia	0.77	0.77	0.77	0.70	0.61	0.50	0.71	0.69	0.70	0.55	0.49	0.42
	SN Ia 91bg	0.25	0.27	0.30	0.16	0.20	0.21	0.23	0.21	0.17	0.28	0.24	0.20
	SN Iax	0.33	0.35	0.37	0.26	0.20	0.13	0.32	0.26	0.26	0.21	0.21	0.16
	SN II	0.79	0.79	0.79	0.55	0.67	0.78	0.75	0.76	0.77	0.65	0.68	0.70
	SN Ibc	0.64	0.61	0.57	0.37	0.47	0.43	0.50	0.53	0.57	0.36	0.32	0.37
	SL SN I	0.91	0.91	0.91	0.79	0.75	0.82	0.82	0.84	0.88	0.76	0.81	0.88
F1 Score	SN Ia	0.78	0.78	0.78	0.65	0.62	0.58	0.72	0.71	0.72	0.57	0.54	0.48
	SN Ia 91bg	0.39	0.40	0.44	0.24	0.29	0.32	0.35	0.33	0.28	0.35	0.33	0.29
	SN Iax	0.42	0.43	0.45	0.31	0.26	0.21	0.40	0.35	0.34	0.26	0.27	0.22
	SN II	0.76	0.77	0.77	0.56	0.62	0.64	0.70	0.70	0.72	0.58	0.59	0.60
	SN Ibc	0.49	0.48	0.48	0.33	0.38	0.36	0.41	0.41	0.43	0.34	0.32	0.34
	SL SN I	0.73	0.73	0.72	0.58	0.62	0.57	0.65	0.66	0.68	0.60	0.58	0.60

Table 4.23: Summary table of the best result for the 4 algorithms with all the features, the 96 and the 78 best ones. RMSProp has the same pair of best values for the decay and the learning rate for every parameter space; 10^{-4} and 0.001, respectively.

Also in the latter use case, Random Forest confirmed itself as the best classifier in all 3 parameter spaces used. Furthermore, both the optimized parameter space and that derived from our analysis are equivalent to the space with all the features, providing solidity to the applied methodology. As regards the classification results, the SNIa91bg are difficult to classify having a high purity but poor completeness, indeed, as can be seen from the first confusion matrices B.5 in Appendix B, both types Ia91bg and Iax are strongly confused for SNIbc, which in fact has a low purity. SLSNI are very complete, although purity is reduced by the presence of SNIbc and SNIax. Peculiar classification for the SNIa and SNII which are confused only with each other despite the difference that these two types of objects should have.

4.9 Direct approach

On SNPhotCC dataset, also a direct approach was performed, using 5088 SNIa and 12027 SNII light curves for LSTM experiments. In this case, for every class, 20% of data composed the test set, and the remaining 80% was divided into training and validation sets in the $80 - 20$ proportions.

Class	Training	Validation	Test	Total
SNIa	3257	814	1017	5088
SNII	7696	1924	2404	12027
Total	10953	2738	3421	

Table 4.24: Dataset composition for LSTM experiments.

First type of data augmentation for LSTM

Since classes within the dataset were not well balanced, both training and validation of SNe Ia have been cloned by doubling their number. Therefore, both class sets have been augmented five times before starting the algorithm. The augmentation was based on the generation of a single timeline for all bands using their MJD (Mean Julian Day). Where the flux value was missing in a band, it was randomly calculated between the closest values in that band. We set our algorithm with 2 layers of 16 neurons each and a cyclic learning process over 200 epochs.

ID	MJD	Flux							
		Before				After			
		G	R	I	Z	G	R	I	Z
1	59820.0015	−25.154	0.284	−20.104	0.204	−25.154	0.284	−20.104	0.204
1	59822.0238	15.458		−5.144	0.215	15.458	1.748	−5.144	0.215
1	59823.1234	−5.848	2.285		6.385	−5.848	2.285	−13.845	6.385
1	59825.4451	−20.548	−0.243	−25.154	0.284	−20.548	−0.243	−25.154	0.284
1	59827.8251		−6.215			−5.234	−6.215	−33.845	2.278
1	59828.0234		9.555	−35.054	5.365	0.554	9.555	−35.054	5.365
1	59829.6234	10.854	0.215	−9.104	0.254	10.854	0.215	−9.104	0.254

Table 4.25: Data augmentation example.

Selection of the number of bands

To understand if one of more bands provide better or worst results, we executed a series of experiments, starting from a single band and increasing their number in order to see if there was one carrying higher performance contributions. The best result is shown in table 4.26.

	Type	griz
Accuracy (%)	-	80
Purity	Ia	0.62
	II	0.93
Completeness	Ia	0.87
	II	0.77
F1 Score	Ia	0.72
	II	0.84

Table 4.26: LSTM best result for different amounts of bands on SNPhotCC dataset. We had set dropout at 0.5, 16 neurons, 200 epochs and only one kind of augmentation.

So we could say that more bands involve better results and they are directly proportional to the number of bands.

LSTM Algorithm optimization

With all four bands, we start testing by investigating if a different augmentation could improve the accuracy of the experiments. This second type of augmentation was composed by adding to old augmentation a random component of ±5‰.

Then two experiments were performed, one with the old augmentation five times augmented and the second one with both ten times augmented. The second approach turned out to be better than the first one, but since it is very computing expensive, only the last experiment was carried out with this setup. After this test, it has been added a multi-step function to halve the learning rate in many epochs, while it has been increased its starting value to give more flexibility to the algorithm. The first experiment is done with one step at 2/3 of the epochs, while the second one have two steps at 2/3 and 9/10 of the epochs. Also in this case, the second option turned out to perform better. For the algorithm optimization the dropout had to be set. We tested the algorithm with a dropout step by 0.1 starting from 0.5 until 0.3 and then the intermediate point of 0.35, following the trend of the accuracy. Finally, we improved the number of neurons of the hidden layers from 16 to 25 and increase the number of training epochs from 200 to 250.

The result is reported in table 4.27.

To verify that the results were not influenced by a favorable selections of light curves, we reversed the test set with the validation set. The result was approximately the same. Finally we launched a test with the only second kind of augmentation and the ten augmentation test for the best configuration but the results was worst.

	type	Best
Accuracy (%)	-	91
Purity	Ia	0.81
	II	0.96
Completeness	Ia	0.90
	II	0.91
F1 Score	Ia	0.85
	II	0.93

Table 4.27: LSTM best configuration experiment on SNPhotCC dataset; 25 neurons, 250 epochs, dropout at 0.35, two step for learning rate (2/3 and 9/10) and only one kind of augmentation.

LSTM on PLAsTiCC dataset

Despite the good results obtained by this algorithm on this dataset, on the PLAsTiCC dataset the performances were not repeated. The algorithm fails to classify in most of the test cases, where the accuracy does not reach the 50%. Such behaviour was strange. Further future analyzes will be performed to analyze the reasons.

Chapter 5

Analysis

The present work is related to the important problem of classification of astrophysical variable sources, with special emphasis to SNe. Their relevance in terms of cosmological implications is well known, causing a special attention to the problem of recognizing different types of such explosive astronomical events. To face this challenge, the SNPhotCC dataset and the PLAsTiCC dataset have been chosen to have a statistical sample, albeit of simulations, as wide as possible. Based on the objects in the datasets, a test campaign with increasing complexity has drawn up. To approach the problem we have chosen to use the machine learning methods, in particular with 4 algorithms that require a statistical approach and with 1 that requires the direct one to the light curves. For the statistical approach we have relied on a set of features already tested in previous works on real objects. In the formation of statistical datasets, the presence of negative fluxes in the observations had to be solved, which prevented the creation of the features based on magnitude. Overcoming this obstacle, we wanted to optimize the space of the parameters to eliminate any redundant entropic information. To do this, the ΦLAB algorithm was used on whose results a statistical analysis of the occurrences was performed.

Negative fluxes problem analysis:

A negative flux has no physical sense. The negative values are related to observations made with an calibrated instrument in unfavorable weather conditions, and therefore to zero settings higher than the values of the astrophysical sources to be observed. This requires treating negative fluxes as real values. In the direct approach, the light curve shape is relevant, so their presence is not important because it is possible to positive translate the curve along the ordinate axis. In the statistical approach instead, since there are features that require magnitude and since the positive translation would alter the features values in an unequal way from curve to curve, to find an alternative way is necessary. To solve this problem we have tried, as already seen, three approaches. In the first, the atmospheric and setting conditions of the instruments were respected, considering the observations made on the same day in groups; this introduced noise as the substituted values were different and did not belong with certainty to the same phase of the light curve. The second, which proved to be the best, replace them with a positive value by maintaining the sampling, and since the value is always the same, the noise introduced is lower. Finally the third method removes the observations with negative fluxes but sub-sampled the light curves that had already been reduced during the preprocessing. From the results obtained it can be deduced that the deformations undergone by the light curves are not such as to alter their original nature nor to significantly reduce the algorithms performance in both datasets.

Parameter space analysis:

The parameter space analysis was performed to verify if the algorithms were afflicted by the so-called problem of the curse of dimensionality and if it was possible to optimize the parameter space by removing any redundant entropic information. For this purpose, the ΦLAB algorithm was applied which provided optimized spaces for the various use cases. Already with the results of these two parameter spaces it can be said that the algorithms are free from the curse and that moreover the amount of information contained in the total parameter space is preserved in the optimized one. Starting from the ΦLAB results, a statistical analysis was performed which highlighted the relevant physical characteristics for the transients classification and in particular of the SNe. The Amplitude feature, which represents the semi-difference between the minimum and maximum of the light curve, is the most relevant of all, having 50% of occurrences in the first quartile of normalized cumulative importance. Since the various classes of SNe have different light peaks, the semidifference of the amplitude of the curve is characteristic of the different type of object. Also important are the standard deviation, the MAD, and all the features that use the percentiles or that define the light curve shape such as Skewness and Kurtosys. The percentiles importance is related to the different decay time of the light radiation for the various types of SNe. Although SN is not a periodic phenomenon, the feature relative to the Lomb-Scargle periodogram has a good importance, because with its frequency peak tends to classify the SNe with a different periods of light decay. On the other hand, those characteristics that set thresholds on the number of points around the median are completely rejected, such as the Rcb, Mr and Mbrp, since their average values are close to their limit values. The test results on this additional parameter space confirm the physical importance of these characteristics and the amount of informations they bring, equaling the two spaces of the previous parameters.

Transients classification analysis:

The first step of the tests campaign consists in the Periodic Vs Non Periodic use case in which with all the algorithms, with the entire space of parameters and with the most noisy replacement method of negative fluxes, almost 100% of accuracy was achieved (Figure 5.1). The ability to distinguish a periodic object from a non-periodic one is confirmed, so the periodic objects were excluded to concentrate the tests on transients and in particular SNe. The high completeness (93%) in classifying SNe (Figure 5.2), in the SN Vs All use case, is a sign that algorithms, in particular Random Forest, distinguish them from other transients. We wanted to verify in the remaining 7%, which of the different SNe was the most contaminating; both SNe Ia91bg and SNe Iax were found to have the highest misclassification rate (12% of their test set). Moreover, from this analysis it was revealed that SNIa and SNII have an error rate of 1 per thousand, a remarkable result compared to the other SNe error rate. For completeness, the contamination was also verified for the All class and it was found that the AGN have an error rate of 1 per thousand, while the M-Dwarf and the TDE are the classes with the highest error rates ever (16%). Experiments should be carried out to identify the SNe class or classes with which these two different types of transients are confused and to verify which features play a key role in their classification. With the SNe and AGN contamination data, histograms were created in which test distributions and those of badly classified objects are superimposed according to the redshift. A correlation of badly classified objects with the redshift was expected, but this did not occur; the distributions perfectly follow those of the test sets regardless of the redshift range. Proceeding in the test campaign towards the classic SNIa Vs SNII, a peculiar result was obtained. On the PLAsTiCC dataset, both with 4 and 6 bands an accuracy close to 80% was reached (Figures 5.3 and 5.4), while on the SNPhotCC dataset, with all classifiers, at least 92% (RF 96%) was obtained (Figure 5.5). If on the one hand with a newer dataset the performances are supposed to be better, on the other the simulations are expected to be more complex and realistic and therefore with lower performances. In any case, a 17% performance difference over a relatively simple case given the different nature of the two objects surprised us. The penultimate element of this analysis is composed of the SLSNe Vs SNIa mixed use case, where the SLSNe are very recognizable to the classifiers with a completeness, in the best case, of 95% (Figure 5.6). To investigate the relatively low completeness of SNe Ia mixed (80%), a contamination test was also carried out in this use case, and it was found that while ”pure” Ia have a very low error percentage (1 per thousand), the other two types Ia91bg and Iax have a very high error percentage (24% and 35%, respectively). This data complies with the results obtained in the SN Vs All use case in which these classes were the worst. In the last test, in which we verify our ability to classify SNe finely, a not excellent result was expected. More than the result, it was interesting to observe the contaminations between the SNe, indeed, the types Ia91bg and Iax were confirmed to be the worst ever, classifying half of their objects as SNIbc; the SLSNe have a completeness of over 90% (Figure 5.8) and are confirmed as well classifiable objects. Peculiar is the classification of SNIa and SNII, since in spite of expectations, they have almost entirely the false negatives of the other. Depending on this result, it is possible to look with a different light at the classification obtained in the SNIa Vs SNII case, in which the overall accuracy was lower than expected. From a physical point of view it is not clear how this is possible, but it should be remembered that these are simulations and therefore a small defection in the creation of these two classes of objects is not to be excluded. Downstream of this classification process it can be said that we possess the features necessary for the classification of SNeIa, SNeII and SLSNe; however, those needed to classify the types Ia91bg and Iax are missing. Tests with only these two classes should be conducted, with at most the addition of SNIbc to evaluate which features contain the greatest importance for these objects and if the result is not satisfactory, find other more effective ones.

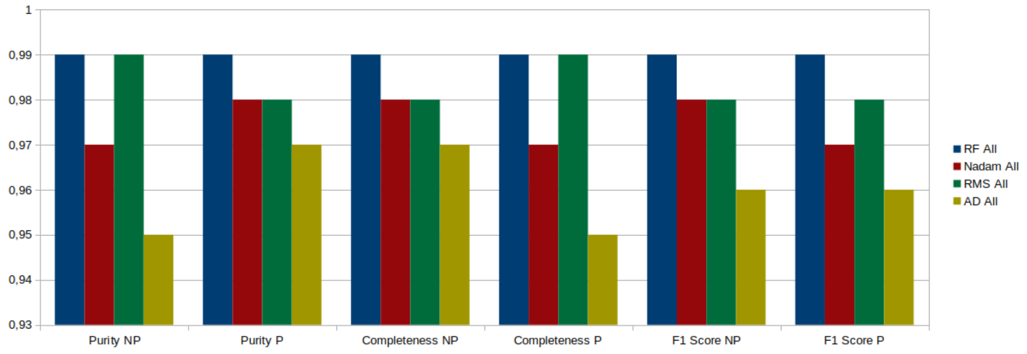


Figure 5.1: Summary histogram of the use case Periodic Vs non Periodic.

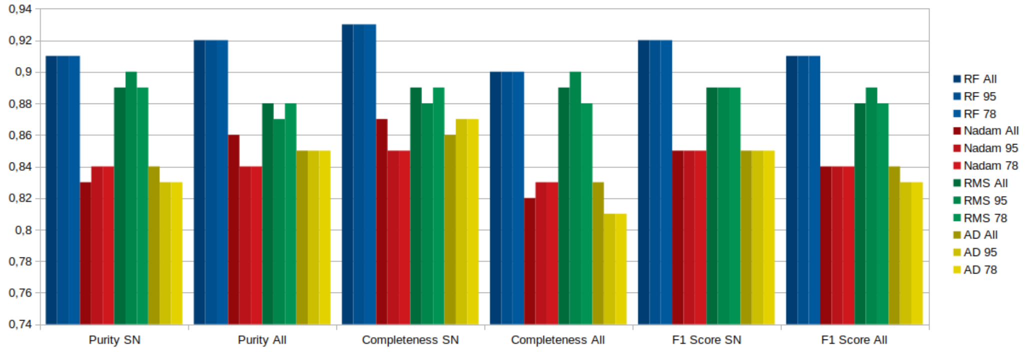


Figure 5.2: Summary histogram of the use case SN Vs All.

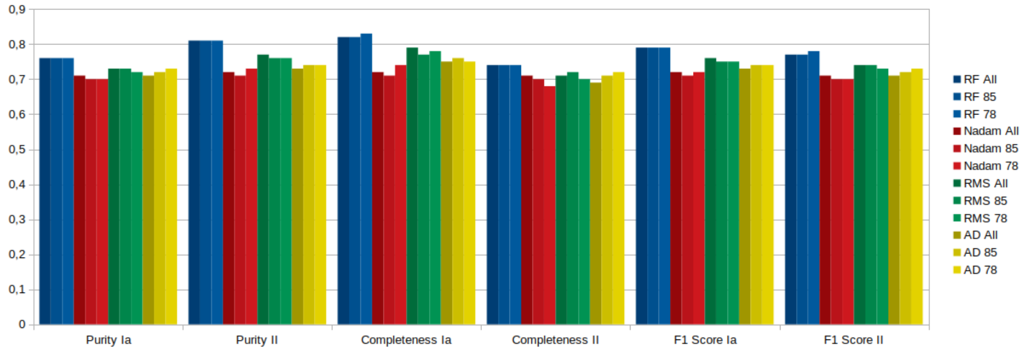


Figure 5.3: Summary histogram of the use case SNIa Vs SNII (UGRIZY).

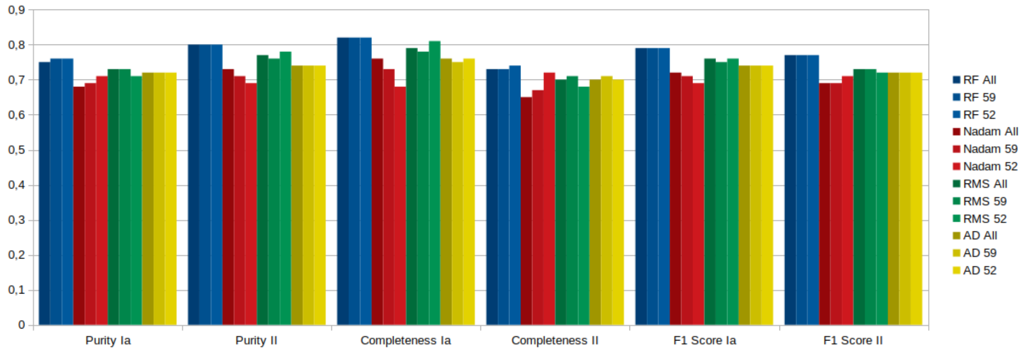


Figure 5.4: Summary histogram of the use case SNIa Vs SNII (GRIZ).

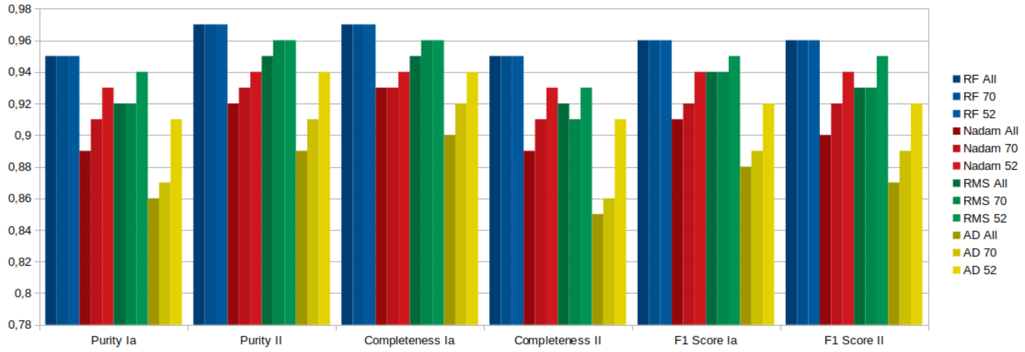


Figure 5.5: Summary histogram of the use case SNIa Vs SNII (GRIZ) on SNPhotCC dataset.

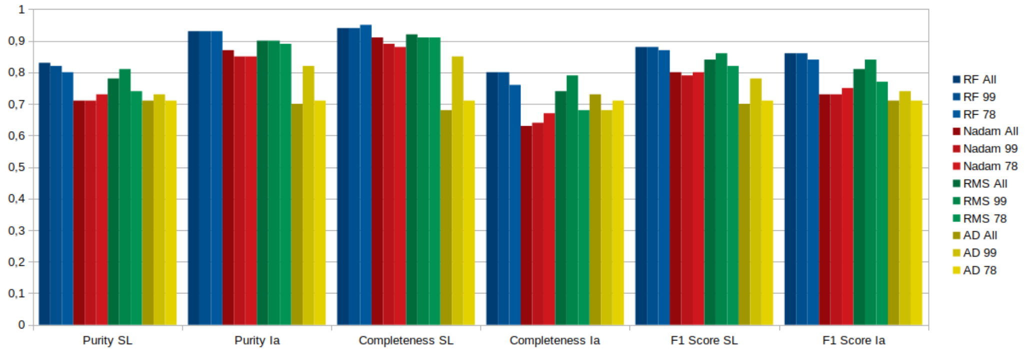


Figure 5.6: Summary histogram of the use case SLSNI Vs SNIa mixed.

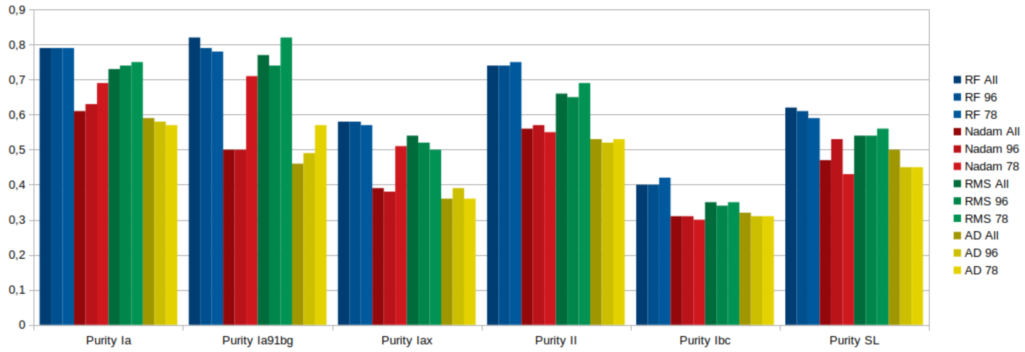


Figure 5.7: Summary histogram of the purity in the six class problem use case.

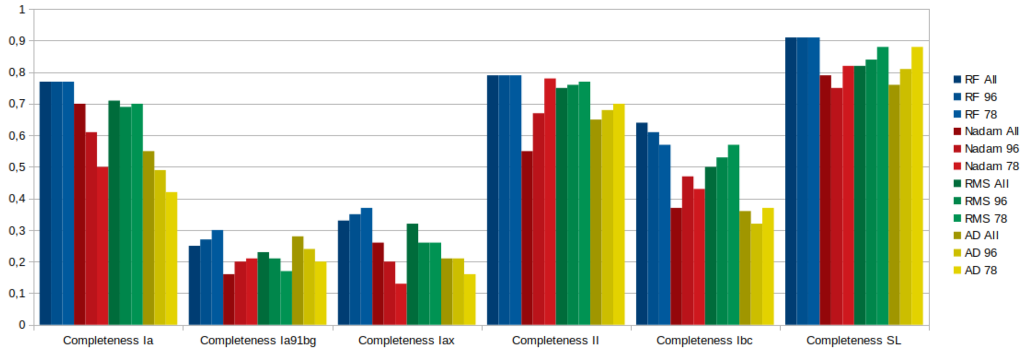


Figure 5.8: Summary histogram of the completeness in the six class problem use case.

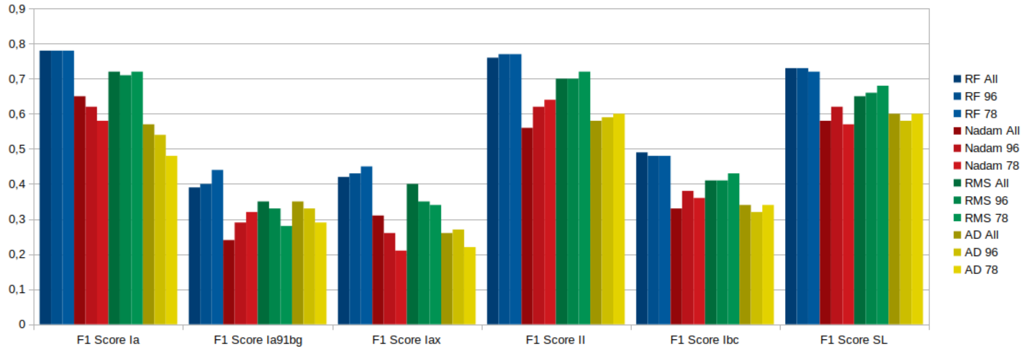


Figure 5.9: Summary histogram of the F1 Score in the six class problem use case.

Conclusions

With the incoming large amounts of Time Domain Astronomy data, foreseen by dedicated surveys like LSST, the urgent requirement is to be able to efficiently process and analyze data in an automatic way. This work was focused on astronomical objects classification, in particular Supernovae, due to their intrinsic importance in modern Cosmology. Two kinds of simulated datasets, named respectively, SNPhotCC and PLAsTiCC (the latter referred to a specific LSST data Challenge), have been chosen to estimate the classification accuracy, the possibility to optimize the multi-band light curve parameter space and to analyze the different performance of several machine learning based models on a series of interesting use cases, derived by proceeding through an increasing level of classification complexity and refinement:

- Periodic Vs non periodic;
- SNe Vs All;
- SNe Ia Vs SNe II;
- Super luminous SNe Vs SNe Ia mixed types;
- multi-classification among six classes of SNe;

Such classification use cases have been approached using four machine learning models, respectively, Random Forest, Nadam, RMSProp and Adadelta on objects whose original light curve has been transformed into a set of statistical parameters. A fifth classifier, Long Short Time Memory (LSTM), was applied only on the SNPhotCC dataset for the use case SNI to Vs SNII in which the original light curves have been used as input space.

Our analysis has highlighted some problematic aspects that require a preprocessing of data also in the real cases. First of all, the presence of negative fluxes within the light curves have a potential confusing impact on the training of machine learning based models, thus requiring a proper way to replace them with values able to avoid confusion, but at the same time without introducing additional noise, which may alter the light curve original properties. We managed to resolve the issue of negative fluxes with a substitution method that minimizes the curve deformation, retaining their original profile, without compromising the classification.

Second, the parameter space made by statistical features could have intrinsic quotes of redundancy, in terms of informative entropy carried by each parameter. Moreover, the large amount of dimensions may cause the occurrence of the so-called problem of curse of dimensionality. In order to circumvent such problems, a statistical analysis and selection of the relevant features was carried out, which led us to highlight which properties are more relevant for the classification of the SNe in different use cases. One of the most interesting outcomes of such investigation was to find frequent commonalities among the retained features in different cases, thus revealing the presence of statistical features able to represent main properties of transient sources.

Third, from the test campaign an overall picture was obtained in which some types of SNe, such as the type Ia91bg and the type Iax, are difficult to classify with completeness values always lower than those of the other types of SNe. Always high completeness ($\geq 90\%$), on the other hand, for SLSNe. As regards the distinction between SNe and other types of variable objects, over 90% results have been obtained both in purity and completeness for both classes, a sign that the algorithms, in the spaces of the parameters provided, are able to separate the two classes. In the SNIa Vs SNII use case, the PLAsTiCC and SNPhotCC datasets were compared, and on the latter, much better results were obtained. The reasons for this disparity will be the subject of future studies.

Appendix A

Details of feature selection

In this appendix, the feature importance ranking produced by ΦLab algorithm for all use cases was reported, except Periodic Vs Non Periodic, and the statistical analysis tables created downstream of its results .

Tables A.1 through A.5 are those generated by ΦLAB for each use case. The rejected features are highlighted in each table. The relevant features, besides their relative importance, have two columns of additional values; the first contains the cumulative importance of all the previous features including its own, while the second contains the normalized value of the cumulative importance compared to the cumulative importance of the last relevant features.

Tables from A.6 to A.10 show the occurrences of the features in the different quartiles of cumulative importance normalized for each use case. Table A.11 is the sum of the occurrence tables, and allows you to have an overview of the overall occurrence of a feature on all use cases. Table A.12 shows the common occurrences between the SNPhotCC dataset and the PLAsTiCC (GRIZ) dataset in the SNIa Vs SNII use case, in order to see the contact points of the two datasets from the point of view of the features.

Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum
pdfp_i	0,0371814660	0,04	0,05	pst_i	0,0070201488	0,57	0,71	$\mathrm{b1std}_r$	0,0033538880	0,75	0,94
ampl_u	0,0338237358	0,07	0,09	lt_z	0,0068909908	0,58	0,72	$\mathrm{fpr20}_u$	0,0032815518	0,76	0,94
ampl_y	0,0328036849	0,10	0,13	pst_z	0,0066251295	0,59	0,73	$\mathrm{fpr80}_r$	0,0032282084	0,76	0,95
pdfp_r	0,0311903111	0,13	0,17	$\mathrm{fpr50}_i$	0,0065837196	0,59	0,74	ms_y	0,0031521486	0,76	0,95
MAD_u	0,0247370610	0,16	0,20	pdfp_u	0,0062529070	0,60	0,75	ms_r	0,0030282367	0,77	0,95
ampl_z	0,0244773504	0,18	0,23	$\mathrm{fpr65}_g$	0,0062209880	0,61	0,75	$\mathrm{fpr20}_r$	0,0030102236	0,77	0,96
pdfp_z	0,0218756921	0,21	0,26	ls_r	0,0061703031	0,61	0,76	$\mathrm{fpr50}_y$	0,0029951743	0,77	0,96
MAD_g	0,0211819071	0,23	0,28	rcb_i	0,0061599942	0,62	0,77	ms_i	0,0028212316	0,78	0,97
std_r	0,0176111254	0,24	0,30	$\mathrm{fpr20}_z$	0,0059463044	0,62	0,78	$\mathrm{fpr35}_y$	0,0028200876	0,78	0,97
ampl_i	0,0168994331	0,26	0,33	$\mathrm{fpr65}_u$	0,005924636	0,63	0,78	kurt_g	0,0026735993	0,78	0,97
std_u	0,0146550311	0,28	0,34	$\mathrm{b1std}_z$	0,0059171911	0,64	0,79	$\mathrm{fpr65}_y$	0,0026673282	0,78	0,98
std_y	0,0142086803	0,29	0,36	$\mathrm{fpr50}_g$	0,0058999249	0,64	0,80	$\mathrm{b1std}_u$	0,0026083272	0,79	0,98
ampl_r	0,0138681062	0,30	0,38	pdfp_y	0,0057419087	0,65	0,81	$\mathrm{fpr20}_y$	0,0026038967	0,79	0,98
std_i	0,0137422819	0,32	0,40	rcb_r	0,0054917098	0,65	0,81	ls_u	0,0024826393	0,79	0,99
ls_i	0,0133292938	0,33	0,41	$\mathrm{fpr65}_z$	0,0054018176	0,66	0,82	lt_g	0,0024780054	0,79	0,99
skew_i	0,0127832995	0,34	0,43	lt_r	0,0053396838	0,66	0,83	$\mathrm{fpr80}_z$	0,0024640467	0,80	0,99
skew_r	0,0126662110	0,36	0,44	$\mathrm{fpr35}_g$	0,0052201272	0,67	0,83	ms_g	0,0022940202	0,80	0,99
skew_z	0,0125664279	0,37	0,46	kurt_y	0,0050879155	0,67	0,84	ls_g	0,0022098577	0,80	1,00
kurt_u	0,0124673349	0,38	0,48	$\mathrm{fpr50}_u$	0,0045345523	0,68	0,85	$\mathrm{fpr80}_y$	0,0021026658	0,80	1,00
pdfp_g	0,0121835862	0,39	0,49	$\mathrm{b1std}_y$	0,0043693833	0,68	0,85	pst_y	0,0032528848	-	-
kurt_z	0,0120057347	0,41	0,51	rcb_z	0,0043473349	0,69	0,86	mr_u	0,0022681871	-	-
std_z	0,0118971464	0,42	0,52	$\mathrm{fpr80}_g$	0,0041234879	0,69	0,86	rcb_y	0,0020005397	-	-
MAD_r	0,0118373976	0,43	0,53	ls_y	0,0040231389	0,70	0,87	mbrp_u	0,0019376063	-	-
MAD_z	0,0116012711	0,44	0,55	skew_y	0,00397431	0,70	0,87	pst_u	0,0019325138	-	-
ampl_g	0,011191885	0,45	0,56	$\mathrm{fpr35}_r$	0,0039497715	0,70	0,88	rcb_u	0,001871349	-	-
std_g	0,0102097268	0,46	0,58	lt_u	0,0039012591	0,71	0,88	rcb_g	0,0017755379	-	-
MAD_i	0,0101543118	0,47	0,59	$\mathrm{fpr20}_g$	0,0038938011	0,71	0,89	pst_g	0,001691790	-	-
ls_z	0,0101466114	0,48	0,60	$\mathrm{b1std}_i$	0,0038689585	0,72	0,89	$\mathrm{b1std}_g$	0,0015279611	-	-
lt_i	0,0099716438	0,49	0,61	$\mathrm{fpr35}_u$	0,0037100893	0,72	0,89	mbrp_y	0,0013337451	-	-
kurt_r	0,0094904909	0,50	0,63	ms_u	0,0037037269	0,72	0,90	mbrp_z	0,0013238730	-	-
pst_r	0,0094210599	0,51	0,64	MAD_y	0,0036772519	0,73	0,90	mbrp_i	0,0011859214	-	-
kurt_i	0,0088512421	0,52	0,65	skew_g	0,0035846557	0,73	0,91	mbrp_r	0,0011541231	-	-
skew_u	0,0085398861	0,53	0,66	$\mathrm{fpr80}_i$	0,0035839347	0,73	0,91	mr_y	0,0010940601	-	-
$\mathrm{fpr35}_z$	0,007742440	0,54	0,67	$\mathrm{fpr65}_i$	0,0035675536	0,74	0,92	mbrp_g	0,0008487359	-	-
$\mathrm{fpr50}_z$	0,0075270085	0,54	0,68	$\mathrm{fpr65}_r$	0,003525520	0,74	0,92	mr_z	0,0008280167	-	-
$\mathrm{fpr80}_u$	0,0075216618	0,55	0,69	ms_z	0,0035158621	0,74	0,93	mr_g	0,0008143458	-	-
$\mathrm{fpr20}_i$	0,007235119	0,56	0,70	$\mathrm{fpr50}_r$	0,0035072734	0,75	0,93	mr_i	0,0007833098	-	-
$\mathrm{fpr35}_i$	0,0072347931	0,57	0,71	lt_y	0,0034903813	0,75	0,93	mr_r	0,0007228613	-	-

Table A.1: Summary table of the feature importance for use case of SN Vs All. Features in red are rejected.

Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum
fpr65$_i$	0,0364737135	0,04	0,05	fpr20$_r$	0,0060775899	0,46	0,69	lt$_g$	0,0039939014	0,65	0,95
fpr50$_i$	0,0319834041	0,07	0,10	lt$_y$	0,0060228691	0,47	0,69	fpr35$_y$	0,0039931642	0,65	0,96
fpr80$_i$	0,0264234489	0,09	0,14	fpr65$_u$	0,0059477008	0,48	0,70	fpr50$_y$	0,0039714445	0,65	0,97
ampl$_i$	0,0248142172	0,12	0,18	ls$_y$	0,0058980548	0,48	0,71	ms$_z$	0,0039542367	0,66	0,97
fpr65$_r$	0,0214046077	0,14	0,21	skew$_z$	0,0058023548	0,49	0,72	ms$_y$	0,0039003685	0,66	0,98
fpr35$_i$	0,0188794956	0,16	0,24	skew$_i$	0,0056624576	0,49	0,73	fpr20$_y$	0,0038196315	0,67	0,98
fpr80$_r$	0,0183874379	0,18	0,26	pdfp$_r$	0,0055762458	0,50	0,74	skew$_u$	0,0038014156	0,67	0,99
fpr65$_z$	0,0180753415	0,20	0,29	fpr20$_g$	0,0055327766	0,50	0,75	MAD$_u$	0,0036932833	0,67	0,99
ampl$_r$	0,0180138119	0,21	0,32	std$_y$	0,0054363434	0,51	0,75	b1std$_i$	0,0036483657	0,68	1,00
fpr80$_z$	0,0155997423	0,23	0,34	fpr80$_u$	0,0053997573	0,52	0,76	ls$_u$	0,0032063276	-	-
fpr50$_r$	0,0142444590	0,24	0,36	MAD$_i$	0,0051946021	0,52	0,77	b1std$_z$	0,0030422291	-	-
ampl$_g$	0,0134437756	0,26	0,38	lt$_r$	0,0051194852	0,53	0,78	fpr20$_u$	0,0029935197	-	-
ampl$_z$	0,0113523636	0,27	0,40	skew$_g$	0,0050822755	0,53	0,78	b1std$_y$	0,0028951163	-	-
fpr50$_z$	0,0110952372	0,28	0,41	kurt$_i$	0,0050410685	0,54	0,79	pst$_z$	0,0028938256	-	-
ls$_r$	0,0092197908	0,29	0,43	fpr20$_z$	0,0050354031	0,54	0,80	pst$_i$	0,0028762013	-	-
ls$_z$	0,0090265278	0,30	0,44	fpr80$_y$	0,0049283046	0,55	0,81	pst$_y$	0,0028462081	-	-
ls$_i$	0,0087881287	0,31	0,45	fpr50$_u$	0,0048594166	0,55	0,81	b1std$_g$	0,0026849415	-	-
fpr50$_g$	0,0085745518	0,32	0,47	kurt$_g$	0,0047945740	0,56	0,82	b1std$_u$	0,0026437420	-	-
std$_g$	0,0085072456	0,32	0,48	skew$_y$	0,0047835870	0,56	0,83	rcb$_r$	0,0026322572	-	-
fpr65$_g$	0,0084664394	0,33	0,49	pdfp$_z$	0,0047805521	0,57	0,83	rcb$_z$	0,0026251495	-	-
std$_r$	0,0078799011	0,34	0,50	std$_u$	0,0047190085	0,57	0,84	rcb$_i$	0,0024917955	-	-
ampl$_u$	0,0077928380	0,35	0,51	MAD$_g$	0,0046995026	0,57	0,85	pst$_u$	0,0024356385	-	-
fpr35$_g$	0,0077329914	0,36	0,53	lt$_i$	0,0046728864	0,58	0,86	pst$_g$	0,0023851032	-	-
fpr35$_r$	0,0077235791	0,36	0,54	b1std$_r$	0,0046464484	0,58	0,86	rcb$_y$	0,0022799237	-	-
kurt$_r$	0,0076964016	0,37	0,55	fpr65$_y$	0,0046085738	0,59	0,87	mbrp$_r$	0,0022189659	-	-
fpr20$_i$	0,0076739719	0,38	0,56	kurt$_z$	0,0045224243	0,59	0,88	mbrp$_y$	0,0020949108	-	-
MAD$_r$	0,0071550146	0,39	0,57	fpr35$_u$	0,0042977987	0,60	0,88	mbrp$_z$	0,0020916067	-	-
lt$_z$	0,0070864089	0,39	0,58	kurt$_y$	0,0042005268	0,60	0,89	mbrp$_i$	0,0020884482	-	-
fpr35$_z$	0,0068892111	0,40	0,59	ms$_r$	0,0041784468	0,61	0,89	mbrp$_g$	0,0019577736	-	-
std$_i$	0,0067064066	0,41	0,60	ms$_u$	0,0041306668	0,61	0,90	mbrp$_u$	0,0017945800	-	-
std$_z$	0,0066396419	0,41	0,61	ls$_g$	0,0041230913	0,61	0,91	rcb$_g$	0,0017359774	-	-
skew$_r$	0,0065828295	0,42	0,62	pdfp$_y$	0,0041150684	0,62	0,91	rcb$_u$	0,0015162150	-	-
pdfp$_i$	0,0065809151	0,43	0,63	ms$_g$	0,0041066446	0,62	0,92	mr$_y$	0,0012921274	-	-
MAD$_y$	0,0064277793	0,43	0,64	kurt$_u$	0,0041027922	0,63	0,92	mr$_i$	0,0012894561	-	-
ampl$_y$	0,0063273329	0,44	0,65	pdfp$_u$	0,0040879600	0,63	0,93	mr$_r$	0,0012721738	-	-
pdfp$_g$	0,0062464267	0,45	0,66	lt$_u$	0,0040637578	0,63	0,94	mr$_z$	0,0012557126	-	-
MAD$_z$	0,0062419253	0,45	0,67	pst$_r$	0,0040362523	0,64	0,94	mr$_g$	0,0010842292	-	-
fpr80$_g$	0,0060871589	0,46	0,68	ms$_i$	0,0040204258	0,64	0,95	mr$_u$	0,0008505400	-	-

Table A.2: Summary table of the feature importance for use case of SNIa Vs SNII. Features in red are rejected.

Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum
ampl$_y$	0,0320255827	0,03	0,04	fpr20$_g$	0,0060730133	0,53	0,71	pst$_z$	0,0034020471	0,69	0,91
ampl$_u$	0,0281304187	0,06	0,08	skew$_i$	0,0055867573	0,54	0,71	pst$_y$	0,0033785937	0,69	0,92
MAD$_y$	0,0260573859	0,09	0,11	MAD$_g$	0,0055041850	0,55	0,72	kurt$_g$	0,0033373606	0,70	0,92
std$_y$	0,0253585894	0,11	0,15	skew$_r$	0,0055008290	0,55	0,73	fpr65$_g$	0,0033088861	0,70	0,93
MAD$_z$	0,0236064541	0,14	0,18	fpr50$_g$	0,0049396156	0,56	0,73	rcb$_y$	0,0032861712	0,70	0,93
ampl$_i$	0,0229316775	0,16	0,21	pdfp$_z$	0,0049330562	0,56	0,74	ms$_z$	0,0032254120	0,71	0,93
ls$_z$	0,0228695666	0,18	0,24	lt$_i$	0,0048983866	0,57	0,75	fpr80$_u$	0,0032053230	0,71	0,94
ampl$_z$	0,0227410316	0,20	0,27	fpr50$_r$	0,0047334098	0,57	0,75	fpr35$_y$	0,0031043992	0,71	0,94
ampl$_g$	0,0199478209	0,22	0,30	b1std$_y$	0,0046033004	0,57	0,76	ms$_g$	0,0030955433	0,72	0,95
std$_r$	0,0187173361	0,24	0,32	kurt$_i$	0,0045588077	0,58	0,77	lt$_u$	0,0030665720	0,72	0,95
ls$_y$	0,0185247897	0,26	0,34	pdfp$_y$	0,0045416009	0,58	0,77	ls$_g$	0,0030132181	0,72	0,95
fpr80$_i$	0,0167955240	0,28	0,37	pst$_r$	0,0045093068	0,59	0,78	kurt$_u$	0,0030070020	0,73	0,96
mr$_y$	0,0157952062	0,29	0,39	fpr65$_z$	0,0044791352	0,59	0,78	mbrp$_y$	0,0030041276	0,73	0,96
fpr80$_r$	0,0155222646	0,31	0,41	fpr65$_y$	0,0044411521	0,60	0,79	pst$_u$	0,0030007378	0,73	0,97
std$_u$	0,0151828694	0,32	0,43	fpr20$_z$	0,0044359990	0,60	0,80	b1std$_r$	0,0029908754	0,73	0,97
ampl$_r$	0,0139228262	0,34	0,45	fpr35$_z$	0,0042552973	0,61	0,80	fpr20$_y$	0,0029643907	0,74	0,97
std$_z$	0,0132031087	0,35	0,46	ms$_u$	0,0041140348	0,61	0,81	pst$_i$	0,0029246804	0,74	0,98
std$_g$	0,0131458693	0,36	0,48	lt$_r$	0,0041108692	0,61	0,81	fpr20$_i$	0,0029091586	0,74	0,98
fpr65$_i$	0,0118705273	0,38	0,50	fpr50$_u$	0,0040568966	0,62	0,82	fpr20$_u$	0,0028995886	0,75	0,99
std$_i$	0,0105095710	0,39	0,51	pdfp$_i$	0,0038630014	0,62	0,82	ms$_i$	0,0027426016	0,75	0,99
MAD$_i$	0,0102435115	0,40	0,52	b1std$_z$	0,0038541677	0,63	0,83	ms$_r$	0,0026988165	0,75	0,99
ls$_i$	0,0099111562	0,41	0,54	fpr35$_i$	0,0037541203	0,63	0,83	skew$_u$	0,0026882788	0,75	1,00
fpr65$_r$	0,0095833575	0,42	0,55	ms$_y$	0,0037373931	0,63	0,84	rcb$_i$	0,0026807914	0,76	1,00
ls$_r$	0,0095633557	0,43	0,56	kurt$_r$	0,0037245447	0,64	0,84	ls$_u$	0,0024952290	-	-
skew$_y$	0,0083979277	0,43	0,57	fpr20$_r$	0,0036783956	0,64	0,85	rcb$_r$	0,0024812505	-	-
lt$_z$	0,0083647938	0,44	0,59	pdfp$_u$	0,0036764153	0,64	0,85	mr$_u$	0,0023307836	-	-
fpr80$_z$	0,0082932774	0,45	0,60	fpr50$_y$	0,0036285316	0,65	0,86	b1std$_u$	0,0021138089	-	-
lt$_y$	0,0075860477	0,46	0,61	fpr65$_u$	0,0036088446	0,65	0,86	b1std$_g$	0,0019728031	-	-
pdfp$_g$	0,0075061657	0,47	0,62	MAD$_u$	0,0035979538	0,66	0,87	pst$_g$	0,0019223527	-	-
kurt$_z$	0,0074538668	0,47	0,63	fpr35$_u$	0,0035651692	0,66	0,87	rcb$_u$	0,0018751162	-	-
skew$_z$	0,0071499082	0,48	0,64	b1std$_i$	0,0035633072	0,66	0,88	rcb$_g$	0,0017223046	-	-
fpr50$_i$	0,0071310123	0,49	0,64	mr$_i$	0,0035415763	0,67	0,88	mbrp$_u$	0,0017204870	-	-
kurt$_y$	0,0070286933	0,50	0,65	skew$_g$	0,0035329114	0,67	0,88	mbrp$_z$	0,0016557219	-	-
mr$_z$	0,0070073810	0,50	0,66	lt$_g$	0,0035011092	0,67	0,89	mbrp$_i$	0,0016070604	-	-
MAD$_r$	0,0068288541	0,51	0,67	fpr35$_r$	0,0034842157	0,68	0,89	mbrp$_r$	0,0015646905	-	-
fpr35$_g$	0,0065967623	0,52	0,68	fpr50$_z$	0,0034569277	0,68	0,90	mr$_g$	0,0015307315	-	-
fpr80$_y$	0,0062993253	0,52	0,69	rcb$_z$	0,0034455195	0,68	0,90	mr$_r$	0,0012955470	-	-
pdfp$_r$	0,0061443148	0,53	0,70	fpr80$_g$	0,0034301776	0,69	0,91	mbrp$_g$	0,0011272735	-	-

Table A.3: Summary table of the feature importance for use case of SLSNI Vs SNIa mixed. Features in red are rejected.

Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum
ampl$_z$	0,0371665073	0,04	0,05	fpr50$_z$	0,0059655657	0,48	0,67	ms$_i$	0,0036188454	0,65	0,91
ampl$_y$	0,0328580929	0,07	0,10	fpr35$_i$	0,0059581076	0,48	0,68	pdfp$_u$	0,0036067467	0,65	0,92
ampl$_i$	0,0308841707	0,10	0,14	ms$_z$	0,0058324001	0,49	0,69	b1std$_r$	0,0036034863	0,65	0,92
std$_y$	0,0234914357	0,12	0,18	MAD$_u$	0,0056959454	0,49	0,70	rcb$_y$	0,0035873927	0,66	0,93
ampl$_r$	0,0226115177	0,15	0,21	fpr50$_g$	0,0054023676	0,50	0,70	fpr35$_y$	0,0035757727	0,66	0,93
ampl$_u$	0,0194837363	0,17	0,23	lt$_z$	0,0052597053	0,50	0,71	fpr80$_u$	0,0035536713	0,66	0,94
std$_u$	0,0193241253	0,19	0,26	kurt$_r$	0,0050588454	0,51	0,72	lt$_u$	0,0035401394	0,67	0,94
std$_r$	0,0167887774	0,20	0,29	pst$_y$	0,0050130364	0,51	0,72	lt$_g$	0,0035389862	0,67	0,95
std$_i$	0,0161267071	0,22	0,31	fpr35$_g$	0,0049656188	0,52	0,73	fpr65$_u$	0,0034835752	0,67	0,95
MAD$_y$	0,0143818273	0,23	0,33	fpr65$_g$	0,0049544934	0,52	0,74	pst$_i$	0,0034519072	0,68	0,95
std$_z$	0,0141385402	0,25	0,35	lt$_y$	0,0048773515	0,53	0,75	fpr50$_u$	0,0033781569	0,68	0,96
ampl$_g$	0,0133533197	0,26	0,37	fpr35$_z$	0,0048383981	0,53	0,75	ls$_g$	0,0033416387	0,69	0,96
MAD$_z$	0,0125141207	0,27	0,38	pdfp$_z$	0,0046635168	0,54	0,76	kurt$_g$	0,0033282592	0,69	0,97
fpr65$_i$	0,0121984706	0,29	0,40	fpr35$_r$	0,0045618536	0,54	0,77	fpr20$_y$	0,0033213773	0,69	0,97
std$_g$	0,0109808067	0,30	0,42	pst$_r$	0,0045135272	0,55	0,77	kurt$_u$	0,0031626114	0,69	0,98
fpr80$_i$	0,0100090151	0,31	0,43	pdfp$_i$	0,0044670326	0,55	0,78	rcb$_z$	0,0031498680	0,70	0,98
fpr50$_i$	0,0096862539	0,32	0,44	pdfp$_y$	0,0044529600	0,56	0,78	fpr35$_u$	0,0031461182	0,70	0,99
fpr65$_r$	0,0095170482	0,33	0,46	lt$_r$	0,0044277816	0,56	0,79	fpr20$_u$	0,0030871890	0,70	0,99
ls$_z$	0,0093559303	0,33	0,47	kurt$_i$	0,0044267098	0,57	0,80	skew$_u$	0,0030740108	0,71	1,00
fpr80$_z$	0,0092912653	0,34	0,48	fpr80$_g$	0,0043213177	0,57	0,80	ls$_u$	0,0030264735	0,71	1,00
MAD$_i$	0,0085083931	0,35	0,50	ms$_u$	0,0043142515	0,57	0,81	rcb$_i$	0,0028884259	-	-
MAD$_r$	0,0081948894	0,36	0,51	pdfp$_g$	0,0042819020	0,58	0,82	rcb$_r$	0,0028149133	-	-
b1std$_y$	0,0080898889	0,37	0,52	pdfp$_r$	0,0042227370	0,58	0,82	rcb$_g$	0,0027381999	-	-
fpr80$_r$	0,0076550705	0,38	0,53	fpr20$_i$	0,0042213405	0,59	0,83	pst$_u$	0,0025119081	-	-
ls$_y$	0,0075746749	0,38	0,54	fpr80$_y$	0,0041868297	0,59	0,83	mr$_u$	0,0024738807	-	-
fpr65$_z$	0,0073660747	0,39	0,55	b1std$_i$	0,0041599087	0,60	0,84	mbrp$_u$	0,0024067044	-	-
skew$_y$	0,0072702645	0,40	0,56	mr$_y$	0,0041000020	0,60	0,84	mr$_z$	0,0023741062	-	-
kurt$_y$	0,0070406228	0,41	0,57	fpr20$_r$	0,0040968447	0,60	0,85	mbrp$_y$	0,0022491790	-	-
skew$_i$	0,0069320168	0,41	0,58	fpr20$_g$	0,0040790757	0,61	0,86	pst$_g$	0,0022055124	-	-
fpr50$_r$	0,0068260005	0,42	0,59	lt$_i$	0,0040778245	0,61	0,86	b1std$_u$	0,0022054805	-	-
ls$_r$	0,0067814491	0,43	0,60	fpr20$_z$	0,0040591997	0,62	0,87	b1std$_g$	0,0021536689	-	-
skew$_r$	0,0067122310	0,43	0,61	ms$_g$	0,0040185697	0,62	0,87	mbrp$_z$	0,0020138005	-	-
ms$_y$	0,0067046154	0,44	0,62	skew$_g$	0,0039004274	0,62	0,88	mbrp$_g$	0,0018607871	-	-
skew$_z$	0,0066126008	0,45	0,63	fpr65$_y$	0,0038633520	0,63	0,88	mbrp$_i$	0,0018299366	-	-
MAD$_g$	0,0062706517	0,45	0,64	ms$_r$	0,0037920950	0,63	0,89	mbrp$_r$	0,0017838205	-	-
ls$_i$	0,0060773819	0,46	0,65	pst$_z$	0,0037138496	0,64	0,89	mr$_i$	0,0017571855	-	-
b1std$_z$	0,0060360428	0,46	0,65	fpr50$_y$	0,0036678623	0,64	0,90	mr$_r$	0,0013230339	-	-
kurt$_z$	0,0059666115	0,47	0,66	rcb$_u$	0,0036363377	0,64	0,90	mr$_g$	0,0012514640	-	-

Table A.4: Summary table of the feature importance for use case of Six Class Problem. Features in red are rejected.

Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum	Feat.	Imp.	Sum	N.Sum
MAD$_z$	0,0932826356	0,09	0,11	fpr50$_r$	0,0082528931	0,68	0,78	rcb$_z$	0,0036409820	0,82	0,94
pdfp$_i$	0,0715069042	0,16	0,19	ls$_z$	0,0078390007	0,69	0,79	fpr50$_g$	0,0034709517	0,82	0,95
MAD$_i$	0,0700383281	0,23	0,27	ampl$_g$	0,0077315636	0,69	0,80	fpr65$_z$	0,0034700518	0,83	0,95
ampl$_i$	0,0501079730	0,28	0,33	std$_g$	0,0072947051	0,70	0,81	ms$_g$	0,0033146934	0,83	0,96
ampl$_z$	0,0402354693	0,33	0,37	ls$_g$	0,0070692247	0,71	0,81	pst$_i$	0,0031706262	0,83	0,96
MAD$_r$	0,0393216910	0,36	0,42	b1std$_i$	0,0070622914	0,71	0,82	fpr65$_g$	0,0031194234	0,84	0,96
pdfp$_z$	0,0344661999	0,40	0,46	ls$_i$	0,0068975218	0,72	0,83	fpr35$_g$	0,0029538266	0,84	0,97
pdfp$_r$	0,0322476328	0,43	0,50	kurt$_i$	0,0067366189	0,73	0,84	mbrp$_g$	0,0029466912	0,84	0,97
lt$_z$	0,0226726435	0,45	0,52	fpr50$_z$	0,0066711905	0,73	0,84	pst$_r$	0,0028317328	0,85	0,97
ampl$_r$	0,0215126662	0,48	0,55	fpr35$_z$	0,0065425625	0,74	0,85	rcb$_i$	0,0028195905	0,85	0,98
MAD$_g$	0,0183406884	0,49	0,57	skew$_z$	0,0064736969	0,75	0,86	fpr20$_g$	0,0027687884	0,85	0,98
lt$_i$	0,0179415019	0,51	0,59	b1std$_r$	0,0058639021	0,75	0,87	kurt$_g$	0,0027620353	0,85	0,98
std$_i$	0,0163199023	0,53	0,61	ms$_z$	0,0057974767	0,76	0,87	b1std$_g$	0,0026661840	0,86	0,99
std$_z$	0,0154092393	0,54	0,62	fpr20$_r$	0,0054197795	0,77	0,88	fpr80$_r$	0,0026329648	0,86	0,99
lt$_r$	0,0142618431	0,56	0,64	fpr65$_r$	0,0052304092	0,77	0,89	fpr80$_g$	0,0026301232	0,86	0,99
fpr35$_i$	0,0122161066	0,57	0,66	mbrp$_z$	0,0051357125	0,78	0,89	rcb$_r$	0,0026253246	0,86	0,99
ls$_r$	0,0118180991	0,58	0,67	mbrp$_i$	0,0046101311	0,78	0,90	fpr80$_i$	0,0024535842	0,87	1,00
fpr50$_i$	0,0114344990	0,59	0,68	ms$_i$	0,0045328953	0,78	0,90	fpr80$_z$	0,0024075737	0,87	1,00
skew$_g$	0,0112242768	0,60	0,69	fpr20$_z$	0,0044204302	0,79	0,91	mr$_g$	0,0021954649	-	-
skew$_i$	0,0103066096	0,61	0,71	fpr65$_i$	0,0044111571	0,79	0,91	mbrp$_r$	0,0019964184	-	-
std$_r$	0,0100680233	0,62	0,72	kurt$_z$	0,0041564953	0,80	0,92	pst$_g$	0,0017612288	-	-
fpr20$_i$	0,0095596391	0,63	0,73	b1std$_z$	0,0041178339	0,80	0,92	mr$_z$	0,0015543084	-	-
skew$_r$	0,0092835200	0,64	0,74	kurt$_r$	0,0040797917	0,81	0,93	mr$_i$	0,0014818076	-	-
lt$_g$	0,0090571308	0,65	0,75	ms$_r$	0,0038564026	0,81	0,93	mr$_r$	0,0013886065	-	-
pdfp$_g$	0,0084317309	0,66	0,76	pst$_z$	0,0038429183	0,81	0,94	-	-	-	-
fpr35$_r$	0,0083728954	0,67	0,77	rcb$_g$	0,0036812692	0,82	0,94	-	-	-	-

Table A.5: Summary table of the feature importance for use case of SNIa Vs SNII on SNPhotCC dataset. Features in red are rejected.

		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	6	3	50.0%	5	83.3%	6	100.0%	6	100.0%	0	0.0%
Pdfp$_X$	6	3	50.0%	4	66.7%	5	83.3%	6	100.0%	0	0.0%
MAD$_X$	6	1	16.7%	2	33.3%	5	83.3%	6	100.0%	0	0.0%
Std$_X$	6		0.0%	5	83.3%	6	100.0%	6	100.0%	0	0.0%
Skew$_X$	6		0.0%	3	50.0%	4	66.7%	6	100.0%	0	0.0%
FprYY$_X$	30		0.0%		0.0%	7	23.3%	30	100.0%	0	0.0%
Kurt$_X$	6		0.0%	2	33.3%	4	66.7%	6	100.0%	0	0.0%
Lt$_X$	6		0.0%	1	16.7%	2	33.3%	6	100.0%	0	0.0%
Ls$_X$	6		0.0%		0.0%	2	33.3%	6	100.0%	0	0.0%
Ms$_X$	6		0.0%		0.0%		0.0%	6	100%	0	0.0%
B1std$_X$	6		0.0%		0.0%		0.0%	5	83.3%	1	16.7.0%
Pst$_X$	6		0.0%		0.0%	3	50.0%	3	50.0%	3	50.0%
Rcb$_X$	6		0.0%		0.0%		0.0%	3	50.0%	3	50.0%
Mr$_X$	6		0.0%		0.0%		0.0%		0.0%	6	100.0%
Mbrp$_X$	6		0.0%		0.0%		0.0%		0.0%	6	100.0%
Total	114	7		22		44		95		19	

Table A.6: Summary table of the features belonging to the use case SN Vs All, in the PLAsTiCC dataset, divided by quartiles of importance.

		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	6	1	16.7%	5	83.3%	6	100.0%	6	100.0%	0	0.0%
FprYY$_X$	30	6	20.0%	12	40.0%	20	66.7%	29	96.7%	1	3.3%
Std$_X$	6		0.0%	2	33.3%	5	83.3%	6	100.0%	0	0.0%
Pdfp$_X$	6		0.0%		0.0%	3	50.0%	6	100.0%	0	0.0%
MAD$_X$	6		0.0%		0.0%	3	50.0%	6	100.0%	0	0.0%
Skew$_X$	6		0.0%		0.0%	3	50.0%	6	100.0%	0	0.0%
Ls$_X$	6		0.0%	3	50.0%	4	66.7%	5	83.3%	1	16.7%
Kurt$_X$	6		0.0%		0.0%	1	16.7%	6	100.0%	0	0.0%
Lt$_X$	6		0.0%		0.0%	2	33.3%	6	100.0%	0	0.0%
Ms$_X$	6		0.0%		0.0%		0.0%	6	100%	0	0.0%
B1std$_X$	6		0.0%		0.0%		0.0%	2	33.3%	4	66.7.0%
Pst$_X$	6		0.0%		0.0%		0.0%	1	16.7%	5	83.3%
Rcb$_X$	6		0.0%		0.0%		0.0%		0.0%	6	100.0%
Mr$_X$	6		0.0%		0.0%		0.0%		0.0%	6	100.0%
Mbrp$_X$	6		0.0%		0.0%		0.0%		0.0%	6	100.0%
Total	114	7		22		47		85		29	

Table A.7: Summary table of the features belonging to the use case SNIa Vs SNII, in the PLAsTiCC dataset, divided by quartiles of importance.

		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	6	3	50.0%	6	100.0%	6	100.0%	6	100.0%	0	0.0%
FprYY$_X$	30		0.0%	3	10.0%	11	36.7%	30	100.0%	0	0.0%
Std$_X$	6	1	16.7%	6	100.0%	6	100.0%	6	100.0%	0	0.0%
MAD$_X$	6	2	33.3%	3	50.0%	5	83.3%	6	100.0%	0	0.0%
Pdfp$_X$	6		0.0%		0.0%	3	50.0%	6	100.0%	0	0.0%
Ls$_X$	6	1	16.7%	2	33.3%	4	66.7%	5	83.3%	1	16.7%
Skew$_X$	6		0.0%		0.0%	4	66.7%	6	100.0%	0	0.0%
Lt$_X$	6		0.0%		0.0%	3	50.0%	6	100.0%	0	0.0%
Kurt$_X$	6		0.0%		0.0%	2	33.3%	6	100.0%	0	0.0%
Ms$_X$	6		0.0%		0.0%		0.0%	6	100%	0	0.0%
B1std$_X$	6		0.0%		0.0%		0.0%	4	67.7%	2	33.3%
Pst$_X$	6		0.0%		0.0%		0.0%	5	83.3%	1	16.7%
Rcb$_X$	6		0.0%		0.0%		0.0%	3	50.0%	3	50.0%
Mr$_X$	6		0.0%	1	16.7%	2	33.3%	3	50.0%	3	50.0%
Mbrp$_X$	6		0.0%		0.0%		0.0%	1	16.7%	5	83.3%
Total	114	7		21		46		99		15	

Table A.8: Summary table of the features belonging to the use case SL SN I Vs SNIa mixed, in the PLAsTiCC dataset, divided by quartiles of importance.

		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	6	5	83.3%	6	100.0%	6	100.0%	6	100.0%	0	0.0%
Std$_X$	6	2	33.3%	6	100.0%	6	100.0%	6	100.0%	0	0.0%
FprYY$_X$	30		0.0%	5	16.7%	14	46.7%	30	100.0%	0	0.0%
MAD$_X$	6		0.0%	4	66.7%	6	100.0%	6	100.0%	0	0.0%
Ls$_X$	6		0.0%	1	16.7%	4	67.7%	6	100.0%	0	0.0%
Skew$_X$	6		0.0%		0.0%	4	66.7%	6	100.0%	0	0.0%
Kurt$_X$	6		0.0%		0.0%	3	50.0%	6	100.0%	0	0.0%
Lt$_X$	6		0.0%		0.0%	2	33.3%	6	100.0%	0	0.0%
Ms$_X$	6		0.0%		0.0%	2	33.3%	6	100.0%	0	0.0%
Pdfp$_X$	6		0.0%		0.0%		0.0%	6	100%	0	0.0%
B1std$_X$	6		0.0%	1	16.7%	2	33.3%	4	67.7%	2	33.3%
Pst$_X$	6		0.0%		0.0%	1	16.7%	4	66.7%	2	33.3%
Rcb$_X$	6		0.0%		0.0%		0.0%	3	50.0%	3	50.0%
Mr$_X$	6		0.0%		0.0%		0.0%	1	16.7%	5	83.3%
Mbrp$_X$	6		0.0%		0.0%		0.0%		0.0%	6	100.0%
Total	114	7		21		46		99		15	

Table A.9: Summary table of the features belonging to the use case six class problem, in the PLAsTiCC dataset, divided by quartiles of importance.

		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	4		0.0%	2	50.0%	3	75.0%	4	100.0%	0	0.0%
Std$_X$	4		0.0%		0.0%	3	75.0%	4	100.0%	0	0.0%
FprYY$_X$	20		0.0%		0.0%	3	15.0%	20	100.0%	0	0.0%
Pdfp$_X$	4	1	25.0%	3	75.0%	3	75.0%	4	100.0%	0	0.0%
MAD$_X$	4	2	50.0%	3	75.0%	4	100.0%	4	100.0%	0	0.0%
Skew$_X$	4		0.0%		0.0%	3	75.0%	4	100.0%	0	0.0%
Kurt$_X$	4		0.0%		0.0%		0.0%	4	100.0%	0	0.0%
Lt$_X$	4		0.0%	1	25.0%	4	100.0%	4	100.0%	0	0.0%
Ls$_X$	4		0.0%		0.0%	1	25.0%	4	100.0%	0	0.0%
Ms$_X$	4		0.0%		0.0%		0.0%	4	100%	0	0.0%
B1std$_X$	4		0.0%		0.0%		0.0%	4	100.0%	0	0.0%
Pst$_X$	4		0.0%		0.0%		0.0%	3	75.0%	1	25.0%
Rcb$_X$	4		0.0%		0.0%		0.0%	4	100.0%	0	0.0%
Mr$_X$	4		0.0%		0.0%		0.0%		0.0%	4	100.0%
Mbrp$_X$	4		0.0%		0.0%		0.0%	3	75.0%	1	25.0%
Total	76	3		9		24		70		6	

Table A.10: Summary table of the features belonging to the use case SNIa Vs SNII, in the SNPhotCC dataset, divided by quartiles of importance.

		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	24	12	50.0%	22	91.7%	24	100.0%	24	100.0%	0	0.0%
Std$_X$	24	3	12.5%	19	79.2%	23	95.8%	24	100.0%	0	0.0%
FprYY$_X$	120	6	5.0%	20	16.7%	52	43.3%	119	99.2%	1	0.8%
Pdfp$_X$	24	3	12.5%	4	16.7%	11	45.8%	24	100.0%	0	0.0%
MAD$_X$	24	3	12.5%	9	37.5%	19	79.2%	24	100.0%	0	0.0%
Skew$_X$	24		0.0%	3	12.5%	15	62.5%	24	100.0%	0	0.0%
Kurt$_X$	24		0.0%	2	8.3%	10	41.7%	24	100.0%	0	0.0%
Lt$_X$	24		0.0%		0.0%	9	37.5%	24	100.0%	0	0.0%
Ls$_X$	24	1	4.2%	7	29.2%	14	58.3%	22	91.7%	2	8.3%
Ms$_X$	24		0.0%		0.0%	2	8.3%	24	100%	0	0.0%
B1std$_X$	24		0.0%	1	4.2%	2	8.3%	15	62.5%	9	37.5%
Pst$_X$	24		0.0%		0.0%	4	16.7%	15	62.5%	9	37.5%
Rcb$_X$	24		0.0%		0.0%		0.0%	10	41.7%	14	58.3%
Mr$_X$	24		0.0%	1	4.2%	2	8.3%	4	16.7%	20	83.3%
Mbrp$_X$	24		0.0%		0.0%		0.0%	1	4.2%	23	95.8%
Total	456	28		88		187		378		78	

Table A.11: Cumulative table of the features belonging to the 4 use cases on the PLAsTiCC dataset, divided by quartiles of importance.

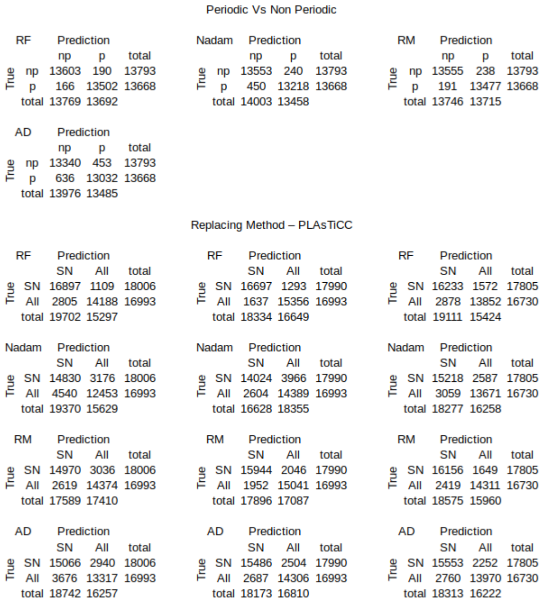
		25%		50%		75%		100%			
Feature	Total	Occur.	%	Occur.	%	Occur.	%	Occur.	%	Rejected	%
Ampl$_X$	4		0.0%	2	50.0%	3	75.0%	4	100.0%	0	0.0%
Std$_X$	4		0.0%		0.0%	3	75.0%	4	100.0%	0	0.0%
FprYY$_X$	20		0.0%		0.0%	3	15.0%	20	100.0%	0	0.0%
Pdfp$_X$	4		0.0%		0.0%	2	50.0%	4	100.0%	0	0.0%
MAD$_X$	4		0.0%		0.0%	2	50.0%	4	100.0%	0	0.0%
Skew$_X$	4		0.0%		0.0%	2	50.0%	4	100.0%	0	0.0%
Kurt$_X$	4		0.0%		0.0%		0.0%	4	100.0%	0	0.0%
Lt$_X$	4		0.0%		0.0%	1	25.0%	4	100.0%	0	0.0%
Ls$_X$	4		0.0%		0.0%	1	25.0%	4	100.0%	0	0.0%
Ms$_X$	4		0.0%		0.0%		0.0%	4	100%	0	0.0%
B1std$_X$	4		0.0%		0.0%		0.0%	2	50.0%	0	0.0%
Pst$_X$	4		0.0%		0.0%		0.0%	3	75.0%	1	25.0%
Rcb$_X$	4		0.0%		0.0%		0.0%		0.0%	0	0.0%
Mr$_X$	4		0.0%		0.0%		0.0%		0.0%	4	100.0%
Mbrp$_X$	4		0.0%		0.0%		0.0%		0.0%	1	25.0%
Total	76	0		2		17		61		6	

Table A.12: Comparative table of the features belonging to the use case SNIa Vs SNII, in the SNPhotCC and PLAsTiCC datasets on GRIZ bands, divided by quartiles of importance.

Appendix B

Confusion Matrices

A confusion matrix is a table where expected and predicted values meet. The horizontal sum of the values gives the real composition of the related class, while the vertical sum gives the composition of the class given by the classification. The elements belonging to the main diagonal represent the correctly classified objects and are called True Positive. All the elements of the first row, with the exception of the true positive, are the False Negatives of the first class and represent the objects which belong to the first class but which have been classified badly. All the elements of the first column, with the exception of the true positive, are the False Positives of the first class and represent objects belonging to the other classes but incorrectly classified in the first. The same goes for the other classes and their rows and columns. Related model is shown in the upper left corner of each confusion matrix with the following acronyms: Random Forest(RF), Nadam(Nadam), RMSProp(RM), Adadelta(AD). For each use case and for each parameter space used in it, there are the confusion matrices of all four models in the best configuration. In the caption the columns and the relative parameter spaces which they refer are indicated in parentheses.



Periodic Vs Non Periodic

RF		Prediction		
		np	p	total
True	np	13603	190	13793
	p	166	13502	13668
	total	13769	13692	

Nadam		Prediction		
		np	p	total
True	np	13553	240	13793
	p	450	13218	13668
	total	14003	13458	

RM		Prediction		
		np	p	total
True	np	13555	238	13793
	p	191	13477	13668
	total	13746	13715	

AD		Prediction		
		np	p	total
True	np	13340	453	13793
	p	636	13032	13668
	total	13976	13485	

Replacing Method – PLAsTiCC

RF		Prediction		
		SN	All	total
True	SN	16897	1109	18006
	All	2805	14188	16993
	total	19702	15297	

RF		Prediction		
		SN	All	total
True	SN	16697	1293	17990
	All	1637	15356	16993
	total	18334	16649	

RF		Prediction		
		SN	All	total
True	SN	16233	1572	17805
	All	2878	13852	16730
	total	19111	15424	

Nadam		Prediction		
		SN	All	total
True	SN	14830	3176	18006
	All	4540	12453	16993
	total	19370	15629	

Nadam		Prediction		
		SN	All	total
True	SN	14024	3966	17990
	All	2604	14389	16993
	total	16628	18355	

Nadam		Prediction		
		SN	All	total
True	SN	15218	2587	17805
	All	3059	13671	16730
	total	18277	16258	

RM		Prediction		
		SN	All	total
True	SN	14970	3036	18006
	All	2619	14374	16993
	total	17589	17410	

RM		Prediction		
		SN	All	total
True	SN	15944	2046	17990
	All	1952	15041	16993
	total	17896	17087	

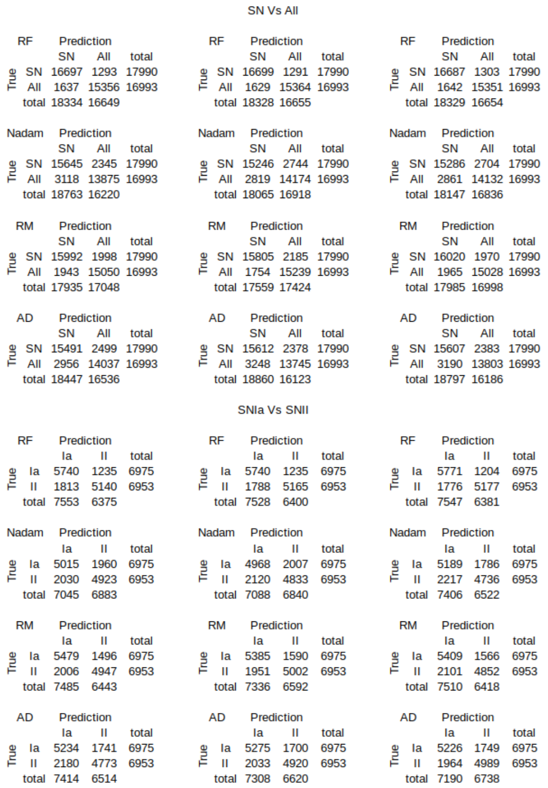
RM		Prediction		
		SN	All	total
True	SN	16156	1649	17805
	All	2419	14311	16730
	total	18575	15960	

AD		Prediction		
		SN	All	total
True	SN	15066	2940	18006
	All	3676	13317	16993
	total	18742	16257	

AD		Prediction		
		SN	All	total
True	SN	15486	2504	17990
	All	2687	14306	16993
	total	18173	16810	

AD		Prediction		
		SN	All	total
True	SN	15553	2252	17805
	All	2760	13970	16730
	total	18313	16222	

Table B.1: Confusion matrices of Periodic Vs Non Periodic (114) use case and replacing methods on PLAsTiCC dataset of SN Vs All (1°-1°method; 2°-2°method; 3°-3°method) use case.



SN Vs All

RF		Prediction		
		SN	All	total
True	SN	16697	1293	17990
	All	1637	15356	16993
	total	18334	16649	

RF		Prediction		
		SN	All	total
True	SN	16699	1291	17990
	All	1629	15364	16993
	total	18328	16655	

RF		Prediction		
		SN	All	total
True	SN	16687	1303	17990
	All	1642	15351	16993
	total	18329	16654	

Nadam		Prediction		
		SN	All	total
True	SN	15645	2345	17990
	All	3118	13875	16993
	total	18763	16220	

Nadam		Prediction		
		SN	All	total
True	SN	15246	2744	17990
	All	2819	14174	16993
	total	18065	16918	

Nadam		Prediction		
		SN	All	total
True	SN	15286	2704	17990
	All	2861	14132	16993
	total	18147	16836	

RM		Prediction		
		SN	All	total
True	SN	15992	1998	17990
	All	1943	15050	16993
	total	17935	17048	

RM		Prediction		
		SN	All	total
True	SN	15805	2185	17990
	All	1754	15239	16993
	total	17559	17424	

RM		Prediction		
		SN	All	total
True	SN	16020	1970	17990
	All	1965	15028	16993
	total	17985	16998	

AD		Prediction		
		SN	All	total
True	SN	15491	2499	17990
	All	2956	14037	16993
	total	18447	16536	

AD		Prediction		
		SN	All	total
True	SN	15612	2378	17990
	All	3248	13745	16993
	total	18860	16123	

AD		Prediction		
		SN	All	total
True	SN	15607	2383	17990
	All	3190	13803	16993
	total	18797	16186	

SNIa Vs SNII

RF		Prediction		
		Ia	II	total
True	Ia	5740	1235	6975
	II	1813	5140	6953
	total	7553	6375	

RF		Prediction		
		Ia	II	total
True	Ia	5740	1235	6975
	II	1788	5165	6953
	total	7528	6400	

RF		Prediction		
		Ia	II	total
True	Ia	5771	1204	6975
	II	1776	5177	6953
	total	7547	6381	

Nadam		Prediction		
		Ia	II	total
True	Ia	5015	1960	6975
	II	2030	4923	6953
	total	7045	6883	

Nadam		Prediction		
		Ia	II	total
True	Ia	4968	2007	6975
	II	2120	4833	6953
	total	7088	6840	

Nadam		Prediction		
		Ia	II	total
True	Ia	5189	1786	6975
	II	2217	4736	6953
	total	7406	6522	

RM		Prediction		
		Ia	II	total
True	Ia	5479	1496	6975
	II	2006	4947	6953
	total	7485	6443	

RM		Prediction		
		Ia	II	total
True	Ia	5385	1590	6975
	II	1951	5002	6953
	total	7336	6592	

RM		Prediction		
		Ia	II	total
True	Ia	5409	1566	6975
	II	2101	4852	6953
	total	7510	6418	

AD		Prediction		
		Ia	II	total
True	Ia	5234	1741	6975
	II	2180	4773	6953
	total	7414	6514	

AD		Prediction		
		Ia	II	total
True	Ia	5275	1700	6975
	II	2033	4920	6953
	total	7308	6620	

AD		Prediction		
		Ia	II	total
True	Ia	5226	1749	6975
	II	1964	4989	6953
	total	7190	6738	

Table B.2: Confusion matrices of SN Vs All (1°-114; 2°-95; 3°-78) and SNIa Vs SNII (1°-114; 2°-85; 3°-78) use cases on PLAsTiCC dataset.

SNIa Vs SNII (GRIZ)

RF		Prediction		
		Ia	II	total
True	Ia	5733	1242	6975
	II	1861	5092	6953
	total	7594	6334	

RF		Prediction		
		Ia	II	total
True	Ia	5721	1254	6975
	II	1843	5110	6953
	total	7564	6364	

RF		Prediction		
		Ia	II	total
True	Ia	5721	1254	6975
	II	1839	5114	6953
	total	7560	6368	

Nadam		Prediction		
		Ia	II	total
True	Ia	5325	1650	6975
	II	2465	4488	6953
	total	7790	6138	

Nadam		Prediction		
		Ia	II	total
True	Ia	5117	1858	6975
	II	2302	4651	6953
	total	7419	6509	

Nadam		Prediction		
		Ia	II	total
True	Ia	4718	2257	6975
	II	1917	5036	6953
	total	6635	7293	

RM		Prediction		
		Ia	II	total
True	Ia	5486	1489	6975
	II	2053	4900	6953
	total	7539	6389	

RM		Prediction		
		Ia	II	total
True	Ia	5417	1558	6975
	II	2033	4920	6953
	total	7450	6478	

RM		Prediction		
		Ia	II	total
True	Ia	5616	1359	6975
	II	2255	4698	6953
	total	7871	6057	

AD		Prediction		
		Ia	II	total
True	Ia	5273	1702	6975
	II	2052	4901	6953
	total	7325	6603	

AD		Prediction		
		Ia	II	total
True	Ia	5240	1735	6975
	II	2024	4929	6953
	total	7264	6664	

AD		Prediction		
		Ia	II	total
True	Ia	5288	1687	6975
	II	2101	4852	6953
	total	7389	6539	

Replacing Method – SNPhotCC

RF		Prediction		
		Ia	II	total
True	Ia	959	58	1017
	II	94	923	1017
	total	1053	981	

RF		Prediction		
		Ia	II	total
True	Ia	990	27	1017
	II	52	965	1017
	total	1042	992	

RF		Prediction		
		Ia	II	total
True	Ia	947	69	1016
	II	95	920	1015
	total	1042	989	

Nadam		Prediction		
		Ia	II	total
True	Ia	934	83	1017
	II	147	870	1017
	total	1081	953	

Nadam		Prediction		
		Ia	II	total
True	Ia	939	78	1017
	II	96	921	1017
	total	1035	999	

Nadam		Prediction		
		Ia	II	total
True	Ia	959	57	1016
	II	83	932	1015
	total	1042	989	

RM		Prediction		
		Ia	II	total
True	Ia	948	69	1017
	II	93	924	1017
	total	1041	993	

RM		Prediction		
		Ia	II	total
True	Ia	977	40	1017
	II	86	931	1017
	total	1063	971	

RM		Prediction		
		Ia	II	total
True	Ia	959	57	1016
	II	67	948	1015
	total	1026	1005	

AD		Prediction		
		Ia	II	total
True	Ia	938	79	1017
	II	113	904	1017
	total	1051	983	

AD		Prediction		
		Ia	II	total
True	Ia	894	123	1017
	II	148	869	1017
	total	1042	992	

AD		Prediction		
		Ia	II	total
True	Ia	937	79	1016
	II	78	937	1015
	total	1015	1016	

Table B.3: Confusion matrices of SNIa Vs SNII (GRIZ) (1°-76; 2°-59; 3°-52) use case on PLAsTiCC dataset and replacing methods on SNPhotCC dataset of SNIa Vs SNII (1°-1°method; 2°-2°method; 3°-3°method) use case.

SNIa Vs SNII – SNPhotCC

RF		Prediction		
		Ia	II	total
True	Ia	990	27	1017
	II	52	965	1017
	total	1042	992	

RF		Prediction		
		Ia	II	total
True	Ia	990	27	1017
	II	52	965	1017
	total	1042	992	

RF		Prediction		
		Ia	II	total
True	Ia	991	26	1017
	II	50	967	1017
	total	1041	993	

Nadam		Prediction		
		Ia	II	total
True	Ia	941	76	1017
	II	114	903	1017
	total	1055	979	

Nadam		Prediction		
		Ia	II	total
True	Ia	945	72	1017
	II	95	922	1017
	total	1040	994	

Nadam		Prediction		
		Ia	II	total
True	Ia	957	60	1017
	II	70	947	1017
	total	1027	1007	

RM		Prediction		
		Ia	II	total
True	Ia	967	50	1017
	II	82	935	1017
	total	1049	985	

RM		Prediction		
		Ia	II	total
True	Ia	977	40	1017
	II	90	927	1017
	total	1067	967	

RM		Prediction		
		Ia	II	total
True	Ia	980	37	1017
	II	68	949	1017
	total	1048	986	

AD		Prediction		
		Ia	II	total
True	Ia	911	106	1017
	II	150	867	1017
	total	1061	973	

AD		Prediction		
		Ia	II	total
True	Ia	934	83	1017
	II	138	879	1017
	total	1072	962	

AD		Prediction		
		Ia	II	total
True	Ia	953	64	1017
	II	91	926	1017
	total	1044	990	

SL SN I Vs SNIa mixed

RF		Prediction		
		SL	Ia	total
True	SL	6553	439	6992
	Ia	1385	5608	6993
	total	7938	6047	

RF		Prediction		
		SL	Ia	total
True	SL	6552	440	6992
	Ia	1409	5584	6993
	total	7961	6024	

RF		Prediction		
		SL	Ia	total
True	SL	6621	371	6992
	Ia	1661	5332	6993
	total	8282	5703	

Nadam		Prediction		
		SL	Ia	total
True	SL	6365	627	6992
	Ia	2609	4384	6993
	total	8974	5011	

Nadam		Prediction		
		SL	Ia	total
True	SL	6222	770	6992
	Ia	2520	4473	6993
	total	8742	5243	

Nadam		Prediction		
		SL	Ia	total
True	SL	6181	811	6992
	Ia	2287	4706	6993
	total	8468	5517	

RM		Prediction		
		SL	Ia	total
True	SL	6422	570	6992
	Ia	1815	5178	6993
	total	8237	5748	

RM		Prediction		
		SL	Ia	total
True	SL	6360	632	6992
	Ia	1464	5529	6993
	total	7824	6161	

RM		Prediction		
		SL	Ia	total
True	SL	6379	613	6992
	Ia	2218	4775	6993
	total	8597	5388	

AD		Prediction		
		SL	Ia	total
True	SL	4774	2218	6992
	Ia	1916	5077	6993
	total	6690	7295	

AD		Prediction		
		SL	Ia	total
True	SL	5953	1039	6992
	Ia	2228	4765	6993
	total	8181	5804	

AD		Prediction		
		SL	Ia	total
True	SL	4972	2020	6992
	Ia	2056	4937	6993
	total	7028	6957	

Table B.4: Confusion matrices of SNIa Vs SNII (1°-76; 2°-70; 3°-52) use case on SNPhotCC dataset and SLSNe Vs SNIa mixed (1°-114; 2°-99; 3°-78) use case on PLAsTiCC dataset.

Six Class Problem

RF		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	5396	0	0	1576	0	7	6979
	Iabg	0	1766	1225	0	3324	670	6985
	Iax	0	177	2282	1	3157	1379	6996
	II	1459	0	0	5474	1	21	6955
	Ibc	5	210	418	2	4467	1885	6987
	Sl	6	6	19	319	290	6352	6992
	total	6866	2159	3944	7372	11239	10314	

RF		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	5395	0	0	1577	0	7	6979
	Iabg	0	1876	1249	0	3144	716	6985
	Iax	0	217	2433	1	2921	1424	6996
	II	1442	0	0	5492	1	20	6955
	Ibc	3	286	511	4	4234	1949	6987
	Sl	9	7	23	310	266	6377	6992
	total	6849	2386	4216	7384	###	10493	

RF		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	5403	0	0	1567	1	8	6979
	Iabg	0	2111	1319	0	2765	790	6985
	Iax	0	272	2615	1	2576	1532	6996
	II	1421	0	0	5513	1	20	6955
	Ibc	4	312	612	2	3994	2063	6987
	Sl	10	10	29	311	236	6396	6992
	total	6838	2705	4575	7394	9573	10809	

Nadam		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	4895	14	3	1845	19	203	6979
	Iabg	29	1083	1837	7	2868	1161	6985
	Iax	30	447	1801	24	2702	1992	6996
	II	2952	7	3	3831	8	154	6955
	Ibc	65	572	941	36	2577	2796	6987
	Sl	72	32	69	1065	265	5489	6992
	total	8043	2155	4654	6808	8439	11795	

Nadam		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	4241	16	1	2560	41	120	6979
	Iabg	34	1408	1563	10	3182	788	6985
	Iax	70	617	1389	20	3419	1481	6996
	II	2112	11	3	4688	13	128	6955
	Ibc	95	718	679	29	3300	2166	6987
	Sl	159	37	31	968	537	5260	6992
	total	6711	2807	3666	8275	10492	9943	

Nadam		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	3554	1	0	3241	5	178	6979
	Iabg	32	1441	630	23	3345	1514	6985
	Iax	46	257	933	52	3321	2387	6996
	II	1352	0	0	5401	4	198	6955
	Ibc	80	312	245	97	3024	3229	6987
	Sl	58	9	9	940	220	5756	6992
	total	5122	2020	1817	9754	9919	13262	

RM		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	4956	1	1	1960	3	58	6979
	Iabg	19	1606	1301	1	3164	894	6985
	Iax	27	200	2265	5	3079	1420	6996
	II	1669	1	0	5217	0	68	6955
	Ibc	53	265	628	9	3515	2517	6987
	Sl	98	7	35	770	342	5740	6992
	total	6822	2080	4230	7962	10103	10697	

RM		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	4785	1	1	2099	9	84	6979
	Iabg	18	1453	1189	1	3410	914	6985
	Iax	25	202	1841	6	3535	1387	6996
	II	1554	0	0	5281	6	114	6955
	Ibc	47	289	496	12	3733	2410	6987
	Sl	58	9	21	710	345	5849	6992
	total	6487	1954	3548	8109	11038	10758	

RM		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	4861	0	1	2016	7	94	6979
	Iabg	13	1178	1329	2	3644	819	6985
	Iax	23	94	1784	4	3562	1529	6996
	II	1472	0	0	5329	4	150	6955
	Ibc	36	169	406	12	4017	2347	6987
	Sl	41	3	24	416	326	6182	6992
	total	6446	1444	3544	7779	11560	11121	

AD		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	3956	40	7	2924	6	46	6979
	Iabg	34	1937	1597	18	2304	1095	6985
	Iax	61	994	1443	34	2577	1887	6996
	II	2377	12	3	4534	2	27	6955
	Ibc	112	1132	846	93	2536	2268	6987
	Sl	59	79	107	916	547	5284	6992
	total	6599	4194	4003	8519	7972	10607	

AD		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	3506	22	3	3353	16	79	6979
	Iabg	39	1706	1579	12	2205	1444	6985
	Iax	57	861	1460	30	2279	2309	6996
	II	2166	12	0	4714	2	61	6955
	Ibc	132	859	703	73	2213	3007	6987
	Sl	58	36	38	856	332	5672	6992
	total	5958	3496	3783	9038	7047	12572	

AD		Prediction						
		Ia	Iabg	Iax	II	Ibc	SL	total
True	Ia	2910	1	1	3894	10	163	6979
	Iabg	27	1370	1424	4	2750	1410	6985
	Iax	37	565	1114	15	2882	2383	6996
	II	1923	0	1	4902	5	124	6955
	Ibc	95	456	516	20	2616	3284	6987
	Sl	100	11	17	494	240	6130	6992
	total	5092	2403	3073	9329	8503	13494	

Table B.5: Confusion matrices of Six Class Problem use case on PLAsTiCC dataset. For each model there are the following confusion matrices, ordered according to the lines: 114-96-78.

SNIa VsSNII – LSTM

		Prediction		
		np	p	total
True	np	885	132	1017
	p	553	1851	2404
	total	1438	1983	

		Prediction		
		np	p	total
True	np	914	103	1017
	p	218	2186	2404
	total	1132	2289	

Table B.6: Confusion matrices of SNIa Vs SNII use case on SNPhotCC dataset with direct approach. Band number test (left) and best model optimization test (right).

Appendix C

Setup of Models

In this appendix we report the parameter setup for all the classification models, used for the experiments.

LSTM	Parameters
Dropout (*)	0.5 or 0.35
Hidden layers number	2
Neuron hidden layers	16 or 25
Learning rate	0.01 or 0.02
Iterations (I)	200 or 250
Step Learning rate	$2/3 * I$ and/or $9/10 * I$
Gamma	0.5
Augmentation	5 or 10
Weight Decay	10^{-5}

Table C.1: Summary table of the LSTM parameters. (*) The dropout value of 0.5 was used to perform the test on the number of bands, while the value 0.35 is the best value obtained after the optimization of the model.

Classifier	Parameter	SNPhotCC	PLAsTiCC
Random Forest	Features number	76-70-52	(*)
	Trees number	1000	1000
	Max depht	None	None
	Min samples split	2	2
	Min samples leaf	1	1
Nadam - RMSProp - Adadelta	Features number (n)	76-70-52	(*)
	Hidden layers number	2	2
	First hidden layer	2n+1	2n+1
	Dropout first layer	0.1	0.1
	Second hidden layer	n-1	n-1
	Dropout second layer	0.1	0.1
	Max iterations	1000	1000
	Learning rate(°)	0.001	$0.001 - 0.0005$
	Decay(+)	$10^{-5} - 10^{-7}$	$10^{-4} - 10^{-5}$
	Epsilon	0.01	0.01
	Beta 1	0.9	0.9
	Beta 2	0.999	0.999

Table C.2: Summary table of the Random Forest, Nadam, RMSProp and Adadelta parameters. (*) The number of features changes in different use cases: P Vs NP (114), SN Vs All (114-95-78), SNIa Vs SNII (114-85-78), SNIa Vs SNII (GRIZ)(76-59-52), SLSNI Vs SNIa mixed (114-99-78), Six class problem (114-96-78). (°) In the PLAsTiCC dataset the best value of the learning rate has varied in the different use cases and in the various spaces of the parameters. The value 0.001 has been adopted for SNIa Vs SNII (85), SNIa Vs SNII (GRIZ) (76-59-52), Six class problem (114-96-78). The value 0.0005 has been adopted for SN Vs All (114-95-78), SNIa Vs SNII (114-78), SLSNe Vs SNIa mixed (114-99-78). (+) In the PLAsTiCC dataset the best decay value has varied in the different use cases and in the various parameter spaces. The value 10^{-4} has been adopted for SLSN Vs SNIa mixed (114-78), Six class problem (114-96-78). The value 10^{-5} in all other cases. In the SNPhotCC dataset there is the decay value 10^{-7} in the case SNIa Vs SNII (70) and 10^{-5} in the cases SNIa Vs SNII (76-52).

Appendix D

Training and test set distributions

In this appendix we show a series of examples of histograms for every kind of experiments, with a superimposed distribution of training and test sets on G band, with respect to the Ampl statistical feature, extracted from light curves. The y axis is in logarithmic scale. The distributions show that, as expected, the training and test sets are uniformly distributed. Same behaviour is present for every distributions in the other bands and for all statistical features, derived from light curves of data samples used.

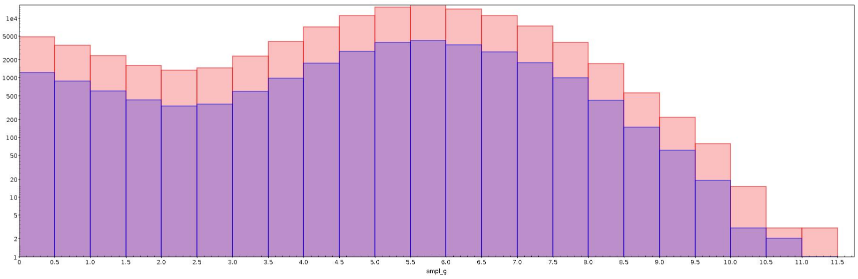


Figure D.1: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for Periodic Vs No Periodic experiments.

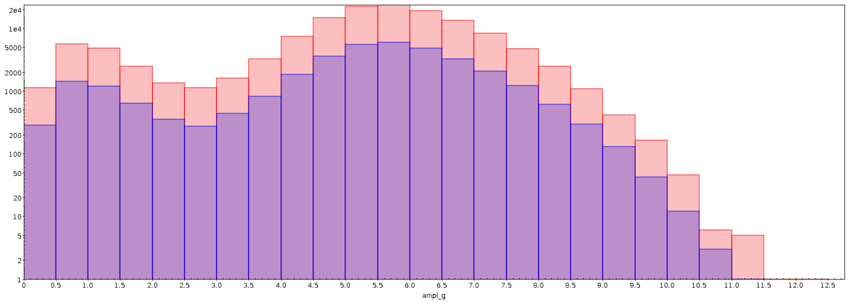


Figure D.2: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for SNe Vs All experiments with the first negative fluxes replacing method.

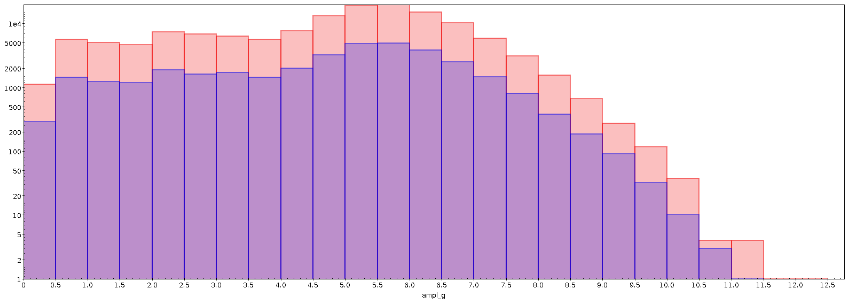


Figure D.3: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for SNe Vs All experiments with the second negative fluxes replacing method.

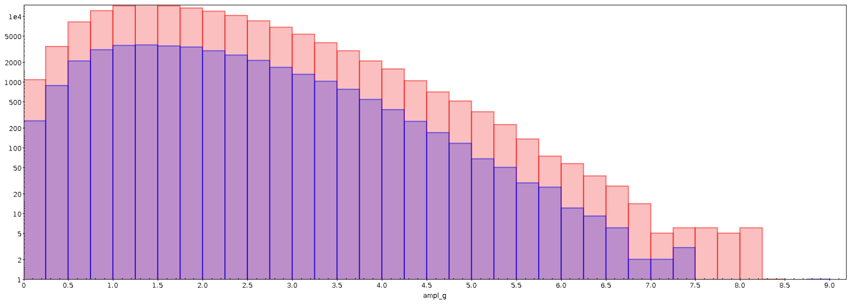


Figure D.4: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for SNe Vs All experiments with the third negative fluxes replacing method.

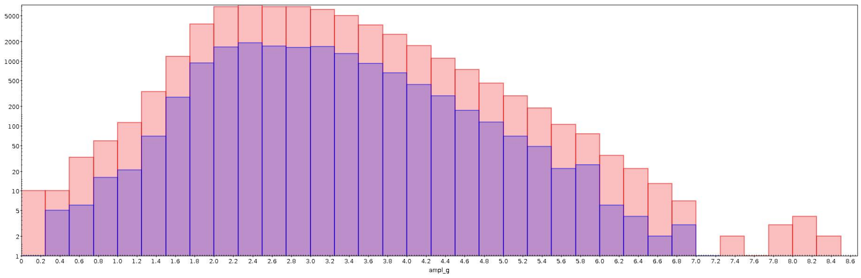


Figure D.5: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for SN Ia Vs SN II experiments.

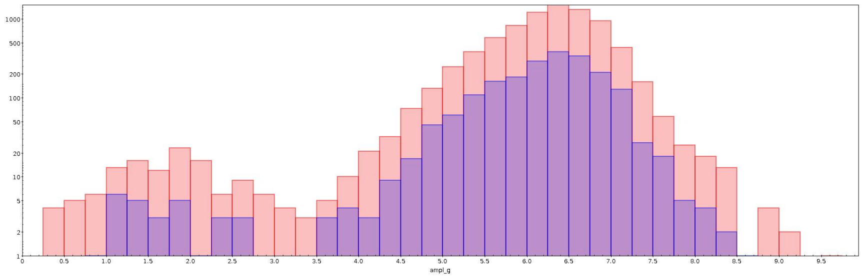


Figure D.6: Data distributions of training (red) and test (blue) sets, on SNPhotCC dataset, for the first negative fluxes replacing method.

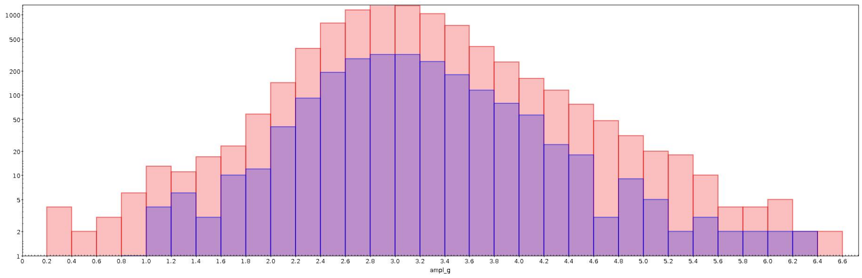


Figure D.7: Data distributions of training (red) and test (blue) sets, on SNPhotCC dataset, for the second negative fluxes replacing method.

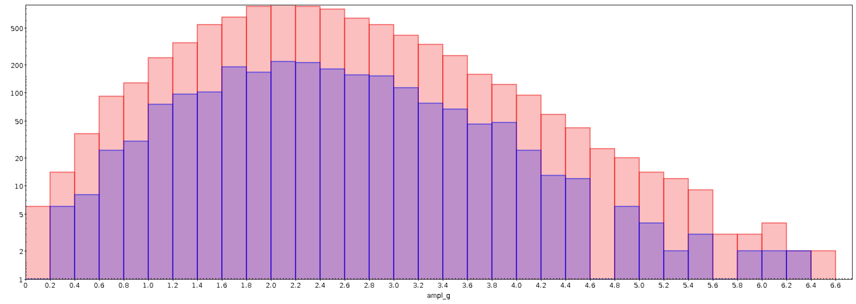


Figure D.8: Data distributions of training (red) and test (blue) sets, on SNPhotCC dataset, for the third negative fluxes replacing method.

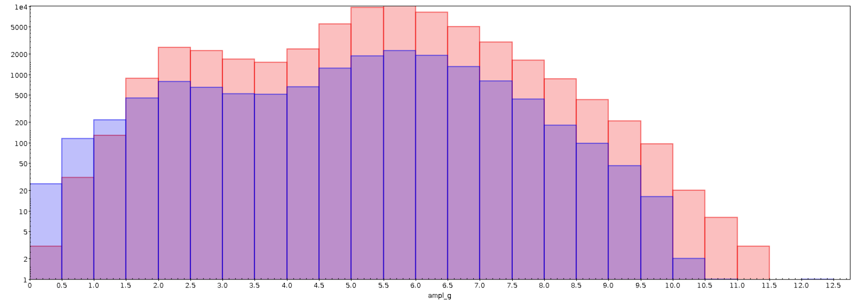


Figure D.9: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for Superluminous SN I Vs SN Ia mixed experiments.

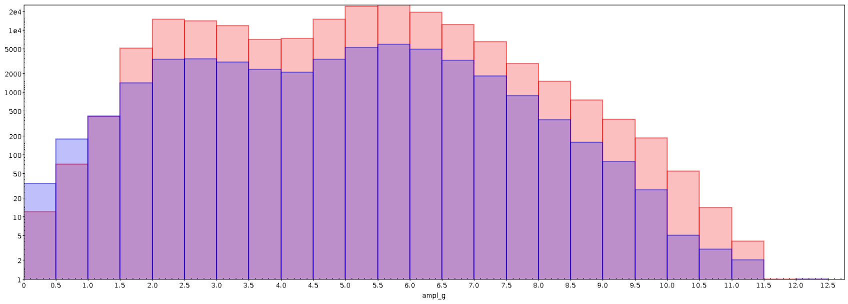


Figure D.10: Data distributions of training (red) and test (blue) sets, on PLAsTiCC dataset, for Six Class Problem experiments.

Bibliography

[1] Baklanov P.V. et al.,, *Study of Supernovae Important for Cosmology*, arXiv:1502.06798.

[2] Bazin G. et al., *The Core-collapse rate from the Supernova Legacy Survey*, arXiv:0904.1066v1, 2009.

[3] Branch D., Wheeler J.C., *Supernovae Explosion*, Springer, Berlin, 2017.

[4] Breiman L., *Random Forests*, Machine Learning, Vol. 45, pp. 5-32, 2001.

[5] Brescia M. et al., *MNRAS*, 489, 1, pp. 663-680, 2019.

[6] Carroll B.W., Ostlie D.A., *An Introduction to Modern Astrophysics*, Pearson, San Francisco, 2007.

[7] Cerda-Duran P., Elias-Rosa N., *Neutron stars formation and Core Collapse Supernovae*, arXiv:1806.07267v1, 2018.

[8] Charnock T., Moss A., *Deep Recurrent Neural Networks for Supernovae Classification*, arXiv:160607442v2, 2017.

[9] Cox A.N., *Allen's Astrophysical Quantities* cap.18, Springer, New York, 2002.

[10] D'Isanto A. et al., *An analysis of feature relevance in the classification of astronomical transients with machine learning methods*, arXiv:1601.03931v1, 2016.

[11] Dai L. et al., *A unified model for tidal disruption events*, arXiv:1803.03265v1, 2018.

[12] Delli Veneri M. et al., *MNRAS*, 486, 1, pp. 1377-1391, 2019.

[13] Dilday B. et al., *A Measurement of the Rate of Type Ia Supernovae at Redshift* $z \sim 0.1$ *from the First Season of the SDSS-II Supernova Survey*, The Astrophysical Journal, 682:262Y282, 2008.

[14] Gal-Yam A. *Luminous Supernovae*, arXiv: 1208.3217v1, 2012.

[15] Girola-Schneider R., *Type Ia supernovas and fusion of black holes: Do they complement each other in measuring the expansion of the universe?*, PoS(BHCB2018)024.

[16] Goobar A., Leibundgut B., *Supernova cosmology: legacy and future*, arXiv:1102.1431v1, 2011.

[17] Hensler G., *The Supernova - ISM/Star-formation interplay*, arXiv:1402.0117v1, 2014.

[18] Huber S. et al., *Strongly lensed SNe Ia in the era of LSST: observing cadence for lens discoveries and time delay measurements*, arXiv:1903.00510v2, 2019

[19] Jha S.W., *Type Iax Supernovae*, arXiv:1707.01110v2, 2017.

[20] Z. et al., *ApJ*, 873, 2, 111, 2019.

[21] Jones D.O. et al., *The Foundation Supernova Survey: Measuring Cosmological Parameters with Supernovae from a Single Telescope*, The Astrophysical Journal, 881:19.

[22] Kessler R. et al., *Results from the Supernova Photometric Classification Challenge*, The Astronomical Society of the Pacific, 2010.

[23] Knobel C., *An Introduction into the Theory of Cosmological Structure Formation*, arXiv:1208.5931v2.

[24] Komossa S., Zensus J.A., *Compact object mergers: observations of supermassive binary black holes and stellar tidal disruption events*, Cambridge University Press, Cambridge, 2014.

[25] Longhair M.S., *High Energy Astrophysics*, Cambridge University Press, Cambridge, 2011.

[26] LSST Science Collaborations, *LSST Science Book*, 2009.

[27] Lyman J.D. et al., *Investigating the diversity of supernovae type Iax: A MUSE and NOT spectroscopic study of their environments*, arXiv:1707.04270v1.

[28] Mao S., *Astrophysical Applications of Gravitational Microlensing*, Research in Astronomy and Astrophysics, 2012.

[29] Metzger B.D. et al., *Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei*, arXiv:1001.5029v2, 2010.

[30] Scolnic D. et al., *Astro2020 Science White Paper The Next Generation of Cosmological Measurements with Type Ia Supernovae*, arXiv:1903.05128v2

[31] Tan P.N., Steinbach M. and Kumar V., *Introduction to Data Mining*,Pearson,U.S.A., 2014.

[32] Taubenberger S., *The Extremes of Thermonuclear Supernovae*, arXiv 1703.00528, 2017.

[33] The PLAsTiCC Team, *The Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC): Selection of a performance metric for classification probabilities balancing diverse science goals*, arXiv:1809.11145v1, 2018.

[34] The PLAsTiCC Team, *The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Data set*, arXiv:1810.00001v1, 2018.

[35] Whitelock P.A., *"Real-Time" Evolution in Mira Variables*, arXiv:9801002v1, 1998.

[36] Yang H. et al., *The Flaring Activity of M Dwarfs in the Kepler Field*, The Astrophysical Journal, 849:36 (15pp), 2017.

[37] Zeiler M.D., *Adadelta: An adaptive learning rate method*, arXiv:1212.5701v1, 2012.

Sitography

[38] Castellani V., *Astrofisica Stellare.*
http://astrofisica.altervista.org/doku.php?id=start

[39] Olah C., *Understanding LSTMs.*
http://colah.github.io/posts/2015-08-Understanding-LSTMs

[40] Pytorch documentation
https://pytorch.org/docs/stable/nn.html

[41] Ray Li, *History of data mining*
https://hackerbits.com/data/history-of-data-mining

[42] The PLAsTiCC Team
https://www.kaggle.com/michaelapers/the-plasticc-astronomy-starter-kit