



PHYSICAL SCIENCES

Intelligent Supernovae Classification Systems in the KDUST context

LUÍS R. ARANTES FILHO, REINALDO R. ROSA & LAMARTINE N.F. GUIMARÃES

Abstract: With the advent of large astronomical surveys plus multi-messenger astronomy, both automatic detection and classification of Type Ia supernovae have been addressed by different machine learning techniques. In this article we present three solutions aimed at the future spectrometer of the KDUST project, within a scope of benchmark, considering three different methodologies. The systems presented here are the following: CINTIA (based on hierarchical neural network architecture), SUZAN (which incorporates the solution known as fuzzy systems) and DANI (based on Deep Learning with Convolutional Neural Networks). The characteristics of the systems are presented and the benchmark is performed considering a data set containing 15,134 spectra. The best performance is obtained by the DANI architecture which provides 96% accuracy in the classification of Type Ia supernovae in relation to other spectral types.

Key words: Deep Learning, Artificial Intelligence, Supernovae, Type Ia Supernova, Classification.

INTRODUCTION

One of the most important types of extreme cosmic events is the Supernovae (SNe) explosions. According to Filippenko (1997) and Horvath (2011) the supernovae represent the explosive ending of a star and it releases vast amounts of energy and luminosity in the process (as bright as the host galaxy). Regarding the importance of these explosions, we highlight the thermonuclear supernovae called Type Ia Supernovae (SNIa) (Filippenko 1997). Actually, the automatic search of SNIa is feasible because both the total luminosity of the explosion and the spectral data follows typical patterns. Moreover, these characteristics make this object a standard candle for measurements of cosmological distances (Perlmutter et al. 1999, Riess et al. 1998).

Other classic types of SNe are related to the core-collapse of massive stars, they are classified as Type Ib Supernovae (SN Ib), Type Ic (SN Ic), and Type II (SN II) (Filippenko 1997, Horvath 2011, Blondin et al. 2012). The difference between SNIa and the other types, in addition to the explosion mechanism, is the presence of Hydrogen (H) or Helium (He) in the spectral data, which is observed only in the spectra of core-collapse supernovae.

The study of SNe involves also the understanding of the accelerated expansion of the universe. Some important works that are related to it are the *Dark Energy Survey* (Brenna 2005), *Supernova Legacy Survey* (Astier et al. 2006), and the *ESSENCE* (Wood-Vasey et al. 2007). For these parameters to be improved, structures are required that involve instruments with a high capacity for detecting extreme cosmic events

and data analysis models that accurately identify these events.

Important Instruments like *Large Synoptic Survey Telescope* (LSST) (Huber et al. 2019) and *Kunlun Dark Universe Survey Telescope* (KDUST) (Li et al. 2019) will have the ability to produce a significant amount of data, producing about Terabytes of data per hour (Graham et al. 2019). A such amount of data requires computational solutions for data science based on machine learning approaches.

The KDUST telescopes (Yuan et al. 2012, Burton et al. 2016) to be installed in 2022-2025 at the Chinese Antarctic Kunlun Station, located on the Antarctic plateau, has as one of its research focuses on the detection and analysis of extreme cosmic events (Yuan et al. 2012, Burton et al. 2016) based on machine learning solutions. This telescope has as one of its main objectives the study of SNIa to provide new insights into the dark energy research.

The KDUST will have a good capacity for observations ranging from optical to infrared & sub-mm wavelengths. Regarding its instrumentation, this telescope has a proposed diameter of 2.5 m. KDUST will adopt an innovative optical system that can deliver very good image quality over a 2 square degree flat field.

In this work we address the development of a technique based on Deep Learning, for identification, treatment, and classification of SNe automatically into the context of KDUST telescopes. We present CINTIA and SUZAN systems that classify SNe using machine learning solutions and we propose a new model that uses a deep learning solution.

The proposed model attempts to evaluate SNe data with an adaptation of Convolutional Neural Networks (CNN) for mapping spectral features and thus allowing a more accurate and consistent analysis. The model was named DANI

acronym for Deep Architecture for superNovae Identification.

Data from different collections were submitted to DANI system, such as the collection of the Online Supernova Spectrum Archive (Richardson et al. 2002), the Open Supernova Catalog (Guillochon et al. 2017) and the CfA Supernova Archive (CfA 2018) resulting in the amount of 15.134 spectra of the classic types of SNe. Concerning the concept of multi-messenger astrophysics, we briefly present a data structure that deals with multiple sources from extreme cosmic events.

The manuscript is organized as follows. In the section Machine Learning Solutions, we detail some concepts about the supernovae classification and present the machine learning solutions for SNe classification. In the section Deep Learning Solutions, we detail some concepts about deep learning methods and outline the DANI system algorithm. Next we discuss a data analysis for the spectral data and present a data structure for Multi-Messenger Astronomy. In the Results section, we describe the benchmark results regarding the performance of DANI, CINTIA, and SUZAN systems. The last section presents the main concluding remarks of this study.

MACHINE LEARNING SOLUTIONS

Supernovae Classification

The evolution of the supernova stage can be verified in a deep analysis of the data collected after the explosion of a star that corresponds to either the radiation flux spectrum and the light curve.

This analysis results in the verification of the type of explosion that can be thermonuclear, so that, related to the mass accretion reactions in white dwarfs or by the core collapse of massive stars. The Figure 1 shows the data of an

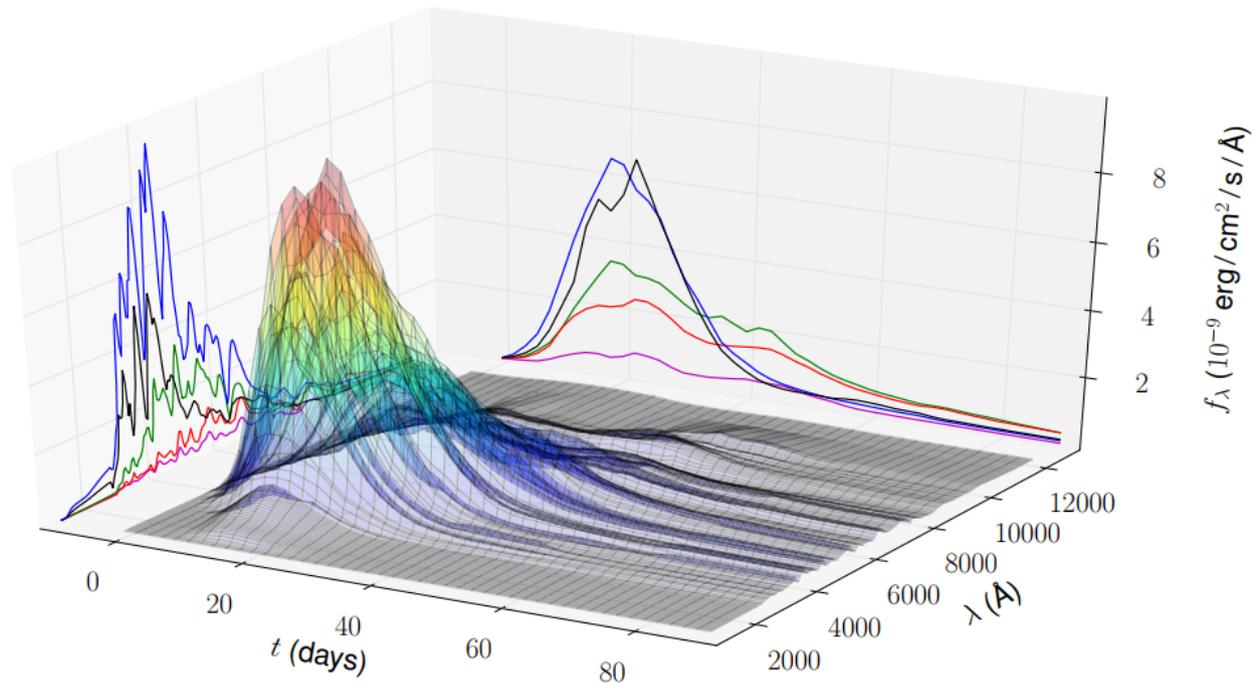


Figure 1. 3D Representation for a SNIa. Adapted from Coelho et al. (2014).

SNIa indicating in a space of 3 dimensions the luminosity, wavelength, and radiation flux as a function of time, showing both the light curves and the light spectra of this object.

Thermonuclear supernovae are generated by explosions related to mass accretion reactions in white dwarf stars (composed essentially of Carbon and Oxygen in degenerate conditions), this type of reaction occurs in multiple systems where a white dwarf absorbs mass from a companion (which can be a star at the main sequence, red giants or white dwarfs). At a certain point in the process, when the white dwarf reaches the mass of $\approx 1.4M_{\odot}$, the star collapses, triggering thermonuclear reactions that destroy the star, this explosion is called thermonuclear supernovae, classified as **SNIa** (Filippenko 1997, Horvath 2011).

The stars have a mechanism that allows the balance between the hydrogen fusion processes and the gravitational force. When a star consumes a large part of Hydrogen (fuel for

internal fusion processes), instability is created between pressure (caused by nuclear fusion) and gravitational force, contributing to the expulsion of matter in space and the fusion of other elements such as Helium, Carbon, Oxygen until it reaches the Iron core (Horvath 2011).

When the fusion process of the Iron core is started, the star collapses exploding. This process is irreversible and the entire envelope of the star (outer layers composed by Hydrogen, Helium, Carbon, Oxygen, etc.) collides against its core, which in turn ricochets all matter into space. This type of explosion is called core collapse supernovae and occurs in massive stars. Core collapse supernovae are classified into three main types: **SNIb**, **SNIc**, **SNIIf** (Filippenko 1997, Blondin et al. 2011, Modjaz et al. 2014).

Intelligent Systems for SNe Classification

In this section, we present some works that played an important role in the development of the DANI supernovae classifier system. These

works are part of an important initiative by INPE and the Institute for Advanced Studies (IEAv) in the development of autonomous and intelligent systems to classify SNe. CINTIA (counting its respective versions, CInt1a and CINTIA 2) (Módolo et al. 2015, do Nascimento et al. 2019) and SUZAN (Arantes Filho et al. 2019, 2020) systems were developed following different models in the Artificial Intelligence field. CINTIA (an acronym for Classificador Inteligente de superNovas Tipo Ia, in English Intelligent Classifier of Type Ia Supernovae) and SUZAN (an acronym for Sistema fUZzy Avaliador de superNovas, in English Fuzzy System for Supernovae Evaluation) were based on the Artificial Neural Networks and Fuzzy Logic paradigms respectively.

CINTIA. This system uses 4 artificial neural networks to classify SNIa, SNIb, SNIc, and SNII (do Nascimento et al. 2019). CINTIA has more diversity in its learning, provided by a hierarchical learning structure that connects Artificial Neural Networks in an integrated system that allows a more secure and unambiguous classification. CINTIA has a computational improvement that includes a new approach to filtering and processing spectral data, the Double Filtering System **SDF-SG** (Arantes Filho et al. 2019), ensuring a better quality of the information to be trained on the neural networks. CINTIA classified about 9000 SNe spectra from several databases and reach good precision and accuracy scores.

SUZAN. This system evaluates the supernovae spectrum identifying in the spectrum the basic chemical elements that allow the classification of classic types of supernovae. This system explores the spectral lines of absorption and emission for the elements Silicon, Sulfur, Hydrogen, Helium, Iron, Oxygen, etc. to separate the thermonuclear SNIa and core collapse SNe. SUZAN can simulate a specialist astronomer who deals with the classification of supernovae

as explained by (Turatto et al. 2007), using Fuzzy rules to identify in the spectral lines its corresponding elements. All the parameters (intensity of spectral lines and equivalent width) found by SUZAN can be modeled by fuzzy functions and all classic types can be classified by their chemical elements (Arantes Filho et al. 2019, 2020). SUZAN classifies about 3082 SNe spectra, obtaining more expressive results for SNIa classification for spectra near the time of the maximum luminosity of the explosion. Like CINTIA, SUZAN also uses optimized spectra by the SDF-SG system.

Double Filtering System SDF-SG. The SDF-SG (an acronym for Sistema de Dupla Filtragem pelo filtro de Savitzky-Golay, in English Double Filtering System by Savitzky-Golay filter) is not defined as a supernovae classifier, but as a step of data optimization. The good results reached in the SNe classification made by CINTIA and SUZAN were improved by a previous optimization stage done into raw spectral data. This system can remove inconsistencies and noise from raw spectral data performing a normalization of the SNe spectra that consists of double filtering by the Savitzky Golay filter (Savitzky & Golay 1964), thus, with the filtered spectrum the main spectral lines sensitive to the classification of SNe become more evident (Arantes Filho et al. 2019). The Figure 2 shows the optimization process made by **SDF-SG** system.

Related Solutions in the Machine Learning Field

In Markel & Bayless (2019), the authors present a method using Random Forest Algorithms to perform a Binary classification of Supernovae, identifying Type Ia supernovae and core collapse supernovae by analyzing the light curve. Santos et al. (2020) in turn seeks the exhaustive test of several Machine Learning techniques, analyzing supernova light curves and developing

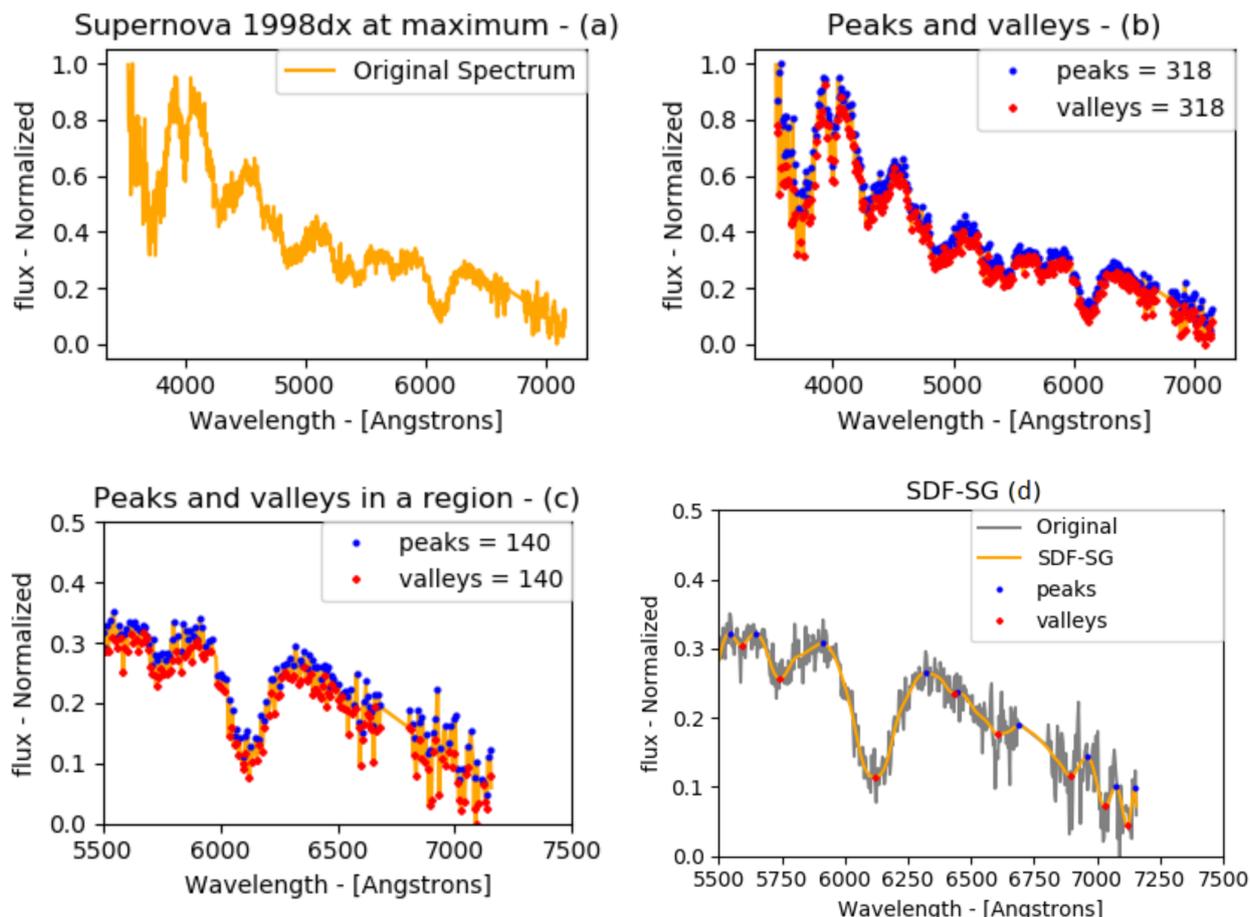


Figure 2. SDF-SG operation. Adapted from Arantes Filho et al. (2019). In (a) we show a raw SNIa spectrum, in (b) we made a search for the peaks and valleys (absorption and emission lines) of this spectrum. In (c) we show a region of this spectrum that shows a large number of peaks and valleys in a noisy area. In (d) we show this same region with the treatment by the SDF-SG system that reveals only the real peaks and valleys of the supernova spectrum.

a decision tree method for separating the classic types.

DEEP LEARNING SOLUTIONS

The techniques and fundamentals of Deep Learning are related to artificial neural networks (they are the new generation of neural networks), to explore and amplify the power of nonlinear data analysis using a large number of intermediate layers of processing. This concept is strictly related to the areas of Machine Learning (ML) and Computational Intelligence (CI).

According to Burkov (2019) ML can be understood as computational processes

developed to perform tasks in such a way as to simulate the human ability to obtain the best solution for a given problem. The techniques related to CI are similar to ML techniques, however, they have different inspirations, according to Keller et al. (2016) CI techniques are inspired by systems of nature and biological behaviors for the development of computational models able to performing tasks and generating intelligent solutions for several problems.

Deep Learning can be defined as a class of techniques in ML and CI that exploit a non-linear analysis in many layers hierarchically. Deep Learning methods use supervised learning

(where data are labeled and identified by classes), unsupervised learning (where data has no labels and are grouped by similarity checks), and Hybrid (junction of supervised and unsupervised learning) to perform different learning tasks (Fausett et al. 1994, Haykin 2001, Mohri et al. 2018, Manaswi et al. 2018).

Related Solutions in the Deep Learning field

The DASH system (Deep Automated Supernova and Host classifier) proposed by Muthukrishna et al. (2019), is an automatic system that seeks to classify the type, age, redshift, and host galaxy of supernovae. This system performs the SNe classification based on characteristics learned by convolutional neural networks over a set of 3899 spectra from 403 SNe. Still in the field of CNN networks Brunel et al. (2019) present a system adapted to classify supernovae by their light curves inferring the classic types of SNe. Kimura et al. (2017) in turn, follows a line for the classification of SNe images presenting a method for classifying SNIa simply from single-epoch observation images without any complex measurements into the standard photometric approach.

Convolutional Neural Networks

Convolutional Neural Networks (CNN), are neural networks that have a deep and hierarchical architecture, that is, CNN networks can extract information from raw data and represent it at many levels of information, so that, from the simplest representations to the most complex. This type of neural network is commonly applied to problems such as image classification, object recognition, and other problems related to computer vision.

The first networks with the concept of deep architectures and convolution operations were proposed by LeCun et al. (1995, 1998) and were

called LeNets. LeNets networks were developed for the recognition of patterns in images, specifically for the recognition of characters. This neural network generated good results reaching 99% precision and accuracy for the classification of the MNIST database (Kim 2014). The MNIST database corresponds to a sequence of handwritten digits, so that the same character can have several representations. The Figure 3 shows the LeNet architecture (LeCun et al. 1998) indicating its main components.

The LeNet neural network described in Figure 3 has a 7-layer architecture (the input layer is not counted), with three layers for convolution operation, two layers for sampling, and two fully connected layers that include the output layer. Convolutional neural networks have similarities to the classic model of neural networks and therefore have a final layer, called a fully connected layer. The CNN architecture (Figure 3) can be described in four components:

1. The input layer is defined as a multidimensional matrix, describing the data. This layer can hold data with 1D, 2D, and 3D.
2. Convolutional layers (C_n) to handling features;
3. Sampling layers (Pooling Layers) (S_n) to reduce the features obtained by the convolutional layers;
4. Fully Connected Layers (F_n) that receive the information processed by the previous layers.

Convolution Operation

The convolutional layer is responsible for processing raw data to retrieve information from this data, generally, this layer consists of filters and mappings on the data, to obtain local

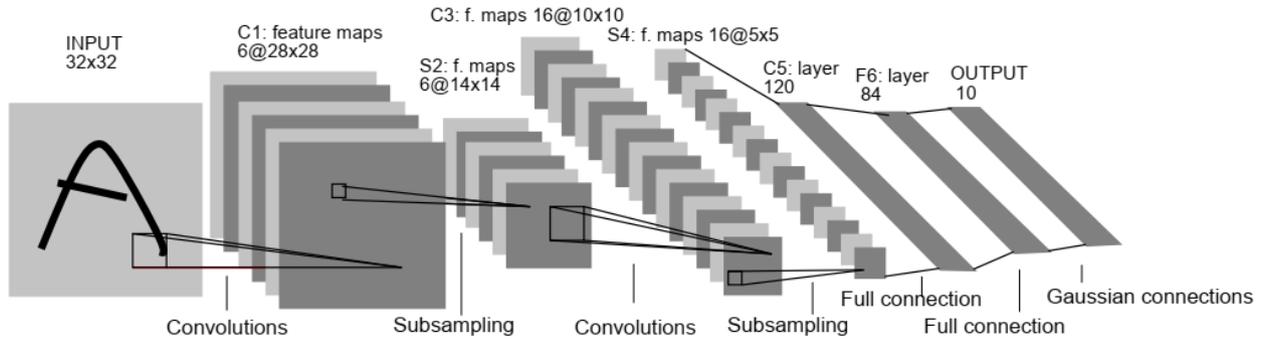


Figure 3. Convolutional Neural Network Lecun Lenet. Adapted from LeCun et al. (1995).

patterns from multidimensional data (Chollet et al. 2015).

The convolution operation can extract features from data input preserving the spatial relationships between pixels, learning local patterns using small matrices of synaptic weights (W^k) with previously defined sizes that can extract edge information, color, intensity, etc. These matrices are called **Filters**, the Filters perform the feature mapping in multidimensional data. Figure 4 illustrates how Filters can perform operations on the input data.

The Figure 4 shows the input data as a 5x3x3 dimension matrix (an image with height, width, and depth) and a 2x2x3 dimension Filter. Convolution is performed through the scalar product between a region of the data (with the dimensions of the filter, that is, a region of size 2x2x3) and the Filter. Then, the Filter is moved to another region and the scalar product is performed again until the entire data can be covered. The Equation 1 indicates how the scalar product is calculated. This same operation can be reproduced for data of different dimensions as it is explored in this work.

$$A \cdot B = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \quad (1)$$

Pooling Operation

The **Pooling** layer samples and reduces the output values from the feature mapping made by the convolutional layer. The purpose of this operation is to reduce the size of the mapping obtaining only the most important features from the data. This operation represents a local pattern learned in single output value.

The Max-pooling operation performs the sampling on the map of features generated by the convolution, partitioning it into regions (matrices with predefined dimensions) and calculating the maximum value for each region. The Figure 5 shows a pooling operation (Max-pooling) considering the result of a convolution operation as a feature matrix of dimensions 4x4 and a Filter with 2x2 dimension. In each of the defined regions, the maximum values are extracted by the application of the maximum function $Max(x)$ (Figure 5 (a)). In addition to the max-pooling operation, it is also possible to use the average function (Average Pooling) (Figure 5(b)) (Vasilev et al. 2019).

Flatten Layer and Fully Connected Layer

After the stage of feature extraction and mapping made by the convolutional and pooling layers, all these learned parameters are inserted in Fully Connected layers and Flatten layers. These layers behave like classic neural networks.

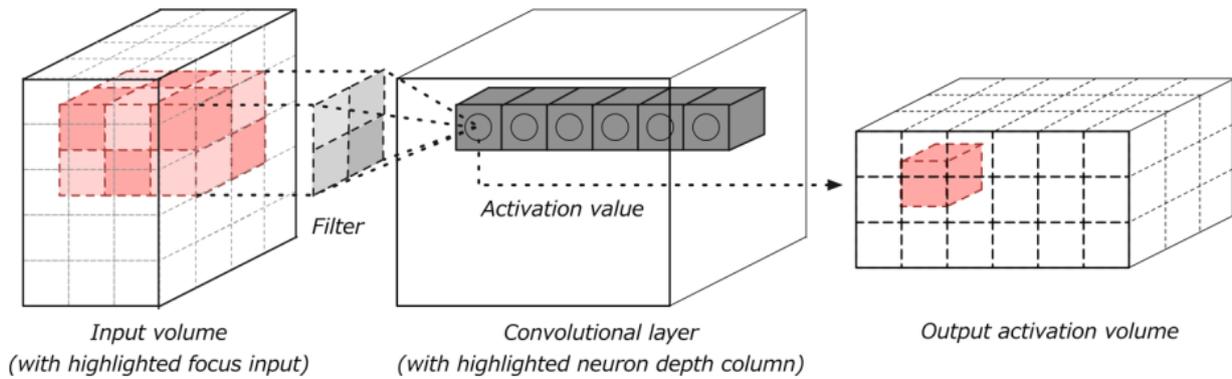


Figure 4. Convolutional layer and Filter operations. Adapted from Patterson & Gibson (2017).

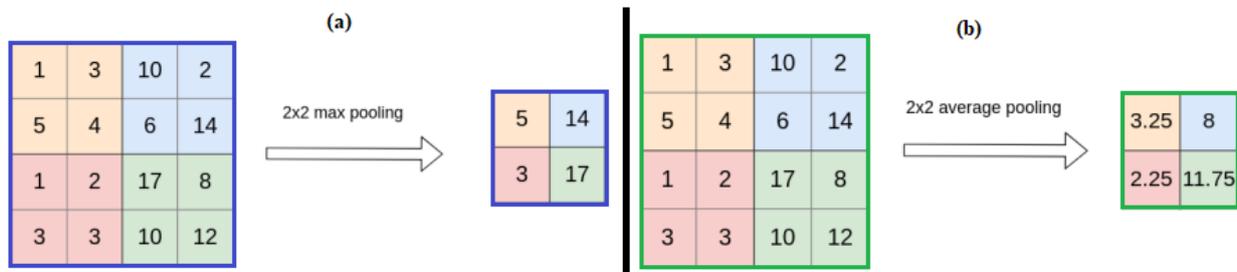


Figure 5. Pooling Operations. Adapted from Vasilev et al. (2019).

A flatten layer collapses the spatial dimensions of the input into the channel dimension (Chollet et al. 2015). For example, if the resultant of the convolutional and pooling layers have the 2x2x2 dimension this layer reduces the dimensionality to a 1D dimension vector (Fausett et al. 1994, Chollet et al. 2015).

Deep Architecture for superNovae Identification - DANI

Convolutional neural networks, as well as the concept of Deep Learning adopted in this work, were chosen instead of the observation made in previous works that indicated the need to adopt a more robust alternative than the fuzzy logic and classical artificial neural networks. This decision derives from how classic methods can extract attributes from raw data and which differs considerably from what is produced by Convolutional Neural Networks.

The Deep Learning method was chosen because the convolution operation made by CNN can be more efficient than classic features (obtained, for example, by descriptor algorithms, clustering algorithms, independent component analysis algorithms, and principal component analysis algorithms) to represent the most important characteristics in multidimensional data (Goodfellow et al. 2016, Keller et al. 2016, Patterson & Gibson 2017). CNN models can extract features from the raw data allowing that small details can be observed, which at some point cannot be perceived by classical methods of feature extraction. The choice of this model is due to the automatic way in which the features are extracted, which makes this model suitable when there are no specialists to directly handle the data, as they occur in autonomous stations in inhospitable places.

Convolutional Neural Networks for 1D data

To find alternatives to the learning of convolutional neural networks that are operated on image data, we explore the operation of convolution over sequential data, that is, over 1D dimensional data. Conventional 1D models are classically used in text and voice recognition, achieving good results described in the literature (Chollet et al. 2015). The choice of this alternative came from the consideration of preserving the original structure of the input data.

Using CNN 1D for sequential data classification or time series classification has relevant performance in the feature extraction process. This type of operation can extract relevant information directly from the raw data of the time series without the need for extensive knowledge about the problem domain. The feature extraction made by CNN 1D can be better than handcrafted features extracted by mathematical models or specialists since peculiarities that may go unnoticed by specialists can be identified by models of CNN 1D neural networks. The Figure 6 shows a simple operation for CNN 1D, these operations are similar to CNN conventional operations.

DANI Modelling

Three learning strategies for supernovae classification were developed, each one having different dimensions for the convolutional layers. The models were developed with convolution layers for 16 points, 32 points and 64 points of the spectra, that is, different point windows. In this way, each model can extract features in different ways. Each model was inserted in a single neural network model, called Multiple Window Convolutional Neural Network, which through matrix operations concatenates the weight matrices generated by the other

models in a single neural network, as shown in Figure 7.

This model is composed of 20 layers, each layer is described in the items below:

1. Input layer with 1000 x 1 dimension, this layer receives all 1D supernovae spectral data;
2. 1st Window: Composed of two 1D convolutional layers with size equal to 16 and filter with size equal to 8. Two pooling layers with filter size equal to 2, doing the operation of Max Pooling. Activation function ReLU (Rectified Linear Unit (Chollet et al. 2015, Patterson & Gibson 2017)). A Flatten layer at the end of the 1st window to reduce the dimensionality of the product from the previous layers.
3. 2nd Window: Composed of two 1D convolutional layers with size equal to 32 and filter with size equal to 12. Two pooling layers with filter size equal to 2, doing the operation of Max Pooling. Activation function ReLU (Rectified Linear Unit (Chollet et al. 2015, Patterson & Gibson 2017)). A Flatten layer at the end of the 2nd window to reduce the dimensionality of the product from the previous layers.
4. 3rd Window: Composed of two 1D convolutional layers with size equal to 64 and filter with size equal to 16. Two pooling layers with filter size equal to 2, doing the operation of Max Pooling. Activation function ReLU (Rectified Linear Unit (Chollet et al. 2015, Patterson & Gibson 2017)). A Flatten layer at the end of the 3rd window to reduce the dimensionality of the product from the previous layers.
5. Concatenation Layer: This layer receives as input a list of tensors (feature Maps obtained from the previous layers), and

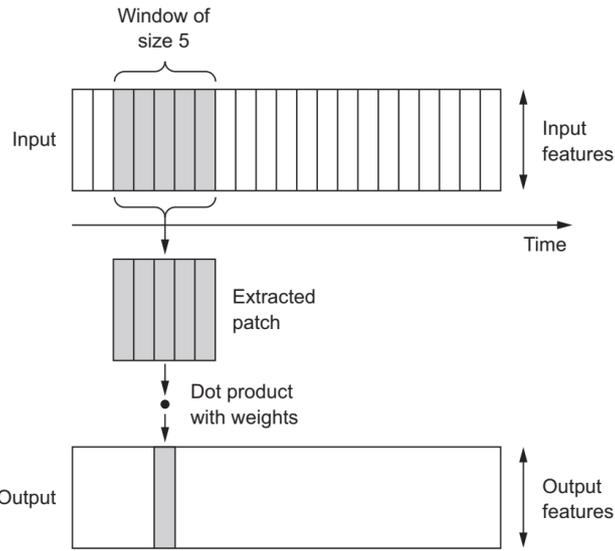


Figure 6. CNN 1D operations for sequential and time series data. Adapted from Chollet et al. (2015).

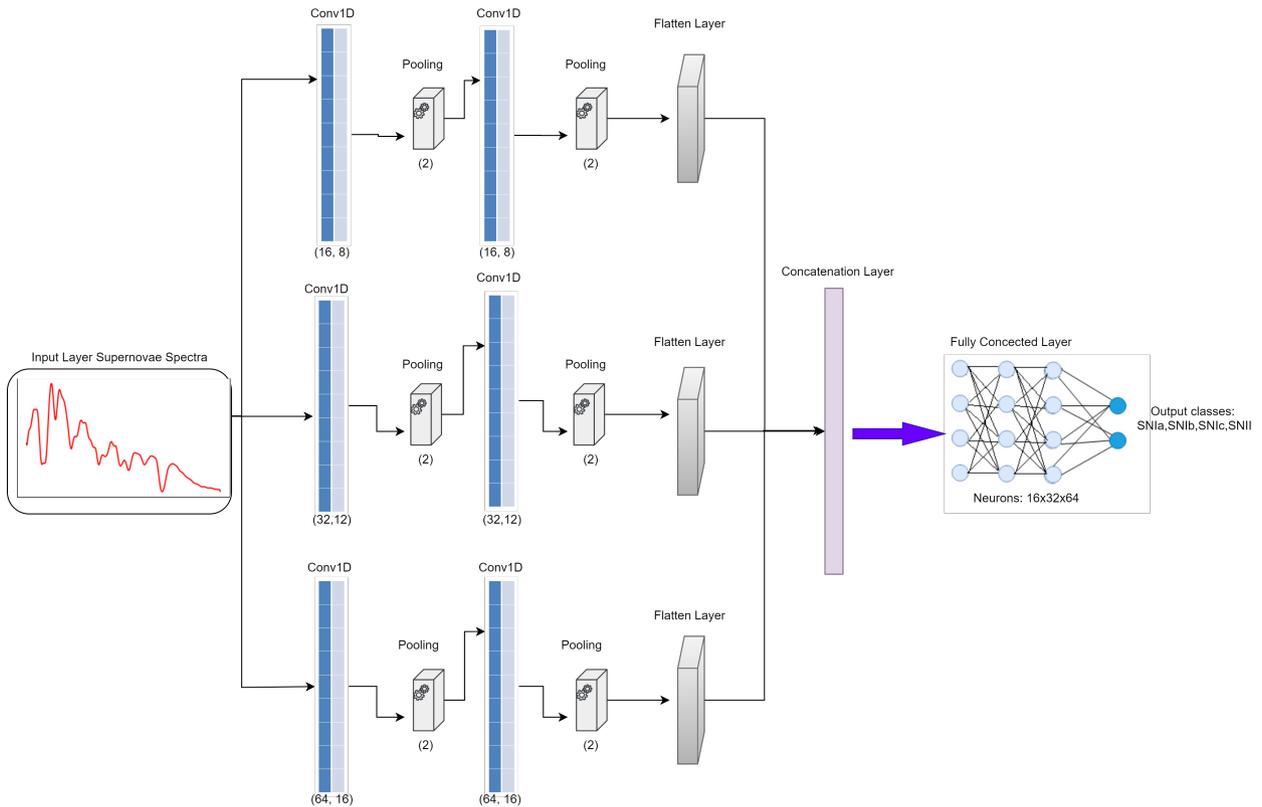


Figure 7. DANI's Multiple Window Convolutional Neural Network to spectral classification.

returns a single tensor (Feature Map). A tensor can be defined as an array of n-dimensional arrays.

6. 17th Layer: Fully connected layer with 16 neurons. Activation function ReLU.
7. 18th Layer: Fully connected layer with 32 neurons. Activation function ReLU.
8. 19th Layer: Fully connected layer with 64 neurons. Activation function ReLU.
9. 20th Layer: Output layer (Fully connected layer) with 4 neurons for multiclass classification. Activation function Softmax (See it in detail in (Chollet et al. 2015, Patterson & Gibson 2017)). The output classes are the classic SN types: SNIa, SNIb, SNIc and SNI.

SUPERNOVAE DATA ANALYSIS AND MMA-SUPERNOVAE APPROACH

To handle the data in the collections (the Online Supernova Spectrum Archive (Richardson et al. 2002), the Open Supernova Catalog (Guillochon et al. 2017), and the CfA Supernova Archive (CfA 2018)) was developed an structure for data organization and easy access. This structure was called **MMA-Supernovae Protocol** (*Multi-Messenger Astrophysics for Supernovae Protocol*).

MMA-Supernovae protocol is attached to the concept of Multi-Messenger Astrophysics, in which the objective is to provide the collection and analysis of astronomical objects and their different sources of data. The analysis of data information from multiple sources, obtained through high-resolution instrumental measures, has become a fundamental task in all scientific areas.

As an objective, this protocol aims to obtain different information about supernovae, collect

their several sources of events, and provide easy access and analysis of these sources. The motivation to create this protocol comes from the experience obtained in analyzing data from the works of CINTIA and SUZAN (do Nascimento et al. 2019, Arantes Filho et al. 2020). The Figure 8 shows the **MMA-Supernovae** process of information extraction.

The **MMA-Supernovae** protocol approach consists of algorithms that handle the JSON data files available in the Open Supernova Catalog. Each supernova available in this catalog has two types of JSON files, one containing information from the supernovae and the other containing files from its different sources. Each of these files was handled by an algorithm written in the Python programming language that accessed the data online and made it available in a dataframe structure (Python tabular data structure).

The protocol structure is designed in levels. At the first level, basic information about each supernova is available, such as the supernova name, the instruments, the supernova type, etc. At the second level, information about each supernova source is available, such as spectral data, light curves, and other sources collected from the SNe explosion, such as the gravitational waves data, neutrinos, etc. The Figure 9 shows the first level.

This structure provided an easy way to treat the supernovae data used in DANI system, and in this way, it was possible to access and handle data from different collections in a single structure. This protocol also supports the analysis of supernovae data on the concept of Multi-Messenger Astrophysics, allowing the use of several data sources generated by this event. In this work we focused essentially on spectral data analysis, however, the analysis of multiple sources is one of the points that we intend to develop.

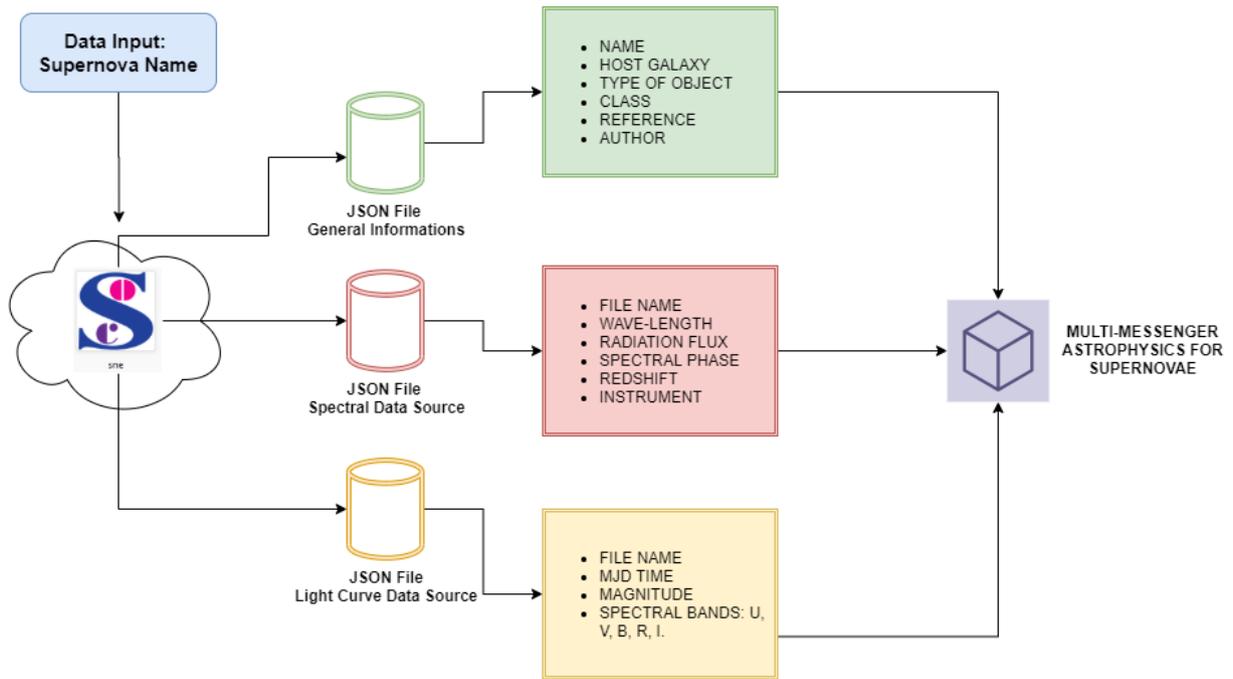


Figure 8. MMA-Supernovae Protocol.

df_MMA - DataFrame

Index	SN_NAME	HOST_GALAXY	OBJECT_TYPE	CATALOG_NAME	AUTHOR_REFERENCE	SPECTRA	LIGHT_CURVE
532	SN2008bo	NGC 6643	IIb	The Cfa Supernova Ar...	Modjaz M., Blondin S., Kirshner R.P., Mathe...	FILE_NAME ...	OBJECT_NAME TIME_MJD_BAN...
533	SN2008cw	A163238+4127	IIb	The Cfa Supernova Ar...	Modjaz M., Blondin S., Kirshner R.P., Mathe...	FILE_NAME ...	OBJECT_NAME TIME_MJD_BAN...
534	SN2008D	NGC 2770	Ib	The Cfa Supernova Ar...	Modjaz M., Blondin S., Kirshner R.P., Mathe...	FILE_NAME ...	OBJECT_NAME ...
535	SN2009er	A153929+2426	Ib Pec	The Cfa Supernova Ar...	Modjaz M., Blondin S., Kirshner R.P., Mathe...	FILE_NAME ...	OBJECT_NAME TIME_MJD_BAN...
536	SN2009iz	UGC 2175	Ib	The Cfa Supernova Ar...	Modjaz M., Blondin S., Kirshner R.P., Mathe...	FILE_NAME ...	OBJECT_NAME TIME_MJD_BAN...
537	SN2009jf	NGC 7479	Ib	The Cfa Supernova Ar...	Modjaz M., Blondin S., Kirshner R.P., Mathe...	FILE_NAME ...	OBJECT_NAME ...
538	SN2009ip	NGC 7259	IIIn	sne	Howerton	FILE...	OBJECT_NAME ...
539	SN2012fr	NGC 1365	Ia	sne	Klotz et al.	FILE...	OBJECT_NAME ...
540	SN2011fe	M101	Ia	sne	Nugent et al.	FILE_N...	OBJECT_NAME ...
541	SN2014J	NGC 3034	Ia	sne	Fossey et al.	FILE_NA...	OBJECT_NAME ...
542	SN2013ej	NGC 628	II-P/L	sne	LOSS	FILE...	OBJECT_NAME ...
543	SN1999em	NGC 1637	II P	sne	LOSS	FILE_N...	OBJECT_NAME ...
544	PTF12gzk	SDSS J221241.53+0...	Ic Pec	sne	PTF	FILE_N...	OBJECT_NAME ...
545	SN2011dh	NGC 5194	IIb	sne	Reiland, Griga, Riou, Lamotte Bailey	FILE_N...	OBJECT_NAME ...
546	SN1998S	NGC 3877	IIIn	sne	BAOSS	FILE_NA...	OBJECT_NAME ...

Formato Redimensionar Cor de fundo Min/max de coluna Salvar e Fechar Fechar

Figure 9. MMA-Supernova Protocol first level.

Data Normalization

The data preparation for feeding DANI's 1D convolution model followed the steps:

1. **Redshift Adjust (z):** This adjustment was done to evaluate the supernovae spectral lines as if the objects were at rest. The redshift adjustment is done to approximate the supernovae spectral lines to the spectral lines of chemical elements that are obtained from the laboratory. The redshift (z) values for each supernova are in the collections and was inserted in the MMA-Supernovae structure. The redshift was done as the Equation 2.

$$\lambda_0 = \frac{\lambda}{z + 1} \quad (2)$$

where:

- λ_0 : wavelength of the object at rest;
- λ : observed wavelength;
- z : redshift.

2. **SNe flux values normalization.** This normalization was done to put all SNe flux values (y values) in the range of 0 to 1, as explained by the Equation 3.

$$\text{Normalization}(y) = \frac{y - \min(y)}{\max(y) - \min(y)} \quad (3)$$

Where $\min(y)$ is the minimum SNe flux value and $\max(y)$ is the maximum SNe flux value.

3. **Linear Interpolation** of the spectra in 1000 points. We did this process to put all spectra with the same size, this way all the database spectra are adjusted with the same number of points;
4. **Application of SDF-SG** this process can be seen with major details in (Arantes Filho et al. 2019). This step consists of two

successive filterings by the Savitzky-Golay filter with window size equal to 71 and a polynomial degree equal to 9.

This data normalization was similar to the same steps made in CINTIA and SUZAN systems, however, we do not define a range for the wavelength of the spectra in order to observe the characteristics of the supernovae in the region of the Infrared spectrum. CINTIA and SUZAN used spectra delimited in the range of wavelengths from 4000 to 7000 angstroms. Each spectrum trained by the DANI system has 1000 points, the trained values correspond only to the supernova radiation flux values (y).

RESULTS

The selected data from the catalogs indicated 26423 instances of different spectra and light curves. These instances are associated with 5197 different supernovae. Some important information collected from the MMA-Supernovae are listed below:

1. Number of different SNe types : 73;
2. Number of Instruments (telescopes and spectrographs) : 215;
3. Number of different host galaxies: 3273;

As a preliminary evaluation step, we selected only the classic types of supernovae in the database. It's important to mention that we found some types of supernovae that have been identified with only one instance and several others defined as peculiar. Thus, in the DANI model, we are restricted to classic types only. The Figure 10 shows the SNe type distribution in the database.

The spectra distribution of the classic types of supernovae (SNIa, SNIb, SNIc, SNIId) used in DANI is illustrated in the graph of the Figure 11. All SNe spectra of the classical types were

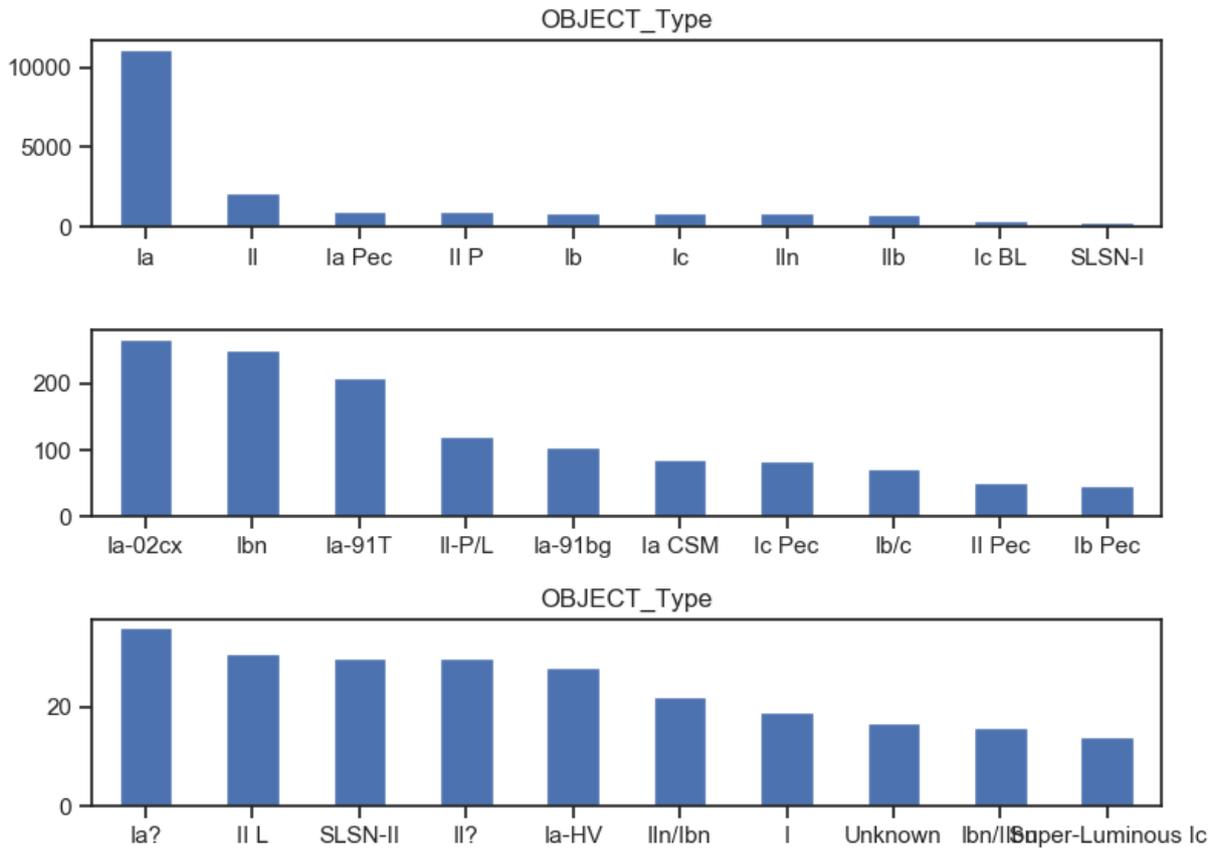


Figure 10. SNe type distribution.

used for the training and validation of the model. From 26423 initial instances analyzed, we selected about 15134 spectra of the classic types of SNe, 11142 of the selection correspond to SNIa, and 3992 correspond to core-collapse SNe (SNIb, SNIc, and SNIId), thus resulting in binary classification.

The training of the Multiple Window model consumed 80% of the available data and covered a total of 100 training epochs. The validation of this model, that is, the model’s ability to classify data that has never been seen before, is done on the other 20% of the total data sample. The Table I shows the performance of the DANI system for new data samples and its performance for the classification of the classic types of SNe. Table I indicates the results of Precision, Recall and

F1-Score, which in turn can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal.

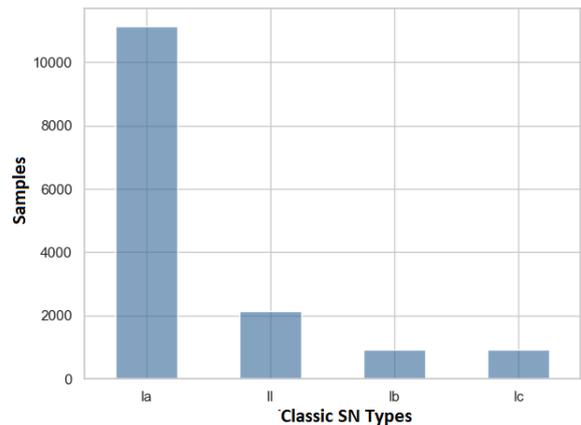


Figure 11. Classical SNe Types distribution.

Table I. Results for the validation of the DANI model for the classification of the classic types of SNe.

SNe Type	Total SNe Spectra	Precision	Recall	F1-Score
Ia	2263	0.93	0.96	0.94
Ib	174	0.58	0.64	0.61
Ic	176	0.61	0.57	0.59
II	414	0.88	0.69	0.77

To compare the approaches made in previous works by SUZAN and CINTIA, the capacity of the DANI system to classify SNe was highlighted, covering the entire available dataset. The comparison made for these systems is important to check the ability to separate the SNIa data from other SNe types. The Table II shows the results for the entire dataset and the performance comparisons with the other systems.

We indicate that the discussed results are from an initial version of the DANI system and that it needs fine adjustments concerning the choice of architectures and training parameters. This need is reflected in the number of training epochs, as it would be impractical to close a model for training in many training epochs without fine-tuning parameters. However, the results of this model stand out, as they surpass the predecessor systems in performance both in the number of evaluated spectra and in the accuracy score to separate SNIa from other types. Table III shows the obtained results in the classification made for all classical SNe types in the dataset.

CONCLUSIONS

The machine learning solutions discussed in this work, essentially DANI, present important contributions in the treatment, identification, and classification of supernovae data. This computational structure shows potential

applicability in instruments and autonomous systems such as KDUST, which requires automatic and precise classification methods.

The spectral patterns learned by DANI combine characteristics present in sequences of 16, 32, and 64 points of the supernova spectra. In this way, it was possible to identify the spectral line patterns (emission and absorption) in different wavelengths.

The DANI performance in the classification of SNIa and SNII are relevant, reaching a score of 97% for correct classification on 13.299 SNe spectra. This achievement indicates that the CINTIA, SUZAN, and DANI can precisely distinguish the SNIa from other types. The classification performance for SNIb and SNIc supernovae has a deficiency when comparing the validation data. At this point, DANI's precision performance is around 60%, indicating the need to improve the training criteria for more accurate classification of these supernovae.

An important appointment is that the DANI, CINTIA, and SUZAN analyses supernovae data from several instruments (telescopes and spectrographs), that is, can influence the learning process of the spectral patterns, essentially for the SNIb and SNIc supernovae, since each instrument carries its peculiarities, such as different calibrations or scales. Considering the SNIb and SNIc explosion mechanisms, the conditions for the

Table II. Benchmark for Supernovae Intelligent Classification Systems.

SNe Classification Systems	Total SNIa Spectra	Total Non-SNIa Spectra	Precision	Recall	Accuracy	F1-Score
DANI	11142	3992	0.92	0.90	0.96	0.91
CINTIA	6644	2512	0.97	0.76	0.95	0.85
SUZAN	3082	615	0.93	0.72	0.73	0.83

Table III. Full Data Results for SNe classified by DANI.

SNe Type	Total SNe Spectra	Precision	Recall	F1-Score
Ia	11142	0.97	0.99	0.98
Ib	918	0.86	0.88	0.87
Ic	917	0.88	0.87	0.88
II	2157	0.97	0.86	0.91

core-collapse are complex, and this complexity can influence the data.

Another appointment is the quality of the labeled data in the catalog. Inconsistencies and exchanged labels were pointed out by Pruzhinskaya et al. (2019) in their studies related to the OpenSN catalog, in which 33% of objects are considered peculiar, and 1.4% shows anomalies, such as wrong classifications.

Finally, we highlight that the solutions presented in this work aimed to provide automatic and precise classification of supernovae, ensuring accuracy and good performance for data from several instruments. The initial results obtained also indicate a path to be explored concerning intelligent classifiers that can act as autonomous systems for survey in remote telescopes such as KDUST.

Acknowledgments

We thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We also thank INPE and IEAv for supporting research and provide the computational and scientific structure.

REFERENCES

- ARANTES FILHO LR, GUIMARÃES LNF, DO NASCIMENTO FJB & ROSA RR. 2019. Estratégia de dupla filtragem utilizando o filtro de Savitzky-Golay em dados espectrais de supernovas. *Rev Bras Comp Apl* 11(2): 86-99.
- ARANTES FILHO LR, GUIMARÃES LNF, ROSA RR & DO NASCIMENTO FJB. 2020. SUZAN: Modelo Fuzzy para Classificação Espectral de Dados de Supernovas Ia. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* 7(1).
- ASTIER P ET AL. 2006. The Supernova Legacy Survey: measurement of, and w from the first year data set. *Astronomy & Astrophysics* 447(1): 31-48.
- BLONDIN S, MANDEL KS & KIRSHNER RP. 2011. Do spectra improve distance measurements of Type Ia supernovae? *Astronomy & Astrophysics* 526: A81.
- BLONDIN S ET AL. 2012. The spectroscopic diversity of type Ia supernovae. *Astron J* 143(5): 126.
- BRENNA F. 2005. The dark energy survey. *Int J Mod Phys A* 20(14): 3121-3123.
- BRUNEL A, PASQUET J, PASQUET J, RODRIGUEZ N, COMBY F, FOCHEZ D & CHAUMONT M. 2019. A CNN adapted to time series for the classification of Supernovae. *Electronic Imaging* 2019(14): 90-91.
- BURKOV A. 2019. The Hundred-Page Machine Learning Book. "On-line".

- BURTON MG ET AL. 2016. Scientific goals of the kunlun infrared sky survey (kiss). *Publications of the Astronomical Society of Australia* 33.
- CFA. 2018. The CfA Supernova Archive. 60 Garden Street, Cambridge. <https://www.cfa.harvard.edu/supernova/SNarchive.html>. Accessed 19 June 2018.
- CHOLLET F ET AL. 2015. Keras: Deep learning library for theano and tensorflow. URL: [https://keras.io/k7\(8\):T1](https://keras.io/k7(8):T1).
- COELHO RC, CALVÃO MO, REIS RR & SIFFERT BB. 2014. Standardization of type Ia supernovae. *Eur J Phys* 36(1): 015007.
- DO NASCIMENTO FJB, ARANTES FILHO LR & GUIMARÃES LNF. 2019. Intelligent Classification of Supernovae Using Artificial Neural Networks. *Inteligencia Artificial* 22(63): 39-60.
- FAUSETT LV ET AL. 1994. *Fundamentals of neural networks: architectures, algorithms, and applications*. Volume 3. prentice-Hall Englewood Cliffs.
- FILIPPENKO AV. 1997. Optical spectra of supernovae. *Annual Review of Astronomy and Astrophysics* 35(1): 309-355.
- GOODFELLOW I, BENGIO Y & COURVILLE A. 2016. Deep learning. MIT press.
- GRAHAM M, BELLM E, GUY L, SLATER C & DUBOIS-FELSMANN G. 2019. LSST Alerts: Key Numbers.
- GUILLOCHON J, PARRENT J, KELLEY LZ & MARGUTTI R. 2017. An open catalog for supernova data. *Astrophys J* 835(1): 64.
- HAYKIN S. 2001. *Redes Neurais Artificiais: Princípios e Práticas*. 2ª edição. Editora Bookman, Porto Alegre, 900p.
- HORVATH JE. 2011. *Fundamentos da Evolução Estelar, Supernovas e Objetos Compactos*. Editora Livraria da Física, 1ª ed.
- HUBER S ET AL. 2019. Strongly lensed SNe Ia in the era of LSST: observing cadence for lens discoveries and time-delay measurements. arXiv preprint arXiv:190300510.
- KELLER JM, FOGEL DB & LIU D. 2016. *Fundamentals of computational intelligence: neural networks, fuzzy systems, and evolutionary computation*. John Wiley & Sons.
- KIM Y. 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:14085882.
- KIMURA A, TAKAHASHI I, TANAKA M, YASUDA N, UEDA N & YOSHIDA N. 2017. Single-epoch supernova classification with deep convolutional neural networks. In: 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), p. 354-359.
- LECUN Y ET AL. 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361(10): 1995.
- LECUN Y ET AL. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11): 2278-2324.
- LI X, YUAN X, GU B, YANG S, LI Z & DU F. 2019. Chinese Antarctic Astronomical Optical Telescopes. In: *Revista Mexicana de Astronomia y Astrofisica Conference Series*. Volume 51, p. 135.
- MANASWI NK, MANASWI NK & JOHN S. 2018. *Deep Learning with Applications Using Python*. Springer.
- MARKEL J & BAYLESS AJ. 2019. Performance of Random Forest Machine Learning Algorithms in Binary Supernovae Classification. arXiv preprint arXiv:190700088.
- MÓDOLO M, GUIMARÃES LNF & ROSA RR. 2015. An expert supernova spectral classification using artificial neural networks. *J Comp Int* 6(2): 81-90.
- MODJAZ M ET AL. 2014. Optical spectra of 73 stripped-envelope core-collapse supernovae. *Astron J* 147(5): 99.
- MOHRI M, ROSTAMIZADEH A & TALWALKAR A. 2018. *Foundations of machine learning*. MIT press.
- MUTHUKRISHNA D, PARKINSON D & TUCKER BE. 2019. DASH: Deep Learning for the Automated Spectral Classification of Supernovae and Their Hosts. *Astrophys J* 885(1): 85.
- PATTERSON J & GIBSON A. 2017. *Deep learning: A practitioner's approach*. O'Reilly Media, Inc.
- PERLMUTTER S ET AL. 1999. Measurements of Ω and Λ from 42 high-redshift supernovae. *Astrophys J* 517(2): 565.
- PRUZHINSKAYA MV, MALANCHEV KL, KORNILOV MV, ISHIDA EE, MONDON F, VOLNOVA AA & KOROLEV VS. 2019. Anomaly Detection in the Open Supernova Catalog. arXiv preprint arXiv:190511516.
- RICHARDSON D, THOMAS R, CASEBEER D, BRANCH D & BARON E. 2002. SUSPECT, The Online Supernova Spectrum Archive: Year Two. *AAS* 201: 56-09.
- RIESS AG ET AL. 1998. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astron J* 116(3): 1009.
- SANTOS MVD, QUARTIN M & REIS RR. 2020. On the cosmological performance of photometrically classified supernovae with machine learning. *Mon Not R Astron Soc* 497(3): 2974-2991.
- SAVITZKY A & GOLAY MJE. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analyt Chem* 36(8): 1627-1639.
- TURATTO M, BENETTI S & PASTORELLO A. 2007. Supernova classes and subclasses. *AIP Conference Proceedings* 937(1): 187-197.
- VASILEV I ET AL. 2019. *Python Deep Learning Exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow*. Packt Publishing Ltd.

WOOD-VASEY WM ET AL. 2007. Observational constraints on the nature of dark energy: first cosmological results from the essence supernova survey. *Astrophys J* 666(2): 694.

YUAN X, CUI X, SU DQ, ZHU Y, WANG L, GU B, GONG X & LI X. 2012. Preliminary design of the kunlun dark universe survey telescope (kdust). *Proc Int Astron Union* 8(S288): 271-274.

How to cite

ARANTES FILHO LR, ROSA RR & GUIMARÃES LNF. 2021. Intelligent Supernovae Classification Systems in the KDUST context. *An Acad Bras Cienc* 93: e20200862. DOI 10.1590/0001-3765202120200862.

LUÍS R. ARANTES FILHO¹

<https://orcid.org/0000-0002-8470-8377>

REINALDO R. ROSA²

<https://orcid.org/0000-0002-2962-4322>

LAMARTINE N.F. GUIMARÃES³

<https://orcid.org/0000-0002-0302-9162>

¹National Institute for Space Research (INPE), Applied Computing Graduate Program (CAP), Av. dos Astronautas, 1758, Jd. da Granja, 12227-010 São José dos Campos, SP, Brazil

²National Institute for Space Research (INPE), Lab for Computing and Applied Mathematics (LABAC), Av. dos Astronautas, 1758, Jd. da Granja, 12227-010 São José dos Campos, SP, Brazil

³Department of Aerospace Science and Technology (DCTA), Trevo Cel. Av. J. A. A. do Amarante, Putim, 12228-001 São José dos Campos, SP, Brazil

Correspondence to: **Luís Ricardo Arantes Filho**

E-mail: luisricardoengcomp@gmail.com

Authors contributions

All authors listed have made an important, precise, and intellectual contribution to the work.

Manuscript received on June 3, 2020; accepted for publication on September 29, 2020

