CARRERA: ESPECIALIZACIÓN EN CIENCIA DE DATOS

TALLER: TRABAJO FINAL INTEGRADOR

"Comparativa de rendimiento de clasificacion de espectrogramas de ondas gravitacionales mediante la utilizacion de tecnicas de machine learning"

Nombre y Apellido del Alumno/a: Msc. Ezequiel H. Martinez

Título de grado o posgrado (último): MBA

Profesores:

Dra. Alicia Mon

Lugar y Fecha: Buenos Aires, 04 de Junio de 2020

Table of contents

1. Introduction

The Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) has opened the field of gravitational wave astronomy through the direct detection of signals predicted by Einstein's General Theory of Relativity. Advanced LIGO's first observing run (O1) saw the first detections of binary black hole mergers. Advanced LIGO and Virgo's second observing run (O2) included both binary black hole and binary neutron star mergers. Data around these discoveries are publicly available, along with associated software libraries. This work is based in a paper published over the classification of "glitches" that affects those waves' readings and the project developed around it: Gravity Spy. This project was designed for scientists and students pursuing research in this field, both inside and outside the LIGO Scientific Collaboration. It offers a dataset of pre-classified spectrograms over these waves. They were classified by a public science effort made by crowd classification[1]. This dataset gave support for the paper "Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science"[2], and its dataset will be the base for this work. This paper identifies a particular kind of "glitch": the "chirp", as a gravitational wave. These spectrograms are representations of real readings, being some of them gravitational waves. This paper asses the performance of different "Machines Learning" techniques in order to classify these spectrograms as "gravitational waves". We will then train a model that will classify this dataset trying to approximate it at the results that the author of the paper achieved.

2. Conceptual Framework

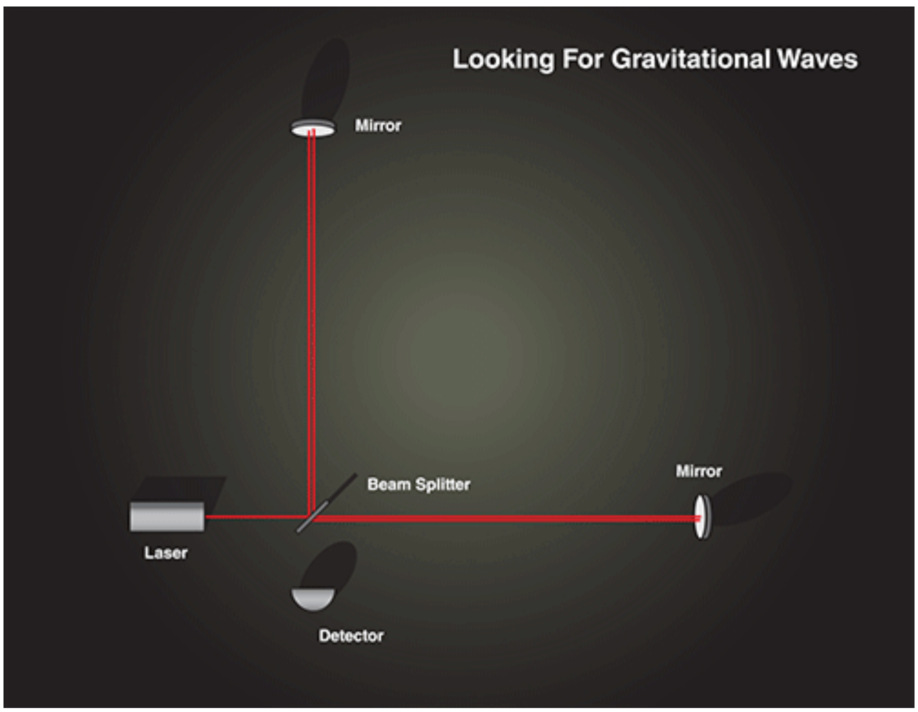
Gravitational waves are signals emited by objects of high mass that resides in deep space. Usually, they can be emitted in the occurence of a celestial event. They constitute disturbances in the space-time fabric. These object are typically massive and of the size of a star. These signals occurs in moment in which there is activity within these, which could be:

- A Supernova explosion
- A merging of two binary stars.
- A merge of two binary black holes.

These signals are catched by an specific type of observatory named LIGO[3]. The first succesful emitions taken were during 2015[4]. These measurements are organized in blocks of data that are being structured in different datasets. Some of these datasets are freely available and are related to an specific time. Despite the fact that not all the data is openly available, we count today within the scope of these interferometers, around four years of measurements[5] organized in serveral rounds of observations.

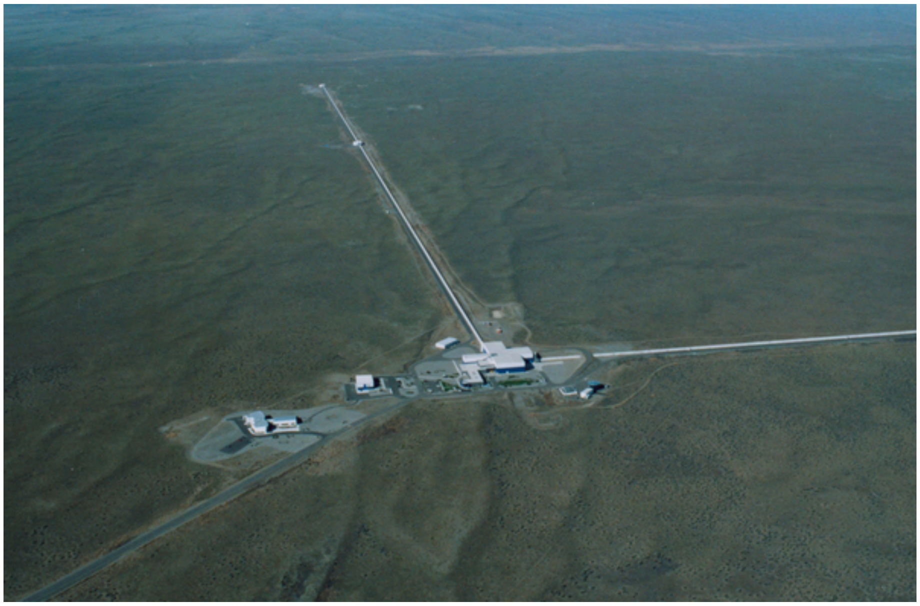
These interferometers[6] work inside the LIGO observatories. An "interferometer", in essence, is the sensor in charge of taking these signals from space. It is analogous to an telescope for electromagnetic signals. As the name says, an interferometer is a sensor designed to identify interferences. There are of several sizes, but the ones you can find in Livignston and Handford observatories are particularly big; their intent is to catch the specific interferences produced by a gravitational wave.

Figure 2.1 - The general internal organization of a LIGO interferometer. [7]



As we can see in figure 2.1, they are build with a single laser beam that is being divided between two rays. The variations in lenght and frequency of any of those two arms in respects to the other produces the detection of a possible signal.

Figure 2.2 - LIGO is made up of two observatories: one in Louisiana and one in Washington. Each observatory has two long “arms” that are each more than 2 miles.[8]



Since an interferometer catches interferences, they are highly sensitive to all sort of white noise. There are techniquer and challenges associated with there measurements, specifically with the filtering and clearence of the signal[9] The signals grabed by the sensors are digitalized for its later analysis. They are structured in databrust that are measured by second[10].

Some of the historical data is available. The current dataset isn't open to the public. What is available is a public subset that is avaiable and refers to past observation run. These are data already classified and treated. In this sense, we will be able to work with data that is already classified and tagged by other scientist.

Currently, the data is obtained from four different observatories:

- `'G1'` - GEO600
- `'H1'` - LIGO-Hanford
- `'L1'` - LIGO-Livingston
- `'V1'` - (Advanced) Virgo

These sensors create datasets from the same events, measured by different latitudes and longitudes in the earth. These varied measurements help to apply techniques that could help in the identification of miss readings or false positives in the identification of waves. There are some specific signals that looks very closely to what a gravitaional wave might look like, they are referred as glitches.

The glitches can be defined as peculiarities in the signal that simulate o are similar to what a gravitational wave might look like. [11] These gitches in the readings are false positives in the signal reading and supposes one of the artifacts most dreaded in the study of these events.

These measurements are standarized and are resolved in standard datesets that can be consumed by a Python. There are in place several efforts to work in the measurements and clean the signal out of noise[12]. We understand that there are a variation of methods to test and apply for this to be accomplished; we are going to be centering our efforts to make use of a technique denominated Deep Filtering[13] which is making usage of convolutional networks and other machine learning applications.

3. Problem definition

There was a challenge held for the LIGO open data worshop of 2020[14], in which with a limited and open held data they tried to classify several binary black holes generated by the scientists that held the workshop. The aim was to allow students and senior scientists to learn about the datasets structure and software tools[15]. In the same workshop, during 2019, the results for the four challenges were mixed, being the fourt challenge the only one that didn't received any points[16]. The problem, in essence, was to develop a method to identify as many signals as possible within the given dataset. The difficulty to classify these waves is high, given that the line and the readings come loaded of noise. Therefore, developing a technique that could classify gravitational waves could be of great value for the average astrophycisist.

We got a dataset of pre-classified spectrograms that we might use as a base study to show and reproduce how a project like Gravity Spy could have been developed. Since the dataset for the same paper is publicitly available; to reproduce these findings between different techniques could end being of great value for the people involved in the classification of these signals.

4. Study justification

Application of some of the ideas of machine learning and deep filtering[17] to the identification of the gravitational signals within the scope of the given dataset and evaluate its performance could be of great use for the further development of the field. We understand that this could give us a better understanding of how different techniques could behave in relation with such a complex and noisy signal. If we reach the correct identification of a subset of the events, we'll be placed a first steps into the classification of such signals. Such applications could hit onto the following fields:

- Filtering of noise signals obtained from other deep space sources.
- Identification of patterns in sensors of other sorts in noisy environments.

- Voice filtering in noisy environments.
- Surveillance and the development of better sensors in the building of satellites.
- Identification of different human verbal languages and their translations.

These signals provide us a challenge to identify them. Developing or reproducing techniques on such a dataset could help us to apply the same models to different problems where the signals are dim or subceptible to heavy noise, but where we know "what we are looking for".

However, we are not going to classify all the universe of these signals available. We will suscribe our research to the dataset freely available and try to classify the spectrogram of these signals. The value is in the comparison with the results obtained in the Gravity Spy's paper and see if there are differences in the classification performance of two or more constrasted techniques.

5. Study scope

Working with the dataset trainingsetv1d1.h5, available from the site https://zenodo.org/record/1476551#.Xu6TaZNKjNw, we will apply techniques of deep learning in order to identify at least one of the "Chrips" classified in the "validation" set of the 0.5 seconds observation range. This dataset is a real observation obtained during the second observation run denominated O2[18]. The scope of this work is to try to train a support vector machine or a recurrent neural network capable of identify at least one of these events. The dataset is fully pre-classified and ready for its division in training, validation and test sets.

6.Hypothesis

The conjecture is that given the pre-classified dataset, we will be able to train at least two machine learning algorithms capable of classifying at least one of the gravitational events introduced (a "Chirp"). For this to happen, we will be able to use the pyspark[19], pandas[20], keras[21] and numpy[22] libraries provided by the body of tools found in the Python's universe. The result of this work will be one or more trained models; these will be capable of classifying a spectrogram of a signal into as a gravitational wave with a certain degree of accuracy. The aim of this work is to compare the performance of these models between them and then check it against the benchmark set by the Gravity Spy's paper.

7. Objectives

Use at least two machine learning models capable of identifying one or more gravitational signals found in the spectrograms available in the given dataset for the range of 0.5 seconds. These are supposed to be preclassified signal for the adjusted dataset provided for the Livingston and Hanford interferometers during the O2 series of observations. For each model trained, we will list:

- The spectrogram IDs found as Chrips.
- The ROC percentage for the model.

All of this will be achieved with techniques related to convolutional neural networks, SVMs and others.

This objective will be restricted to the data file "trainingsetv1d1.h5" openly found[23] on the Zenodo repositories. The timespan we will operate with will be 0.5 seconds. These are real adjusted LIGO data from O2 observation run.

Sub-goals:

○ Download the datasets trainingsetv1d1.h5; place them into a spark architecture for its later treatment.

○ Train a "Chrip signal classificator" throught a trained model, like a linear SVM.

○ Train a "Chrip signal classificator" throught a trained model, like a Recurrent Neural Network or a GAN.

○ Compare the ROC curves; use other methods to assest their performance.

○ Compare it with the classification done by the Zooniverse effort for Gravity Spy and draw conclussions over the classifiers performance over the dataset.

8. Methodology

The main methodology used for this work involves different machine learning and deep learning algorithms comparatives. We are going to use the following algorithms and measure its performance in a binary classification of spectrograms of half seconds for signals taken from deep space:

- Convolutional Neural Networks
- Linear SVC
- Recurrent Neural Networks
- Generative Adversarial Networks

We are going to work in training and testing the dataset in the methods suggested by François Chollet[24], which entails the training of different deep learning algorithms by dividing the dataset in three layers:

- Training set: 5587 images. (41 gravitational waves)

- Testing set: 1179 images (10 gravitational waves)
- Validation set: 1200 images (9 gravitational waves)

(We are going to use the division sets used in the "Gravity Spy"[25] paper, since the same comes already divided and tagged in these three subsets.)

After this step is taken, we are going to configure the different algorithms, train them, save the model and performance during metrics for then on measure their performance in accordance with the book's metrics and suggestions over the test and validation sets.

The general metrics we have in mind to evaluate performance are the following:

- To verify that the trained model classifies a "Chirp" class (as the minimum).
- Usage of similar "Loss" and "Accuracy" functions in order to estimate how "good" the trained model is. (they are a percentage and therefore could be compared)

8.2 Convolutional Neural Network Approach

We applied here the given sub-set of training pre-classified samples; they were 5587 images, everyone constituent of these O2 measurements. We configured the network with different layers; which were mere adaptations of an exercise from the Deep Learning book[26] example for classification of digits of the MNIST classic dataset:

1. Conv2D(32, (3,3), activate='relu', input_shape=(140,170,1))
2. MaxPooling((2,2))
3. Conv2D(64, (3,3), activate='relu')
4. MaxPooling((2,2))
5. Conv2D(64, (3,3), activate='relu')
6. Flatten()
7. Dense(64, activate='relu')
8. Dense(2, activate='softmax')

We configured the optimisers and loss functions as "RMS" and "Binary Cross Entropy" and trained over the same dataset over five epochs.

This network failed to classify the images; in spite the fact that the "accuracy" metric stayed over 0.992 with a "loss" below 0.007. Not going deep into the analysis of "why" this were the case the best conjecture were that the metrics remained sound because the network was good at classifying the bast amount of images that “weren’t” the ones. The small amount of error remained still higher than the percentage of “gravitational waves” spectrograms; which were below the 0.008 of the total. Even training over 50 epochs weren’t enough for the network to identify one of the images in the validation or testing sets.

We found that the network “deepness” was the problem. With the given positives’ set (41 in 5587), we needed to build up a bigger network. We changed the model for a similar one found in the same book[27] for classification of a small set of images of “dogs” vs “cats” (The dataset was small, roughly 2000 images, but the categories remained binary as well). We adapted such a model in the following way:

1. Conv2D(32,(3,3), activation="relu", input_shape=(140,170,1))
2. MaxPooling2D(2,2)
3. Conv2D(64,(3,3), activation="relu")
4. MaxPooling2D(2,2)
5. Conv2D(128,(3,3), activation="relu")
6. MaxPooling2D(2,2)
7. Conv2D(128,(3,3), activation="relu")
8. MaxPooling2D(2,2)
9. Flatten()
10. Dense(512, activation='relu')
11. Dense(1, activation=‘sigmoid’)

We added three more layers; and changed the activation of the last layer to “sigmoid”. This network showed the following performance during training:

Figure 8.2.1 - Generated during the training session for the Convolution Neural Network approach.

```
Train on 5587 samples, validate on 1200 samples
Epoch 1/30
5587/5587 [==============================] - 384s 69ms/step - loss: 0.0864 - binary_accuracy: 0.9755 - val_loss: 0.0488 - val_binary_accuracy: 0.9925
Epoch 2/30
5587/5587 [==============================] - 384s 69ms/step - loss: 0.0369 - binary_accuracy: 0.9927 - val_loss: 0.0212 - val_binary_accuracy: 0.9925
Epoch 3/30
5587/5587 [==============================] - 385s 69ms/step - loss: 0.0101 - binary_accuracy: 0.9970 - val_loss: 0.0014 - val_binary_accuracy: 1.0000
Epoch 4/30
5587/5587 [==============================] - 384s 69ms/step - loss: 0.0058 - binary_accuracy: 0.9980 - val_loss: 0.0011 - val_binary_accuracy: 1.0000
Epoch 5/30
5587/5587 [==============================] - 382s 68ms/step - loss: 0.0025 - binary_accuracy: 0.9989 - val_loss: 1.2844e-04 - val_binary_accuracy: 1.0000
Epoch 6/30
5587/5587 [==============================] - 383s 68ms/step - loss: 0.0045 - binary_accuracy: 0.9986 - val_loss: 1.0823e-04 - val_binary_accuracy: 1.0000
Epoch 7/30
5587/5587 [==============================] - 383s 69ms/step - loss: 4.6107e-04 - binary_accuracy: 0.9998 - val_loss: 0.1149 - val_binary_accuracy: 0.9425
Epoch 8/30
5587/5587 [==============================] - 383s 69ms/step - loss: 0.0020 - binary_accuracy: 0.9991 - val_loss: 7.0463e-05 - val_binary_accuracy: 1.0000
Epoch 9/30
5587/5587 [==============================] - 384s 69ms/step - loss: 2.4350e-04 - binary_accuracy: 1.0000 - val_loss: 6.9949e-06 - val_binary_accuracy: 1.0000
Epoch 10/30
5587/5587 [==============================] - 381s 68ms/step - loss: 8.0051e-04 - binary_accuracy: 0.9995 - val_loss: 6.9947e-06 - val_binary_accuracy: 1.0000
Epoch 11/30
5587/5587 [==============================] - 383s 69ms/step - loss: 1.5719e-04 - binary_accuracy: 1.0000 - val_loss: 6.8409e-06 - val_binary_accuracy: 1.0000
Epoch 12/30
5587/5587 [==============================] - 381s 68ms/step - loss: 1.8200e-05 - binary_accuracy: 1.0000 - val_loss: 8.6281e-07 - val_binary_accuracy: 1.0000
Epoch 13/30
5587/5587 [==============================] - 382s 68ms/step - loss: 5.4967e-04 - binary_accuracy: 0.9996 - val_loss: 8.9405e-07 - val_binary_accuracy: 1.0000
Epoch 14/30
5587/5587 [==============================] - 384s 69ms/step - loss: 1.6957e-06 - binary_accuracy: 1.0000 - val_loss: 4.5648e-06 - val_binary_accuracy: 1.0000
Epoch 15/30
5587/5587 [==============================] - 382s 68ms/step - loss: 8.3811e-04 - binary_accuracy: 0.9996 - val_loss: 1.5488e-06 - val_binary_accuracy: 1.0000
1179/1179 [==============================] - 23s 20ms/step
Convolutional Network model LOSS: 0.0007566791790672206
Convolutional Network model ACCURACY: 0.9991518259048462
```

As we can see in figure 8.2.1, we intended to train the network for 30 epochs, with a validation set and a clause that stopped the training as soon as the network ceased to show any improvements.

The history for these epochs were as follows:

Figure 8.2.2 - ConvNet accuracy vs validation set

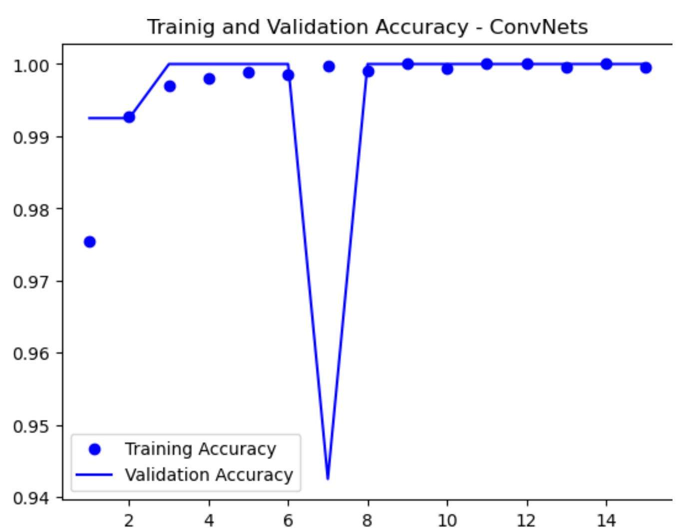
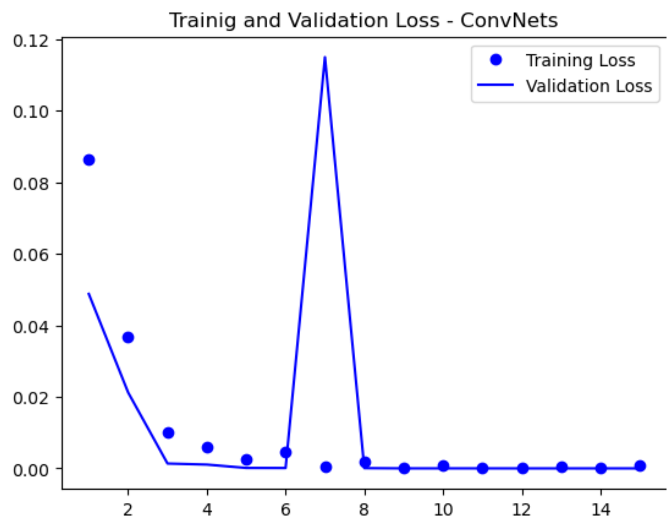


Figure 8.2.3 - ConvNet loss vs validation set



As can be appreciated in both images (Figure 8.2.2 and 8.2.3), the training and validation sets converges during cross validation training.

Therefore, we tested the model against the "testing set", which thought the following results:

• Convolutional Network model LOSS: 0.0007566791790672206

• Convolutional Network model ACCURACY: 0.9991518259048462

The current convolutional neural network reached an asymptotic curve around epoch 5; we saw afterwards that the network classified all the validation sets gravitational waves (9 out of 9) and almost all the same images from the tests set (9 out of 10). Given that the aim of this work was to classify "at least one image", we think that this model is appropriately trained for the scope of this work (in spite its possible issues with overfitting).

8.3 Linear SVC Approach

The Linear SVC (Linear Support Vector Classifier) it's a kind of support vector machine designed to deal with binary classification of complex data.

8.1 Techniques

8.2 Tools

9. References

[1] Zooniverse (2020), "Help scientists at LIGO search for gravitational waves, the elusive ripples of spacetime.". Recovered from https://www.zooniverse.org/projects/zooniverse/gravity-spy

[2] "Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science" M Zevin1, S Coughlin1, S Bahaadini2, E Besler2, N Rohani2, S Allen3, M Cabero4, K Crowston5, A K Katsaggelos2, S L Larson1,3, T K Lee6, C Lintott7, T B Littenberg8, A Lundgren4, C Østerlund5, J R Smith9, L Trouille1,3 and V Kalogera1Hide (February, 2017) doi: https://doi.org/10.1088/1361-6382/aa5cea

[3] Caltech, MIT, NSF (2020), "What is LIGO" [LIGO official site] recovered from https://www.ligo.caltech.edu/page/what-is-ligo

[4] Caltech, MIT, NSF (2020), "What is LIGO" [LIGO official site] recovered from https://www.ligo.caltech.edu/page/what-is-ligo

[5] NSF, LIGO, INFN, CNRS, LSC, EGO, VIRGO (2020). "Gravitational Wave Open Science Center" [Official site explaining how their interferometers works] recovered from https://www.gw-openscience.org/about/

[6] Caltech, MIT, NSF (2020), "What is an interferometer", recovered from https://www.ligo.caltech.edu/page/what-is-interferometer

[7] NASA (2020), "What is a gravitational Wave?" [NASA's official Site] recovered from https://spaceplace.nasa.gov/gravitational-waves/en/

[8] NASA (2020), "What is a gravitational Wave?" [NASA's official Site] recovered from https://spaceplace.nasa.gov/gravitational-waves/en/

[9] N. Mavalvala, "Precision measurement at the quantum limit in gravitational wave detectors," *2014 Conference on Lasers and Electro-Optics (CLEO) - Laser Science to Photonic Applications*, San Jose, CA, 2014, pp. 1-2.

[10] NSF, LIGO, INFN, CNRS, LSC, EGO, VIRGO (2020). (Datasets and coding libraries). Recovered from https://www.gw-openscience.org/archive/O2_4KHZ_R1/

[11] NSF, LIGO, INFN, CNRS, LSC, EGO, VIRGO (2020). (Datasets and coding libraries). Recovered from https://www.gw-openscience.org/yellow_box/

[12] "Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science" M Zevin1, S Coughlin1, S Bahaadini2, E Besler2, N Rohani2, S Allen3, M Cabero4, K Crowston5, A K Katsaggelos2, S L Larson1,3, T K Lee6, C Lintott7, T B Littenberg8, A Lundgren4, C Østerlund5, J R Smith9, L Trouille1,3 and V Kalogera1Hide (February, 2017) doi: https://doi.org/10.1088/1361-6382/aa5cea

[13] Daniel George, E.A. Huerta, (2017) "Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data", doi: 10.1016/j.physletb.2017.12.053

[14] GWOSC (2020), "Gravitational-Wave Open Data Workshop #3" [dataset and coding libraries], recovered fromhttps://github.com/gw-odw/odw-2019/blob/master/Challenge/CHALLENGE.md

[15] GWOSC (2020), "Gravitational-Wave Open Data Workshop #3" [Introduction to the workshop], recovered from https://www.gw-openscience.org/s/workshop3/

[16] GWOSC (2019), "GW-ODW2 -- Hands-on session" [Contests unofficial results document], recovered from https://bit.ly/2WRxjay

[17] Daniel George, E.A. Huerta, (2017) "Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data", doi: 10.1016/j.physletb.2017.12.053

[18] GWOSC(2016), "The O2 database release" [Dataset and coding libraries], recovered from https://www.gw-openscience.org/O2/

[19] Apache Spark (Version 2.5) [Software Library], recovered from https://spark.apache.org/

[20] The pandas development team (Version 1.0.3), recovered from https://pandas.pydata.org/docs/

[21] Keras (Version 2.0) [software library] recovered from https://keras.io/

[22] Numpy (Version 1.18) [Software library] recovered from https://numpy.org/

[23] Zenodo(2018), "Updated Gravity Spy Data Set" [Dataset and coding libraries] recovered from: https://zenodo.org/record/1476551#.Xu6TaZNKjNw, DOI: 10.5281/zenodo.1476551

[24] Chollet, F. (2018). *Deep learning with Python*. Shelter Island, NY: Manning Publications.

[25] "Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science" M Zevin1, S Coughlin1, S Bahaadini2, E Besler2, N Rohani2, S Allen3, M Cabero4, K Crowston5, A K Katsaggelos2, S L Larson1,3, T K Lee6, C Lintott7, T B Littenberg8, A Lundgren4, C Østerlund5, J R Smith9, L Trouille1,3 and V Kalogera1Hide (February, 2017) doi: https://doi.org/10.1088/1361-6382/aa5cea

[26] Chollet, F. (2018). *Deep learning with Python*. Shelter Island, NY: Manning Publications, pag. 120-122

[27]Chollet, F. (2018). *Deep learning with Python*. Shelter Island, NY: Manning Publications, pag. 133-138