Wavelet-based classification of transient signals for Gravitational Wave detectors

1st Elena Cuoco
European Gravitational Observatory
EGO and Scuola Normale Superiore
Pisa, Italy
elena.cuoco@ego-gw.it

2nd Massimiliano Razzano
Department of Physics, University of Pisa
Pisa University and INFN Pisa
Pisa, Italy
massimiliano.razzano@unipi.it

3rd Andrei Utina
Institute for Gravitational Research
Glasgow University
Glasgow, UK
2074809U@student.gla.ac.uk

***Abstract*—The detection of gravitational waves opened a new window on the cosmos. The Advanced LIGO and Advanced Virgo interferometers will probe a larger volume of Universe and discover new gravitational wave emitters. Characterizing these detectors is of primary importance in order to recognize the main sources of noise and optimize the sensitivity of the searches. Glitches are transient noise events that can impact the data quality of the interferometers and their classification is an important task for detector characterization. In this paper we present a classification method for short transient signals based on a Wavelet decomposition and de-noising and a classification of the extracted features based on XGBoost algorithm. Although the results show the accuracy is lower than that obtained with the use of deep learning, this method which extracts features while detecting signals in real time, can be configured as a fast classification system.**

***Index Terms*—signal processing, wavelet decomposition, machine learning classification**

I. Introduction

The detection of gravitational waves (GW) has opened a new window on the Universe. The first detected event, GW150914[1] and most of the following ones are interpreted as the coalescence and merging of binary black hole systems. However, the detection of GW170817 [2] has been produced by a pair of coalescing neutron stars and its electromagnetic counterpart has been detected at various wavelengths, thus inaugurating a new era of multi-messenger astronomy. Although coalescences of binary systems are a primary source of gravitational waves, other cosmic sources are believed to produce gravitational wave emission, such as core-collapse supernovae or rotating pulsars. Furthermore, a continuous stochastic background is expected due to the superposition of sources of different kind. Advanced LIGO [3] and Advanced Virgo [4] are second-generation laser interferometers that will be ten times more sensitive than previous LIGO and Virgo detectors. This jump in sensitivity will allow us to explore a larger volume of Universe and increase the rate of detections. The upgrades to Advanced LIGO and Advanced Virgo are related to all major detector subsystems, including optics, suspensions, and seismic attenuation systems. The construction of Advanced Virgo has been completed at the end of 2016 and the interferometer has joined LIGO for its second observing run (O2) on 1 August 2017 up to the end of the run on 25 August.

In order to detect transient gravitational wave signals, two main methods are used. A matched filtering technique is the preferred method when the gravitational waveform is well-known and modeled. In this case, a waveform model is correlated with the detector strain data [5][6].If the expected waveform is not known, an agnostic, excess-power strategy is adopted. The non-stationary, frequency-dependent noise in the detector limits the sensitivity of the searches. In particular, noise transients called *glitches* affect data quality in the interferometers and could also mimic the time-frequency behavior of astrophysical signals. Usually transient signals from astrophysical sources produce consistent delayed waveforms in the different detectors due to the propagation time of the waves, and so they can be discriminated from noise by comparing data in independent detectors [7]. However, near-simultaneous glitches can mimic astrophysical signals or some glitches could result coincident with a near-threshold Gaussian noise trigger, and thus it is important to monitor the auxiliary channels of each detector, usually not sensitive to gravitational waves but rich in information about the status of the environment and of the interferometer. Characterizing the glitches is an important task to reduce the impact of transient noise on the detectors. In particular, the classification of glitches can help to perform further investigations to establish its origin and prepare custom data quality flags in order to reduce their impact on the detector performance [7]. Inspecting glitches manually is a time-consuming and error-prone task. Furthermore, the increase of sensitivity in advanced detectors will lead to new classes of glitches. The use of machine learning looks to be a promising way to tackle the classification of glitches. In this paper we present a method to extract features while detecting signals in the GW main output channel, followed by XGBoost[8], a supervised technique for multi-label or binary classification.

II. Data Simulations

In order to test the algorithm, we simulated a set of glitches of different waveform and intensities. With this approach [9], [10] we can evaluate the classification performance depending on the parameters of the simulated glitches. The simulated

ISBN 978-90-827970-1-5

transients have been added to a Gaussian colored by the sensitivity curve (the power spectral density of the detector strain) of real interferometers. We used the design sensitivity of LIGO Hanford detector (H1) available online[1]. Each glitch is described by the strain time series h(t), and its injected SNR is the SNR obtained by the Wiener filer with that given waveform. We have simulated six families of signals, that approximate the time-frequency evolution of main noise transients observed in the real interferometers [7]. The simulated families are those described in [11] and are Gaussian glitches (GAUSS), sine-gaussian (SG), ringdown (RD), whistle-like and scattered light-like. We also included a family of chirp-like transients (CHIRPLIKE), in order to show the potential of the machine learning pipeline in distinguishing transients of astrophysical origin and of known waveform, such as those produced by coalescing binaries. Figure 1 shows the distributions of the simulated glitches.

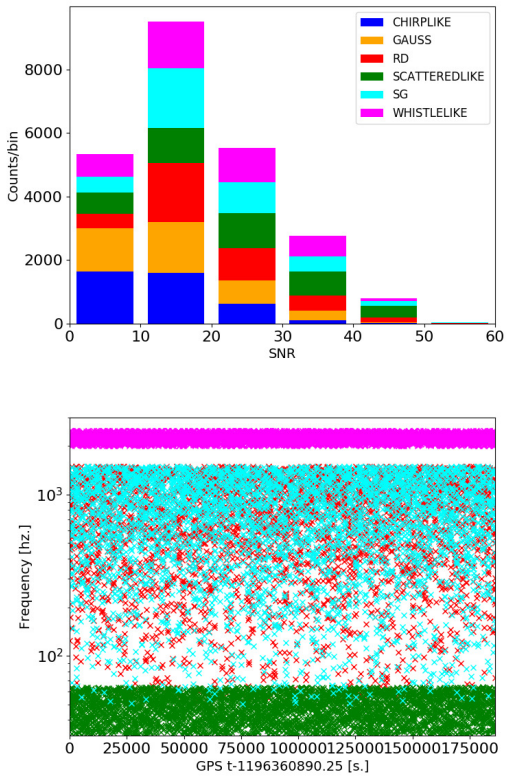

Fig. 1. Distribution of simulated glitches. Top: Distribution of Signal-to-noise ratios. Bottom: Time-frequency distribution. The SNR distribution is the result of our choice to have an uniform spacing in the parameter space of the simulated glitches. This results in a larger number of low-SNR glitches, which is good since we want to have a large statistics of faint glitches, more difficult to detect and classify.

III. Pipeline: WDFX

The Wavelet Detection Filter (WDF), developed by one of the author (Cuoco), was used for the first time in the analysis of the association between gamma-ray bursts and gravitational wave signal for the GRB 050915a detected by the Swift satellite in 2005, when the Virgo detector was engaged in one of its science runs, namely the C7 run [12]. WDF was used also in the first approaches for unsupervised classification for glitches [9], [13]. In this work we set up a new version of the pipeline based on a first stage where the WDF algorithm is applied and a second stage where the XGBoost [8] supervised classifier was used to identify the main types of signals. In this implementation the WDF run as multiprocess pipeline, in order to obtain a quasi real time pipeline [2].

1) Wavelet Detection Filter.: Wavelet-based algorithms are well tuned for the identification of transients signals; as different wavelet types could better match different waveform morphologies, WDF performs wavelet decomposition using different types of wavelet basis, including the Daubechies, Haar and spline wavelets [14]–[16]. The wavelet transform of a signal $f(t)$ is defined as the projection of f on the wavelet basis

$$Wf(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{b}} \psi^* \left(\frac{t-a}{b} \right) \quad dt, \quad (1)$$

where ψ^* is the complex conjugate of the mother wavelet. The wavelet transform has a time frequency resolution that depends on the scale b. The time spread is proportional to b, and the frequency spread is proportional to the inverse of b. The factor a represents a scale, which dilates or compress the signal.

A. Data Conditioning: Whitening procedure

Before applying any detection and classification pipeline, the data should be conditioned. The whitening procedure [17] can remove the contribution of Gaussian colored noise. Many pipelines use whitening in the frequency domain, while the whitening applied here is based on a time-domain procedure, using an Auto Regressive (AR) fit to the data as described in [17], [18]. The whitening procedure is based on a Linear Predictor Filter, whose parameters are estimated through a parametric Auto Regressive (AR) model fit to the noise Power Spectral Density (PSD), as described in [17]. In this work we used a model with 4000 AR parameters to fit the noise and whiten the data.

B. Wavelet de-noising procedure

A digital signal x_i that is corrupted by additive Gaussian random noise $n_i \sim \mathcal{N}(0, \sigma^2)$ is given by

$$x_i = h_i + n_i, \quad i = 0, 1, ...N-1, \quad (2)$$

where h_i is the transient signal. The signal x_i is used to find an approximation $\hat{h}_i$ to the original h_i, which minimizes the mean squared error

[1]Data publicly available at the LIGO Open Science Center

[2]Usually 1 second of raw data is analyzed in less than 3 seconds on a virtual machine 16 VCPU Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.4GHz equipped with 32GB RAM

ISBN 978-90-827970-1-5 © EURASIP 2018

$$\|\mathbf{h} - \hat{\mathbf{h}}\|^2 = \frac{1}{N}\sum_{i=0}^{N-1} |h_i - \hat{h}_i|^2. \quad (3)$$

If an orthogonal wavelet transform W is applied to the sequence of data x_i, we obtain $W(x_i) = W(h_i) + W(n_i)$. For a given threshold t and wavelet coefficient w, the wavelet coefficient is retained if $|w| > t$, or is set to zero if $|w| < t$. This removes wavelet coefficients that are due to background noise and retains wavelet coefficients that are due to the transient waveforms. WDF uses the universal Donoho and Johnstone threshold method [19], where $t = \sqrt{2\log N}\hat{\sigma}$, N is the number of data points, and $\hat{\sigma}$ is an estimate of the noise level σ, estimated during the AR parametric fit to the data. Given wavelet thresholding function t, the threshold based de-noising can be written as

$$\hat{h_i} = W^{-1}(t[W(x_i)]). \quad (4)$$

The thresholding function is applied to the wavelet transform of the noisy signal, then the output is inverted and the wavelet transformed. The effectiveness of the technique is dependent upon the choice of wavelet used, the decomposition level, and the amplitude of the threshold value. The wavelet coefficients contain the energy of the transient at different scales. After the wavelet thresholding procedure is applied, only the highest coefficients of the wavelet transform remain. These coefficients are expected to contain only features of the transient waveforms. The energy of the transient is given by the sum of the square of the coefficients above the threshold value. The SNR is then given by the energy divided by $\hat{\sigma}$. A newer implementation of WDF with respects to the work [9], [13] was applied in this paper. The clustering procedure of the event detected by the WDF was applied directly in the Wavelet domain, and only the coefficient above the threshold, whose relative distance differ in index values less than 10% in the wavelet plane are retained. This will help the signal waveform reconstruction, avoiding the insertion of noisy coefficients. The meta parameters (SNR, central frequency, duration) describing the detected waveform are estimated on the reconstructed waveform.

WDF produces a list of triggers, which include the maximum SNR and frequency, a GPS starting time for the transient, the transient duration, the name of the wavelet family which triggered the event, and the full list of the wavelet coefficients after the de-noising procedure.

For WDF to correctly identify the glitch, the choice of window size and overlapping parameter between two consecutive sliding windows becomes important. For this data set we select a window size of 1024 points,that is is 0.125 seconds, being the data sampled at 8192 Hz. We select a small overlapping window, since we assume all the glitches we are looking for are contained in 1 window. This is generally not true, since we simulated events longer than 1 second, but we plan to insert in the pipeline a second stage for signal reconstruction, linking the signals found in consecutive windows.

C. Supervised Machine Learning with XGBoost classifier

The concept of Machine Learning (ML) refers to a class of mathematical functions that are designed to learn from examples in order to solve new problems. In Machine Learning, two main branches are present: Supervised and Unsupervised Learning. In the Unsupervised Learning the algorithm needs to find correlation from an input of unstructured data while Supervised Learning means that the algorithm performs correlation tests after it is trained with labeled data.

Supervised learning takes an input variable containing sets of data of the form $X = (x_N, y_N)$ where x represents the feature vector while y represents the label vector. The process needs to be trained on this set of data to perform a mapping of the output variable $Y = f(X)$

The goal of the supervised learning, and in this case supervised classification, is to reach a high accuracy in predicting the output variable on a given set of input data.

EXtreme Gradient Boosting (XGBoost [8]) implements the gradient boosting framework. The gradient boosting decision tree algorithm is a decision tree implementation in which the objects or classes are classified by moving them down the tree from the root, to some leaf node where the classification is generated. At each node of the tree there is a test placed on that object and, based on this the object can take different classifications corresponding to each branch starting at the node. Boosting in this framework consists of having a large number of weak learners or trees and train them multiple times in order to minimize the loss function of the mapping.

A model present in the algorithm, for a total number of trees K, consists of a sum of decision trees, each with a specific prediction given by f_K. Using all the decision trees, a prediction y_i can be expressed as:

$$y_n = \sum_{1}^{K} f_K(x_n), \quad (5)$$

where x_n represents the feature vector of the nth component of the data series. The loss function, to be optimized when the model is trained, is the logarithmic loss [20]:

$$L = -\frac{1}{N}\sum_{1}^{N}((y_i \log(p_i) + (1 - y_i)(log(1 - p_i)) + \Omega \quad (6)$$

being p_i the predicted values and y_i the ground_truth value and Ω a regularization function added to avoid the problem of overfitting [8] This characteristic of the algorithm combines the predictive power given by the loss function and the linearity from the regularization term. Optimizing this objective is a key in learning problems and in XGBoost this is performed using gradient training descending calculations. After the calculations, the final form of the model contains parameters that can be used for fine-tuning.

IV. Results

As a first step we have created our training data, by running WDF on the simulated data sets. The second step is finding the signals coincident within 0.1 secs with the injected ones.

ISBN 978-90-827970-1-5 © EURASIP 2018

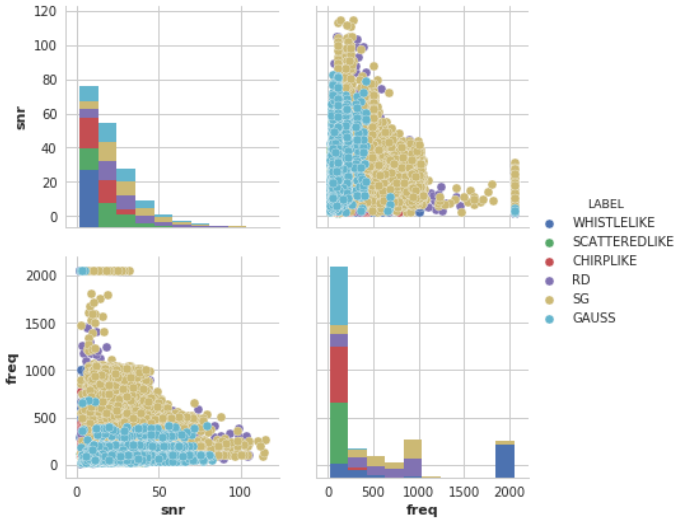

Fig. 2. The SNR and frequency distribution of the detected triggers are reported in the diagonal of the plot. Upper right the scatter plot of SNR versus frequency, bottom left the scatter plot of frequency versus SNR

WDF with a threshold of $SNR = 2$ was able to detect 97% of the injected signals, considering all the signals injected at different SNR, which is a very high efficiency, considering that also signals with $SNR = 1$ were injected in the data. The average of accidental coincidences, regardless the values of injected SNR, is around 10%, with a time-shift of 1 sec. A more detailed study about the efficiency and performance for WDF as event trigger generation for distinct type of signals is in progress.

In fig. 2 we reported the meta-parameter values of the recovered signals by WDF. We used a hyper-parameters selection method for the XGBoost classifier using a grid search cross validation [21]. The parameter selection could be further tuned, using also some features selection [22], which have not been applied in this work. We performed 2 main classification tasks:

- Binary classification: Chirp-like signal versus the rest,
- Multi-label classification.

To prepare our training data we selected only the coincident triggers, adding almost 2000 not coincident glitches labeled as 'NOISE', for a total of 25000 triggers. We then create train/validation/test sets in the ratio 70/15/15, by random shuffling the input data set. In table I we reported some of the parameters used in both tasks. In fig. 3 we reported the

TABLE I
XGBOOST PARAMETERS USED FOR CLASSIFICATION.

Classification task	**xgboost main parameters**			
	Classes	***Learning rate***	***max_depth***	***estimators***
Binary	2	0.01	7	5000
Multilabel	7	0.01	10	6000

loss function versus the number of estimators, for training and validation data sets. With the selected parameters for XGBoost, we kept under control the over-fitting. We obtained

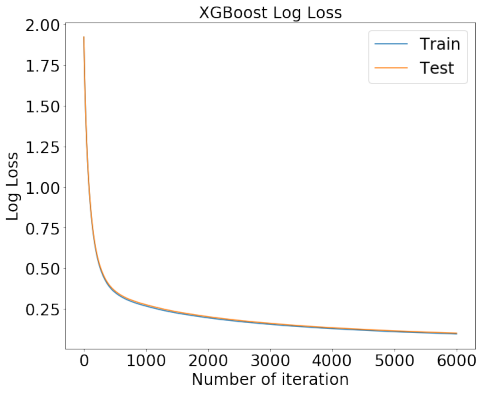

Fig. 3. Log loss function for training and validation data set for XGBoost multi-classification task versus number of estimators. With the selected parameters we had no over-fitting

an overall accuracy for binary classification of more than 95%, with a detailed confusion matrix reported in fig. 4. The noise events set was made by 'glitches' simulated events different by chirp-like and some random noise glitches. The chirp-like events have been correctly identified for 81% of the cases. While noise events for 98% ones. In the multi-class case, the

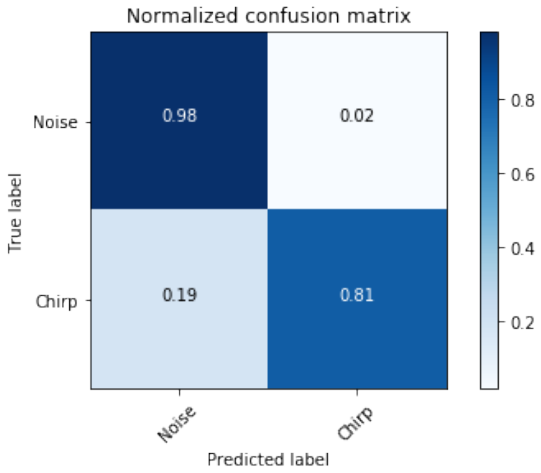

Fig. 4. XGBoost binary classification confusion matrix.

overall accuracy obtained was more than 82%, detailed in the confusion matrix of fig. 5. Longer signals classification have higher accuracy with respect to the short ones, in particular for the Sine Gaussian waveform. This could be due to the selection of a small overlap between two consecutive windows, in the event a signal falls between them. Work is in progress to tune the parameters used by WDF and the ones used for the classifier.

V. CONCLUSION

In this work we have tested the XGBoost algorithm efficiency in constructing a binary and a multi label classifier. The binary classifier has the purpose to accurately distinguish

ISBN 978-90-827970-1-5 © EURASIP 2018

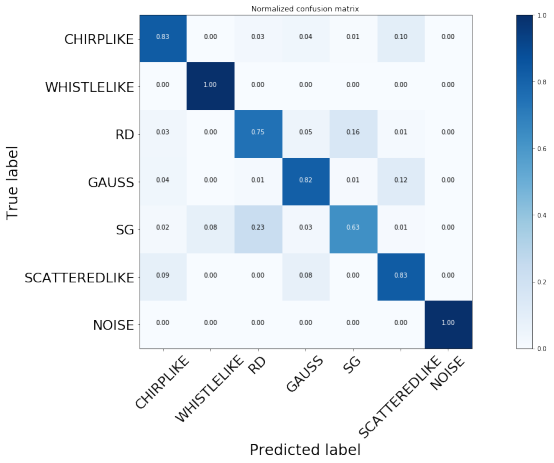

Fig. 5. XGBoost multi label classification confusion matrix.

between a "chirp-like" morphology which can be associated with a true compact binary coalescence event and other classes of "glitch" morphologies which might be caused by non-Gaussian external noise events. It has been shown in this work that the WDF extracted features and the XGBoost algorithm can classify different datasets of simulated waveforms a with an accuracy $> 90\%$. The power of the methods relies on the possibility to have one in-time classifier for the data acquired by the Gravitational Wave detectors. Further tuning of the parameters is needed to avoid the lack of the signals on the edges of the analyzing windows.

Acknowledgment

The authors thank the internal referee Jade Powell for the useful insights.

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Physical Review Letters*, vol. 116, no. 6, 061102, p. 061 102, Feb. 2016.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, "Gw170817: Observation of gravitational waves from a binary neutron star inspiral," *Phys. Rev. Lett.*, vol. 119, p. 161 101, 16 Oct. 2017.

[3] LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, *et al.*, "Advanced LIGO," *Classical and Quantum Gravity*, vol. 32, no. 7, 074001, p. 074 001, Apr. 2015.

[4] F. Acernese, M. Agathos, K. Agatsuma, *et al.*, "Advanced Virgo: a second-generation interferometric gravitational wave detector," *Classical and Quantum Gravity*, vol. 32, no. 2, 024001, p. 024 001, Jan. 2015.

[5] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley, 1949.

[6] B. S. Sathyaprakash and S. V. Dhurandhar, "Choice of filters for the detection of gravitational waves from coalescing binaries," *Phys. Rev. D*, vol. 44, pp. 3819–3834, 12 1991.

[7] N. Christensen, LIGO Scientific Collaboration, and Virgo Collaboration, "LIGO S6 detector characterization studies," *Classical and Quantum Gravity*, vol. 27, no. 19, 194010, p. 194 010, Oct. 2010.

[8] C. G. Tianqi Chen, "Xgboost: A scalable tree boosting system.," in *KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

[9] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglià, "Classification methods for noise transients in advanced gravitational-wave detectors," *Classical and Quantum Gravity*, vol. 32, no. 21, 215012, p. 215 012, Nov. 2015.

[10] N. Mukund, S. Abraham, S. Kandhasamy, S. Mitra, and N. S. Philip, "Transient classification in LIGO data using difference boosting neural network," *Physical Review D*, vol. 95, no. 10, 104059, p. 104 059, May 2017.

[11] M. Razzano and E. Cuoco, "Image-based deep learning for classification of noise transients in gravitational wave detectors," *Classical and Quantum Gravity*, vol. 35, no. 9, p. 095 016,

[12] Acernese, F. et al, "Gravitational waves by gamma-ray bursts and the virgo detector: The case of grb 050915a," vol. 24, S671, Sep. 2007.

[13] Powell J. et al, "Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data," *Classical and Quantum Gravity*, vol. 34, no. 3, 034002, p. 034 002, Feb. 2017.

[14] I. Daubechies *et al.*, *Ten lectures on wavelets*. SIAM, 1992, vol. 61.

[15] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1998.

[16] M. Unser, "Ten good reasons for using spline wavelets," in *Proc. SPIE Vol. 3169, Wavelets Applications in Signal and Image Processing*, 422431.

[17] Cuoco E. et al, "On-line power spectra identification and whitening for the noise in interferometric gravitational wave detectors," *Classical and Quantum Gravity*, vol. 18, pp. 1727–1751, May 2001.

[18] Cuoco et al, "Noise parametric identification and whitening for LIGO 40-m interferometer data," *Physical Review D*, vol. 64, no. 12, p. 122 002, Dec. 2001.

[19] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[20] C. M. Bishop, *Pattern recognition and machine learning (information science and statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006, ISBN: 0387310738.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[22] J. F. T. Hastie R. Tibshirani, *The elements of statistical learning*. Springer, 2001.