Modelling, Detection, and Classification: Applying Machine Learning to Gravitational Wave Astronomy

Angela Y. Zheng
angelayang.zheng@mail.utoronto.ca
University of Toronto
Toronto, Ontario, Canada

Abstract

Gravitational wave (GW) astronomy is an emerging field of astrophysics research that is focused on the ripples in spacetime caused by cataclysmic events. GW detectors such as the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO) interferometers, which first detected gravitational waves in 2015, continue to collect constant streams of GW data. In this paper, we will outline the current state of machine learning (ML) algorithms in GW research. Many forms of ML frameworks have been proposed with the purpose of improving data quality through noise detection and filtering, detecting GW signals in real time, and estimating the parameters of the GW sources through the development of theoretical models of waveforms. Moreover, there is strong potential for the use of ML in analyzing resulting GW data to support related research, such as the Baryon Acoustic Oscillation (BAO) scale. Preliminary trials of ML algorithms on real data have shown promising results, and researchers have further goals of putting the proposed algorithms into practice for real-time analysis as scientific instruments continue to improve.

1 Introduction

An emerging field of astrophysics research is gravitational wave (GW) astronomy, the study of ripples in spacetime caused by cataclysmic events. These include the merging of binary black holes (BBH), collisions of neutron stars, and the continuous spin of massive objects. First predicted by Einstein through his general theory of relativity in 1916, gravitational waves were not confirmed to exist until Hulse and Taylor [1975] discovered a binary pulsar in 1974.

Today's observations are done by the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO) detectors, which are 4 km long cavity arms with good sensitivity across a wise band centered around 100 Hz and located in Hanford, Washington and Livingston, Louisiana [8]. The first observation of gravitational waves was made at LIGO in 2015 [1] and the observatory has since continued to be at the frontier of gravitational wave astronomy. After replacing the initial LIGO detectors, the advanced LIGO detectors have an order of magnitude better sensitivity in its most sensitive band and improved noise levels. Its detections "could include supernovas, compact object mergers, neutron star instabilities and cusps in cosmic strings" [8]. Most importantly, its continued operation around the clock provides a continuous stream of valuable GW data for analysis.

In the past, the size of astronomical data was manageable such that the main form of research could be done using "qualitative inspection supported by quantitative analysis" [5]. However, the sheer volume and complexity of data now makes this impossible, leading to the adoption of machine learning (ML) algorithms and artificial intelligence (AI), as "processing and analyzing the increased rate of detections in future observing runs will require researchers to streamline current search pipelines" [4].

Fluke and Jacobs [2020] identified seven main categories of activity—classification, regression, clustering, forecasting, generation, discovery, and the development of new scientific insights—and three stages of maturity—emerging, progressing, and established—in their survey of ML and AI in astronomy. They identified that many ML techniques employ artificial neural networks (ANNs), which first appeared in the 1980s. Specifically, convolutional neural networks (CNNs), an extension of ANNs, contain convolutional layers that are sensitive to specific features that may have undergone transformations. The final stage of a CNN is a fully-connected ANN which can generate a classification or numerical prediction, and is often used for "a variety of image-based classification, regression, and discovery activities" [5] including gravitational wave events.

A GW detector's output "is a temporal series of the detector strain" [4]. A major obstacle in the analysis of GW data is the presence of noise, including seismic noise, thermal noise due to Brownian motion, shot noise due to quantum uncertainties in laser light, and non-astrophysical triggers in data, known as glitches [12]. The automatic classification of these types of noise, "extracting features from each glitch time series and mapping these features to the target glitch types" [4], would significantly increase the speed at which data can be passed on to analysis. Multiple ML algorithms were proposed by Powell et al. [2015], which have then been tested on Year 1 LIGO data [11]. Zevin et al. [2017] proposed Gravity Spy, a project leveraging human pattern recognition skills to categorize images of glitches to use as training data for ML algorithms.

Additionally, ML models can detect GW signals in real time and estimate the parameters of the source through the development of theoretical models of waveforms [13]. Deep neural networks (DNNs) can be trained on time-series inputs for "rapid detection and characterization of gravitational wave signals" [7]. George and Huerta's [2018] proposed Deep Filtering algorithm is able to both detect the presence of a signal in the input and estimate the parameters of the source of the signal, creating a "single, robust, efficient data analysis pipeline for GW detectors".

Finally, in the applications of GW research, there exists potential for ML to help analyze resulting data. From the first year of the Dark Energy Survey (DES), Abbott et al. [2019] converted samples of galaxies into a "three dimensional map in 'photometric redshift space' " to predict the Baryon Acoustic Oscillation (BAO) scale. BAO is an important standard for cosmology and crucial for the understanding of dark energy, and its study uses data collected by GW detectors.

2 Improving Gravitational-Wave Data Quality

Currently, one of the most common applications of ML in gravitational wave astronomy is to remove noise from the Advanced LIGO measurements, as “the sensitivity of advanced gravitational-wave detectors [is] limited by multiple sources of noise from the hardware subsystems and the environment” [12] and “the output signal measured by [the] interferometers is the sum of noise and GW signals” [4].

Some sources of noise for Advanced LIGO are: “quantum noise [due to] statistical fluctuations in detected photon arrival”; Brownian noise from test mass thermal noise; thermal noise in the test mass suspension, “due to loss in the fused silica fibres”; and gravity gradients due to seismic waves [14].

The term “glitches” refers to “non-astrophysical triggers in science data” caused by “instrumental and environmental disturbances” [12]. To prevent these glitches from being mistaken as GW detections, searches for unmodelled GW signals “currently combine the noise from multiple detectors coherently” [12]. This is done because it is assumed that the detectors are independent from one another, meaning that non-astrophysical noise would generally only occur in one channel and be discarded.

However, glitches can still occur at the same time in multiple detectors. These events are frequent enough to necessitate the classification of glitches in order to eliminate noise transients, which are short-duration, sudden fluctuations in the amplitude of the GW signal that are often non-stationary and non-Gaussian. Previously, visually inspecting the glitches’ time series was done to classify them, but this is a slow and inefficient method that would not keep up with the Advanced LIGO detectors. In Figure 1 and Figure 2, Powell et al. [2017] outlined common glitch types in time-domain spectrograms from the Livingston and Hanford Advanced LIGO interferometers, with a clear visual difference between morphological transient classes.

Zevin et al. [15] described the three main ways that glitches impact data analysis: (1) increasing the loudness of the background in GW searches, thereby reducing the significance of candidate events, (2) impacting the recovery of astrophysical parameters from a GW source, and (3) reducing the amount of usable data.

Powell et al. [12] and Zevin et al. [15] both propose ML algorithms that can identify and classify these transient waveforms, allowing them to be removed appropriately from the data in real-time, as well as use crowdsourcing to further improve the training datasets for the ML models. ML models excel at being able to handle “huge data sets, proving invaluable in analyzing auxiliary channel data” [4], which contain data from a multitude of sensors. Figure 3 shows an example of a signal spike which is observed in multiple auxiliary channels, therefore indicating that this spike was non-astrophysical in origin.

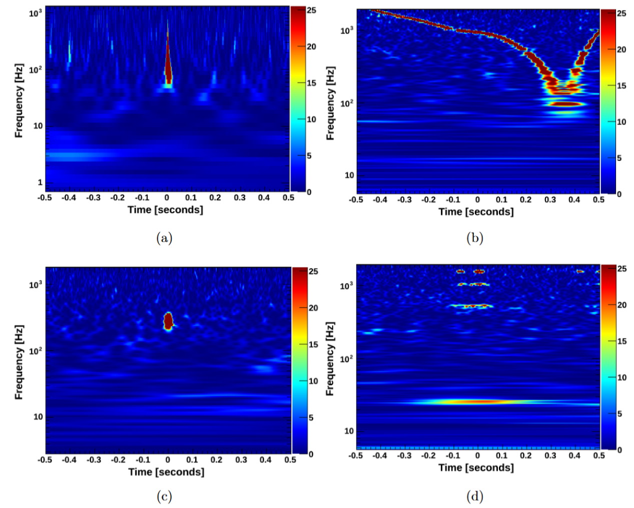

Figure 1: Examples of typical transient types in spectrograms from Advanced LIGO Livingston ER7 data. (a) A tear drop shape. (b) A whistle glitch. (c) A hardware injection. (d) A type that has high frequency lines and lower frequency features (reproduced from [11]).

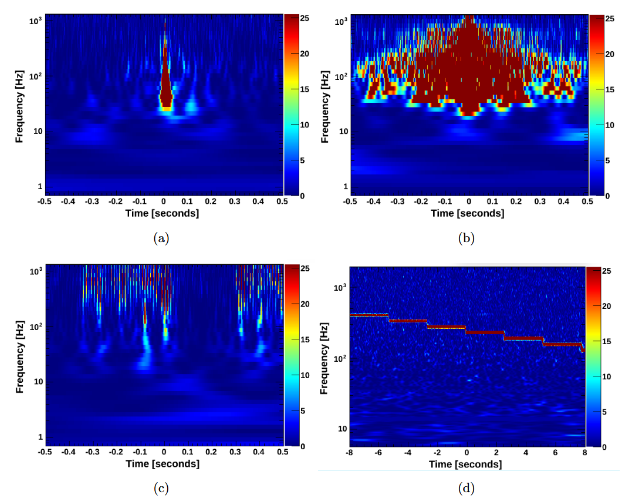

Figure 2: Examples of typical transient types in spectrograms from Advanced LIGO Hanford ER7 data. (a) A tear drop shape. (b) A transient type with large signal-to-noise ratio and duration. (c) A high frequency type. (d) A longer duration line (reproduced from [11]).

2.1 Proposed Algorithms for Real-Time Noise Classification

Powell et al. [2015] developed “three methods that can be used for the fast classification of advanced detector noise transients” and tested them against real LIGO data in 2016, since “prompt characterization of instrumental and environmental noise transients will be critical for improving the sensitivity of the advanced detectors”. The main use case for their proposed algorithms are for glitches, which are classified based on their characteristics in time-series spectrograms.

The three algorithms are “Principal Component Analysis for Transients (PCAT), an adoption of LALInference Burst (LIB), and a combination of a trigger generator called Wavelet Detection Filter

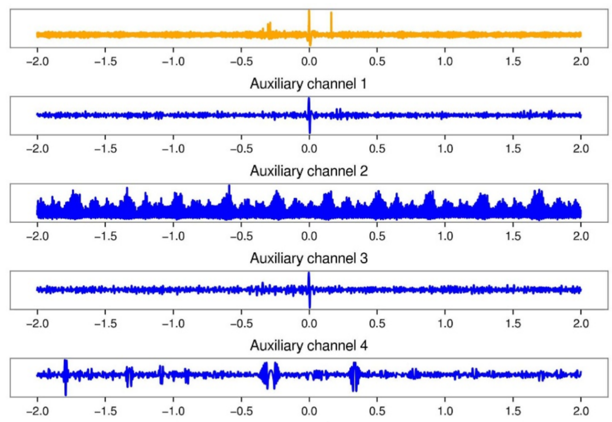

Figure 3: An example of time series auxiliary channels, with the top one monitoring GW signals and the others monitoring sources not sensitive to GW. The spike at $t = 0$ occurring in multiple channels indicates that this trigger is not caused by an astrophysical source (reproduced from [4]).

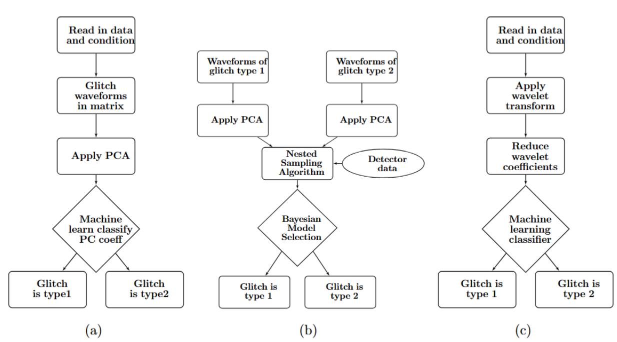

Figure 4: A workflow diagram displaying the classification methods and procedures for the three algorithms proposed by Powell et al [2017, 2015]. (a) PCAT. (b) PC-LIB. (c) WDF-ML (reproduced from [12]).

and Machine Learning techniques (WDF-ML)" [12]. The general workflow for each algorithm is depicted in Figure 4.

2.1.1 Principal Component Analysis for Transients (PCAT). PCAT is based on the use of Principal Component Analysis (PCA), with Powell et al's [2015] version using Python, and identifies and classifies noise transients. PCA "consists of a linear orthogonal transformation of a set of (possibly correlated) variables into a set of linearly uncorrelated variables, called Principal Components (PCs)" which "define the direction of greatest variation in the data" [12]. The PCs indicate the time variability of the LIGO channel and specify the properties of the transients, in this case, the noise transients in the time-domain.

In the pre-processing phase, the data is down-sampled, high passed with a Butterworth 4th order filter, then whitened along with the Fast Fourier Transform (FFT). Noise transients are identified when the channel amplitude exceeds a chosen threshold, after which it is transformed into a standardized data matrix. This data matrix is decomposed into multiple matrices, one of which contains the PCs in its rows, sorted by decreasing eigenvalues. Once the data matrix is projected onto the PC basis, the first few coefficients (those with the greatest eigenvalues) "identify the most important features of the glitch waveforms that can be separated from the noise" [12]. Finally, PCAT "uses the scikit-learn Gaussian Mixture Model (GMM) algorithm to cluster the PCA-reduced data" [12].

Importantly, the number of PCs must be chosen carefully, as a low number of PCs implies insufficient information and a high number of PCs includes noise features in the reduced dataset, both leading to the clustering algorithm performing poorly. Although there is no existing method to compute the ideal number of PCs directly, researchers commonly set a threshold on the explained variance, which is a statistical measure computed from the eigenvalues.

2.1.2 LALInference Burst (PC-LIB). LALInference Burst (LIB) "is a Bayesian parameter estimation algorithm for parameter estimation or model selection for gravitational-wave burst signals" [12]. Powell et al. [2015] adopted the PCA approach taken by Logue et al. [10] for the classification of glitches, taking the time series of fifty glitches, applying a high pass filter, and FFT the waveforms (since "LIB performs model selection in the frequency domain" [12]). PCA is applied to the waveforms similarly to the PCAT algorithm (Section 2.1.1), where the linear combination of the PCs (using the PC coefficients) is considered the new signal model. Finally, these signal models for each glitch population are used for Bayesian model selection, which "can determine the type of each new noise transient that is detected" [12]. The number of PCs is chosen so that they accurately describe "a large percentage ($\gtrsim$ 70%)" [12] of the explained variance in the data, similar to the method described in Section 2.1.1.

2.1.3 Wavelet Detection Filter and Machine Learning (WDF-ML). Wavelet-based algorithms "decompose the data into multiple time-frequency resolution maps" [12] in a method similar to a Fourier transform. It performs "wavelet domain decomposition using different types of wavelet basis" [12], decomposing the signal into a mutually orthogonal set of wavelets.

The original signal is projected onto this wavelet basis, thus the wavelet coefficients (the resulting coefficients of the projection mapping) "contain the energy of the transient at different scales" [12]. After thresholding is applied, "only the highest coefficients of the wavelet transform" [12] remain. These highest coefficients should only contain features of the transient waveforms, allowing for the correct identification of the glitch.

WDF-ML initially used an unsupervised classification procedure, as there was no labelled dataset at the time. A supervised ML algorithm is planned to be implemented using auxiliary monitoring channel data. The unsupervised clustering algorithm identifies classes of events in the parameter space created by the wavelet decomposition, and finally applies the GMM classification algorithm.

2.1.4 Evaluation of Proposed Algorithms. Powell et al. [11] followed up with analysis done on the 7th Advanced LIGO engineering run (ER7), which was conducted during 2015. The two interferometers at Hanford, Washington (H1) and Livingston, Louisiana (L1) collected two separate data sets.

From the L1 detector, out of the noise transients that were not missed, PCAT classified 95% of them correctly and PC-LIB classified 98% correctly. WDF-ML classified all of the transients and 95% were correct. From the H1 detector, PCAT classified 99% of non-missed transients correctly with 95% for PC-LIB. WDF-ML again classified all transients with 92% accuracy. They concluded that all three proposed algorithms can classify transient waveforms and have “a high efficiency in real non-stationary and non-Gaussian detector noise” [11].

The three algorithms have different strengths and weaknesses. PCAT can “classify new types of transients as soon as they appear in the data” [11], thus creating transient waveforms for PC-LIB’s signal models. PC-LIB has a longer analysis window, and so it better classifies longer duration transients. Finally, WDF-ML has the ability to classify lower frequency transients unreachable by the other two algorithms.

In their future work, they plan to include multiple auxiliary channels in the classification procedure, use a training set of pre-classified waveforms, and explore the use of dictionary learning algorithms for glitch classification.

These three algorithms are significant in being a highly useful application of ML in GW research, allowing glitch classification to be performed in real time in accordance with the Advanced LIGO data collection streams. “As real noise from the advanced detectors is non-stationary and non-Gaussian, a better understanding of how [their] methods will perform during the upcoming observation runs of the advanced detectors is required” [11].

2.2 Gravity Spy: Crowdsourced Training for Machine Learning Algorithms

When training ML algorithms, the lack of high-quality, pre-classified data can be a limitation. Zevin et al. [2017] proposed Gravity Spy, a project combining crowdsourcing with ML for gravitational noise classification. The core idea lies in “engaging and recruiting volunteers from the public to categorize images of time-frequency representations of glitches into pre-identified morphological classes and to discover new classes that appear as the detectors evolve” [15]. ML algorithms then classify images after being trained on the human-classified datasets.

Using the Zooniverse platform, they leverage human pattern recognition skills by providing labeled morphological classes as training data. The volunteers can filter based on the duration, frequency, and whether the glitch is evolving in time. After being classified by these volunteers, the data then gets passed on to the ML algorithm which classifies the data in full. It determines confidence scores for each classification and returns the most questionable glitches back to the citizen scientists for further analysis. The full project workflow is demonstrated in Figure 5.

Interestingly, the learning pipeline is not one-way; Zevin et al. [15] proposes the use of ML results to train new human volunteers, as shown in Figure 6. The platform moves new volunteers through a sequence of levels and are presented with increasing amounts of glitch classes and feature sophistication, allowing them to improve their ability to classify glitches. Volunteers get trained on “gold standard images”, which are a subset of the dataset that had already

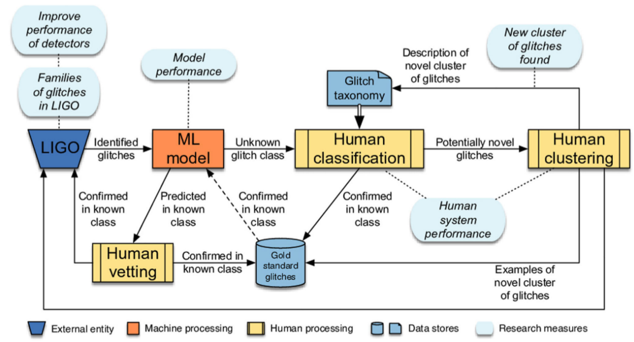

Figure 5: A workflow diagram for the Gravity Spy system architecture, especially between the interdisciplinary components (reproduced from [15]).

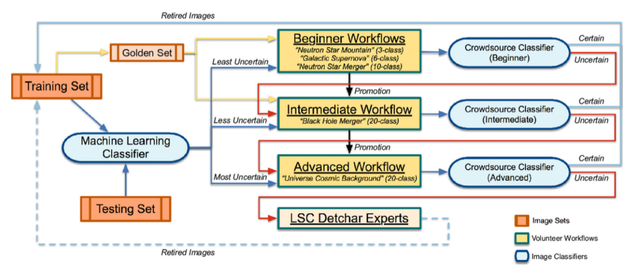

Figure 6: A workflow diagram showing the movement of images and volunteers through the Gravity Spy project (reproduced from [15]).

been classified by experts, after which they get immediate feedback on the accuracy of their chosen class. Finally, the algorithm assigns tasks based on the capability and historical accuracy of each volunteer, as it keeps track of “confusion matrices” for each volunteer.

The ML algorithm used is a deep model with CNN layers, which is typical for image classification, along with a rectified linear unit (ReLU) as an activation function.

The initial training set for the algorithm was determined after consulting with experts, and it contained 7,718 glitches grouped into 20 classes. The Gravity Spy project was launched publicly on October 12, 2016 for the “initial renditions of machine learning models and beta-testing of the human interface” [15]. The ML algorithm was able to achieve an average accuracy of 97.1% for the testing data.

Before this launch, beta testing in 2016 “engaged over 14,000 users and delivered over 45,000 glitch classifications” and incredibly led to the “discovery of multiple new and substantial glitch categories from LIGO first observing run” [15]. Two of these glitch classes were later named “Paired Doves” and “Helix”. Moreover, Gravity Spy caused hundreds of conversation threads on the Zooniverse talk forum and “fostered excitement and intrigue for the nascent field of GW astrophysics” [15].

A major form of testing for the future is experimenting with various parameters relating to the training and performance of the volunteers, including the speed at which to promote volunteers to higher levels, the types of motivational messages provided, and the information shown about each type of glitch. A concern noted by Zevin et al. [2017] was "how much the volunteers [would] need to know about gravitational-wave astrophysics and the workings of the detectors that produce the glitches", so they incorporated a mini-course on GW and Advanced LIGO into the platform.

In the future, the Gravity Spy project will continue to operate and engage with community volunteers with additional tools and "interaction between project scientists and volunteers on the Talk forum" [15]. It will also incorporate data from the multiple new interferometers joining the Advanced LIGO network. Zevin et al. [2017] concluded that "the integration of human and computer classification schemes will maintain citizen science as a prolific scientific tool and allow it to scale with the ever-increasing datasets of the future".

3 Real-Time Gravitational Wave Signal Detection and Parameter Estimation

In addition to the elimination of transient noise in GW data, there have also been proposals for ML models to be used in the entire pipeline of detecting GW signals. This specifically involves developing accurate theoretical models of GW signals in order to be able to approximate and determine the location and parameters of the GW sources and, later, the clustering of similar signals. These proposals would allow the analysis of interferometer data to happen in real time, as researchers develop increasingly efficient computational models.

3.1 Developing Theoretical Models of Waveforms

GW analysis depends on accurate "theoretical models of the gravitational-wave signal that is emitted as binaries coalesce" [13], as the data from the detectors are "filtered with many theoretically predicted waveforms with varying binary parameters" [13]. These waveform models are crucial to determining accurate parameters for the GW sources.

Setyawati et al. [2020] reviewed the current algorithms being used, which largely fall under two categories: analytical and numerical relativity (NR). The main analytical model, Post-Newtonian (PN) expansion, breaks down in certain parts of the signal. Meanwhile, NR waveforms "require high computational resources" [13]. The two new methods being developed to model full waveforms are effective-one-body (EOB) and phenomenological, which "start from a reformulation of PN results and calibrate the model to a select number of NR simulations" [13]. Both rely on interpolation—"typically [fitting] free coefficients to a set of NR data"—and "traditional methods originated from approximation theory and numerical analysis in mathematics" [13].

Machine learning techniques generally involve Gaussian process regression (GPR) and ANNs. In their study, Setyawati et al. [2020] trained a multi-layer perceptron (MLP) model on data from binary black hole (BBH) systems and found that "there is no guarantee that many neurons [in the neural network] yield smaller error than fewer neurons" [13], with extra neurons leading to overfitting and fewer neurons leading to underfitting. While both traditional and ML models performed better with more training data, Setyawati et al. [2020] concluded that "one should critically evaluate the performance of approximation methods and understand the features of the method that are necessary for the data of interest", as simpler methods can save significantly on computation time.

3.2 Deep Learning for Real-Time Gravitational Wave Detection

Another such proposal was made by George and Huerta [2018], involving deep neural networks (DNNs), specifically two CNNs that take time-series inputs for the "rapid detection and characterization of gravitational wave signals". One CNN acts as a "classifier" to detect the presence of a signal, providing a confidence level for the detection, and the other acts as a "predictor" to estimate the parameters of the signal, such as the component masses of a BBH. Their algorithm takes time-series inputs for both classification and regression.

Currently, Advanced LIGO uses a time-domain matched-filtering algorithm, the most sensitive GW detection algorithm available. It targets a 3 parameter space ("compact binary sources with spin-aligned components on quasi-circular orbits" [7]) and currently requires over 2 s to analyze 1 s inputs.

Meanwhile, the DNN was tested with real LIGO noise, with different realizations of noise randomly sampled from data. This includes "injections of GW templates originating from quasi-circular, non-spinning, stellar-mass BBH systems" [7], with an example of an injection depicted in Figure 7. They found that the Deep Filtering algorithm could automatically recognize glitches and classify them as noise, enabling the "detection of new classes of GW sources", as well as "interpolate between waveform templates, in a similar manner to [GPR]" [7]. When the signal-to-noise ratio (SNR) was greater than 10, the sensitivity was 100% and the false alarm rate was able to be tuned below 1%, which can be decreased further to 0.01% when the classifier is applied independently to each of the two detectors and coincidence is enforced.

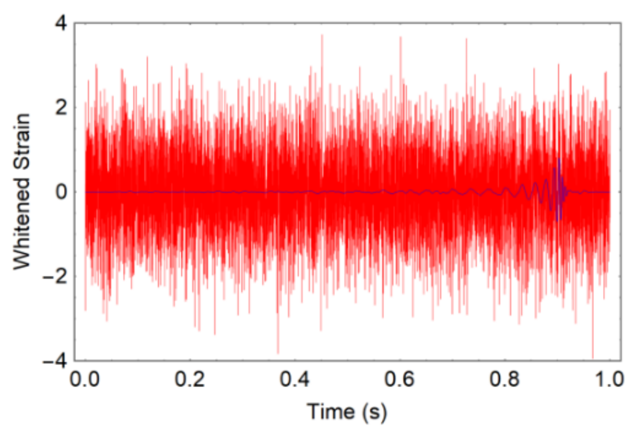

Figure 7: A hidden BBH GW signal (blue) injected into real LIGO noise. The GW signal was detected with over 99% sensitivity by the Deep Filtering algorithm (reproduced from [7]).

The significance is not only the accuracy and potential of this algorithm, but also the speed and size; "Deep Filtering achieves similar sensitivities and lower errors compared to matched-filtering while being more computationally efficient and more resilient to glitches" [7], as seen in Figure 8. The Deep Filtering algorithm takes only 85 ms to analyze 1 s of input data on a single CPU core, or 540 µs on a GPU, allowing for faster-than-real-time analysis. Both CNNs are only 23 MB in size, and encode all relevant information from 2500 GW templates and several GB of noise used to generate the training data. Also, while the time for matched-filtering grows exponentially with the number of parameters, the Deep Filtering algorithm requires only a one-time training process; afterward, its analysis is performed in constant time.

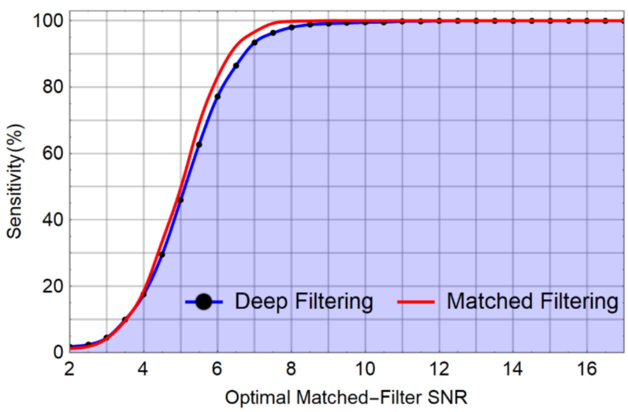

Figure 8: A comparison of the proposed Deep Filtering algorithm and the currently used Matched Filtering algorithm in sensitivity of GW signal detection (reproduced from [7]).

This leads to the potential of creating a unified pipeline for GW detectors, completing glitch classification, signal detection, parameter estimation, and clustering in real-time. George and Huerta [2018] conclude that the Deep Filtering algorithm could benefit from training on other researchers' existing data, such as those found in the Gravity Spy project [15], as described in Section 2.2.

3.3 Potential Pitfalls of ML in the Detection of GW Signals

Gebhard et al. [2019] analyzed how CNNs can be used to search for GWs from compact binary coalescences (CBCs). They particularly point out possible challenges with the data generation process that "may lead to unfair comparisons" [6], demonstrating with concrete examples that their "architecture is also prone to adversarial attacks" [6].

Referencing George and Huerta [2018] (Section 3.2) as well as other previous works on CNNs used for GW astronomy, Gehbard et al. [2019] list some potential pitfalls:

(1) CNNs' success rate with new inputs "depends also on the relative frequencies of positive and negative examples in the training set" [6], and the false positive rate calculated on training data is not necessarily representative of the true confidence level for real data.
(2) For an "arbitrarily long input time series that contains a signal" [6], the model must determine the step size of the sliding window for streaming data, which contrasts with the often discrete isolated time series data used in existing literature.
(3) Overfitting the model to the data due to specific ways of constructing training data means that CNNs could perform worse on real GW data.

Subsequently, Gebhard et al. [2019] "develop a promising proof of concept implementation" that avoids these issues. They outline their PyCBC search procedure: first, a bank of simulated waveforms is created through simulation, then an SNR time series is calculated, and finally, the peaks in these time series are clustered alongside the parameters to become a trigger.

By adding injections, which are "a simulated waveform that is added into a piece of background noise ... to emulate a real gravitational-wave signal", Gebhard et al. [2019] could evaluate the detection and false alarm rate of the search pipeline. In the training procedure, they chose a fully convolutional architecture, being more computationally efficient than a sliding window approach due to the overlaps in steps performing redundant computations. Finally, the network output is smoothed (through rolling average) and thresholded (mapped from a continuous range to discrete values based on a set threshold).

For the full test set, the model was able to successfully recover 89% of injections, with the smoothing and thresholding procedures acting as tuning knobs by "trading off the sensitivity and the false positive rate" [6].

In their published data generation pipeline, they highlighted a selection of " 'failure modes' of [their] model which are typical for deep convolutional neural networks" [6], where signals that are structurally different from real GW are detected as a false positive. Despite their notes of caution, Gebhard et al. [2019] conclude that CNNs are still a promising tool to produce real-time triggers due to their ability to scale up the number of waveforms presented in training with fast and efficient computation times.

4 Machine Learning in Applications of Gravitational Wave Astronomy

Machine learning is not only useful for the detection of gravitational waves, but also for the related applications of GW astronomy.

One important use for the data from GW detectors is to measure the BAO scale with increased precision. The "signature of BAO can be observed in the distribution of tracers in the matter density field and [is] used to measure the expansion history of the universe" [2]. The BAO acts as a standard ruler, "an object of a known size at a single redshift, z, or a population of objects at different redshifts whose size changes in a well-known way (or is actually constant) with redshift" [3]. This is especially important for any study in cosmology, refining the Hubble parameter, and reconstructing "the underlying dark matter distribution from discrete tracers such as galaxies" [3].

Abbott et al. [2019] used imaging data from the first year (Y1) of the Dark Energy Survey (DES) to present "angular diameter distance measurements", including a sample of over 1.3 million galaxies. To validate their methodology, they first used 1,800 simulations that approximated their data sample.

Then, they obtained two different photometric redshift estimates by converting the galaxy sample into a "three dimensional map in 'photometric redshift space' by converting angles and redshifts to physical distances" [2], and obtained results using two different photometric redshift estimates. One was machine learning based (DNF) while the other was template based (BPZ), yielding results that matched within 0.37σ. Even after conducting stress tests and varying their methodology, they found that there were no significant changes in the BAO measurements.

Ultimately, Abbott et al.'s [2019] measurement of DES Y1 BAO is consistent with their simulations of the DES Y1 dataset and are the first BAO measurement results to be obtained from DES. Their future plans include using a more realistic modeling of photometric redshift errors, reducing the "uncertainty in the mean of the redshift distributions", and "further [investigating] the dependence of the covariance matrix on cosmological parameters and other choices" [2]. However, none of these sources of error would result in systematic uncertainties comparable to the expected future statistical precision, and they expect to be obtaining results with smaller statistical uncertainties by a factor of 2 with DES year 3 (Y3) data [2].

5 Conclusion

As the field of gravitational wave astronomy continues to develop, the possibilities of machine learning applications will only increase. Advanced LIGO continuously collects new data, providing a growing GW dataset for analysis.

Already, there are many ways that machine learning algorithms are being used to support GW astronomy. In the field of noise classification and elimination, Powell et al. [2017, 2015] developed and tested three types of ML algorithms that can classify glitches. Additionally, the Gravity Spy project [15] crowdsources glitch image classification from online volunteers, showcasing the potential of citizen science projects. Moreover, theoretical waveform models are being developed [13], while George and Huerta [2018] proposed a Deep Filtering algorithm which is both more computationally efficient and more resilient than the current matched-filtering algorithm. It integrates the entire pipeline into one, consisting of signal detection, noise classification, and parameter estimation. Finally, gravitational wave data is used for other related studies, such as measuring the Baryon Acoustic Oscillations scale for the fields of cosmology and dark energy. Abbott et al. [2019] were able to use machine learning to produce a photometric redshift estimate of the BAO scale that matched their template-based estimate.

In the future, researchers plan to unify different proposed machine learning algorithms into a single pipeline that can detect GW signals, classify and eliminate noise, and perform data analysis. Additionally, the continued innovation in scientific instruments will allow Advanced LIGO and its partner sites to detect more GW signals, and researchers hope to implement ML algorithms in the production chain for real-time analysis. Ultimately, machine learning is an innovative technique that is becoming crucial for data analysis in the growing field of gravitational wave astronomy.

References

[1] Benjamin P. Abbott et al. 2016. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters* 116, 6 (2016), 061102. doi:10.1103/PhysRevLett.116.061102 (LIGO Scientific Collaboration and Virgo Collaboration).

[2] Timothy M. C. Abbott et al. 2019. Dark Energy Survey Year 1 results: measurement of the baryon acoustic oscillation scale in the distribution of galaxies to redshift 1. *Monthly Notices of the Royal Astronomical Society* 483, 4 (2019), 4866–4883. doi:10.1093/mnras/sty3351 (The Dark Energy Survey Collaboration).

[3] Bruce A. Bassett and Renée Hložek. 2010. Baryon Acoustic Oscillations. In *Dark Energy: Observational and Theoretical Approaches*, Pilar Ruiz-Lapuente (Ed.). Cambridge University Press, 151–176. doi:10.1017/CBO9780511779398.009

[4] Elena Cuoco, Jade Powell, Marco Cavaglià, Kendall Ackley, Michał Bejger, Chayan Chatterjee, Michael Coughlin, Scott Coughlin, Paul Easter, Reed Essick, Hunter Gabbard, Timothy Gebhard, Shaon Ghosh, Leïla Haegel, Alberto Iess, David Keitel, Zsuzsa Márka, Szabolcs Márka, Filip Morawski, Tri Nguyen, Rich Ormiston, Michael Pürrer, Massimiliano Razzano, Kai Staats, Gabriele Vajente, and Daniel Williams. 2020. Enhancing gravitational-wave science with machine learning. *Machine Learning: Science and Technology* 2, 1 (Dec. 2020), 011002. doi:10.1088/2632-2153/abb93a

[5] Christopher J. Fluke and Colin Jacobs. 2020. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge Discovery* 10, 2 (March 2020), e1349. doi:10.1002/widm.1349

[6] Timothy D. Gebhard, Niki Kilbertus, Ian Harry, and Bernhard Schölkopf. 2019. Convolutional neural networks: A magic bullet for gravitational-wave detection? *Physical Review D* 100, 6 (Sept. 2019), 063015. doi:10.1103/PhysRevD.100.063015

[7] Daniel George and Eliu A. Huerta. 2018. Deep Filtering: real-time gravitational wave astrophysics with deep learning. *Physical Review D* 97, 4 (2018), 044039. doi:10.1103/PhysRevD.97.044039

[8] Gregory M. Harry et al. 2010. Advanced LIGO: the next generation of gravitational wave detectors. *Classical and Quantum Gravity* 27, 8 (2010), 084006. doi:10.1088/0264-9381/27/8/084006 (LIGO Scientific Collaboration).

[9] Russell A. Hulse and Joseph H. Taylor. 1975. Discovery of a pulsar in a binary system. *The Astrophysical Journal* 195 (Jan. 1975), L51. doi:10.1086/181708

[10] Joshua Logue, Christian David Ott, Ik Siong Heng, Peter M. Kalmus, and James H. C. Scargill. 2012. Inferring core-collapse supernova physics with gravitational waves. *Physical Review D* 86, 4 (Aug. 2012), 044023. doi:10.1103/PhysRevD.86.044023

[11] Jade Powell, Alejandro Torres-Forné, Ryan Lynch, Daniele Trifirò, Elena Cuoco, Marco Cavaglià, Ik Siong Heng, and José A. Font. 2017. Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data. *Classical and Quantum Gravity* 34, 3 (2017), 034002. doi:10.1088/1361-6382/34/3/034002

[12] Jade Powell, Daniele Trifirò, Elena Cuoco, Ik Siong Heng, and Marco Cavaglià. 2015. Classification methods for noise transients in advanced gravitational-wave detectors. *Classical and Quantum Gravity* 32, 21 (2015), 215012. doi:10.1088/0264-9381/32/21/215012

[13] Yoshinta Setyawati, Michael Pürrer, and Frank Ohme. 2020. Regression methods in waveform modeling: a comparative study. *Classical and Quantum Gravity* 37, 7 (2020), 075012. doi:10.1088/1361-6382/ab693b

[14] The LIGO Scientific Collaboration et al. 2015. Advanced LIGO. *Classical and Quantum Gravity* 32, 7 (March 2015), 074001. doi:10.1088/0264-9381/32/7/074001

[15] Michael Zevin, Scott Coughlin, Sara Bahaadini, Emre Besler, Neda Rohani, Sarah Allen, Miriam Cabero, Kevin Crowston, Aggelos K. Katsaggelos, Shane L. Larson, Tae Kyoung Lee, Chris Lintott, Tyson B. Littenberg, Andrew Lundgren, Carsten Oesterlund, Joshua R. Smith, Laura Trouille, and Vicky Kalogera. 2017. Gravity Spy: integrating Advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity* 34, 6 (2017), 064003. doi:10.1088/1361-6382/aa5cea