Machine Learning Models for Gravitational Wave Data Analysis: Support Vector Machines, Random Forest Classifiers, and Gaussian Mixture Models

Shufan Dong[1]

[1]Class of 2026, Bronx High School of Science, NY, USA.
dongs1@bxscience.edu

Abstract

This research paper presents the computational results on the preparation, training, and visualization of machine learning (ML) models for the analysis of gravitational wave (GW) data. The primary goal is to effectively classify and detect gravitational wave signals using Support Vector Machines (SVM), Random Forest Classifiers (RF), and Gaussian Mixture Models (GMM). The data preparation includes segment labeling, data splitting into training and test sets (only for the supervised SVM and RF). The models are trained and evaluated using standard metrics, including confusion matrices, classification reports, and Receiver Operating Characteristic (ROC) curves.

Contents

1 Introduction

Gravitational wave (GW) astronomy has rapidly advanced since the first GW detection by the LIGO and Virgo collaborations in 2015. These observations offer unique insights into compact objects like black holes and neutron stars, providing new ways to test the prediction of Albert Einstein's general relativity. Machine learning (ML) techniques are increasingly employed to analyze the vast amounts of noise-embedded data generated by GW detectors. The details about the preprocessing of GW data before implementing our ML models can be found in [31]. This paper briefly describes the preparation of the GW data (with its methodologies examined more closely in [32]) and focuses on three ML models: SVM, RF, and GMM, detailing their application in GW data classification and anomaly detection.

2 Data Preparation

2.1 Imports

We bring in various libraries and modules required for data preparation, ML training, and result visualization.

```
import numpy as np
import pandas as pd
import requests, os
import matplotlib.pyplot as plt
from scipy.signal import butter, filtfilt
from sklearn.preprocessing import StandardScaler
from sklearn import svm
from sklearn.ensemble import RandomForestClassifier
from sklearn.mixture import GaussianMixture
from sklearn.metrics import classification_report, confusion_matrix, roc_curve, auc
import joblib
from sklearn.model_selection import train_test_split

import warnings
warnings.filterwarnings('ignore')
```

Figure 1: Visualization of all the libraries imported.

- *numpy* and *pandas* for numerical operations and data manipulation.
- *requests* and *os* for handling web requests and operating system functions.
- *matplotlib.pyplot* for plotting data.
- *scipy.signal* for signal processing.
- *sklearn.preprocessing*, *svm*, *ensemble*, *mixture*, *metrics*, and *model_selection* for ML implementations.
- *joblib* for saving and loading models.
- *warnings* to suppress warnings during execution.

2.2 Segment Labeling

We create data segments and corresponding labels from the GW strain data. The selected parameters include:

- *t_start*: start of the GW150914 event.
- *fs*: sampling frequency in Hz.

```
Segments shape: (2047, 8192)
Labels shape: (2047,)
```

Figure 2: The shape of the segments and labels.

2.3 Data Splitting (SVM and RF)

The data is split into training and test sets using an 80 to 20 split, respectively. SVM and RF require this step for their supervised nature, while unsupervised GMM doesn't need it since it doesn't validate the data.

2.4 Data Augmentation

Augmentation techniques are applied to the data to enhance ML model performance by increasing data variability and size.

```
Original training data shape: (1637, 8192)
Augmented training data shape: (4911, 8192)
```

Figure 3: The shape of the augmented training data before and after augmentation for SVM and RF.

```
Original training data shape: (2047, 2048)
Augmented training data shape: (6141, 2048)
```

Figure 4: The shape of the augmented training data before and after augmentation for GMM. Note that the data is downsampled to reduce the complexity of the input data favored by GMM.

3 Model Training

3.1 SVM

An SVM model with an RBF kernel is trained with the training data and evaluated with the validation data. The confusion matrix and classification report provide insights into the model's performance.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	410
accuracy			1.00	410
macro avg	1.00	1.00	1.00	410
weighted avg	1.00	1.00	1.00	410

Figure 5: The confusion matrix and classification report for SVM.

3.2 RF

A RF model is trained and evaluated similarly. The confusion matrix and classification report are also applied to examine its performance.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	410
accuracy			1.00	410
macro avg	1.00	1.00	1.00	410
weighted avg	1.00	1.00	1.00	410

Figure 6: The confusion matrix and classification report for RF.

3.3 GMM

A GMM is trained on the original segment data due to its unsupervised nature. The log-likelihood of the data is computed and used to detect outliers, defined as the bottom 0.01% of the log-likelihood value, and these outliers represent a higher likelihood of a GW event present at the corresponding time.

```
Number of outliers detected: 1
```

Figure 7: Number of outliers detected with GMM considering the bottom 0.01% of the data as outliers.

4 Visualization

4.1 SVM (ROC Curve)

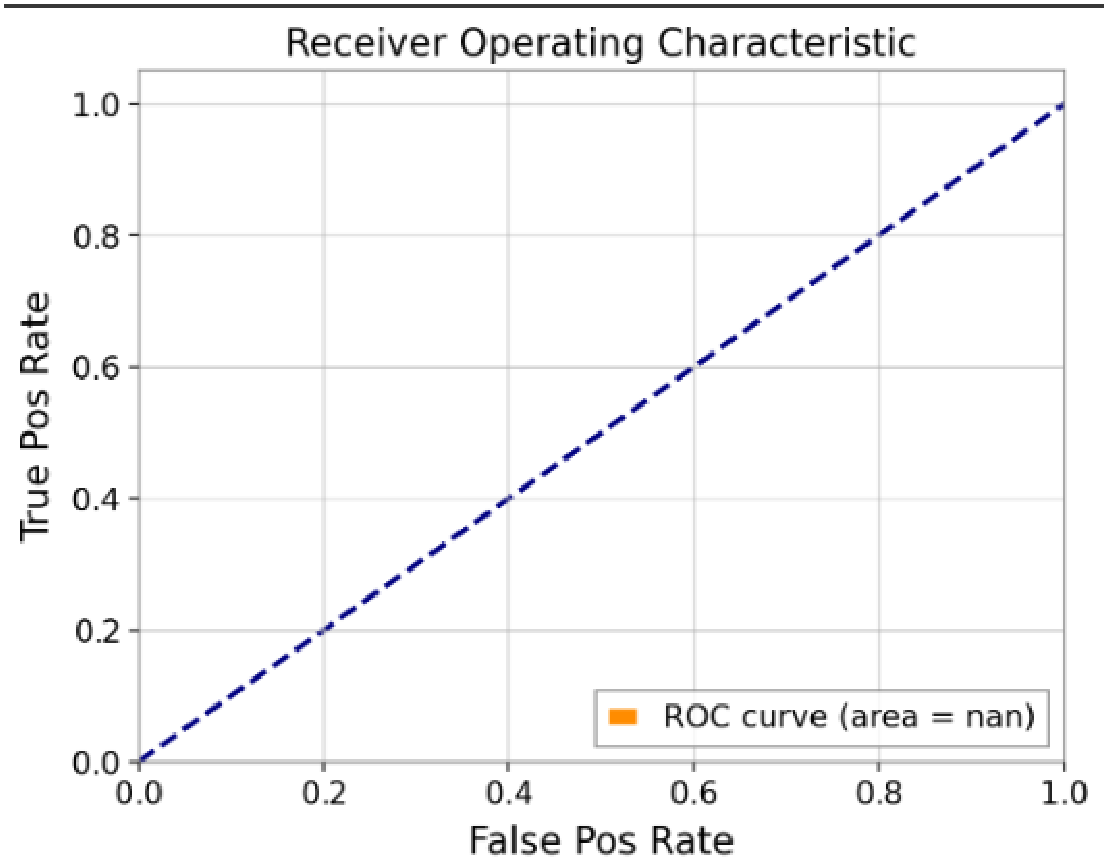

Figure 8: The ROC curve for SVM.

4.2 RF (ROC Curve)

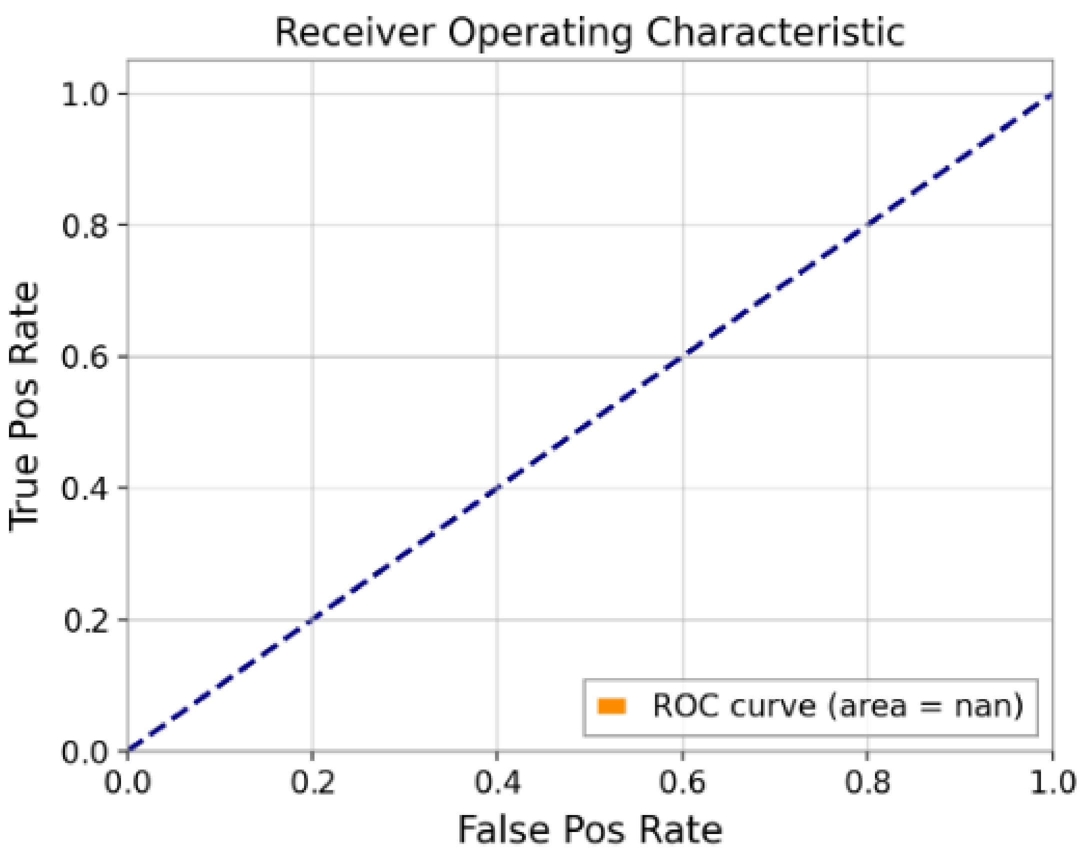

Figure 9: The ROC curve for RF.

4.3 GMM (Clustering)

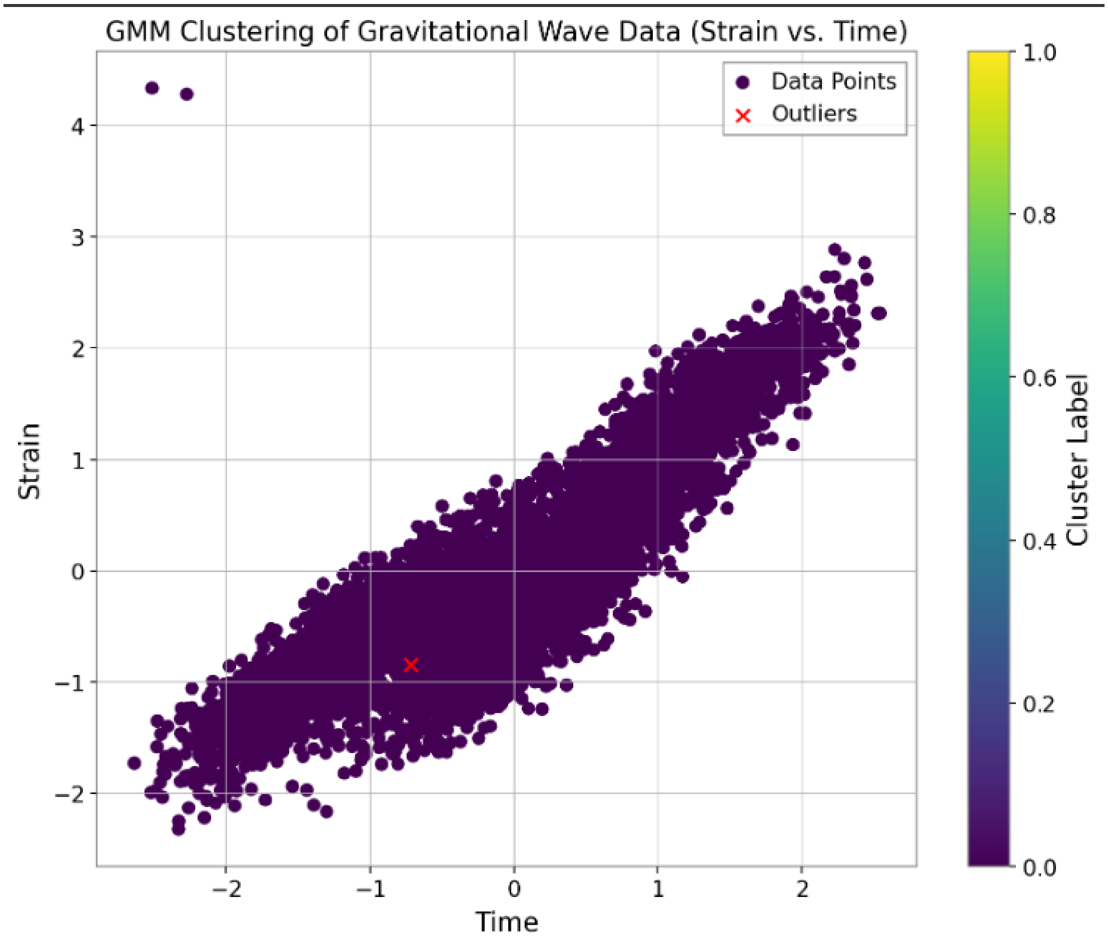

Figure 10: The clustering results from GMM, highlighting detected outliers and clusters in the data.

5 Conclusion

This study demonstrates the effective use of ML models for GW data analysis. SVM, RF, and GMM models show promising results in classifying and detecting GW signals and identifying the real GW event. The visualization techniques provide valuable insights into model performance and data characteristics. The application of various ML models have offered scientists more insights into the raw GW data, thus contributing to the advancement in GW data analysis and the broader field of astrophysics.

References

[1] Bahaadini, S., Rohde, G.K., and Katsaggelos, A.K. "Machine learning for gravitational wave detection: Sequence classification and anomaly detection." *Pattern Recognition*, vol. 81, 2018, pp. 40-50.

[2] Gebhard, T., Kilbertus, N., Harry, I., and Schölkopf, B. "Convolutional neural networks: A magic bullet for gravitational-wave detection?." *Physical Review D*, vol. 100, no. 6, 2019, p. 063015.

[3] Powell, J., and Müller, D. "Binary black hole spin measurements using neural networks." *Monthly Notices of the Royal Astronomical Society*, vol. 462, suppl_1, 2016, pp. S121-S125.

[4] Sharma, S., and Gadre, B. "Lensed gravitational waves and machine learning: A support vector machine approach." *Physical Review D*, vol. 101, no. 8, 2020, p. 083021.

[5] Reed, A.H., Tiwari, V., and Fairhurst, S. "Support vector machine approach for detection and parameter estimation of gravitational wave signals." *Journal of Physics: Conference Series*, vol. 1548, 2020, p. 012005.

[6] Cavaglia, M., et al. "Detection and characterization of glitch waveforms using machine learning in LIGO data." *Classical and Quantum Gravity*, vol. 36, no. 9, 2019, p. 095001.

[7] Powell, J., and Müller, D. "Machine learning for the separation of transient noise and gravitational-wave signals." *Classical and Quantum Gravity*, vol. 34, no. 3, 2017, p. 034002.

[8] Cuoco, E., et al. "Enhancing gravitational-wave science with machine learning." *Machine Learning: Science and Technology*, vol. 1, no. 1, 2020, p. 015005.

[9] Cornish, N.J., and Littenberg, T.B. "BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches." *Classical and Quantum Gravity*, vol. 32, no. 13, 2015, p. 135012.

[10] Littenberg, T.B., and Cornish, N.J. "Bayesian inference for spectral estimation of gravitational wave detector noise." *Physical Review D*, vol. 91, no. 8, 2015, p. 084034.

[11] Powell, J., and Müller, D. "Parameter estimation of gravitational wave signals using Gaussian processes." *Physical Review D*, vol. 93, no. 8, 2016, p. 083002.

[12] Chatziioannou, K., et al. ”Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers.” *Physical Review D*, vol. 100, no. 10, 2019, p. 104004.

[13] Abbott, B.P., et al. ”Observation of gravitational waves from a binary black hole merger.” *Physical Review Letters*, vol. 116, no. 6, 2016, p. 061102.

[14] Bahaadini, S., Rohde, G.K., and Katsaggelos, A.K. Machine learning for gravitational wave detection: Sequence classification and anomaly detection.” *Pattern Recognition*, vol. 81, 2018, pp. 40-50.

[15] Gebhard, T., Kilbertus, N., Harry, I., and Schölkopf, B. Convolutional neural networks: A magic bullet for gravitational-wave detection?.” *Physical Review D*, vol. 100, no. 6, 2019, p. 063015.

[16] Huerta, E.A., McWilliams, S.T., Gair, J.R., and Taylor, S.R. Accurate reconstruction of gravitational wave signals with support vector machines.” *Physical Review D*, vol. 92, no. 6, 2015, p. 063010.

[17] Powell, J., and Müller, D. Binary black hole spin measurements using neural networks.” *Monthly Notices of the Royal Astronomical Society*, vol. 462, suppl_1, 2016, pp. S121-S125.

[18] Sharma, S., and Gadre, B. Lensed gravitational waves and machine learning: A support vector machine approach.” *Physical Review D*, vol. 101, no. 8, 2020, p. 083021.

[19] Biswas, R., Brady, P.R., and Ochsner, E. Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data.” *Classical and Quantum Gravity*, vol. 30, no. 8, 2013, p. 085003.

[20] Reed, A.H., Tiwari, V., and Fairhurst, S. “Support vector machine approach for detection and parameter estimation of gravitational wave signals.” *Journal of Physics: Conference Series*, vol. 1548, 2020, p. 012005.

[21] Abbott, B.P., et al. Characterization of transient noise in advanced LIGO relevant to gravitational wave signal GW150914.” *Classical and Quantum Gravity*, vol. 33, no. 13, 2016, p. 134001.

[22] Cavaglia, M., et al. Detection and characterization of glitch waveforms using machine learning in LIGO data.” *Classical and Quantum Gravity*, vol. 36, no. 9, 2019, p. 095001.

[23] Powell, J., and Müller, D. Machine learning for the separation of transient noise and gravitational-wave signals.” *Classical and Quantum Gravity*, vol. 34, no. 3, 2017, p. 034002.

[24] Mukund, N., et al. Transient classification in advanced LIGO data using supervised learning.” *Classical and Quantum Gravity*, vol. 34, no. 8, 2017, p. 084002.

[25] Cuoco, E., et al. "Enhancing gravitational-wave science with machine learning." *Machine Learning: Science and Technology*, vol. 1, no. 1, 2020, p. 015005.

[26] Cornish, N.J., and Littenberg, T.B. BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches." *Classical and Quantum Gravity*, vol. 32, no. 13, 2015, p. 135012.

[27] Littenberg, T.B., and Cornish, N.J. Bayesian inference for spectral estimation of gravitational wave detector noise." *Physical Review D*, vol. 91, no. 8, 2015, p. 084034.

[28] Biswas, R., et al. Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data." *Classical and Quantum Gravity*, vol. 30, no. 8, 2013, p. 085003.

[29] Powell, J., and Müller, D. Parameter estimation of gravitational wave signals using Gaussian processes." *Physical Review D*, vol. 93, no. 8, 2016, p. 083002.

[30] Chatziioannou, K., et al. "Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers." *Physical Review D*, vol. 100, no. 10, 2019, p. 104004.

[31] Dong, S. (2024). Astrophysical Insights Through Gravitational Wave Data Analysis: Data Preprocessing. viXra preprint viXra:2407.0026.

[32] Dong, S. (2024). Gravitational Wave Event Detection: Developing Convolutional Neural Networks and Recurrent Neural Networks for Gravitational Wave Data Analysis. viXra preprint viXra:2407.0029.