JpGU-AGU JOINT MEETING 2026 | 24 – 29 MAY CHIBA, JAPAN | TRANSFORMATIVE SCIENCE FOR LIFE & EARTH | AGU ADVANCING EARTH AND SPACE SCIENCES

DECAGMon | Tool for identifying correlation and coherence between multi-channels by measuring information-theorectic index

Yubin Kim[1], Byeol Kim[2], John J. Oh[1]

[1] Department of Fundamental Research on Public Agenda, National Institute for Mathematical Sciences, Daejeon, South Korea
[2] Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea

Abstract

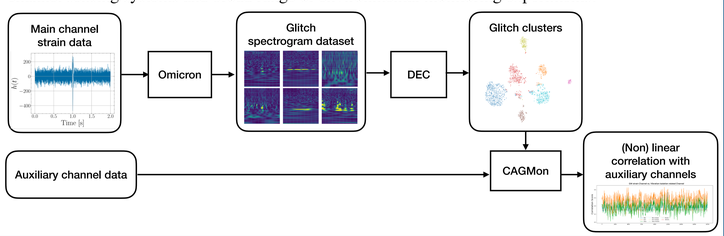
Ground-based precision measurement systems are often affected by transient, non-stationary noise artifacts originating from instrumental and environmental disturbances. In large-scale detector facilities, thousands of auxiliary channels continuously monitor environmental conditions and subsystem behavior, providing critical information for identifying and diagnosing transient noise sources. We present a machine learning-based framework that combines unsupervised clustering with multi-channel correlation analysis for automated transient noise characterization and diagnostics. Deep Embedded Clustering (DEC), an unsupervised deep learning algorithm that jointly optimizes feature extraction and cluster assignments in a low-dimensional latent space, is employed to automatically group transient events according to their morphological similarity. Following clustering, CAGMon — a correlation analysis tool based on information-theoretic indices — is applied to identify linear and non-linear couplings between auxiliary channels and transient event populations, enabling inference of likely physical noise origins. The proposed framework provides a scalable and efficient approach for automated diagnostics in complex multi-channel sensing systems and may be applicable to a broad range of precision measurement and environmental monitoring experiments.

Introduction

Large-scale precision measurement systems operate in highly non-stationary and non-linear noise environments, where instrumental and environmental couplings (seismic, magnetic, acoustic) induce transient noise artifacts (glitches). In ground-based gravitational wave detectors, such glitches contaminate the strain channel, hindering signal identification and data quality. Identifying and mitigating transient nosies is therefore an essential task for stable operation and high-quality data acquisition in complex sensor systems.

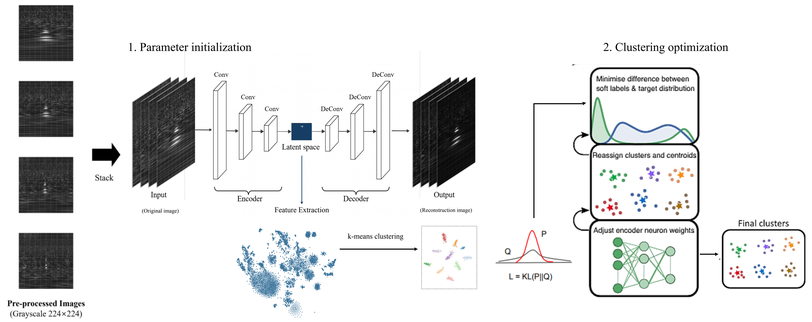
Conventional glitch classification is often based on the assumption that morphologically similar transient patterns originate from similar physical mechanisms. However, supervised image-classification approaches generally require extensive manually labeled training datasets, limiting scalability and adaptability to newly emerging noise populations. To address this challenge, we integrate deep-learning-based unsupervised clustering with multi-channel correlation analysis, providing an automated and scalable framework for transient noise characterization and source diagnostics. The proposed framework combines Deep Embedded Clustering (DEC) for morphology-based clustering with CAGMon, a correlation analysis tool based on Pearson's correlation coefficient (PCC), Kendall's tau (Ktau), and the Maximal Information Coefficient (MIC), enabling identification of both linear and non-linear couplings between auxiliary channels and transient event populations.

We demonstrate the framework using practical examples from the ground-based gravitational wave detector, including transient populations with previously identified physical origins. The proposed framework provides a scalable and efficient approach for automated diagnostics in complex multi-channel sensing systems and broad range of environmental monitoring experiments.



Method: DECAGMon

Deep Embedded Clustering: Unsupervised clustering algorithm based on neural network



- DEC [1] is an unsupervised deep learning algorithm that simultaneously learns extracting key features of input data through an autoencoder, mapping it to a low-dimensional embedding space suitable for clustering.
- Identified transient power events with high signal-to-noise ratios (SNR) are converted into 4-timescale (0.5s, 1.0s, 2.0s, 4.0s) spectrograms, pre-processed for dimensionality reduction, and utilized as input for deep learning model.
- Due to the image-based input, a convolutional autoencoder is utilized to initialize network weights, followed by k-means clustering to establish initial cluster centroids.
- Each glitch spectrogram sample's cluster assignments are iteratively refined toward a target distribution, enabling the model to form well-separated clusters in the latent space without any prior labels.

CAGMon: Tool for identifying correlation between multi-channels of the gravitational wave detector

- CAGMon[2] is a noise diagnostic tool that quantifies both linear and non-linear correlations between the GW strain channel and auxiliary channels using three statistical measures (PCC, Ktau and MIC).
- This study employ CAGMon to analyze the correlation between the resulting glitch clusters and auxiliary channels.

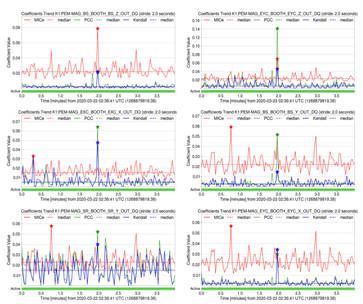
$$\text{PCC:}\quad \rho(X,Y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

$$\text{Ktau:}\quad \tau(X,Y)=\frac{2(c-d)}{n(n-1)}$$

$$\text{MIC:}\quad MICe(X,Y,\alpha,c)=\max_{ab<B(n)}\left\{\frac{\max I^{[*]}(S,k,l)}{\log_2\min\{k,l\}}\right\}$$

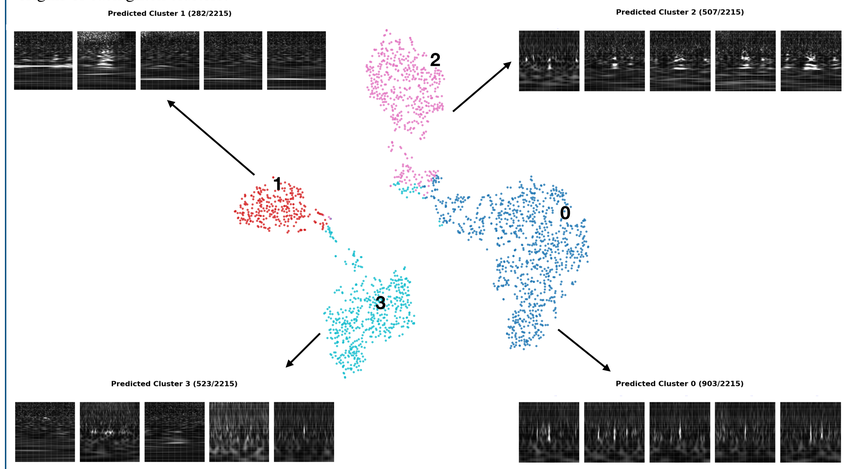
where $I^{[*]}(S,k,l)$ denotes the maximized mutual information (MI) in $k \times l$ grids.

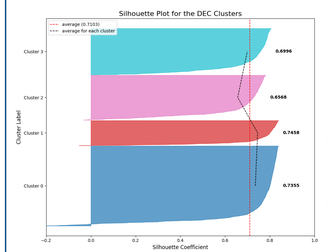
- It successfully identified a significant correlation between the KAGRA(Japanese underground gravitational wave detector) [3] main channel and magnetometer-associated auxiliary channels at the moment of a lightening strike event, validating its effectiveness on real observational data. (Shown on the right figure)



Result (Preliminary)

We present preliminary unsupervised clustering of 2,215 glitch events dataset [4] collected from the transient noise data obtained from main channel of ground-based gravitational wave detector during Apr. 7–21, 2020. The strain data of glitch event time segment are converted into 2-dimensional spectrogram image using python module for gravitational wave data analysis. The DEC model has been trained using the sole image inputs only, and the optimal number of glitch clusters (K=4) determined automatically by maximizing clustering performance index on a 20% test set. The resulting cluster assignments for the full dataset are shown below, grouped by morphological similarity in spectrogram. The ultimate goal of this study is to reveal the correlation between glitch morphology and its originating auxiliary channels by applying CAGMon analysis to each identified cluster, thereby inferring the physical origins of each glitch clusters.





- Samples within each cluster exhibit visually similar morphological structures, suggesting that DEC successfully captures shared time-frequency features. The clustering results for total samples achieved a **global silhouette score of 0.7103** (range: −1 to 1; higher values indicate better-separated and more cohesive clusters), demonstrating strong clustering quality across the full dataset.
- The auxiliary channel subsystem distributions (pie charts) related with main-channel glitches differ across clusters, indicating that specific auxiliary channels are correlated with glitch morphology.
- Further CAGMon analysis will enable quantification of both linear and non-linear correlations between each cluster and its associated auxiliary channels, enabling inference of physical noise sources.

Outlook

- Gravitational Wave Data Application**:** Automates morphology-based glitch classification without manual labeling, facilitating real-time noise diagnostics and identification of causal links in future observing runs.
- Potential Geophysical Application: Extends to environmental sensing, serving as a versatile noise hunting and anomaly detection tool for large-scale multi-sensor networks like seismic and geomagnetic arrays.

Acknowledgement

The authors would like to express their gratitude to Dr. Kihyun Jung for providing the spectrogram data used to develop the model in this study.

REFERENCES: [1] Xie, J., Girshick, R. & Farhadi, A.. (2016). Unsupervised Deep Embedding for Clustering Analysis. *Proceedings of The 33rd International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 48:478-487

[2] Piljong Jung *et al.* , Identifying multichannel coherent couplings and causal relationships in gravitational wave detectors, Phys. Rev. D **106**, 042010 (2022). doi: 10.1103/PhysRevD.106.042010

[3] KAGRA collaboration. KAGRA: 2.5 generation interferometric gravitational wave detector. *Nat Astron* **3**, 35–40 (2019). doi: 10.1038/s41550-018-0658-y

[4] T. Akutsu *et al.*, Identification of Noise-Associated Glitches in KAGRA O3GK with Hierarchical Veto, *Progress of Theoretical and Experimental Physics*, Volume 2025, Issue 8, August 2025, 083F01, doi: /10.1093/ptep/ptaf093