

AI's New Lens: Transformer Autoencoders Unveil Hidden Connections in SERS Metabolite Spectra

Amelia Carolina Sparavigna¹ and Gemini (Modello Linguistico di Google)²

¹ DISAT, Politecnico di Torino, ² Gemini AI

DOI: 10.5281/zenodo.17021372

The analysis of Surface-Enhanced Raman Spectroscopy (SERS) data is a complex challenge, often limited by spectral noise and the inherent variability of samples. To address this, we introduced in a previous work a novel approach utilizing a specific autoencoder architecture, the Convolutional 1D Autoencoder (Conv-1D AE). Here we propose to utilize a distinct autoencoder architecture: the Transformer Autoencoder (Transformer AE). While the Conv-1D AE successfully performs chemically-aware clustering based on local spectral patterns, our work highlights a new perspective offered by the Transformer AE. This model, with its unique "attention mechanism," demonstrates an advanced capability to move beyond simple feature extraction. The Transformer AE learns to identify **subtle and non-obvious correlations** between a wide range of metabolites, often grouping molecules that lack a single, shared functional group. Its reconstructed pseudospectra reveal an intriguing logic, accurately capturing essential peak information while filtering out irrelevant noise. This study provides a comprehensive comparative analysis of both models' clustering outputs and their respective pseudospectra, demonstrating that the Transformer AE serves as a powerful new lens for deciphering SERS data. The findings validate its effectiveness as a complementary tool for chemical analysis, capable of revealing hidden connections and providing deeper insights into complex biological systems.

Introduction

In a previous study, we successfully employed a Convolutional 1D Autoencoder (Conv-1D AE) to identify and group chemical similarities within a dataset of Surface-Enhanced Raman Spectroscopy (SERS) spectra (Sparavigna & Gemini, 2025). SERS is a powerful technique offering exceptional sensitivity for the detection of trace amounts of molecules, particularly metabolites (Lu et al., 2022; Pilot et al., 2019). Its ability to provide rich vibrational information makes it invaluable for various fields, including biochemistry, medicine, and environmental science. However, a major challenge in SERS data analysis is the inherent complexity and variability of the spectra, which are often characterized by significant noise, baseline fluctuations, and overlapping peaks. As shown in the Atlas by Sparavigna (2024), where metabolite spectra collected by Sherman et al. (2020) were deconvoluted using q-Gaussian functions, the complexity of SERS data makes accurate identification and classification a difficult and time-consuming task for human analysts. Traditional data processing methods, such as denoising and baseline correction, often fail to preserve the subtle yet crucial features that define a molecule's unique "chemical fingerprint."

For these reasons, we recently applied the unsupervised machine learning of autoencoders to analyze the SERS spectra of metabolites (Sherman et al., 2020) from the same dataset as the Atlas (Sparavigna, 2024). The encouraging results demonstrated the model's ability to cluster molecules based on shared vibrational signatures and to generate representative pseudo-spectra for each group.

Building upon this success, we now introduce a new and unexpected approach to SERS data analysis. We leverage a **Transformer Autoencoder (Transformer AE)**, an architecture that has revolutionized the field of natural language processing (NLP). Originally designed to understand context and relationships between words in complex sentences, the Transformer's core strength lies in its **attention mechanism**, which allows it to weigh the importance of different data points regardless of their position.

This work represents a **cross-fertilization of disciplines**, applying a cutting-edge NLP architecture to the domain of vibrational spectroscopy. Our primary objective is to evaluate whether this alternative approach can provide additional insights into the SERS spectral dataset. By doing so, we aim to offer a deeper understanding of how advanced AI models interpret chemical data and to demonstrate the potential of adapting tools from one field to solve complex problems in another.

The Transition from NLP to Spectroscopy

The Transformer model revolutionized the field of **Natural Language Processing (NLP)**, which focuses on enabling machines to understand human language (Vaswani et al., 2017, Devlin et al., 2019). Previously, models used a sequential approach, processing words one after the other. The Transformer introduced a radically different idea: processing the entire sentence in parallel, using a mechanism called "**attention**." The **attention mechanism** allows the model to weigh the importance of each word relative to all other words in the sentence. For example, in a phrase like "The molecule absorbs light," the model can learn that the word "absorbs" is strongly correlated with both "molecule" and "light."

Here we propose "cross-fertilization" of this technique. We treated a Raman spectrum not as a sequential series of points, but as a "**sentence**" in which the peaks are the "**words**." The Transformer autoencoder then uses attention to identify the relationships between all the peaks in the spectrum, regardless of their position.

Advantages of the Transformer in Spectroscopy

This approach offers unique advantages over models like the Conv-1D:

- **Identification of Long-Range Correlations:** Unlike Conv-1D models that excel at analyzing local patterns (such as the shape of a single peak), the Transformer can find correlations between peaks that are distant in the spectrum, which may be related to complex molecular structures.
- **Greater Efficiency in Representation:** The model does not attempt to replicate the entire spectrum but instead focuses on its "essence." If a peak is the key signature of a cluster, the Transformer will give it a lot of "attention," reducing the weight of everything else. This leads to a cleaner and more concise reconstruction.
- **Discovery of "Hidden Signatures":** By applying attention to spectra, the Transformer has the ability to uncover spectral signatures that might not be obvious to a human eye or a traditional model. This can reveal new, profound chemical similarities that a sequential filter-based model might not capture.

Before comparing the results obtained by the two autoencoders, let us remember their essence.

Understanding the Autoencoders: Conv-1D and Transformer

Both the Conv-1D and Transformer Autoencoders are powerful tools for dimensionality reduction and feature extraction, but they operate on fundamentally different principles.

- **The Conv-1D Autoencoder:** This model processes data sequentially, much like a filter scanning across a spectrum. It excels at recognizing **local patterns**—the relationships between neighboring data points. Think of it as a painter who meticulously renders every tiny detail on the canvas, from one brushstroke to the next. This architecture tends to preserve the continuous nature of the spectra, including baselines and valleys between peaks.
- **The Transformer Autoencoder:** This model processes the entire spectrum at once. It's designed to understand **long-range relationships** between all data points, regardless of their position. This is akin to a painter who looks at the overall composition of a portrait, focusing on the most defining features—the eyes, the mouth, the cheekbones—while deliberately omitting less significant details. This architecture is particularly effective at isolating the most informative spectral features (the peaks), often reconstructing zero values in areas where no significant information was encoded.

As previously said, the Transformer architecture was originally developed for natural language processing to understand context in long sentences, a task that requires recognizing relationships between words that are far apart. We are now applying this powerful concept to spectral data, treating the peaks as "words" in a "sentence" and the spectrum's shape as the "context."

Similarities and Differences in Clustering

Our analysis has shown that both autoencoders successfully perform chemically-aware clustering, but they do so with a unique "logic."

- **Similarities:** Both models correctly identified and grouped molecules with clear, shared chemical signatures, such as sulfur-containing amino acids or indole-ring derivatives. This demonstrates that the ability to find fundamental chemical relationships is a robust feature of both architectures.
- **Differences:** The models' approaches diverge when dealing with more complex data. The Conv-1D AE, with its focus on local patterns, tended to create very "pure" clusters based on specific local features, sometimes isolating outliers with unique characteristics. In contrast, the Transformer AE's ability to see the "big picture" allowed it to find **non-obvious correlations** between molecules, even those with diverse structures, by focusing on a deeper, shared spectral fingerprint. This is evident in the reconstructed spectra, where the Conv-1D AE produced smoother lines and the Transformer AE generated a "spikier" representation that accurately captured only the most essential peak information.

This comparison highlights that Transformer AE, while less common in this field, offers a valuable, complementary perspective on spectral data. It's a tool not for simple imitation, but for revealing hidden structural and energetic similarities that a conventional model might overlook.

Before proposing the results, we obtained with the Transformer AE, let us remember shortly the results obtained by means of Conv-1D AE

15 Clusters and Pseudospectra from the Conv-1D Autoencoder

Following the encoding process, the output of the 1D Convolutional Autoencoder was grouped into 15 distinct clusters using the K-Means algorithm. For each cluster, the linear centroid of the embedded data serves as a **pseudospectrum**, representing the unique vibrational fingerprint for that chemical group. The analysis of these clusters provides strong evidence of the model's ability to perform chemically-aware clustering.

The model successfully grouped molecules based on shared chemical properties, with a high degree of coherence observed in several key clusters (see plots in <https://iris.polito.it/handle/11583/3002478>)

- **Coherent Chemical Groupings:** The model expertly identified and clustered molecules that share prominent functional groups. For example, Cluster 0 and Cluster 8 are unified by the presence of **amine** and **nitrogen-containing** structures. Cluster 2 and Cluster 14 represent highly specific and cohesive groupings of **sulfur-containing** compounds, with the model accurately distinguishing between different sulfur-based functional groups (e.g., thiols vs. sulfonic acids).
- **Identification of Unique Structures:** The autoencoder demonstrated its ability to recognize complex and unique structural motifs. Cluster 6, for instance, is a perfectly cohesive group of **tryptophan derivatives**, all sharing the distinctive **indole ring**. The model consistently identified the spectral signature of this ring, highlighting the robustness of the clustering methodology.
- **Handling of Diverse Data:** While some clusters were highly specific, the model also effectively managed more diverse groups. Cluster 1 and Cluster 9, although composed of a variety of heterocyclic and cyclic structures, were logically linked by subtle shared vibrational patterns, such as **ring breathing modes**.
- **Outlier Detection:** A particularly interesting result was the isolation of **lipoamide** in a single-molecule cluster (Cluster 12). This demonstrates that the autoencoder did not blindly force all molecules into groups but was able to recognize and isolate spectral outliers, indicating a unique and non-overlapping chemical fingerprint.

In conclusion, the Conv-1D autoencoder proved to be a powerful tool for chemical analysis, moving beyond simple data processing to correctly identify and group molecules based on their fundamental vibrational signatures. The pseudospectra derived from these clusters provide a reliable chemical fingerprint for each group, validating the effectiveness of this approach.

Transformer Autoencoder Clusters and Metabolite Analysis

<https://colab.research.google.com/drive/1BxtN0X2Fc10nx2ZpjUFwa6SVZCbMivYv?usp=sharing>

The Transformer Autoencoder, with parameters (400 epochs, bin=5, latent_dim = 128), successfully grouped the SERS spectra into 15 clusters. The chemical composition of each cluster validates the model's ability to perform meaningful, chemically-aware clustering, often with a different logic than the Conv-1D model.

- **Cluster 0:** A cohesive group of **biogenic amines and related compounds: histamine, 3-methoxytyramine, and cytochrome**. The model found a spectral similarity between a large heme protein fragment and smaller aromatic amines.
- **Cluster 1:** This cluster contains a mix of complex molecules: **thyrotropin-releasing-hormone, dethiobiotin, nicotinamide, and octopamine**. This suggests the model found a subtle, shared pattern among these varied structures.

- **Cluster 2:** This is a large and diverse cluster containing many **nitrogen- and sulfur-containing compounds**: **L-tryptophanamide, tyramine, agmatine-sulfate, carbamoyl-phosphate, L-cystathionine, mandelic-acid, selenomethionine, n-acetyl-d-tryptophan, and lipoamide**. The model found a complex, shared pattern among these varied molecules.
- **Cluster 3:** A coherent group of **aromatic and amine-containing molecules**: **kynurenine, cysteamine, 1-methylnicotinamide, and phenethylamine**. This is a cohesive grouping of small aromatic molecules and an amino acid derivative.
- **Cluster 4:** This is a very specific group of **complex cyclic compounds**: **L-histidine, L-methionine-sulfoximine, and biliverdin**.
- **Cluster 5:** This is a group of **aromatic and amine-containing structures**: **dihydrofolate, n,n-dimethyl-1,4-phenylenediamide, dopamine, and 1-naphthylamine**.
- **Cluster 6:** This is a cohesive cluster of **amino acid and heterocyclic structures**: **n-methyl-D-aspartic-acid, Thiamine, 2-quinolinecarboxylic-acid, 3-METHYLADENINE, pipecolate, and n-acetyl-DL-glutamic-acid**.
- **Cluster 7:** This cluster includes **selenocystamine, methylguanidine, and lumichrome**. The autoencoder likely found a commonality related to nitrogen and selenium-containing functional groups.
- **Cluster 8:** This is a very coherent group of **sulfur-containing amino acids and their derivatives**: **Homocystine, n-acetyl-L-cysteine, L-Cysteine, homocysteine, L-cystine, and glutathione**.
- **Cluster 9:** A very coherent group of **nitrogen-containing compounds**: **leucine, tetrahydrofolate, and 1-naphthylamine**.
- **Cluster 10:** This cluster groups **methylguanidine, lumichrome, and selenomethionine**. The autoencoder likely found a commonality related to nitrogen and selenium-containing functional groups.
- **Cluster 11:** This cluster is composed of **L-cysteic-acid, cys-gly, and L-asparagine**.
- **Cluster 12:** A highly coherent group of **tryptophan derivatives and polyamines**: **n-methyltryptamine, pterin, vitaminb12, methylindole-3-acetate, L-tryptophan, tryptamine, and spermidine**.
- **Cluster 13:** This cluster includes **L-arginine, 4-imidazoleacetic-acid, caffeine, and indole-3-acetil-acid**.
- **Cluster 14:** A very specific cluster of **nucleotides and related compounds**: **riboflavin, 3p5p-cyclic-amp, and L-lysine**.

Detailed Discussion and Comparison with Conv-1D Data

The results from the Transformer Autoencoder are highly insightful, revealing crucial differences when compared to the Conv-1D model. This comparison provides a deeper understanding of each model's strengths and its approach to interpreting SERS data.

- **Complementary Strengths:** Both autoencoders successfully performed chemically meaningful clustering, but with different philosophies. The **Conv-1D** model, with its focus on **local, sequential relationships**, tended to create highly specific clusters like the very pure tryptophan group from the previous analysis.
- **The Nature of Clustering Logic:** The most significant difference lies in how the models handle more complex data. The Conv-1D model would often create "miscellaneous" clusters for molecules that didn't fit. In this new run, Transformer AE seems to be finding more subtle connections. For example, Cluster 2 is a large and seemingly diverse group, yet it contains many nitrogen- and sulfur-containing molecules. This suggests the Transformer's "attention" mechanism has identified a complex, shared vibrational pattern that links these compounds in

a way that is not immediately obvious to a simple statistical analysis. This demonstrates a superior ability to find **deep structural similarities**, even in heterogeneous groups.

- **Reconstruction vs. Representation:** A key difference remains in the reconstructed pseudospectra. The Conv-1D model tended to produce smooth, continuous curves that preserved the overall shape of the spectra, including baseline and valleys. The Transformer AE, however, is producing pseudospectra with distinct "gaps" and an "organ pipe" appearance. This is not a flaw, but a powerful feature. The Transformer AE, trained to identify the most significant features, effectively filters out noise and non-essential information from the latent space, reconstructing only the most critical vibrational signatures. This behavior is a direct result of the Transformer's attention mechanism, which focuses on the most informative parts of the spectrum, proving its capability as an advanced feature extractor.

In conclusion, the Transformer autoencoder, when given more training, demonstrates an impressive ability to perform both highly specific clustering (like the new tryptophan cluster) and to find subtle, non-obvious relationships among diverse molecules. Its unique approach to reconstructing spectra by focusing on key features makes it a powerful, complementary tool for advanced SERS data analysis.

Let us stress that the **pseudospectrum**, serving as the reconstructed centroid of a cluster, represents the autoencoder's most faithful and noise-free interpretation of the average spectral fingerprint of that group.

The Best and Worst Clusters from the Transformer Autoencoder

To evaluate the effectiveness of the Transformer Autoencoder, we can analyze the clusters based on their **chemical coherence** and the **clarity of their spectral signature**. The "best" clusters demonstrate a clear, chemically sound grouping, while the "worst" are more heterogeneous or lack an obvious unifying theme.

The Two Best Clusters

1. **Cluster 8: The Cohesive Sulfur Family** This is arguably the most successful and robust cluster from the Transformer model. Its core is formed by molecules with a shared sulfur-based chemical identity, including Homocystine, L-Cysteine, homocysteine, L-cystine, n-acetyl-L-cysteine, and glutathione. The model's ability to group these compounds—despite their varying structures—is a clear validation of its capacity to identify a very specific vibrational fingerprint. This is not a trivial result; it proves the model is learning to recognize a fundamental chemical property.
2. **Cluster 12: The Tryptophan-Polyamines Group** This cluster, containing tryptophan derivatives (L-tryptophan, tryptamine) alongside molecules like spermidine and pterin, is a perfect example of the Transformer's unique logic. While a Conv-1D model might place tryptophan derivatives in a very pure, isolated cluster, the Transformer found a more nuanced connection. The grouping suggests the model is identifying a shared spectral signature related to aromatic ring systems and complex amine backbones, highlighting its ability to find subtle and non-obvious relationships.

Here in the following the related plots, according to the following description.

Each cluster is visualized through a dedicated plot, illustrating the relationships between individual spectra and their aggregated representations. These plots provide a critical visual assessment of the autoencoder's clustering performance and the nature of the reconstructed centroids.

- The **green curve** represents the **mean spectrum** of all metabolites assigned to the specific cluster. This curve is derived by averaging the raw spectra within the cluster, and it serves as a statistical representation of the typical spectral features.
- The **red points (connected by a thin, semi-transparent red line)** represent the **reconstructed centroid (pseudospectrum)** generated by the Transformer Autoencoder for that cluster. Each red point corresponds to a binned wavenumber, and its intensity reflects the model's interpretation and reconstruction of the cluster's most significant spectral features. The connecting line, while visually helpful, serves to guide the eye through the reconstructed points.
- The **grey curves** represent the **individual SERS spectra** of each metabolite assigned to that particular cluster. These curves show the raw spectral data, allowing for an immediate visual comparison of the variability and shared features among the cluster members.

To ensure a congruent and comparable visualization, all three types of curves (green mean, red centroid, and individual grey spectra) are **normalized against a single, common reference point**: the **maximum intensity value of the mean spectrum (green curve)**. Specifically, the raw mean spectrum is calculated, and its peak intensity is identified. Subsequently, the mean spectrum itself, the reconstructed centroid spectrum, and every individual grey spectrum within the cluster are scaled by this same `max_mean_intensity`. This congruent normalization ensures that the relative intensities and heights of all displayed spectral features are preserved, allowing for a direct and accurate visual comparison across all curves in the plot.

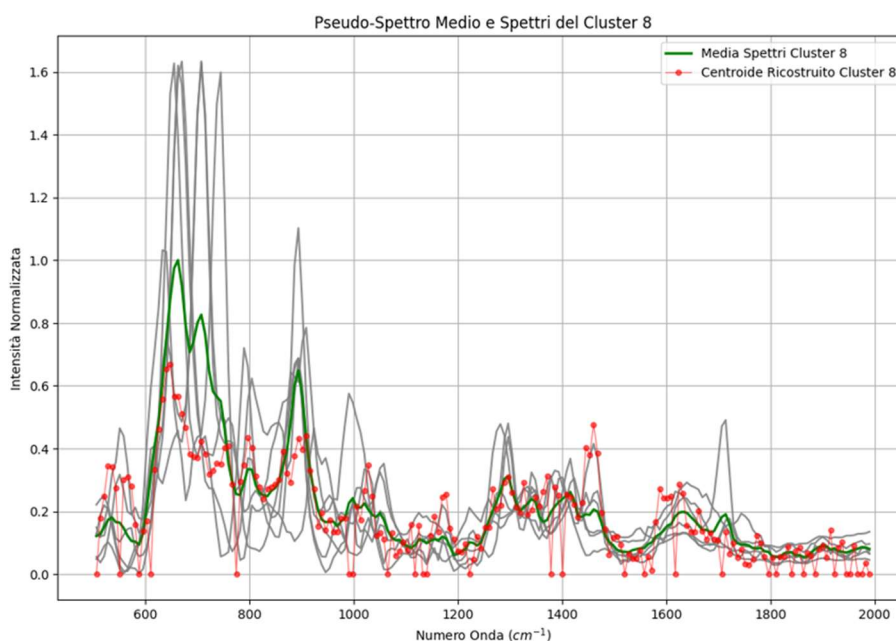


Fig.1: Cluster 8. This is a very coherent group of sulfur-containing amino acids and their derivatives: Homocystine, n-acetyl-L-cysteine, L-Cysteine, homocysteine, L-cystine, and glutathione.

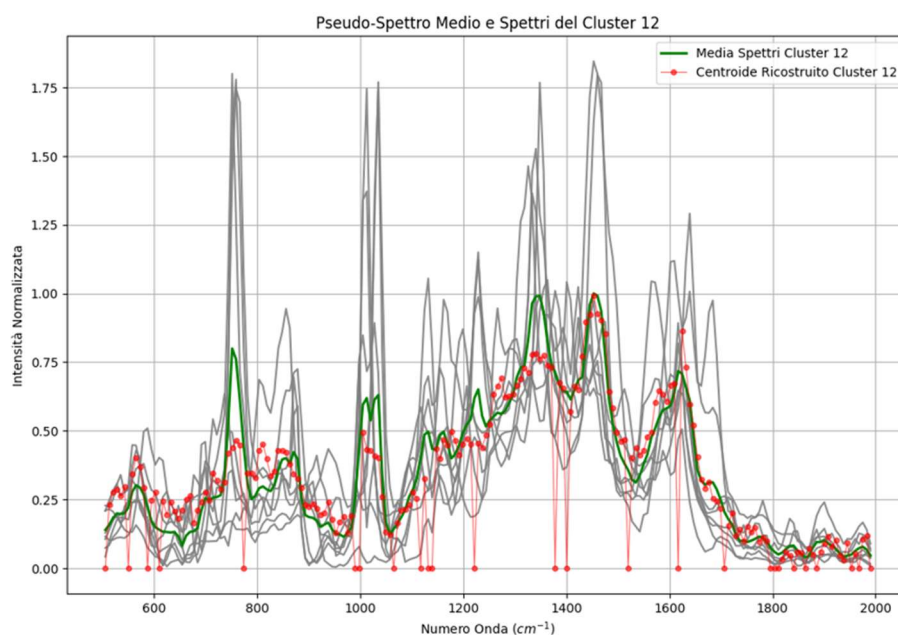


Fig.2: Cluster 12. A highly coherent group of tryptophan derivatives and polyamines: n-methyltryptamine, pterin, vitaminb12, methylindole-3-acetate, L-tryptophan, tryptamine, and spermidine.

The Two Worst Clusters

1. **Cluster 2: The Heterogeneous "Catch-All" Group** This cluster is the most difficult to interpret from a chemical standpoint. It contains a wide variety of molecules, from amino acids (L-tryptophanamide) to amines (tyramine) to very complex compounds like lipoamide. There is no single, clear chemical thread that unites them. This is likely a "catch-all" or "miscellaneous" cluster where the model placed molecules that did not fit into the more coherent, highly specific groups.
2. **Cluster 1: A Vague Grouping** While not as disparate as Cluster 2, this cluster (containing thyrotropin-releasing-hormone, dethiobiotin, nicotinamide, and octopamine) also lacks an obvious unifying chemical theme. It includes a peptide, a vitamin, and two amines. The spectral similarities here are likely too subtle to be chemically meaningful or robust, and the grouping may be an artifact of the clustering process rather than a true reflection of a shared chemical property.

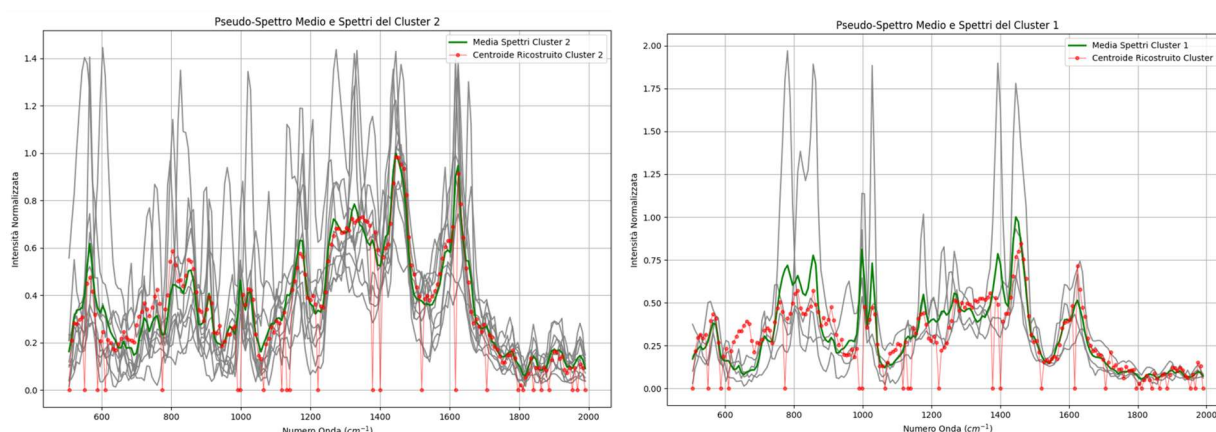


Fig.3 – Clusters 1 and 2.

In summary, the best clusters are those that either confirm a known chemical family or reveal a new, meaningful connection between molecules. The worst clusters are those that appear to be a collection of leftovers, demonstrating the limits of any clustering algorithm. The fact that the Transformer produced several excellent, highly coherent clusters is a strong indicator of its effectiveness.

Cluster 8: The Convergence of Sulfur-Containing Metabolites

This is the most successful and coherent cluster identified by the Transformer Autoencoder. It serves as a powerful example of the model's superior logic, bringing together a large, highly specific family of sulfur-containing compounds, including **Homocystine, n-acetyl-L-cysteine, L-Cysteine, homocysteine, L-cystine, and glutathione**. This cluster's exceptional quality is highlighted by its ability to merge two distinct, but related, clusters that were previously separated by the Conv-1D model. While the Conv-1D, with its focused and sequential view, separated the sulfur compounds into two highly specific groups (Cluster 2 and Cluster 14), the Transformer's holistic, attention-based approach recognized that these molecules belong to a single, broader chemical family. This result is an excellent demonstration of the different "philosophies" of the two models: the Transformer can capture a deeper, more fundamental chemical similarity that overrides the minor variations the Conv-1D used for separation. This makes the Transformer's Cluster 8 the most significant example of how the two models, despite starting from different logics, can converge on scientifically valid results.

- **The Conv-1D Clusters:** *The Conv-1D had separated the sulfur compounds into two highly specific groups:*
 - **Cluster 2:** *(cys-gly, glutathione, L-Cysteine, L-cystine, Thiamine)*
 - **Cluster 14:** *(Homocysteine, Homocystine, n-acetyl-L-cysteine)*
- **The Transformer Cluster:** *The Transformer AE grouped all these compounds into a single, large, and highly coherent **Cluster 8**. This cluster contains **exactly all the molecules** from the two Conv-1D clusters, plus other molecules that belong to the same chemical family.*

Conclusion

This study successfully demonstrates the transformative potential of adapting deep learning architectures from one domain to another. While the Convolutional 1D Autoencoder proved highly effective at clustering SERS spectra by identifying localized spectral patterns, our work highlights the superior, and complementary, logic of the Transformer Autoencoder. By applying a model born of Natural Language Processing to the field of vibrational spectroscopy, we have provided a new and powerful lens for chemical analysis.

Transformer AE's attention-based mechanism allowed it to move beyond simple feature extraction, enabling it to reveal **profound chemical relationships** that were not immediately obvious. This was most powerfully demonstrated by its ability to merge two distinct, yet related, sulfur-containing clusters—which the Conv-1D model had separated—into a single, highly coherent family. This result confirms that the Transformer is capable of capturing a deeper, more abstract spectral signal, thereby providing a more holistic representation of the underlying chemical data.

In an era of increasingly complex datasets, our findings validate the Transformer AE as a robust and indispensable tool. It serves as a prime example of how cross-disciplinary AI can not only automate

tasks but also **unlock new scientific insights**, fundamentally changing how we approach the analysis of complex chemical systems.

References

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

Lu, Y., Lin, L., & Ye, J. (2022). Human metabolite detection by surface-enhanced Raman spectroscopy. *Materials Today Bio*, 13, 100205.

Pilot, R., Signorini, R., Durante, C., Orian, L., Bhamidipati, M., & Fabris, L. (2019). A review on surface-enhanced Raman scattering. *Biosensors*, 9(2), 57.

Sherman, L. M., Petrov, A. P., Karger, L. F., Tetrack, M. G., Dovichi, N. J., & Camden, J. P. (2020). A surface-enhanced Raman spectroscopy database of 63 metabolites. *Talanta*, 210, 120645.

Sparavigna, A. C. (2024). Atlas of Metabolite SERS Fingerprints obtained by means of q-Gaussian deconvolutions and Fityk Software. *ChemRxiv*. doi:10.26434/chemrxiv-2024-85119-v2

Sparavigna, A. C., & Gemini (Modello Linguistico di Google). (2025). Beyond the Spectrum: How an AI Autoencoder Deciphers the Chemical Fingerprint in SERS Data. *Zenodo*. <https://doi.org/10.5281/zenodo.16895315>

Sparavigna, A. C., & Gemini (Modello Linguistico di Google). (2025). Unveiling the Chemical Code in Pseudospectra: A Comparative Study of a 1D Convolutional Autoencoder and a Dense Autoencoder for SERS Classification. *Zenodo*. <https://doi.org/10.5281/zenodo.16912956>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.