ACCELERON AEROSPACE JOURNAL (AAJ)
E-ISSN: 2583-9942 (Online) | Accessible at www.acceleron.org.in
Volume 6, Issue 1, (pp.1677-1681)
Article ID: AAJ.11.2106-2604
Short Communications
https://doi.org/10.61359/11.2106-2604

How SETI Uses Machine Learning for Analyzing Radio Signals

Aakaash A*

Department of Computer Science, Vellore Institute of Technology, Kelambakkam - Vandalur Rd, Rajan Nagar, Chennai, Tamil Nadu 600127.

Abstract: There are more stars in the universe than the number of grains of sand on Earth. And orbiting each of these are more worlds than we could ever visit, more chances than we could ever count. And yet, we have heard nothing - no messages, no signals, no signs. Just absolute silence, stretching across billions of years and light years alike. But as Carl Sagan himself says, "The absence of evidence is not evidence of absence". Maybe they don't know we're here. Maybe they're too far to reach us. Maybe they're choosing not to speak. Or even wilder - they've been speaking all along, and we haven't known how to listen. So, we built ears as wide as continents and tuned our them to the language of numbers - with structured signals, repeating patterns, the rhythms hidden in randomness, the fingerprints of thought etched in the static. And this brings us to SETI, the Search for Extraterrestrial Intelligence.

Table of Contents

1. Introduction

For decades, the SETI mission has looked to the stars with a sense of wonder, using radio telescopes to listen for any faint signal that might hint at intelligent life beyond Earth. In the early years, researchers relied on hand-crafted rules and traditional signal processing to scan through the vast amounts of data they collected. Although it was a painstaking and meticulous work, fueled by hope and curiosity, we were able to pull it off. But the times have changed. Today's observatories produce data at a staggering scale with petabytes every hour. The sheer volume, complexity, subtlety and ambiguity of these signals have outpaced what manual analysis and simple algorithms can handle. The old methods, while groundbreaking in their time, are no longer enough and this has made manual inspection and simple thresholding insufficient. In response, the SETI community has turned to a new kind of tool, named machine learning. These adaptive algorithms don't need every rule spelled out. Instead, they learn from data, uncover patterns, and flag anomalies that might otherwise be missed.

*Department of Computer Science, Vellore Institute of Technology, Kelambakkam - Vandalur Rd, Rajan Nagar, Chennai, Tamil Nadu 600127.
Corresponding Author: Aakaash A.
Article History: Received: 29-Jun-2025 || Revised: 31-Dec-2025 || Accepted: 30-Jan-2026 || Published Online: 30-March-2026.

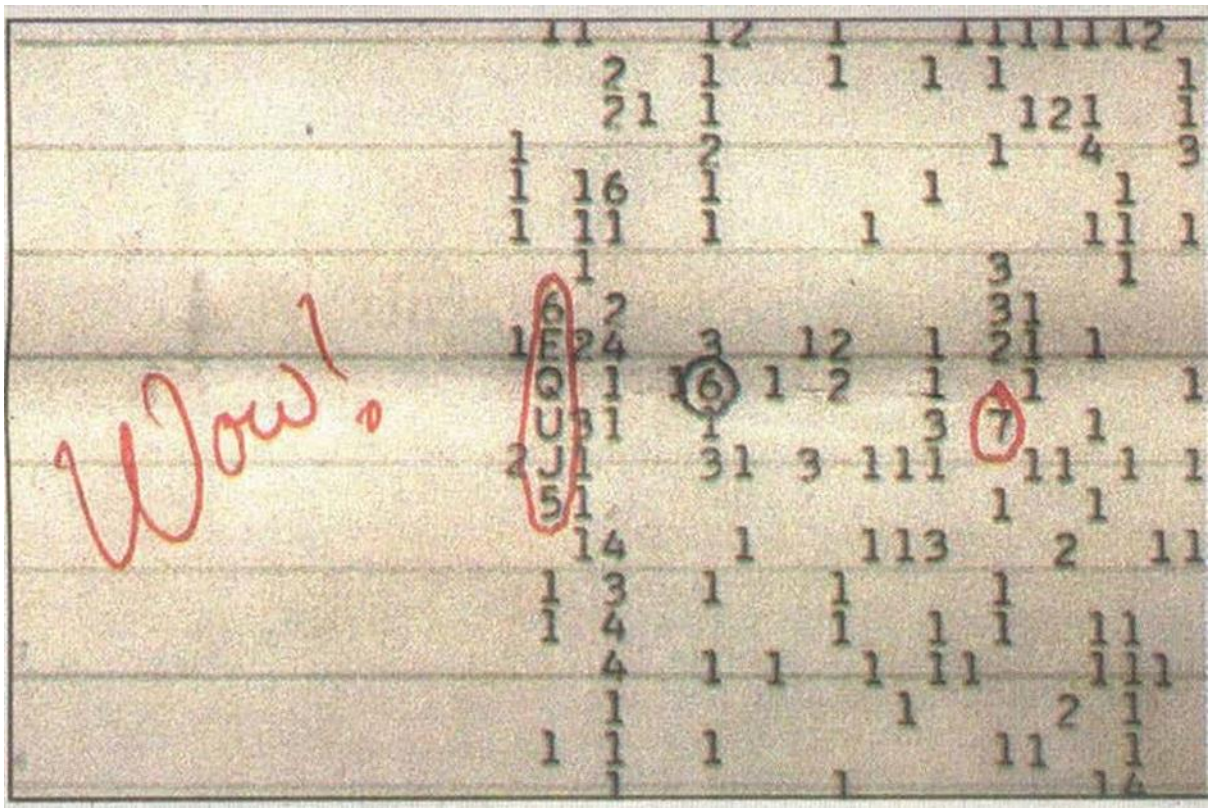

Figure-1. The famous 'Wow! signal' detected in 1977 by the Big Ear radio telescope - a 72 second burst of radio waves from deep space. Image credit: Big Ear Radio Observatory and North American Astrophysical Observatory (NAAPO).

2. The Need for Machine Learning (ML)

Modern radio telescopes collect huge amounts of data as they keep scanning billions of frequency channels and can gather up to a petabyte of information in a single day. For instance, the Breakthrough Listen initiative collects around 500 gigabytes of data every single hour. But this data not only includes signals from space, but also interference from satellites, planes, radio towers, or even microwave ovens. One major problem here is Radio Frequency Interference (RFI). These interferences can look like real signals or can even hide a real one completely. That makes it really hard to find possible signs of alien life, also known as techno signatures. On top of that, within that sea of information, the types of signals that SETI is interested in are rare, faint, and fleeting. Most of these useful signals are narrowband signals, meaning they occupy only a tiny slice of the frequency spectrum. Some exhibit Doppler drift, where the signal's frequency shifts slightly over time because of the motion of the source. And to make things worse, most of these signals last only for few milliseconds, and they have to be captured before they are lost forever. Traditional signal processing methods like thresholding or simple bandpass filters often miss these kinds of events entirely, or mistake Earth based interference for a real candidate. To put things in a nutshell, we are not just looking for a needle in a haystack but also looking for a very weird needle in a pile of hay, gravel and what not.

3. Data Preprocessing

Let's go through the procedure by which these signals. First, the information must be cleaned up and reshaped before machine learning models could do anything useful. That begins with getting the raw measurements and transforming them into something called waterfall plots and 2D spectrograms which are graphical plots that indicate how signal strength varies as a function of time and frequency. They resemble grayscale images where time runs along one axis, frequency runs along the other, and the brightness of each pixel indicates signal strength. Eventually these spectrograms become the pictures that machine learning models analyze.

Acceleron Aerospace Journal || AAJ.11.2106-2604
Volume 6, Issue 1, pp (1677-1681)
E-ISSN- 2583-9942
https://dx.doi.org/10.61359/11.2106-2604

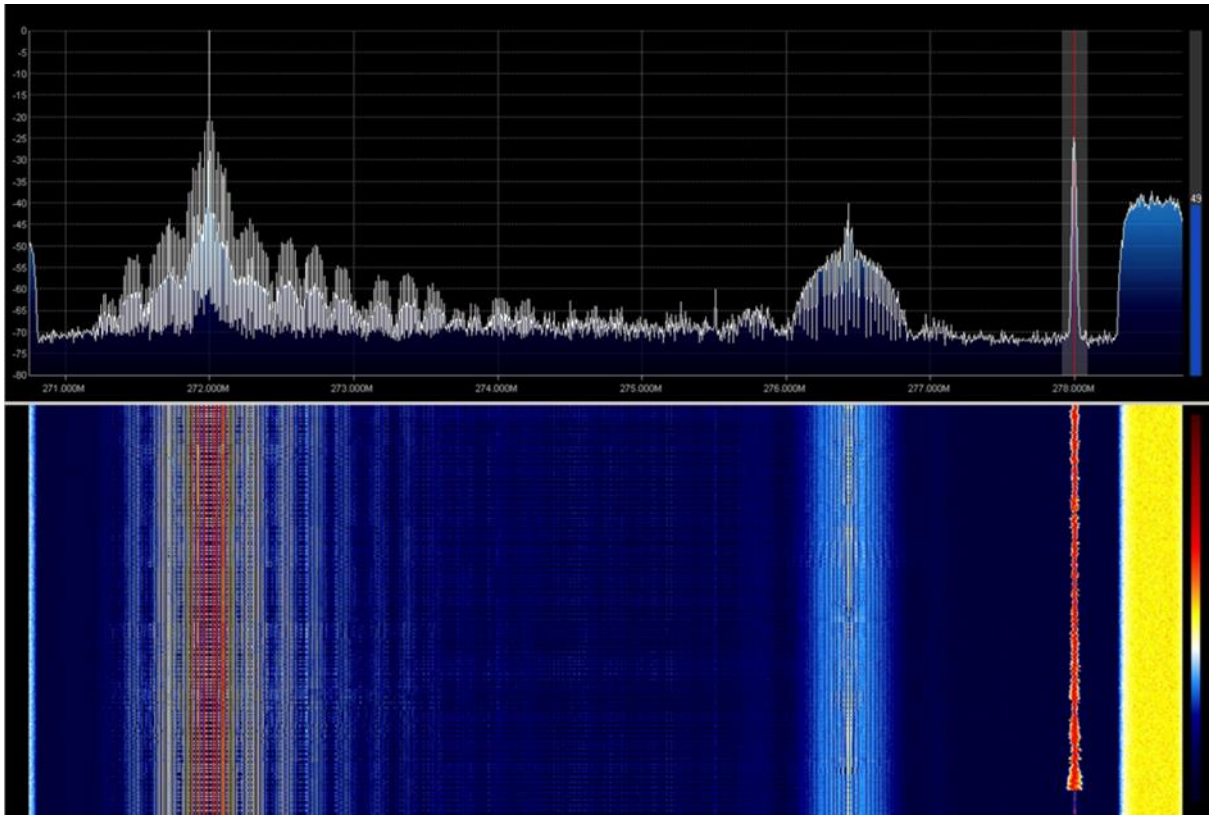

Figure-2. Spectrum above and waterfall diagram below of an 8MHz wide PAL-I Television signal

But before the models get to work, the spectrograms need to be cleaned. This is done through something called bandpass correction, which helps us to smoothen out uneven frequency sensitivities that emerge as a result of interferences and noise from the hardware. And in this stage, we flatten the background noise so that any potential signal stands out more clearly. After that comes the challenge of removing RFI, the radio frequency interference. One technique used is called DBSCAN, a clustering algorithm that can group together dense clusters of signal like features and isolate the ones that don't belong. The assumption is that RFI often repeats or clusters over time and direction, while a genuine extraterrestrial signal would likely appear in isolation, although this is an overgeneralized assumption.

4. Pattern Recognition

After cleaning and converting SETI's raw radio data into spectrograms, the subsequent challenge is instructing the machine to identify significant patterns. It all starts with feature extraction. These spectrograms, handled as images, are examined for several types of visual and statistical features that can be used to distinguish natural from artificial signals. One of the most important methods is the Gray Level Co-occurrence Matrix (GLCM), which derives the texture of a picture by quantifying the way pixel intensities correspond to each other. Texture has the ability to show minute patterns of modulation which suggest engineered origins. When a signal appears as a straight line, possibly the signature of a Doppler-shifted narrowband transmission, the Hough Transform is employed in order to find linear features in the image, typically the sign of something that is worth looking into. Alongside, statistical descriptors like mean brightness, variance, and symmetry summarize broad signal properties, while fine grained pixel level texture descriptions by Local Binary Patterns (LBP) allow the model to differentiate between smooth cosmic background noise and sudden, artificial edges.

These feature extractions create high-dimensional representations of data that are not easy to directly interpret, and hence dimensionality reduction methods take center stage. Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) project high-dimensional data into lower-dimensional spaces where the structure and shape of the dataset are easier to see. Clustering algorithms such as K-Means, in the lower space, recognize groups of similar signals. Anomaly detection algorithms such as Local Outlier Factor (LOF) mark those data points that are not consistent with known patterns. These outliers, which are different in structure or behavior are usually the most fascinating, since they might be new phenomena or even new kinds of signals that have never been observed before. With this multi-layered process, machine learning allows SETI not just to categorize established types of signals but also highlight the unusual and unexpected, where the most surprising finds might be.

5. Deep Learning

Once the system knows what to look for, deep learning models - specifically convolutional neural networks or CNNs - can take over. CNNs have been hugely successful in image recognition tasks, so it makes sense to apply them to spectrograms. Back in 2017, SETI researchers held a public machine learning challenge where participants trained CNNs on labeled spectrograms with identified signal types, such as simulated representations of what an extraterrestrial message could look like. Some of the models achieved up to 95% accuracy, illustrating just how capable these tools are once they're given the correct data.

But there was an issue - when scientists attempted to use pre-existing CNN architectures such as VGG16, AlexNet, or GoogLeNet, the outcome was not always good. These models had been trained on natural images, not spectrograms. Some of them performed poorly with hardly 14% accuracy in most cases because they were too sophisticated and had overfit the data. The best results came from custom-built models, carefully designed to match the structure and scale of SETI's unique data.

But of course, not every signal fits neatly into a pre-defined category. In fact, some of the most interesting ones are the ones that don't. This is where unsupervised learning becomes invaluable. Autoencoders, for example, are a type of neural network that learn to compress and then reconstruct their input data. If a spectrogram can't be reconstructed well, that's a sign that the input was unusual or unexpected. These anomalies are exactly what SETI wants to investigate.

6. Real-time Detection

Timing is everything in the world of SETI. Signals are fleeting, occurring in a flash once and then never again. So having the ability to capture and analyze them in the moment isn't a nicety - it's a necessity. To achieve this, SETI researchers employ streaming platforms such as Apache Kafka to handle the stream of data. Inference engines such as TensorRT and ONNX make models execute quickly on GPUs, enabling near-instant analysis. Signals are ranked in accordance with a set of criteria, such as their signal-to-noise, how surprising they are, and how confident the model is in its classification. The strongest signals are passed directly to human operators through dashboards to be scrutinized by astronomers. The combined approach ensures that any new signals are not missed simply because they are in the context of a backlog.

Breakthrough Listen has applied many ideas and placed them in a successful framework. Their processing pipeline combines supervised and unsupervised methods. Convolutional neural networks trained on real and synthetic data help identify known types of signals, and autoencoders find anomalies that do not fit learned patterns. How effective their results are is clear. Their system can automatically reject more than 90 percent of false positives, drastically reducing the number of signals that will have to be examined by human beings. In addition, by incorporating artificial signals like technosignatures into the dataset, they prime the system to remain aware of the kind of signals scientists anticipate discovering in the future. This is a powerful demonstration of the advancements in the search for SETI, from passive listening to the creation of systems that are able to actively shed light on potential discoveries.

7. Conclusions

In the end, nobody really knows what a message from another world would look like. It might not be anything like we expect. It might be buried in noise, or so faint it barely exists. For a long time, we have been trying to find it using rules we wrote ourselves, hoping that we'd spot something unusual. Now, the mission is the same, but the methodology has changed, and we have got tools that can look deeper, faster, and more openly in seconds than we ever could on our own in our entire lifetime. Machine learning isn't magic and can't guarantee we'll find anything. But it gives us a better shot. It lets us notice the weird, the rare, the things that don't fit. And in a search like this, that's what matters most. We're still listening. We're still waiting. But now, we might finally be ready to hear something - if there's anything out there to hear.

Acceleron Aerospace Journal || AAJ.11.2106-2604
Volume 6, Issue 1, pp (1677-1681)
E-ISSN- 2583-9942
https://dx.doi.org/10.61359/11.2106-2604

8. References

[1] SETI Institute. SETI Institute. (n.d.). SETI research overview. https://www.seti.org
[2] Breakthrough Listen.Worden, S. P., Drew, J., Siemion, A., Werthimer, D., et al. (2017). Breakthrough Listen—A new search for life in the universe. Acta Astronautica, 139, 98–101. https://doi.org/10.1016/j.actaastro.2017.06.008
[3] Enriquez, J. E., Siemion, A. P. V., Foster, G., et al. (2017). The Breakthrough Listen search for intelligent life: 1.1–1.9 GHz observations of 692 nearby stars. The Astrophysical Journal, 849(2), 104. https://doi.org/10.3847/1538-4357/aa8d1a
[4] Zhang, Y. G., Siemion, A. P. V., Foster, G., et al. (2020). Fast radio burst 121102 pulse detection and periodicity: A machine learning approach. The Astrophysical Journal, 891(2), 174. https://doi.org/10.3847/1538-4357/ab6a9a
[5] NASA. (2023). Technosignatures: The search for signs of technology in the universe. https://www.nasa.gov/technosignatures
[6] Margot, J.-L., Benford, J., Benford, G., et al. (2019). SETI and technosignatures: A report from the NASA technosignatures workshop. Acta Astronautica, 168, 1–13. https://doi.org/10.1016/j.actaastro.2019.11.013
[7] Radio Frequency Interference. Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. (2012). A morphological algorithm for improving radio-frequency interference detection. Astronomy & Astrophysics, 539, A95. https://doi.org/10.1051/0004-6361/201118497
[8] Akeret, J., Chang, C., Lucchi, A., & Refregier, A. (2017). Radio frequency interference mitigation using deep convolutional neural networks. Astronomy and Computing, 18, 35–39. https://doi.org/10.1016/j.ascom.2016.12.002
[9] Convolutional Neural Networks. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
[10] Spectrogram. Cohen, L. (1995). Time-frequency analysis. Prentice Hall.
[11] DBSCAN. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of KDD (pp. 226–231).
[12] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202
[13] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507. https://doi.org/10.1126/science.1127647
[14] Duda, R. O., & Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM, 15(1), 11–15. https://doi.org/10.1145/361237.361242
[15] Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1), 51–59. https://doi.org/10.1016/0031-3203(95)00067-4
[16] Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314
[17] Kreps, J., Narkhede, N., & Rao, J. (2011). Kafka: A distributed messaging system for log processing. Proceedings of NetDB.
[18] TensorRT. NVIDIA. (2023). TensorRT developer guide. https://docs.nvidia.com/deeplearning/tensorrt
[19] ONNX. Bai, J., Lu, F., Zhang, K., et al. (2019). ONNX: Open neural network exchange. https://onnx.ai

9. Conflict of Interest

The author declares no competing conflict of interest.

10. Funding

No funding was issued for this research.