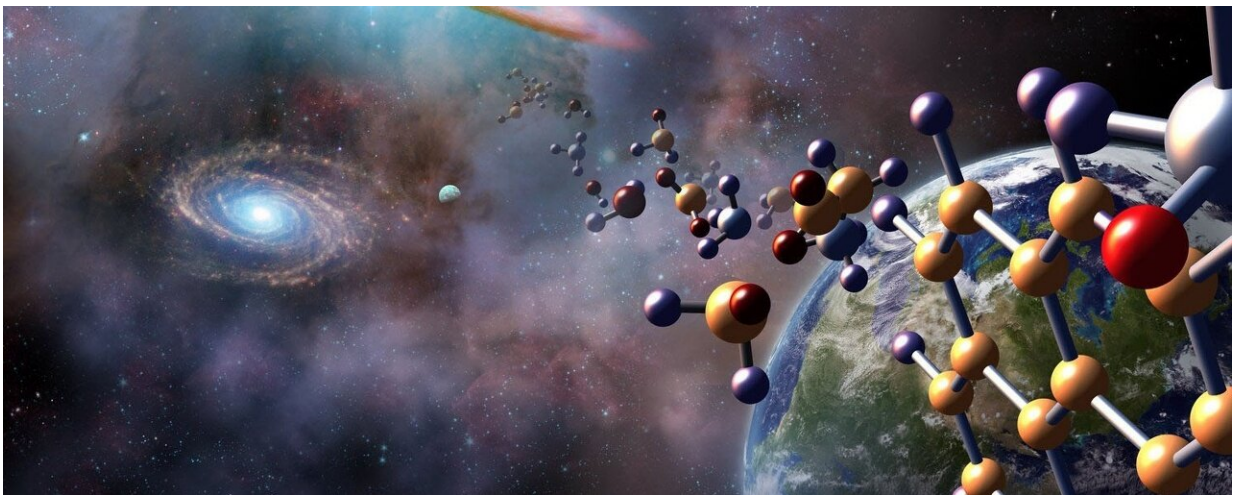# Researchers suggest the most compelling places to search for life will look like 'anomalies'

August 23 2023, by Matt Williams

Will it be possible someday for astrobiologists to search for life "as we don't know it"? Credit: NASA/Jenny Mottar

In the past two and a half years, two next-generation telescopes have been sent to space: NASA's James Webb Space Telescope (JWST) and the ESA's Euclid Observatory. Before the decade is over, they will be joined by NASA's Nancy Grace Roman Space Telescope (RST), Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx), and the ESA's PLAnetary Transits and Oscillations of stars (PLATO) and ARIEL telescopes.

These observatories will rely on advanced optics and instruments to aid in the search and characterization of exoplanets with the ultimate goal of finding habitable planets.

Along with still operational missions, these observatories will gather massive volumes of high-resolution spectroscopic data.

Sorting through this data will require cutting-edge machine-learning techniques to look for indications of life and biological processes (aka. biosignatures).

In a recent paper, a team of scientists from the Institute for Fundamental Theory at the University of Florida (UF-IFL) recommended that future surveys use machine learning to look for anomalies in the spectra, which could reveal unusual chemical signatures and unknown biosignatures.

The study was conducted by a mix of physicists and machine learning experts, including Associate Professor Katia Matcheva, physics graduate student Roy T. Forestano, Professor Konstantin T. Matchev, and Ph.D. student Eyup B. Unlu.

A preprint of their paper, "Searching for Novel Chemistry in Exoplanetary Atmospheres using Machine Learning for Anomaly Detection," appeared online on the pre-print server *arXiv* and is being reviewed for publication in the *Astrophysical Journal*. As they explained, the central premise of their paper is that what constitutes "life" remains an open question for scientists, and it would be advantageous to expand the scope of our search.

First off, it is important to acknowledge how far the study of exoplanets has come in recent decades. The first confirmed detection did not take place until 1992, which consisted of two Super-Earths (Poltergeist and Phobitor) observed around a pulsar (PSR B1257+12, aka. Lich) located

2,300 light-years from Earth. While scientists firmly believed that most stars had their own system of planets, they had no incontrovertible proof before this discovery. And until the Kepler Space Telescope launched in 2009, exoplanet discoveries were being added at a rate of a few per year.

Since then, a total of 5,496 exoplanets have been confirmed in 4,096 systems, with another 9,820 candidates awaiting confirmation. In recent years, the process has shifted from the process of discovery towards characterization, where improved instruments and methods have enabled astronomers to analyze exoplanet atmospheres directly to measure their potential habitability. As Prof. Matcheva explained to Universe Today via email:

"The instruments are getting better and better: better spectral resolution, exceptional signal-to-noise level, wider wavelength coverage. In addition to JWST, which has returned some exceptional spectroscopic observations of several exoplanets, ESA is planning a dedicated exoplanet space telescope ARIEL that will observe 1000 planets. Analyzing this data will keep scientists busy for a long time."

According to Matcheva, the fields of exoplanet studies and astrobiology are incredibly fascinating because of the sheer potential involved. Currently, the field is largely concerned with constraining "habitability" through the targeted search for biosignatures: evidence of life and organic processes.

Using Earth as a template, the only planet where we know life exists, the most highly-sought biosignatures include nitrogen gas (N_2), oxygen gas (O_2), carbon dioxide (CO_2), methane (CH_4), ammonia (NH_3), and water (H_2O).

This constitutes the "low-hanging fruit approach," where scientists are looking for life that conforms to terrestrial standards. This is not an

accident, nor is it a lazy approach. It's simply because it is exceedingly difficult to search for signs of life that we are completely unfamiliar with. But this also presents an opportunity to contemplate the possibilities and expand the range of what we know.

"Do we know what to search for?" Matcheva added. "Do we know where to search? Would we recognize it if we saw it? The exoplanet science community always works with these questions in mind."

For their study, Matcheva and her colleagues investigated how machine learning could be trained to look for "anomalies" in transit spectra. This refers to light curves obtained by observing distant stars for periodic dips in luminosity, which could indicate the presence of a planet passing in front of the star relative to the observer. This is known as Transit Spectroscopy (or the Transit Method), which remains the most effective and widely-used method for detecting exoplanets. In addition to detection, this method allows astronomers to occasionally observe light passing through the planet's atmosphere.

When measured with a spectrometer, these observations will reveal data on the atmosphere's chemical composition, which could include telltale biosignatures. In the coming years, the combination of next-generation telescopes and machine learning (ML) will allow astronomers to more accurately determine the potential habitability of exoplanets. "We believe that ML methods in astrophysics can be a game changer in how we process data in terms of speed, volume, and methodology, said Matcheva. "And we see that across all fields of science."

For their purposes, Matcheva and her team used two popular anomaly-detection machine learning methods—Local Outlier Factor (LOF) and One-Class Support Vector Machine (OCSVM) to analyze a large public database of synthetic spectra.

This database was developed by the ESA ARIEL science team in anticipation of the mission (scheduled to launch in 2029) and contains more than 100,000 computer-generated spectra signals of exoplanets. The team also used Receiver Operating Characteristic (ROC) curves to quantify and compare the performance of the two ML techniques. The process and results, as Matcheva related, were both fascinating:

"The spectra are calculated with current models, assuming that the atmosphere of each planet is a mixture of five different gases in different proportions. As an experiment, we treated one of the absorbers (for example, H_2O) as a 'mystery' absorber. We trained the ML algorithm on a subset of the data that is deficient in H_2O and tested if it will correctly flag planets with water as anomalous."

"We repeated the experiment for four of the gases. We used both LOF and OCSVM. Both methods did an outstanding job in finding the anomalous planets when no noise or very little noise (~10 ppm) is present, even for very small amounts of the 'mystery' gas. Unsurprisingly, the ML model starts making mistakes when the noise level increases too much."

As Matcheva indicated, their paper demonstrated that LOF and OCSVM methods are very robust, even in the presence of signal noise. These results offer a taste of what could be possible in the near future, where literally thousands of exoplanets can be analyzed rapidly and systematically using ML methods to identify anomalous planets for follow-up investigations. These examinations will likely be very educational, given that inconsistencies between theoretical models and observations are often how the most exciting discoveries are made.

"Although looking for biosignatures was not a primary goal of this paper, it is a very interesting outcome, and we are very excited about the potential of the method," said Matcheva. "Looking for signatures of life

in the universe is more like looking for a needle in a haystack rather than for a smoking gun. It is actually even more challenging because we do not know what the needle looks like. The novelty detection methods are designed exactly for that: rare events [where] we do not know what they look, smell, or sound like."

As noted earlier, the search for extraterrestrial life—and indeed, the search for extraterrestrial intelligence (SETI)—can be summarized as searching for life "as we know it." But if life is very rare in the universe or very "exotic" in nature (meaning that it can arise from all sorts of chemicals and conditions), then it makes sense to cast a wider net. After all, if our frame of reference is an impediment to our astrobiology efforts (one could certainly argue as much), expanding it could be the difference between finding evidence that we are not alone and leaving the question unanswered for another generation. Said Matcheva:

"The astrobiology community has been working on a definition of 'life' for a long time, but we have no idea what aliens really look like and how they would interact with their environments. We are biased by our human experience, and the current strategies are to search for life in the 'habitable zone,' which by definition is human (or terrestrial life) friendly."

"So how do you search for something when you don't know what it looks like? That is where the novelty detection machine learning techniques come in—they can flag data points that are inconsistent with the training data, i.e., do not agree with the current theoretical models. So indeed, in that sense, our method is searching for life 'as we don't know it.'"

As Isaac Asimov famously said, "The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny.'"

Provided by Universe Today