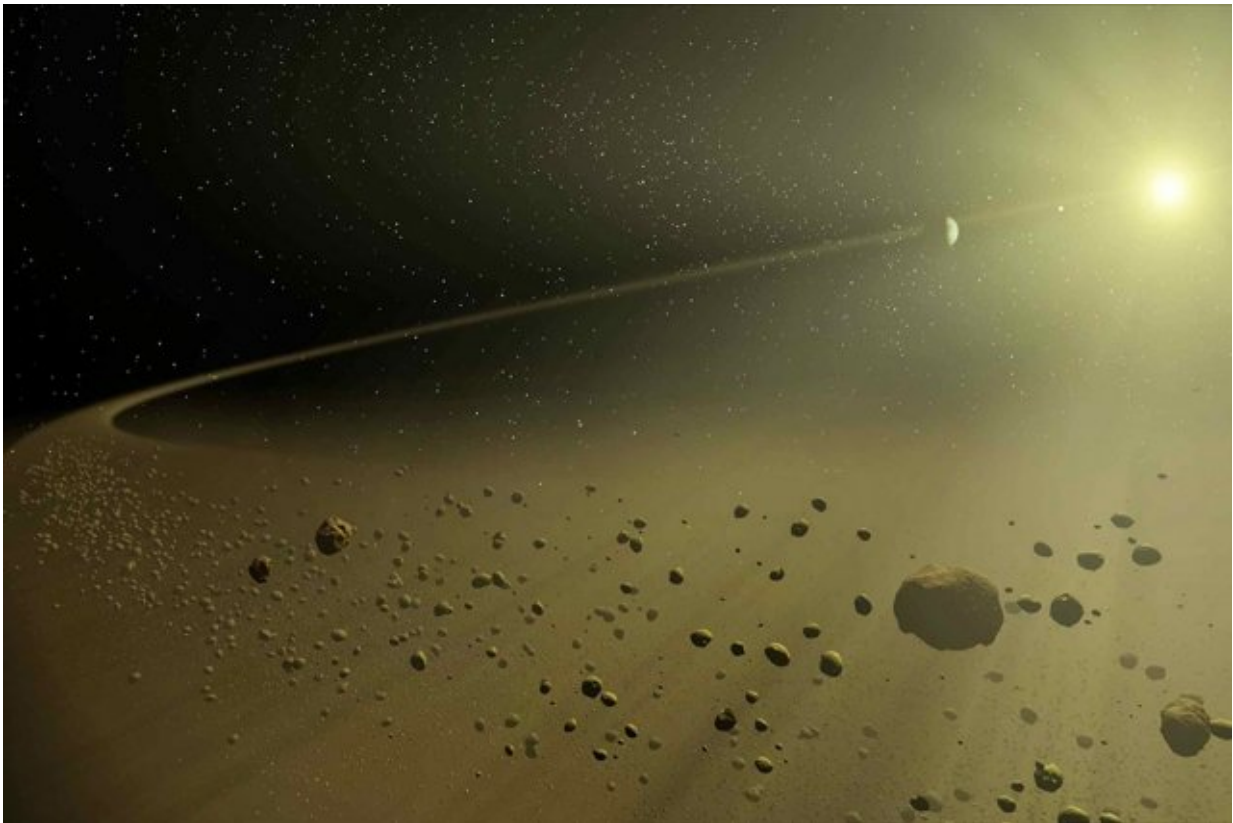# Computer searches telescope data for evidence of distant planets

March 30 2018, by Larry Hardesty

A young sun-like star encircled by its planet-forming disk of gas and dust. Credit: NASA/JPL-Caltech

As part of an effort to identify distant planets hospitable to life, NASA has established a crowdsourcing project in which volunteers search

telescopic images for evidence of debris disks around stars, which are good indicators of exoplanets.

Using the results of that project, researchers at MIT have now trained a machine-learning system to search for debris disks itself. The scale of the search demands automation: There are nearly 750 million possible light sources in the data accumulated through NASA's Wide-Field Infrared Survey Explorer (WISE) mission alone.

In tests, the machine-learning system agreed with human identifications of debris disks 97 percent of the time. The researchers also trained their system to rate debris disks according to their likelihood of containing detectable exoplanets. In a paper describing the new work in the journal *Astronomy and Computing*, the MIT researchers report that their system identified 367 previously unexamined celestial objects as particularly promising candidates for further study.

The work represents an unusual approach to machine learning, which has been championed by one of the paper's coauthors, Victor Pankratius, a principal research scientist at MIT's Haystack Observatory. Typically, a machine-learning system will comb through a wealth of training data, looking for consistent correlations between features of the data and some label applied by a human analyst—in this case, stars circled by debris disks.

But Pankratius argues that in the sciences, machine-learning systems would be more useful if they explicitly incorporated a little bit of scientific understanding, to help guide their searches for correlations or identify deviations from the norm that could be of scientific interest.

"The main vision is to go beyond what A.I. is focusing on today," Pankratius says. "Today, we're collecting data, and we're trying to find features in the data. You end up with billions and billions of features. So

what are you doing with them? What you want to know as a scientist is not that the computer tells you that certain pixels are certain features. You want to know 'Oh, this is a physically relevant thing, and here are the physics parameters of the thing.'"

Classroom conception

The new paper grew out of an MIT seminar that Pankratius co-taught with Sara Seager, the Class of 1941 Professor of Earth, Atmospheric, and Planetary Sciences, who is well-known for her exoplanet research. The seminar, Astroinformatics for Exoplanets, introduced students to data science techniques that could be useful for interpreting the flood of data generated by new astronomical instruments. After mastering the techniques, the students were asked to apply them to outstanding astronomical questions.

For her final project, Tam Nguyen, a graduate student in aeronautics and astronautics, chose the problem of training a machine-learning system to identify debris disks, and the new paper is an outgrowth of that work. Nguyen is first author on the paper, and she's joined by Seager, Pankratius, and Laura Eckman, an undergraduate majoring in electrical engineering and computer science.

From the NASA crowdsourcing project, the researchers had the celestial coordinates of the light sources that human volunteers had identified as featuring debris disks. The disks are recognizable as ellipses of light with slightly brighter ellipses at their centers. The researchers also used the raw astronomical data generated by the WISE mission.

To prepare the data for the machine-learning system, Nguyen carved it up into small chunks, then used standard signal-processing techniques to filter out artifacts caused by the imaging instruments or by ambient light. Next, she identified those chunks with light sources at their centers, and

used existing image-segmentation algorithms to remove any additional sources of light. These types of procedures are typical in any computer-vision machine-learning project.

Coded intuitions

But Nguyen used basic principles of physics to prune the data further. For one thing, she looked at the variation in the intensity of the light emitted by the light sources across four different frequency bands. She also used standard metrics to evaluate the position, symmetry, and scale of the light sources, establishing thresholds for inclusion in her data set.

In addition to the tagged debris disks from NASA's crowdsourcing project, the researchers also had a short list of stars that astronomers had identified as probably hosting exoplanets. From that information, their system also inferred characteristics of debris disks that were correlated with the presence of exoplanets, to select the 367 candidates for further study.

"Given the scalability challenges with big data, leveraging crowdsourcing and citizen science to develop training data sets for machine-learning classifiers for astronomical observations and associated objects is an innovative way to address challenges not only in astronomy but also several different data-intensive science areas," says Dan Crichton, who leads the Center for Data Science and Technology at NAASA's Jet Propulsion Laboratory. "The use of the computer-aided discovery pipeline described to automate the extraction, classification, and validation process is going to be helpful for systematizing how these capabilities can be brought together. The paper does a nice job of discussing the effectiveness of this approach as applied to debris disk candidates. The lessons learned are going to be important for generalizing the techniques to other astronomy and different discipline applications."

"The Disk Detective science team has been working on its own machine-learning project, and now that this paper is out, we're going to have to get together and compare notes," says Marc Kuchner, a senior astrophysicist at NASA's Goddard Space Flight Center and leader of the crowdsourcing disk-detection project known as Disk Detective. "I'm really glad that Nguyen is looking into this because I really think that this kind of machine-human cooperation is going to be crucial for analyzing the big data sets of the future."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology