Classifying Agnostic Biosignatures using Raman, VNIR, and Elemental Data

Aarya Mishra[1], Tao Sheng[2], Aivaras Vilutis[3], Paxton Tomko[4], Michael Furlong[3], Jesse Murray[3], Sunanda Sharma[5], and Diana Gentry[3]

[1]University of San Francisco
[2]University of Pittsburgh
[3]NASA Ames Research Center
[4]Purdue University
[5]Massachusetts Institute of Technology

November 22, 2022

Abstract

How can we use our current wealth of terrestrial data, encompassing biogenic and abiogenic systems, to determine the distinguishing properties of life? SCOBI (Statistical Classification of Biosignature Information) uses machine learning techniques to algorithmically identify combinations of measurements that are "indicative of life". A set of ˜1000 observations, comprising elemental abundance, isotopic fractionation, VNIR reflectance, and (in progress) Raman spectra, have been assembled from existing literature and databases. The observations cover systems classified as "indicative alive" (e.g., cells, vegetation), "indicative non-alive" (e.g., fossils, teeth), "mixed indicative" (e.g., soil, pond water), or "non-indicative" (e.g., rocks, meteorites). VNIR data was preprocessed by linear interpolation from 400-2100 nm and smoothed with a Savitzky-Golay filter. To limit the amount of Earth-biochemistry-specific (non-agnostic) information included, the first five spectral features extracted were number of peaks, number of troughs, mean reflectance, mean peak width, and broadest peak width. To help further emphasize agnostic biosignatures, Earth-specific features such as chlorophylls have been manually flagged so that feature importance with and without them can be compared. Classifiers including k-nearest neighbors (KNN), Gaussian Naïve Bayes (GNB), logistic regression (LR), random forest (RF), and support vector machine (SVM) were implemented, as was a combination voting classifier. Performance metrics included false positive rates, false negative rates, and AUC with 50-50 test/train splits (Monte Carlo simulations). Key takeaways from this stage, prior to the inclusion of Raman spectra, are (1) the overall success rate of 0.933 AUC was most heavily influenced by the elemental abundance data; and (2) VNIR reflectance had the lowest classification performance with 0.52 AUC (58% of objects correctly classified). The next steps are to complete integration of Raman spectral data and to improve the approach to pre-processing and feature extraction for both types of spectral data, such as automated baseline removal, whole spectrum matching, and dimensionality reduction.

National Aeronautics and Space Administration

Classifying Agnostic Biosignatures using Raman, VNIR, and Elemental Data

Aarya Mishra[1], Tao Sheng[2], Aivaras Vilutis[3], Paxton Tomko[4], Michael Furlong[5], Jesse Murray[6], Sunanda Sharma[7], Diana Gentry[5]

[1]University of San Francisco, [2]University of Pittsburgh, [3]NASA International Internships Program, [4]Purdue University, [5]NASA Ames Research Center, [6]University of Oxford, [7]Massachusetts Institute of Technology

Introduction and Background

How can we use our current wealth of terrestrial data, encompassing living, non-alive but biogenic, and abiogenic systems, to determine the distinguishing properties of life? SCOBI (Statistical Classification of Biosignature Information) is an effort to develop an algorithm for identifying “astrobiologically interesting” observations. Agnostic biosignatures, signs of life that are not specific to a particular biochemistry, are considered a particularly high standard for life detection.

The goal of this work is to classify and organize ‘agnosticized’ data on terrestrial systems collated from existing literature and databases. This allows a machine learning algorithm to be provided with an adequate teaching data set classifying samples as “indicative of life” or “not indicative of life”; if successful, the results will provide insight into promising targets for future astrobiology missions.

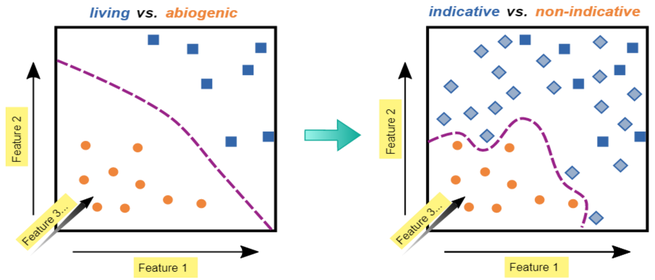

Figure 1. Graphic representations of a binary classifier using observations of features to decide which side of the boundary a system would likely be on. For SCOBI, objects on one side of the boundary would be "indicative of life" or “not indicative of life” [1].

Investigations

Verifying the current organizational systems in place involved several types of investigations. These investigations aimed to either:

- improve biochemistry agnosticism (e.g., remove features specific to ATP or chlorophyll); or
- verify that systems were correctly classified as “indicative alive” (cell), “indicative non-alive” (fossil), “mixed indicative” (soil) or “non-indicative” (rock).

Data Type	# of Samples	Data Type Examples
Non-indicative	255	Meteorites, carbonite, sand
Indicative alive	42	Pigmented microorganisms, vegetation
Indicative non-alive	23	Coral skeleton, limestone
Indicative mixed with non-indicative	12	Biofilm and microbial mats, clay

www.nasa.gov

Data Type Comparisons

Combine Silt and Clay Data?

Investigation:
- Compared spectral data within silt and clay
- Noted one graphical outlier (Silt fgt017) in the silt category; doesn't follow other silt spectra reflectance trend

- **Results:**
  - Silt and clay spectral data are similar
  - Missing isotopic data for silt and clay, elemental data for silt
  - Unable to combine silt and clay categories based on only available spectral data

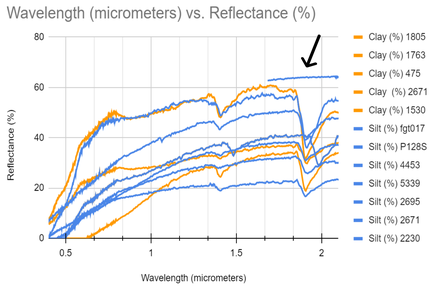

Figure 2. A comparison of spectra in samples tagged as silt versus samples tagged as clay. The features are substantially the same. Arrow points to outlier, Silt fgt017.

Combine Meteorite Classes?

Investigation:
- Explored elemental, spectral, and isotopic differences between stony and iron meteorites
- Compared stony and iron meteorites data collection methodology , specifically, isotopic data
- Further explored stony class and 2 of its subcategories: chondrites and achondrites

- **Results:**
  - Homogenous types of 3 main classes of meteorites have different compositions
  - Chondrites have larger range of O isotopes than achondrites
  - "Group level difficult to determine using bulk elemental composition” [2]
  - Will split stony and iron meteorite categories

Figure 3. (L) Chondrite [3] and (R) Achondrite [4]. Both are stony class meteorites yet have different isotopic ranges for various elements like oxygen.

Agnostically Reclassify Microbes?

Investigation:
- Prev. photosynthetic vs. non-photosynthetic
- Separated each microbe in SCOBI data set based on phylum
- Identified key secondary metabolites (pigments) that could be used as markers for life; confirmed key pigments in microbes

- **Results:**
  - Potentially non-agnostic pigments identified include sunscreen pigments, certain carotenoids, chlorophylls, and bacteriochlorophylls
  - Specific carotenoids and chlorophylls could be better agnostic classifiers for microorganisms

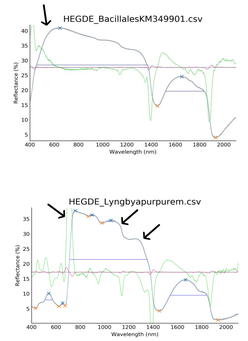

Figure 4. (L) Reflectance spectra of *Bacillales* (KM349901), a non-photosynthetic microorganism. (R) Reflectance spectra of *Lyngbya purpurem*, a photosynthetic microorganism. Arrows point to differences in spectra, and to spectral features such as peak number and slope, which suggest more in-depth classification is needed [5].

Future Work

Possible Ways to Improve Data Coverage and Classification Confidence:

1. Complete addition of Raman spectroscopy:
   1. Common, publicly available
   2. Provides info on chemical bonds (new data type)
2. Add more mixed samples to increase confidence in mixed indicative samples (ex: snow mixed with vegetation)
3. Add more sterilized/artificial samples in non-indicative categories
4. Explore more examples of secondary metabolites as alternative classifiers for life detection (ex: pigments and toxins)

References & Acknowledgements

[1] Gentry, D. Statistical Classification of Biosignature Information: Improving Life Detection Confidence-Bio-Engineering & Instrumentation Group (BeING) Laboratory. *Collaborative Biosciences Seminar.* NASA Ames Research Center. 8 April 2021.
[2] Miyamoto, Hideaki, et al. “Cluster Analysis on the Bulk Elemental Compositions of Antarctic Stony Meteorites.” Meteoritics & Planetary Science, vol. 51, no. 5, May 2016, pp. 906–19. doi:10.111/maps.12634.
[3] Gronstal, Aaron. “A Unique Metal-Rich Chondrite”. 21 September 2019.
[4] NASA Planetary Data System. https://pds.nasa.gov
[5] Sharma, Sunanda. SCOBI. NASA Ames Research Center, 2021.
Thank you to Gabriela Pena Carmona and the rest of the BeING Lab for your support. Thank you to the NASA VIP Program coordinator, Porsche Parker.