

# Algorithmic detection of elemental biosignatures

Murray Jesse<sup>1</sup>

<sup>1</sup>University of Oxford

November 16, 2022

## Abstract

Machine learning (ML) models that classify a sample as non-indicative or indicative of life can play an important role in planning life-detection missions. They are based on clearly defined and consistent algorithms, regardless of sample type or origin, and make their predictions from weighted combinations of multiple features rather than from any singular feature. These weighted combinations can reveal the most informative measurements within the operational constraints of a life-detection mission. The Ladder of Life Detection (Neveu 2018) identifies the need for an understanding of how combinations of multiple biosignatures affect overall confidence. The present work provides a starting point to answer this need, and future work will expand the data types to obtain even more predictive combinations of features. Elemental composition and isotope fractionation were chosen as the data types, as they are available for both biogenic and abiogenic systems and not unique to Earth biochemistry. Measurements of these data types across a wide range of unambiguously non-indicative or indicative samples were gathered from published literature. The varied sample measurements were then integrated into twenty-one representative samples. The ML models only made binary classifications of non-indicative or indicative of life. Nonetheless, the indicative samples broadly fell into three categories: mixed, non-alive, and alive. Four classification algorithms were trained and tested with Monte Carlo simulations using a 70:30 train to validation ratio. Between the models, around 75% of the test samples were correctly classified, with variations in sensitivity and specificity of the models. For elemental abundances predictive of a non-indicative of life sample: all models found Ti and Si as strong and Fe, Al, Mn, and Mg as medium. For predicting an indicative of life sample, all models found C, N, and Carbon-13 as strong and K, H, P, and Ca as medium. A weighted combination of multiple biosignatures is shown to be a more effective approach to classifying sample-data than relying on any individual biosignature or on an unweighted group of biosignatures. Different models also made different chronic misclassifications, suggesting that combining the outputs of multiple models may be more effective than relying on the output of a singular model. Which type of model to use may depend on the application, e.g., higher sensitivity models might be preferred in first-pass situations where false-negatives are more costly than false-positives. Lastly, the weighted combination of measurements in a model suggests how to combine biosignatures to affect the overall confidence of the classification. These results provide evidence of elemental biosignatures beyond the CHNOPS of Earth-based life and serve as a proof of concept for algorithmic biosignature classification.

# Algorithmic detection of elemental biosignatures



## Algorithmic detection of elemental biosignatures

Jesse Murray (1, 2), Aivaras Vilutis (3, 4), Thomas Stucky (5, 6), Michael Furlong (6, 7), Jessica E. Koehne (6), David Mauro (6, 8), Annmarie Schramm (6), Diana Gentry (6)

(1) University of Oxford, (2) NASA Ames Research Center, NASA Internships, Fellowships & Scholarships, (3) Vilnius University, (4) NASA Ames Research Center, NASA International Internships Program, (5) SETI Institute Mountain View, (6) NASA Ames Research Center, (7) Stinger Ghaffarian Technologies, (8) Millennium Engineering & Integration



### Introduction

Machine learning (ML) models that classify a sample as *non-indicative* or *indicative* of life can play an important role in engineering life-detection systems.

There is not currently a general classification scheme for potential biosignature data. ML models can fill this gap by modeling concrete, specific features, i.e., measurements, that identify a sample as *non-indicative* or *indicative* of life. ML models have the following advantages:

1. They are based on clearly defined and consistent algorithms, regardless of sample type or origin.
2. They make these predictions from weighted combinations of multiple features, which can be more predictive than any singular feature.

### Modeling

The basic approach of the modeling was to algorithmically classify a representative sample as *non-indicative* or *indicative* of life from its elemental composition and isotopic fractionation.

The representative samples were initially explored through principal component analysis (PCA) (shown in Figure 1), and principal components (PCs) were used to describe the most informative components, which are perpendicular vectors through the variance of the data. The vector that explained the largest proportion of the variance of the data (PC1) is usually also diagonal during separation between *non-indicative* and *indicative* of life samples. This vector can represent a weighted sum of multiple measurements, providing an early example of how a weighted combination of measurements was most effective at differentiating the *non-indicative* and *indicative* samples. It is clear from Figure 2, that a single measurement provides a "winning" over classification.

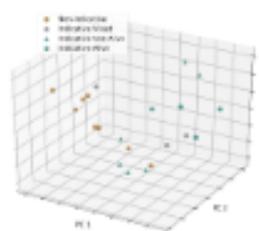


Figure 2: The three leading PCA vectors of all the representative samples. These three vectors explained 87% of the variance of the data.

### Results

Elements predictive of a *non-indicative* of life sample:

- All models based Si and S abundances to be strong predictors.
- All models based Fe, B, Mn, and Mg abundances to be moderate predictors.

Elements predictive of an *indicative* of life sample:

- All models based C and Mn abundances to be strong predictors.
- All models based K, H, P and Ca abundances to be moderate predictors.

Elements with varied predictive abilities:

- Si (strong indicator), Na, Cl (strong non-indicative)
- All models based Carbon 13 abundance to be *non-indicative* of the

### Data Collection and Standardization

Elemental composition and isotopic fractionation measurements are desirable, as they are available for both biogenic and abiogenic systems and not unique to both biosignatures. Specifically, the isotopic ratios were:

- Carbon 13 to Carbon 12
- Oxygen 18 to Oxygen 16
- Hydrogen 2 (Deuterium) to Hydrogen 1.

Measurements of these data types across a wide range of heterogeneity *non-indicative* or *indicative* samples were gathered from published literature. The varied sample measurements were then integrated into fairly *non-representative* samples.

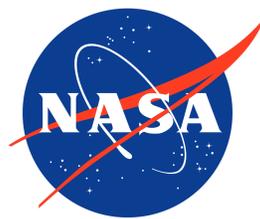
Table 2: The classification of each representative sample, and the number of papers used for the elemental composition and isotopic fractionation data.

Classification	Representative Sample	Compositional Isotope Papers
<i>Non-indicative</i>	Si	1
	S	2
<i>Indicative</i>	C	1
	Mn	2

[LIVE SESSION](#)
[AUTHOR INFORMATION](#)
[ABSTRACT](#)
[CONTACT AUTHOR](#)
[PRINT](#)
[GET POSTER](#)

Jesse Murray (1, 2), Aivaras Vilutis (3, 4), Thomas Stucky (5, 6), Michael Furlong (6, 7), Jessica E. Koehne (6), David Mauro (6, 8), Annmarie Schramm (6), Diana Gentry (6)

(1) University of Oxford, (2) NASA Ames Research Center, NASA Internships, Fellowships & Scholarships, (3) Vilnius University, (4) NASA Ames Research Center, NASA International Internships Program, (5) SETI Institute Mountain View, (6) NASA Ames Research Center, (7) Stinger Ghaffarian Technologies, (8) Millennium Engineering & Integration



PRESENTED AT:

# AGU FALL MEETING

Online Everywhere | 1-17 December 2020

## INTRODUCTION

Machine learning (ML) models that classify a sample as **non-indicative** or **indicative** of life can **play an important role in planning life-detection missions**.

There is not currently a general classification scheme for potential biosignature data. ML models can fill this gap by revealing common agnostic features, i.e., measurements, that identify a sample as non-indicative or indicative of life. ML models have the following advantages:

1. They are based on clearly defined and consistent algorithms, regardless of sample type or origin.
2. They make their predictions from weighted combinations of multiple features, which can be more predictive than any singular feature.

This last capability can be used to understand how combinations of different biosignatures affect overall confidence, which was identified as a key gap in The Ladder of Life Detection (Neveu 2018).

ML models can improve the instrumentation protocols within the operational constraints of a planetary exploration mission. For detecting the presence of life at a particular location, e.g., Europa, one possible approach is to down-select samples using two passes.

In the first pass, a set of instruments could be used to provide a preliminary indication of the presence of life. These instruments would ideally not be constrained to a limited number of measurements over the course of the mission. It is also more important that the classification they provide be sensitive rather than specific. If this first-pass yields positive results, a second, more specific pass could be made by supply-constrained assays, e.g., those involving reagents.

By showing how combinations of different measurements affect the confidence of a classification, ML models can suggest the most informative set of measurements for this first-pass. Furthermore, certain instruments might not improve the performance of classifications, in which case their exclusion from a mission would improve the effective allocation of available payload space.

## DATA COLLECTION AND STANDARDIZATION

Elemental composition and isotope fractionation were chosen as the data types, as they are available for both biogenic and abiogenic systems and not unique to Earth biochemistry. Specifically, the isotopic ratios were:

- Carbon-13 to Carbon-12
- Oxygen-18 to Oxygen-16
- Hydrogen-2 (Deuterium) to Hydrogen-1.

Measurements of these data types across a wide range of unambiguously **non-indicative** or **indicative** samples were gathered from published literature. The varied sample measurements were then integrated into twenty-one **representative samples**.

**Table 1:** The classification of each representative sample, and the number of academic papers sourced for the element composition and isotope fractionation data.

Classification	Representative Sample	Composition Papers	Isotope Papers	
<b>Non-Indicative</b>	Basalt	5	1	
	Carbonatite	1	2	
	Lunar Rock	2	1	
	Magnetite	1	2	
	Mars Atmosphere	1	4	
	Mars Rock	2	3	
	Meteorite	1	2	
	Sand	3	1	
<b>Indicative - Mixed</b>	Earth Atmosphere	1	5	
	Seawater	1	3	
	Soil	1	1	
	<b>Non-Alive</b>	Bone	2	2
		Chalk	1	1
		Coal	2	4
		Coral Skeleton	1	1
		Crude Oil	1	3
		Limestone	1	2
		<b>Alive</b>	Bacteria	3
Biofilm	2		3	
Human	1		1	
Plant	2		3	

The ML models only made binary classifications of **non-indicative** or **indicative** of life. Nonetheless, the indicative samples broadly fell into three categories: mixed, non-alive, and alive.

- The mixed samples were largely non-biological, with small concentrations of bacteria - potentially similar to a future mission sample.
- The non-alive samples were primarily fossilized remains, including skeletal coral fragments (decades old) and marine sediments of biological origin (millions of years old). The bone samples included modern and fossil teeth, as well as other historic bone samples.
- The non-indicative samples included a range of mineral samples and the atmosphere of Mars. The mineral-types could not have been affected by the oxygenation of the Earth's atmosphere and were unambiguously formed via geological

processes.

X-ray diffraction, mass spectrometry, and other measurements were standardized to a simulated limit of detection of 300 ppm. Only the elements for which 30% of samples had measurements above this level were included in the analysis, which resulted in sixteen (non-trace) elements, shown in Figure 2. Abundances below 300 ppm were imputed with 15 ppm, which was the geometric mean of the measurements of the sixteen elements below 300 ppm. Twenty-one percent of the final element abundance measurements were imputed this way.

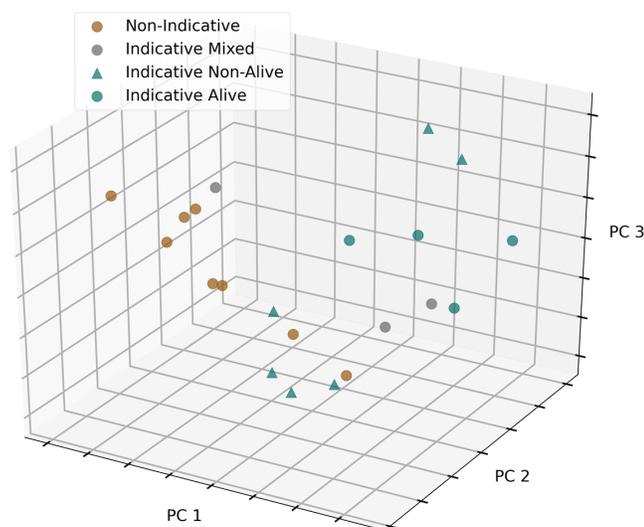
Oxygen-18 fractionation was not available for two representative samples (Coral Skeleton and Human) and deuterium fractionation was not available for ten representative samples. Missing isotopic measurements were imputed with the overall means, and the models were run both including and excluding deuterium (Figure 3 is shown with deuterium excluded).

## MODELING

The basic approach of the modeling was to algorithmically classify a representative sample as **non-indicative** or **indicative** of life from its elemental composition and isotope fractionation.

The representative samples were initially explored through principal component analysis (PCA) shown in Figure 1, and strip-plots shown in Figure 2. PCA reduces multi-dimensional data into the most informative components, which are perpendicular vectors through the measurement space. The vector that explained the largest proportion of the variance of the data (30%) incidentally also displayed strong separation between non-indicative and indicative of life samples.

This vector corresponded to a weighted sum of multiple measurements, providing an early example of how a weighted combination of measurements was most effective at differentiating the non-indicative and indicative samples. It is clear from Figure 2, that no single measurement provides a 'smoking-gun' classification.



**Figure 1:** The three leading PCA vectors of all the representative samples. These three vectors explained 57% of the variance of the data.

Four classification algorithms were used, which offer different ways of making predictions:

- K-nearest neighbors (KNN)
- Logistic regression (LR)
- Linear support vector machines (SVM)
- Gaussian naïve Bayes (GNB)

The models were trained and tested on 1,000 random splits of the data (70% train, 30% test). For the KNN algorithm, the training data was reduced into the three leading principal components, similar to Figure 1. LR and SVM were run with regularization (L1 and L2) in order to reduce overfitting and distinguish the most predictive features.

**Figure 2:** Strip-plots of the elemental abundances and isotope fractionations across the four representative sample types.

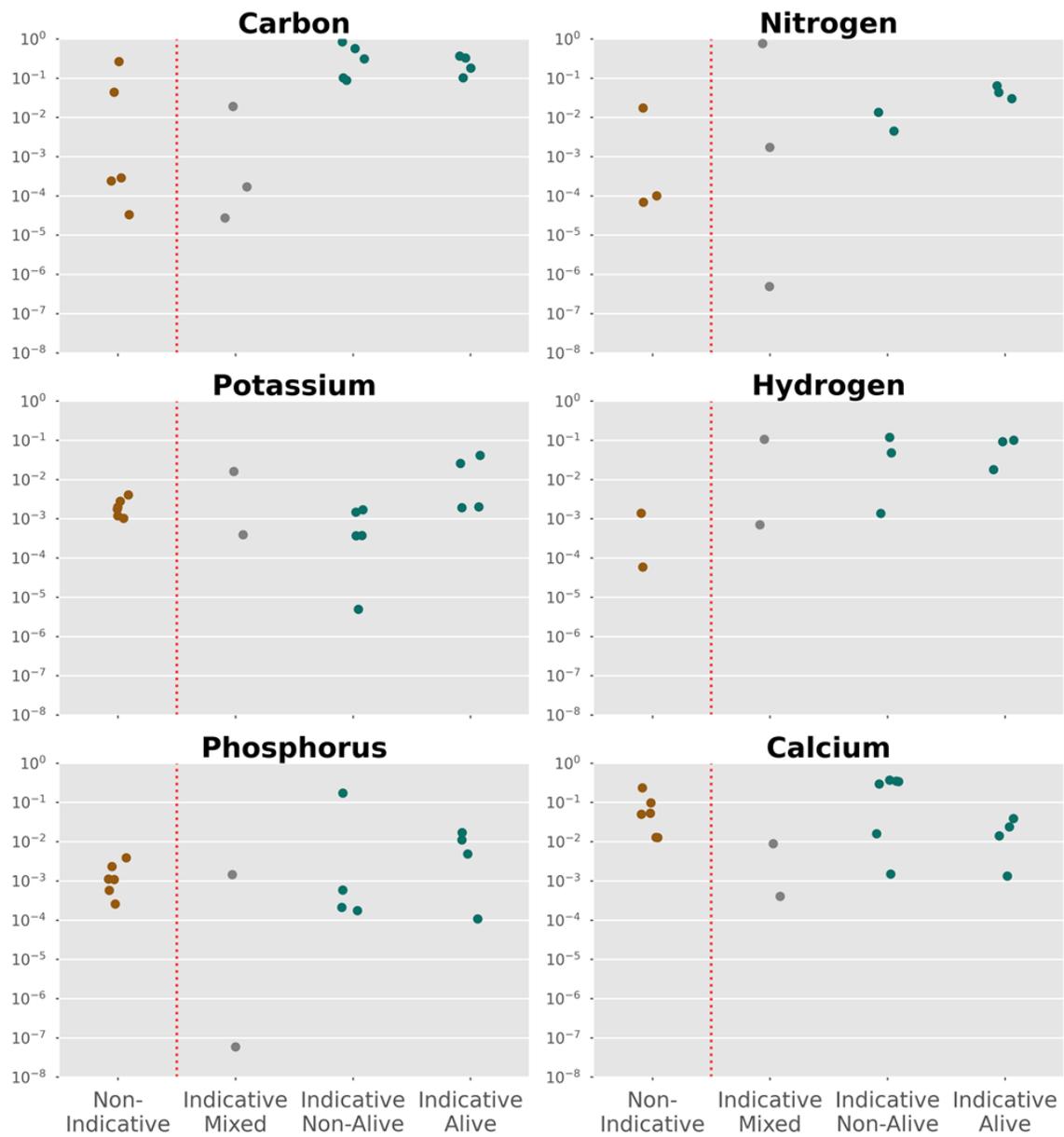


Figure 2a: The mass abundances, e.g. 0.1 = 10%, of elements found by the models to be predictive of an *indicative of life sample*.

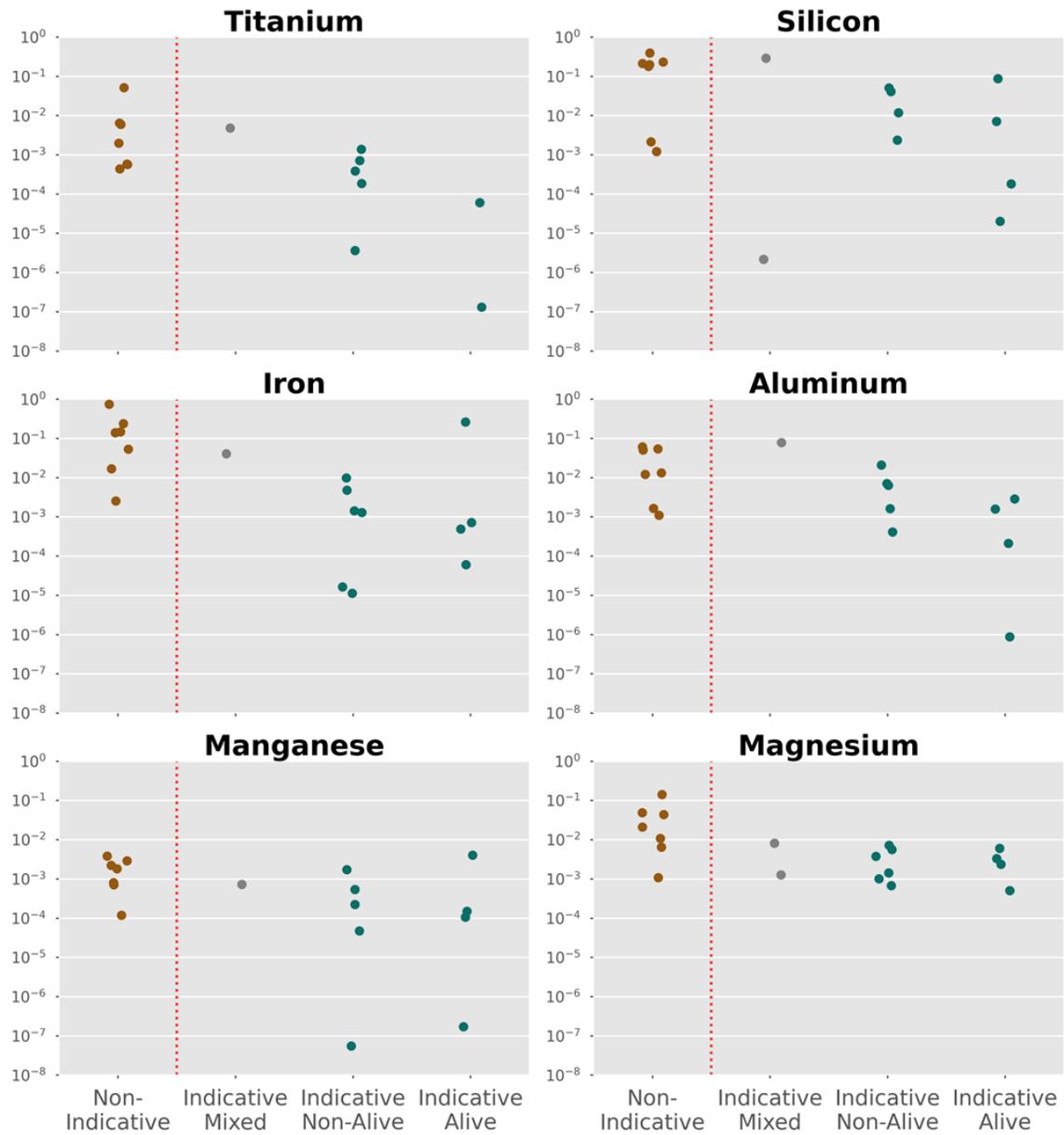


Figure 2b: The mass abundances of elements found to be predictive of a *non-indicative of life* sample.

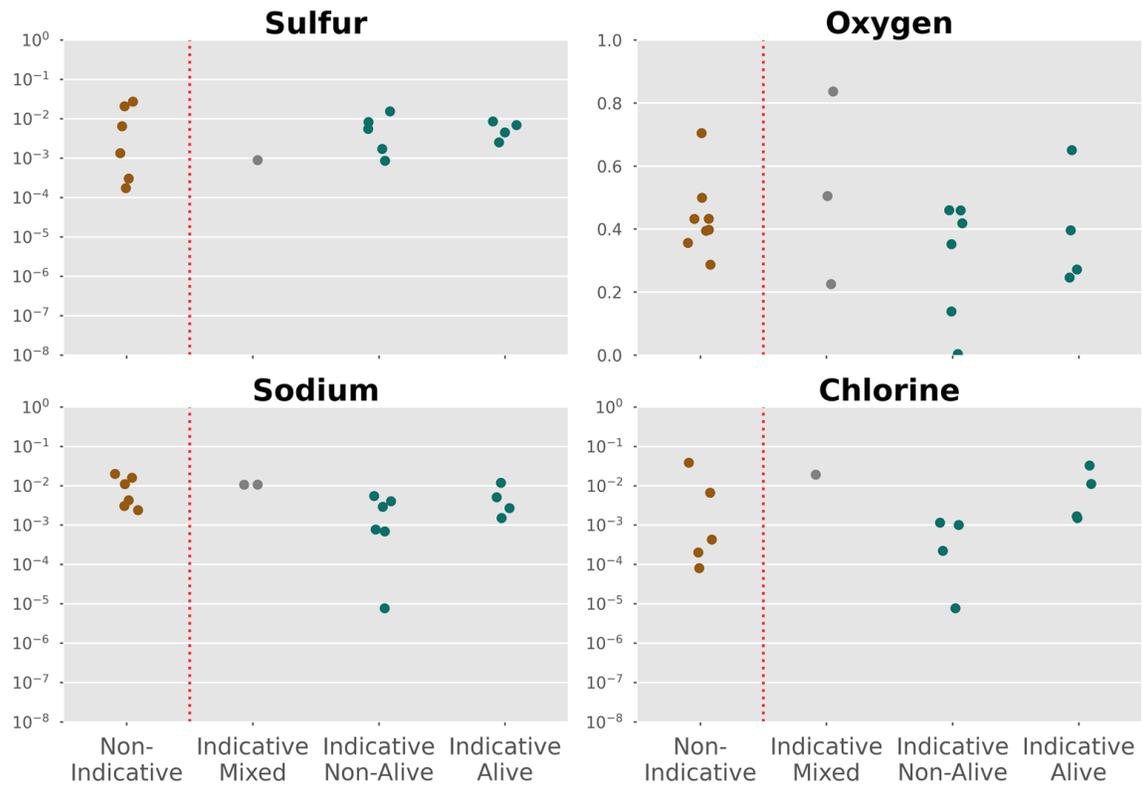


Figure 2c: The mass abundances of elements with varied predictions.

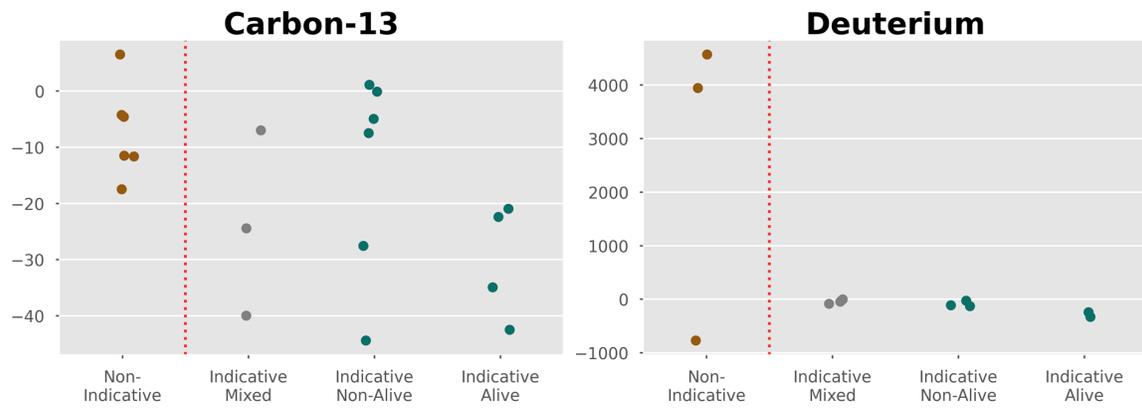


Figure 2d: Isotope fractionations found to be predictive, measured in per mil (‰) relative to reference standards.

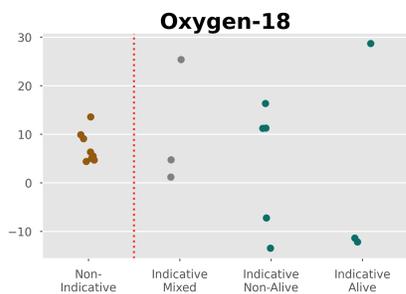


Figure 2e: Isotope fractionation not found to be predictive.



## RESULTS

Elemental abundances predictive of a **non-indicative of life sample**:

- All models found Ti and Si to be strong predictors.
- All models found Fe, Al, Mn, and Mg to be moderate predictors.

Elemental abundances predictive of an **indicative of life sample**:

- All models found C and N to be strong predictors.
- All models found K, H, P, and Ca to be moderate predictors.

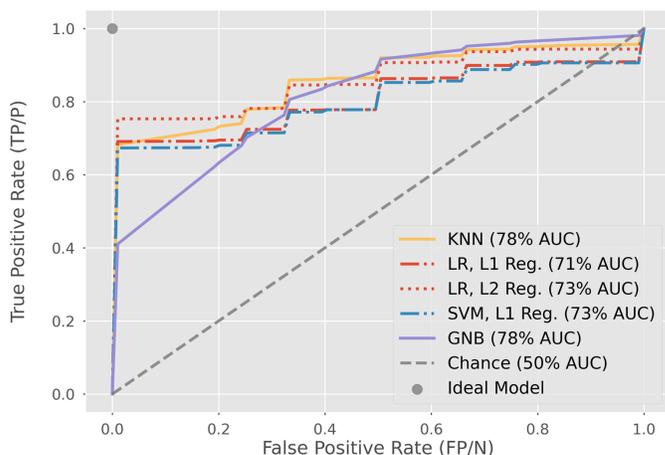
Elemental abundances with varied prediction directions:

- S (leaning indicative), Na, Cl (leaning non-indicative)

Isotope fractionation predictions:

- All models found Carbon-13 abundance to be non-indicative of life.
- All models found Hydrogen-2 abundance to be moderately non-indicative of life.
- Oxygen-18 fractionation was not predictive.

On average, the models correctly classified 75% of the test samples. While LR was more specific, GNB was more sensitive. It can be seen in Figure 3 that under the constraint of the same low false-positive rate, LR performed best (most true positives). Likewise, under the constraint of a low false-negative rate, i.e., a high true-positive rate, GNB performed best (fewest false positives).



**Figure 3:** Mean receiver operating characteristics of the four models.

False-positive misclassifications:

- Carbonatite was *almost always* (~90%) misclassified by all the models.
- Mars Rock was almost always misclassified by LR, though *rarely* (~10%) by the other models.
- Mars Atmosphere was rarely misclassified by LR, though almost always by the other models.
- A few other non-indicative samples were misclassified at extremely low rates.

False-negative misclassifications:

- Soil was almost always misclassified by the models.
- Earth Atmosphere was *frequently* (~80%) misclassified by LR and SVM, though rarely by KNN and GNB.
- Chalk was frequently misclassified by LR, though rarely by the other models.

- Biofilm was rarely misclassified by LR, though frequently by the other models.
- Coral skeleton was misclassified by the models at relatively low rates.
- A few other samples were misclassified at extremely low rates.

## DISCUSSION AND FUTURE WORK

Our results suggest that a weighted combination of multiple biosignatures is a more effective approach to classifying sample-data as **non-indicative** or **indicative** of life than relying on any individual biosignature or on an unweighted group of biosignatures.

The different ML models 'learned' the available data in different ways, resulting in their differing chronic misclassifications. Therefore, combining the outputs of multiple models may be more effective than relying on the output of a singular model.

Which type of model to use depends on the application. In the first-pass of a location, false-negatives are more costly than false-positives. Therefore, the important features of the GNB model would be the recommended set of measurements to make for the first-pass. Furthermore, the weighted combination in the GNB model suggests how those measurements might be combined to affect the overall confidence of the classification.

These results serve as a proof of concept for algorithmic biosignature classification and provide evidence of elemental biosignatures beyond the CHNOPS of Earth-based life. In future work, we intend to expand the models to include other types of data, e.g., spectral information and free energy.

# AUTHOR INFORMATION

Jesse Murray<sup>1</sup>, [jesse.murray@stats.ox.ac.uk](mailto:jesse.murray@stats.ox.ac.uk)

## ABSTRACT

Machine learning (ML) models that classify a sample as non-indicative or indicative of life can play an important role in planning life-detection missions. They are based on clearly defined and consistent algorithms, regardless of sample type or origin, and make their predictions from weighted combinations of multiple features rather than from any singular feature. These weighted combinations can reveal the most informative measurements within the operational constraints of a life-detection mission. The Ladder of Life Detection (Neveu 2018) identifies the need for an understanding of how combinations of multiple biosignatures affect overall confidence. The present work provides a starting point to answer this need, and future work will expand the data types to obtain even more predictive combinations of features.

Elemental composition and isotope fractionation were chosen as the data types, as they are available for both biogenic and abiogenic systems and not unique to Earth biochemistry. Measurements of these data types across a wide range of unambiguously non-indicative or indicative samples were gathered from published literature. The varied sample measurements were then integrated into twenty-one representative samples. The ML models only made binary classifications of non-indicative or indicative of life. Nonetheless, the indicative samples broadly fell into three categories: mixed, non-alive, and alive. Four classification algorithms were trained and tested with Monte Carlo simulations using a 70:30 train to validation ratio. Between the models, around 75% of the test samples were correctly classified, with variations in sensitivity and specificity of the models. For elemental abundances predictive of a non-indicative of life sample: all models found Ti and Si as strong and Fe, Al, Mn, and Mg as medium. For predicting an indicative of life sample, all models found C, N, and Carbon-13 as strong and K, H, P, and Ca as medium.

A weighted combination of multiple biosignatures is shown to be a more effective approach to classifying sample-data than relying on any individual biosignature or on an unweighted group of biosignatures. Different models also made different chronic misclassifications, suggesting that combining the outputs of multiple models may be more effective than relying on the output of a singular model. Which type of model to use may depend on the application, e.g., higher sensitivity models might be preferred in first-pass situations where false-negatives are more costly than false-positives. Lastly, the weighted combination of measurements in a model suggests how to combine biosignatures to affect the overall confidence of the classification. These results provide evidence of elemental biosignatures beyond the CHNOPS of Earth-based life and serve as a proof of concept for algorithmic biosignature classification.

