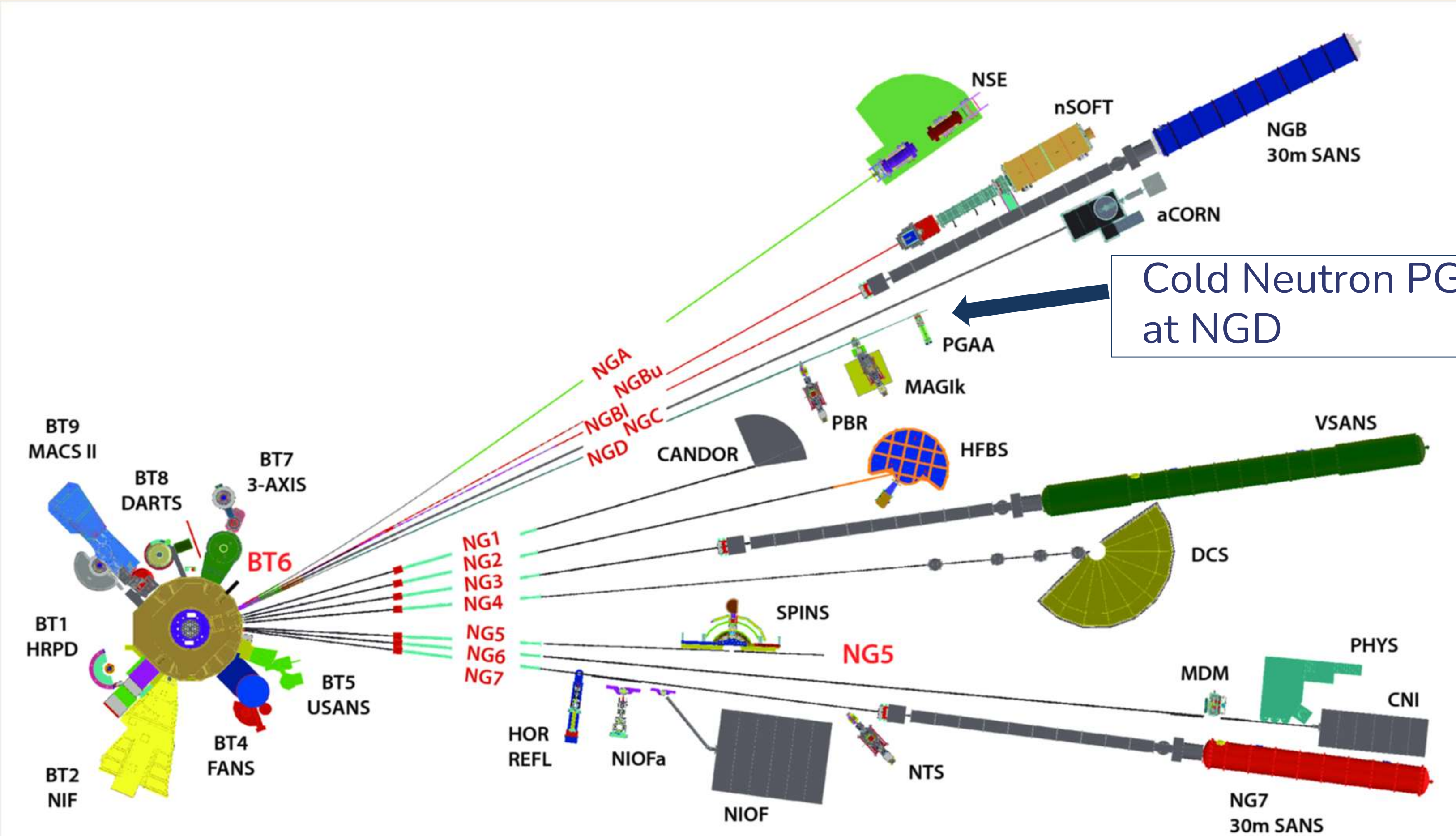
Summer 2026

Classifying and Quantifying Materials from PGAA Spectra with Machine Learning

NIST's Center for Neutron Research

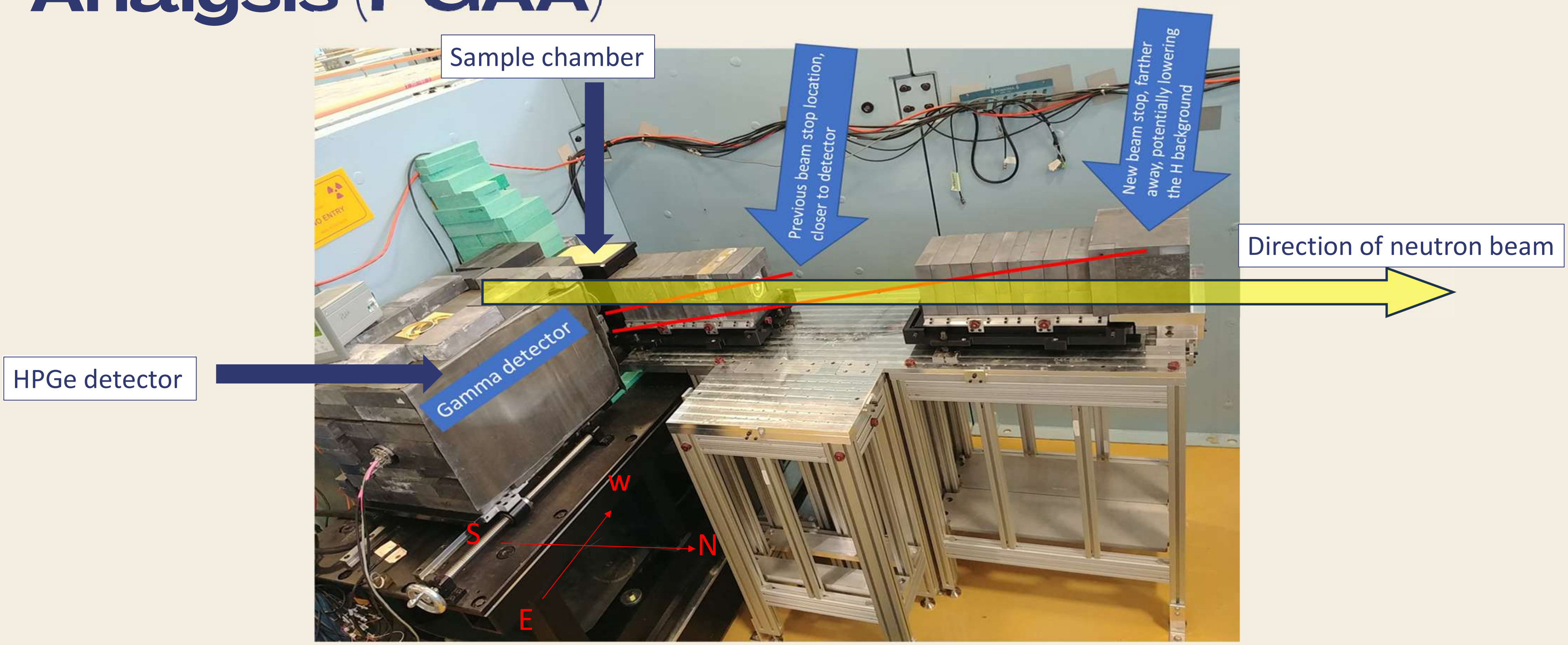
Presented by Cecilia Cohen, The Holton-Arms School
Mentors: Heather Chen-Mayer, Ibrahim Oksuz

NCNR cold neutron beam guide hall



Cold Neutron PGAA at NGD

Prompt Gamma Activation Analysis (PGAA)

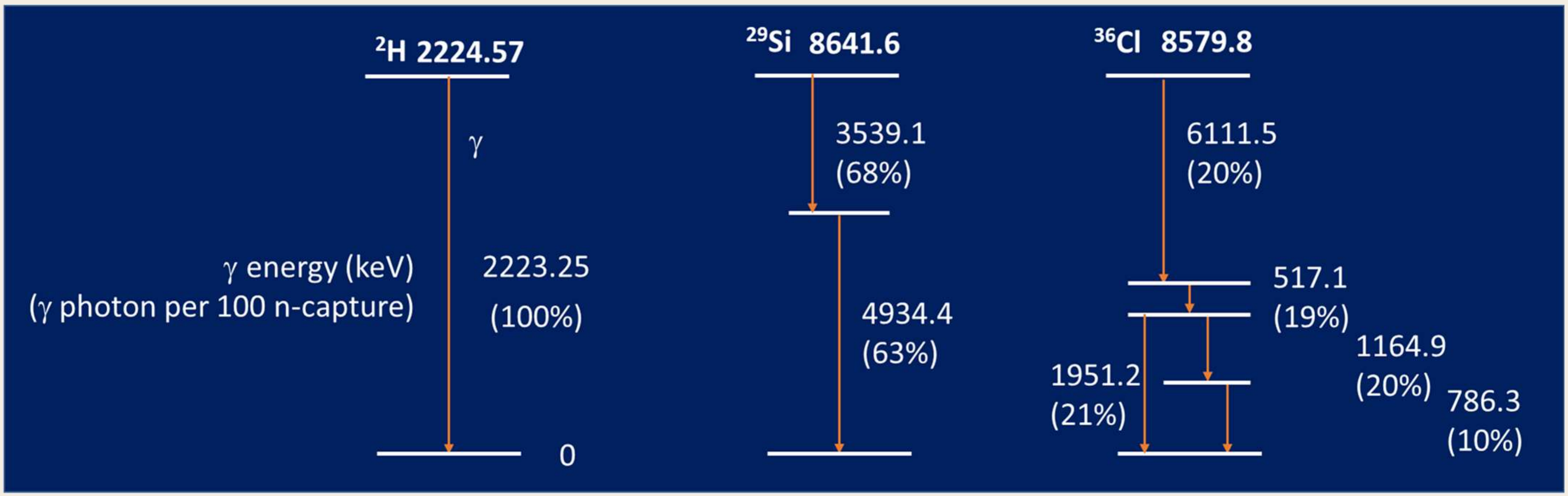


Radiative Neutron Capture

$$ {}^{A}_{Z}X(n,\gamma)\ {}^{A+1}_{Z}X^{*} \longrightarrow {}^{A+1}_{Z}X $$

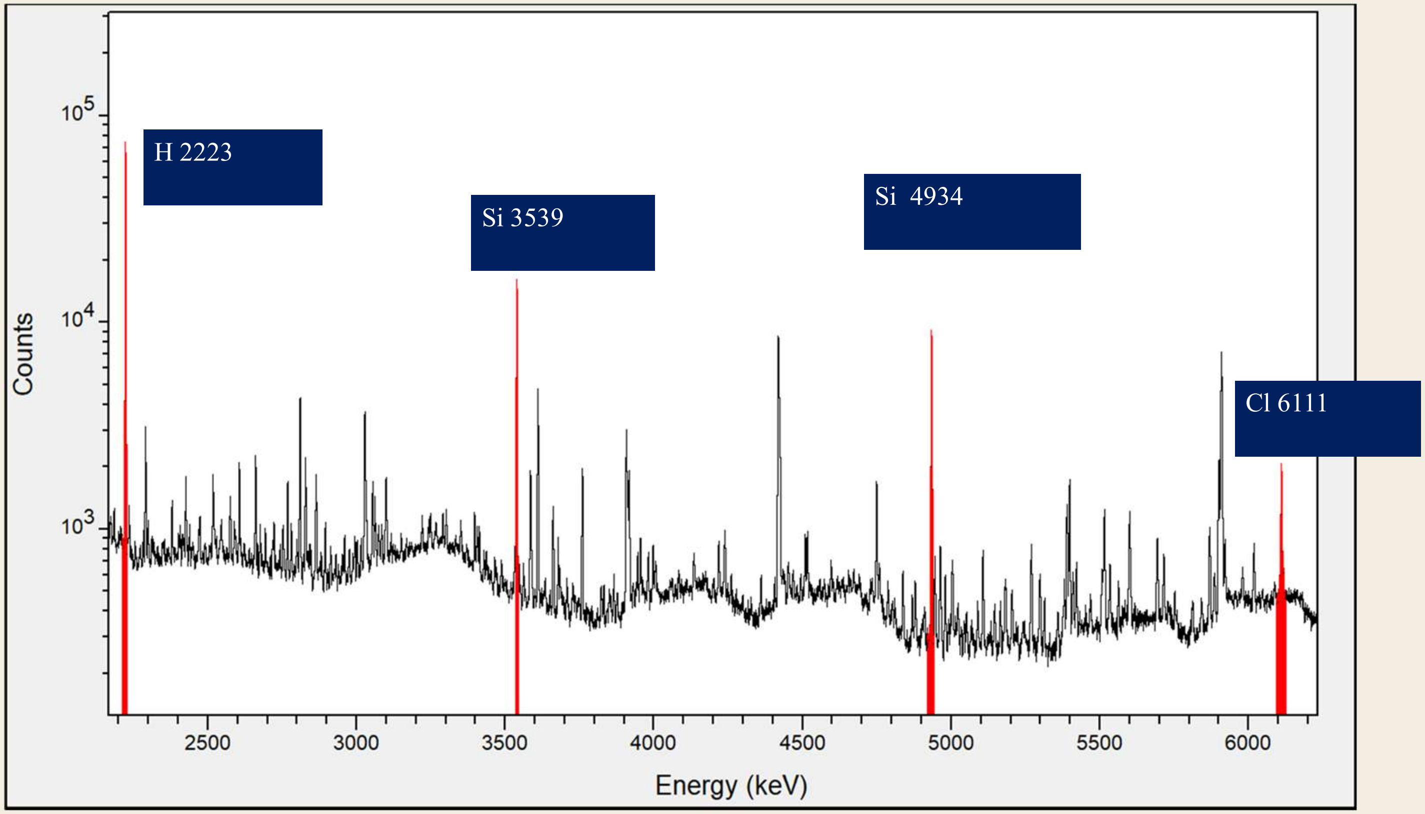
"de-excitation"

"Prompt" -- transition occurs 10^{-12} s to 10^{-9} s



Example nuclear transition levels and prompt gamma ray emission probability

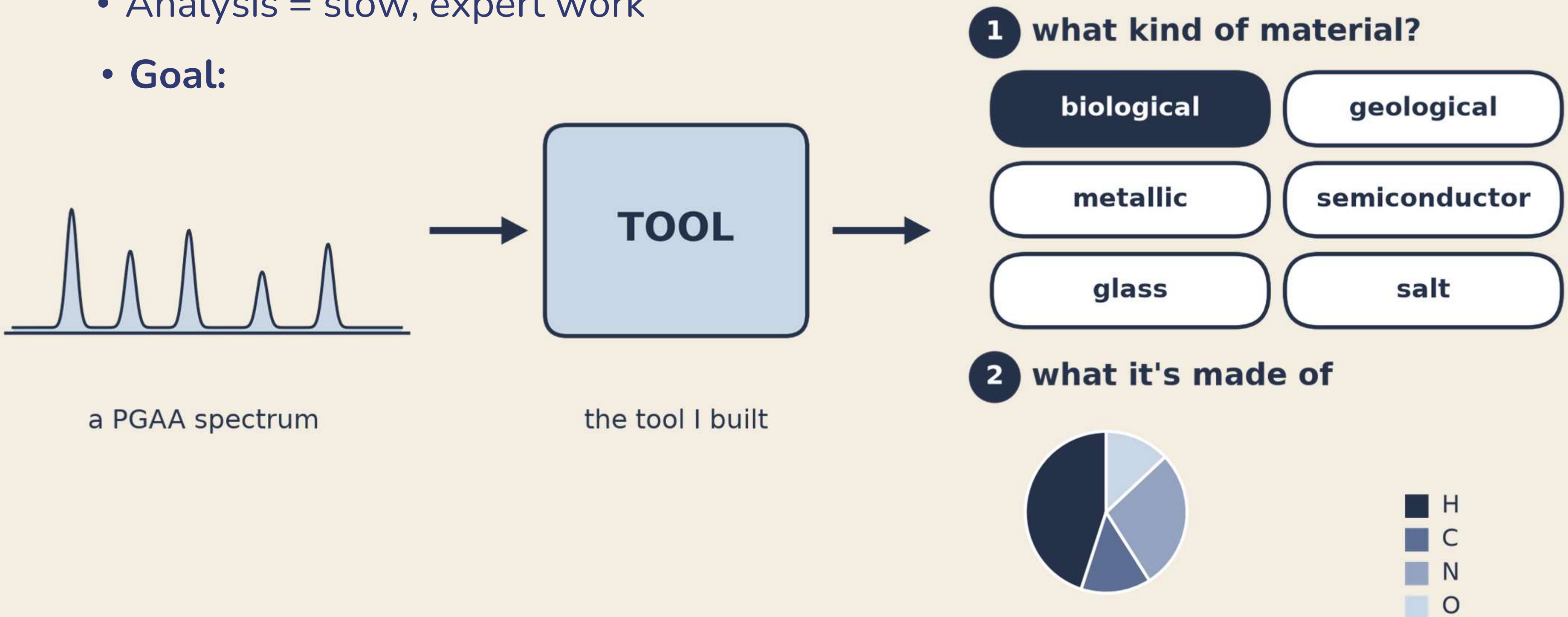
Gamma emission energy range ~ 50 keV to 10 MeV



Partial gamma ray energy spectrum (H, Si, and Cl in concrete matrix)

The Problem

- Analysis = slow, expert work
- **Goal:**



Applications + Motivations

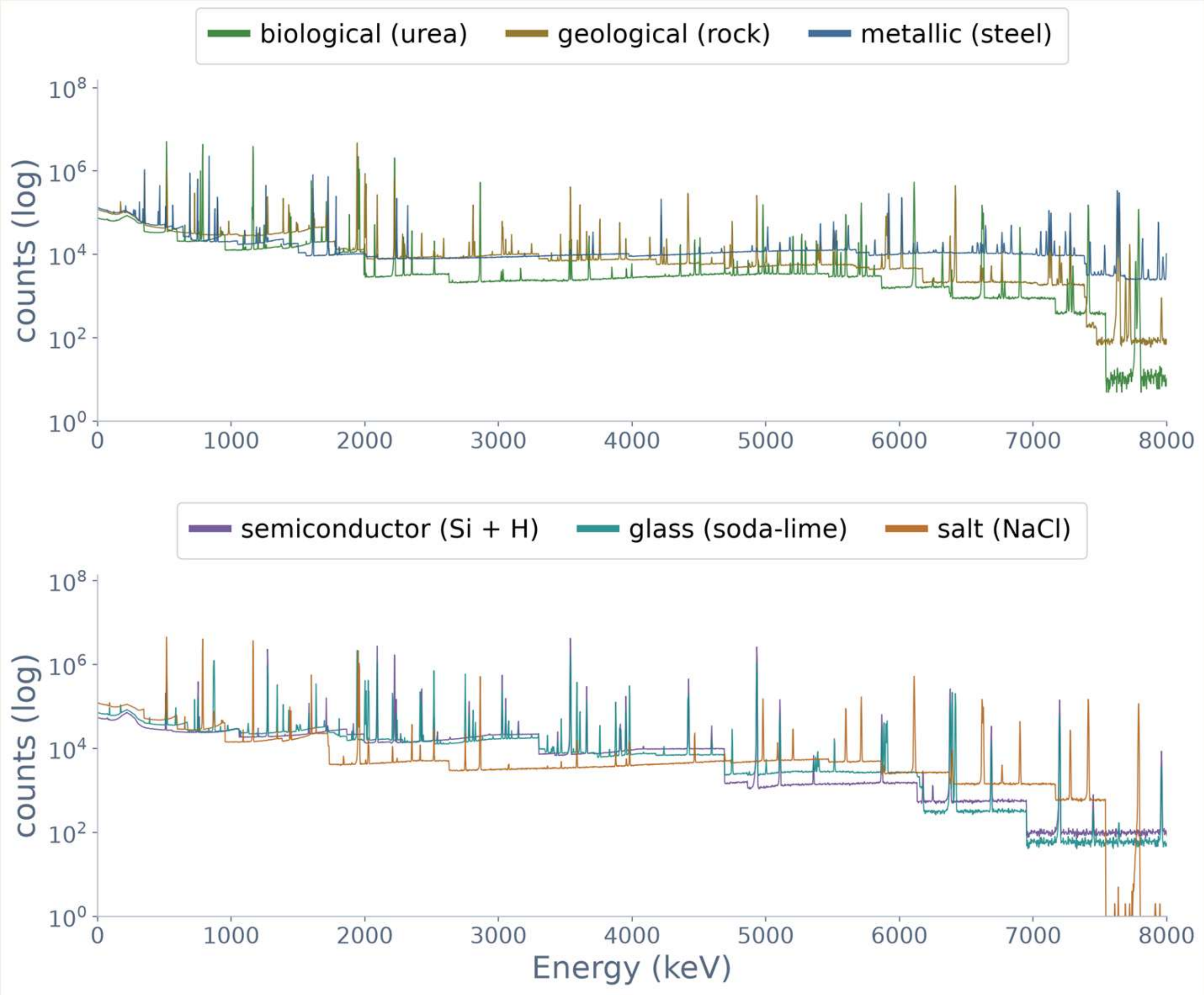
Why use PGAA:	Real-world Applications	Motivation
• Non-destructive, multi-element composition analysis • Highly penetrating neutrons/gammas • Advantages over mass spectrometry and x-ray fluorescence • Enables rapid classification and potentially quantitative analysis	• **Infrastructure**: Cl in concrete • **Geological/industrial materials**: soils, rocks, coal/mineral exploration • **Oil & gas**: on site elemental analysis in wells • **Security**: proposed identification of contraband and explosives	• Portable PGAA systems are technically capable of these applications • Creates a need for fast, reliable material classification and ultimately quantitative composition analysis

The Core Idea

- Training a model requires thousands of labeled spectra
- Time-consuming on hardware
- Software to generate training spectra
  - Use hardware just to validate if the synthetic spectra match reality

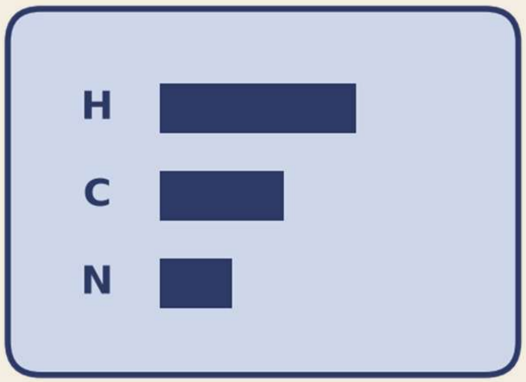
What I Built

- Synthetic spectrum generator
- Analysis engine: classifier, composition finder
- Desktop app (GUI)



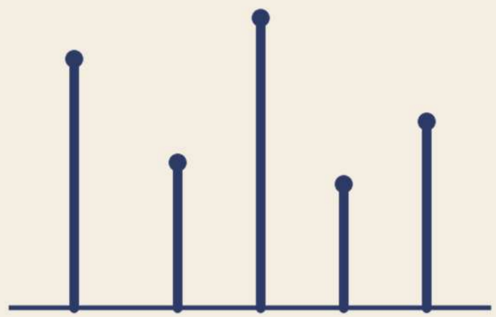
Generating Spectra

1. Pick a composition



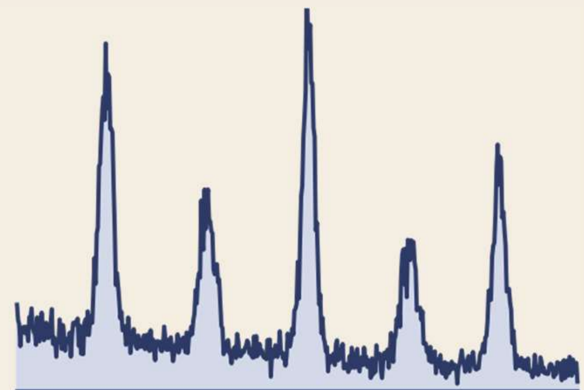
Which elements, how much

2. Add the gamma peaks



One per element line, sized by physics

3. Add realism + noise



Continuum, background, blur

x 9000 spectra

Add the Gamma Peaks

The clean spectrum is a matrix product: $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$

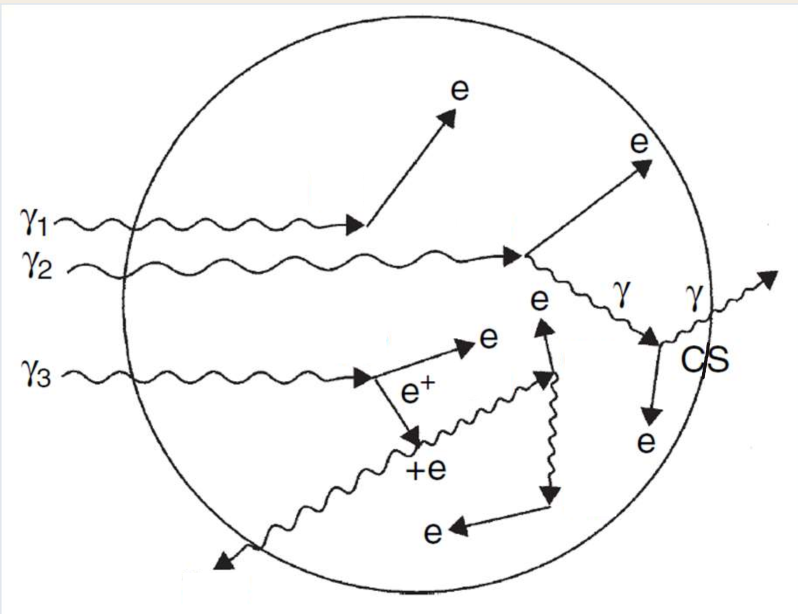
Counts at each energy = (how strongly the element emits there) x (how much of it)

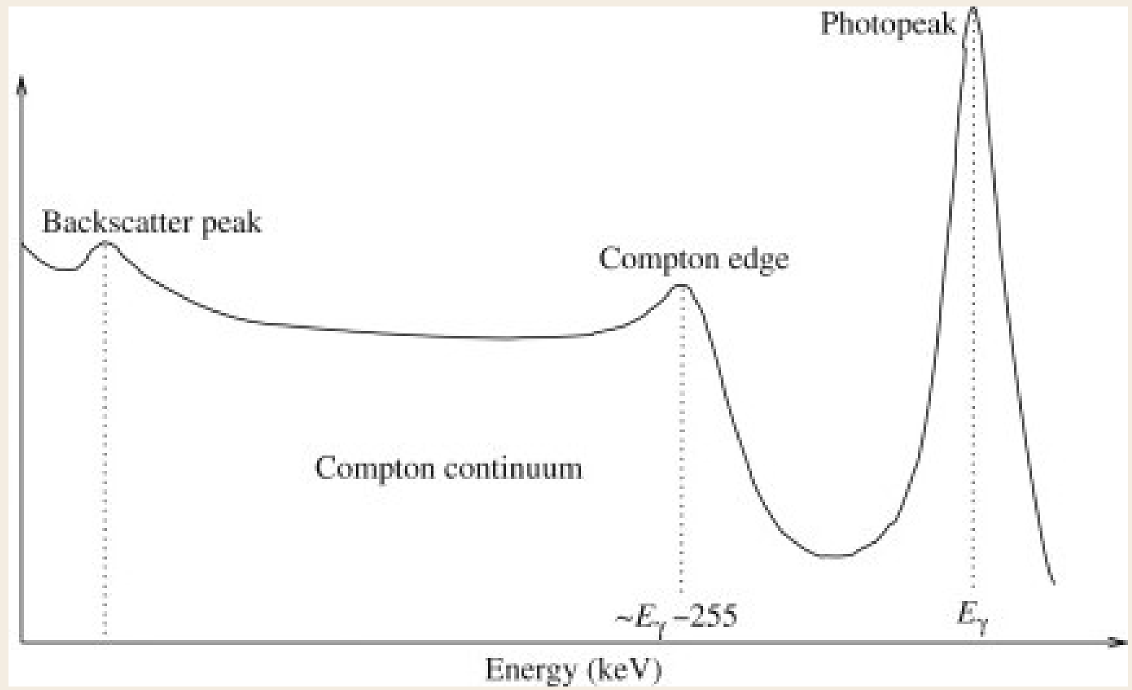
$$
\underset{\mathbf{A}}{\begin{array}{cc}
\text{Energy (keV)} & \begin{array}{ccc} \text{H} & \text{N} & \text{Cl} \end{array} \\
\begin{array}{c} 2223 \\ 1885 \\ 1951 \\ 5269 \end{array} & \begin{bmatrix} 10 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 30 \\ 0 & 3 & 2 \end{bmatrix}
\end{array}}
\cdot
\underset{\mathbf{x}}{\begin{array}{cc}
\begin{array}{c} \text{H} \\ \text{N} \\ \text{Cl} \end{array} & \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix}
\end{array}}
=
\underset{\mathbf{b}}{\begin{array}{cc}
\begin{bmatrix} 6.0 \\ 1.2 \\ 3.0 \\ 1.1 \end{bmatrix} & \begin{array}{c} 2223\ \text{keV} \\ 1885\ \text{keV} \\ 1951\ \text{keV} \\ 5269\ \text{keV} \end{array}
\end{array}}
$$

Composition (fractions)

Counts at each line

Add Realism

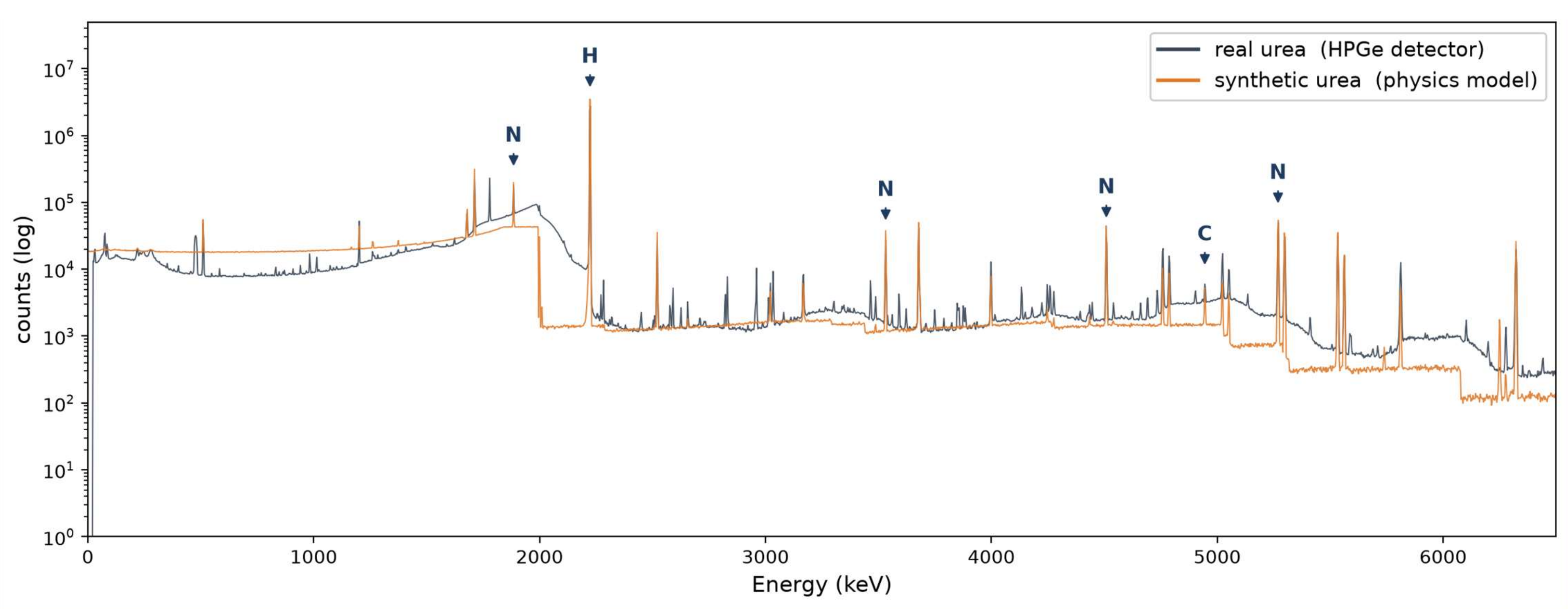




https://www.sciencedirect.com/topics/physics-and-astronomy/compton-scattering

$$compton\ edge = \frac{2E^2}{511 + 2E} \qquad continuum\ shape = \frac{d\sigma}{dT}$$

Synthetic vs Real Urea



The Classifier

- Algorithm: Random Forest
  - 2 RFs: fine and coarse
- Trained on 7200 spectra (80% of 9000)

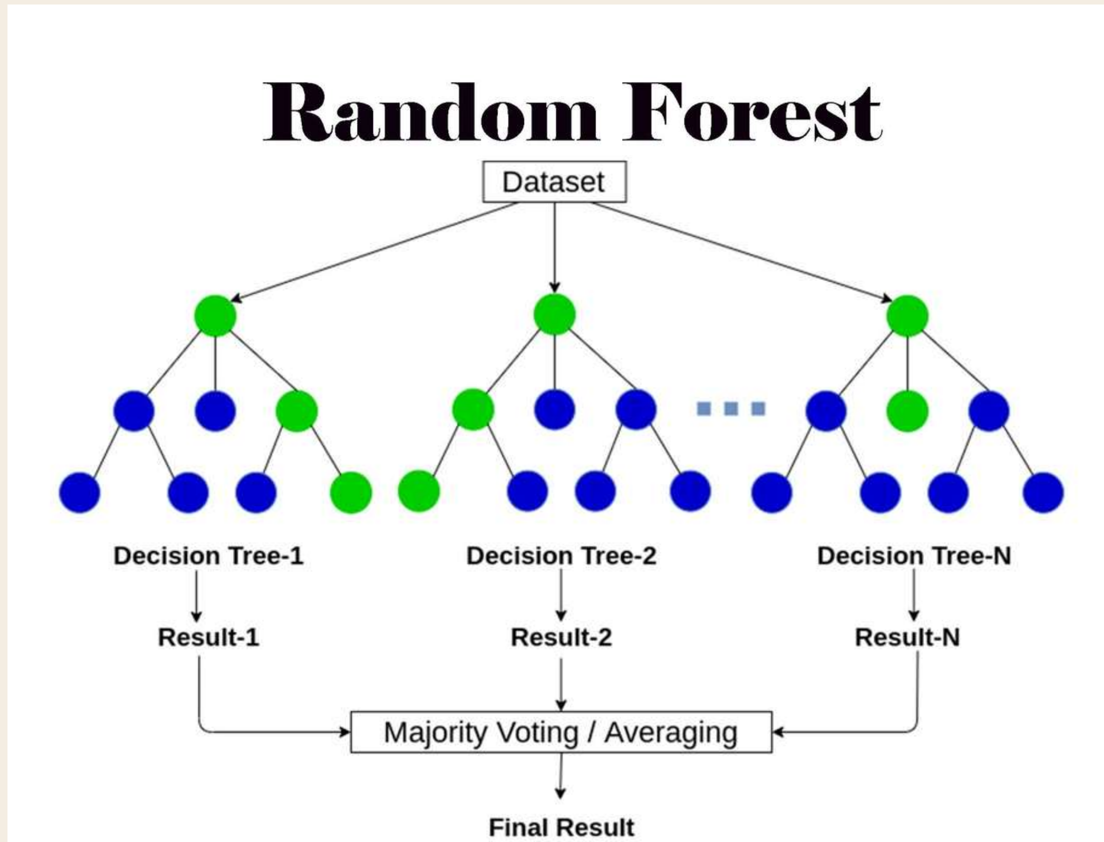
Input: the spectrum as 4,095 energy-bin counts

RF Classifier (coarse, 256 bins)

RF Classifier (fine, 4095 bins)

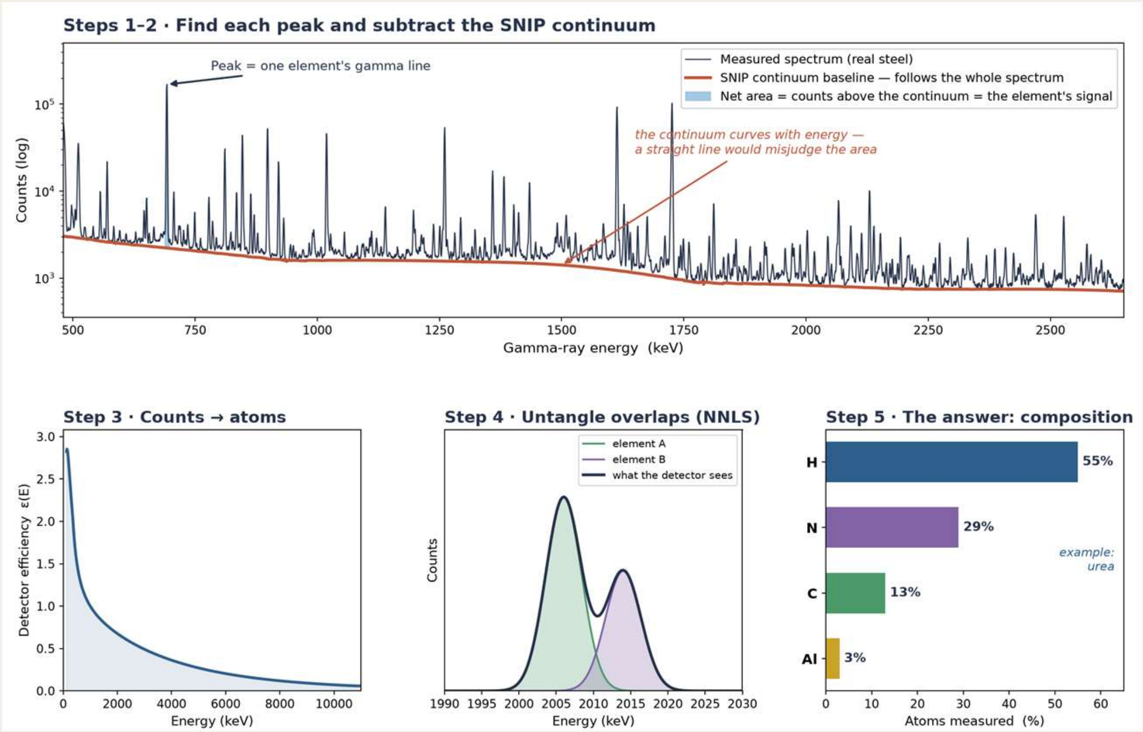
*Average predictions

Output: category



https://medium.com/@fraidoonomarzai99/random-forest-in-depth-f0556817c40b

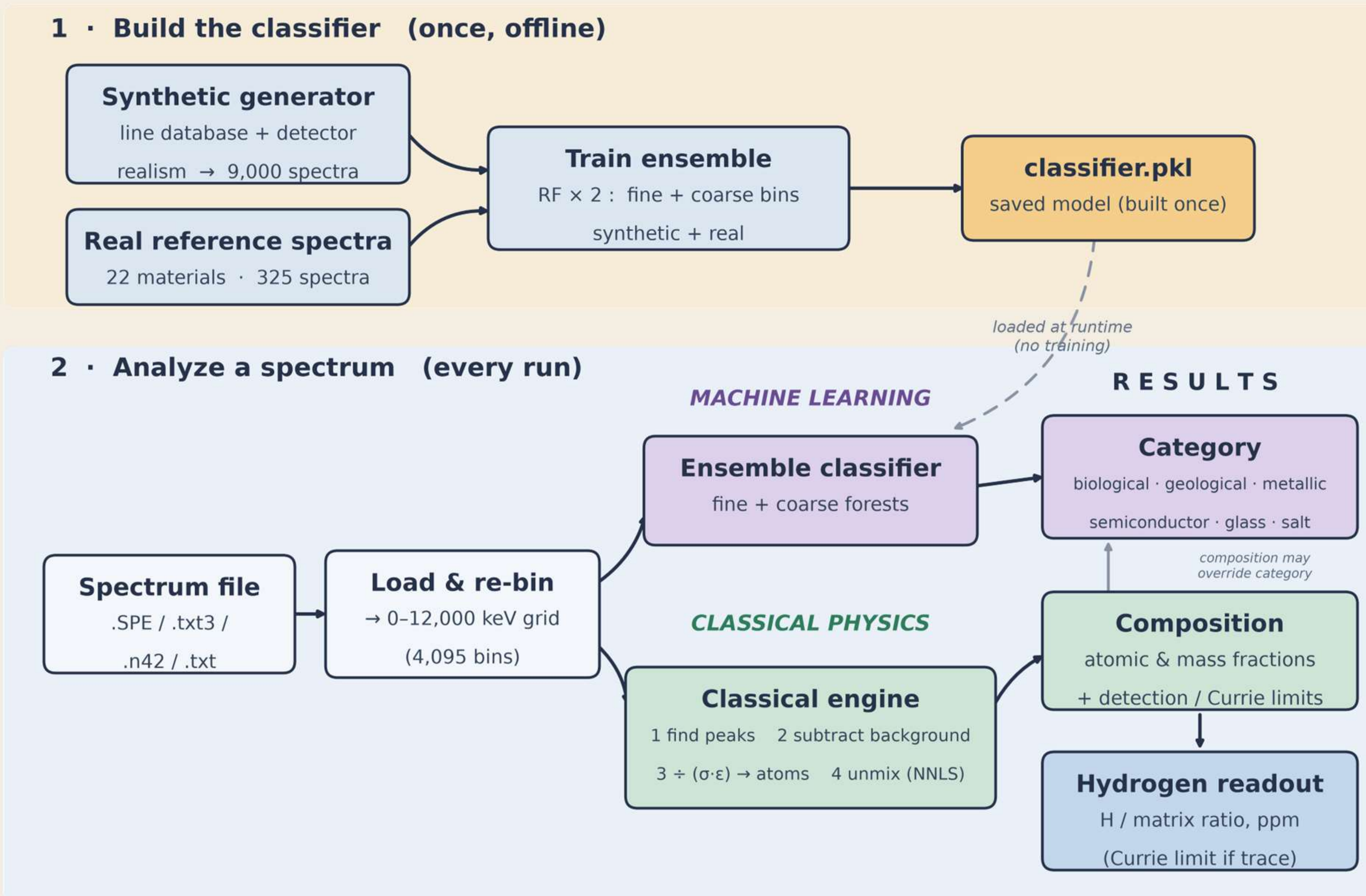
Composition finder (Classical method)



$$atoms \propto \frac{net\ peak\ area}{\sigma \times \varepsilon(E)}$$

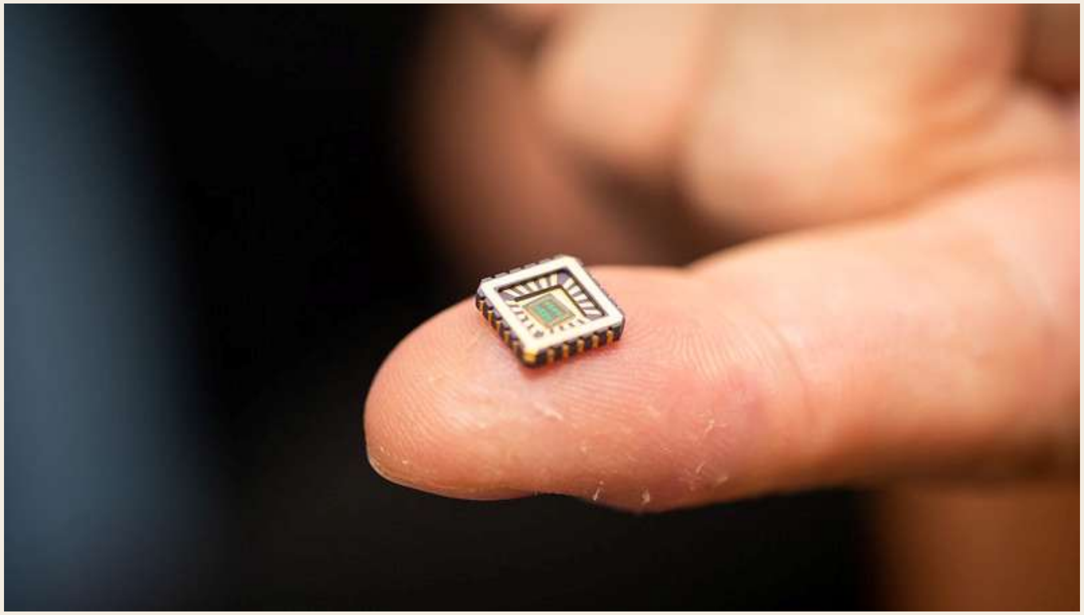
- σ = probability that a specific characteristic prompt gamma ray is emitted upon neutron capture
- $\varepsilon(E)$ = how well the detector catches the specific ray

Pipeline


1 · Build the classifier (once, offline)
Synthetic generator
line database + detector realism → 9,000 spectra
Real reference spectra
22 materials · 325 spectra
Train ensemble
RF × 2 : fine + coarse bins
synthetic + real
classifier.pkl
saved model (built once)
loaded at runtime (no training)
2 · Analyze a spectrum (every run)
MACHINE LEARNING
RESULTS
Spectrum file
.SPE / .txt3 / .n42 / .txt
Load & re-bin
→ 0–12,000 keV grid
(4,095 bins)
Ensemble classifier
fine + coarse forests
Category
biological · geological · metallic
semiconductor · glass · salt
composition may override category
CLASSICAL PHYSICS
Classical engine
1 find peaks 2 subtract background
3 ÷ (σ·ε) → atoms 4 unmix (NNLS)
Composition
atomic & mass fractions
+ detection / Currie limits
Hydrogen readout
H / matrix ratio, ppm
(Currie limit if trace)

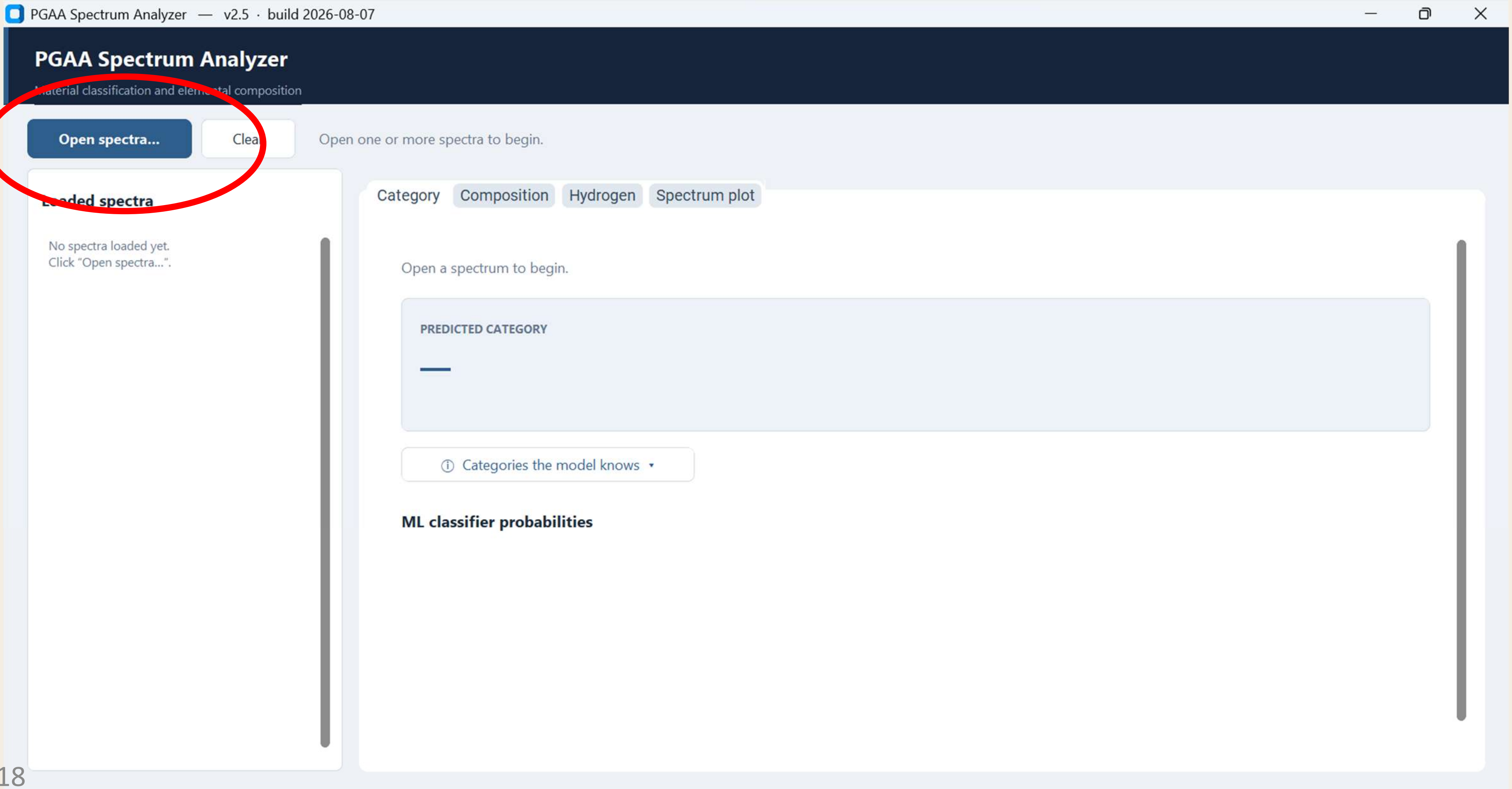
H-content

- Hydrogen degrades silicon chips
- Tool returns H : Si ratio
  - Atomic ppm
  - Weight ppm

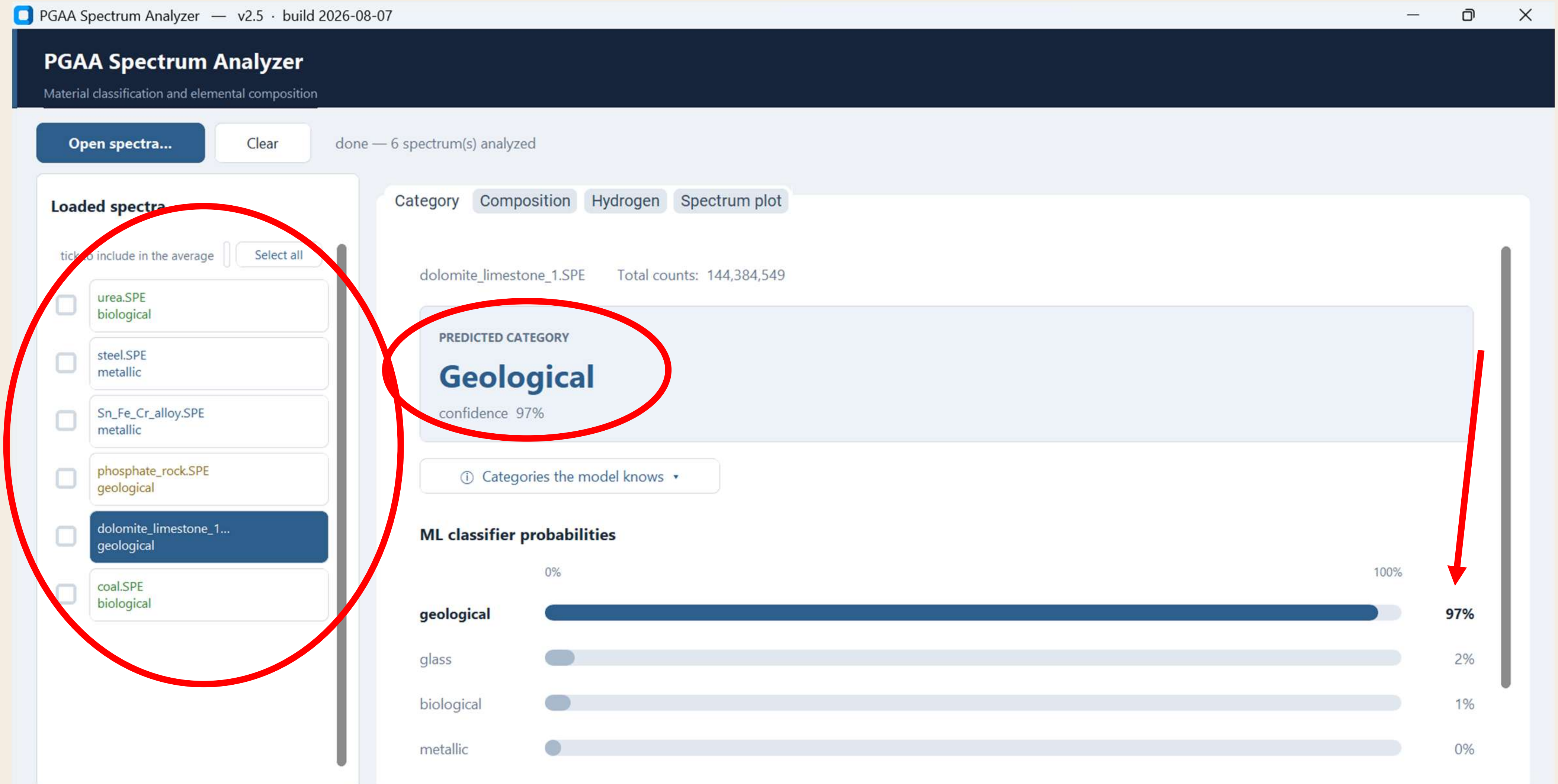


https://www.technologyreview.com/2019/12/04/131742/a-silicon-chip-that-mimics-the-brains-neurons-could-help-fight-paralysis/

Tkinter GUI

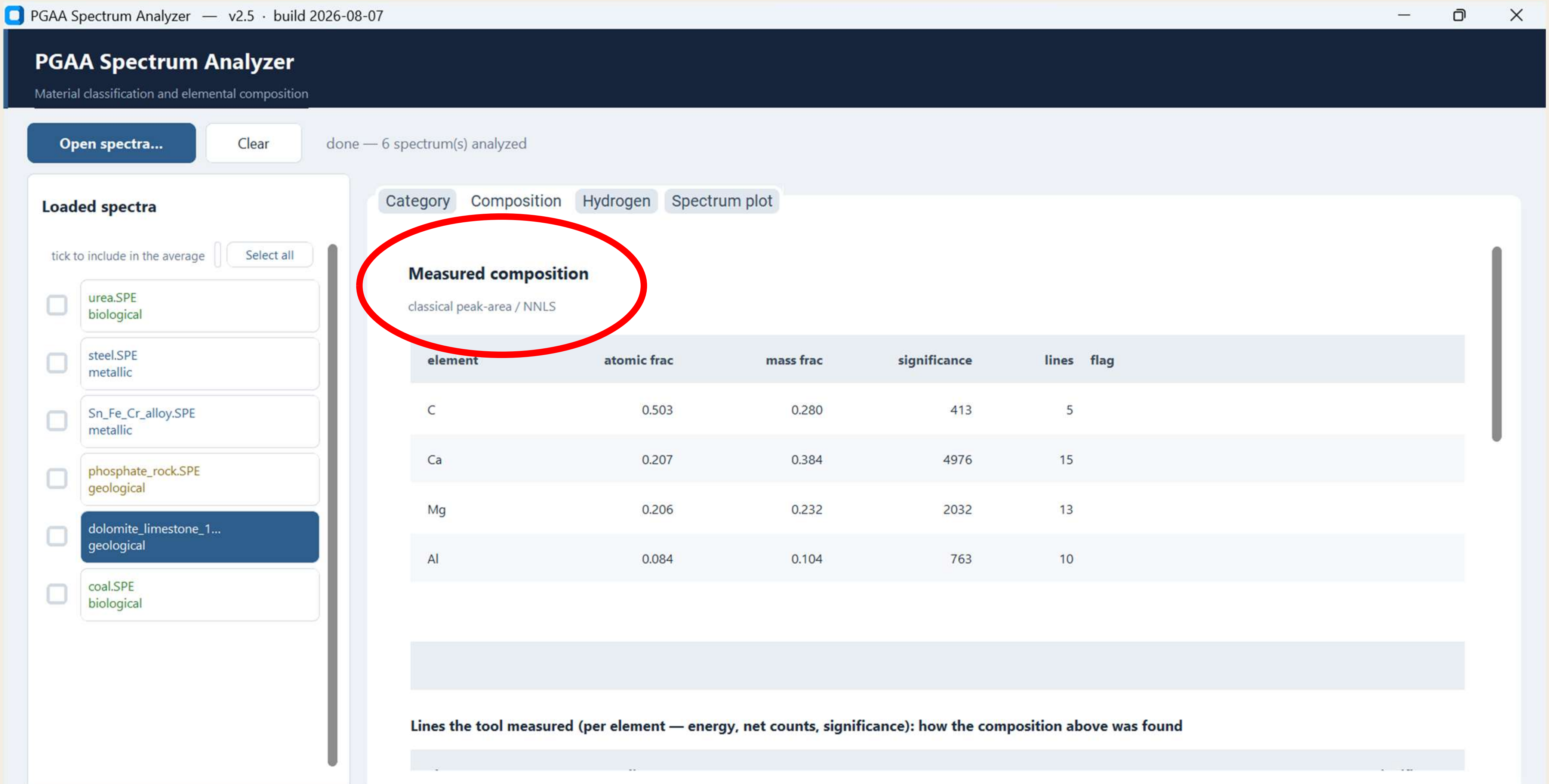

PGAA Spectrum Analyzer — v2.5 · build 2026-08-07
PGAA Spectrum Analyzer
Open spectra...
Open one or more spectra to begin.
No spectra loaded yet.
Click "Open spectra...".
Category
Composition
Hydrogen
Spectrum plot
Open a spectrum to begin.
PREDICTED CATEGORY
—
Categories the model knows
ML classifier probabilities

Tkinter GUI

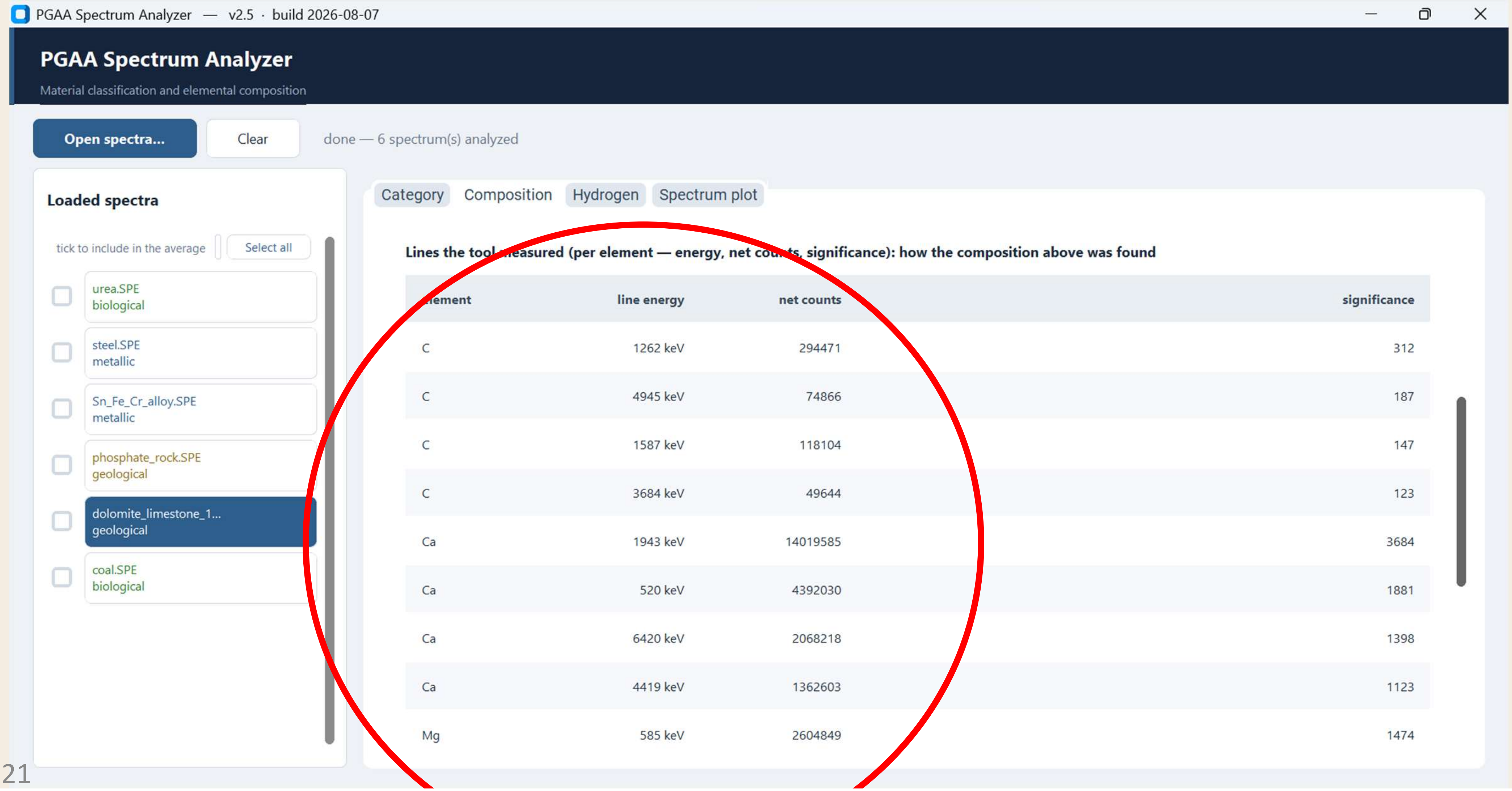


PGAA Spectrum Analyzer — v2.5 · build 2026-08-07
PGAA Spectrum Analyzer
Material classification and elemental composition
Open spectra...
Clear
done — 6 spectrum(s) analyzed
Loaded spectra
tick to include in the average
Select all
urea.SPE
biological
steel.SPE
metallic
Sn_Fe_Cr_alloy.SPE
metallic
phosphate_rock.SPE
geological
dolomite_limestone_1...
geological
coal.SPE
biological
Category
Composition
Hydrogen
Spectrum plot
dolomite_limestone_1.SPE
Total counts: 144,384,549
PREDICTED CATEGORY
Geological
confidence 97%
Categories the model knows
ML classifier probabilities
0%
100%
geological
97%
glass
2%
biological
1%
metallic
0%

Tkinter GUI

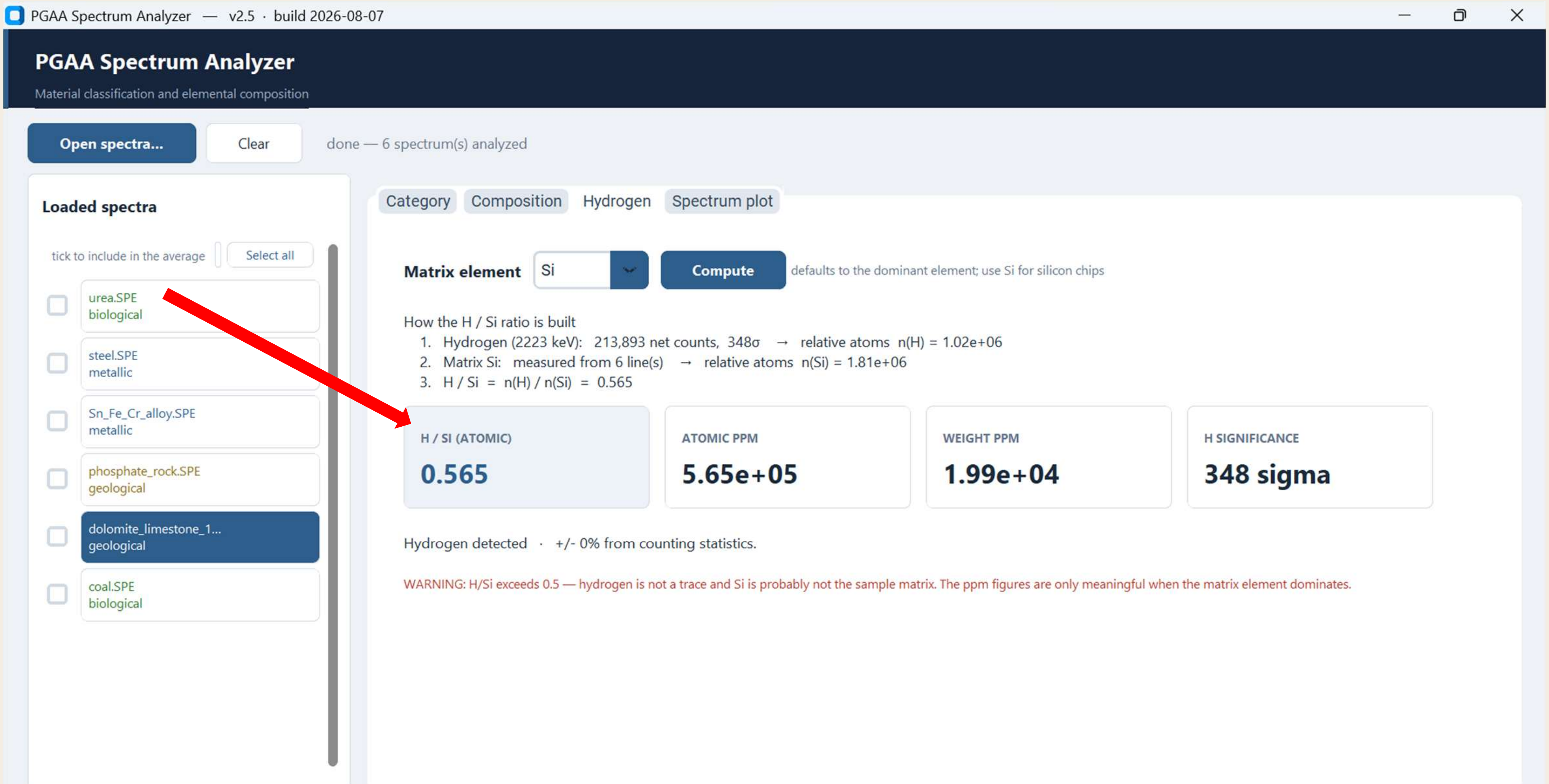

PGAA Spectrum Analyzer — v2.5 · build 2026-08-07
PGAA Spectrum Analyzer
Material classification and elemental composition
Open spectra...
Clear
done — 6 spectrum(s) analyzed
Loaded spectra
tick to include in the average
Select all
urea.SPE
biological
steel.SPE
metallic
Sn_Fe_Cr_alloy.SPE
metallic
phosphate_rock.SPE
geological
dolomite_limestone_1...
geological
coal.SPE
biological
Category
Composition
Hydrogen
Spectrum plot
Measured composition
classical peak-area / NNLS
element atomic frac mass frac significance lines flag
C 0.503 0.280 413 5
Ca 0.207 0.384 4976 15
Mg 0.206 0.232 2032 13
Al 0.084 0.104 763 10
Lines the tool measured (per element — energy, net counts, significance): how the composition above was found

Tkinter GUI

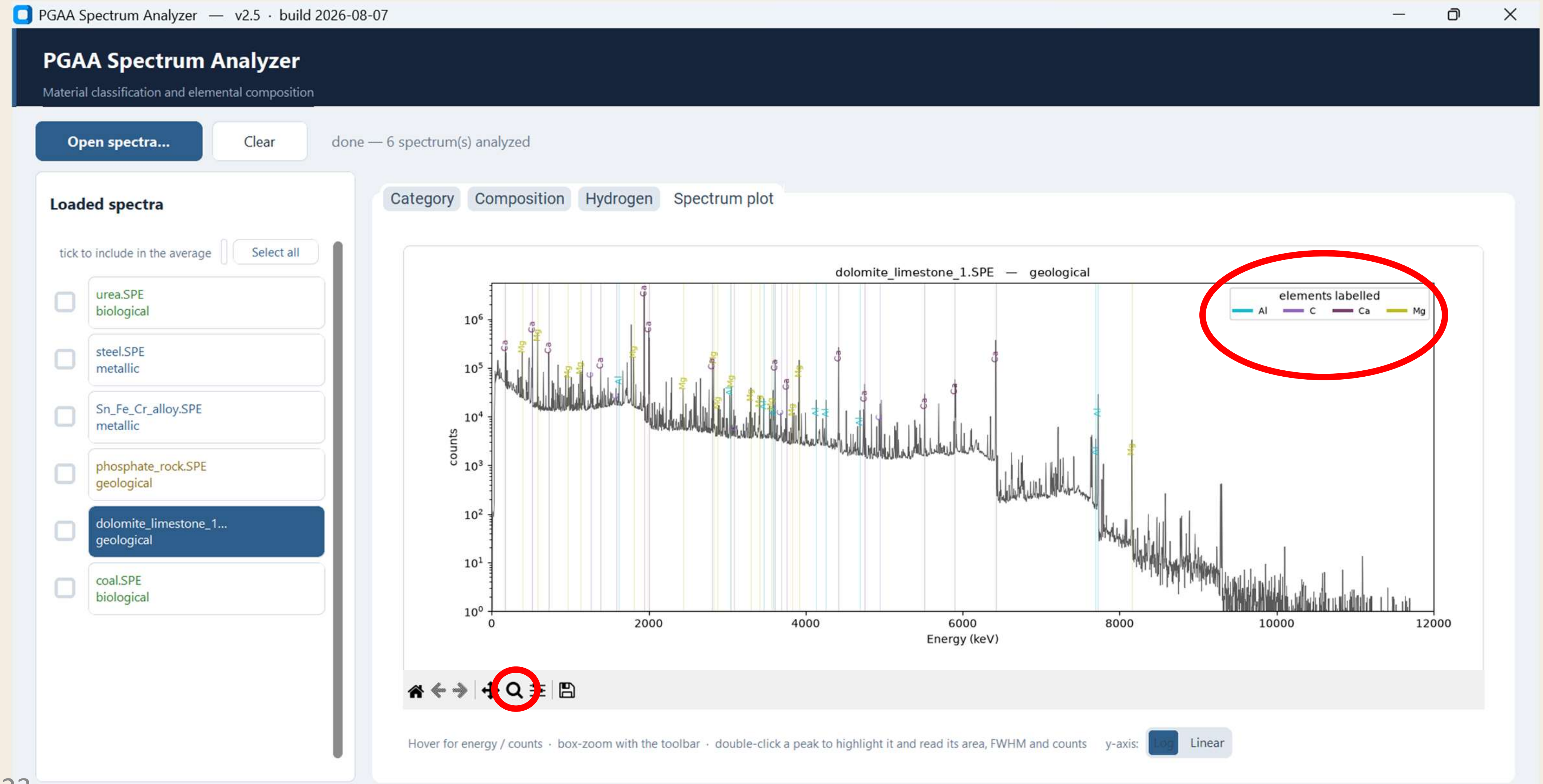


element	line energy	net counts	significance
C	1262 keV	294471	312
C	4945 keV	74866	187
C	1587 keV	118104	147
C	3684 keV	49644	123
Ca	1943 keV	14019585	3684
Ca	520 keV	4392030	1881
Ca	6420 keV	2068218	1398
Ca	4419 keV	1362603	1123
Mg	585 keV	2604849	1474

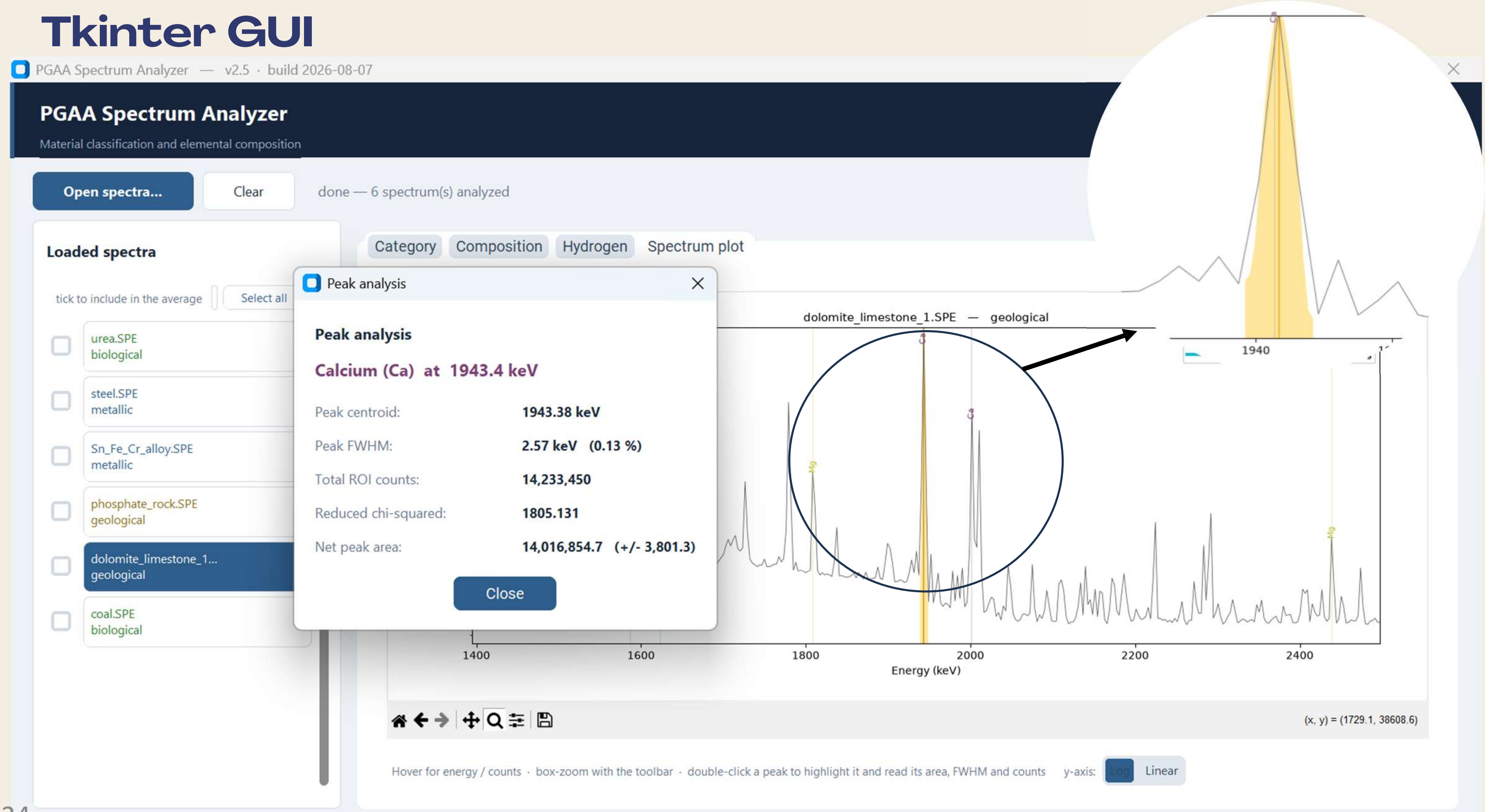
Tkinter GUI


PGAA Spectrum Analyzer — v2.5 · build 2026-08-07
PGAA Spectrum Analyzer
Material classification and elemental composition
Open spectra...
Clear
done — 6 spectrum(s) analyzed
Loaded spectra
tick to include in the average
Select all
urea.SPE
biological
steel.SPE
metallic
Sn_Fe_Cr_alloy.SPE
metallic
phosphate_rock.SPE
geological
dolomite_limestone_1...
geological
coal.SPE
biological
Category
Composition
Hydrogen
Spectrum plot
Matrix element
Si
Compute
defaults to the dominant element; use Si for silicon chips
How the H / Si ratio is built
1. Hydrogen (2223 keV): 213,893 net counts, 348σ → relative atoms n(H) = 1.02e+06
2. Matrix Si: measured from 6 line(s) → relative atoms n(Si) = 1.81e+06
3. H / Si = n(H) / n(Si) = 0.565
H / SI (ATOMIC)
0.565
ATOMIC PPM
5.65e+05
WEIGHT PPM
1.99e+04
H SIGNIFICANCE
348 sigma
Hydrogen detected · +/- 0% from counting statistics.
WARNING: H/Si exceeds 0.5 — hydrogen is not a trace and Si is probably not the sample matrix. The ppm figures are only meaningful when the matrix element dominates.

Tkinter GUI


PGAA Spectrum Analyzer — v2.5 · build 2026-08-07
PGAA Spectrum Analyzer
Material classification and elemental composition
Open spectra...
Clear
done — 6 spectrum(s) analyzed
Loaded spectra
tick to include in the average
Select all
urea.SPE
biological
steel.SPE
metallic
Sn_Fe_Cr_alloy.SPE
metallic
phosphate_rock.SPE
geological
dolomite_limestone_1...
geological
coal.SPE
biological
Category
Composition
Hydrogen
Spectrum plot
dolomite_limestone_1.SPE — geological
elements labelled
Al
C
Ca
Mg
counts
Energy (keV)
Hover for energy / counts · box-zoom with the toolbar · double-click a peak to highlight it and read its area, FWHM and counts
y-axis:
Log
Linear

Tkinter GUI



PGAA Spectrum Analyzer — v2.5 · build 2026-08-07
PGAA Spectrum Analyzer
Material classification and elemental composition
Open spectra...
Clear
done — 6 spectrum(s) analyzed
Loaded spectra
tick to include in the average
Select all
urea.SPE
biological
steel.SPE
metallic
Sn_Fe_Cr_alloy.SPE
metallic
phosphate_rock.SPE
geological
dolomite_limestone_1...
geological
coal.SPE
biological
Category
Composition
Hydrogen
Spectrum plot
Peak analysis
Peak analysis
Calcium (Ca) at 1943.4 keV
Peak centroid: 1943.38 keV
Peak FWHM: 2.57 keV (0.13 %)
Total ROI counts: 14,233,450
Reduced chi-squared: 1805.131
Net peak area: 14,016,854.7 (+/- 3,801.3)
Close
dolomite_limestone_1.SPE — geological
1400
1600
1800
2000
2200
2400
Energy (keV)
1940
(x, y) = (1729.1, 38608.6)
Hover for energy / counts · box-zoom with the toolbar · double-click a peak to highlight it and read its area, FWHM and counts
y-axis:
Log
Linear

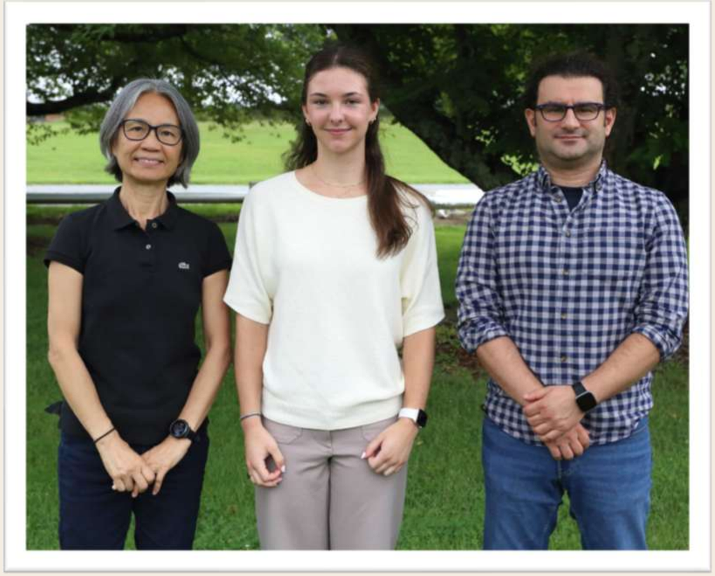
Conclusion and Future Work

- **Conclusion**
  - Built a PGAA tool that does two things from one gamma spectrum: an ML classifier **names the material class**, and a classical physics engine **measures the elemental composition**.
  - All my original code.
  - Trained entirely on **9,000 physics-based synthetic spectra**, then tested on **real reactor spectra it never saw.**
- **Results:**
  - Classifies **5 of 6** real reference materials correctly, across **6 classes** (biological, geological, metallic, semiconductor, glass, salt).
  - Recovers urea's **H / C / N composition to a mean absolute error of 0.009** vs. the known formula.
- **Future Work**
  - Improve the **high-energy continuum** (multiple-scatter modeling)
  - Extend to **other detectors** (user-supplied efficiency curve)
  - Deepen the line library

Summer 2026

THANK YOU

Any questions?



ML Quantifier → Classical Method

Attempt #:

1. RF on raw bins
2. RF on peak area features
3. Classical peak-area / NNLS engine

$$\text{skill} = \frac{\text{error of the class-average guess} - \text{error of the model}}{\text{error of the class-average guess}}$$

Skill > 0 → model is reading the spectrum

Skill ≈ 0 → reciting class average

Skill < 0 → doing worse than the class-average guess

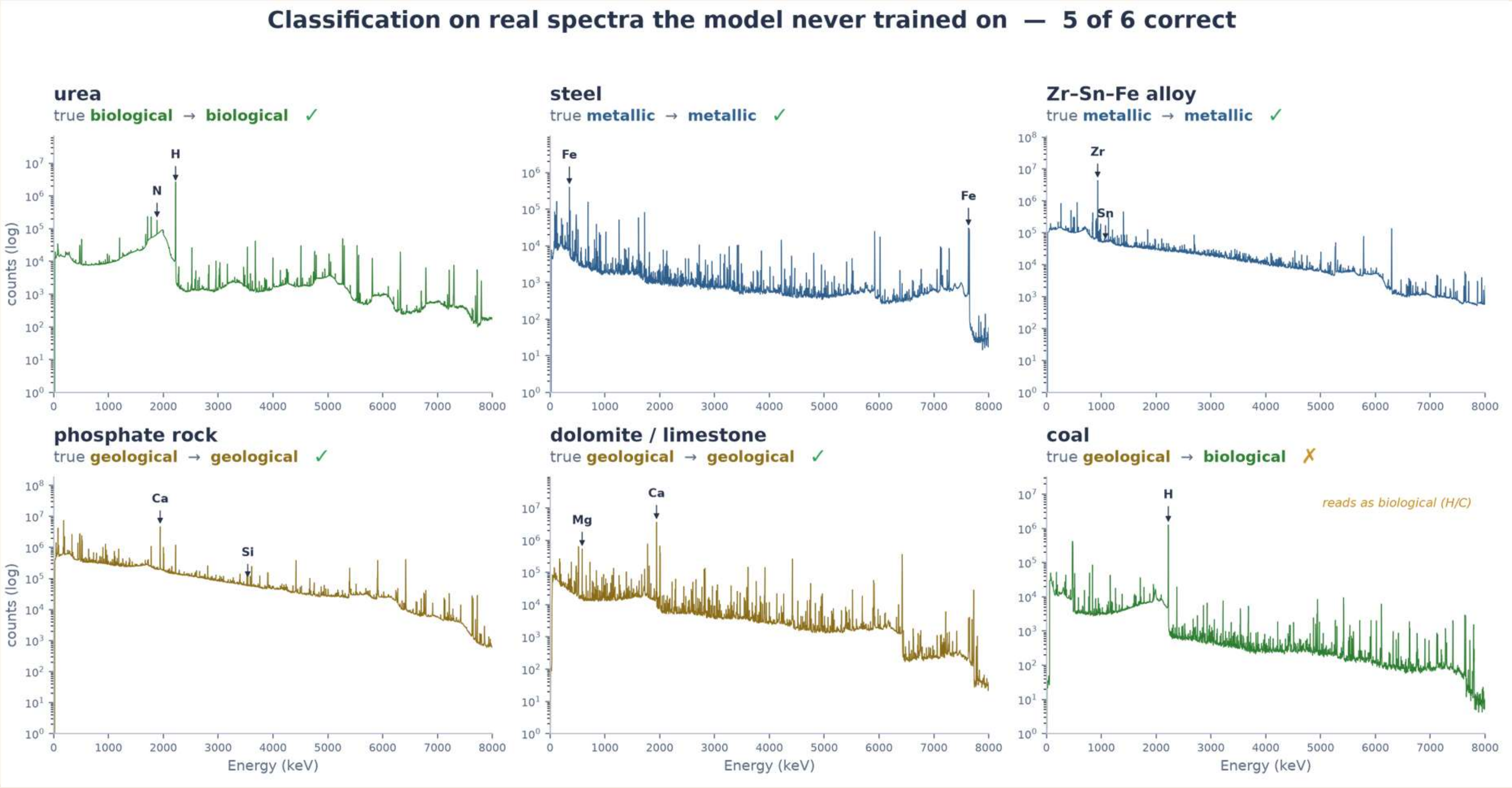
Approach	Classifies (real)	Urea skill	Phosphate skill
Classical (peak-area / NNLS)	4/5	**+0.88**	**+0.29**
RF on peak-area features	2/5	+0.27	**−0.37**
RF on raw bins	4/5	**−0.01**	**−0.39**

PG Line Library

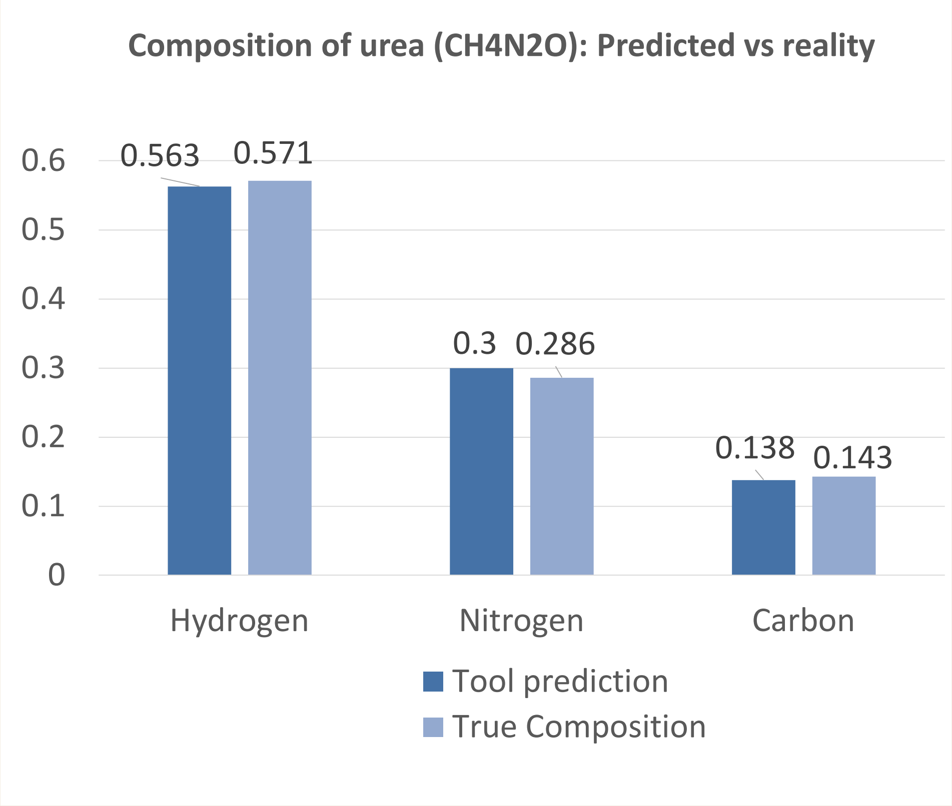
	A	B	C	D	E	F	G	H	I	J	K	L
1	A	Isotope	Z	T	E(γ)		Sigma			K_0		
2					Text	Value	Text	Value	Uncertainty	Text	Value	Uncertainty
3	1	2-H	1	p	2223.24835(9)	2223.248	0.3326(7)	0.3326	7.00E-04	1.0000(21)	1.00E+00	2.10E-03
4	2	3-H	1	p	6250.243(3)	6250.243	0.000519(7)	0.000519	7.00E-06	0.001560(21)	1.56E-03	2.10E-05
5	3	4-He	2	p	20520.46	20520.46	4.2E-11(12)	4.20E-11	1.20E-11	3.2E-11(9)	3.18E-11	9.09E-12
6	6	7-Li	3	p	477.595(3)	477.595	0.00153(8)	0.00153	8.00E-05	0.00067(4)	6.68E-04	3.49E-05
7	6	7-Li	3	p	6768.81(4)	6768.81	0.00151(9)	0.00151	9.00E-05	0.00066(4)	6.59E-04	3.93E-05
8	6	7-Li	3	p	7245.91(4)	7245.91	0.00247(14)	0.00247	1.40E-04	0.00108(6)	1.08E-03	6.11E-05
9	7	8-Li	3	p	980.53(7)	980.53	0.00415(13)	0.00415	1.30E-04	0.00181(6)	1.81E-03	5.68E-05
10	7	8-Li	3	p	1051.90(7)	1051.9	0.00414(12)	0.00414	1.20E-04	0.00181(5)	1.81E-03	5.24E-05
11	7	8-Li	3	p	2032.30(4)	2032.3	0.0381(8)	0.0381	8.00E-04	0.0166(4)	1.66E-02	3.49E-04
12	9	10-Be	4	p	219.40(10)	219.4	4.5E-06(6)	0.0000045	6.00E-07	1.51E-06(20)	1.51E-06	2.02E-07
13	9	10-Be	4	p	304.24(13)	304.24	2.3E-06(5)	0.0000023	5.00E-07	7.7E-07(17)	7.73E-07	1.68E-07
14	9	10-Be	4	p	547.55(4)	547.55	1.09E-05(10)	0.0000109	1.00E-06	3.7E-06(3)	3.67E-06	3.36E-07
15	9	10-Be	4	p	631.92(4)	631.92	1.8E-05(4)	0.000018	4.00E-06	6.1E-06(13)	6.05E-06	1.35E-06
16	9	10-Be	4	p	853.630(12)	853.63	0.00208(24)	0.00208	2.40E-04	0.00070(8)	6.99E-04	8.07E-05
17	9	10-Be	4	p	2590.014(19)	2590.014	0.00191(15)	0.00191	1.50E-04	0.00064(5)	6.42E-04	5.04E-05
18	9	10-Be	4	p	2811.68(5)	2811.68	1.05E-05(12)	0.0000105	1.20E-06	3.5E-06(4)	3.53E-06	4.04E-07
19	9	10-Be	4	p	2896.02(4)	2896.02	1.16E-05(13)	0.0000116	1.30E-06	3.9E-06(4)	3.90E-06	4.37E-07
20	9	10-Be	4	p	3367.448(25)	3367.448	0.00285(22)	0.00285	2.20E-04	0.00096(7)	9.58E-04	7.40E-05
21	9	10-Be	4	p	3443.406(20)	3443.406	0.00098(7)	0.00098	7.00E-05	0.000330(24)	3.30E-04	2.35E-05
22	9	10-Be	4	p	5956.53(3)	5956.53	1.46E-04(12)	0.000146	1.20E-05	4.9E-05(4)	4.91E-05	4.04E-06
23	9	10-Be	4	p	5958.85(12)	5958.85	6.8E-06(8)	0.0000068	8.00E-07	2.3E-06(3)	2.29E-06	2.69E-07
24	9	10-Be	4	p	6809.61(3)	6809.61	0.0058(5)	0.0058	5.00E-04	0.00195(17)	1.95E-03	1.68E-04
25	10	11-B	5	p	477.595(3)	477.595	716(25)	716	2.50E+01	201(7)	2.01E+02	7.01E+00
26	10	11-B	5	p	2296.63(13)	2296.63	0.0054(12)	0.0054	1.20E-03	0.0015(3)	1.51E-03	3.36E-04
27	10	11-B	5	p	2533.35(18)	2533.35	0.0083(12)	0.0083	1.20E-03	0.0023(3)	2.33E-03	3.36E-04
28	10	11-B	5	p	4444.03(11)	4444.03	0.039(3)	0.039	3.00E-03	0.0109(8)	1.09E-02	8.41E-04

Prompt-gamma library used throughout project. We only look at the top 15 lines by cross section (σ) for each element.

Classification Results



Urea Comp


Composition of urea (CH4N2O): Predicted vs reality
0.6
0.5
0.4
0.3
0.2
0.1
0
0.563
0.571
0.3
0.286
0.138
0.143
Hydrogen
Nitrogen
Carbon
Tool prediction
True Composition

Composition Results

Spectrum	Actual composition	Tool prediction	Accuracy
urea (CH_4N_2O)	H 0.57 N 0.29 C 0.14	H 0.56 N 0.30 C 0.14	99%
steel	Fe 0.97 Cr 0.02 Ni 0.01	C 0.72 Fe 0.26 N 0.02	26%
Zr-Sn-Fe alloy	Zr 0.98 Sn 0.01 Fe 0.01	Zr 0.95 Sn 0.03	97%
phosphate rock (apatite)	Ca 0.62 P 0.38	Ca 0.36 P 0.24 H 0.17 Al 0.13	60%
dolomite ($CaMg(CO_3)_2$)	C 0.50 Ca 0.25 Mg 0.25	C 0.50 Ca 0.21 Mg 0.21 Al 0.08	91%
coal (organic)	C 0.60 H 0.40	C 0.53 H 0.45 N 0.02	93%