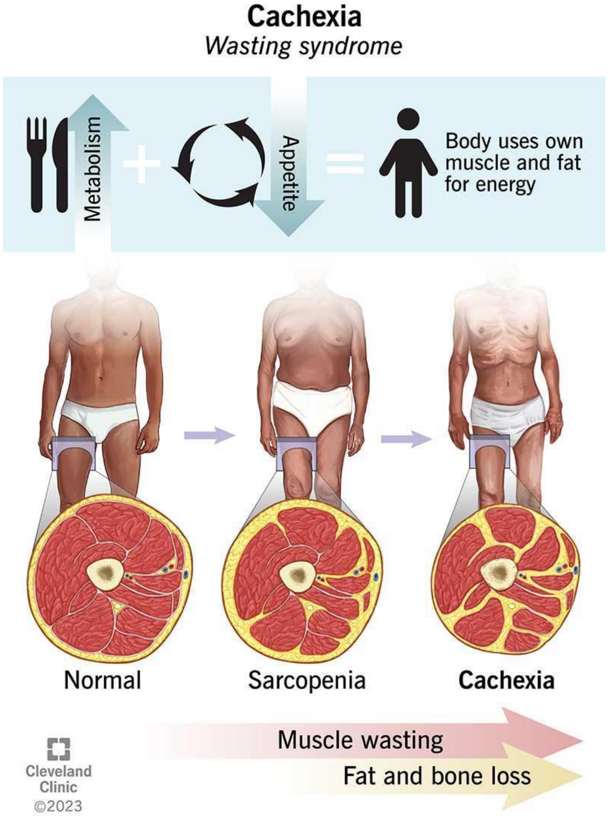
Stage Classification of Cancer-Associated Cachexia Through Hidden Markov Models: Exploratory Investigation

Lizelle Ocfemia, Sam Keene, Ed Reznik

Memorial Sloan Kettering Cancer Center

THE COOPER UNION

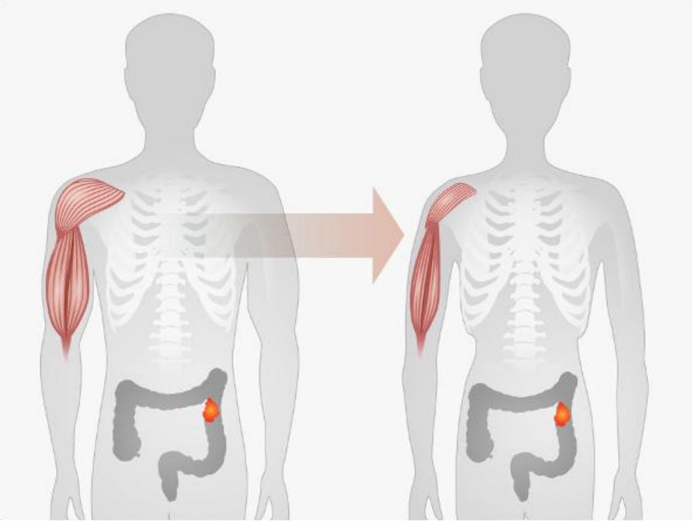
What is Cachexia?



ca·chex·i·a
/kəˈkeksēə/

Cancer-Associated Cachexia

- Estimated to occur in up to 80% of people with advanced cancer
- Cause of death for ~20% of all cancer patients
- Decreased quality of life
- Poor response to chemotherapy



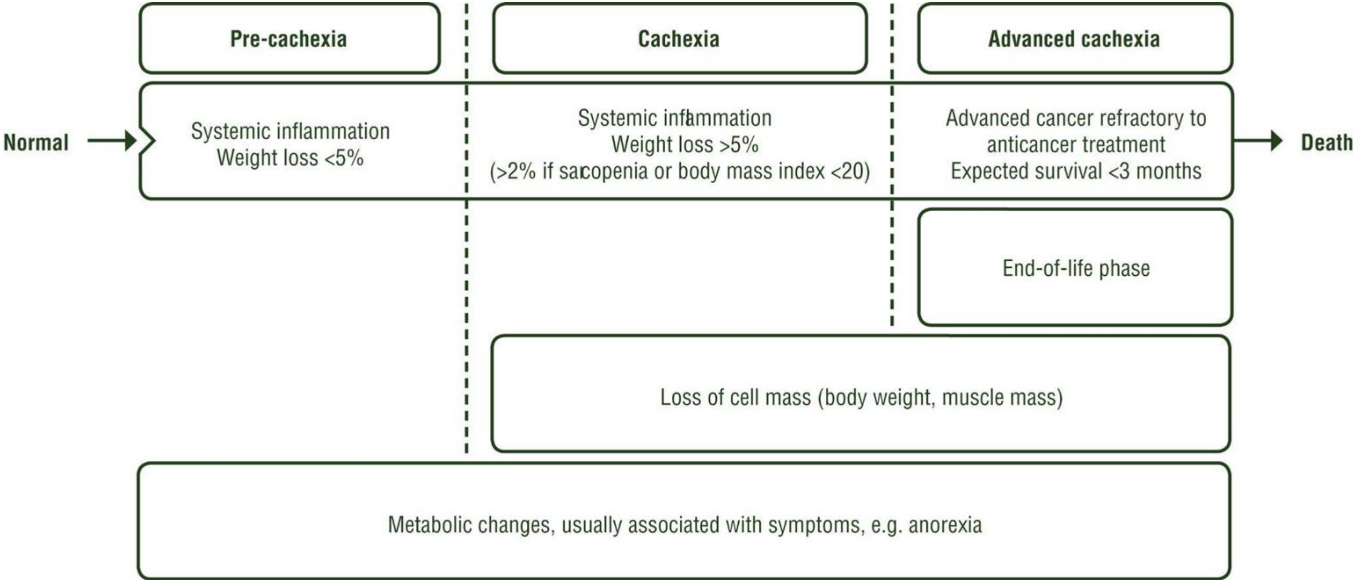
Credit: National Cancer Institute

Staging Criteria

Either of the following conditions:

- **Weight loss>5% over past 6 months** (in absence of simple starvation);
- **BMI<20** and any degree of weight loss>2%;
- Appendicular skeletal muscle index consistent with sarcopenia (male<7.26kg/m2; female<5.45kg/m2) and any degree of weight loss>2%

Fearon et al (2011)



Normal → Pre-cachexia → Cachexia → Advanced cachexia → Death

Pre-cachexia
Systemic inflammation
Weight loss <5%

Cachexia
Systemic inflammation
Weight loss >5%
(>2% if sarcopenia or body mass index <20)

Advanced cachexia
Advanced cancer refractory to anticancer treatment
Expected survival <3 months

End-of-life phase

Loss of cell mass (body weight, muscle mass)

Metabolic changes, usually associated with symptoms, e.g. anorexia

Problem + Motivation

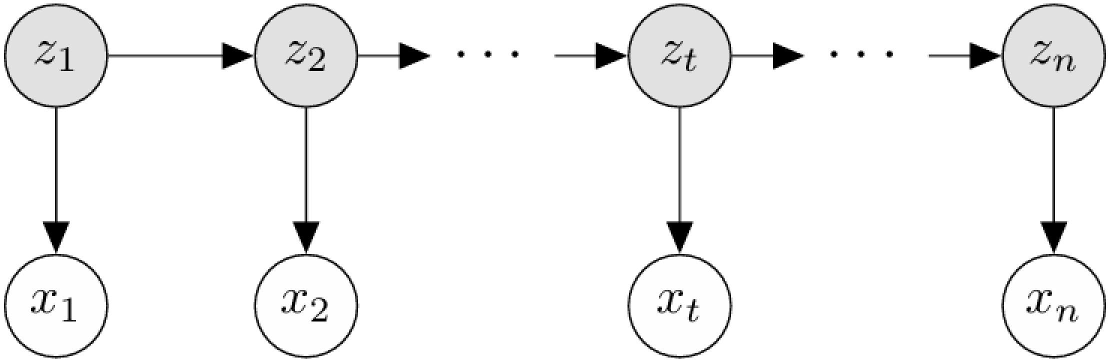
- Mechanisms of cachexia are still unclear.
- Symptoms are similar to those caused by cancer treatments or other issues like malnutrition.
- No direct treatments to stop/reverse CC.
- Early detection of CC can enable interventions to slow muscle wasting, improve metabolic function, and enhance the patient's quality of life.

- Multimodal deep learning-based model on skeletal muscle area in CT scans.
- Various machine learning models based on *static* data, not longitudinal.
- *Supervised* classifiers.

Can Hidden Markov Models classify stages of cancer-associated cachexia from longitudinal biomarkers?

Why use Hidden Markov Models?

- Used in speech recognition, gene finding, gesture tracking, etc.



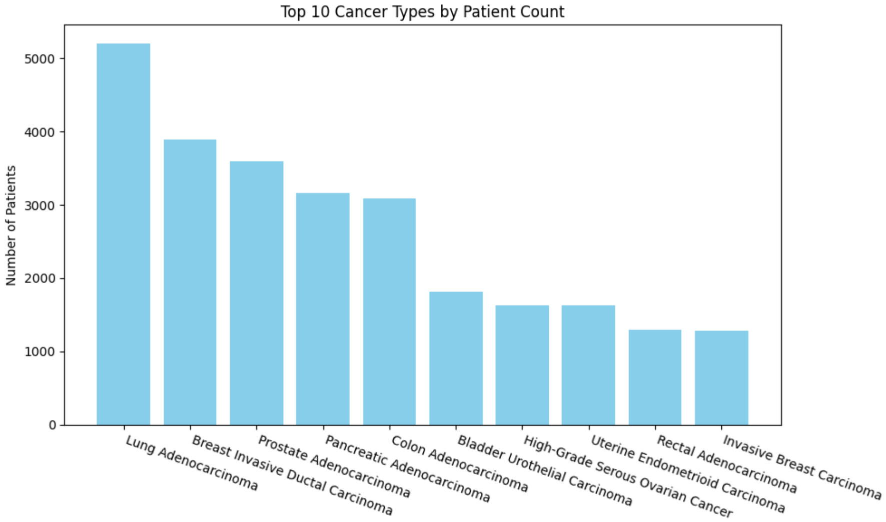
Latent state of CC at time *t*

Observed features of patient at time *t*

Data

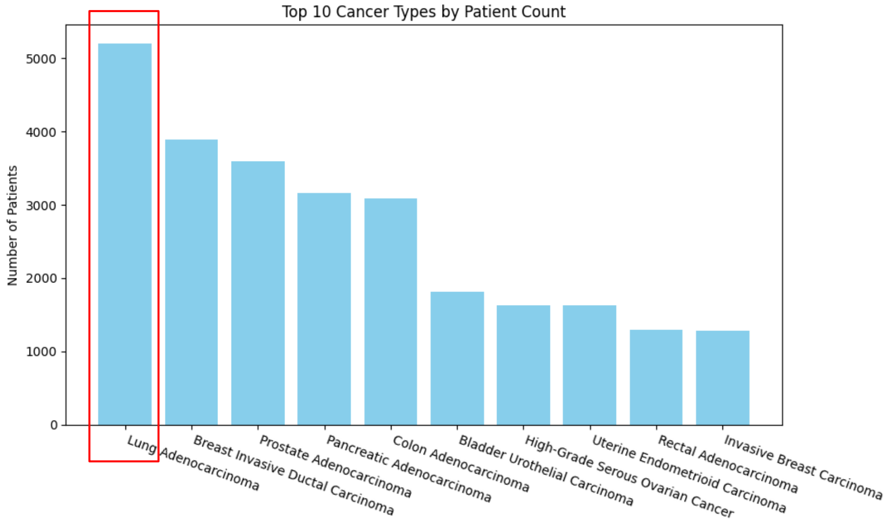
MSK-IMPACT Cohort

- MSK-IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets) is a targeted tumor-sequencing test available to MSK patients.
- 63,008 patients in the cohort with recorded body mass index (BMI) and blood lab values.



MSK-IMPACT Cohort

- MSK-IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets) is a targeted tumor-sequencing test available to MSK patients.
- 63,008 patients in the cohort with recorded body mass index (BMI) and blood lab values.



Blood Lab Values

Metabolic & Organ Function

- *Albumin, Total Protein*
    - Nutritional status
- *ALT, AST, ALK, Bilirubin*
    - Liver function
- *Creatinine, BUN, CO_2*
    - Kidney & acid-base balance
- *Glucose, Anion Gap, Electrolytes (Na, K, Cl, Ca)*: Metabolism & fluid balance

Platelets

- *Platelet count*
    - Clotting and inflammation

Immune & Inflammatory Markers

- *WBC, Neutrophils, Lymphocytes, Monocytes, Eosinophils, Basophils*
    - *Major immune cell types*
- *Absolute counts & Immature Granulocytes*:
    - Infection & immune activity

Red Blood Cell Indices (Anemia)

- *HGB, HCT, RBC*
    - Oxygen transport
- *MCV, MCH, MCHC, RDW*
    - Red cell size & content

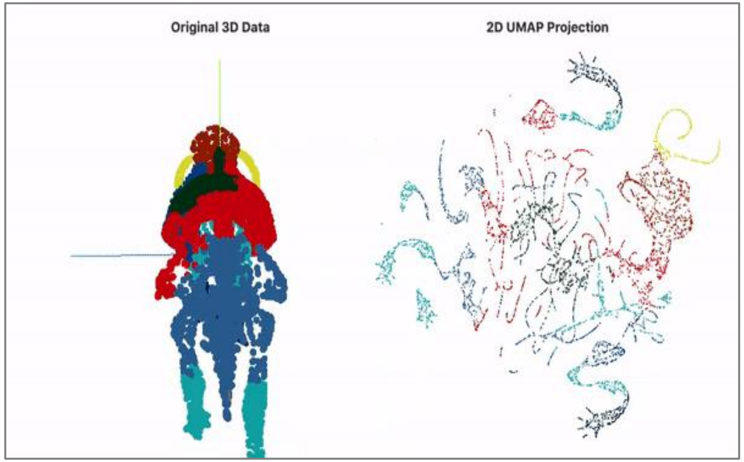
Data Preprocessing

1. Filter to include only patients diagnosed with **Lung Adenocarcinoma**
2. Filter to include only **deceased** patients
3. Filter to patients with **>20 visits**

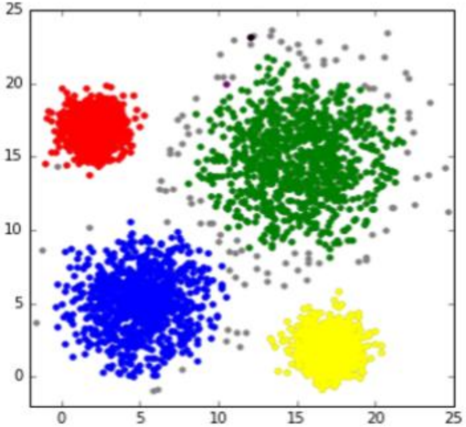
796 patients x 37 features (blood lab values), 20-234 visits

But how do we decide the number of states to choose for our HMM?

Clustering Algorithm



Uniform Manifold Approximation and Projection (UMAP)



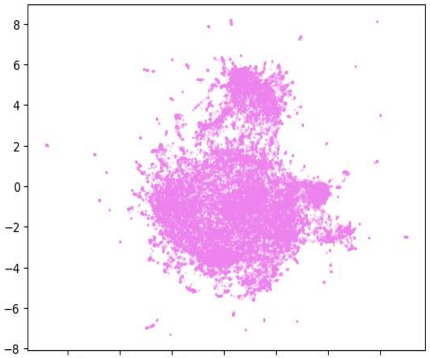
Hierarchical DBSCAN (HDBSCAN)

Clustering Results

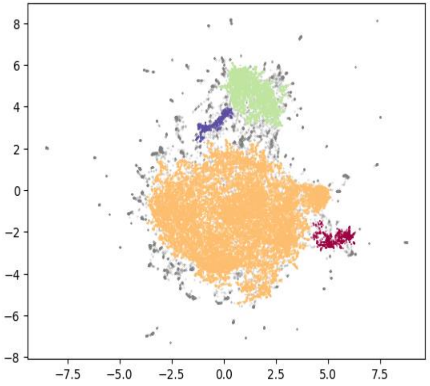
- UMAP+HDBSCAN reveals that **four** clusters is optimal as the number of hidden states for our HMM.

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,37} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,37} \\ \vdots & \vdots & \ddots & \vdots \\ x_{796,1} & x_{796,2} & \cdots & x_{796,37} \end{bmatrix} \in \mathbb{R}^{796\times37}$$

Raw Patient Values



UMAP Embedding



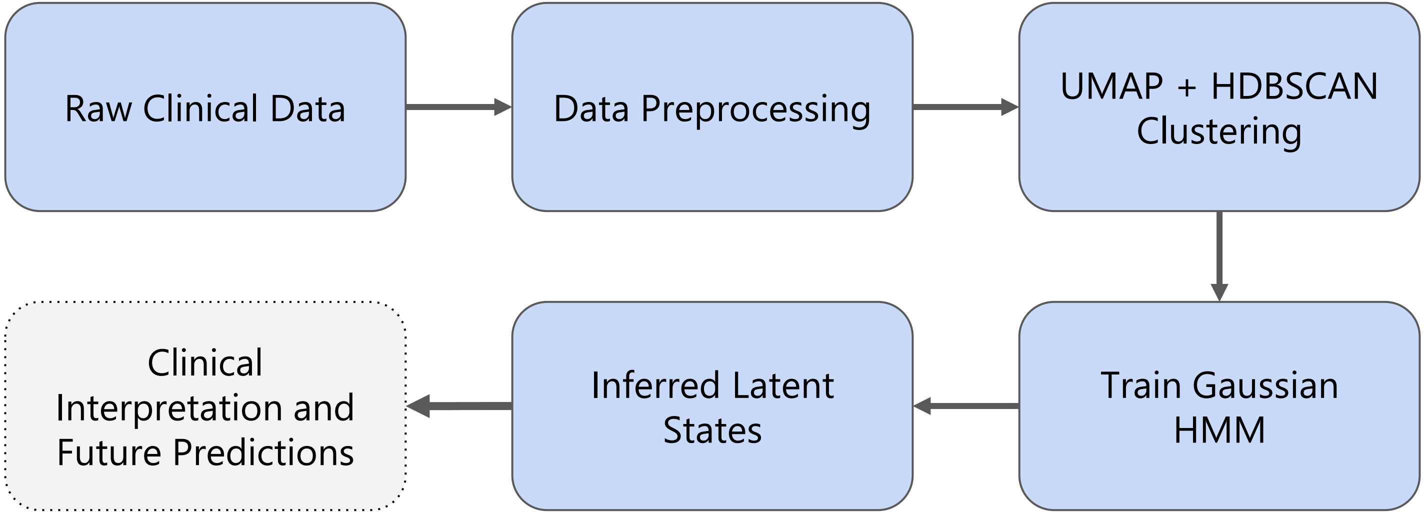
HDBSCAN Clusters

HMM Model Initialization

- Emission means $\boldsymbol{\mu}$ for each state initialized by sampling from a uniform distribution over [0,1).
- *Full* covariance matrix $\boldsymbol{\Sigma}$ initialized for each state.

$$\mathbf{A} = \begin{bmatrix} 0.90 & 0.04 & 0.03 & 0.03 \\ 0.03 & 0.90 & 0.04 & 0.03 \\ 0.03 & 0.03 & 0.90 & 0.04 \\ 0.01 & 0.01 & 0.01 & 0.97 \end{bmatrix} \qquad \boldsymbol{\pi} = [0.33,\ 0.33,\ 0.33,\ 0.01]$$

Methodology



Raw Clinical Data

Data Preprocessing

UMAP + HDBSCAN Clustering

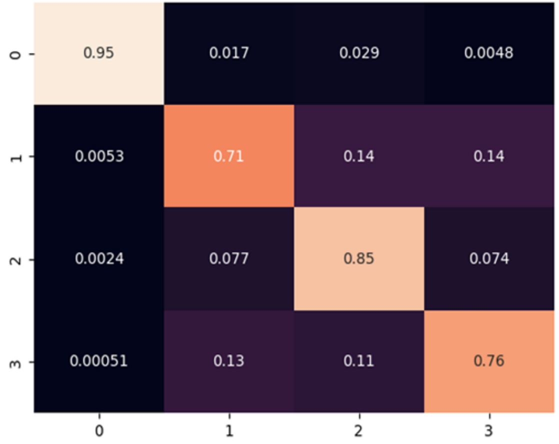
Train Gaussian HMM

Inferred Latent States

Clinical Interpretation and Future Predictions

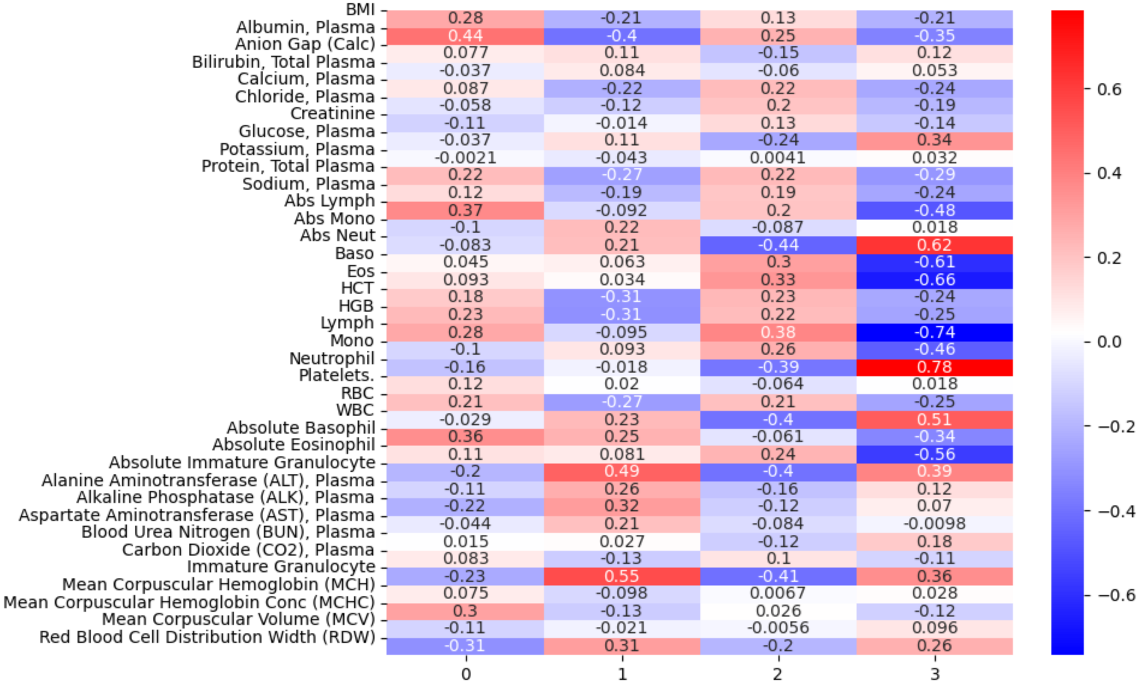
Results

Learned Parameters



	0	1	2	3
0	0.95	0.017	0.029	0.0048
1	0.0053	0.71	0.14	0.14
2	0.0024	0.077	0.85	0.074
3	0.00051	0.13	0.11	0.76

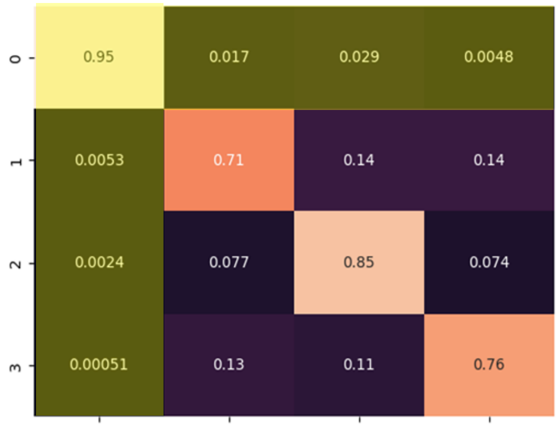
Final Transition Matrix **A**



	0	1	2	3
BMI	0.28	-0.21	0.13	-0.21
Albumin, Plasma	0.44	-0.4	0.25	-0.35
Anion Gap (Calc)	0.077	0.11	-0.15	0.12
Bilirubin, Total Plasma	-0.037	0.084	-0.06	0.053
Calcium, Plasma	0.087	-0.22	0.22	-0.24
Chloride, Plasma	-0.058	-0.12	0.2	-0.19
Creatinine	-0.11	-0.014	0.13	-0.14
Glucose, Plasma	-0.037	0.11	-0.24	0.34
Potassium, Plasma	-0.0021	-0.043	0.0041	0.032
Protein, Total Plasma	0.22	-0.27	0.22	-0.29
Sodium, Plasma	0.12	-0.19	0.19	-0.24
Abs Lymph	0.37	-0.092	0.2	-0.48
Abs Mono	-0.1	0.22	-0.087	0.018
Abs Neut	-0.083	0.21	-0.44	0.62
Baso	0.045	0.063	0.3	-0.61
Eos	0.093	0.034	0.33	-0.66
HCT	0.18	-0.31	0.23	-0.24
HGB	0.23	-0.31	0.22	-0.25
Lymph	0.28	-0.095	0.38	-0.74
Mono	-0.1	0.093	0.26	-0.46
Neutrophil	-0.16	-0.018	-0.39	0.78
Platelets.	0.12	0.02	-0.064	0.018
RBC	0.21	-0.27	0.21	-0.25
WBC	-0.029	0.23	-0.4	0.51
Absolute Basophil	0.36	0.25	-0.061	-0.34
Absolute Eosinophil	0.11	0.081	0.24	-0.56
Absolute Immature Granulocyte	-0.2	0.49	-0.4	0.39
Alanine Aminotransferase (ALT), Plasma	-0.11	0.26	-0.16	0.12
Alkaline Phosphatase (ALK), Plasma	-0.22	0.32	-0.12	0.07
Aspartate Aminotransferase (AST), Plasma	-0.044	0.21	-0.084	-0.0098
Blood Urea Nitrogen (BUN), Plasma	0.015	0.027	-0.12	0.18
Carbon Dioxide (CO2), Plasma	0.083	-0.13	0.1	-0.11
Immature Granulocyte	-0.23	0.55	-0.41	0.36
Mean Corpuscular Hemoglobin (MCH)	0.075	-0.098	0.0067	0.028
Mean Corpuscular Hemoglobin Conc (MCHC)	0.3	-0.13	0.026	-0.12
Mean Corpuscular Volume (MCV)	-0.11	-0.021	-0.0056	0.096
Red Blood Cell Distribution Width (RDW)	-0.31	0.31	-0.2	0.26

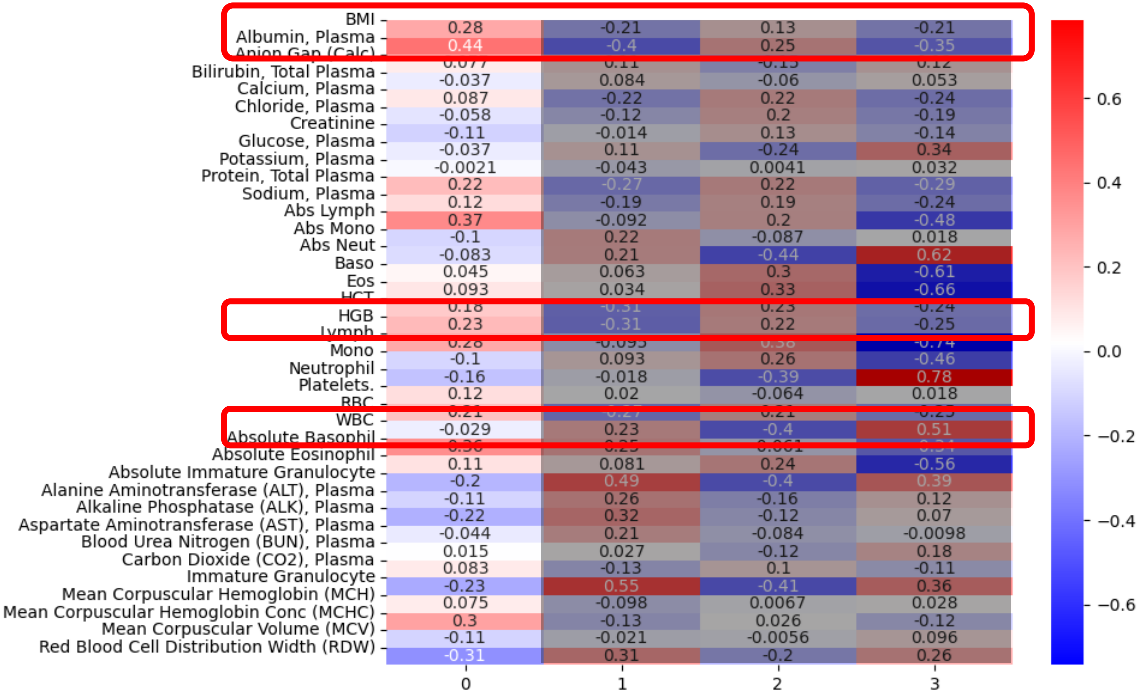
HMM Feature Averages per Hidden State

State 0 → Pre-Cachexia



	0	1	2	3
0	0.95	0.017	0.029	0.0048
1	0.0053	0.71	0.14	0.14
2	0.0024	0.077	0.85	0.074
3	0.00051	0.13	0.11	0.76

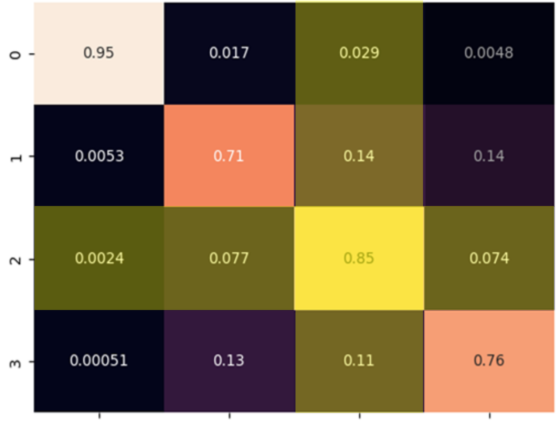
Final Transition Matrix **A**



	0	1	2	3
BMI	0.28	-0.21	0.13	-0.21
Albumin, Plasma	0.44	-0.4	0.25	-0.35
Anion Gap (Calc)	0.077	0.11	-0.13	0.12
Bilirubin, Total Plasma	-0.037	0.084	-0.06	0.053
Calcium, Plasma	0.087	-0.22	0.22	-0.24
Chloride, Plasma	-0.058	-0.12	0.2	-0.19
Creatinine	-0.11	-0.014	0.13	-0.14
Glucose, Plasma	-0.037	0.11	-0.24	0.34
Potassium, Plasma	-0.0021	-0.043	0.0041	0.032
Protein, Total Plasma	0.22	-0.27	0.22	-0.29
Sodium, Plasma	0.12	-0.19	0.19	-0.24
Abs Lymph	0.37	-0.092	0.2	-0.48
Abs Mono	-0.1	0.22	-0.087	0.018
Abs Neut	-0.083	0.21	-0.44	0.62
Baso	0.045	0.063	0.3	-0.61
Eos	0.093	0.034	0.33	-0.66
HCT	0.18	-0.31	0.23	-0.24
HGB	0.23	-0.31	0.22	-0.25
Lymph	0.28	-0.095	0.38	-0.74
Mono	-0.1	0.093	0.26	-0.46
Neutrophil	-0.16	-0.018	-0.39	0.78
Platelets.	0.12	0.02	-0.064	0.018
RBC	0.21	-0.27	0.21	-0.23
WBC	-0.029	0.23	-0.4	0.51
Absolute Basophil	0.36	0.25	-0.061	-0.34
Absolute Eosinophil	0.11	0.081	0.24	-0.56
Absolute Immature Granulocyte	-0.2	0.49	-0.4	0.39
Alanine Aminotransferase (ALT), Plasma	-0.11	0.26	-0.16	0.12
Alkaline Phosphatase (ALK), Plasma	-0.22	0.32	-0.12	0.07
Aspartate Aminotransferase (AST), Plasma	-0.044	0.21	-0.084	-0.0098
Blood Urea Nitrogen (BUN), Plasma	0.015	0.027	-0.12	0.18
Carbon Dioxide (CO2), Plasma	0.083	-0.13	0.1	-0.11
Immature Granulocyte	-0.23	0.55	-0.41	0.36
Mean Corpuscular Hemoglobin (MCH)	0.075	-0.098	0.0067	0.028
Mean Corpuscular Hemoglobin Conc (MCHC)	0.3	-0.13	0.026	-0.12
Mean Corpuscular Volume (MCV)	-0.11	-0.021	-0.0056	0.096
Red Blood Cell Distribution Width (RDW)	-0.31	0.31	-0.2	0.26

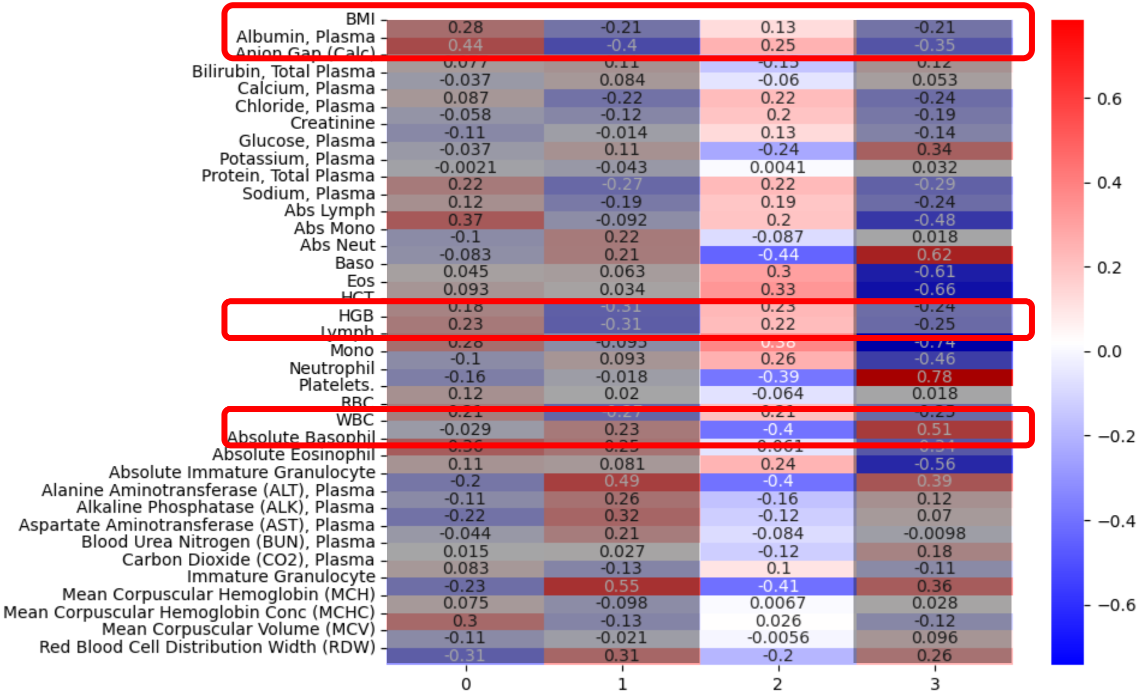
HMM Feature Averages per Hidden State

State 2 → Early Cachexia



	0	1	2	3
0	0.95	0.017	0.029	0.0048
1	0.0053	0.71	0.14	0.14
2	0.0024	0.077	0.85	0.074
3	0.00051	0.13	0.11	0.76

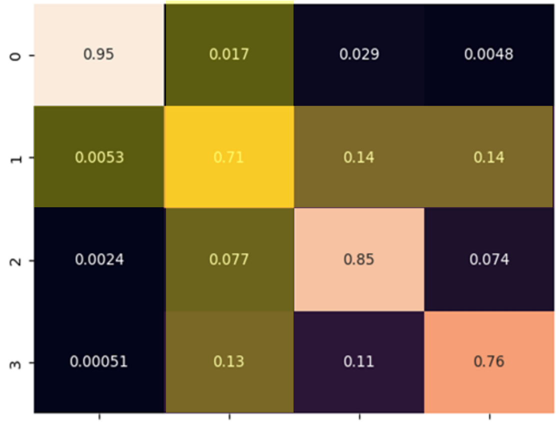
Final Transition Matrix **A**



	0	1	2	3
BMI	0.28	-0.21	0.13	-0.21
Albumin, Plasma	0.44	-0.4	0.25	-0.35
Anion Gap (Calc)	0.077	0.11	-0.13	0.12
Bilirubin, Total Plasma	-0.037	0.084	-0.06	0.053
Calcium, Plasma	0.087	-0.22	0.22	-0.24
Chloride, Plasma	-0.058	-0.12	0.2	-0.19
Creatinine	-0.11	-0.014	0.13	-0.14
Glucose, Plasma	-0.037	0.11	-0.24	0.34
Potassium, Plasma	-0.0021	-0.043	0.0041	0.032
Protein, Total Plasma	0.22	-0.27	0.22	-0.29
Sodium, Plasma	0.12	-0.19	0.19	-0.24
Abs Lymph	0.37	-0.092	0.2	-0.48
Abs Mono	-0.1	0.22	-0.087	0.018
Abs Neut	-0.083	0.21	-0.44	0.62
Baso	0.045	0.063	0.3	-0.61
Eos	0.093	0.034	0.33	-0.66
HCT	0.18	-0.31	0.23	-0.24
HGB	0.23	-0.31	0.22	-0.25
Lymph	0.28	-0.095	0.38	-0.74
Mono	-0.1	0.093	0.26	-0.46
Neutrophil	-0.16	-0.018	-0.39	0.78
Platelets.	0.12	0.02	-0.064	0.018
RBC	0.21	-0.27	0.21	-0.23
WBC	-0.029	0.23	-0.4	0.51
Absolute Basophil	0.36	0.25	-0.061	-0.34
Absolute Eosinophil	0.11	0.081	0.24	-0.56
Absolute Immature Granulocyte	-0.2	0.49	-0.4	0.39
Alanine Aminotransferase (ALT), Plasma	-0.11	0.26	-0.16	0.12
Alkaline Phosphatase (ALK), Plasma	-0.22	0.32	-0.12	0.07
Aspartate Aminotransferase (AST), Plasma	-0.044	0.21	-0.084	-0.0098
Blood Urea Nitrogen (BUN), Plasma	0.015	0.027	-0.12	0.18
Carbon Dioxide (CO2), Plasma	0.083	-0.13	0.1	-0.11
Immature Granulocyte	-0.23	0.55	-0.41	0.36
Mean Corpuscular Hemoglobin (MCH)	0.075	-0.098	0.0067	0.028
Mean Corpuscular Hemoglobin Conc (MCHC)	0.3	-0.13	0.026	-0.12
Mean Corpuscular Volume (MCV)	-0.11	-0.021	-0.0056	0.096
Red Blood Cell Distribution Width (RDW)	-0.31	0.31	-0.2	0.26

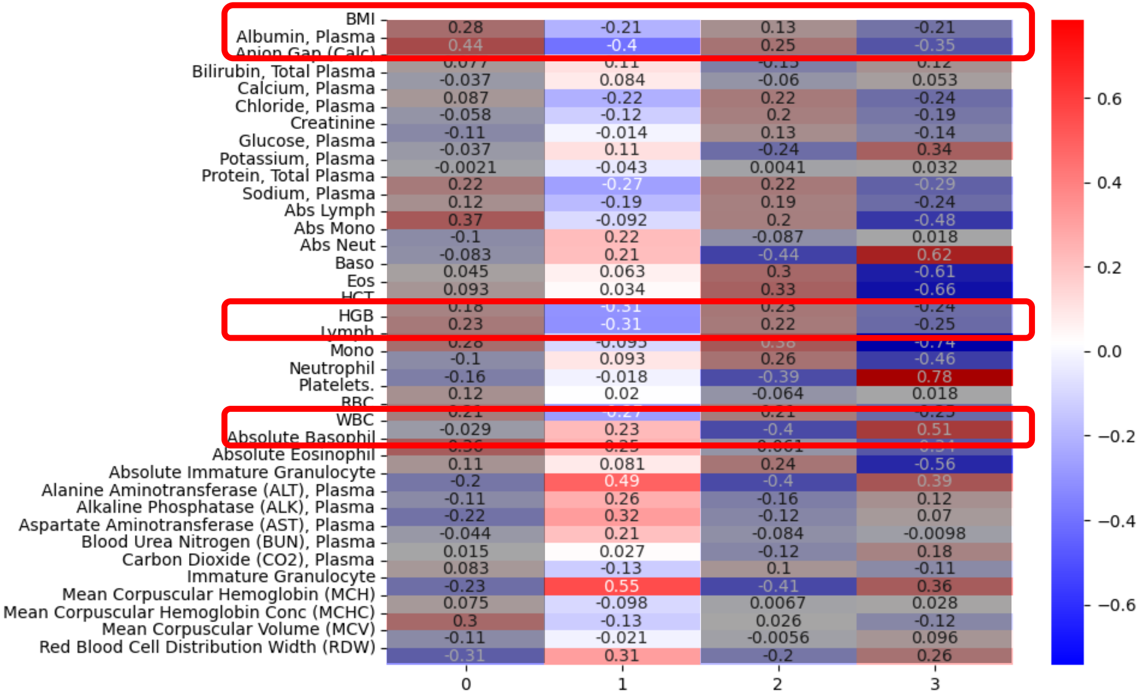
HMM Feature Averages per Hidden State

State 1 → Intermediate Cachexia



	0	1	2	3
0	0.95	0.017	0.029	0.0048
1	0.0053	0.71	0.14	0.14
2	0.0024	0.077	0.85	0.074
3	0.00051	0.13	0.11	0.76

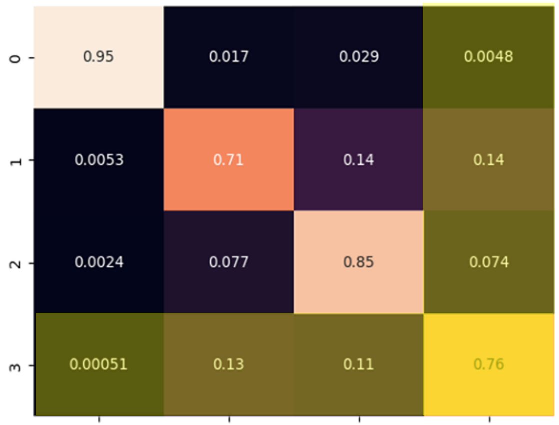
Final Transition Matrix **A**



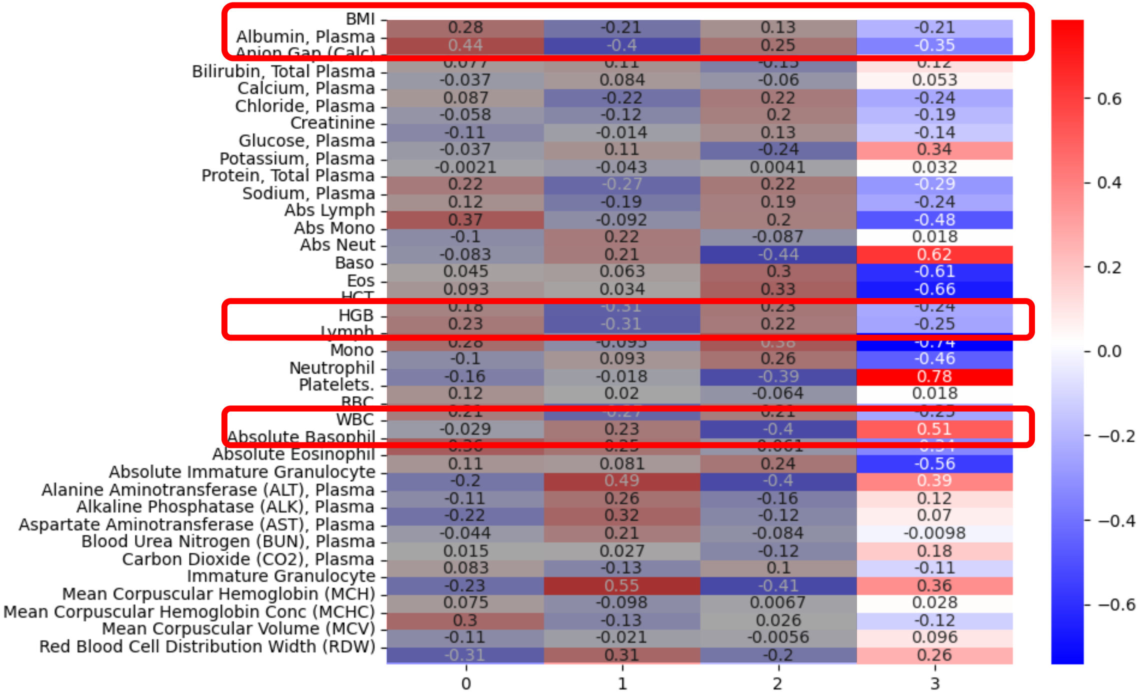
	0	1	2	3
BMI	0.28	-0.21	0.13	-0.21
Albumin, Plasma	0.44	-0.4	0.25	-0.35
Anion Gap (Calc)	0.077	0.11	-0.13	0.12
Bilirubin, Total Plasma	-0.037	0.084	-0.06	0.053
Calcium, Plasma	0.087	-0.22	0.22	-0.24
Chloride, Plasma	-0.058	-0.12	0.2	-0.19
Creatinine	-0.11	-0.014	0.13	-0.14
Glucose, Plasma	-0.037	0.11	-0.24	0.34
Potassium, Plasma	-0.0021	-0.043	0.0041	0.032
Protein, Total Plasma	0.22	-0.27	0.22	-0.29
Sodium, Plasma	0.12	-0.19	0.19	-0.24
Abs Lymph	0.37	-0.092	0.2	-0.48
Abs Mono	-0.1	0.22	-0.087	0.018
Abs Neut	-0.083	0.21	-0.44	0.62
Baso	0.045	0.063	0.3	-0.61
Eos	0.093	0.034	0.33	-0.66
HCT	0.18	-0.31	0.23	-0.24
HGB	0.23	-0.31	0.22	-0.25
Lymph	0.28	-0.095	0.38	-0.74
Mono	-0.1	0.093	0.26	-0.46
Neutrophil	-0.16	-0.018	-0.39	0.78
Platelets.	0.12	0.02	-0.064	0.018
RBC	0.21	-0.27	0.21	-0.23
WBC	-0.029	0.23	-0.4	0.51
Absolute Basophil	0.36	0.25	-0.061	-0.34
Absolute Eosinophil	0.11	0.081	0.24	-0.56
Absolute Immature Granulocyte	-0.2	0.49	-0.4	0.39
Alanine Aminotransferase (ALT), Plasma	-0.11	0.26	-0.16	0.12
Alkaline Phosphatase (ALK), Plasma	-0.22	0.32	-0.12	0.07
Aspartate Aminotransferase (AST), Plasma	-0.044	0.21	-0.084	-0.0098
Blood Urea Nitrogen (BUN), Plasma	0.015	0.027	-0.12	0.18
Carbon Dioxide (CO2), Plasma	0.083	-0.13	0.1	-0.11
Immature Granulocyte	-0.23	0.55	-0.41	0.36
Mean Corpuscular Hemoglobin (MCH)	0.075	-0.098	0.0067	0.028
Mean Corpuscular Hemoglobin Conc (MCHC)	0.3	-0.13	0.026	-0.12
Mean Corpuscular Volume (MCV)	-0.11	-0.021	-0.0056	0.096
Red Blood Cell Distribution Width (RDW)	-0.31	0.31	-0.2	0.26

HMM Feature Averages per Hidden State

State 3 → Refractory Cachexia

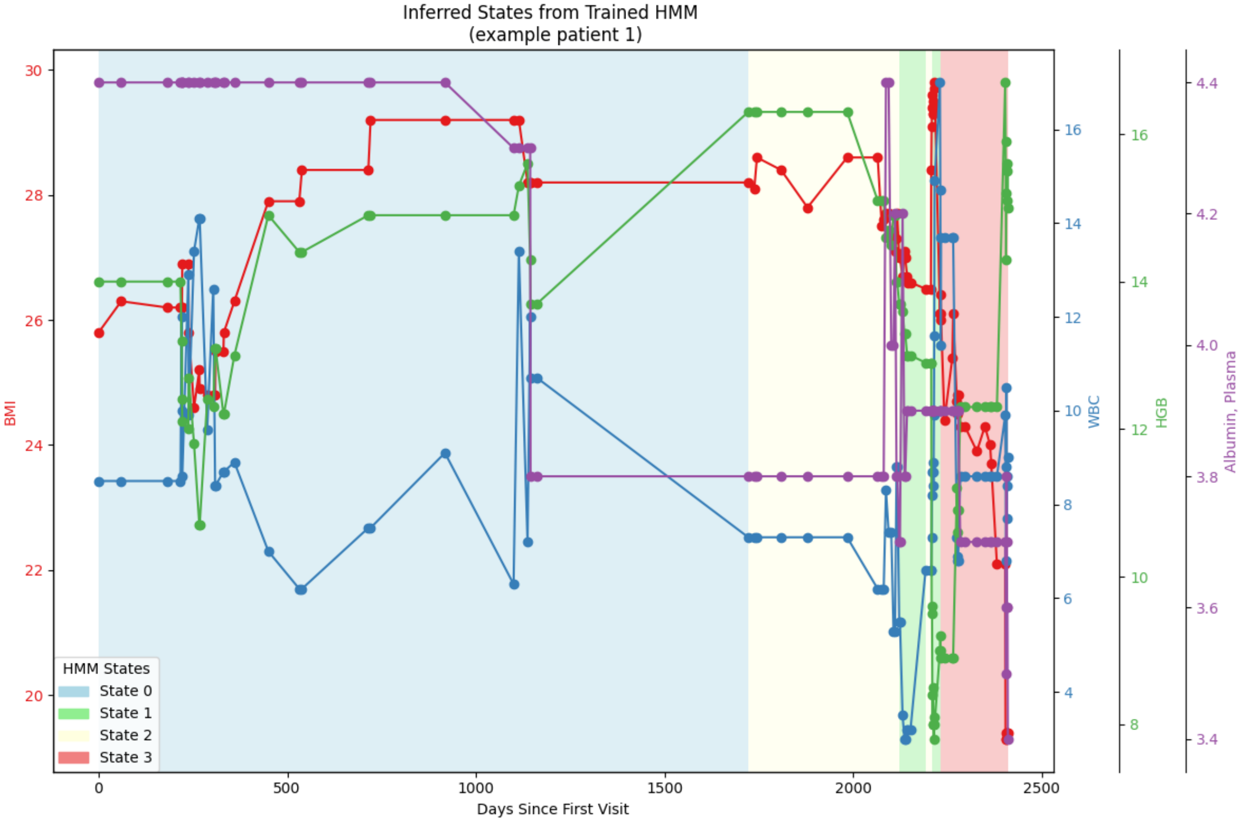


Final Transition Matrix **A**

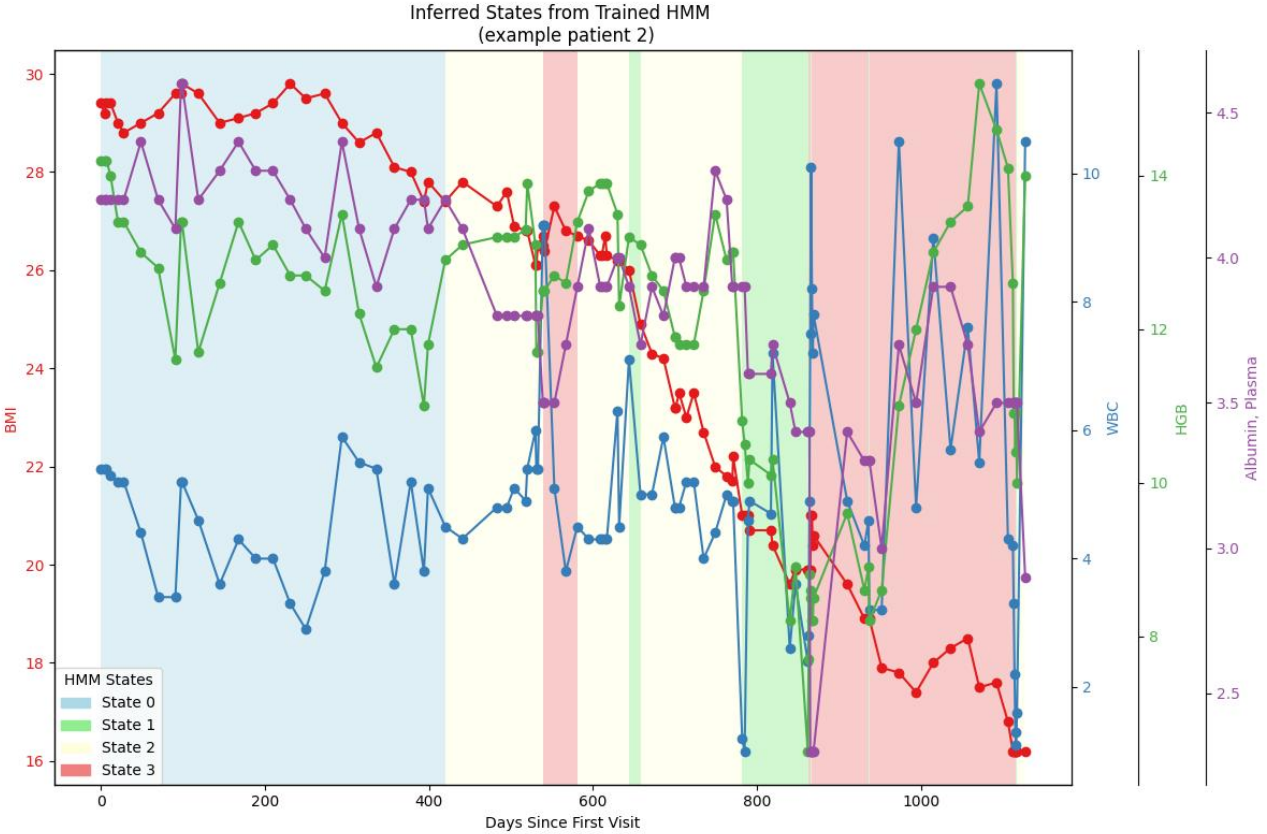


HMM Feature Averages per Hidden State

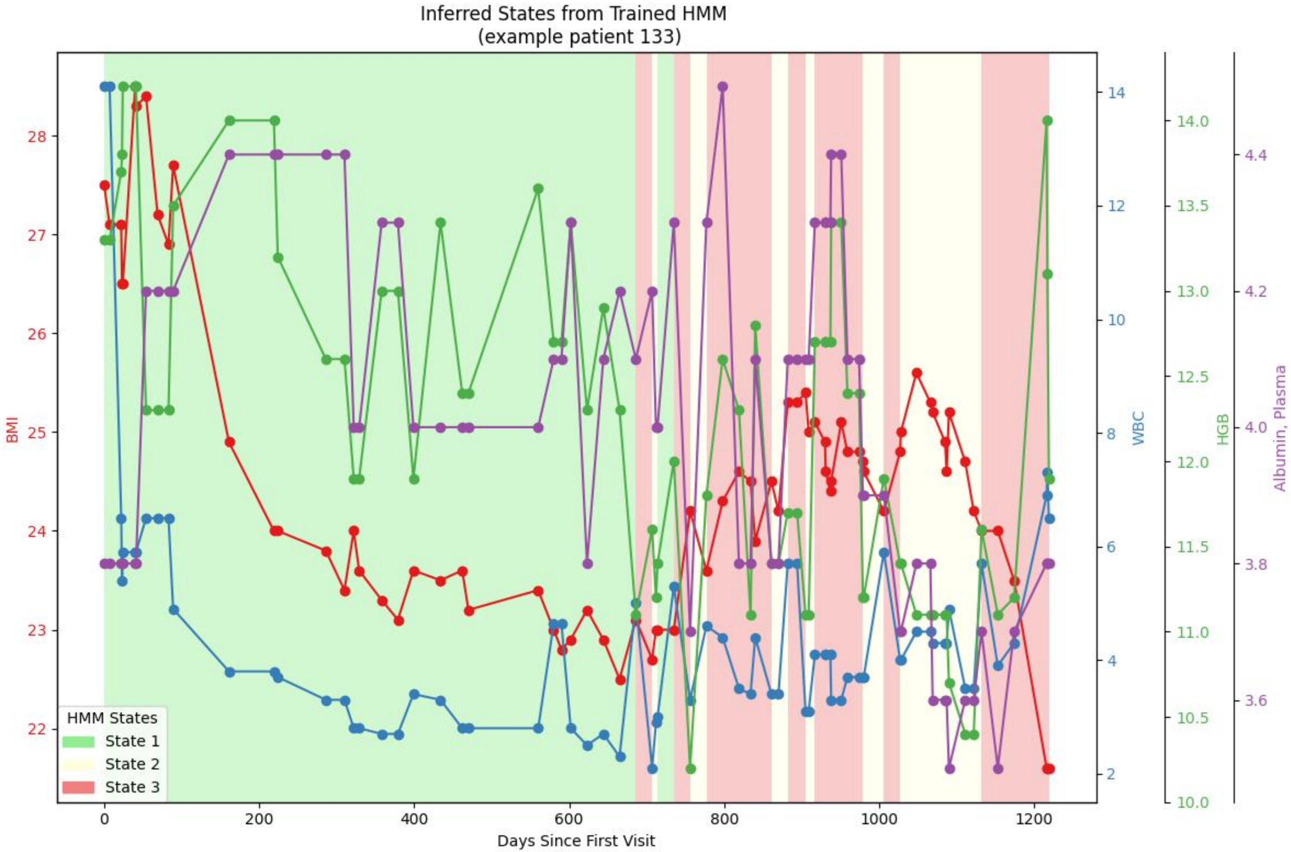
Example Patient 1



Example Patient 2



Unexpected Behaviors



Areas of improvement

- States seem to change too frequently in transitional cachexia states.
- Learned states dependent on initialized $\boldsymbol{A}$, $\boldsymbol{\pi}$, and covariance type.
- States are only classified during date of visit.
- Difficulty quantifying the uncertainty and precision of predicted stages.

Future Work

- Incorporate medical imaging data like MRI or CT scans.
- Improve clinical interpretation of learned parameters.
- *Evaluate the model against some ground truth?*

Conclusion

- Deployed various clustering algorithms to optimize the number of hidden states of initialized HMM.
- Trained Gaussian HMM to model the progression of cachexia using longitudinal biomarkers available.
- *HMMs can be a useful tool to identifying stages of cachexia.*

Thank you!
Q&A