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

L. Ocfemia[1], E. Reznik[2] and S. Keene[1]

1. The Cooper Union, New York NY, USA
2. Memorial Sloan-Kettering Cancer Center, New York NY, USA
{lizelle.ocfemia, keene}@cooper.edu, reznike@mskcc.org

***Abstract*— Cancer cachexia is a complex, multifactorial syndrome characterized by severe muscle wasting, weight loss, and systemic inflammation, significantly affecting patient prognosis and quality of life. Cachexia is estimated to occur in up to 80% of people with advanced cancer, depending on cancer type and response to cancer treatment. Although it is said to be the direct cause of 30% of cancer deaths, there are currently no effective treatments for cachexia. The clinical definition of the syndrome is also loosely defined as losing more than 5% of body weight in the past six to twelve months. Accurate classification of cancer cachexia stages is crucial for timely intervention and personalized treatment strategies. Hidden Markov Models (HMMs), with their ability to model temporal sequences and hidden states, offer a robust approach for classifying the stages of cancer cachexia. This study explores the application of HMMs in analyzing longitudinal BMI and blood lab values from the MSK-IMPACT cohort to identify and classify the different stages of cachexia in cancer patients. The results demonstrate that HMMs can effectively distinguish between early, intermediate, and late stages of cachexia, thereby offering valuable insights for clinicians to optimize treatment regimens and improve patient outcomes. This approach improves the precision of cachexia staging and contributes to the broader field of predictive modeling in computational oncology.**

Keywords*— *Cachexia Stage Classification, Hidden Markov Models.

I. Introduction

Cancer is a leading cause of death globally, with the World Health Organization (WHO) reporting 9.6 million deaths and 18.1 million new cases in 2018 [1]. In cancer patients, cachexia is responsible for significant morbidity, contributing to weight loss, decreased performance status, and resistance to cancer therapies. Cancer cachexia has been identified in a wide range of cancers, including pancreatic, colorectal, gastric, and lung cancers, among others. It presents a critical concern for oncologists due to its negative impact on both the patient's physical health and their ability to tolerate cancer treatments.

The syndrome is multifactorial, involving complex interactions between cancer cells, the host's immune system, and metabolic pathways. Although the symptoms of cachexia are well recognized, the mechanisms driving the condition remain incompletely understood, complicating efforts to prevent or treat it effectively. Early identification of cachexia is crucial to improving patient outcomes, as timely intervention can help mitigate some of the devastating effects of muscle wasting and metabolic dysfunction. However, current diagnostic criteria often fail to catch cachexia in its early stages, leading to missed opportunities for intervention.

In this paper, we aim to address the gap in early-stage diagnosis and classification of cancer-associated cachexia by applying advanced computational techniques to better understand its progression. Specifically, we utilize hidden Markov models (HMMs) to model the dynamic changes in cachexia progression over time, based on clinical data collected from cancer patients of Memorial Sloan Kettering Cancer Center (MSK or MSKCC). Our approach focuses on classifying patients into different stages of cachexia using longitudinal data, including key biomarkers such as Body Mass Index (BMI), Albumin, Hemoglobin, and inflammatory markers like White Blood Cells (WBC).

By applying HMMs, we can capture *interpretable* latent states underlying the observed clinical data – an advantage not easily achievable with deep learning approaches – thereby identifying distinct physiological trajectories of cachexia progression. This method allows us to predict transitions between states, offering a clearer understanding of disease progression and providing a new tool for researchers.

II. Definitions of Cancer Cachexia

Cachexia is a disorder marked by involuntary weight loss and a disruption of homeostatic control over energy and protein balance [2]. It is commonly associated with malignant diseases but also occurs in various chronic nonmalignant conditions, including heart failure, kidney disease, chronic obstructive pulmonary disease, neurological disorders, and AIDS. The syndrome arises from a complex interaction of reduced food intake and metabolic alterations, such as increased energy expenditure, anorexia, and systemic inflammation. Unlike starvation or simple malnutrition—both of which can be reversed with adequate nutrition—cachexia persists despite nutritional support.

In 2011, an international Delphi consensus definition and classification of cancer cachexia was published, provisionally defining cancer cachexia as $> 5\%$ weight

loss in the previous 6 months or 2%-5% weight loss with either a BMI of less than 20 kg/m^2 or reduced muscle mass [3]. It was also proposed that cancer cachexia is a multi-factorial syndrome defined by an ongoing loss of skeletal muscle mass (with or without loss of fat mass) that can be partially but not entirely reversed by conventional nutritional support. Depletion of skeletal muscle and adipose tissue (body fat) are key features of cancer-associated cachexia [4], and its consequences include reduced antitumor efficacy [5], increased chemotherapy toxicity [6–8], complications from cancer surgery [8], and mortality. The physiological characteristics of the disease are negative nitrogen balance and negative energy balance due to reduced food intake and abnormally high metabolism. The diagnostic criteria defined by Evans [9] are applicable to all types of chronic disease-related cachexia, when taking metabolism and nutrition into account. The diagnostic criteria presented by Fearon [3] specifically target cancer-related cachexia, emphasizing weight loss factors and reduced muscle loss. Overall, weight loss, loss of appetite, growth disorders, and decreased muscle mass are the main symptoms of cachexia.

Globally, half of all cancer-related deaths ($\sim$8.2 million annually) [1] occur in cancers most frequently associated with cachexia, including pancreatic (0.33 million deaths), esophageal (0.40 million), gastric (0.72 million), pulmonary (1.59 million), hepatic (0.75 million), and colorectal (0.69 million) cancers [10]. Despite its high prevalence, cancer-associated cachexia is not included in national cancer statistics and is rarely cited as a direct cause of death. However, it is strongly linked to incurable disease and is highly prevalent in end-of-life stages [11]. While cachexia is most commonly associated with advanced cancers, it can also occur in curable cases and may be reversed if the underlying cancer is successfully treated [12]. Therefore, active prevention, diagnosis, and treatment of cachexia and its precursor states are necessary to reverse or minimize the negative impact on patient outcomes.

Fearon's conceptual framework for diagnosing cancer cachexia has since gained wide acceptance and will be the main framework used in this thesis. In this framework, progressive stages of cachexia were also defined: pre-cachexia, cachexia, and refractory cachexia, though patients may not experience all three stages [3].

- **Pre-cachexia:** In this stage, the patient experiences weight loss despite maintaining a normal diet. Weight loss over the past six to twelve months is less than 5% of total body weight. For patients with obesity, weight loss may not be as apparent, but metabolic changes detected by tests indicate loss of muscle and fat.
- **Cachexia:** The patient has lost more than 5% of body weight within the past six to twelve months or more than 2% of body weight with a body mass index (BMI) below 20 kg/m^2. Diagnostic tests may reveal signs of muscle loss and inflammation.
- **Refractory cachexia:** The patient has lost more than 20% of usual body weight, with a BMI lower than 27 kg/m^2. This significant weight loss leads to a noticeable change in appearance and may result in insufficient strength to undergo cancer treatment. In some cases, cancer treatments become ineffective.

The incidence and severity of cachexia are highly diverse and depend on the type, location, and stage of the tumor. Currently, there are no specific biomarkers for early-stage cachexia identification. Staging is determined according to the clinical manifestations and characteristics of the patient. The refractory cachexia phase is determined by the underlying disease of the patient and the overall condition; diagnosis of this stage requires a low WHO performance status score and a survival period of less than 3 months [3]. The focus of treatment for refractory cachexia goes from the objective of cure and control to the maintenance of the patient's quality of life. This type of grading system can provide patients with more suitable treatment options at all stages of disease development and allows for targeted research and treatment for each stage.

III. Prior Research

Numerous machine learning techniques have been applied to the study of cancer cachexia, particularly those that utilize supervised classifiers to predict the presence of cachexia in patients [13–15]. These models typically rely on predefined labels, classifying patients as either cachectic or non-cachectic. However, this approach has inherent limitations, primarily because the definition of cachexia is commonly based on a threshold value of BMI, which is a simplistic measure that may not fully capture the complex nature of the syndrome. Cachexia, as defined by BMI, is particularly susceptible to external factors such as dietary changes or recent surgeries, which can cause rapid shifts in a patient's weight, potentially leading to misclassification. As a result, BMI-based classification does not account for other critical clinical markers such as muscle wasting, inflammation, or metabolic dysfunction, which are essential components of cachexia progression.

Furthermore, many existing machine learning models use static data, such as a snapshot of a patient's clinical status at a particular time point [16]. While this may provide useful information, these models are unable to capture the dynamic, longitudinal nature of cachexia progression. Cachexia typically evolves over time, with subtle changes in various biomarkers, and the ability to track these changes across multiple visits can offer deeper insights into disease progression and response to

treatment. Machine learning models that fail to incorporate this temporal aspect are therefore limited in their ability to infer the stage of cachexia at any given time, making them less reliable for monitoring patients over the course of their illness.

Additionally, many state-of-the-art machine learning models in this field integrate medical imaging data, such as CT or MRI scans, to evaluate a patient's skeletal muscle area [17–19]. While these imaging techniques are considered the gold standard for diagnosing cachexia by directly visualizing muscle loss, they come with significant drawbacks, making them less feasible for regular monitoring, especially for patients undergoing long-term treatment.

These limitations highlight the need for more accessible and cost-effective methods to classify and monitor cachexia progression. Approaches that leverage longitudinal data—such as hidden Markov models used in this study—can offer a more nuanced understanding of cachexia by considering temporal dynamics without relying on expensive imaging techniques. By focusing on easily accessible clinical data, such as BMI, albumin, and inflammatory markers, we can develop models that are both effective and practical for real-world applications, allowing continuous monitoring of cachexia progression over time.

IV. Dataset

The datasets used for the research in this paper were collected from patients in the MSK-IMPACT cohort. MSK-IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets) is a targeted tumor-sequencing test available to MSK patients [20]. It is used to detect mutations and other critical changes in the genes of both rare and common cancers. There are a total of 63,008 patients in the cohort with recorded body mass index (BMI) and surgery data.

Body Mass Index (BMI) is a standardized metric used to assess body fat composition based on an individual's weight and height [21]. It is a clinically relevant indicator of nutritional status and physical health, often used as a screening tool for malnutrition and disease severity in cancer patients. In the context of cachexia, BMI serves as a critical biomarker for disease progression and prognosis [22]. For this study, BMI values were extracted from the longitudinal patient records, with each entry corresponding to a patient's clinical visit to MSK.

In addition to BMI, 36 longitudinal laboratory features were used to form the multivariate time series input for the HMM. The features chosen were recommended by clinicians at MSKCC as likely to be clinically relevant to cachexia. These features span a variety of physiological systems and were selected for their clinical relevance to cancer-associated cachexia, including markers of inflammation, nutrition, liver and kidney function, and red blood cell dynamics. Future research could further investigate feature importance. Table 1 categorizes the features by physiological system.

V. Preprocessing Methods

Prior to training the HMM, the MSK health record data was filtered, aligned, and normalized to construct consistent multivariate time-series sequences for each patient. This section describes the preprocessing pipeline used to extract valid longitudinal clinical trajectories from the dataset.

V-A. Patient Filtering Criteria

The following filtering criteria were applied to the dataset to ensure meaningful cachexia state predictions:

- Patients were restricted to those diagnosed with a specified cancer type (e.g., Lung Adenocarcinoma) and confirmed deceased.
- Records with missing or malformed timestamps were removed.
- For each patient, a new column `days_since_first` was computed, representing the number of days since the patient's first recorded visit.
- Patients with fewer than 20 visits were excluded from the dataset to ensure sufficient temporal resolution for modeling disease progression.

V-B. Sequence Construction and Normalization

Missing values within each patient trajectory were imputed using forward-fill and backward-fill methods to preserve temporal consistency without introducing external information.

Each remaining patient sequence was then standardized using z-score normalization. Each patient trajectory was represented as a time series matrix $\mathbf{X}^{(i)} \in \mathbb{R}^{T_i \times D}$, where T_i is the number of visits and $D = 37$ is the number of standardized clinical features. All patient sequences were concatenated into a single array $\mathbf{X} \in \mathbb{R}^{N \times D}$ for training, where $N = \sum_i T_i$.

VI. Determining the Optimal Number of Hidden States

To determine the optimal number of hidden states for the HMM implementation, we employed various clustering algorithms such as K-means, Spectral Clustering, UMAP and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HBDSCAN) to detect naturally occurring clusters in the dataset. HDBSCAN and UMAP were chosen because this combination meant we did not have to specify the number of clusters The time of the clinical visit was excluded as an input

Table 1. Categories of Laboratory Features Used in trained HMM

Category	Blood Lab Value Features
Nutritional Markers	Albumin, Total Protein, Calcium
Inflammatory Markers	WBC, Neutrophils, Monocytes, Immature Granulocytes
Liver Function Tests	ALT, AST, ALK, Bilirubin
Kidney Function Tests	Creatinine, BUN
Electrolytes	Sodium, Potassium, Chloride, CO_2
Red Blood Cell Dynamics	RBC, HGB, HCT, MCV, MCH, MCHC, RDW
White Blood Cell Differentiation	Lymphocytes, Basophils, Eosinophils, Platelets

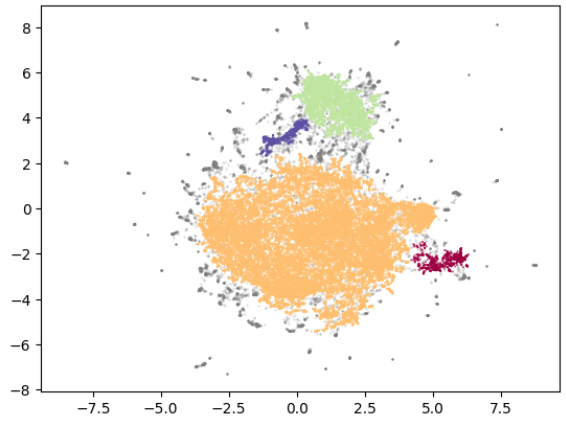

Figure 1. HDBSCAN Clusters of UMAP Embedding.

to the clustering models. This allowed us to focus solely on the attributes that are representative of the patients' conditions and allow the HMM to learn the transitions between these states.

VI-0a. HDBSCAN

The method that gave the best clustering was to use UMAP for dimensionality reduction on our dataset followed by clustering through HDBSCAN. UMAP was used to reduce the dimensionality of the data to two components while preserving the global structure.

We refined the UMAP embedding by adjusting the hyperparameters to enhance the clustering capability. We increased the number of neighbors to 30 and set the minimum distance to 0.0, which resulted in an embedding that produced good clusters. This transformation highlighted regions of the data that are more suitable for clustering, allowing for clear differentiation between data points.

To identify clusters, we then applied HDBSCAN, with a minimum samples for a cluster of 10 and a minimum cluster size of 500 to ensure that only sufficiently dense regions were considered valid clusters. The HDBSCAN algorithm assigned cluster labels to each data point. In Figure 1, the points that were identified as belonging to clusters are colored based on their assigned labels, while noise points are shown in gray. This approach allowed us to identify a natural formation of clusters in the data, with the most coherent results achieved with four clusters.

VII. HMM Initialization and Training

A multivariate Gaussian HMM was used to model latent clinical states underlying longitudinal changes in patient BMI and blood laboratory values. The model was implemented using the `GaussianHMM` class from `hmmlearn` Python library.

VII-A. Transition Matrix and Initial Probabilities

The transition matrix $\mathbf{A}$ and initial state probability vector π were explicitly initialized to encourage biologically plausible dynamics. Specifically, the matrix was structured to favor self-transitions while allowing low-probability transitions to adjacent states:

$$\mathbf{A} = \begin{bmatrix} 0.90 & 0.04 & 0.03 & 0.03 \\ 0.03 & 0.90 & 0.04 & 0.03 \\ 0.03 & 0.03 & 0.90 & 0.04 \\ 0.01 & 0.01 & 0.01 & 0.97 \end{bmatrix}, \tag{1}$$

$$\pi = [0.33,\ 0.33,\ 0.33,\ 0.01]. \tag{2}$$

Each row of A corresponds to the current state, and each column represents the probability of transitioning to the next state. The elements of π indicate the probability of starting in each state.These choices reflect a model where patients tend to remain in the same state over time, with State 3 representing a terminal or stable absorbing condition (e.g., advanced cachexia), given its high self-transition probability.

VII-A1. Emission Parameter Initialization

Each hidden state in the Gaussian HMM emits observations drawn from a multivariate normal distribution parameterized by a state-specific mean vector μ_k and covariance matrix Σ_k. Proper initialization of these parameters is crucial for efficient convergence of the Expectation-Maximization (EM) algorithm.

The emission means $\mu_k \in \mathbb{R}^D$ for each state k were initialized using random draws from a uniform distribution over $[0,1]$, independently for each feature dimension. We ensured valid full-rank covariance matrices for each state.

VII-B. Model Training

Model fitting was performed using the Expectation-Maximization (EM) algorithm via the `fit` method, with the complete training dataset and corresponding sequence lengths provided. The model iteratively updated the emission and transition probabilities to max-

imize the likelihood of the observed data under the current parameters. The final trained model produced a set of learned transition probabilities, emission means and covariances, and state sequences for each patient trajectory, which are analyzed in subsequent sections.

We present an analysis on deceased patients who suffered from lung adenocarcinoma, a dataset consisting of 2,345 patients. This section shows our findings for each of this cohort.

VIII. Learned Model Parameters

After training the HMM on the preprocessed longitudinal clinical data, the model converged to a set of four latent states that capture underlying patterns in the progression of cancer-associated cachexia. The learned transition probability matrix $\mathbf{A}$ is shown below:

$$\mathbf{A} = \begin{bmatrix} 0.95 & 0.017 & 0.029 & 0.0048 \\ 0.0053 & 0.71 & 0.14 & 0.14 \\ 0.00024 & 0.077 & 0.85 & 0.07 \\ 0.00051 & 0.13 & 0.11 & 0.76 \end{bmatrix}. \quad (3)$$

In particular, we see that State 0 exhibits a high self-transition probability of 95%. This suggests a stable condition where patients tend to remain for extended periods. State 1, 2, and 3 also have relatively high self-transition probabilities, but allow movement to adjacent states.

The emission distributions of each state are defined by Gaussian distributions with learned means and covariances across the 37 clinical features. Figure 2 summarizes the normalized (z-scored) mean values for key features across the four states.

IX. State Characterization

The four hidden states discovered by the HMM represent distinct physiological phenotypes that align with progressive stages of cancer-associated cachexia. These latent states were interpretable through distinct differences in key clinical biomarkers, most notably BMI, Albumin, and Hemoglobin (HGB) —– core indicators of nutritional status, anemia, and systemic inflammation. The transition probabilities and mean feature values offer a biologically coherent interpretation of these states, which we label as: pre-cachexia (State 0), early cachexia (State 2), intermediate cachexia (State 1), and terminal cachexia (State 3).

IX-0a. State 0 – Pre-Cachexia.

State 0 is distinguished by a high self-transition probability (94.9%) and very low transition probabilities to all other states, indicating that patients in this state tend to remain physiologically stable over time. Clinically, this state is characterized by the highest mean values for BMI ($+0.280$), Albumin ($+0.442$), HGB ($+0.228$), Hematocrit (HCT) ($+0.177$), and Red Blood Cell Count (RBC), indicating robust nutritional and hematologic status. Inflammatory markers such as White Blood Cell (WBC) count (-0.029), Neutrophils (-0.162), and Immature Granulocytes are also at their lowest in this state, consistent with the absence of systemic inflammation [23–25]. It is also interesting to note that the Red Cell Distribution Width (RDW) is the lowest in this state (-0.310). These features are aligned with the international consensus definition of pre-cachexia, which describes early metabolic changes without significant weight or muscle loss [3].

IX-0b. States 1 and 2 – Transitional Cachexia.

States 1 and 2 represent transitional stages along the cachexia trajectory. The learned transition matrix suggests that patients in State 0 are more likely to transition to State 2 (2.9%) than to State 1 (1.7%), indicating that State 2 precedes State 1 in disease progression. State 2, interpreted as early cachexia, maintains relatively preserved BMI ($+0.128$), Albumin ($+0.250$), and HGB ($+0.222$) compared to more deteriorated states, suggesting only initial signs of weight loss and anemia. Inflammatory markers such as WBC and Neutrophils remain moderately elevated, but not extreme. These features are consistent with early cachexia, which involves metabolic derangements and mild inflammation.

State 1, in contrast, displays more advanced deterioration. It has significantly lower levels of BMI (-0.212), Albumin (-0.401), HGB (-0.313), and HCT (-0.306) than State 2. Inflammatory markers and liver enzymes (e.g., ALT, ALK) are also mildly elevated, suggesting a worsening inflammatory and catabolic profile. Furthermore, State 1 receives frequent transitions from both State 2 (14.2%) and State 3 (12.9%), indicating its role as an intermediate and potentially unstable phase within the disease trajectory. Patients appear to pass through this state on the way to either improvement or further deterioration, reflecting what may be considered intermediate cachexia.

IX-0c. State 3 – Terminal Cachexia.

State 3 is most consistent with terminal or refractory cachexia. While patients in this state exhibit the not as low values for BMI, Albumin, HGB, and HCT in comparison to that of State 1, (-0.208, -0.352, -0.249, and -0.238 respectively), there are more markers that correlate with late cancer infection. The mean values of lymphocytes (-0.744), basophils (-0.606), eosinophils (-0.662), and absolute eosinophils (-0.559) are all significantly decreased, representing signs of immune system exhaustion. These immune perturbations, when combined with elevated inflammatory markers such as neutrophils ($+0.785$), WBC ($+0.508$), and immature granulocytes (0.360), strongly suggest that State 3 corresponds to a terminal phase of disease where immune exhaustion, inflammation, and metabolic decline converge.

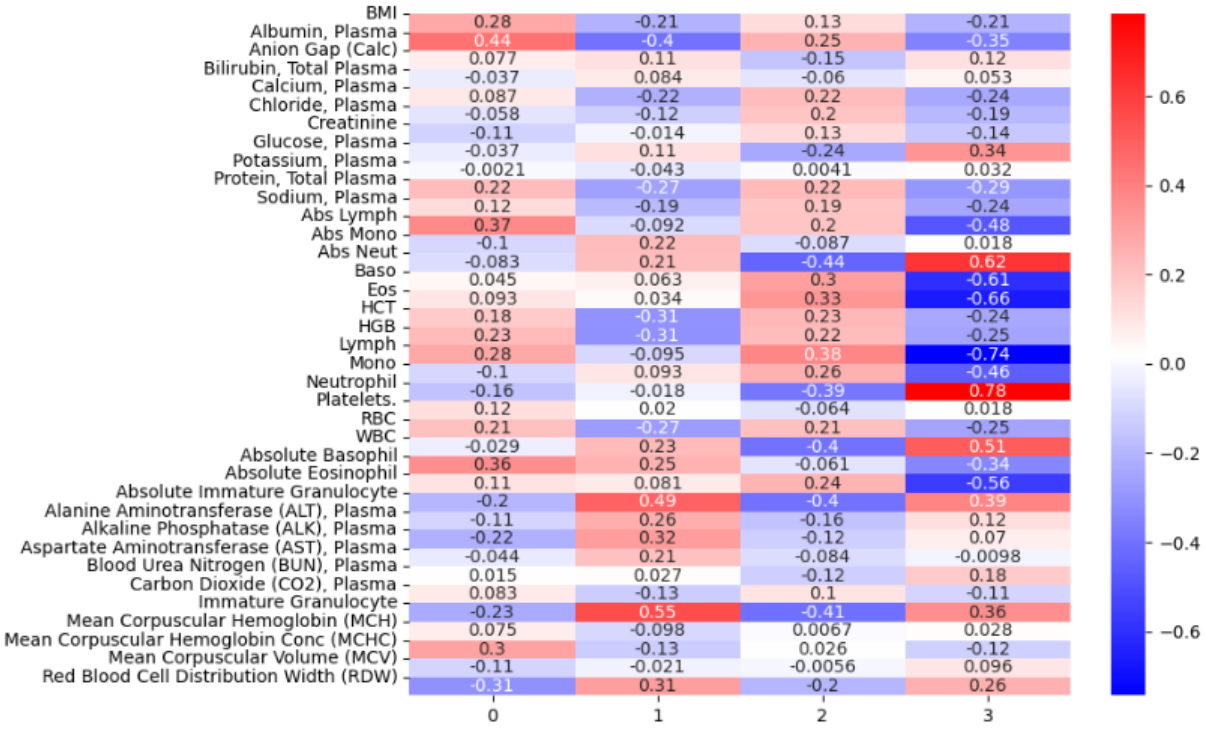

Figure 2. HMM Feature Averages per Hidden State

With a self-transition probability of 76.2%, the patterns shown in State 3 align with definitions of refractory cachexia, characterized by systemic inflammation, significant functional decline, and unresponsiveness to nutritional support or anticancer therapy [3, 26]. Thus, State 3 likely represents the terminal phase of disease progression.

X. Example Patient Trajectory

In this section, we examine one patient trajectoriy where the inferred transitions between cachexia states align closely with the hypothesized progression of the disease. This trajectory follow sthe expected order of events: starting with a relatively stable state (pre-cachexia), progressing to cachexia (where significant weight loss and muscle wasting occur), and eventually reaching refractory cachexia (marked by severe weight loss and potential treatment resistance). These examples illustrate the validity of our model in capturing the typical disease progression for certain patients.

Figure 3 shows the trajectory which correctly follows the hypothesized stages of cachexia. The patient begins in State 0, with a stable BMI of around 28, and transitions to State 1 as the BMI starts to decrease and inflammatory markers show signs of progression. The transition from State 1 to State 2, indicative of the cachexia phase, is clear, as BMI continues to decline while WBC levels rise and Albumin levels decrease. The final transition to State 3, representing refractory cachexia, aligns with a further drop in BMI and stabilization of the inflammatory markers, signaling the patient's progression into a more severe stage of cachexia.

XI. Conclusions and Future Work

This paper presents a detailed analysis of cancer-associated cachexia through the application of HMMs, aiming to provide insights into the stages of cachexia and the clinical trajectories of affected patients. By defining distinct states—pre-cachexia, early cachexia, intermediate cachexia, and terminal cachexia—this research has successfully demonstrated how HMMs can model the progression of cachexia using biomarkers such as BMI, Albumin, and White Blood Cell (WBC) count.

Through the use of various clustering algorithms, including K-means, spectral clustering, and HDBSCAN, we optimized the number of hidden states, leading to a more accurate and dynamic representation of patient trajectories. These models not only reveal the temporal transitions between states but also allow for the characterization of specific patient pathways, highlighting the variability in disease progression. The results provide valuable insights for healthcare providers, offering a more structured framework for understanding cachexia's stages and the factors influencing these transitions. Such

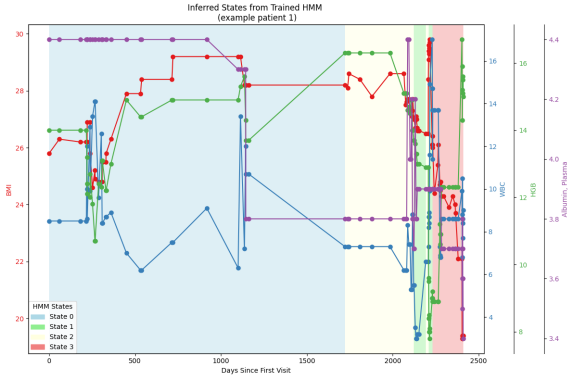

Figure 3. Inferred Cachexia Trajectory of Example Patient 1

a framework could significantly improve early detection, intervention strategies, and personalized treatment for cancer patients suffering from cachexia. The integration of HMMs with clustering methods opens avenues for further research into predictive models for cancer cachexia and similar complex conditions.

Our primary limitation lies in the difficulty of quantifying the uncertainty and precision of our predicted stages of cancer cachexia. Although the hidden Markov model provides a powerful framework for classifying the progression of cachexia, the predictions of the model are highly dependent on clinical data such as BMI, albumin levels, and hemoglobin counts. However, without direct measures of muscle mass decline, such as using imaging techniques such as CT or MRI, we cannot definitively verify the presence or stage of cachexia in individual patients. Future research will involve quantifying the performance of our model by systematically investigating the convergence of the EM algorithm, clustering scores, and the sensitivity of the model to initialization of parameters. As the number of parameters is large, a complete study of regularization should also be conducted. We also note that the model is not yet intended to be used as a predictive model, but instead as a tool to assist researchers in understanding the progression of cachexia in patients.

Incorporating natural language processing (NLP) to analyze unstructured data from doctor's notes would offer valuable insights into the patient's subjective experience, providing a more comprehensive understanding of cachexia progression. Expanding the dataset to include genetic, muscular, and lifestyle factors would also be beneficial, enabling the model to capture additional risk factors for cachexia and potentially improving early detection and personalized treatment strategies. Future work could also include evaluating muscle mass including CT scans specifically; however, an advantage of our method is that we are able to characterize cachexia trajectories without the use of expensive imaging.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[2] W. D. DeWys, "Pathophysiology of cancer cachexia: current understanding and areas for future research," *Cancer Res*, vol. 42, no. 2 Suppl, pp. 721s–726s, 1982.

[3] K. Fearon, F. Strasser, S. D. Anker, I. Bosaeus, E. Bruera, R. L. Fainsinger, A. Jatoi, C. Loprinzi, N. MacDonald, G. Mantovani, M. Davis, M. Muscaritoli, F. Ottery, L. Radbruch, P. Ravasco, D. Walsh, A. Wilcock, S. Kaasa, and V. E. Baracos, "Definition and classification of cancer cachexia: an international consensus," *The Lancet Oncology*, vol. 12, no. 5, pp. 489–495, 2011.

[4] J. M. Argilés, F. J. López-Soriano, B. Stemmler, and S. Busquets, "Cancer-associated cachexia - understanding the tumour macroenvironment and microenvironment to improve management," *Nat. Rev. Clin. Oncol.*, vol. 20, no. 4, pp. 250–264, Apr. 2023.

[5] P. Senesse, E. Assenat, S. Schneider, C. Chargari, N. Magné, D. Azria, and X. Hébuterne, "Nutritional support during oncologic treatment of patients with gastrointestinal cancer: who could benefit?" *Cancer Treat. Rev.*, vol. 34, no. 6, pp. 568–575, Oct. 2008.

[6] I. M. G. da Rocha, A. Marcadenti, G. O. C. de Medeiros, R. A. Bezerra, J. F. d. M. Rego, M. C. Gonzalez, and A. P. T. Fayh, "Is cachexia associated with chemotherapy toxicities in gastrointestinal cancer patients? a prospective study," *J. Cachexia Sarcopenia Muscle*, vol. 10, no. 2, pp. 445–454, Apr. 2019.

[7] V. E. Baracos, V. C. Mazurak, and A. S. Bhullar, "Cancer cachexia is defined by an ongoing loss of skeletal muscle mass," *Ann. Palliat. Med.*, vol. 8, no. 1, pp. 3–12, Jan. 2019.

[8] A. Nicolini, P. Ferrari, M. C. Masoni, M. Fini, S. Pagani, O. Giampietro, and A. Carpi, "Malnutrition, anorexia and cachexia in cancer patients: A mini-review on pathogenesis and treatment," *Biomed. Pharmacother.*, vol. 67, no. 8, pp. 807–817, Oct. 2013.

[9] W. J. Evans, J. E. Morley, J. Argilés, C. Bales, V. Baracos, D. Guttridge, A. Jatoi, K. Kalantar-Zadeh, H. Lochs, G. Mantovani, D. Marks, W. E. Mitch, M. Muscaritoli, A. Najand, P. Ponikowski, F. Rossi Fanelli, M. Schambelan, A. Schols, M. Schuster, D. Thomas, R. Wolfe, and S. D. Anker, "Cachexia: a new definition," *Clin. Nutr.*, vol. 27, no. 6, pp. 793–799, Dec. 2008.

[10] V. E. Baracos, L. Martin, M. Korc, D. C. Guttridge, and K. C. H. Fearon, "Cancer-associated cachexia," *Nat. Rev. Dis. Primers*, vol. 4, no. 1, p. 17105, Jan. 2018.

[11] K. Amano, I. Maeda, T. Morita, M. Baba, T. Miura, T. Hama, I. Mori, N. Nakajima, T. Nishi, H. Sakurai, S. Shimoyama, T. Shinjo, H. Shirayama, T. Yamada, S. Ono, T. Ozawa, R. Yamamoto, N. Yamamoto, H. Shishido, and H. Kinoshita, "C-reactive protein, symptoms and activity of daily living in patients with advanced cancer receiving palliative care," *J. Cachexia Sarcopenia Muscle*, vol. 8, no. 3, pp. 457–465, Jun. 2017.

[12] J. Arends, P. Bachmann, V. Baracos, N. Barthelemy, H. Bertz, F. Bozzetti, K. Fearon, E. Hütterer, E. Isenring, S. Kaasa, Z. Krznaric, B. Laird, M. Larsson, A. Laviano, S. Mühlebach, M. Muscaritoli, L. Oldervoll, P. Ravasco, T. Solheim, F. Strasser, M. de van der Schueren, and J.-C. Preiser, "ESPEN guidelines on nutrition in cancer patients," *Clin. Nutr.*, vol. 36, no. 1, pp. 11–48, Feb. 2017.

[13] Y. Chen, C. Liu, X. Zheng, T. Liu, H. Xie, S.-Q. Lin, H. Zhang, J. Shi, X. Liu, Z. Wang, L. Deng, and H. Shi, "Machine learning to identify precachexia and cachexia: a multicenter, retrospective cohort study," *Support. Care Cancer*, vol. 32, no. 10, p. 630, Sep. 2024.

[14] L. Yin, J. Cui, X. Lin, N. Li, Y. Fan, L. Zhang, J. Liu, F. Chong, C. Wang, T. Liang, X. Liu, L. Deng, M. Yang, J. Yu, X. Wang, M. Cong, Z. Li, M. Weng, Q. Yao, P. Jia, Z. Guo, W. Li, C. Song, H. Shi, and H. Xu, "Identifying cancer cachexia in patients without weight loss information: machine learning approaches to address a real-world challenge," *Am. J. Clin. Nutr.*, vol. 116, no. 5, pp. 1229–1239, Nov. 2022.

[15] D. Ferrara, E. M. Abenavoli, T. Beyer, S. Gruenert, M. Hacker, S. Hesse, L. Hofmann, S. Pusitz, M. Rullmann, O. Sabri, R. Sciagrà, L. K. S. Sundar, A. Tönjes, H. Wirtz, J. Yu, and A. Frille, "Detection of cancer-associated cachexia in lung cancer patients using whole-body [18F]FDG-PET/CT imaging: A multi-centre study," *J. Cachexia Sarcopenia Muscle*, vol. 15, no. 6, pp. 2375–2386, Dec. 2024.

[16] P. Jia, Q. Zhao, X. Wu, F. Shen, K. Sun, and X. Wang, "Identification of cachexia in lung cancer patients with an ensemble learning approach," *Front. Nutr.*, vol. 11, p. 1380949, May 2024.

[17] S. Ahmed, N. Parker, M. Park, E. W. Davis, J. B. Permuth, M. B. Schabath, Y. Yilmaz, and G. Rasool, "Multimodal AI-driven biomarker for early detection of cancer cachexia," 2025.

[18] W.-H. Hsu, A.-T. Ko, C.-S. Weng, C.-L. Chang, Y.-T. Jan, J.-B. Lin, H.-J. Chien, W.-C. Lin, F.-J. Sun, K.-P. Wu, and J. Lee, "Explainable machine learning model for predicting skeletal muscle loss during surgery and adjuvant chemotherapy in ovarian cancer," *J. Cachexia Sarcopenia Muscle*, vol. 14, no. 5, pp. 2044–2053, Oct. 2023.

[19] W. Mu, E. Katsoulakis, C. J. Whelan, K. L. Gage, M. B. Schabath, and R. J. Gillies, "Radiomics predicts risk of cachexia in advanced NSCLC patients treated with immune checkpoint inhibitors," *Br. J. Cancer*, vol. 125, no. 2, pp. 229–239, Jul. 2021.

[20] D. T. Cheng, T. N. Mitchell, A. Zehir, R. H. Shah, R. Benayed, A. Syed, R. Chandramohan, Z. Y. Liu, H. H. Won, S. N. Scott *et al.*, "Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (msk-impact): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology," *The Journal of molecular diagnostics*, vol. 17, no. 3, pp. 251–264, 2015.

[21] F. Q. Nuttall, "Body mass index: obesity, bmi, and health: a critical review," *Nutrition today*, vol. 50, no. 3, pp. 117–128, 2015.

[22] B. H. Tan and K. C. Fearon, "Cachexia: prevalence and impact in medicine," *Current Opinion in Clinical Nutrition & Metabolic Care*, vol. 11, no. 4, pp. 400–407, 2008.

[23] M. A. Farhangi, S.-A. Keshavarz, M. Eshraghian, A. Ostadrahimi, and A.-A. Saboor-Yaraghi, "White blood cell count in women: relation to inflammatory biomarkers, haematological profiles, visceral adiposity, and other cardiovascular risk factors," *Journal of health, population, and nutrition*, vol. 31, no. 1, p. 58, 2013.

[24] M. Komiyama, Y. Ozaki, Y. Miyazaki, Y. Katanasaka, Y. Sunagawa, M. Funamoto, K. Shimizu, H. Yamakage, N. Sato-Asahara, A. Yasoda, H. Wada, T. Morimoto, and K. Hasegawa, "Neutrophil/lymphocyte ratio is correlated with levels of inflammatory markers and is significantly reduced by smoking cessation," *J. Int. Med. Res.*, vol. 49, no. 6, p. 3000605211019223, Jun. 2021.

[25] S. Incir, H. K. Calti, and K. E. Palaoglu, "The role of immature granulocytes and inflammatory hemogram indices in the inflammation," *Int J Med Biochem*, vol. 3, no. 3, pp. 125–30, 2020.

[26] J.-X. Huang, X. Zhang, M. Tang, Q. Zhang, L. Deng, C.-H. Song, W. Li, H.-P. Shi, and M.-H. Cong, "Comprehensive evaluation of serum hepatic proteins in predicting prognosis among cancer patients with cachexia: an observational cohort study," *BMC Cancer*, vol. 24, no. 1, p. 293, Mar. 2024.