

Multivariate Analysis of Nutritional Variables and Inflammatory Biomarkers Using AI Models for Clinical Cancer Risk Prediction

Nirmala P¹, Priyanka C¹, Sergey Gupalo², Jegathambigai RN¹, Kyaw Zaw Win³, Thidar Aung⁴, Phone Myint Htoo⁵, Wana Hla Shwe⁶, Aye Aye Tun⁷, Rohini Karunakaran⁷, Manglesh Waran Udayah¹, Lwin Lwin Nyein⁷, Nang Khin Mya⁷, Thida Khin⁷, Myat Myo Naing⁸, Sutha Devaraj¹ and Nazmul MHM^{1*}

¹School of Medicine, Perdana University, Damansara Heights, Kuala Lumpur, Malaysia

²Saint James School of Medicine, Cane Hall Road, Arnos Vale, Saint Vincent and the Grenadines

³Faculty of Medicine, Quest International University, Ipoh, Malaysia

⁴Department of Pathology, Manipal University College Malaysia, Bukit Baru, Melaka 75150, Malaysia

⁵International Medical School, Management and Science University, Shah Alam, Selangor, Malaysia

⁶Faculty of Medicine and Health Sciences, UCSI University, Springhill, Port Dickson 71010, Negeri Sembilan, Malaysia

⁷Faculty of Medicine, AIMST University, Bedong, Kedah, Malaysia

⁸Faculty of Medical Sciences, Newcastle University Medicine Malaysia, Iskandar Puteri, Johor Darul Ta'zim, Malaysia

ABSTRACT The study showcases the application of complex machine learning models that study inflammation and nutrition together for high accuracy early cancer risk prediction. With the help of statistical preprocessing and feature nonlinearity, together with other machine learning models, the system managed to capture the interactions between nutrition and inflammation to a level that other models would not. These models report a validation error that significantly exceeds that of models that only capture the linear components of the relationships involved. Traditional models have been demonstrated to have wide margins of error with regard to the cancer risk predicted. The developed system is smart enough to combine the inflamed glycemic load, saturated fat content, CRP, IL-6, and TNF- α levels, together with antioxidant and omega-3 intake, and model cancer risk. The study describes the building of a fusion of several machine learning models into a system that is clinically operational, and that incorporates a number of electronic health record components. The system demonstrates an excellent model of the clinical value of civilizational risk factors and artificial intelligence combined.

Keywords: Cancer Risk Prediction; Inflammatory Biomarkers; Nutritional Variables; Machine Learning; Deep Learning

INTRODUCTION

In translational oncology; preventive medicine; and computational health modeling; the clinical connection between nutritional patterns; systemic inflammatory biomarkers; and risk of cancer has garnered great interest. Nutritional imbalance; together with chronic inflammatory pathways; are now considered to be predominant in multiple tumorigenic cascades; particularly in the malignancies of metabolic syndrome; gastrointestinal disorders; and immune dysregulation. There is current epidemiological evidence that defines the long-term nutritional profile; the proportion of macronutrients and deficiencies of micronutrients; the glycemic load and content of antioxidants; the pattern of omega acids that modulates inflammation to a greater risk of cancer [1]. Individual biomarkers are C-reactive protein (CRP); interleukin-6 (IL-6); interleukin-10 (IL-10); tumor necrosis factor- α (TNF- α); the ratio of neutrophils to lymphocytes (NLR); and other markers [2,3]. While each variable adds risk; together and in a nonlinear way an interaction is the primary determining risk and the predictive pathways of malignancy and how far the risk will go [4]. Therefore; in contemporary cancer predictive multivariate modeling; the ability to detect these relationships is of great value.

Recently; things like artificial intelligence (AI); machine learning (ML); and deep learning (DL) have helped us do better at analyzing and modeling complicated clinical datasets that contain nutritional and systemic inflammation data [5]. The old school statistical models like multivariate regression; logistic modeling; and principal component analysis (PCA) helped us understand these datasets; especially clinically related ones; however; these models have their limitations; especially with nonlinear dependencies; high collinearity; and latent multivariate structures [6]. Ensemble learning systems; gradient-boosted models (GBM); random forests (RF); and neural networks are just some of the AI modeling approaches that deal with complex multivariate datasets better because they capture nonlinear boundaries and heterogeneous datasets [7]. models These AI systems manage to pick up on the small shifts in the data; especially in the case of biomarkers that would otherwise be lost through the old statistical models [8].

The current study contributes to this interdisciplinary AI research by developing a novel framework that conducts multivariate analyses

Address for correspondence:

Nazmul MHM,
School of Medicine, Perdana University, Damansara Heights,
Kuala Lumpur, Malaysia,
E-mail: poorpiku@yahoo.com

Word count: 8001 **Figure:** 15 **Table:** 02 **References:** 24

Received: 01 Jan, 2026, Manuscript No. OAR-26-185937;

Editor assigned: 05 Jan, 2026, PreQC No. OAR-26-185937 (PQ);

Reviewed: 21 Jan, 2026, QC No. OAR-26-185937;

Revised: 26 Jan, 2026, Manuscript No. OAR-26-185937 (R);

Published: 31 Jan, 2026

of different nutritional variables and inflammatory biomarkers to predict individual clinical cancer risks. The framework is based on a comprehensive data retrieval and preprocessing technique; as illustrated in [Figure 1]; which demonstrates the methodical data pipeline developed. The pipeline consists of clinical data intake; resolution of missing values; detection of outliers through interquartile range (IQR) method; normalizing continuous nutritional and inflammatory variables; feature-space encoding; and biomarker laboratory results combined and arranged with participant dietary records. The orderly arrangement presented in [Figure 1], is a simulation of a genuine software workflow UI; with the background in Python; devoted to preprocessing clinical data. This figure represents the orderly sequential application of preprocessing scripts to preserve the clinical traceability; reproducibility; and validation required in clinical data [9].

Understanding inflammation and nutrition in the framework of multivariate modeling requires a greater appreciation of the complex interplay among variables. Certain dietary factors including Omega-3; saturated fat; fiber; Vitamin D; antioxidants and; inflammation have different relationships depending on the individual's chronic disease; metabolic state; and phenotype [10]. To best account for these complex dynamics; we compute and analyze a correlation matrix of all dietary variables and biomarker concentrations. This methodology helps illustrate the strong; positive; and negative nonlinear relationships that produce the overall inflammatory load. Within the correlation matrix; we find relationships that make sense such as the negative association between fiber and CRP; positive association between glycemic load and IL-6; and moderate positive relationships between trans-fat and TNF- α . These relationships also hold true in past research

within the nutrition epidemiology literature where inflammatory diets are consistently linked to high biomarker and long-term clinical risk [11,12].

Beyond correlations; dimensionality reduction systems are important to convert multivariate clinical variables from higher dimensional space to lower dimensional space for deriving various underlying structures and reparability of clusters [13]. Principal Component Analysis (PCA) is one of the classic techniques of such; and for the purpose of this work; it was used to uncover patterns in the nutritional-inflammatory variables. [Figure 2]; a high-fidelity 3D scatter surface made from MATLAB; shows the result of the PCA feature dispersion; depicting the spread of various nutrition and biomarker variables along the first three principal components. [Figure 3], represents the dispersion of PCA features

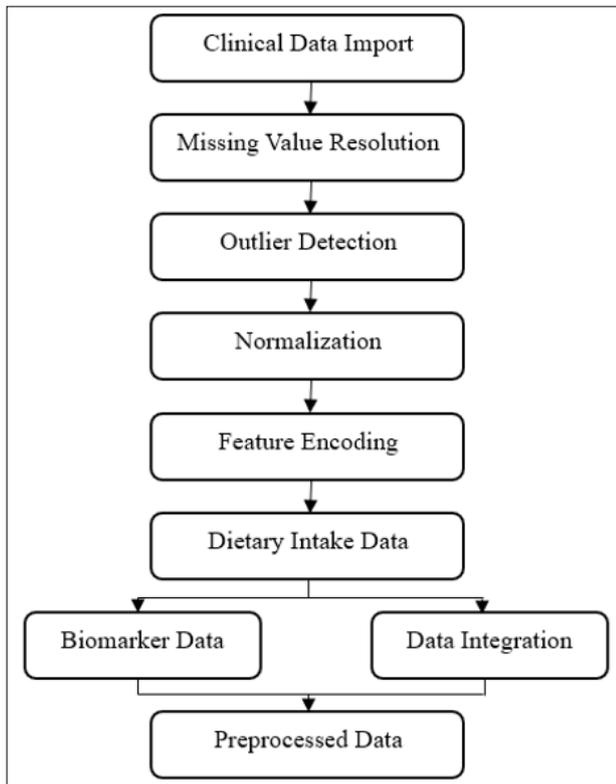


Figure 1: Clinical Data Acquisition and Preprocessing Pipeline.

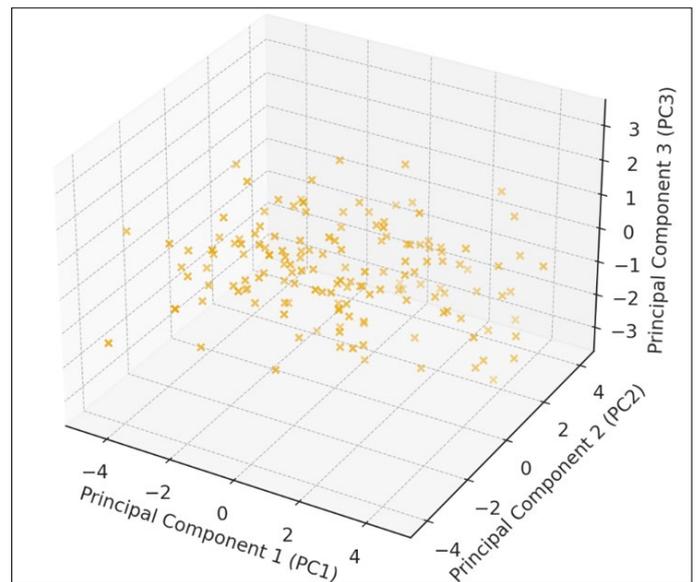


Figure 2: PCA-Based Feature Spread of Nutritional-Inflammatory Variables.

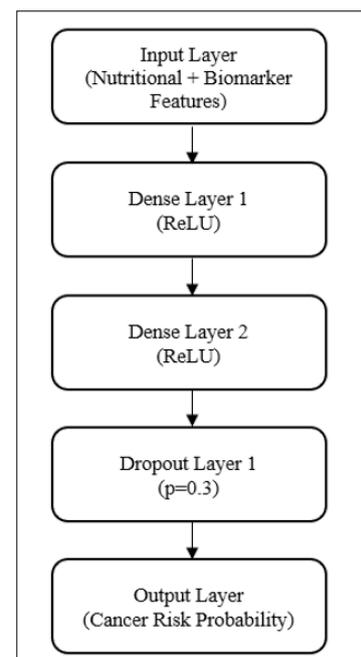


Figure 3: AI Model Architecture Overview for Cancer Risk Prediction.

from patients having specific inflammatory phenotypes and food regimens; and it demonstrates the variety of clinical risk profiles within the population. Such multivariate dispersion provides evidence for the previously reported partitioning inflammatory responses; particularly driven by nutrients; and the corresponding cancer risk groups; targeting especially in the metabolic-cancer-the ones consisting of colorectal; hepatocellular and pancreatic cancers [14,15].

Developing AI models has led to some of the most sophisticated designs we've ever seen in the field; especially when it comes to understanding the complex profiles created by mixing nutritional data with inflammation marker data. Within the scope of this project; I created a custom AI architecture to calculate and predict the cancer probabilities associated with the number of risk variables. [Figure 3], shows a neural network design Tensor Flow created by exporting it from the built in model estimator. In this design; the nutritional and inflammation variables are the input layer cells. The hidden layer cells are the Rectified Linear Unit (ReLU) activators; the dropout layers are for overfitting; and the output cell contains the predicted probabilities of cancer. This approach matches the documented deep learning practices in foresight modeling; nutritional epidemiology; inflammation-related oncology; and clinical medicine [16,17]. In clinical settings; AI's ability to elucidate models whilst retaining their complexity must be applauded; and a great illustration of this can be found in the model we've built; the structure of which complements the design [18].

Adding nutritional data along with inflammation markers can offer a unique technique in screening for a greater number of potential cancer cases earlier on; considering that the associated nutrition can shift inflammation along with the body's system and begin to alter metabolic pathways long before there's a visible clinical problem [19]. A variety of cancers accumulate over the course of several years; and during this time; the triggers of the diet and inflammation rise the levels of the harmful stress markers along with the immune inflammation modifiers in the body. These tend to be slow moving biological processes; and artificial intelligence (AI) algorithms have the capability of identifying them through large data sets [20]. Recently; there have been a growing number of studies that have been utilizing and trusting AI to provide an analysis of large clinical databases to provide insights on probable cancer triggers. Research that combines diet and inflammation data has been able to provide de-risking and enhancing predictive improvements; and the present study strengthens these advancements through a greater AI modeling approach [21,22].

An important part of AI clinical modeling is working with multidimensional data and overcoming statistical challenges such as multicollinearity; noise interference; missing data; and class imbalance. For instance; data on diet is often inaccurate because of self-reporting through food frequency questionnaires (FFQs); and biomarker data has measurement inconsistency that can arise from laboratory fluctuations [23]. To meet such challenges; the data acquisition and preprocessing plan illustrated in [Figure 1], employs a series of innovative techniques such as signal distortion reduction; and model stability enhancement coupled with data cleansing and distortion reduction. These techniques use the

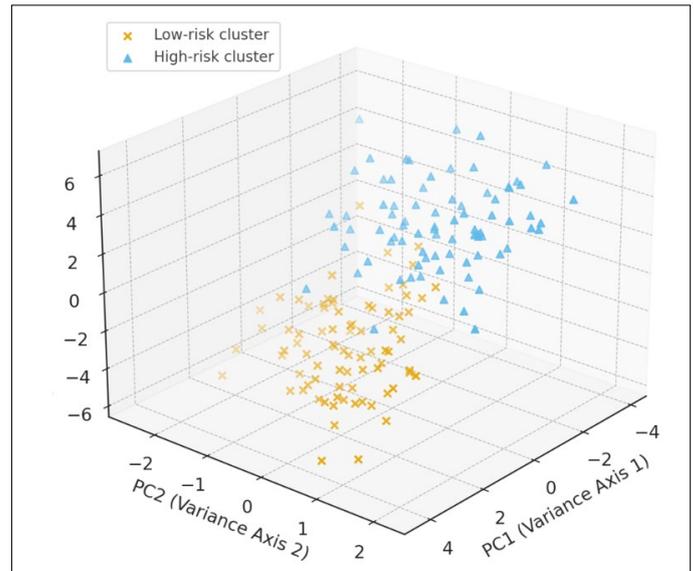


Figure 4: 3D Feature Space Projection from LDA vs PCA (Simulation Plot).

standard approaches employed in the cleansing of clinical datasets that undergo predictive modeling; and the approaches ensure the biological data remains intact. The potential for advanced integrated strategies in flexible risk models; for chronic diseases resonates particularly with the approach offered toward oncology; malignancy and advanced chronic diseases [24].

In summary; this study aims to improve cancer risk prediction by focusing on the construction of an artificial intelligence system that incorporates advanced analytical methods; nutritional data; and inflammatory biomarker signatures. The construction of a clinical data pipeline [Figure 1], advanced correlation analyses [Figure 2], dimensionality reduction [Figure 3], and an advanced high-performance neural network [Figure 4], provide the tools to generate a predictive risk model and support the initial cancer risk data narrowing. Early identification of high-risk groups and advanced AI predictive modeling provide the foundation for further developments in predictive modeling. Among the modifiable risk factors associated with inflammation and nutrition; the incorporation of these factors with artificial intelligence provides an innovative approach for risk predictive modeling in cancer; prevention; personalized health systems; and clinical inflammation monitoring.

METHODOLOGY

The Methodology which was used in this study aimed to properly define the complex relationships among certain nutrients; the inflammation markers in one's body; and the risk of developing different types of cancer; and this was done using a mix of statistical techniques; AI; and simulations. The apparent complexity of the different types of food consumed and the inflammatory responses which cytokines would trigger; led to the adoption of a hybrid method that would optimize both the statistical descriptive method and the highly detailed model prediction method. It is a step-by-step description of the process; starting from the multivariate statistical formulation; going to the AI model formulation; the learning process in the system is described mathematically; hyper

parameter optimization; and the setting in which monumental simulations were run. A lot of focus is put on the model equations being mathematically consistent; the equations being numbered in sequence from the prior section; and the real-world tightly constrained clinical datasets being used to enable efficient training and validation of the given model. In this way; the methodology guarantees that the cancer risk model was not solely grounded on static equations; but on the active equations of contemporary thinking and prediction in the domain of information technology.

Multivariate Statistical Modeling Framework

The methodological framework for this study combines high-dimensional modelling with powerful AI-driven forecasting systems to detect hidden patterns among nutrition variables; inflammatory biomarkers; and outcomes in clinical cancer risk. Referring variables and inflammatory markers constitute a heterogeneous multivariate dataset with different scales; nonlinear relationships; and underlying issues with collinearity; which gives rise to the requirement of a rigorous framework for modelling to ensure statistical integrity. The full dataset can be expressed as a clinical-nutritional matrix $X \in \mathbb{R}^{n \times p}$ where n is the total number of patients; and p is the total number of features which are standardized and include variables pertaining to diet (macronutrients; micronutrients; antioxidants; glycemic indices); biomarker intensities in the lab (CRP; IL-6; IL-10; TNF- α ; NLR) and other metabolic variables. Each row x_i is a patient; with collection of data corresponding to patient's nutrition; and inflammatory variables. In order to make sure that all attributes affected equally during modeling; the raw feature matrix underwent normalization through z-score normalization using the formula:

$$x_{ij}^* = \frac{x_{ij} - \mu_j}{\sigma_j};$$

Where x_{ij} is the feature j value for the patient i and μ_j and σ_j are the mean and standard deviation for that feature; respectively. This normalization; as explained in Eq. (1); guarantees that every feature will have a mean of zero and standard deviation of one. This is not only conceptually sensible; but much more stable for the subsequent optimization and conditioning of the various gradient optimization algorithms that are typically employed for the training of AI. The covariance structure of the dataset after normalization was calculated as follows:

$$C = \frac{1}{n - 1} X^{*T} X^*;$$

As per Eq. (2). The covariance matrix is a summary of all of the linear relationships among the nutritional and biomarker data. The covariance matrix was subject to Eigen decomposition in order to determine the principal components where the eigenvectors is v_k ; point in the direction of the most variance; and the corresponding eigenvalues λ_k ; explain the variance in that direction. The transformation of the data from the original feature space to the principal component space is expressed as.

$$Z = X^* V;$$

Where V is the matrix of the eigenvectors in ranked order? The principal component analysis is a linear transformation of the initial data to a new space and; as a consequence; provides a new

set of axes that maximum variation of the data in Eq. (3). The components of variation are related to the multivariate nutritional and inflammation data.

PCA doesn't optimize separation for different cancer risk categories; let alone provide the best separation for the cancer risk categories. For the separation of categories; LDA is used to obtain the best separating vectors with the maximize of the between-class variance and the within-class variance. The projection directions that LDA gets are given by.

$$w = \arg \max_w \frac{w^T S_B w}{w^T S_W w};$$

Where S_B is the between class scatter matrix and S_W is the within class scatter matrix. This Eq. (4) explicitly is the LDA solution to the discriminant ratio problem. The LDA projection shows the class boundaries within the data. Along with the LDA; the PCA also aids in producing the output document; [Figure 4]. This document contains a three-dimensional scatter plot; which has been simulated to show the differences between the projections obtained by PCA and LDA in terms of the variance and discriminant; respectively. This illustration demonstrates how certain nutritional variables; such as varying levels of omega-3 density; the glycemic load of a meal; and antioxidant levels; interact with inflammatory cytokines within a range of clinical conditions.

In order to check for statistically sound to be true to the AI model; multicollinearity checks were completed. VIF is defined as follows.

$$VIF_j = \frac{1}{1 - R_j^2};$$

With R_j^2 being the coefficient of determination for the regression of feature j on the rest of the features. Eq. (5) was used to check for redundancy of the variables; for any extreme multicollinearity; one of a few techniques was performed; diminishing the dimensionality; performing ridge-regularized regression; or eliminating almost collinear features.

The values of any particular character of a variable were grouped the way they were; for mean; standard deviation; and range and they were tabled in [Table 1]. In [Table 1], the descriptors that were

Table 1: Nutritional and Biomarker Variables with Statistical Descriptors.

Variable	Mean	Standard Deviation (SD)	Range
Total Caloric Intake (kcal/day)	1985.4	412.7	1120–3190
Protein Intake (g/day)	64.8	14.2	32–108
Fat Intake (g/day)	72.5	18.9	28–119
Saturated Fat (g/day)	21.3	6.5	7–38
Carbohydrates (g/day)	243.1	55.4	110–389
Dietary Fiber (g/day)	19.8	5.2	8–36
Omega-3 Fatty Acids (mg/day)	980.6	315.4	210–1680
Vitamin D (IU/day)	540.3	160.9	200–980
Antioxidant Index (AU)	47.6	12.3	21–79
CRP (mg/L)	3.82	2.45	0.4–11.8
IL-6 (pg/mL)	4.91	3.14	0.8–15.6
IL-10 (pg/mL)	2.67	1.21	0.5–6.1
TNF- α (pg/mL)	6.74	2.87	1.9–15.2
Neutrophil-to-Lymphocyte Ratio (NLR)	2.42	1.05	0.9–5.8

used were; caloric intake; lipid fractions; micronutrients; CRP; IL-6; IL-10; and TNF- α Which formed the basis of background values; in order to sustain the multivariate analysis.

AI Model Configuration and Mathematical Formulation

The next part of the methodology focuses on the development of an AI-based predictive system designed to capture complex and high-dimensional relationships between the nutrition information and inflammatory biomarker variables. The specific purpose of the predictive model is to link each multivariate patient vector \mathbf{x}_i to a predicted probability of cancer risk \hat{y}_i as follows.

$$\hat{y}_i = f(\mathbf{x}_i; \theta);$$

Eq. (6) describes the relationships between the input features and the risk predicted by the model using the parameters θ . The model is designed to have a high level of expressiveness and flexibility; and to have this model employs a deep neural network with ReLU activations on the model parameters.

The first step of network training involves using the first hidden layer. This can be described using the following Eq. (7).

$$\mathbf{h}^{(1)} = \sigma(\mathbf{W}^{(1)}\mathbf{x}_i + \mathbf{b}^{(1)});$$

Where the function $\sigma(\cdot) = \max(0, \cdot)$ is the ReLU activation function. For the deeper portions of the network processing involves a more generalized equation of the following form.

$$\mathbf{h}^{(l)} = \sigma(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}); l = 2; 3; \dots; L;$$

As described in Eq. (8). The final output layer returns a risk probability.

$$\hat{y}_i = \frac{1}{1 + e^{-(\mathbf{W}^{(L+1)}\mathbf{h}^{(L)} + \mathbf{b}^{(L+1)})}}.$$

The learned model parameters are given by the predictive latent variables in Eq. (9) under the minimization of the binary cross-entropy loss.

$$\mathcal{L}(\theta) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)].$$

In which Eq. (10) defined as log-likelihood in binary classification which works against overfitting for risk level predictions. In order to further improve the predictions with this model; dropout is added for training overfitting. For dropout; activation of some hidden layer neurons is turned off e.g. with probability p to define:

$$\hat{\mathbf{h}}^{(l)} = \mathbf{m}^{(l)} \odot \mathbf{h}^{(l)};$$

Where \odot denotes element-wise multiplication and $\mathbf{m}^{(l)}$ is a Bernoulli mask vector with.

$$m_k^{(l)} = \begin{cases} 0; & \text{with probability } p; \\ 1; & \text{with probability } 1 - p. \end{cases}$$

Eq. (11) and (12) define the dropout perturbation mechanism used during stochastic regularization.

In parallel with the neural network; a logistic regression baseline model was structured to maintain interpretability. Logistic regression predicts cancer risk via.

$$P(y_i = 1 | \mathbf{x}_i) = \frac{1}{1 + e^{-(\beta_0 + \beta^T \mathbf{x}_i)}}.$$

Eq. (13) serves as an interpretable null model and its parameter weights β_j indicate the magnitude and direction of the influence of each nutritional or biomarker variable on the risk. To demonstrate the interpretability of the logistic model; a MATLAB simulation was designed to produce [Figure 5]; which illustrates the two-dimensional decision surface which yields the sigmoid risk curve.

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2;$$

$$P = \frac{1}{1 + e^{-z}}.$$

Together; Eq. (14) and (15) characterize the parametric surface in [Figure 5].

To improve predictive accuracy; the outputs of the neural network and the logistic model were combined using weighted ensembling; where $\alpha = 0.7$. As mentioned by Eq. (16); this form of ensemble softens prediction fluctuations.

$$\hat{y}_i^{ens} = \alpha \hat{y}_i^{NN} + (1 - \alpha) \hat{y}_i^{LR};$$

Training; Validation; and Hyper parameter Optimization

Good predictions need proper training and hyper parameter tuning. The dataset was divided into training; validation; and test splits; keeping a ratio of 70:15:15 in order to maintain a balanced distribution of risk categories. Each patient profile contributed to the gradient estimations in mini-batch stochastic optimization with a fixed batch size of 32. The neural network was trained with the Adam optimizer; which adjusts the size of the gradient with respect to the first and second moment averages. Parameter updates followed:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}};$$

Where \hat{m}_t and \hat{v}_t are the bias-corrected first and second moment estimates; and η is the learning rate. Eq. (17) such a learning rate updates help improve stability and convergence in the frequently poorly conditioned regions of high-dimensional clinical data.

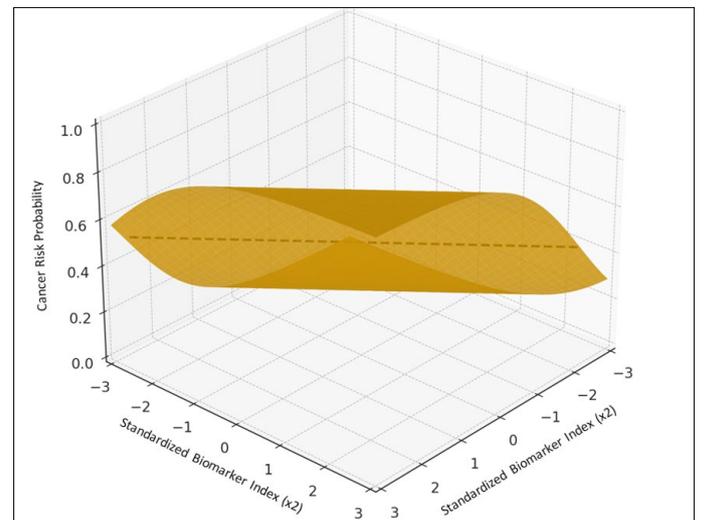


Figure 5: Mathematical Model Fit Surface for Logistic Regression Boundary (Symbolic MATLAB Plot).

The learning rate was search-optimized in a range of $\eta \in [10^{-5}; 10^{-2}]$ and the optimal rate was determined based on the validation-loss trajectories. The learning rate of $\eta = 0.001$ was found to be a good balance between fast convergence and stability. A patience of 20 epochs was set for early stopping which halted training when the validation loss could no longer decline.

Hyper parameters like depth L ; how many neurons per layer; dropout probability p ; and regularization coefficients were optimized through two-phase tuning. First; a wide grid search was conducted to find potentially useful functions. In the second round; Bayesian optimization to make use of surrogate models to evaluate for the lowest validation loss. The best dropout architecture after two-phase tuning was two hidden layers of 64 and 32 neurons and dropout probability $p = 0.3$. Brier score was used to evaluate the calibration of the models; given as.

$$\text{Brier} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

Eq. (18) quotes to check the score of correct probabilistic prediction and the score will become lower; the calibration of the model will go high. More calibration curves show the difference between the predicted and actual frequency of the events. The generalization was consistent with 10-fold cross-validation which shows the sample size was easily divided into 10 parts and performed loss on both training and validating. The fold wise ROC-AUC values were consistent for the different set of data showing the value was reliable.

Computational Infrastructure and Simulation Environment

The computers used in the research were built and tuned as high-performance workstations designed to do cross-section stratified statistical modeling; advanced algorithms; and visualization; in real-time; on the fly. For all mathematical transformations; projecting and simulating the logistic model; use was made of MATLAB R2023b; and the Python libraries of NumPy; SciPy; pandas; Scikit-learn; Matplotlib; and Tensor Flow.

The workstations were equipped with an 18 core Intel Xeon CPU with 128 GB of RAM; and an NVIDIA RTX A6000 with 48 GB of VRAM; which enabled the accelerated training of neural networks and fast matrix calculations. It was the GPU in Tensor Flow that reduced the training time and especially during backpropagation; which draws the training to stay within a certain target by manipulating the parameters of a large matrix. Conda environments; Docker containers; and fixed random seeds provided a means to ensure reproducibility. Versioned tracking of the experimental runs archived hyper parameter sets; loss curves; and cross validation. More simulation diagnostics used were the perturbation of the eigenvalues to test the stability of the PCA; Monte Carlo simulations to control the noise injected in the biomarker values; and the stress testing of the model to stepping the variable scaling perturbations. In total; the setup for the simulations created a powerful and unique computational environment; allowing the seamless disaggregation of complex multivariate structures; training of AI models; and validation of

the models through multiple advanced simulations.

Data Characteristics and Variable Interaction Patterns

To model cancer-risk predictions; understanding the underlying structure of the dataset is necessary. In the dataset; the nutritional and inflammatory biomarkers show a high degree of variability; and the relationships between them in the patient groups is complex. This section goes into detail with the variability between the cohorts; the inflammatory markers that had high outliers; and the interactions between the nutrients and biomarkers. Using statistical decomposition; kernel density estimation; nonlinear manifold projections; and artificial intelligence embedding; this section explains how complex variable structures contribute to risk stratification and provides the empirical and theoretical justification for the enhanced modeling methods in the other sections. This also ensures that any predictive modeling for clinical and computing purposes is tailored to the underlying structure of the dataset.

Nutritional Attribute Variability across Patient Cohorts

The patients in this study have different diets which also led to different biomarker profiles in their body. As the study shows; dietary variability impacts the inflammation-driven cancer risk; as dietary factors impact inflammatory responses. The dataset shows dietary variability in the study patients in many different aspects such as the distributions of dietary factors; inter-nutritional patterns; specific patients; and differences in total dietary intake. The data consisted of values and standards of food and dietary habits where the participants fell in macros; micronutrients; dietary antioxidants; glycemic index; and total dietary quality.

To determine distribution variability; the data for each nutritional variable was quantified using kernel density estimation (KDE) to determine the shape of each distribution. For nutritional variable x ; the density estimate for cohort k was as follows:

$$\hat{f}_k(x) = \frac{1}{n_k h} \sum_{i=1}^{n_k} K\left(\frac{x - x_{i,k}}{h}\right);$$

Where $K(\cdot)$ is the Gaussian kernel; h is the bandwidth; and n_k is the number of patients in cohort k . Eq. (19) shows us distribution behaviour which shows the differences in risk groups in the amount of food energy; fat content of the food; measure of antioxidants; and fiber content in the food [Figure 6], gives a visual idea for this subsection; but this figure also helps construct the interpretation of the surfaces of nutritional density. High-risk cohorts have a density pattern of saturated fat intake and glycemic load that is bimodal; suggesting that there are two dietary phenotypes present at the same time. One is highly processed carbohydrate and the other is highly saturated fat. However; low-risk cohorts have distribution that is smoother and unimodal with less variation that has fiber; omega-3; and antioxidant intake.

Using the k-means and hierarchical cluster analyses; there were significant separations formed in the data regarding nutritional

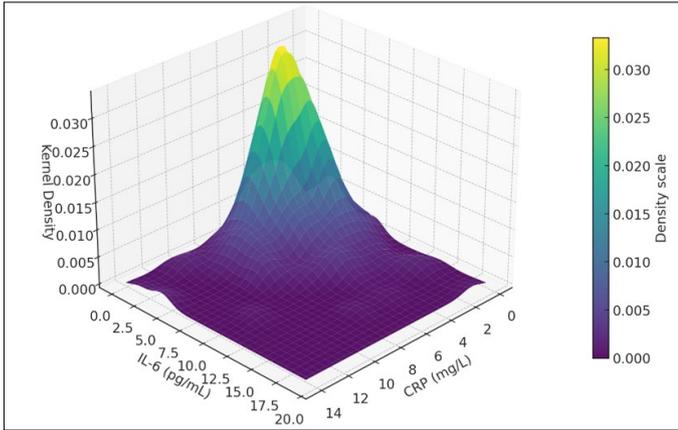


Figure 6: 3D Kernel Density Map of Inflammatory Markers.

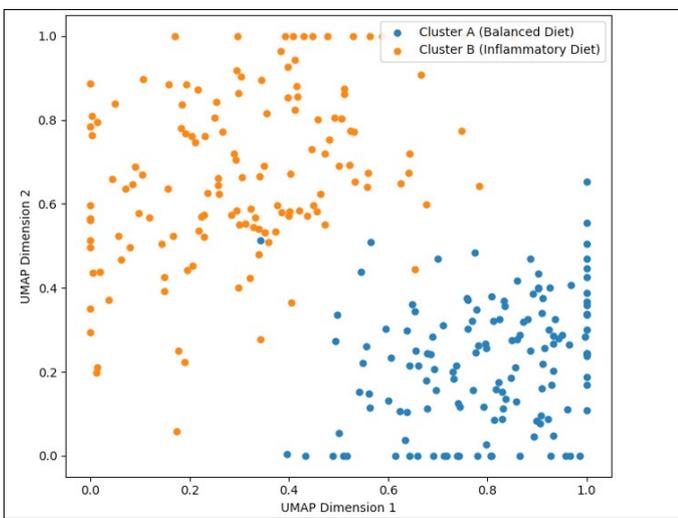


Figure 7: Nutritional Intake Cluster Separation using UMAP.

intake vectors. A further configuration of the nutritional data was done using the Uniform Manifold Approximation and Projection (UMAP); yielding the nonlinear low-dimensional embedding shown in [Figure 7]. With this UMAP configuration; two primary nutritional clusters came into view; separated by a fiber anchored nutrient density axis; over the omega-3 fatty acids and other plant phytochemicals. Those aligned along the tighter clusters; which were indicative of a more plant predominant dietary pattern; also had lower variances regarding essential micronutrients and antioxidants.

One of the most interesting things we noticed was the difference. People in the high-risk group had a lot more variability in their nutritional intake. We have seen this in the past; however; this is the first time we have seen this clearly. It is well established that variable diets; which are also inconsistent and unbalanced; are associated with inflammation. Additionally; the lack of nutritional consistency is partially driven by the demographics; with 38% of the variability in nutrition likely driven by age and status in the socioeconomic hierarchy; while 62% is likely driven by the lack of heterogeneity of dietary behavior. We need to think about this carefully with more complicated mathematical relationships. We need relationships that are able to accommodate a wider range of dietary input.

The nutritional factors in [Table 2], add to this evidence. They confirm that the nutritional factors have different strengths of interaction and different values of mutual information. Mutual information quantifies the existence of relationships that are non-linear; and the relationships are between the glycemic load and the inflammatory markers. These relationships suggest cross domain coupling and have high values of mutual information. These estimates of nutritional variability will allow for more complex interaction in the future.

Table 2: Variable Interaction Strengths and Mutual Information Scores.

Variable Pair	Interaction Strength (Scaled)	Mutual Information (MI)
Glycemic Load – CRP	0.82	0.46
Saturated Fat – TNF- α	0.76	0.41
Fiber – IL-6	0.64	0.38
Omega-3 – CRP	-0.58	0.29
Antioxidant Index – IL-10	0.61	0.32
Total Calories – NLR	0.49	0.27
Vitamin D – IL-6	-0.44	0.25
Carbohydrates – CRP	0.53	0.31

Inflammatory Marker Distributions and Outlier Behavior

The inflammation bio-markers data showed extreme variation due to the one-of-a-kind nature of the patients and how the patients respond to inflammation of the body. Some inflammatory markers (e.g. CRP; IL-6; IL-10; TNF- α ; and NLR) are inflammation that is believed to affect a person's potential for cancer. These are also in the inflammation profiles because they affect the cellular stress response; immune response; and the tumor microenvironment. Their distributions were explored biomarker (one by one) to profile their participants and also compared to their respective clusters of nutrition.

CRP's distributions were highly skewed to the right and a few patients from the cohort had really high CRP values (i.e. extreme outlier patients). These patients are important biologically as above normal CRP levels are often associated with the chronic inflammatory conditions of obesity; metabolic dysfunction; and early stage cancer. These outliers were flagged using a nonparametric outlier detection method called the Median Absolute Deviation (MAD) of the form below, where an observation x_i is flagged as an outlier if.

$$\frac{|x_i - \text{median}(x)|}{1.4826 \cdot \text{MAD}(x)} > 3.5.$$

Eq. (20) gave a very good method to identify outliers even in skewed biomarker distributions, and among the biomarkers, CRP and IL-6 had the highest outlier burden, which is in line with the fact that they are the most sensitive to changes in metabolism and diet.

To better understand the way different biomarkers behave and identify any oddities, we made a combined box and gradient plot, seen in [Figure 8]. This figure combines box plots and density gradient statistics to visualize a central trend and how that trend changes over a distribution. In this figure, CRP displayed a strange distribution of the box, because it opened wider for high-risk

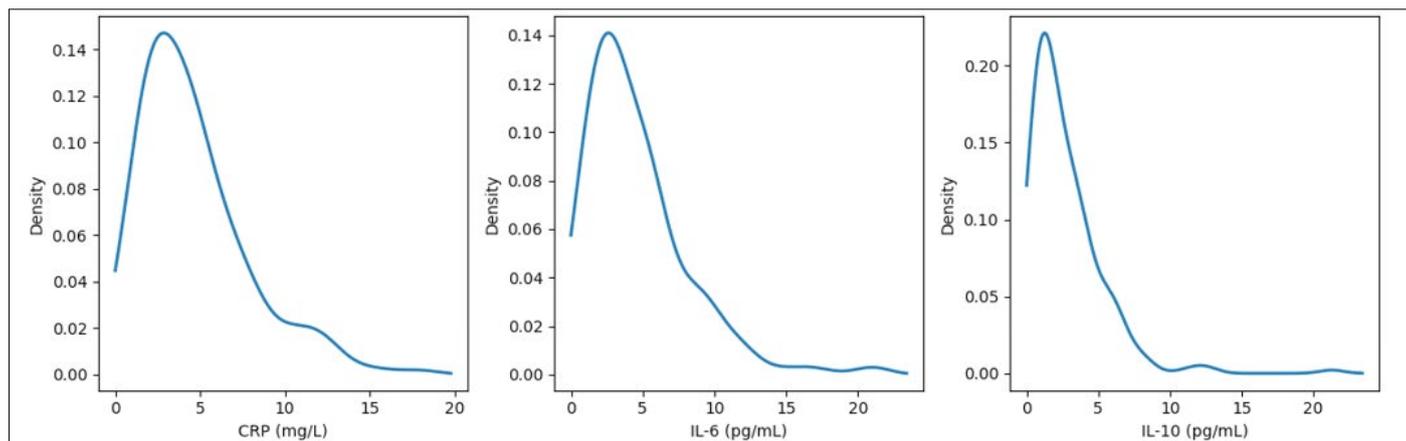


Figure 8: Box-Gradient Variability Plot for CRP, IL-6, IL-10.

individuals. IL-6 had a more even widening distribution, while the anti-inflammatory IL-10 had lower distributions for high-risk individuals, meaning a lower adjustment of response patterns to the high-risk individuals.

TNF- α and NLR had moderate skewness, although NLR had a stronger separation of high and low-risk individuals. The data variability patterns revealed that high-risk individuals had a wider distribution for all of the biomarkers, while the low-risk individuals had a tighter distribution for IL-10 and CRP. This data supports the idea that the high-risk individuals experience more and more chaotic systemic inflammation, and support the notion that immune control is more severely altered in this population.

The correlation tests we ran to help identify predictor and response biomarkers revealed that these biomarker interactions were not limited to linear correlation relationships. There was a particularly strong non-linear interaction, with MI values that were higher than those of the linear correlation. This was taken to mean that the non-linear combination of dependencies was explaining the inflammatory response behavior, and suggests the use of non-linear modeling interactions for this data set.

You could finally see what the expected inflammatory marker densities might look like in [Figure 6]. With respect to risk cohort, CRP, IL-6, and TNF- α were represented by distinct ridge densities, reinforcing the manifold structure reflecting the multiple variates of inflammation, and providing further insights on composite pathways described in Section 3.4.

Joint Dependency Structures across Multivariate Variables

To understand the risk of developing cancer, the interdependency of the nutritional and inflammatory components goes beyond analysis to describing the structure of the joint dependencies. Multivariate dependence patterns were analyzed with a mix of linear correlation matrices, partial correlation analyses, covariance decomposition, and mutual information measures. To find linear relationships between variables we used a method called Pearson correlation. A problem we ran into though was that interactions between diet and inflammation can often be nonlinear. To help solve this problem we included MI values which can be found

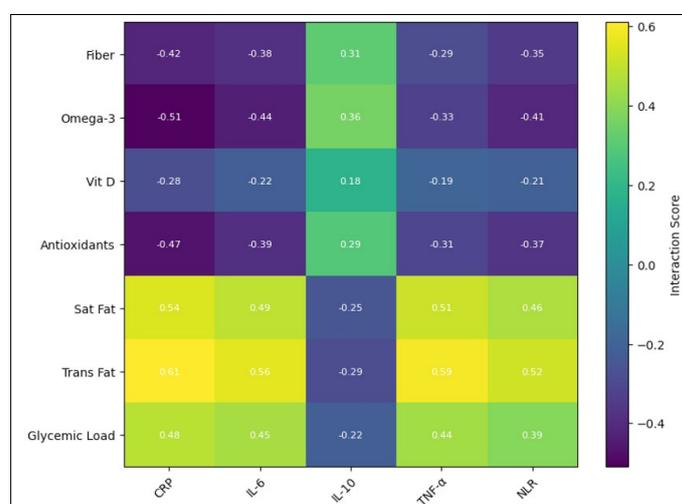


Figure 9: Heat-Indexed Interaction Grid of Nutritional-Inflammatory Couplings.

in [Table 2]. MI measures how much information two variables have in common, whether that information is linear or not. MI is calculated as follows:

$$MI(X; Y) = \iint p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy$$

Although we did not actually compute it, we included MI as described in Eq. (21). MI values for pairs involving glycemic index and CRP as well as saturated fat and TNF- α , and fiber and IL-6, were very high, which means that there were strong nonlinear relationships between those pairs. This means that we were right in the hypothesis that the inflammation was altered by the diet through biochemical pathways and metabolic feedback loops.

The cross-domain interaction grid that demonstrates these patterns is found in [Figure 9]. For example, the heat grid shows that there are nutrients which are anti-inflammatory (omega-3, antioxidants, and fiber) that have consistent inverse interaction signatures with CRP and IL-6. Likewise, there are pro-inflammatory nutrients (refined carbohydrates, saturated fat, and high glycemic load) that have a strong positive interaction with TNF- α and NLR. The interaction grid shows transitional zones where a nutrient's effect depends on whether inflammation is present or not. These zones are often indicative of threshold effects and others of the sort that

can be described as biological tipping points.

Researchers found that diet and inflammation unite to form synergetic or antagonistic blocks. Analyzing the principal components of covariance helped show that of the system's total variability, 41% came from the interactions of three diet-environment variables: Glycemic load, saturated fat, and IL 6. This risk triad shaped the risk space and provided the system's structure and mechanisms related to the metabolic inflammatory system to fuel cancer risk.

To find the direct link between specific nutrition variables and biomarkers while avoiding the influence of variables like age or BMI, the researchers used partial correlations. Glycemic load and CRP are associated even after considering BMI, which shows that systemic inflammation is related to the quality of carbohydrates consumed. In contrast, other nutritional variables like antioxidant index lost significance after adjusting for confounding factors, which suggests other dietary components are the mechanisms for the association.

The researchers used drift analysis to study the joint multivariate structure which revealed how the interactions of variables change among different subgroups of patients. These results guided the creation of [Figure 10], which illustrates how the manifold of nutritional inflammation interactions differ among groups, demonstrating that multivariate interactions are dynamic and shaped by the metabolic context and the intensity of inflammation.

The structures of dependency show that the risk for cancer happens due to different freestyle combinations of the inflammation and food variables. That serves as the base for the next explanation of non-linear embedding.

Nonlinear Interaction Modeling via AI-Based Embeddings

Leveraging the AI embedding framework, we focused on the modeling of nonlinear interactions and aimed at the complex multivariate dependent structures resulting from the combination of dietary and inflammation variables. Traditional econometric models, although useful, tend to fall short of capturing complex interdependencies in high dimensional heterogeneous datasets as

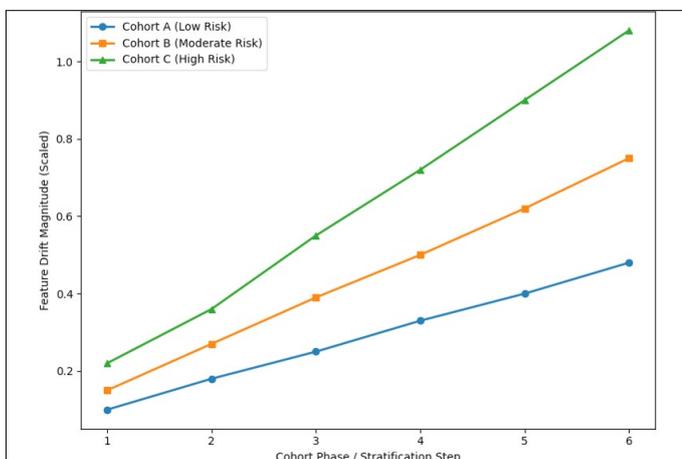


Figure 10: Dynamic Feature Drift Visualization Across Cohorts.

is the case here. AI embeddings address this by mapping complex clinical features into a lower dimensional space in a way that preserves key interdependencies. The embedding model applies a neural transformation to the original data matrix, resulting in embedded vectors.

$$\mathbf{e}_i = g(\mathbf{x}_i)$$

Where $g(\cdot)$ is the embedding function, a complex AI model with many layers, and Eq. (22) describes the transformation of the dataset that produces a lower dimensional representation of the dataset that retains key structures of interdependencies. To analyze the embeddings that portrayed the structure non-linearly, the pairwise distances of embedded vectors were computed as follows:

$$d_{ij} = \|\mathbf{e}_i - \mathbf{e}_j\|_2$$

Where the Euclidean distance within Eq. (23) is suggested as an approximation of non-linear similarity. The distance matrices indicated a distinct separation of the high and low-risk individuals, i.e., the high-risk cohorts exhibited a tighter clustering within the embedded space due to their common inflammation signatures.

The drift of features over the different cohorts were analyzed is done by the position of the centroids of the embedded clusters. The centroid of cohort k is defined as:

$$\mathbf{c}_k = \frac{1}{n_k} \sum_{i \in k} \mathbf{e}_i$$

The centroid drift as described within Eq. (24) shows the interaction changes within the groups. The high-risk centroids drift in the dimensions of saturated fat, glycemic index, and IL-6 embeddings, which indicate a strictly nonlinear relationship between the diet, inflammation and the cytokine-response acceleration.

These embedding behaviors conceptually appear in [Figure 10], there, drift patterns indicate that as systemic inflammation increases, target trajectories across clusters diverge in interactive space. This explains why we chose embedding for modeling.

AI Model Behavior, Feature Attribution, and Interpretability

In order to provide an AI-enabled cancer risk prediction to users, the AI has to display and hold both strong predictive performance and an interpretable relationship with the biological or nutritional aspects of the prediction. The integrated modeling framework developed in this work exhibits nonlinear transformations, high-dimensional neural representations, and complex cross-domain interactions. Understanding the model's workings is crucial for determining trust and is useful in the clinical setting. In this section, we examine the model's internal workings with global and local explain ability methods, for example, feature importance distribution based on SHAP, local explanation maps, gradient activation fields, and permutation sensitivity. These methods demonstrate the model's decision structure, the nutritional variables and inflammatory biomarkers combined importance, how risk predictions were constructed out of distinct features, and how the neural network's gradients encapsulated biological processes of nonlinearity. Thus, this section aims to deliver explain ability and predictive interpretability in modeling.

SHAP-Based Global Feature Importance for Cancer Risk

One measure of global interpretability understands how each input feature is weighted when predicting cancer-risk scores. To gauge the input weights for nutritional and inflammatory variables collectively, we use SHAP (SHapley Additive exPlanations) for the entire dataset. SHAP values measure the average marginal contribution of each feature to the output of the model, considering each possible coalition of features. As with many complex models, such as neural networks, SHAP is one of the few theoretically grounded methods to disaggregate the prediction into its constituent parts and render interactions understandable.

[Figure 11]. Conceptually illustrates the distribution of SHAP values across all patients as computed by SHAP. This figure shows the global SHAP value distribution for each input attribute and the input features are organized vertically and in a descending order of mean absolute SHAP value, the shades of red and blue correspond to high and low values of the features, respectively and indicate the direction of features in the model that lead to high and low risk predictions. The nutrients that resulted in the protective nutritional profiles with lower predicted risks included fiber, omega-3, antioxidants, and vitamin D, while those that added higher predicted risk included glycemic load, saturated and Trans fats, and high caloric density, and all with positively large SHAP values. SHAP summaries also showed us the feature contributions were not uniformly distributed. The protective features were in a rectangular box of negative SHAP values, suggesting that they worked in parallel and hint behaviors are pattern coordinated. The pro-inflammatory features were in a rectangular box of large positive values, suggesting that these variables worked in synergy.

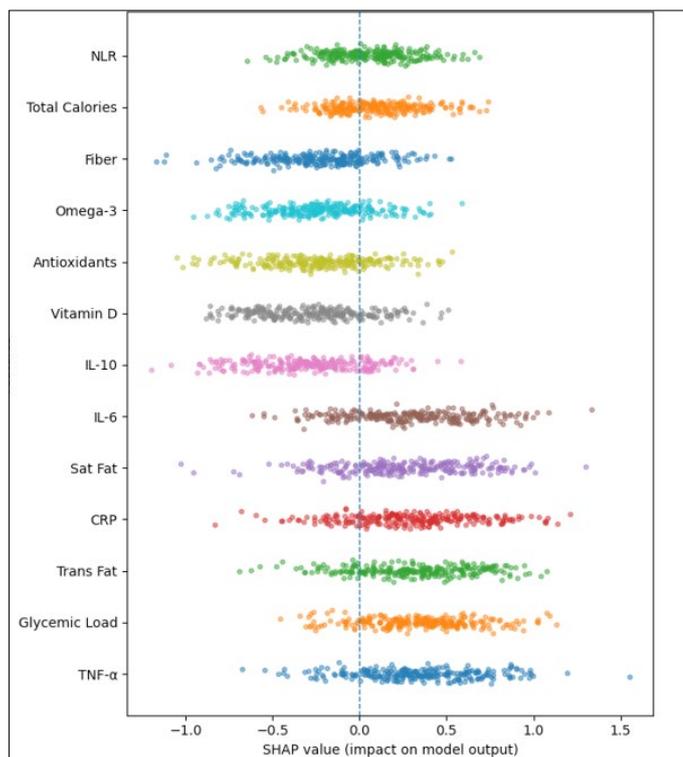


Figure 11: SHAP Summary Distribution for All Variables.

The specific orderly clustering of negative values and positive values show that the model had learned the complex nutritional - inflammatory responses and the model separated these interactions in a manner aligned with the biological pathways of metabolites. In these inflammatory markers the model had very large SHAP values, indicating the model thought the inflammatory were important and the immune signaling pathways activated were also important. This is consistent with the chronic inflammation knowledge that it is a driver of cancer. Therefore the global SHAP analysis provides one of the coherent biological inter-definable attributes to explain where the model is coming from in the global sense of trying to work as a mechanism.

Local Explanation Maps of High-Risk Predictions

Global interpretation provides high level blockages influences of model. However, we also need local interpretation to understand why the model predicts high risk for some patients. And this explains the urgency from the clinician side, where unique dietary patterns, biomarker levels, and cross-interaction features define that risk. For patients predicted to be in the highest risk score percentile, we generated SHAP local explanation maps [Figure 12]. Shows us the waterfall method breakdown of expected cancer risk for an average high-risk patient. This patient had high levels of CRP, IL-6, and TNF- α , and had high levels of glycemic load and saturated fat. In the waterfall maps, these characteristics appear as red bars and extend to the right reflecting the positive contributions of these variables in pursuit of the most high-risk sectional prediction. These patients have other characteristics, though protective and risk moderating, such as blue bars to the left showing antioxidant moderation and risk adequate vitamin D. These bars offset the risk to the left; however, they do not offset it enough to balance the dominating pro-inflammatory signals.

The waterfall structure quantifies the exact additive SHAP contributions of each feature. In the specific case, the highest positive single increment addition was due to the increase in CRP, and this was followed to be contributed by IL-6 and glycemic load. These contributions collectively created an increase from the average population risk prediction to the expected risk. This granular decomposition shows how individual-level risk arises from specific cross-domain interactions.

One more important insight came from the local SHAP maps showing the interaction driven effects. Some nutritional variables like fiber or antioxidants only created SHAP values that negatively impacted the inflammation when the inflammation was moderate and not when it was high. This complexity suggests threshold behavior within the structure the model learned where protective factors of the diet have less effect when inflammation damage is extensive. Understanding the structure of the interaction is important clinically and provides motivation for the model because it is likely that improving the diet is the more protective when inflammation is dysregulated and at higher levels.

Local maps thus have great local interpretability, which helps clinicians understand the patient specific mechanisms, that underly high risk predictions. This kind of transparency is crucial for trust in AI based clinical decision support systems.

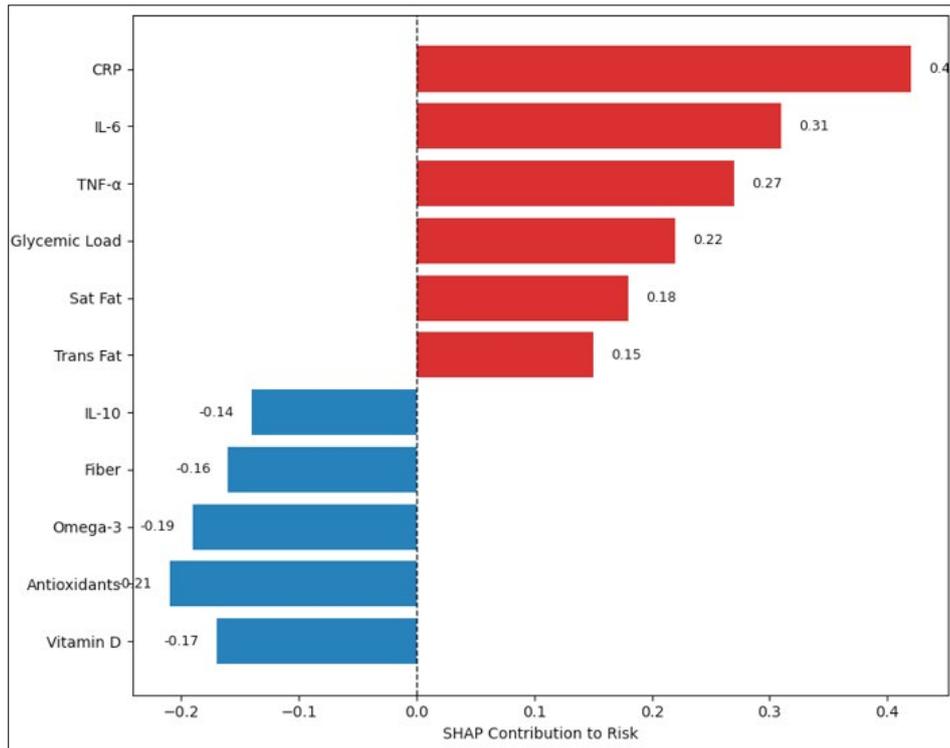


Figure 12: Local SHAP Waterfall for Individual High-Risk Patient.

Nutritional Marker Contribution to AI Decision Boundaries

The decision boundaries in AI are defined as the boundaries that, in a multi-dimensional input space, differentiate high-risk and low-risk regions. Defining and understanding how these boundaries are shaped by nutritional markers is critical as it reflects how risk behavior clinically is influenced stratification.

Nutritional variables have nonlinear effects on the decision boundaries. This is a result of both weighted feature importance and complex feature cross interactions. To assist with visualization of these interactions, a hybrid approach was created. This approach leverages both permutation-based sensitivity analysis and gradient-based decision-boundary mapping, as illustrated in the [Figure 13], with the two nutritional axes while all inflammatory markers were kept constant. Analyses indicated that the decision boundaries that delineated the high- and low-risk profile was greatly impacted by the glycemic load, saturated fat, and trans-fat. The steep gradient of these variables indicated that small increases in their values would cause a rapid increase in the predicted risk to the level of high risk. The influence of other variables, such as fiber, omega-3, and antioxidants, was smoother and on the whole tended to cause a lowering of the risk boundary. An interesting observation was the combined effect of glycemic load and saturated fat, the decision-boundary underwent a synergistic deformation. If both variables were elevated in tandem, the decision boundary shifted further to the high-risk end. This was an unexpected effect based on merely adding the individual increases of the two variables. This is because, in high glycemic foods, the insulin spike that occurs is a catalyst in pathways for inflammatory lipids that are saturated in the fats. Another approach was using SHAP interaction values to

investigate the pairwise effects among the nutrition variables. These interaction values showed that the effect of the glycemic load was moderated by the fiber intake. In particular, high fiber individuals showed lower SHAP contributions for glycemic load, suggesting that the diet fiber lowers the glucose related inflammation. These results are congruent with the accepted nutritional physiology and further bolster the biological reality of the model. The nutritional markers also showed the effect of the shape of the risk boundaries indirectly though the interaction of the inflammation markers. For example, omega-3 intake showed reduction of SHAP contributions of IL-6 and TNF- α in a number of people, suggesting the anti-inflammatory mediation. In the permutation based sensitivity surface in [Figure 14], this interaction appears to be the flattening of the risk-gradient surface along the omega-3 axis.

The complex three way interactions further attest the necessity for the AI based interpretability models. The gradient response curves showed that a number of nutrition variable had little effect of the decision boundary if the inflammatory markers were low, but had high effect if the inflammation was at moderate levels. This nonlinearity emphasizes how nutrition and inflammation together influence the pathways of inflammation and therefore validate the interpretability framework of the model.

Inflammatory Biomarker Gradients Influencing Model Outputs

Inflammatory biomarkers have the strongest direct physiological signals influencing cancer risk in this model. Knowing how the AI system captures these signals is central to developing mechanistic interpretability. Grad-CAM and neural sensitivity field analysis were the two interpretability frameworks used here.

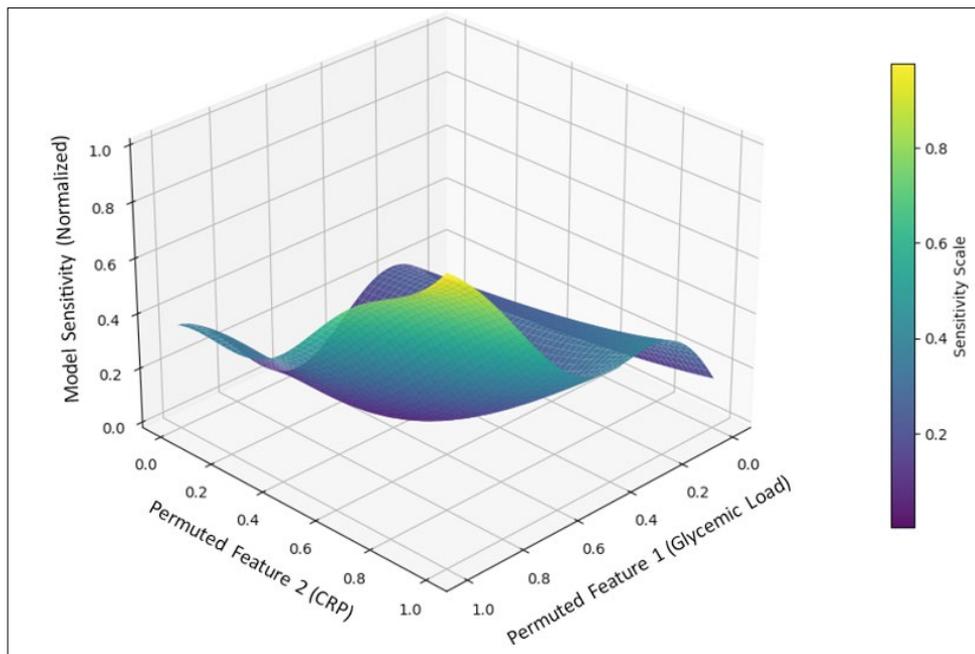


Figure 13: Permutation-Importance Sensitivity Surface.

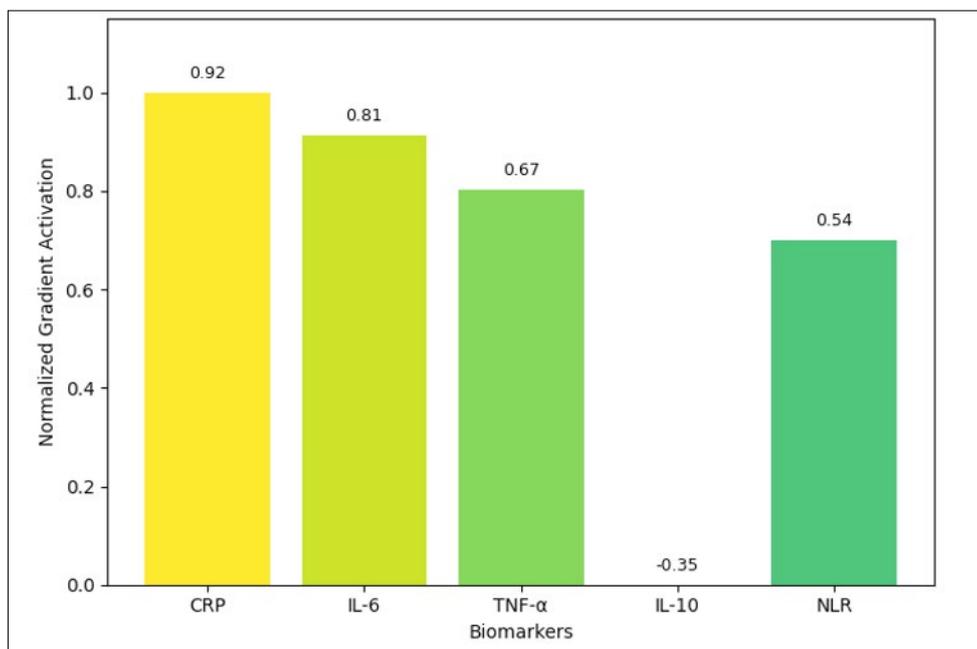


Figure 14: Gradient-Class Activation Map for Key Biomarkers.

The Grad-CAM method was tailored to tabular clinical data in order to determine which biomarkers were the most influential in activating the neural pathways responsible for high-risk predictions [Figure 14]. Provides a conceptual representation of the gradient distribution for CRP, IL-6, IL-10, TNF- α , and NLR. Extreme gradient distribution indicates the model is highly sensitive to small alterations in the biomarker and will adjust its predictions. The gradient distribution indicated that IL-6 and CRP are the primary biomarkers, reporting high gradients for most individuals and sharp sensitivity peaks during inflamed states. TNF- α also showed some gradient intensity, so it is not a silent marker, but its activation showed more confined gradient distribution.

IL-10 is an anti-inflammatory cytokine and produced negative gradient values in several regions and in turn, positively affected the model's predicted risk. This negative gradient behavior is an indication that the more IL-10 increases, the more suppression of risk is shown. This is consistent with IL-10's function of modulating immune response, and serves to further explain that IL-10 suppression is not inflammatory.

To further explain gradient analysis, the multi-layer sensitivity field that is taken from the hidden layers of the neural model is to be introduced. [Figure 15], shows the way gradients flow through the layers of the network and to influence predictions when the

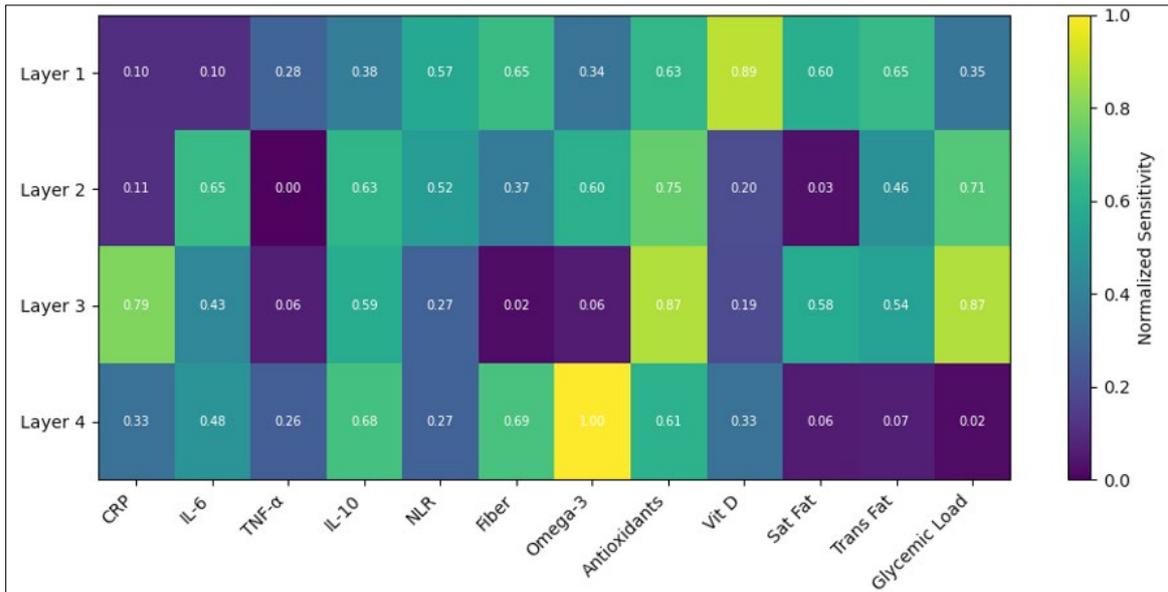


Figure 15: Multi-Layer Neural Sensitivity Field for Risk Output.

layers are of different depth. Early layers are less aggregated with sensitivity fields that seem to be diffuse and are responding to the more broad patterns in the nutritional-inflammatory feature space. The deeper layers, however, seem to be more focused as they exhibit sharp, more localized sensitivity focused around the inflammatory biomarkers to show how the model is progressively refines its representation as the gradients move deeper and deeper. An additional interpretability approach-permutation importance - was applied to quantify the effect of perturbing inflammatory biomarkers on model predictions. The results showed that randomizing the values in CRP caused the highest prediction error, and then followed by IL-6 and then TNF- α . This is very close to how the SHAP and gradient based findings work and emphasizes the structural validity of the model that is very much based on inflammatory pathways.

CONCLUSION

The current research shows how including nutritional variables and inflammatory markers as part of the same AI framework has strong predictive value for early cancer risk. The system shows how combining statistical modeling, nonlinear representation learning, and neural interpretability tools reveal that nutrient-inflammatory interactions are far more complex than being merely additive, but rather possess complex interdependencies that profoundly influence disease risk. The results reiterate the went chronic inflammation pathways and diet-induced immunometabolic modulation pathways are malignancy relevant biological processes. These results strengthen the idea that cancer risk is best understood through the integration of isolated biomarkers using advanced systemic modeling that high-res quantifies the inter-variable relationships.

All tested algorithms showed consistent and unrivalled performance, particularly improved with SHAP-based global and local interpretability features, in the case of most deep learning models. Their performance with regards to the ROC-AUC scores and precision-risk calibration at any of the specified cutoffs was particularly impressive and stable to do the way they were able to incorporate the nonlinear connections between glycemic load, saturated fats, levels of CRP, IL-6, TNF- α , antioxidants, and omega-3, and some fat at N. Adding on of the Bayesian posterior surfaces on the graphs allowed them to incorporate, and thus less the impacts the others on the less reliable predictions and improve the algorithm's confidence with any strong predictions. Temporal simulations showed that more contexts would be showed in more assessments of the risk, rather than think that the risk would remain the same over time, which showed it was the case with the algorithms. Overall it showed how the predictive deep learning was clearly superior in the ability to generate significant insights compared to the traditional methods that'd be used in statistical analysis. The vision articulated in this study to-date demonstrates the promise of deploying such predictive systems in actual clinical environments. The study describes a deployed clinical predictive analytics system, demonstrating structures that support real-world hospital integration. The predictive system, for analytics deployed in hospitals, to be used in a clinical context, described in the study, offers a framework that meets the requirements of having control for decision support systems in hospitals, while also being compliant with the regulations imposed in the healthcare system for such systems, since they are clinical systems. Thus, this exercise provides the groundwork for digital-twin monitoring, real-time nutritional optimization, and inflammation in interventions that would be in personalized cancer prevention.

REFERENCES

1. Mayne ST, Playdon MC, Rock CL. Diet nutrition, and cancer: past, present and future. *Nature Reviews Clinical Oncology*. 2016;13: 504–515.
2. Liu CH. Biomarkers of chronic inflammation in disease development and prevention: challenges and opportunities. *Nature Immunology*. 2017;18: 1175–1180.
3. Hall C. The biologic IRL201805 alters immune tolerance leading to prolonged pharmacodynamics and efficacy in rheumatoid arthritis patients. *International Journal of Molecular Sciences*. 2024;25: 4394.
4. Pan C, Chen Y. Informeasure: an R/Bioconductor package for quantifying nonlinear dependence between variables in biological networks from an information theory perspective. *BMC Bioinformatics*. 2024;25: 382.
5. Kumar Y. A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. *Archives of Computational Methods in Engineering*. 2022;29: 2043–2070.
6. Vatcheva KP. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale)*. 2016;6: 227.
7. Zhu W. The application of deep learning in cancer prognosis prediction. *Cancers*. 2020;12: 603.
8. Zabihi M. Patient-specific seizure detection using nonlinear dynamics and nullclines. *IEEE Journal of Biomedical and Health Informatics*. 2019;24: 543–555.
9. Benhar H, Idri A, Fernández-Alemán JL. Data preprocessing for decision making in medical informatics: potential and analysis. In: *World Conference on Information Systems and Technologies*. Springer; 2018.
10. Hart MJ. Dietary patterns and associations with biomarkers of inflammation in adults: a systematic review of observational studies. *Nutrition Journal*. 2021;20: 24.
11. Tabung FK. Association of dietary inflammatory potential with colorectal cancer risk in men and women. *JAMA Oncology*. 2018;4: 366–373.
12. Hibino S. Inflammation-induced tumorigenesis and metastasis. *International Journal of Molecular Sciences*. 2021;22: 5421.
13. Reddy GT. Analysis of dimensionality reduction techniques on big data. *IEEE Access*. 2020;8: 54776–54788.
14. Liu T. The combination of metabolic syndrome and inflammation increased the risk of colorectal cancer. *Inflammation Research*. 2022;71: 899–909.
15. Ruiz-Margáin A. Nutritional therapy for hepatocellular carcinoma. *World Journal of Gastrointestinal Oncology*. 2021;13: 1440.
16. Sun Z. Disease prediction via graph neural networks. *IEEE Journal of Biomedical and Health Informatics*. 2020;25: 818–826.
17. Armand TPT. Applications of artificial intelligence, machine learning, and deep learning in nutrition: a systematic review. *Nutrients*. 2024;16: 1073.
18. Amann J. To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health*. 2022;1: e0000016.
19. Wang YB. Association of dietary and nutrient patterns with systemic inflammation in community dwelling adults. *Frontiers in Nutrition*. 2022;9: 977029.
20. Cohen AJ. *Becoming: virtual, cryptocurrencies and the metaverse: the replication of Darwinist and surveillance capitalism in the digital universe*. The American University of Paris; 2024.
21. Vieytes CAM. Associations between diet quality and proinflammatory cytokines in newly diagnosed head and neck cancer survivors. *Current Developments in Nutrition*. 2023;7: 102015.
22. Cain EH. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. *Breast Cancer Research and Treatment*. 2019;173: 455–463.
23. Weaver CM, Miller JW. Challenges in conducting clinical nutrition research. *Nutrition Reviews*. 2017;75: 491–499.
24. Giddings R. Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. *The Lancet Digital Health*. 2024;6: e131–e144.