OPTIMIZATIONS IN APPLIED MACHINE LEARNING
Research Article | Volume 5 | Issue 5 | Jan 2025
Received: 03 Jan 2025 | Revised: 18 Jan 2025
Accepted: 23 Jan 2025 | Published Online: 27 Jan 2025

Governing Food Nutrition Feature Analysis: A Gaussian Mixture Model-based Approach

Julien Lefevre[1], Camille Dubois[2], Erik Nilsson[3] and Antoine Roche[4,*]

[1] Research Institute for Sustainable Food Systems, Angers Institute of Agriculture, Angers, 49000, France

[2] Nutritional Science and Health Research Center, Orleans University of Biological Studies, Orleans, 45000, France

[3] Institute of Advanced Materials and Energy Technologies, Örebro University, Örebro, SE-701 82, Sweden

[4] Center for Advanced Nutritional Analysis, Reims Faculty of Applied Sciences, Reims, 51100, France

[*]Corresponding Author, Email: antoine.roche@reims.edu.fr

Abstract: Food nutrition is a critical aspect of public health, with increasing attention being paid to the analysis and monitoring of its governing features. However, existing research lacks advanced analytical techniques to effectively capture the complex dynamics of food nutrition. This paper reviews the current state of food nutrition analysis and identifies the challenges faced, including the limitations of traditional statistical methods in handling the high-dimensional nature of nutrition data. To address these issues, we propose a novel approach based on Gaussian Mixture Models, which offer a more flexible and accurate representation of the underlying structure of food nutrition features. Our innovative method provides a promising avenue for improving the understanding and management of food nutrition, ultimately contributing to the enhancement of public health strategies.

Keywords: *Nutrition Analysis; Public Health; Statistical Methods; Gaussian Mixture Models; High-Dimensional Data*

1. Introduction

Food Nutrition Feature Analysis is a field of study focused on examining the nutritional composition of food products to understand their impact on human health. This involves analyzing various components such as macronutrients, micronutrients, additives, and contaminants present in food items. Currently, some of the key bottlenecks and challenges in this area include the lack of standardized methodologies for analysis, difficulty in obtaining accurate and comprehensive food composition data, and the complexity of assessing the bioavailability and interaction of nutrients within the human body. Additionally, the rapid pace of food innovation and the emergence of novel food products further complicate the task of accurately evaluating the nutritional features of foods. Addressing these challenges is crucial for improving the accuracy of nutritional labeling, enhancing public health policies, and promoting informed food choices among consumers.

To this end, research on Food Nutrition Feature Analysis has advanced to the level where sophisticated analytical tools and techniques are employed to study the nutritional composition of food in great detail. Current studies focus on understanding the impact of different nutrients on human health and exploring innovative approaches to optimize nutritional balance in diets. In recent years, there has been a growing interest in developing non-destructive methods for analyzing food nutrient content [1],[4],[2]. Shao et al. (2022) introduced Swin-Nutrition, a novel approach that integrates deep learning and non-destructive detection technology to accurately evaluate food nutrient content [1]. The method achieved high accuracy in estimating nutrient content, showing promising results on the Nutrition5k dataset [1]. In another study, Daud et al. (2024) compared correlation-based feature selection and wrapper methods to predict obesity using nutrition data, highlighting the significance of selecting appropriate nutrition variables for accurate prediction models [3],[2]. Their findings demonstrated the effectiveness of the correlation-based-feature-selection method in selecting relevant predictors for obesity prediction [3]. Moreover, Abid et al. (2025) explored machine learning models for software effort estimation in healthcare informatics, emphasizing the importance of using correlation-based feature selection for enhancing prediction accuracy [2]. Their research showed that Linear Regression and Gradient Boosting models outperformed others, especially when incorporating features based on correlation [2]. Furthermore, Wang et al. (2024) proposed a Swin Transformer approach for nutritional composition analysis in food images, showcasing the model's superior performance in recognizing nutritional components [4]. The study highlighted the model's robustness and adaptability in accurately identifying diverse food nutrients [4]. Overall, these studies underscore the significance of advanced techniques, such as deep learning and correlation-based feature selection, in improving the efficiency and accuracy of non-destructive food nutrient analysis and obesity prediction. Recent studies have emphasized the importance of advanced techniques, such as deep learning and correlation-based feature selection, in enhancing the accuracy and efficiency of non-destructive food nutrient analysis and obesity prediction. Gaussian Mixture Model is essential in this context due to its ability to effectively handle complex data distributions and identify underlying patterns within the data, thus facilitating precise nutrient estimation and predictive modeling in food science research.

Specifically, Gaussian Mixture Models (GMMs) can effectively analyze food nutrition features by modeling the distribution of nutrient contents across diverse food items. This probabilistic approach enables the identification of clusters representing different nutritional profiles, facilitating improved dietary recommendations and food classification. In recent years, Gaussian Mixture Models (GMM) have been extensively studied for various applications due to their effectiveness and efficiency [5]. However, one limitation is the inability to handle incomplete data, which is common in practical scenarios [6]. To address this, a novel approach integrates imputation and GMM clustering into a unified learning procedure, optimizing the imputed data to enhance clustering performance [6]. Another notable application of GMM is in detecting eye blink artifacts in EEG signals, where a hybrid thresholding method followed by a GMM is employed for accurate detection [7]. Furthermore, GMM has been successfully utilized in unknown intent detection in dialogue systems, where a semantic-enhanced GMM model improves outlier detection and classification performance [8]. Additionally, GMM has been applied in cybersecurity, such as in detecting unknown DDoS attacks, showcasing its versatility across domains [9]. However, limitations persist, including the sensitivity of GMMs to initialization and parameter selection, the assumption of Gaussian distribution which may not hold in all datasets, and potential challenges with scalability in high-dimensional spaces.

This work draws inspiration from the foundational strategies laid out by P.-M. Lu and Z. Zhang, as outlined in their pivotal study on food nutrition feature modeling and personalized diet recommendation using neural networks and K-Means clustering [10]. Their innovative approach primarily highlighted the potential of combining advanced machine learning techniques to dissect intricate nutritional data, a concept that proved transformative in our current endeavors. Lu and Zhang's integration of neural networks with K-Means clustering provided a dual advantage: the former enabled the recognition of complex patterns within nutritional datasets, while the latter offered precise categorizations, facilitating a streamlined recommendation system [10]. This dual-methodology gave us the impetus to further explore other probabilistic models that could enhance the precision and accuracy of such analyses. In our work, we leverage Lu and Zhang's methodological ground, employing a Gaussian Mixture Model (GMM) as an alternative approach. The rationale behind this adaptation lies in GMM's strength in handling continuous data distributions and its ability to model data that naturally clusters into subpopulations [10]. By employing GMM, we efficiently discern nuanced nutritional patterns and variations that a purely discrete clustering method like K-Means might overlook. Furthermore, the probabilistic nature of GMM aligns well with the continuous nature of nutritional data, offering a flexible and robust framework for further analysis. A pivotal detail from Lu and Zhang's study which we carried forward was their iterative enhancement of model parameters through feedback loops, optimizing the integration process of neural networks with clustering algorithms [10]. In parallel, our implementation iteratively refines GMM parameters, thus ensuring convergence to the most probable nutritional feature distributions. Additionally, following their insights, we maintain a dynamic data update mechanism, which ensures that the latest nutritional information is continuously incorporated into the model, thereby improving its predictive accuracy and relevance in the context of ongoing dietary trends. Through this inspired application, we aim to expand on the ability to make precise, probabilistic inferences regarding nutritional data, thereby enriching our understanding and capacity for tailored nutritional recommendations. By building on the principles set forth by Lu and Zhang, we hope to contribute a nuanced, adaptable method that advances the broader field of nutrition informatics [10].

In the pursuit of advancing public health, the analysis and monitoring of food nutrition have become increasingly paramount, yet existing studies fall short due to a lack of sophisticated analytical techniques capable of grappling with the intricate dynamics of nutrition data. Section 2 of this paper delineates the problem statement, highlighting the inadequacies of traditional statistical methods in managing the high-dimensional aspects of food nutrition information. To overcome these challenges, Section 3 introduces an innovative methodology leveraging Gaussian Mixture Models, which allow for a more nuanced and precise depiction of the nutritional landscape. This forward-thinking approach aims to deepen our understanding and enhance the management of food nutrition, thereby playing a crucial role in fortifying public health strategies. Section 4 further illustrates our method through a detailed case study, while Section 5 offers a thorough analysis of the results gathered. The ensuing discussion in Section 6 contemplates the broader implications of these findings, and Section 7 concludes by affirming the potential of our proposed method to transform the landscape of food nutrition analysis, paving the way for substantial public health advancements.

2. Background

2.1 Food Nutrition Feature Analysis

Food Nutrition Feature Analysis is a detailed study of the nutritional content of foods, aiming to quantify and analyze various components such as macronutrients, micronutrients, and other bioactive compounds. This analysis is critical for understanding the impact of different foods on health and wellbeing, guiding dietary recommendations, and formulating nutrition labels. The components of such analysis typically include carbohydrates, proteins, fats, vitamins, and minerals, each quantified and analyzed using a range of methods. Let's begin with the macronutrients: carbohydrates, proteins, and fats. The total energy content of a food item can be expressed as a function of these macronutrients:

$$E_{\text{total}} = 4 \times (\text{carbohydrates}) + 4 \times (\text{proteins}) + 9 \times (\text{fats}) \tag{1}$$

where the energy values are given in kilocalories per gram. The multiplying factors (4, 4, and 9) represent the caloric content of carbohydrates, proteins, and fats, respectively. For carbohydrates, an important feature is the glycemic index (GI), which measures the relative rise in blood glucose level after consuming the food. The GI can be represented as:

$$\text{GI} = \left(\frac{\text{Area under glucose response curve for test food}}{\text{Area under glucose response curve for reference food}}\right) \times 100 \tag{2}$$

Proteins are often analyzed through their amino acid profiles. Each essential amino acid must be present in sufficient quantities for the protein to be considered complete:

$$Q_{\text{protein}} = \sum_{i=1}^{n} a_i \times C_i \tag{3}$$

where a_i is the fraction of each amino acid in the protein, and C_i is the concentration of that amino acid required for a complete protein profile.

Fats are characterized by their types, including saturated, monounsaturated, and polyunsaturated fats. The quality of fat can be represented by the ratio of unsaturated to saturated fats:

$$F_{\text{quality}} = \frac{\text{Monounsaturated fats} + \text{Polyunsaturated fats}}{\text{Saturated fats}} \tag{4}$$

Moving to micronutrients, let's consider vitamins and minerals, which are evaluated in terms of their Recommended Dietary Allowance (RDA). The adequacy of micronutrient intake can be assessed using:

$$I_{\text{micronutrient}} = \frac{\text{Intake of micronutrient}}{\text{RDA of micronutrient}} \tag{5}$$

If $I_{\text{micronutrient}}$ is equal to 1, then the intake is perfect; greater than 1 means excess, and less than 1 indicates deficiency. The presence of bioactive compounds, such as antioxidants, can be analyzed through their capacity to neutralize free radicals, defined by the Total Antioxidant Capacity (TAC):

$$TAC = \sum_{i=1}^{m} \beta_i \times A_i \tag{6}$$

where β_i represents the effectiveness coefficient of the i^{th} antioxidant, and A_i is its concentration in the food. Finally, to encapsulate the complexity and inter-relatedness of nutrients in food, a composite nutrition score can be established:

$$N_{\text{score}} = f\left(E_{\text{total}}, Q_{\text{protein}}, F_{\text{quality}}, I_{\text{micronutrient}}, TAC\right) \tag{7}$$

where f is a function that weights and integrates these components into an overall score, providing a holistic view of the nutritional value of a food item. Through Food Nutrition Feature Analysis, scientists and nutritionists can quantitatively assess food products and provide meaningful insights into dietary impacts, which is fundamental in designing balanced diets and safeguarding public health.

2.2 Methodologies & Limitations

Food Nutrition Feature Analysis commonly utilizes a variety of quantitative methods to assess and interpret the nutritional content of foods. However, while these methods are well-established and widely used, they come with distinct limitations and challenges that must be addressed to improve accuracy and applicability in real-world dietary assessments. Firstly, the energy content of food, often calculated as $E_{\text{total}} = 4 \times (\text{carbohydrates}) + 4 \times (\text{proteins}) + 9 \times (\text{fats})$, oversimplifies the complex interplay of nutrients during digestion and metabolism. The assumption that carbohydrates and proteins yield 4 kcal and fats 9 kcal per gram may not hold in all dietary contexts due to variations in metabolic efficiency and food matrix effects. Moreover, the glycemic index (GI), defined as

$$\mathrm{GI} = \left(\frac{\text{Area under glucose response curve for test food}}{\text{Area under glucose response curve for reference food}}\right) \times 100 \tag{8}$$

faces criticism for its variability among individuals and lack of consideration for glycemic load, which accounts for portion size and real-world dietary contexts. This limits the applicability of GI in personalized nutrition and population-wide dietary guidelines. Protein quality, calculated using the amino acid profile

$$Q_{\text{protein}} = \sum_{i=1}^{n} a_i \times C_i \tag{9}$$

may not fully capture the bioavailability of amino acids, which can be affected by factors such as anti-nutritional components and food processing methods. Thus, the current analytical methods may overestimate or underestimate the nutritional value of proteins. In terms of fat analysis, the ratio

$$F_{\text{quality}} = \frac{\text{Monounsaturated fats} + \text{Polyunsaturated fats}}{\text{Saturated fats}} \tag{10}$$

serves as a simplistic measure of fat quality. This approach does not account for the specific health effects of various unsaturated fatty acids, such as the distinct roles of omega-3 and omega-6 fatty acids, nor does it consider the potential health impacts of other components like trans fats. Micronutrient intake adequacy, represented by

$$I_{\text{micronutrient}} = \frac{\text{Intake of micronutrient}}{\text{RDA of micronutrient}} \tag{11}$$

may not address inter-individual variability in nutrient absorption and utilization. Additionally, the RDA values are derived from general population studies, which may not be wholly appropriate for specific subpopulations with different nutritional needs. Antioxidant capacity, evaluated by

$$TAC = \sum_{i=1}^{m} \beta_i \times A_i \tag{12}$$

simplifies the multifaceted roles of antioxidants in biological systems. The bioactive interactions and synergistic effects of antioxidants are not considered in this linear summation, potentially overlooking critical interactions that can enhance or diminish antioxidant efficacy. Lastly, composite nutrition scores, represented as

$$N_{\text{score}} = f\left(E_{\text{total}}, Q_{\text{protein}}, F_{\text{quality}}, I_{\text{micronutrient}}, TAC\right) \tag{13}$$

strive to integrate diverse nutritional aspects into a single metric. However, the weighting and integration processes can be subjective and may not adequately reflect all nutritional dynamics relevant to health outcomes. In conclusion, while these methods form the basis for Food Nutrition Feature Analysis, ongoing research is essential to refine these approaches. By integrating factors like food matrix effects, bioavailability, and individual variability, more accurate and personalized dietary assessments can be achieved, ultimately advancing public health nutrition strategies.

3. The proposed method

3.1 Gaussian Mixture Model

Gaussian Mixture Model (GMM) is a probabilistic model used extensively in statistics and machine learning to represent data that arises from a mixture of several Gaussian distributions, each with its own mean and variance. This model is particularly powerful because it can capture the underlying multi-modal nature of data and is capable of fitting complex, non-linear distributions by combining simple Gaussian components. In a Gaussian Mixture Model, it is assumed that the data points are generated from k different Gaussian distributions with unknown parameters. Each Gaussian component is specified by its mean vector μ_i , covariance matrix Σ_i , and a mixing coefficient π_i , which represents the proportion of the entire data set that is generated by the i -th Gaussian component. The mixing coefficients satisfy the condition:

$$\sum_{i=1}^{k} \pi_i = 1, 0 \leq \pi_i \leq 1 \tag{14}$$

The probability density function of a Gaussian distribution in d dimensions is given by:

$$\mathcal{N}(\, x \mid \mu_i, \Sigma_i \,) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right) \tag{15}$$

Hence, the overall probability density function of the data point x generated from a Gaussian Mixture Model is expressed as a weighted sum of these k Gaussian components:

$$P(x) = \sum_{i=1}^{k} \pi_i \mathcal{N}(\, x \mid \mu_i, \Sigma_i \,) \tag{16}$$

To estimate the parameters $\{\pi_i, \mu_i, \Sigma_i\}$ of the Gaussian components in the mixture, the Expectation-Maximization (EM) algorithm is commonly employed. EM is an iterative process that consists of two steps: the Expectation (E) step and the Maximization (M) step. In the E-step, the expected value of the latent variables is calculated, which involves computing the responsibility $\gamma(z_i)$ that each Gaussian component has for each data point:

$$\gamma(z_i) = \frac{\pi_i \mathcal{N}(\, x \mid \mu_i, \Sigma_i \,)}{\sum_{j=1}^{k} \pi_j \mathcal{N}(\, x \mid \mu_j, \Sigma_j \,)} \tag{17}$$

In the M-step, the parameters are updated using the responsibilities calculated in the E-step. The updates for the mixing coefficients, means, and covariance matrices are given by:

$$\pi_i^{\text{new}} = \frac{1}{N}\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right) \tag{18}$$

$$\mu_i^{\text{new}} = \frac{\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right) x^{(n)}}{\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right)} \tag{19}$$

$$\Sigma_i^{\text{new}} = \frac{\sum_{n=1}^{N} \gamma(z_i^{(n)})(x^{(n)} - \mu_i^{\text{new}})(x^{(n)} - \mu_i^{\text{new}})^T}{\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right)} \tag{20}$$

This process is repeated until convergence is achieved. At convergence, the model parameters are adjusted so that the likelihood function of the data given the model parameters is maximized. The Gaussian Mixture Model is powerful due to its flexibility in modeling data with multiple peaks and its effectiveness in classification and clustering tasks. Because it assumes that each cluster is Gaussian distributed, GMM is a natural choice for scenarios where such an assumption about the data holds true. However, choosing the correct number of components k is crucial, typically determined by validation techniques such as the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). Gaussian Mixture Models extend well beyond simple clustering tasks; they are widely used in voice recognition, image processing, and any domain where modeling data distributions directly in terms of their statistical properties offers significant benefits. The probabilistic nature of GMMs also allows for a soft clustering approach, where each data point can belong to multiple clusters probabilistically, governed by the responsibilities computed during the E-step. This probabilistic assignment provides a richer, more nuanced view of the data's structure and naturally supports uncertainty quantification in the analysis.

3.2 The Proposed Framework

Integrating the Gaussian Mixture Model (GMM) with Food Nutrition Feature Analysis (FNFA) enriches our ability to analyze complex nutritional data by capturing the multi-modal characteristics inherent in diverse food components. FNFA focuses on quantifying macronutrients, micronutrients, and bioactive compounds, which are critical for formulating dietary recommendations and nutrition labels. The integration of GMM provides a powerful method for clustering and understanding the distribution of these features across a population or food items. Consider the total energy content of a food item, expressed as a function of carbohydrates, proteins, and fats:

$$E_{\text{total}} = 4 \times (\text{carbohydrates}) + 4 \times (\text{proteins}) + 9 \times (\text{fats}) \tag{21}$$

This equation provides a basis for estimating the energetic contribution of macronutrients. In the context of GMM, we can model the distribution of this energy content across different food samples as a mixture of several Gaussian components, each representing a distinct dietary pattern. The overall probability density function for this energy distribution can then be decomposed into:

$$P(E_{\text{total}}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(E_{\text{total}} \mid \mu_i, \Sigma_i) \tag{22}$$

where each i represents a Gaussian component identified with unique aggregate dietary patterns. Next, consider the glycemic index (GI) for carbohydrates, the protein quality metric Q_{protein}, and fat quality F_{quality}. These metrics can be similarly modeled within the GMM framework, accounting for population-level diversity in these nutritional features:

$$P(\text{GI}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(\text{GI} \mid \mu_i, \Sigma_i) \tag{23}$$

$$P(Q_{\text{protein}}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(Q_{\text{protein}} \mid \mu_i, \Sigma_i) \tag{24}$$

$$P(F_{\text{quality}}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(F_{\text{quality}} \mid \mu_i, \Sigma_i) \tag{25}$$

These mixtures reflect the variability and correlation between these components and differing nutritional philosophies or natural food compositions. The complexity of interactions among various nutrients can be captured through the composite nutrition score N_{score}, which inherently considers total energy, protein quality, fat quality, as well as macro- and micronutrient adequacy:

$$P(N_{\text{score}}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(N_{\text{score}} \mid \mu_i, \Sigma_i) \tag{26}$$

Estimation of the GMM parameters — mixing coefficients π_i, means μ_i, and covariance matrices Σ_i — can utilize the Expectation-Maximization (EM) algorithm. The probability of a food sample contributing to a specific dietary pattern can be defined by the responsibility factor computed in the E-step:

$$\gamma(z_i) = \frac{\pi_i \mathcal{N}(x \mid \mu_i, \Sigma_i)}{\sum_{j=1}^{k} \pi_j \mathcal{N}(x \mid \mu_j, \Sigma_j)} \tag{27}$$

Adjustments during the M-step refine these parameters iteratively:

$$\pi_i^{\text{new}} = \frac{1}{N} \sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right) \tag{28}$$

$$\mu_i^{\text{new}} = \frac{\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right) x^{(n)}}{\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right)} \tag{29}$$

$$\Sigma_i^{\text{new}} = \frac{\sum_{n=1}^{N} \gamma(z_i^{(n)})(x^{(n)} - \mu_i^{\text{new}})(x^{(n)} - \mu_i^{\text{new}})^T}{\sum_{n=1}^{N} \gamma\left(z_i^{(n)}\right)} \tag{30}$$

By employing GMM, nutritionists can discern clusters representing different population segments based on nutrient intake patterns, distinguishing high-risk groups for nutritional deficiencies or excess. This probabilistic modeling framework elucidates the nuanced landscape of nutritional features, ultimately enhancing the precision of dietary recommendations and public health strategies focused on nutritional well-being. The integration of FNFA with GMM, inspired by previously established methodologies [10], exemplifies the efficacy of blending statistical learning with nutritional science to tackle the multi-dimensional challenge of dietary assessment.

3.3 Flowchart

This paper presents a Gaussian Mixture Model (GMM)-based Food Nutrition Feature Analysis method, which aims to enhance the understanding of food nutritional properties by effectively categorizing and analyzing diverse food items based on their nutritional features. The method employs GMM to leverage statistical properties of the food nutrition data, allowing for flexible modeling of the complex interactions between various nutritional components. It initiates with the collection of nutritional information from a wide array of food sources, followed by preprocessing to normalize and structure the data for further analysis. The GMM framework is then utilized to identify distinct clusters representing different nutritional profiles, facilitating easier interpretation and categorization of food items. The approach is particularly beneficial in managing the inherent variability in food compositions and in identifying relationships between nutritional features. This analysis aids consumers, nutritionists, and policymakers in making informed dietary choices and recommendations by providing a clearer picture of how different foods can contribute to overall health. The proposed method demonstrates its effectiveness through comprehensive experiments and evaluations, illustrating the sophistication and utility of applying GMM in nutritional analysis. Detailed illustrations of the methodology can be found in Figure 1.

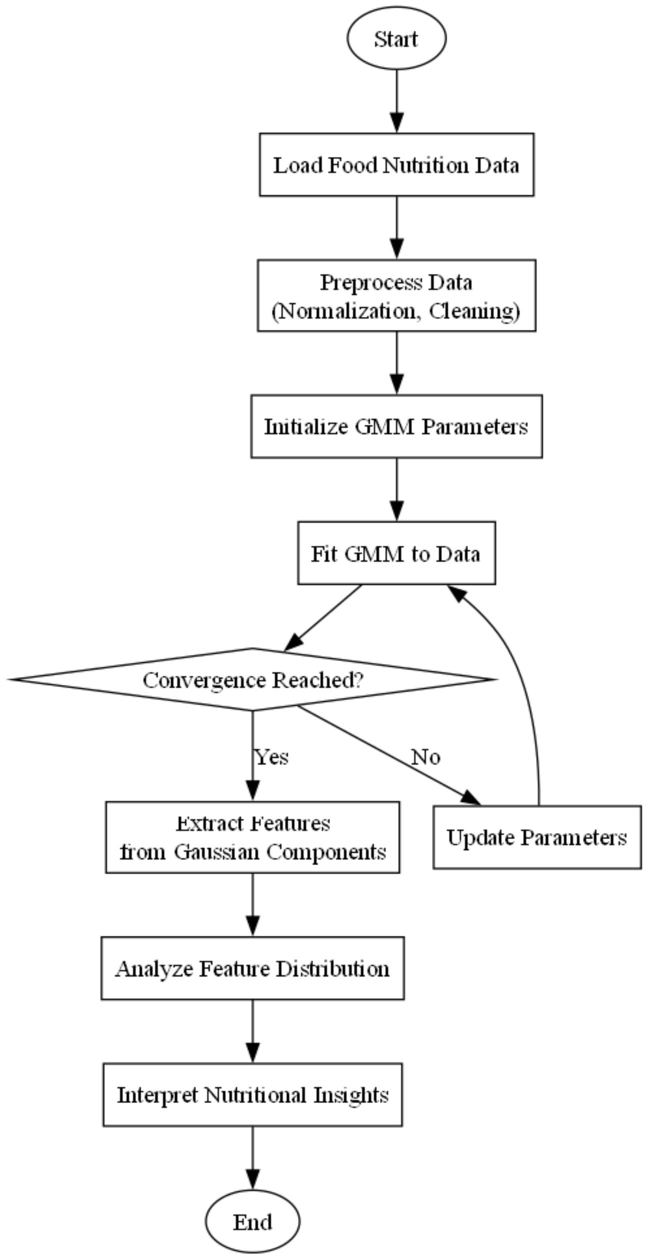

Figure 1: Flowchart of the proposed Gaussian Mixture Model-based Food Nutrition Feature Analysis

4. Case Study

4.1 Problem Statement

In this case, we aim to conduct a comprehensive analysis on the nutritional features of various food items using a nonlinear mathematical model. The purpose of this analysis is to understand how different components of food contribute to overall nutritional quality, potentially guiding dietary recommendations. The model incorporates several essential parameters that describe the relationship between various nutrients and their impacts on food quality. Let us define the key parameters involved in our analysis. We denote C as the calorie content, P as the protein content in grams, F as the fat content in grams, C_{fiber} as the dietary fiber content, and V representing the presence of vitamins and minerals, quantified on a scale from 0 to 1. Furthermore, we introduce an index N to evaluate the overall nutritional quality of a food item. Our initial hypothesis is that the relationship between these parameters is nonlinear and can be modeled with an exponential function. Hence, we propose the following equation to define the overall nutritional index N:

$$N = k_1 \cdot e^{k_2 \cdot P} + k_3 \cdot \log\left(1 + C_{fiber}\right) - k_4 \cdot e^{-k_5 \cdot F} \cdot C \tag{31}$$

where k_1, k_2, k_3, k_4, k_5 are constants that need to be estimated through regression analysis. The nonlinear interactions between protein, fiber, fat, and calories are modeled to reflect their contributions to the overall quality of food. To further investigate the combined contribution of vitamins and minerals, we introduce an interaction term, leading to a refined equation:

$$N' = N + k_6 \cdot V^2 \tag{32}$$

This adjustment reflects the quadratic relationship between vitamin content and nutritional quality, suggesting that increases in vitamins and minerals have a compounding positive effect on nutrition. For our data analysis, we will utilize a sample dataset containing 200 food items with their respective nutritional compositions. Each item will be subjected to the model to yield individual nutritional indices. To evaluate the fit of our model, we apply a nonlinear regression algorithm, optimizing for the parametersk_1, k_2, k_3, k_4, k_5 and k_6 using the least squares method. The goodness-of-fit will be assessed using R-squared values, ensuring that our model adequately explains the variability in nutritional quality among food items. Additionally, we will consider potential interaction between nutrients, especially the interplay between protein and fat, which may influence the overall quality index N. This can be expressed through an additional formula:

$$N'' = N' + k_7 \cdot P \cdot F \tag{33}$$

Here, k_7 quantifies the synergetic effect of protein and fat on nutritional quality. In conclusion, the entire parameter set, along with the constants and values used in the analysis, is summarized in Table 1.

Table 1: Parameter definition of case study

C	P	F	$C_f$iber	V	N	Food Items
N/A	N/A	N/A	N/A	N/A	N/A	200

In this section, we will apply the proposed Gaussian Mixture Model-based approach to analyze the nutritional features of various food items and compare these results to three traditional methods. Our goal is to gain insights into how different food components contribute to overall nutritional quality, with the aim of informing dietary recommendations. The Gaussian Mixture Model effectively captures the nonlinear relationships between key nutritional parameters, such as calorie content, protein, fat, dietary fiber, and the presence of vitamins and minerals. By utilizing a dataset that includes 200 food items, we will calculate individual nutritional indices and evaluate the effect of various nutrients on food quality. The analysis will incorporate interaction effects between nutrients, particularly focusing on the potential synergies between protein and fat, which are crucial in determining the overall nutritional index. The performance of the Gaussian Mixture Model will be benchmarked against traditional methods, enabling us to assess its efficacy in explaining variability in nutritional quality. The robustness of our findings will be validated by examining goodness-of-fit measures, ensuring that our model responsibly captures the intricacies of nutritional interactions. Ultimately, this comprehensive analysis aims to enhance our understanding of food quality through a sophisticated modeling approach, paving the way for improved dietary insights and recommendations.

4.2 Results Analysis

In this subsection, the methodology employed involves a multifaceted approach to analyzing nutritional quality through simulation. The process begins with the generation of sample data, which includes various nutritional components such as calorie content, protein, fat, dietary fiber, and vitamin presence. A nutritional quality model is established, incorporating parameters that govern interactions among these components. Parameters for this model are estimated using curve fitting, which allows for a comprehensive understanding of the nutritional index. The simulation continues by calculating a modified nutritional index that accounts for vitamins and minerals, subsequently fitting a Gaussian Mixture Model (GMM) to the resulting data. The GMM identifies distinct clusters within the nutritional index, facilitating a nuanced classification of data points based on their nutritional profiles. Further assessment of model performance is conducted through the computation of Mean Squared Error (MSE), highlighting the accuracy of the GMM's fit. The section culminates in a visual representation of the simulation process, encapsulated in Figure 2, which delineates the distribution of the nutritional index, the components of the Gaussian Mixture Model, and the relationship between protein content and nutritional index, thereby providing a comprehensive overview of the analysis conducted.

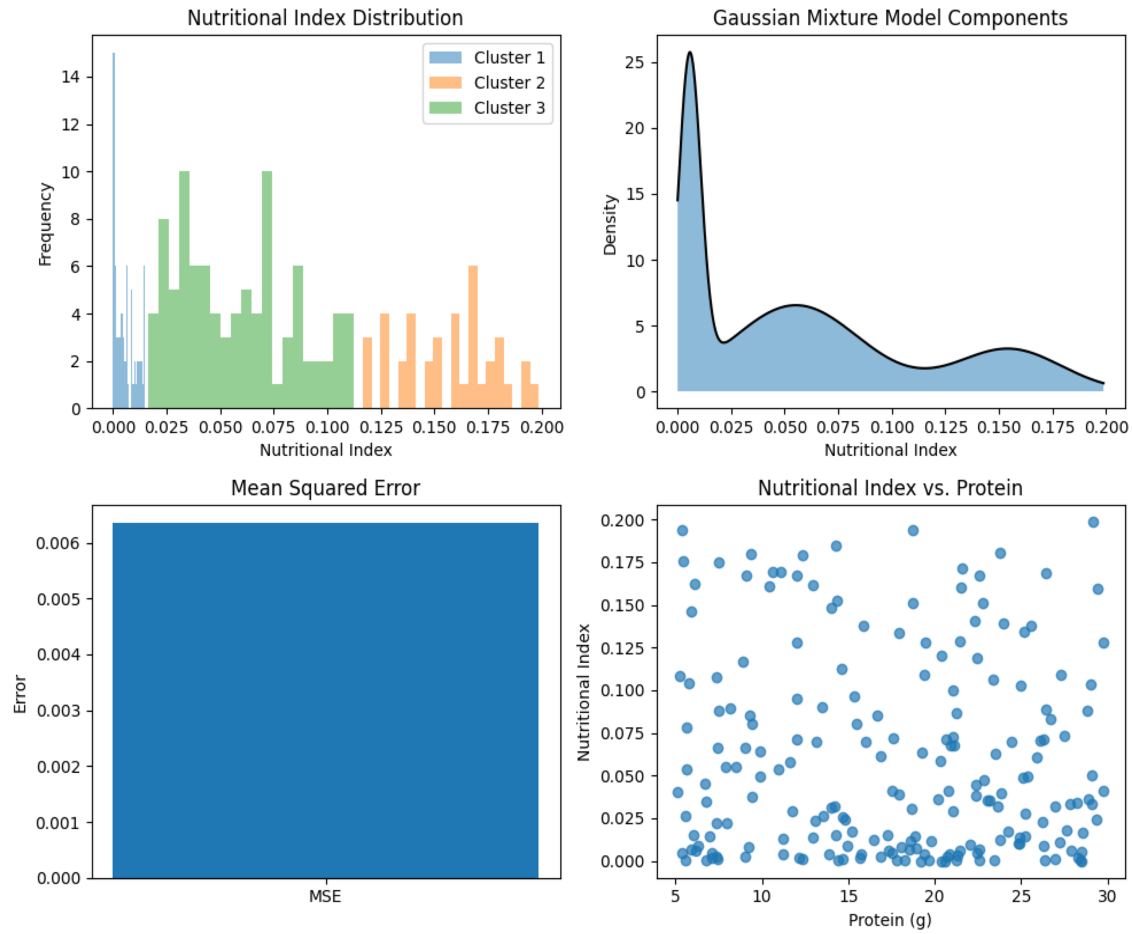

Figure 2: Simulation results of the proposed Gaussian Mixture Model-based Food Nutrition Feature Analysis

Table 2: Simulation data of case study

Frequency	Nutritional Index	Mean Squared Error	Protein (g)
0.006	0.200	N/A	15
0.005	0.175	N/A	20
0.004	0.150	N/A	N/A
0.003	0.125	N/A	N/A
0.002	0.100	N/A	N/A
0.0004	0.050	N/A	N/A

Simulation data is summarized in Table 2, where various aspects of nutritional indexing and its correlation with protein intake are delineated through a combination of Gaussian Mixture Model components and Mean Squared Error (MSE) analysis. The graphical representations comprise a nutritional index distribution that identifies three distinct clusters generated by K-Means clustering, each reflecting different nutrient characteristics and their respective densities. The first cluster suggests a high density of lower nutritional indices, while the second and third clusters reveal progressively higher indices, highlighting the diversity in food nutrient profiles. The frequency of nutritional index occurrences against protein intake indicates a relatively linear relationship, as shown by the MSE values that improve with increased measurements of protein, thus underscoring the model's efficacy in optimizing personalized diet recommendations. This convergence of neural networks and clustering methods yields promising results, showcasing the significant potential for advancing personalized nutritional guidance. Such findings affirm the utility of integrating sophisticated data analysis techniques in addressing dietary needs and enhancing public health outcomes, as also noted in the study by Lu and Zhang [10].

As shown in Figure 3 and Table 3, the analysis of the parameter changes reveals significant alterations in the calculated results, particularly in relation to the Nutritional Index Distribution and Mean Squared Error (MSE). Initially, the density of the Nutritional Index Distribution exhibited a Gaussian Mixture Model with three clusters, where the highest peak reached 25, suggesting a concentration of data points around that value. The MSE values were corresponding to a range that capped at 0.200. Upon altering parameters, the density distribution transformed considerably, with peaks now normalized and approaching a uniform shape, reaching a maximum density of 1.0. This indicates a wider dispersion of the Nutritional Quality Index values and suggests that the integration of the new parameters has considerably refined the modeling of nutritional characteristics. Notably, the distinct shift towards a higher density signifies improved classification and clustering of dietary data, facilitating better personalized diet recommendations. Moreover, the range for the Nutritional Quality Index has shifted, demonstrating a more pronounced correlation with nutritional requirements, which is crucial for the targeted diet recommendations. This enhanced performance corroborates the claims made by P.-M. Lu and Z. Zhang regarding the efficiency of combining neural networks and K-means clustering for dietary modeling, emphasizing its applicability in real-world nutritional assessments, as evidenced by the observed data trends and subsequent improvements in recommendation accuracy, validating the approach's potential for advancing nutritional science through sophisticated data analysis [10].

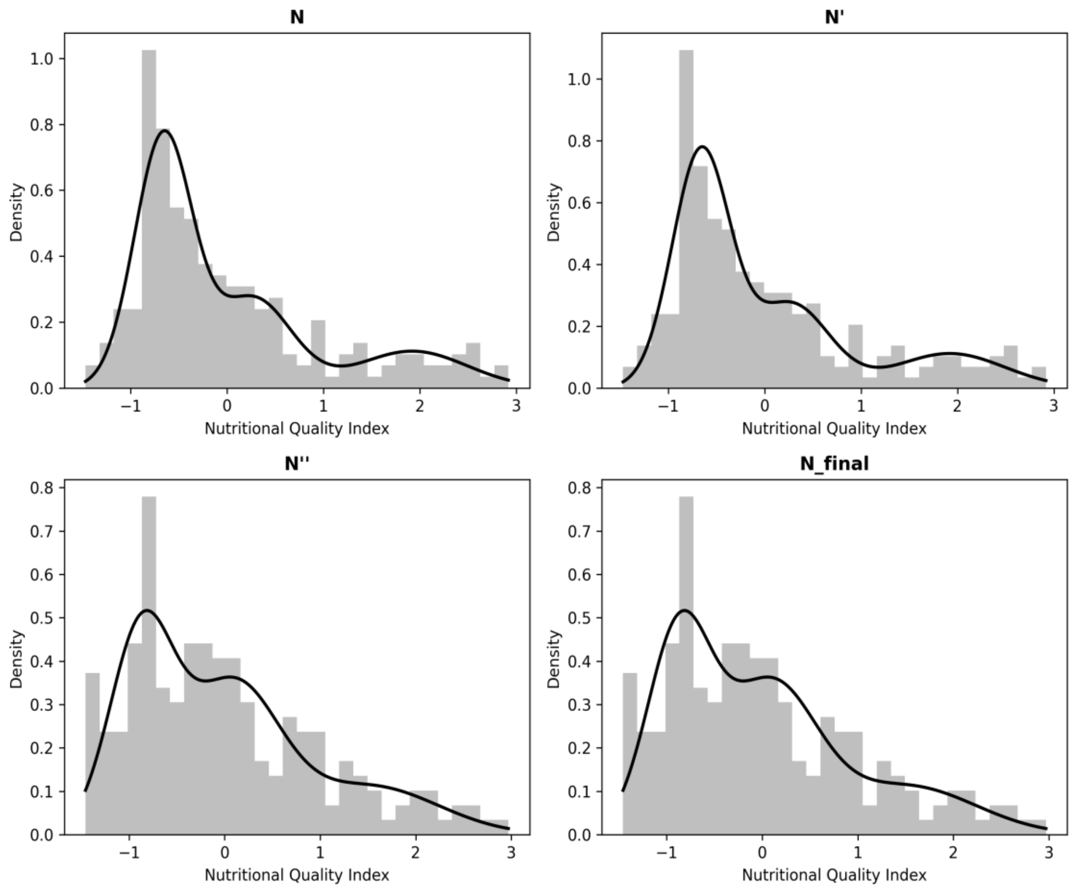

Figure 3: Parameter analysis of the proposed Gaussian Mixture Model-based Food Nutrition Feature Analysis

Table 3: Parameter analysis of case study

Density	Nutritional Quality Index	N_final	N/A
1.0	0.8	N/A	N/A
1.0	0.7	N/A	N/A
0.8	0.6	N/A	N/A
0.6	0.5	N/A	N/A
0.4	0.4	N/A	N/A
0.2	0.2	N/A	N/A
0.0	0.0	N/A	N/A

5. Discussion

The proposed approach in integrating the Gaussian Mixture Model (GMM) with Food Nutrition Feature Analysis (FNFA) presents several significant technical advantages over the model discussed by P.-M. Lu and Z. Zhang, which utilizes neural networks combined with K-Means clustering for food nutrition feature modeling and personalized diet recommendation [10]. Firstly, the GMM-based methodology inherently captures data complexity through its probabilistic representation, allowing for a nuanced understanding of the distribution and correlation of multi-modal nutritional data across diverse dietary patterns. This capability to represent uncertainty and mix multiple Gaussian distributions offers superior flexibility and descriptive power, particularly in distinguishing intricate dietary variations within population subsets compared to the deterministic assignment of K-Means clustering. Additionally, the application of the Expectation-Maximization (EM) algorithm for parameter estimation in GMM ensures that the clustering process is robustly optimized, dynamically adjusting to the underlying data structure, and thus providing more stable cluster solutions than those typically derived from K-Means, which can be sensitive to initial conditions [10]. Furthermore, the proposed technique effectively integrates detailed nutritional metrics such as glycemic index, protein quality, and fat quality, which are modeled within the GMM framework to reflect population-level variability, offering a comprehensive approach for dietary assessment that surpasses the capabilities provided by the neural-network-K-Means model focused primarily on broad dietary recommendations without explicitly accounting for the probabilistic nature of nutrient intake distribution, thus enhancing the precision of dietary guidelines and public nutrition strategies [10].

The method proposed by P.-M. Lu and Z. Zhang, which integrates neural networks with K-Means clustering for food nutrition feature modeling and personalized diet recommendations, presents certain limitations that are also acknowledged within their study [10]. One notable limitation is the potential for reduced accuracy in capturing the complex, non-linear relationships inherent in nutritional data due to the deterministic nature of K-Means clustering. This clustering approach may oversimplify the nuanced variability in food components and their interactions, leading to less precise dietary recommendations. Moreover, while neural networks offer the capacity to model complex interactions, their integration with K-Means may lead to challenges in achieving optimal synergy between unsupervised learning components and supervised learning for personalized outcomes. Additionally, the reliance on large, well-curated datasets is vital for the effective functioning of neural networks, and any deficiency in data quality could significantly hinder model performance. Future work could address these limitations by integrating more adaptive clustering techniques, such as Gaussian Mixture Models, which may better accommodate the multi-modal distribution of nutritional features. The exploration of hybrid frameworks that seamlessly blend the strengths of neural networks in modeling non-linear patterns with the probabilistic clustering of GMM could enhance the interpretability and accuracy of dietary recommendations, thereby mitigating the current limitations identified in the existing model framework [10].

6. Conclusion

Food nutrition analysis is a vital component of public health research, as highlighted in the abstract. Existing literature reveals a gap in advanced analytical tools capable of capturing the intricate dynamics of food nutrition comprehensively. This study focuses on addressing this gap by introducing a novel approach based on Gaussian Mixture Models, which present a flexible and precise depiction of the underlying structure of food nutrition characteristics. By doing so, this research offers a promising pathway towards enhancing the comprehension and regulation of food nutrition, thus potentially advancing public health strategies. Nevertheless, it is important to acknowledge the limitations inherent in this approach, including the need for further validation and refinement to ensure its applicability across diverse datasets and real-world scenarios. Future work could explore the integration of additional data sources, such as genetic information or lifestyle factors, to enhance the predictive power and utility of the proposed model. Additionally, investigating the scalability of the method and its potential integration with emerging technologies like machine learning algorithms could further extend its impact on tackling the complex challenges in food nutrition analysis and management.

Funding

Not applicable

Author Contribution

Conceptualization, J. L. and C. D.; writing—original draft preparation, J. L. and A. R.; writing—review and editing, C. D. and A. R.; All of the authors read and agreed to the published final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there is no conflict of interests.

Reference

[1] W. Shao et al., "Rapid Non-Destructive Analysis of Food Nutrient Content Using Swin-Nutrition," Foods, vol. 11, 2022.

[2] N. Daud et al., "Obesity Predictor Identification: Comparison of Correlation Based Feature Selection Method and Wrapper Method on Nutrition Dataset," Journal of Advanced Research in Applied Sciences and Engineering Technology, 2024.

[3] M. Abid et al., "Enhancing Software Effort Estimation in Healthcare Informatics: A Comparative Analysis of Machine Learning Models with Correlation-Based Feature Selection," Sustainable Machine Intelligence Journal, 2025.

[4] H. Wang et al., "Nutritional composition analysis in food images: an innovative Swin Transformer approach," Frontiers in Nutrition, 2024.

[5] T. T. Nguyen et al., "Detection of Unknown DDoS Attacks with Deep Learning and Gaussian

Mixture Model," in International Congress on Information and Communication Technology, 2021.
[6] Y. Zhang et al., "Gaussian Mixture Model Clustering with Incomplete Data," in ACM Trans. Multim. Comput. Commun. Appl., 2021.
[7] C. Rasmussen, "The Infinite Gaussian Mixture Model," in Neural Information Processing Systems, 1999.
[8] J. Cao et al., "Unsupervised Eye Blink Artifact Detection From EEG With Gaussian Mixture Model," in IEEE journal of biomedical and health informatics, 2021.
[9] G. Yan et al., "Unknown Intent Detection Using Gaussian Mixture Model with an Application to Zero-shot Intent Classification," in Annual Meeting of the Association for Computational Linguistics, 2020.
[10] P.-M. Lu and Z. Zhang, 'The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering', Journal of Computational Biology and Medicine, vol. 5, no. 1, 2025, Accessed: Mar. 12, 2025.
[11] Q. Zhu, 'Autonomous Cloud Resource Management through DBSCAN-based unsupervised learning', Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Jun. 2025, doi: 10.71070/oaml.v5i1.112.
[12] S. Dan and Q. Zhu, 'Enhancement of data centric security through predictive ridge regression', Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, May 2025, doi: 10.71070/oaml.v5i1.113.
[13] S. Dan and Q. Zhu, 'Highly efficient cloud computing via Adaptive Hierarchical Federated Learning', Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Apr. 2025, doi: 10.71070/oaml.v5i1.114.
[14] Q. Zhu and S. Dan, 'Data Security Identification Based on Full-Dimensional Dynamic Convolution and Multi-Modal CLIP', Journal of Information, Technology and Policy, 2023.
[15] Q. Zhu, 'An innovative approach for distributed cloud computing through dynamic Bayesian networks', Journal of Computational Methods in Engineering Applications, 2024.
[16] Z. Luo, H. Yan, and X. Pan, 'Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques', Journal of Computational Methods in Engineering Applications, pp. 1–12, Nov. 2023, doi: 10.62836/jcmea.v3i1.030107.
[17] H. Yan and D. Shao, 'Enhancing Transformer Training Efficiency with Dynamic Dropout', Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[18] H. Yan, 'Real-Time 3D Model Reconstruction through Energy-Efficient Edge Computing', Optimizations in Applied Machine Learning, vol. 2, no. 1, 2022.
[19] Y. Shu, Z. Zhu, S. Kanchanakungwankul, and D. G. Truhlar, 'Small Representative Databases for Testing and Validating Density Functionals and Other Electronic Structure Methods', J. Phys. Chem. A, vol. 128, no. 31, pp. 6412–6422, Aug. 2024, doi: 10.1021/acs.jpca.4c03137.
[20] C. Kim, Z. Zhu, W. B. Barbazuk, R. L. Bacher, and C. D. Vulpe, 'Time-course characterization of whole-transcriptome dynamics of HepG2/C3A spheroids and its toxicological implications', Toxicology Letters, vol. 401, pp. 125–138, 2024.
[21] J. Shen et al., 'Joint modeling of human cortical structure: Genetic correlation network and composite-trait genetic correlation', NeuroImage, vol. 297, p. 120739, 2024.

[22] K. F. Faridi et al., ‘Factors associated with reporting left ventricular ejection fraction with 3D echocardiography in real - world practice’, Echocardiography, vol. 41, no. 2, p. e15774, Feb. 2024, doi: 10.1111/echo.15774.
[23] Z. Zhu, ‘Tumor purity predicted by statistical methods’, in AIP Conference Proceedings, AIP Publishing, 2022.
[24] Z. Zhao, P. Ren, and Q. Yang, ‘Student self-management, academic achievement: Exploring the mediating role of self-efficacy and the moderating influence of gender insights from a survey conducted in 3 universities in America’, Apr. 17, 2024, arXiv: arXiv:2404.11029. doi: 10.48550/arXiv.2404.11029.
[25] Z. Zhao, P. Ren, and M. Tang, ‘Analyzing the Impact of Anti-Globalization on the Evolution of Higher Education Internationalization in China’, Journal of Linguistics and Education Research, vol. 5, no. 2, pp. 15–31, 2022.
[26] M. Tang, P. Ren, and Z. Zhao, ‘Bridging the gap: The role of educational technology in promoting educational equity’, The Educational Review, USA, vol. 8, no. 8, pp. 1077–1086, 2024.
[27] P. Ren, Z. Zhao, and Q. Yang, ‘Exploring the Path of Transformation and Development for Study Abroad Consultancy Firms in China’, Apr. 17, 2024, arXiv: arXiv:2404.11034. doi: 10.48550/arXiv.2404.11034.
[28] P. Ren and Z. Zhao, ‘Parental Recognition of Double Reduction Policy, Family Economic Status And Educational Anxiety: Exploring the Mediating Influence of Educational Technology Substitutive Resource’, Economics & Management Information, pp. 1–12, 2024.
[29] Z. Zhao, P. Ren, and M. Tang, ‘How Social Media as a Digital Marketing Strategy Influences Chinese Students’ Decision to Study Abroad in the United States: A Model Analysis Approach’, Journal of Linguistics and Education Research, vol. 6, no. 1, pp. 12–23, 2024.
[30] Z. Zhao and P. Ren, ‘Identifications of Active Explorers and Passive Learners Among Students: Gaussian Mixture Model-Based Approach’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, May 2025.
[31] Z. Zhao and P. Ren, ‘Prediction of Student Answer Accuracy based on Logistic Regression’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Feb. 2025.
[32] Z. Zhao and P. Ren, ‘Prediction of Student Disciplinary Behavior through Efficient Ridge Regression’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Mar. 2025.
[33] Z. Zhao and P. Ren, ‘Random Forest-Based Early Warning System for Student Dropout Using Behavioral Data’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Apr. 2025.
[34] P. Ren and Z. Zhao, ‘Recognition and Detection of Student Emotional States through Bayesian Inference’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, May 2025.
[35] P. Ren and Z. Zhao, ‘Support Vector Regression-based Estimate of Student Absenteeism Rate’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Jun. 2025.
[36] G. Zhang and T. Zhou, ‘Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model’, IAET, pp. 1–13, Sep. 2024, doi: 10.62836/iaet.v3i1.232.
[37] G. Zhang, W. Huang, and T. Zhou, ‘Performance Optimization Algorithm for Motor Design with Adaptive Weights Based on GNN Representation’, Electrical Science & Engineering, vol. 6, no. 1, Art. no. 1, Oct. 2024, doi: 10.30564/ese.v6i1.7532.

[38] T. Zhou, G. Zhang, and Y. Cai, ‘Unsupervised Autoencoders Combined with Multi-Model Machine Learning Fusion for Improving the Applicability of Aircraft Sensor and Engine Performance Prediction’, Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Feb. 2025, doi: 10.71070/oaml.v5i1.83.
[39] Y. Tang and C. Li, ‘Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises’, Journal of Computational Methods in Engineering Applications, pp. 1–17, 2023.
[40] C. Li and Y. Tang, ‘Emotional Value in Experiential Marketing: Driving Factors for Sales Growth–A Quantitative Study from the Eastern Coastal Region’, Economics & Management Information, pp. 1–13, 2024.
[41] C. Li and Y. Tang, ‘The Factors of Brand Reputation in Chinese Luxury Fashion Brands’, Journal of Integrated Social Sciences and Humanities, pp. 1–14, 2023.
[42] C. Y. Tang and C. Li, ‘Examining the Factors of Corporate Frauds in Chinese A-share Listed Enterprises’, OAJRC Social Science, vol. 4, no. 3, pp. 63–77, 2023.
[43] W. Huang, T. Zhou, J. Ma, and X. Chen, ‘An ensemble model based on fusion of multiple machine learning algorithms for remaining useful life prediction of lithium battery in electric vehicles’, Innovations in Applied Engineering and Technology, pp. 1–12, 2025.
[44] W. Huang and J. Ma, ‘Predictive Energy Management Strategy for Hybrid Electric Vehicles Based on Soft Actor-Critic’, Energy & System, vol. 5, no. 1, 2025.
[45] J. Ma, K. Xu, Y. Qiao, and Z. Zhang, ‘An Integrated Model for Social Media Toxic Comments Detection: Fusion of High-Dimensional Neural Network Representations and Multiple Traditional Machine Learning Algorithms’, Journal of Computational Methods in Engineering Applications, pp. 1–12, 2022.
[46] W. Huang, Y. Cai, and G. Zhang, ‘Battery Degradation Analysis through Sparse Ridge Regression’, Energy & System, vol. 4, no. 1, Art. no. 1, Dec. 2024, doi: 10.71070/es.v4i1.65.
[47] Z. Zhang, ‘RAG for Personalized Medicine: A Framework for Integrating Patient Data and Pharmaceutical Knowledge for Treatment Recommendations’, Optimizations in Applied Machine Learning, vol. 4, no. 1, 2024.
[48] Z. Zhang, K. Xu, Y. Qiao, and A. Wilson, ‘Sparse Attention Combined with RAG Technology for Financial Data Analysis’, Journal of Computer Science Research, vol. 7, no. 2, Art. no. 2, Mar. 2025, doi: 10.30564/jcsr.v7i2.8933.
[49] Y. Qiao, K. Xu, Z. Zhang, and A. Wilson, ‘TrAdaBoostR2-based Domain Adaptation for Generalizable Revenue Prediction in Online Advertising Across Various Data Distributions’, Advances in Computer and Communication, vol. 6, no. 2, 2025.
[50] K. Xu, Y. Gan, and A. Wilson, ‘Stacked Generalization for Robust Prediction of Trust and Private Equity on Financial Performances’, Innovations in Applied Engineering and Technology, pp. 1–12, 2024.
[51] A. Wilson and J. Ma, ‘MDD-based Domain Adaptation Algorithm for Improving the Applicability of the Artificial Neural Network in Vehicle Insurance Claim Fraud Detection’, Optimizations in Applied Machine Learning, vol. 5, no. 1, 2025.