British Journal of Nutrition

cambridge.org/bjn

Horizons in Nutritional Science

Cite this article: Lee H, Shin W-K, Heo D, Min S, Cho S, Lee S-Y, Choi J-Y, Suh B, and Kang D (2026). Machine learning-based dietary pattern analysis: a narrative review and future perspectives. *British Journal of Nutrition*, page 1 of 12. doi: 10.1017/S0007114526108630

Received: 3 April 2026 UTC
Revised: 8 July 2026 UTC
Accepted: 10 August 2026 UTC

Keywords:
Dietary pattern; Dietary assessment; Nutritional epidemiology; Machine learning; Cancer prevention

Abbreviations:
24HR, 24-h recalls; AE, autoencoders; ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention; CA, cluster analysis; PCA, principal component analysis; UMAP, Uniform Manifold Approximation and Projection

Corresponding author:
Daehee Kang; Email: dhkang@snu.ac.kr

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Nutrition Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0007114526108630

Machine learning-based dietary pattern analysis: a narrative review and future perspectives

Hyobin Lee[1,2], Woo-Kyoung Shin[3], Dongseok Heo[4], Sukhong Min[1], Sinyoung Cho[1,5], So-Yoon Lee[1], Ji-Yeob Choi[6,7], Bongwon Suh[8] and Daehee Kang[1,2]

[1]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea; [2]Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Republic of Korea; [3]Division of Food and Pharmaceutical Technology, Mokwon University, Daejeon, Republic of Korea; [4]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea; [5]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea; [6]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Republic of Korea; [7]Cancer Research Institute, Seoul National University, Seoul, Republic of Korea and [8]Department of Intelligence and Information, Seoul National University, Seoul, Republic of Korea

Abstract

Over recent decades, nutritional epidemiology has evolved from nutrient-specific investigations to dietary pattern analysis, complementing traditional dietary assessment methods. With the global aging population and rising chronic disease burden, particularly cancer, identifying modifiable dietary risk factors through holistic dietary assessments is essential. Data-driven dietary pattern analysis methods, including principal component analysis and cluster analysis, have been widely used to derive dietary patterns. However, challenges in cluster quality persist, especially with dietary data. Recent advances in machine learning, including non-linear dimensionality reduction, offer promising strategies to enhance cluster separability in dietary pattern research. This review summarises the evolution of dietary assessment methodologies, outlines the development of dietary pattern analytical approaches and synthesises epidemiological evidence linking clustering-driven dietary patterns to cancer and metabolic disease risk. Furthermore, it highlights methodological gaps and discusses future directions emphasising the integration of machine learning methods for practical application in personalised cancer prevention. Establishing standardised frameworks combining dimensionality reduction and clustering may facilitate the translation of validated dietary patterns into digital platforms for widespread public health impact.

Over the past several decades, nutritional epidemiology has undergone a transition from nutrient-focused investigations to more holistic assessments of overall dietary habits. Traditional dietary assessment methods such as 24-h recalls (24HR) and FFQ have been applied to large-scale cohort studies but remain subject to recall bias and measurement error. To better characterise overall dietary behaviours despite these limitations, data-driven approaches for identifying dietary patterns have been increasingly adopted.

With the consistent rise in the proportion of the elderly, the prevalence of chronic diseases, particularly cancer, is of significant concern. According to estimates from the WHO in 2019, cancer is the first or second leading cause of death before the age of 70 years in 112 of 183 countries[(1)]. Also, the Global Cancer Observatory estimated that the global cancer burden is expected to undergo a 47 % rise in 2040 compared to 2020[(2)]. This escalating burden underscores the urgent need to identify modifiable risk factors, including diet, that contribute to effective cancer prevention.

In this context, dietary pattern analysis provides a more comprehensive framework for understanding the role of diet in chronic disease prevention. By capturing the combined and potentially synergistic effects of multiple foods, it reflects real-world dietary behaviours comprehensively[(3)]. Among available methods, principal component analysis (PCA) and cluster analysis (CA) are the most widely used tools for deriving empirical dietary patterns[(4)]. However, for CA, challenges related to cluster quality persist, particularly in dietary datasets.

Recent methodological advances in computational science – especially machine learning-based techniques such as Uniform Manifold Approximation and Projection (UMAP) and deep learning-based autoencoders (AE) – offer promising solutions for modelling non-linear relationships and improving cluster separability. Integrating these approaches into dietary pattern analysis, together with rigorous dietary assessment and appropriate validation, may contribute to more robust analytical frameworks. If validated in independent populations, these approaches could support the future development of digital health tools for personalised

nutrition and chronic disease prevention, while also informing population-level public health strategies.

However, their successful application depends on the quality of the underlying dietary assessment data and requires careful consideration of dietary measurement error. This perspective is consistent with recent methodological work highlighting that the performance of neural network models for dietary data can be substantially affected by dietary measurement error and therefore requires appropriate handling of error-prone dietary exposures[5].

Existing reviews have separately addressed advanced dietary assessment technologies[6], analytical approaches for deriving dietary patterns[7] or the application of machine learning in personalised nutrition[8]. More recently, a scoping review summarised advances in methods for dietary pattern characterisation, highlighting the growing diversity of analytical approaches while also emphasising the need for greater methodological consistency and standardised reporting[9]. What remains underexplored is a unifying perspective that connects these advances to the challenges of dietary pattern research.

To address this gap, this review aims to (1) summarise the evolution of dietary assessment methodologies, (2) describe the conceptual development and methodological advances in dietary pattern analysis, (3) synthesise findings from prospective cohort studies linking data-driven dietary patterns to cancer and metabolic disease risk and (4) discuss future directions emphasising the integration of advanced machine learning approaches to enhance real-world applicability in dietary pattern research.

To improve transparency, we briefly describe how the literature included in this review was identified. The literature included in this narrative review was identified through a combination of targeted database searches and manual screening of relevant publications and reference lists. To summarise the epidemiological evidence presented in Tables 1 and 2, a structured search of PubMed and Web of Science was conducted using combinations of keywords related to dietary patterns, CA and chronic disease outcomes (Supplementary Table 1). Studies were screened based on their relevance to cluster-derived dietary patterns and cancer or metabolic disease outcomes, with emphasis on prospective cohort studies.

Evolution of dietary assessment methodology

Conventional dietary assessment methods

Accurate assessment of dietary intake in daily life and identification of an individual's representative habitual intake are crucial for developing personalised dietary advice. Conventional dietary assessment methods used in epidemiological studies can largely be categorised into four approaches: dietary records, 24HR, FFQ and dietary biomarkers[10]. Each method possesses intrinsic strengths and limitations, influencing their distinct applications[11]. Dietary records involve real-time recording of all foods consumed, while 24HR relies on retrospective recall with the assistance of trained interviewers. Both methods can yield detailed and accurate intake data but may not capture habitual long-term dietary patterns. FFQ, in contrast, are designed to assess habitual intake over extended periods, typically the past year, by querying the frequency of consumption of approximately 100–150 food items. While FFQ are cost-effective and suitable for large-scale epidemiological studies, they have limitations such as recall bias and difficulties in estimating portion sizes – particularly in Asian dietary contexts, where meals often consist of mixed dishes and shared portions. Furthermore, FFQ have limited ability to capture emerging dietary variables, such as meal timing and ultra-processed food consumption, which are increasingly considered relevant to chronic disease risk. Despite these challenges, FFQ remain the primary tool in prospective cohort studies because of their practicality and ability to capture habitual intake. Dietary biomarkers measured in biological specimens, such as blood and urine, can provide objective indicators of dietary exposure and complement self-reported dietary assessment methods. Examples include urinary nitrogen for protein intake and serum carotenoids for fruit and vegetable intake. Although they cannot comprehensively capture overall dietary patterns, dietary biomarkers play an important role in validating self-reported dietary data.

Advances in dietary assessment tools utilising new technology

Research has been focusing on refining and developing new tools to improve accuracy and reduce the burden of both researchers and participants, particularly by integrating digital tools such as smartphones, cameras and sensors into the four existing dietary assessment methodologies[12,13], as depicted in Figure 1.

Integration of digital technology into dietary assessment can be traced back to the advent of paperless forms of 24HR and FFQ[14]. Examples of this digital integration include the online Automated Self-Administered 24HR[15], along with various mobile device applications as summarised by Schembre *et al.*[16]. More recently, image-based dietary record tools emerged, which utilise digital cameras to capture food consumption via photographs[17]. The objective is to decrease the burden on participants and resolve issues of underestimation and recall bias inherent in traditional methods. Examples of these tools include the Mobile Food Record[18] and the Remote Food Photography Method[19], where participants capture everything they eat during a defined time period by taking a photo before and after each meal. However, challenges such as high costs, lack of database and the need for enhanced algorithms to analyse the images remain.

Moreover, wearable cameras attached to various body parts, such as the SenseCam and e-Button[20], and sensors worn on the neck or wrist detecting eating motions and the tooth and epidermal sensors detecting metabolites are designed to reduce bias and participant burden while capturing continuous real-time data[13]. Dietary records using wearable cameras and sensors represent a novel approach in dietary assessments, with wearable devices continuously capturing dietary events or specific nutrients or metabolites associated with eating. Recent initiatives, such as the Nutrition for Precision Health powered by the All of Us Research Program, have also incorporated emerging technologies, including wearable devices (e.g. the Automatic Ingestion Monitor-2)[21], to improve the objective assessment of dietary behaviours.

Although technology-based dietary assessment tools offer several advantages, including reduced participant burden, automated data processing and the ability to capture eating behaviours in real time, they also present important challenges. Most currently available tools still rely on self-reported dietary intake and therefore remain susceptible to measurement error and reporting bias. In addition, rigorous validation, standardised reporting and transparent documentation of tool development are essential to ensure data quality. Accordingly, these technologies should be viewed as complementary to, rather than replacements for, conventional dietary assessment methods[22]. No single dietary assessment method is optimal for all research settings, and the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

Table 1. Epidemiological studies examining associations between cluster-derived dietary patterns and cancer

Algorithm	Design	Data source (country)	Outcome	Number of participants	Number of cases	Follow-up, years	Number of clusters	Names of the clusters	Results		Author, year
									OR	95 % CI	
Cluster analysis	Prospective study	NHANES (USA)	Cancer mortality	3518 chronic kidney disease patients	137 deaths	7·3	4	C1 (low vitamin)	1·00	ref.	Li, 2024[56]
								C2	0·68	0·36–1.29	
								C3 (vitamin C/K)	0·42	0·20–0.88	
								C4	0·64	0·36–1.14	
	Prospective study	NHANES (USA)	All-cause mortality, Cancer mortality, CVD mortality	12 724	6·99 % cancer deaths 10·47 % CCVD deaths	7·1	3	C1 (high HEI)	1·00	ref.	Liu, 2022[57]
								C2	1·43	1·17–1.75	
								C3 (low HEI)	1·30	1·12–1.41	
	Prospective cohort study	Newfoundland and Labrador colorectal cancer cohort (Canada)	Colorectal cancer survival	532 colorectal cancer patients	170 deaths	6·3	4	C1 (fruits and vegetables)	1·00	ref.	Sharma, 2018[58]
								C2 (meat and dairy products)	2·19	1·03–4.67	
								C3 (grains sugar)	1·95	1·13–3.37	
								C4 (many foods)	1·55	0·92–2.61	
	Prospective cohort study	NIH–AARP Diet and Health Study (USA)	Colorectal cancer incidence	492 306	3110	4·5	4 for men 3 for women	Many foods	1·00	ref.	Wirfalt, 2009[59]
								Vegetables and fruits	0·85	0·76–0.94	
								Fatty meats	0·94	0·80–1.10	
								Fat-reduced foods	0·88	0·70–1.11	
	Prospective cohort study	Whitehall II Study (UK)	All-cause mortality, Cancer mortality	7731	376 deaths	15	4	Unhealthy	1·00	ref.	Brunner, 2008[60]
								Sweet	0·85	0·51–1.43	
								Mediterranean-like	0·88	0·54–1.44	
								Healthy	1·15	0·80–1.64	
MCA and cluster analysis	Prospective cohort study	NRP1A, MONICA (Switzerland)	All-cause mortality, Cardiovascular mortality Cancer mortality	15 936	1460 cancer deaths	Up to 37·9	5	Sausage and vegetables	1·00	ref.	Krieger, 2018[61]
								Meat and salad	0·95	0·82–1.11	
								Fish	0·82	0·68–0.99	
								Traditional	0·93	0·79–1.10	
								High-fibre foods	0·85	0·72–1.00	

(*Continued*)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

Table 1. (*Continued*)

Algorithm	Design	Data source (country)	Outcome	Number of participants	Number of cases	Follow-up, years	Number of clusters	Names of the clusters	Results		Author, year
									OR	95 % CI	
PCA and cluster analysis	Prospective cohort study	EPIC–Norway (Norway)	Total cancer, breast cancer, Gastrointestinal cancer, Colon cancer, Rectal cancer, Colorectal cancer incidence	34 471	1355	6	6	Fish	0·95	0·73–1.23	Engeset, 2009[51]
								Healthy	0·92	0·79–1.08	
								Average	1·00	ref.	
								Western	0·80	0·68–0.95	
								Bread	0·86	0·73–1.00	
								Alcohol	1·10	0·91–1.34	
	Case–control study	Morocco	Colorectal cancer risk	1483	1483	–	2	Dangerous	1·59	1·37–1.83	Qarmiche, 2023[52]
								Prudent	1·00	ref.	
	Case–control study	China	Colorectal cancer risk	2799	2799	–	2	Refined grain	1·00	ref.	Ma, 2023[53]
								Balanced	0·62	0·52–0.74	
	Case–control study	Italy	Breast cancer, Ovarian cancer risk	3413	2569 (breast cancer) 1031 (ovarian cancer)	–	5	G1 (fruits/ vegetables)	0·96	0·84–1.10	Edefonti, 2009[54]
								G2 (meat/eggs/ dairy products)	0·88	0·77–1.02	
								G3 (low intakes)	1·00	ref.	
								G4 (olive oil)	1·00	0·86–1.16	
								G5 (bread/pasta)	1·23	1·08–1.38	

NHANES, National Health and Nutrition Examination Survey; CCVD, cerebrocardiovascular disease; HEI, Healthy Eating Index; NIH–AARP, National Institutes of Health–American Association of Retired Persons; MCA, multiple correspondence analysis; NRP1A, National Research Program 1A; MONICA, Monitoring of Trends and Determinants in Cardiovascular Disease; EPIC, European Prospective Investigation into Cancer and Nutrition; PCA, principal component analysis.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

Table 2. Epidemiological studies examining associations between cluster-derived dietary patterns and metabolic outcomes

Algorithm	Design	Data source (country)	Outcome	Number of participants	Number of cases	Follow-up, years	Number of clusters	Names of the clusters	Results		Author, year
									OR	95 % CI	
UMAP and cluster analysis	Prospective cohort study	Oroshisho study (Japan)	Hypertension incidence	447	64	2	4	Seafood and alcohol	1·00	ref.	Li, 2024[55]
								Dairy products/ vegetables-based	0·43	0·21–0.86	
								Meat-based	0·38	0·16–0.86	
								Low-fibre high-sugar	0·35	0·13–0.92	
FA and cluster analysis	Prospective cohort study	Cardiovascular Disease Association Study (CAVAS) (Korea)	Metabolic syndrome incidence	11 305	2398	4·8	3 for men 4 for women	C1 (low in all factors)	0·80	0·64–1.00	Kim, 2022[62]
								C2 (vegetables/ seaweeds)	1·00	ref.	
								C3 (non-traditional/non-staple foods)	0·71	0·45–0.93	
								C4 (meat/poultry/ seafood)	0·69	0·54–0.88	
Cluster analysis	Prospective cohort study	Ansan-Ansung cohort study (Korea)	Obesity incidence	6539	1812	11	3	White rice and kimchi	1·00	ref.	Tan, 2024[63]
								Multigrain rice and kimchi	0·87	0·79–0.95	
								Barley and kimchi	1·07	0·94–1.22	
	Prospective cohort study	The Malmö Diet and Cancer Study (Sweden)	Diabetes incidence	26 622	4046	18·4	5	High refined carbs/low fruit and vegetables	1·00	ref.	Olsson, 2023[64]
								High vegetables/low added sugar	0·99	0·88–1.12	
								High sugar-sweetened beverages	0·97	0·86–1.08	
								High juice	0·95	0·86–1.05	
								High fruit	0·86	0·78–0.94	
	Prospective cohort study	The Geisinger Rural Aging Study (USA)	Metabolic syndrome incidence	449	146	5	3	Health-conscious	1·00	ref.	Hsiao, 2013[68]
								Sweets and dairy products	1·24	0·69–2.26	
								Western	0·92	0·46–1.85	
	Prospective cohort study	Vasterbotten Intervention Programme (Sweden)	Obesity incidence	6545	535	10	2	Fruits and vegetables	1·30	0·89–1.91	Nyholm, 2013[69]
								Mixed pattern	1·00	ref.	

(*Continued*)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

Table 2. (*Continued*)

Algorithm	Design	Data source (country)	Outcome	Number of participants	Number of cases	Follow-up, years	Number of clusters	Names of the clusters	Results		Author, year
									OR	95 % CI	
	Prospective cohort study	The Framingham Heart Study (USA)	Metabolic syndrome incidence	1146	355	7	5	Empty calorie	1·00	ref.	Kimokoti, 2012[70]
								Higher fat	0·62	0·29–1.32	
								Wine and moderate eating	0·46	0·14–1.50	
								Lighter eating	0·67	0·33–1.34	
								Heart healthier	0·88	0·42–1.87	
	Prospective cohort study	Coronary Artery Risk Development in Young Adults (CARDIA) study (USA)	Metabolic syndrome incidence	3728	–	20	2	Western consumers	1·00	ref.	Duffey, 2012[65]
								Western non-consumers	0·85	0·69–1.07	
								Prudent consumers	0·84	0·64–1.10	
								Prudent non-consumers	0·64	0·50–0.82	
	Prospective cohort study	Shanghai Women's Health Study (SWHS) (China)	Diabetes incidence	64 191	2267	6·9	3	Staples	1·00	ref.	Villegas, 2010[66]
								Dairy products	0·68	0·62–0.75	
								High energy intake	0·95	0·74–1.22	
	Prospective cohort study	Whitehall II Study (UK)	Diabetes incidence	7731	410	15	4	Unhealthy	1·00	ref.	Brunner, 2008[60]
								Sweet	1·16	0·83–1.61	
								Mediterranean-like	1·04	0·75–1.43	
								Healthy	0·74	0·58–0.94	
	Prospective cohort study	The Framingham Nutrition Studies (USA)	Overweight incidence	737	214	12	5	Heart healthy	1·00	ref.	Quatromoni, 2002[67]
								Light eating	1·10	0·80–1.60	
								Wine and moderate eating	0·90	0·40–2.10	
								High fat	1·00	0·70–1.50	
								Empty calorie	1·40	0·90–2.20	

UMAP, Uniform Manifold Approximation and Projection; FA, factor analysis.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

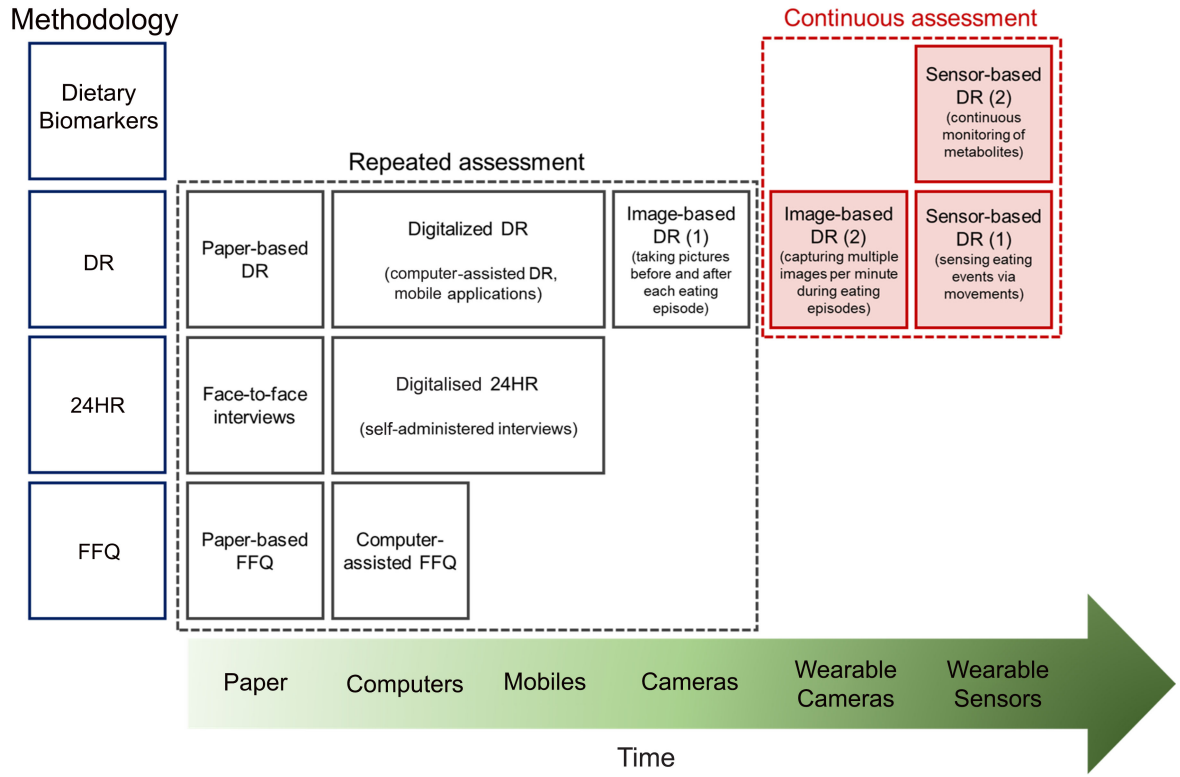


Fig. 1. Evolution of dietary assessment tools from past to future. The *x*-axis represents the progression from manual tools to technology-based tools, while the *y*-axis differentiates the four dietary assessment methods. 24HR, 24-h recall; DR, dietary record.

choice of method should be guided by the specific research objective while balancing considerations such as accuracy, participant burden, feasibility, cost and scalability. Consequently, although technology-based tools may enhance dietary assessment in specific settings, FFQ continue to play a central role in prospective cohort studies because they provide a practical, cost-effective and scalable approach for assessing habitual dietary intake over extended periods in large populations, despite their inherent limitations in measurement error and recall bias.

Dietary pattern analysis: concepts, methods and applications

Historical context and rationale

Nutritional epidemiology has undergone a paradigm shift over the past several decades as depicted in Figure 2. Early research focused primarily on nutrient deficiencies[23]. With the development of FFQ and the establishment of large cohort studies, attention shifted towards evaluating the effects of single nutrients on chronic disease risk. However, major randomised controlled trials, such as the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study, yielded null or inconclusive findings on the benefits of vitamin and mineral supplementation[24], prompting a transition towards examining individual foods and food groups. Also, O'Connor and colleagues[25] systematically reviewed trial evidence to update the US Preventive Services Task Force recommendation on the efficacy of vitamin supplements for primary prevention. The authors concluded that there was no clear evidence of a beneficial effect of supplements on all-cause mortality, CVD or cancer[25]. More recently, the focus has shifted to dietary pattern analysis, which captures the combined and potentially synergistic effects of multiple foods and better reflects real-world dietary behaviours. This evolution is also evident in major cancer prevention reports. While the 2018 report of the World Cancer Research Fund[26] primarily focused on individual foods and nutrients, its 2025 update[27] incorporates dietary patterns and lifestyle patterns as a more holistic approach to cancer prevention. This evolution in dietary pattern research has also been accompanied by growing recognition that diet is a multidimensional and dynamic exposure. Building on this perspective, the National Cancer Institute and the National Institutes of Health convened the 'Extending Methods in Dietary Patterns Research' workshop, which identified several methodological priorities for advancing dietary pattern research, including improved dietary assessment, standardisation of dietary pattern methods, consideration of dietary dynamism across the life course and better approaches for addressing measurement error[28]. These recommendations provide an important conceptual foundation for the continued development of dietary pattern research.

Classification of analytical approaches

Dietary patterns are defined as the quantities, proportions, variety and combinations of different foods and beverages in diets and the frequency with which they are habitually consumed[29,30]. Two primary methodological approaches have been established for analysing dietary patterns: *a priori* methods and *a posteriori* methods[31,32].

A priori methods, such as the Healthy Eating Index[33], rely on predefined scoring systems based on established dietary guidelines. These methods use fixed criteria to assess adherence to recommended dietary behaviours, producing standardised and interpretable scores. Large collaborative efforts, such as the Dietary

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

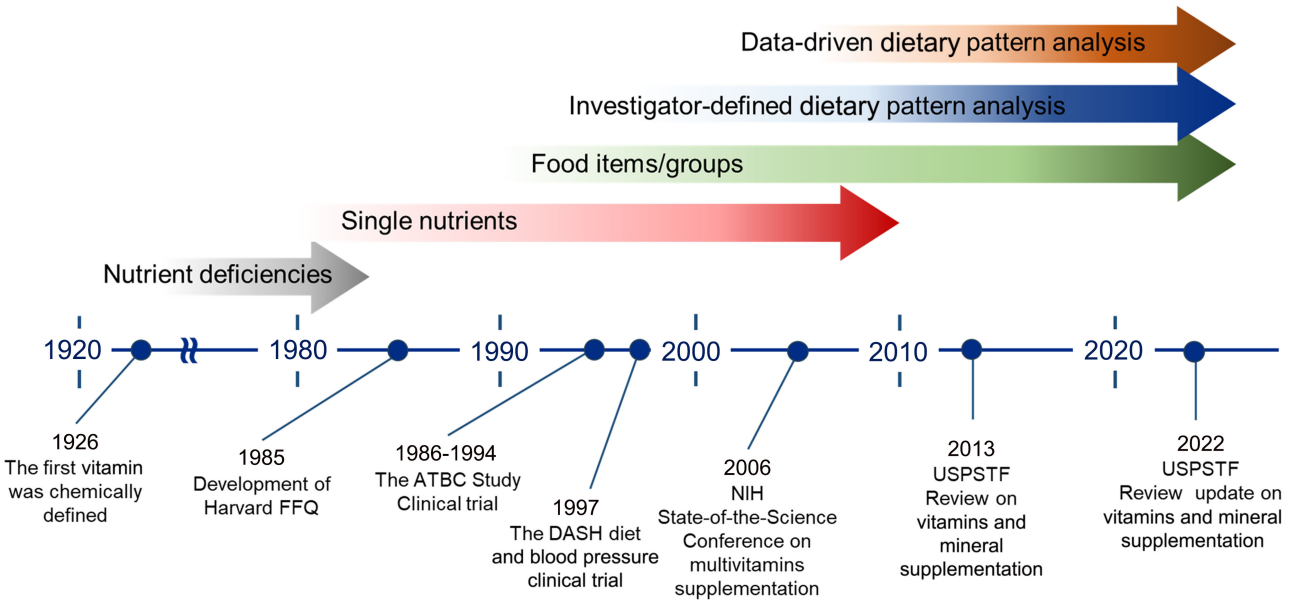


Fig. 2. Shift of the focus of nutritional epidemiology research. ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; DASH, Dietary Approaches to Stop Hypertension; NIH, National Institutes of Health; USPSTF, US Preventive Services Task Force.

Patterns Methods Project, have further demonstrated that major dietary quality indices – including the Healthy Eating Index, Alternative Healthy Eating Index, Mediterranean Diet score and Dietary Approaches to Stop Hypertension score – consistently capture key aspects of healthy dietary patterns and provide a robust basis for dietary guidance[34]. In contrast, *a posteriori* methods are data-driven and include techniques such as PCA, factor analysis and CA. PCA and factor analysis extract continuous dietary patterns – such as the 'Western', 'Prudent' or 'Traditional' diet – and assign each person an adherence score[35]. Although PCA and factor analysis are often used for similar purposes in nutritional epidemiology, they differ conceptually. PCA is a technique that transforms correlated variables into orthogonal components explaining the maximum variance in the observed data[36], whereas factor analysis assumes that the observed correlations arise from a smaller number of underlying latent factors[37]. While these approaches capture variation in dietary intake, the reliance on continuous scores can limit direct interpretability in some contexts. CA instead produces mutually exclusive groups with distinct dietary profiles[35], facilitating clearer subgroup identification and potentially more actionable results. The commonly used method is the k-means clustering algorithm, which partitions the data into k clusters by assigning each data point to the nearest cluster centroid after specifying the desired number of clusters. Recent work has further demonstrated that CA can complement Healthy Eating Index-based assessment by identifying distinct dietary subgroups within individuals with similarly high diet quality scores, highlighting heterogeneity that may not be captured by a single dietary index[38]. Clustering quality is often evaluated using the silhouette coefficient, which compares the average distance between a data point and all other points in the same cluster with the average distance to points in the nearest neighbouring cluster. The coefficient ranges from −1 to 1, with values closer to 1 indicating better-defined and more cohesive clusters[39]. Although the silhouette coefficient is widely used to assess cluster separation, it has limitations because it is sensitive to cluster shape and density and therefore should not be used as the sole criterion for evaluating clustering performance[40]. Hybrid methods, including reduced rank regression[35], combine elements of both approaches by linking dietary patterns to predefined health outcomes or biomarkers. Although this outcome-driven focus can be valuable for elucidating diet-disease relationships, it may be less suited to identifying novel dietary patterns that are not closely aligned with the selected response variables.

Machine learning methods in dietary pattern analysis

Machine learning encompasses a range of algorithms and statistical models designed to automatically learn patterns from data[41]. In dietary pattern research, machine learning offers powerful tools for handling complex, multidimensional datasets[42]. Historically, the field of dietary pattern analysis has relied on unsupervised learning, with PCA and CA being the most traditional techniques[4].

However, generating stable and meaningful clusters from dietary data can be challenging. Outside the field of nutritional research, dimensionality reduction techniques – including PCA and UMAP – have been applied prior to clustering to improve clustering quality, enhance visualisation and reduce noise[43–45]. Beyond conventional machine learning, methods that perform clustering after dimensionality reduction with deep learning models such as AE[46] have been proposed since the 2010s[47]. According to a recent survey, these approaches, often referred to as 'deep clustering', have primarily been applied to image, text, video and graph data[48]. Also, their use has extended to tabular data as well[49,50].

However, evidence supporting their application to dietary pattern analysis remains limited, and their potential benefits in nutritional epidemiology have yet to be systematically evaluated. Most studies perform clustering directly on dietary intake data; only a few have used PCA as a preprocessing step[51–54], and the application of newer methods such as UMAP is limited to a single known example[55].

This gap suggests untapped potential for improving dietary pattern research through integrated workflows that combine dimension reduction with clustering. The literature review in the *Methodological trends and research gaps in epidemiological studies*

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

section will examine these possibilities in greater detail, highlighting both current applications and opportunities for methodological advancement.

Methodological trends and research gaps in epidemiological studies

Numerous epidemiological studies have applied PCA or CA to FFQ-derived dietary data. PCA is more commonly used, as it effectively summarises dietary variation into a smaller set of components. However, unlike PCA, CA offers the advantage of classifying individuals into discrete groups, making the results more interpretable and potentially actionable for public health recommendations. However, ensuring high-quality, well-separated clusters in FFQ data remains challenging.

To date, only a small number of studies have applied CA to derive dietary patterns and examine their associations with cancer risk (Table 1).

Specifically, five studies applied CA alone[56–60], and five studies combined dimensionality reduction with CA[51–54,61]. Among these, only two – by Wirfalt *et al.*[59] and Engeset *et al.*[51] – were prospective cohort studies using cancer incidence as the outcome. Notably, Engeset *et al.* investigated the association between clusters and cancer incidence in the European Prospective Investigation into Cancer and Nutrition–Norway cohort. One case–control study by Qarmiche *et al.*[52] reported a silhouette score for clustering quality; however, the score was low (0·12). Apart from this study, no other research reported quantitative metrics of cluster quality. This scarcity of studies underscores the need for further methodological development and application of CA in cancer epidemiology.

Given the limited number of prospective cohort studies that have applied CA to cancer outcomes, it is informative to examine studies investigating other chronic disease outcomes – particularly metabolic syndrome and its components. Table 2 summarises several prospective cohort studies that have applied CA to examine the risk of metabolic syndrome or its components[55,60,62–70]. In most cases, clustering was conducted without prior dimensionality reduction, which may limit cluster separability and obscure meaningful dietary distinctions. The number of clusters was also typically small (3–5), potentially overlooking dietary heterogeneity. A notable exception is a 2024 study by Li *et al.*[55], which applied UMAP prior to clustering to investigate hypertension risk. The reported silhouette score was 0·4, but higher than in previous studies – indicating a methodological improvement. Nonetheless, the study was constrained by a small sample size and short follow-up period. Across these studies, several recurring methodological limitations are evident. First, quantitative metrics to assess clustering quality – such as silhouette scores – were rarely reported, making it difficult to evaluate the robustness and separation of the identified clusters. Second, clustering was often performed without prior dimensionality reduction, a practice that may reduce cluster separability and obscure meaningful distinctions in dietary patterns. Third, the number of clusters was typically restricted to 3–5, which could oversimplify the diversity of dietary behaviours within a population. Finally, the application of advanced machine learning-based dimensionality reduction techniques capable of capturing non-linear relationships in dietary data remains limited. Addressing these limitations through more systematic analytic frameworks and the adoption of machine learning methods has the potential to enhance the identification of meaningful dietary patterns and improve their applicability across diverse populations.

Despite differences in study design and disease outcomes, several common patterns emerged across the reviewed studies. Overall, the available epidemiological evidence suggests a generally consistent association between cluster-derived dietary patterns and chronic disease risk. Dietary patterns characterised by higher consumption of vegetables, fruits and dairy products were generally associated with lower risks of cancer or metabolic diseases. In contrast, dietary patterns characterised by refined grains or unhealthy dietary habits tended to be associated with increased disease risk. Although the magnitude and statistical significance of these associations varied across studies, the overall direction of the findings was broadly consistent. Differences in study populations, dietary assessment methods, clustering algorithms and outcome definitions likely contributed to the observed heterogeneity.

Discussion and future directions

Advances in machine learning offer promising avenues for dietary pattern analysis. AE can learn non-linear representations of dietary data, potentially capturing complex relationships among foods that are not detectable with linear methods. Coupling such representations with clustering algorithms may enhance clustering quality, even in dietary datasets. Future research should systematically compare multiple dimensionality reduction methods, including PCA, UMAP and more recently, deep learning-based approaches such as AE, to identify the optimal analytical workflow for dietary data clustering.

A priority for future work is to apply these methods across large-scale prospective cohort studies and rigorously evaluate clustering performance using standardised metrics alongside interpretability measures. Validation should extend beyond internal clustering performance to include external validation in independent cohorts, assessment of the reproducibility of dietary patterns across populations, and evaluation of whether the identified patterns demonstrate consistent associations with relevant health outcomes. Such evaluations should aim to determine the most effective combination of dimensionality reduction and clustering algorithms for dietary data, establishing a workflow that can be applied across diverse datasets. Demonstrating consistent associations between the identified dietary patterns and cancer or other chronic disease outcomes will be essential for establishing their epidemiological and public health relevance. Inconsistent findings in diet-cancer studies could stem from exposure misclassification. When dietary patterns are poorly separated, individuals with distinct risks are grouped together, attenuating the observed association. Improving clustering quality can reduce this, allowing for a clearer detection of the significant effects of dietary patterns on cancer risk. In the longer term, if dietary patterns can be consistently identified across cohorts, translating these patterns into digital health tools may support personalised dietary guidance and chronic disease prevention. Integrating such platforms into healthcare and public health systems would bridge the gap between epidemiological research and real-world dietary interventions, enhancing both prevention and precision nutrition efforts.

Despite their promise, machine learning-based approaches also introduce practical challenges for epidemiological research. AE typically require careful tuning of model architecture and hyperparameters, and their flexibility raises the risk of overfitting, particularly when training datasets are relatively small or lack sufficient diversity. Consequently, model performance may depend on sample size and data quality, making robust internal validation essential. In addition, representation learning may be sensitive to cohort-specific dietary distributions, potentially limiting the transferability of learned representations across populations

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

without appropriate harmonisation. Compared with PCA, the latent features generated by AE are also less transparent, making it more difficult to interpret their biological or nutritional meaning. Finally, future studies should therefore prioritise external validation in diverse populations to evaluate the robustness and broader applicability of these approaches. Additionally, cluster quality metrics are necessary but not sufficient; clusters must also be nutritionally meaningful. Translating clusters into meaningful dietary patterns can be strengthened by standardised interpretation steps, such as (i) reporting cluster profiles using relative food-group intake frequencies, (ii) identifying the foods that most distinguish clusters (e.g. top relative intakes or feature-importance approaches applied to cluster assignment) and (iii) applying transparent labelling rules to name patterns consistently across studies. Such practices can improve the interpretability of dietary patterns.

Although methodological advances such as non-linear dimensionality reduction may improve the identification of dietary patterns, the quality of the resulting patterns ultimately depends on the quality of the underlying dietary data. Dietary assessment methods, particularly self-reported instruments such as FFQ, remain susceptible to measurement error, recall bias and misclassification. Consequently, improvements in clustering methodology alone cannot fully compensate for limitations in dietary exposure assessment. Previous methodological studies have also highlighted that measurement error can substantially influence clustering results and that clustering based on corrected estimates of usual dietary intake may better reflect true dietary patterns[71]. Therefore, future methodological developments should integrate advances in machine learning with continued improvements in dietary assessment, measurement error correction and validation of dietary data to ensure that identified dietary patterns are both statistically robust and nutritionally meaningful.

Conclusion

Over the past several decades, dietary pattern analysis has emerged as an essential tool in nutritional epidemiology, enabling the evaluation of complex dietary exposures that cannot be fully captured through single-nutrient or single-food approaches. While large-scale prospective cohort studies have provided valuable insights using PCA and CA, methodological limitations-including low clustering quality, lack of standardised evaluation metrics and minimal application of advanced dimensionality reduction techniques-persist.

Emerging machine learning methods, particularly deep learning-based approaches such as AE, hold promise for overcoming these challenges by capturing non-linear structures in dietary data and improving cluster separability. Systematic comparisons of PCA, UMAP and AE-based workflows across diverse cohorts are needed to identify the most effective and reproducible strategies.

If future studies establish the external validity of these approaches, they may provide a foundation for integrating dietary pattern analysis into digital health platforms capable of supporting personalised dietary guidance and chronic disease prevention. Such integration would not only enhance the precision and applicability of dietary pattern analysis but also strengthen the translation of epidemiological findings into impactful public health interventions.

This review has several limitations. First, the number of published studies applying machine learning-based clustering approaches in nutritional epidemiology remains limited, particularly for prospective cohort studies investigating cancer outcomes. Second, because methodological evaluation metrics and reporting standards vary considerably across studies, direct comparison of clustering performance remains challenging. Third, this review is narrative rather than systematic in nature. Although we performed structured literature searches for the epidemiological studies summarised in Tables 1 and 2, the broader discussion was intended to provide a conceptual overview rather than an exhaustive synthesis of all available evidence. Consequently, some relevant studies may not have been included. Finally, some of the methodological advances discussed in this review originate from the broader data science literature rather than nutritional epidemiology. Although these approaches offer valuable methodological insights, their applicability to dietary data requires further validation in diverse nutritional datasets.

Supplementary material. For supplementary material/s referred to in this article, please visit https://doi.org/10.1017/S0007114526108630.

Data availability statement. No new data were created or analysed in this study.

Acknowledgements. During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist with language editing. All content was reviewed and verified by the authors, who accept full responsibility for the final manuscript.

Authors Contributions. DK: Conceptualisation, Methodology, Supervision, Writing – review & editing; HL: Conceptualisation, Investigation, Methodology, Visualisation, Writing – original draft, Writing – review & editing; BS: Methodology, Supervision, Writing – review & editing; W-KS: Investigation, Writing – original draft, Writing – review & editing; DH: Investigation, Methodology, Writing – review & editing; J-YC: Investigation, Writing – review & editing; SM: Investigation, Writing – review & editing; SC: Investigation, Writing – review & editing; S-YL: Investigation, Writing – review & editing.

Funding. This research received no external funding.

Declaration of interest. The authors declare no conflict of interest.

References

1. World Health Organization (2020) Global health estimates: Deaths by Cause, Age, Sex, by Country and by Region, 2000–2019. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death (accessed January 2025).
2. Sung H, Ferlay J, Siegel RL, *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **71**, 209–249.
3. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
4. Wingrove K, Lawrence MA & McNaughton SA (2022) A systematic review of the methods used to assess and report dietary patterns. *Front Nutr* **9**, 892351.
5. Spicker D, Nazemi A, Hutchinson J, *et al.* (2025) Challenges for predictive modeling with neural network techniques using error-prone dietary intake data. *Stat Med* **44**, e70013.
6. Oliveira Chaves L, Gomes Domingos AL, Louzada Fernandes D, *et al.* (2023) Applicability of machine learning techniques in food intake assessment: a systematic review. *Crit Rev Food Sci Nutr* **63**, 902–919.
7. Zhao J, Li Z, Gao Q, *et al.* (2021) A review of statistical methods for dietary pattern analysis. *Nutr J* **20**, 37.
8. Tsolakidis D, Gymnopoulos LP & Dimitropoulos K (2024) Artificial intelligence and machine learning technologies for personalized nutrition: a review. *Informatics* **11**, 62.
9. Hutchinson JM, Raffoul A, Pepetone A, *et al.* (2025) Advances in methods for characterising dietary patterns: a scoping review. *Br J Nutr* **133**, 1–15.
10. Bailey RL (2021) Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. *Curr Opin Biotechnol* **70**, 91–96.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

11. Moshfegh AJ, Rhodes DG & Martin CL (2022) National food intake assessment: technologies to advance traditional methods. *Annu Rev Nutr* **42**, 401–422.
12. Zhao X, Xu X, Li X, *et al.* (2021) Emerging trends of technology-based dietary assessment: a perspective study. *Eur J Clin Nutr* **75**, 582–587.
13. Skinner A, Toumpakari Z, Stone C, *et al.* (2020) Future directions for integrative objective assessment of eating using wearable sensing technology. *Front Nutr* **7**, 80.
14. Neuhouser ML, Prentice RL, Tinker LF, *et al.* (2023) Enhancing capacity for food and nutrient intake assessment in population sciences research. *Annu Rev Public Health* **44**, 37–54.
15. Subar AF, Kirkpatrick SI, Mittl B, *et al.* (2012) The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. *J Acad Nutr Diet* **112**, 1134–1137.
16. Schembre SM, Liao Y, O'Connor SG, *et al.* (2018) Mobile ecological momentary diet assessment methods for behavioral research: systematic review. *JMIR mHealth uHealth* **6**, e11170.
17. Stumbo PJ (2013) New technology in dietary assessment: a review of digital methods in improving food record accuracy. *Proc Nutr Soc* **72**, 70–76.
18. Boushey CJ, Spoden M, Zhu FM, *et al.* (2017) New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. *Proc Nutr Soc* **76**, 283–294.
19. Martin CK, Correa JB, Han H, *et al.* (2012) Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. *Obesity (Silver Spring)* **20**, 891–899.
20. Sun M, Burke LE, Mao ZH, *et al.* (2014) eButton: a wearable computer for health monitoring and personal assistance. *Proc Des Autom Conf* **2014**, 1–6.
21. Doulah A, Ghosh T, Hossain D, *et al.* (2021) "Automatic ingestion monitor version 2" – a novel wearable device for automatic food intake detection and passive capture of food images. *IEEE J Biomed Health Inform* **25**, 568–576.
22. Eldridge AL, Piernas C, Illner AK, *et al.* (2018) Evaluation of new technology-based tools for dietary intake assessment-an ILSI Europe dietary intake and exposure task force evaluation. *Nutrients* **11**, 55.
23. Willett W (2012) *Overview of Nutritional Epidemiology*. Nutritional Epidemiology. New York, NY: Oxford University Press.
24. Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* **330**, 1029–1035.
25. O'Connor EA, Evans CV, Ivlev I, *et al.* (2022) Vitamin and mineral supplements for the primary prevention of cardiovascular disease and cancer: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* **327**, 2334–2347.
26. World Cancer Research Fund/American Institute for Cancer Research (2018) *Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018*. Washington, DC: American Institute for Cancer Research.
27. World Cancer Research Fund International (2025) *Dietary and Lifestyle Patterns for Cancer Prevention: evidence and Recommendations from CUP Global*. London: World Cancer Research Fund International.
28. Reedy J, Subar AF, George SM, *et al.* (2018) Extending methods in dietary patterns research. *Nutrients* **10**, 571
29. Scientific Report of the 2020 Dietary Guidelines Advisory Committee (2020) *Advisory Report to the Secretary of Agriculture and the Secretary of Health and Human Services*. Washington DC.
30. Reedy J, Krebs-Smith SM, Hammond RA, *et al.* (2017) Advancing the science of dietary patterns research to leverage a complex systems approach. *J Acad Nutr Diet* **117**, 1019–1022.
31. McNaughton SA (2020) Chapter 13 - Dietary patterns. In *Present Knowledge in Nutrition*, pp. 235–248 [Marriott BP, Birt DF, Stallings VA and Yates AA, editors]. (11th ed): London, UK: Academic Press.
32. Zhao J, Li Z, Gao Q, *et al.* (2021) A review of statistical methods for dietary pattern analysis. *Nutr J* **20**, 37.
33. Kennedy ET, Ohls J, Carlson S, *et al.* (1995) The healthy eating index: design and applications. *J Am Diet Assoc* **95**, 1103–1108.
34. Liese AD, Krebs-Smith SM, Subar AF, *et al.* (2015) The Dietary Patterns Methods Project: synthesis of findings across cohorts and relevance to dietary guidance. *J Nutr* **145**, 393–402.
35. Schulz CA, Oluwagbemigun K & Nothlings U (2021) Advances in dietary pattern analysis in nutritional epidemiology. *Eur J Nutr* **60**, 4115–4130.
36. Greenacre M, Groenen PJF, Hastie T, *et al.* (2022) Principal component analysis. *Nat Rev Methods Primers* **2**, 100.
37. Tavakol M & Wetzel A (2020) Factor analysis: a means for theory and instrument development in support of construct validity. *Int J Med Educ* **11**, 245–247.
38. Liese AD, Wambogo E, Lerman JL, *et al.* (2022) Variations in dietary patterns defined by the healthy eating index 2015 and associations with mortality: findings from the dietary patterns methods project. *J Nutr* **152**, 796–804.
39. Kaufman L & Rousseeuw P (1990) *Finding Groups in Data: An Introduction To Cluster Analysis*. New York: John Wiley & Sons
40. Lengyel A & Botta-Dukát Z (2019) Silhouette width using generalized mean – a flexible method for assessing clustering efficiency. *Ecol Evol* **9**, 13231–13243.
41. Nichols JA, Herbert Chan HW & Baker MAB (2019) Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev* **11**, 111–118.
42. Kirk D, Kok E, Tufano M, *et al.* (2022) Machine learning in nutrition research. *Adv Nutr* **13**, 2573–2589.
43. Xiang R, Wang W, Yang L, *et al.* (2021) A comparison for dimensionality reduction methods of single-cell RNA-seq data. *Front Genet* **12**, 646936.
44. Bae HJ, Park JS, Choi JH, *et al.* (2025) Learning model combined with data clustering and dimensionality reduction for short-term electricity load forecasting. *Sci Rep* **15**, 3575.
45. Seo SK, Kim G, Woo Y, *et al.* (2024) Assessment of risk levels in cut-slope using dimensionality reduction and clustering analysis. *J Korean Geotech Soc* **40**, 113–129.
46. Chunfeng Song FL, Yongzhen Huang, Liang Wang, Tieniu Tan, editor Auto-encoder Based Data Clustering, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 26th Iberoamerican Congress, CIARP 2023; 2023 2024; Coimbra, Portugal.
47. Xie J, Girshick R & Farhadi A (2015) Unsupervised Deep Embedding for Clustering Analysis. https://ui.adsabs.harvard.edu/abs/2015arXiv151106335X (accessed January 2025).
48. Zhou S, Xu H, Zheng Z, *et al.* (2022) A Comprehensive Survey on Deep Clustering: Taxonomy, Challenges, and Future Directions June 01, 2022. https://ui.adsabs.harvard.edu/abs/2022arXiv220607579Z (accessed January 2025).
49. Svirsky J & Lindenbaum O (2023) Interpretable Deep Clustering for Tabular Data2023 June 01, 2023. https://ui.adsabs.harvard.edu/abs/2023arXiv230604785S (accessed January 2025).
50. Rabbani SB, Medri I & Samad MD (2025) Deep clustering of tabular data by weighted Gaussian distribution learning. *Neurocomputing* **623**, 129359.
51. Engeset D, Dyachenko A, Ciampi A, *et al.* (2009) Dietary patterns and risk of cancer of various sites in the Norwegian European Prospective Investigation into Cancer and Nutrition cohort: the Norwegian Women and Cancer study. *Eur J Cancer Prev* **18**, 69–75.
52. Qarmiche N, El Kinany K, Otmani N, El Rhazi K, Chaoui NEH (2023) Cluster analysis of dietary patterns associated with colorectal cancer derived from a Moroccan case-control study. *BMJ Health Care Inform* **30**, e100710.
53. Ma T, Tu K, Ou Q, *et al.* (2023) Comparing the associations of dietary patterns identified through principal component analysis and cluster analysis with colorectal cancer risk: a large case-control study in China. *Nutrients* **16**, 147.
54. Edefonti V, Randi G, Decarli A, *et al.* (2009) Clustering dietary habits and the risk of breast and ovarian cancers. *Ann Oncol* **20**, 581–590.
55. Li L, Momma H, Chen H, *et al.* (2024) Dietary patterns associated with the incidence of hypertension among adult Japanese males: application of machine learning to a cohort study. *Eur J Nutr* **63**, 1293–1314.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630

56. Li J, Liu Z, Xie X, *et al.* (2024) Dietary vitamin intake and cancer risk in patients with chronic kidney disease: results from the National Health and Nutrition Examination Survey (2007–2018). *Int Urol Nephrol* **56**, 3101–3109.
57. Liu Z, Xu D, Xu WX, *et al.* (2022) Dietary patterns, nutritional status, and mortality risks among the elderly. *Front Nutr* **9**, 963060.
58. Sharma I, Roebothan B, Zhu Y, *et al.* (2018) Hypothesis and data-driven dietary patterns and colorectal Cancer survival: findings from Newfoundland and Labrador colorectal Cancer cohort. *Nutr J* **17**, 55.
59. Wirfalt E, Midthune D, Reedy J, *et al.* (2009) Associations between food patterns defined by cluster analysis and colorectal cancer incidence in the NIH-AARP diet and health study. *Eur J Clin Nutr* **63**, 707–717.
60. Brunner EJ, Mosdøl A, Witte DR, *et al.* (2008) Dietary patterns and 15-y risks of major coronary events, diabetes, and mortality. *Am J Clin Nutr* **87**, 1414–1421.
61. Krieger JP, Cabaset S, Pestoni G, *et al.* (2018) Dietary patterns are associated with cardiovascular and cancer mortality among Swiss adults in a census-linked cohort. *Nutrients* **10**, 313.
62. Kim Y, Kim YM, Shin MH, *et al.* (2022) Empirically identified dietary patterns and metabolic syndrome risk in a prospective cohort study: the cardiovascular disease association study. *Clin Nutr* **41**, 2156–2162.
63. Tan LJ, Oh SJ, Nasan-Ulzii B, *et al.* (2024) Association between dietary patterns and obesity: a longitudinal prospective cohort study. *Public Health* **234**, 217–223.
64. Olsson K, González-Padilla E, Janzi S, *et al.* (2023) Clusters of carbohydrate-rich foods and associations with type 2 diabetes incidence: a prospective cohort study. *Nutr J* **22**, 71.
65. Duffey KJ, Steffen LM, Van Horn L, *et al.* (2012) Dietary patterns matter: diet beverages and cardiometabolic risks in the longitudinal Coronary Artery Risk Development in Young Adults (CARDIA) study. *Am J Clin Nutr* **95**, 909–915.
66. Villegas R, Yang G, Gao YT, *et al.* (2010) Dietary patterns are associated with lower incidence of type 2 diabetes in middle-aged women: the Shanghai Women's Health Study. *Int J Epidemiol* **39**, 889–899.
67. Quatromoni PA, Copenhafer DL, D'Agostino RB, *et al.* (2002) Dietary patterns predict the development of overweight in women: the Framingham Nutrition Studies. *J Am Diet Assoc* **102**, 1239–1246.
68. Hsiao PY, Mitchell DC, Coffman DL, *et al.* (2013) Dietary patterns and relationship to obesity-related health outcomes and mortality in adults 75 years of age or greater. *J Nutr Health Aging* **17**, 566–572.
69. Nyholm M, Lissner L, Hörnell A, *et al.* (2013) Exploring dietary patterns, obesity and sources of bias: the Västerbotten Intervention Programme (VIP). *Public Health Nutr* **16**, 631–638.
70. Kimokoti RW, Gona P, Zhu L, *et al.* (2012) Dietary patterns of women are associated with incident abdominal obesity but not metabolic syndrome. *J Nutr* **142**, 1720–1727.
71. Su Y, Reedy J & Carroll RJ (2018) Clustering in general measurement error models. *Stat Sin* **28**, 2337–2351.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 11:58:40, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0007114526108630