



Investigating the Influence of Heavy Metals and Environmental Factors on Metabolic Syndrome Risk Based on Nutrient Intake: Machine Learning Analysis of Data from the Eighth Korea National Health and Nutrition Examination Survey (KNHANES)

Seungpil Jeong ¹ and Yean-Jung Choi ^{2,*}

- ¹ Department of Medical Informatics, College of Medicine, Catholic University of Korea, Seoul 06591, Republic of Korea; seungpil720@catholic.ac.kr
- ² Department of Food and Nutrition, Sahmyook University, Seoul 01795, Republic of Korea
- * Correspondence: yjchoi@syu.ac.kr; Tel.: +82-2-3399-1656; Fax: +82-2-3399-1655

Abstract: This study delves into the complex interrelations among nutrient intake, environmental exposures (particularly to heavy metals), and metabolic syndrome. Utilizing data from the Korea National Health and Nutrition Examination Survey (KNHANES), machine learning techniques were applied to analyze associations in a cohort of 5719 participants, categorized into four distinct nutrient intake phenotypes. Our findings reveal that different nutrient intake patterns are associated with varying levels of heavy metal exposure and metabolic health outcomes. Key findings include significant variations in metal levels (Pb, Hg, Cd, Ni) across the clusters, with certain clusters showing heightened levels of specific metals. These variations were associated with distinct metabolic health profiles, including differences in obesity, diabetes prevalence, hypertension, and cholesterol levels. Notably, Cluster 3, characterized by high-energy and nutrient-rich diets, showed the highest levels of Pb and Hg exposure and had the most concerning metabolic health indicators. Moreover, the study highlights the significant impact of lifestyle habits, such as smoking and eating out, on nutrient intake phenotypes and associated health risks. Physical activity emerged as a critical factor, with its absence linked to imbalanced nutrient intake in certain clusters. In conclusion, our research underscores the intricate connections among diet, environmental factors, and metabolic health. The findings emphasize the need for tailored health interventions and policies that consider these complex interplays, potentially informing future strategies to combat metabolic syndrome and related health issues.

Keywords: nutrient intake; environmental exposures; heavy metals; machine learning; phenotypes; metabolic syndrome

1. Introduction

Metabolic syndrome, encompassing a myriad of interrelated conditions, stands as a significant global health threat [1]. This cluster of conditions, which includes elevated blood pressure, heightened blood sugar, abdominal obesity, elevated triglyceride levels, and reduced high-density lipoprotein (HDL) cholesterol levels, presents more than just isolated health issues [2,3]. When viewed collectively, these conditions escalate the risks for severe health complications such as heart disease, stroke, and type 2 diabetes [4]. As individual health concerns, they require rigorous surveillance, but their combined presence indicates potentially dire health trajectories.

The global rise of metabolic syndrome has garnered the attention of significant health organizations, necessitating urgent interventions. The World Health Organization (WHO) and various leading health institutions have documented the escalating prevalence of this syndrome, especially in countries undergoing swift urbanization and experiencing



Citation: Jeong, S.; Choi, Y.-J. Investigating the Influence of Heavy Metals and Environmental Factors on Metabolic Syndrome Risk Based on Nutrient Intake: Machine Learning Analysis of Data from the Eighth Korea National Health and Nutrition Examination Survey (KNHANES). *Nutrients* **2024**, *16*, 724. https:// doi.org/10.3390/nu16050724

Academic Editor: Ewa Jablonska

Received: 28 January 2024 Revised: 27 February 2024 Accepted: 28 February 2024 Published: 2 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). consequential lifestyle shifts [5,6]. This is particularly concerning given the syndrome's links to chronic diseases, such as heart disease and diabetes, both of which rank highly among global mortality causes [7,8].

Despite the multifaceted origins of metabolic syndrome, ranging from genetics to age-related factors, this research focuses on a more specialized domain: the nexus among dietary patterns and external environmental influences, especially emphasizing the role of heavy metal exposure. Numerous studies have identified a significant association between heavy metals, such as cadmium, mercury, lead, and nickel, and their impacts on nutrient absorption and metabolism [9–12]. Concurrently, environmental factors, from economic dynamics to individual lifestyle choices like tobacco use and physical activity, have been documented to considerably shape nutritional behaviors and the resultant health outcomes [13–15].

The cornerstone of this investigation lies in the exhaustive data provided by the Korea National Health and Nutrition Examination Survey (KNHANES). By leveraging this vast dataset, our intent is to elucidate the multifaceted intersections among dietary habits, environmental variables, and metabolic health. What sets our study apart is our commitment to employing machine learning techniques. Our reliance on machine learning promises a nuanced understanding of dietary and environmental interplays with metabolic syndrome, potentially highlighting correlations that may be understated using conventional research methodologies [16–19].

This research represents a pivotal exploration into the intricate web of nutrition, environmental factors, and metabolic health. Our approach, combining intensive data scrutiny with state-of-the-art machine learning tools, aspires to furnish insights pivotal for shaping future health interventions and policy-making endeavors.

2. Methods

2.1. Data Source

Leveraging the vast repository of the Eighth KNHANES, the study aims to delve into the intricate nexus among dietary practices, environmental factors, and metabolic syndrome within the Korean demographic [20]. Renowned for its meticulous aggregation and diverse representation, the KNHANES offers unparalleled insights into Koreans' evolving health and nutritional paradigms. Its wealth of variables provides granular details on health metrics, nutritional nuances, and ensuing health outcomes, cutting across demographic divisions like age, socio-economic strata, and urban-rural gradients [21]. This breadth ensures an all-encompassing vantage, rendering the dataset indispensable for in-depth public health studies. Tailoring our research focus, we have selectively mined this data, emphasizing variables associated with nutrient intake, heavy metal exposure, and pertinent environmental factors. This pivot is informed by emergent studies indicating potential interconnections among dietary habits, environmental interplay, and metabolic syndrome ramifications [22,23]. Therefore, by harnessing the depth of the KNHANES dataset, we aim to illuminate the intricate interplay between dietary practices, environmental influences, and metabolic health in Korea.

2.2. Data Preprocessing

Associations were analyzed using machine learning techniques on data from the KNHANES. This analysis was conducted on a cohort of 5719 participants who were categorized into four distinct nutrient intake phenotypes (Figure 1). To enhance the fidelity of the Eighth KNHANES for our machine learning analysis, we embarked on a rigorous preprocessing journey. Recognizing the pitfalls of missing data, we employed advanced imputation techniques grounded in probabilistic frameworks to ensure coherent value replacement [24]. Outliers, which can jeopardize model accuracy, were identified and rectified using robust statistical methodologies such as the IQR method and Z-score method [25–30]. Furthermore, given the sensitivity of machine learning algorithms to feature scales, normalization processes like Min–Max scaling and Z-score normalization were utilized, ensuring



consistent interpretability and optimization across all variables [31,32]. This meticulous refinement transformed the KNHANES dataset into a precision-primed resource, poised to provide robust and insightful analytical outputs.

Figure 1. Flowchart that depicts the selection process of study participants.

2.3. Machine Learning Approach

Harnessing the power of machine learning in this analytical exploration, the study primarily relied on the K-means clustering algorithm—an unsupervised learning technique celebrated for its precision in parsing multivariate datasets [33,34]. In the expansive realm of nutrition, profiling and categorization can be daunting given the myriad nutritional variables at play. With the KNHANES dataset, this challenge was accentuated, presenting a dense matrix of 25 nutrient dimensions. However, K-means clustering's adaptability and efficiency proved invaluable, enabling us to succinctly categorize individuals into four distinct nutrient intake profiles or clusters [35,36]. Beyond mere categorization, the application of K-means on this dataset illuminated intrinsic patterns, teasing out the subtle interactions among different nutrient intake, ambient environmental factors, and the resultant metabolic health markers. This clustering exercise further underscored the premise that dietary habits, when viewed through the lens of data-driven algorithms like K-means, can shed light on broader metabolic health trajectories, thereby deepening our understanding of the factors exacerbating or mitigating metabolic syndrome risks [37,38].

2.4. Statistical Analyses

Utilizing the robust SAS software (Version 9.4, SAS Institute Inc., Cary, NC, USA), each variable from the Eighth KNHANES was systematically processed using the complex sample design data analysis method, factoring in the survey's inherent clustering, stratification variables, and weights. Initial descriptive analyses yielded statistics such as means, frequencies, and standard deviations for age and BMI, stratified by nutrient intake clusters. The chi-square test addressed categorical variables, while the analysis of covariance (ANCOVA) was deployed for continuous ones. Furthermore, to identify associations with nutrient intake clusters, we employed univariate logistic regression and expanded upon this with a multiple logistic regression analysis, concentrating on the relationships between participants' daily dietary nutrient intake and their respective nutrient intake clusters. We

maintained a rigorous quality threshold, designating a *p*-value of less than 0.05 as our benchmark for statistical significance.

3. Results

3.1. General Characteristics of Participants and Their Nutrient Intake Phenotypes

Table 1 provides a comprehensive breakdown of participants by nutrient intake phenotype, offering insights into demographic distribution, health metrics, lifestyle habits, self-reported health status, and education levels. Each cluster presents a unique profile, indicating the diversity and range of characteristics among the participants.

Out of a total of 5719 participants, the average age was 45.70 ± 16.42 years. The distribution among the different nutrient intake phenotype clusters revealed distinct patterns. Cluster 1, consisting of 1581 participants, had an average age of 49.40 ± 16.75 years. Cluster 2 (n = 1229) had an average age of 44.77 ± 17.19 years. Cluster 3 (n = 501) had the youngest participants with an average age of 41.75 ± 14.49 years, and Cluster 4 (n = 2408) had an average age of 44.56 ± 15.76 years. Regarding gender distribution, males represented 39.41% of the total participants, while females represented 60.59%. Gender distribution varied significantly among the clusters.

Metal levels, such as Pb, Hg, Cd, and Ni, were recorded for each participant. The average Pb level was $1.56 \pm 0.68 \ \mu g/dL$ for the entire cohort. Cluster-wise, Cluster 3 showed the highest Pb level ($1.82 \pm 0.29 \ \mu g/dL$), while Clusters 1, 2, and 4 showed values of 1.44 ± 0.85 , 1.46 ± 0.52 , and $1.65 \pm 0.67 \ \mu g/dL$, respectively. Similar patterns were observed for Hg, Cd, and Ni.

Health metrics revealed substantial differences among clusters. For obesity, Cluster 3 had the highest proportion of obese participants at 59.48%, whereas Cluster 2 had the lowest at 3.25%. For diabetes, Cluster 3 again had the highest prevalence at 31.54%, while Cluster 4 had the lowest at 4.44%. The distribution for high blood cholesterol, high blood pressure, high fasting blood glucose, and high blood triglyceride followed similar patterns.

Regarding lifestyle habits, there was a notable variance among clusters. For instance, Cluster 3 had the highest proportion of participants eating out daily (77.64%). Cluster 4 had the highest percentage of participants who consume dietary supplements in a year (74.79%). Smoking prevalence was highest in Cluster 4 (16.28%), while Cluster 3 had no current smokers.

Self-reported health and education: Self-reported health status varied among clusters, with Cluster 1 having the most participants reporting good health at 65.46%. On the other hand, Cluster 2 had the highest proportion of participants reporting average or poor health (80.55%). In terms of education, the majority of the participants in Cluster 1 (73.75%) had a high school or lower education.

					Nutrient Intake	Phenotypes				
Variables	Total		Cluster	1 (1)	Cluster	2 (2)	Cluster	3 (3)	Cluster	4 (4)
Vallables	(n = 571	19)	(n = 15	81)	(n = 122	29)	(n = 50)1)	(n = 240)8)
	n (%)	р	n (%)	р	n (%)	р	n (%)	р	n (%)	р
Age (yrs.) ⁽⁵⁾	45.70 ± 16.42	< 0.0001	49.40 ± 16.75	< 0.0001	44.77 ± 17.19	< 0.0001	41.75 ± 14.49	< 0.0001	44.56 ± 15.76	< 0.0001
Sex		< 0.0001		0.0528		< 0.0001		< 0.0001		< 0.0001
Male	2254 (39.41%)		752 (47.56%)		133 (10.82%)		443 (88.42%)		926 (38.46%)	
Female	3465 (60.59%)		829 (52.44%)		1096 (89.18%)		58 (11.58%)		1482 (61.54%)	
Pb ($\mu g/dL$) ⁽⁵⁾	1.56 ± 0.68	< 0.0001	1.44 ± 0.85	< 0.0001	1.46 ± 0.52	< 0.0001	1.82 ± 0.29	< 0.0001	1.65 ± 0.67	< 0.0001
Hg (μg/L) ⁽⁵⁾	3.27 ± 2.23	< 0.0001	3.16 ± 2.69	< 0.0001	2.71 ± 1.92	< 0.0001	3.77 ± 1.55	< 0.0001	3.53 ± 2.09	< 0.0001
Cd (µg/L) ⁽⁵⁾	0.87 ± 0.56	< 0.0001	0.99 ± 0.68	< 0.0001	0.84 ± 0.45	< 0.0001	0.57 ± 0.51	< 0.0001	0.86 ± 0.51	< 0.0001
Ni (μg/L) ⁽⁵⁾	0.33 ± 0.07	< 0.0001	0.32 ± 0.09	< 0.0001	0.33 ± 0.06	< 0.0001	0.30 ± 0.06	< 0.0001	0.35 ± 0.07	< 0.0001
Obese		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	1194 (20.88%)		573 (36.24%)		40 (3.25%)		298 (59.48%)		283 (11.75%)	
No	4525 (79.12%)		1008 (63.76%)		1189 (96.75%)		203 (40.52%)		2125 (88.25%)	
Diabetes		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	646 (11.30%)		276 (17.46%)		105 (8.54%)		158 (31.54%)		107 (4.44%)	
No	5073 (88.70%)		1305 (82.54%)		1124 (91.46%)		343 (68.46%)		2301 (95.56%)	
High blood cholesterol		< 0.0001		< 0.0001		< 0.0001		0.0006		< 0.0001
Yes	1715 (29.99%)		473 (29.92%)		293 (23.84%)		212 (42.32%)		737 (30.61%)	
No	4004 (70.01%)		1108 (70.08%)		936 (76.16%)		289 (57.68%)		1671 (69.39%)	
High blood pressure		< 0.0001		<0.0001		< 0.0001		0.0006		< 0.0001
Yes	1369 (23.94%)		293 (18.53%)		80 (6.51%)		212 (42.32%)		784 (32.56%)	
No	4350 (76.06%)		1288 (81.47%)		1149 (93.49%)		289 (57.68%)		1624 (67.44%)	
High Fasting blood glucose		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	2485 (43.45%)		1007 (63.69%)		388 (31.57%)		331 (66.07%)		759 (31.52%)	
No	3234 (56.55%)		574 (36.31%)		841 (68.43%)		170 (33.93%)		1649 (68.48%)	
High blood TG ⁽⁶⁾		< 0.0001		< 0.0001		< 0.0001		0.8934		< 0.0001
Yes	1672 (29.24%)		413 (26.12%)		108 (8.79%)		249 (49.70%)		902 (37.46%)	
No	4047 (70.76%)		1168 (73.88%)		1121 (91.21%)		252 (50.30%)		1506 (62.54%)	

Table 1. General characteristics of participants by nutrient intake phenoty	pes.

Table 1. Cont.

					Nutrient Intake	Phenotypes				
Variables	Total		Cluster	1 (1)	Cluster	2 (2)	Cluster	3 (3)	Cluster	4 (4)
Vallables	(n = 571	19)	(n = 15	81)	(n = 12	29)	(n = 50)1)	(n = 240)8)
	n (%)	р	n (%)	р	n (%)	р	n (%)	р	n (%)	р
Central obesity		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	1060 (18.53%)		321 (20.30%)		310 (25.22%)		61 (12.18%)		368 (15.28%)	
No	4659 (81.47%)		1260 (79.70%)		919 (74.78%)		440 (87.82%)		2040 (84.72%)	
Low HDL ⁽⁷⁾ cholesterol		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	989 (17.29%)		265 (16.76%)		87 (7.08%)		61 (12.18%)		576 (23.92%)	
No	4730 (82.71%)		1316 (83.24%)		1142 (92.92%)		440 (87.82%)		1832 (76.08%)	
Number of MetS ⁽⁸⁾ risk factors		<0.0001		<0.0001		<0.0001		<0.0001		<0.0001
0	1896 (33.15%)		233 (14.74%)		531 (43.21%)		140 (27.94%)		992 (41.20%)	
1	1375 (24.04%)		561 (35.48%)		437 (35.56%)		88 (17.56%)		289 (12.00%)	
2	1261 (22.05%)		623 (39.41%)		247 (20.10%)		54 (10.78%)		337 (14.00%)	
3	1098 (19.20%)		164 (10.37%)		14 (1.14%)		158 (31.54%)		762 (31.64%)	
4	61 (1.07%)		0 (0.00%)		0 (0.00%)		61 (12.18%)		0 (0.00%)	
5	28 (0.49%)		0 (0.00%)		0 (0.00%)		0 (0.00%)		28 (1.16%)	
Heavy drinking		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	627 (10.96%)		116 (7.34%)		95 (7.73%)		84 (16.77%)		332 (13.79%)	
No	5092 (89.04%)		1465 (92.66%)		1134 (92.27%)		417 (83.23%)		2076 (86.21%)	
Current smoking		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	653 (11.42%)		247 (15.62%)		14 (1.14%)		0 (0.00%)		392 (16.28%)	
No	5066 (88.58%)		1334 (84.38%)		1215 (98.86%)		501 (100.00%)		2016 (83.72%)	
Eating out		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
≥1 time/d	1526 (26.68%)		495 (31.31%)		374 (30.43%)		389 (77.64%)		268 (11.13%)	
≥1 time/w	3113 (54.43%)		690 (43.64%)		670 (54.52%)		54 (10.78%)		1699 (70.56%)	
<1 time/w	1080 (18.88%)		396 (25.05%)		185 (15.05%)		58 (11.58%)		441 (18.31%)	

Table 1. Cont.

					Nutrient Intake	Phenotypes				
Variables	Total		Cluster	1 (1)	Cluster	2 (2)	Cluster	3 (3)	Cluster 4	1 ⁽⁴⁾
Vallables	(n = 571	19)	(n = 158	31)	(n = 122	29)	(n = 50)1)	(n = 240)8)
	n (%)	р	n (%)	р	n (%)	р	n (%)	р	n (%)	р
Eating breakfast		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
5–7 times/w	3526 (61.65%)		1093 (69.13%)		605 (49.23%)		270 (53.89%)		1558 (64.70%)	
3–4 times/w	416 (7.27%)		193 (12.21%)		56 (4.56%)		30 (5.99%)		137 (5.69%)	
1–2 times/w	789 (13.80%)		39 (2.47%)		355 (28.89%)		67 (13.37%)		328 (13.62%)	
0 times/w	988 (17.28%)		256 (16.19%)		213 (17.33%)		134 (26.75%)		385 (15.99%)	
Diet therapy		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	1550 (27.10%)		559 (35.56%)		119 (9.68%)		188 (37.52%)		684 (28.41%)	
No	4169 (72.90%)		1022 (64.64%)		1110 (90.32%)		313 (62.48%)		1724 (71.59%)	
Eating dietary supplements in a year		< 0.0001		0.5974		< 0.0001		0.6876		< 0.0001
Yes	3331 (58.24%)		801 (50.66%)		483 (39.30%)		246 (49.10%)		1801 (74.79%)	
No	2388 (41.76%)		780 (49.34%)		746 (60.70%)		255 (50.90%)		607 (25.21%)	
Self-reported health status		< 0.0001		< 0.0001		< 0.0001		0.0049		< 0.0001
Good	2100 (36.72%)		1035 (65.46%)		239 (19.45%)		282 (56.29%)		544 (22.59%)	
Average or poor	3619 (63.28%)		546 (34.54%)		990 (80.55%)		219 (43.71%)		1864 (77.41%)	
Education level		< 0.0001		< 0.0001		< 0.0001		0.0814		< 0.0001
High school or lower	3225 (56.39%)		1166 (73.75%)		696 (56.63%)		270 (53.89%)		1093 (45.39%)	
College or higher	2494 (43.61%)		415 (26.25%)		533 (43.37%)		231 (46.11%)		1315 (54.61%)	
Household income level		< 0.0001		0.0009		< 0.0001		< 0.0001		< 0.0001
Low or mid-low	1636 (28.61%)		557 (35.23%)		494 (40.20%)		67 (13.37%)		518 (21.51%)	
Mid-high	2349 (41.07%)		457 (28.91%)		482 (39.22%)		373 (74.45%)		1037 (43.06%)	
High	1734 (30.32%)		567 (35.86%)		253 (20.59%)		61 (12.18%)		853 (35.42%)	
Economic activity		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Yes	3728 (65.19%)		1121 (70.90%)		710 (57.77%)		417 (83.23%)		1480 (61.46%)	
No	1991 (34.81%)		460 (29.10%)		519 (42.23%)		84 (16.77%)		928 (38.54%)	
Stress awareness		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001
Low	4430 (77.46%)		1399 (88.49%)		865 (70.38%)		434 (86.63%)		1732 (71.93%)	
High	1289 (22.54%)		182 (11.51%)		364 (29.62%)		67 (13.37%)		676 (28.07%)	

Table 1. Cont.

					Nutrient Intake	Phenotypes					
Variables	Total		Cluster	1 (1)	Cluster	2 (2)	Cluster	3 (3)	Cluster	4 (4)	
variables	(n = 571	19)	(n = 158	81)	(n = 12)	29)	(n = 50	01)	(n = 2408)		
	n (%)	р	n (%)	р	n (%)	р	n (%)	р	n (%)	р	
Walking		< 0.0001		< 0.0001		< 0.0001		< 0.0001		0.0035	
<5 days	2085 (36.46%)		325 (20.56%)		768 (62.49%)		61 (12.18%)		931 (38.66%)		
≥5 days	3634 (63.54%)		1256 (79.44%)		461 (37.51%)		440 (87.82%)		1477 (61.34%)		
Moderate intensity physical activity		< 0.0001		<0.0001		<0.0001		<0.0001		<0.0001	
Yes	1628 (28.47%)		497 (31.44%)		131 (10.66%)		67 (13.37%)		933 (38.75%)		
No	4091 (71.53%)		1084 (68.56%)		1098 (89.34%)		434 (86.63%)		1475 (61.25%)		
Physical activity		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001	
<5 days, 30 min	2337 (40.86%)		445 (28.15%)		830 (67.53%)		61 (12.18%)		1001 (41.57%)		
\geq 5 days, 30 min	3382 (59.14%)		1136 (71.85%)		399 (32.47%)		440 (87.82%)		1407 (58.43%)		

(1) Cluster 1: Diverse Nutrient Intake; (2) Cluster 2: Nutrient-Deficient; (3) Cluster 3: High-Energy & Nutrient-Rich; (4) Cluster 4: Balanced Intake; (5) Data were expressed as mean ± SD; (6) TG: triglyceride; (7) HDL: high density lipoprotein; (8) MetS: metabolic syndrome.

3.2. Characteristics of Heavy Metals and Five Metabolic Syndrome Factors According to Nutrient Intake Level

Diving deep into the multifaceted relationship between heavy metal exposure and metabolic syndrome indicators, our analysis discerned distinct patterns across four specific clusters, each uniquely portraying the interplay between environmental heavy metal exposures and metabolic health outcomes (Table 1 and Figure 2).



Figure 2. Key results related to clusters. (**A**) Normalized heavy metal concentrations across clusters. (**B**) Average metabolic syndrome-related indicators across clusters.

Cluster 1 showcased a moderate inclination towards Pb (Lead) exposure, while Hg (Mercury) levels stood out distinctly, overshadowed only by Cluster 3. Interestingly, a spike in cadmium levels was observed here. On the metabolic front, Cluster 1 embodied a moderate hypertension (BP11) level and elevated fasting blood sugar (FBS1)—the latter being second only to Cluster 3. Interestingly, the waist circumference (WAIST1) readings in this cluster hinted at reduced concerns of central obesity as compared to most of its counterparts.

While Cd (Cadmium) exposure echoed patterns seen in Cluster 1, Cluster 2 was characterized by a notable dip in Hg exposure—the lowest among all clusters. Metabolically, this cluster presented a dichotomy: while it showcased the lowest hypertension levels, suggesting better cardiovascular health, it also showed the highest waist circumference and lowest HDL cholesterol (HDL1) levels, implying potential obesity and cholesterol-related challenges.

Alarming insights emerged from Cluster 3. Both Pb and Hg exposures hit peak levels, making this the most exposed cluster. Its metabolic readings resonated with this heightened exposure, revealing peak hypertension and fasting blood sugar levels, flagging substantial cardiovascular and potential diabetes-related risks. In a stark juxtaposition, however, the waist circumference was the lowest.

Cluster 4 reported elevated Pb levels, trailing just behind Cluster 3. Remarkably, this cluster stood out with the highest Ni (Nickel) exposure. In terms of metabolic health, the data painted a mixed picture: while elevated hypertension level readings were a cause for concern, it also showed the peak high-density cholesterol levels, suggesting a relatively healthier cholesterol profile.

3.3. Nutrient Intake of Participants across Four Distinct Clusters

In Table 2, each cluster presents distinct dietary patterns, from a diverse nutrient intake to potential deficiencies, to very high-energy consumptions, and finally to a balanced dietary profile.

Individuals in Cluster 1 generally had a broad range of nutrient intakes. They had a significantly higher consumption of energy at 2281.31 kcal, which was above the total average of 1942.76 kcal. Notably, their intake of water, carbohydrates, protein, and fats were all higher than the average, suggesting a possible higher total food consumption or choices dense in these nutrients. The elevated intake of micronutrients like calcium, vitamin A, and vitamin C indicates a potential varied diet.

Cluster 2 showcased the lowest intake across most nutrients, suggesting a potential deficiency or a lower overall food consumption. Their average energy intake was considerably lower at 1235.07 kcal. Micronutrients like calcium, vitamin A, and vitamin C were also consumed in limited amounts. This cluster might be at a nutritional risk and may benefit from interventions to enhance their nutrient intake.

Individuals in Cluster 3 had a remarkable energy consumption, with an average intake of 3165.82 kcal—the highest among all clusters. Their nutrient profile also stands out, with a high intake of proteins, fats, and carbohydrates. The group seemed to prefer diets high in energy, with nutrients like sodium, potassium, and vitamins (like vitamin A and C) being consumed in large quantities.

The largest cluster, Cluster 4, exhibited what can be described as a balanced nutrient intake. Their values generally hovered around the overall averages, suggesting a balanced and potentially healthier diet. While their energy intake was below the overall average, they seemed to have a consistent intake of other nutrients, not veering too high or too low.

3.4. Cluster Characteristics across Nutrient Intake Levels

Table 2 provides a breakdown of the average nutrient consumption across four distinct groups of individuals. Under the category of energy and core macronutrients, the table details the intake of energy, water, and basic macronutrients like carbohydrates, proteins, and fats. Specifically, the average caloric intake is represented by the Energy value, where Cluster 3 had the highest consumption, and Cluster 2 the lowest. Water intake, indicating hydration from the diet, also shows that Cluster 3 consumed the most, while Cluster 2 consumed the least. As for carbohydrates, proteins, and fats, Cluster 3 led in consumption, pointing to a high-energy diet.

When considering the types of fat, the table categorizes them into saturated, monounsaturated, and polyunsaturated fatty acids, as well as *n*-3 and *n*-6 fatty acids. Across all fat types, Cluster 3 consumed the most, while Cluster 1 consumed the least. The essential fatty acids, *n*-3 and *n*-6, which are vital for heart health and reducing inflammation, were most consumed by Cluster 3 and least by Cluster 2.

Table 2 further highlights other dietary components like cholesterol and dietary fiber. Cluster 1 had the highest cholesterol intake, closely followed by Cluster 3, which can be a potential health concern. On the other hand, dietary fiber, essential for digestion and often found in whole grains, fruits, and vegetables, was most consumed by Cluster 3. Sugar intake was highest in Cluster 1.

Lastly, under vitamins and minerals, the table lists several essential minerals and vitamins. Except for calcium, Cluster 3 consumed the most of these minerals, with Cluster 1 leading in calcium intake. For the vitamins, Cluster 3 also consumed the most, with the exception of Vitamin C and retinol, where Cluster 1 led.

					Nutrient Intake Ph	enotypes				
Variables -	Total		Cluster 1 ⁽¹⁾		Cluster 2 ⁽²)	Cluster 3 ⁽³)	Cluster 4 ⁽⁴⁾)
variables	(n = 5719)		(n = 1581)		(n = 1229)		(n = 501)		(n = 2408)	
-	$\mathbf{Mean} \pm \mathbf{SD}$	р	$\mathbf{Mean} \pm \mathbf{SD}$	р	$\mathbf{Mean} \pm \mathbf{SD}$	р	$\mathbf{Mean} \pm \mathbf{SD}$	р	$\mathbf{Mean} \pm \mathbf{SD}$	р
Energy Water	$\begin{array}{c} 1942.76 \pm 635.85 \\ 1024.28 \pm 425.29 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 2281.31 \pm 362.45 \\ 1318.92 \pm 352.61 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1235.07 \pm 330.48 \\ 633.97 \pm 287.14 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 3165.82 \pm 421.43 \\ 1523.31 \pm 258.77 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1827.21 \pm 354.82 \\ 926.22 \pm 309.54 \end{array}$	<0.0001 <0.0001
Carbohydrate Protein	$\begin{array}{c} 293.73 \pm 98.81 \\ 73.14 \pm 29.63 \end{array}$	<0.0001 <0.0001	$342.17 \pm 87.59 \\ 85.89 \pm 13.58$	<0.0001 <0.0001	$\begin{array}{c} 207.24 \pm 65.99 \\ 43.36 \pm 14.51 \end{array}$	<0.0001 <0.0001	$389.02 \pm 100.92 \\ 139.21 \pm 29.91$	<0.0001 <0.0001	$\begin{array}{c} 286.24 \pm 80.70 \\ 66.21 \pm 12.46 \end{array}$	<0.0001 <0.0001
Fat SFA	$\begin{array}{c} 47.97 \pm 25.71 \\ 15.19 \pm 8.73 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 63.33 \pm 17.52 \\ 20.90 \pm 7.97 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 23.94 \pm 10.94 \\ 7.73 \pm 5.42 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 94.22 \pm 24.50 \\ 25.00 \pm 9.84 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 40.53 \pm 15.26 \\ 13.21 \pm 5.51 \end{array}$	<0.0001 <0.0001
MUFA PUFA	$\begin{array}{c} 15.23 \pm 9.00 \\ 12.82 \pm 8.61 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 19.35 \pm 5.93 \\ 16.34 \pm 5.83 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 7.87 \pm 4.23 \\ 6.00 \pm 2.60 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 34.01 \pm 8.12 \\ 28.00 \pm 14.41 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 12.38 \pm 5.13 \\ 10.82 \pm 4.82 \end{array}$	<0.0001 <0.0001
N3 N6	$\begin{array}{c} 2.00 \pm 1.56 \\ 10.78 \pm 7.67 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 2.45 \pm 1.86 \\ 13.82 \pm 5.31 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1.07 \pm 0.64 \\ 4.88 \pm 2.31 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 3.76 \pm 2.03 \\ 24.30 \pm 12.88 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1.81 \pm 1.08 \\ 8.99 \pm 4.34 \end{array}$	<0.0001 <0.0001
Cholesterol Fiber	$\begin{array}{c} 257.30 \pm 197.97 \\ 26.13 \pm 10.87 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 357.41 \pm 236.22 \\ 33.71 \pm 9.20 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 120.44 \pm 87.06 \\ 15.33 \pm 4.73 \end{array}$	<0.0001 <0.0001	$356.83 \pm 121.84 \\ 43.26 \pm 9.20$	<0.0001 <0.0001	$\begin{array}{c} 240.72 \pm 176.18 \\ 23.10 \pm 5.33 \end{array}$	<0.0001 <0.0001
Sugar Calcium	$\begin{array}{c} 63.20 \pm 36.52 \\ 558.13 \pm 267.47 \end{array}$	<0.0001 <0.0001	$\begin{array}{r} 85.82 \pm 40.13 \\ 833.62 \pm 241.88 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 40.01 \pm 24.30 \\ 297.38 \pm 151.46 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 84.06 \pm 37.56 \\ 681.87 \pm 153.30 \end{array}$	<0.0001 <0.0001	$55.85 \pm 27.60 \\ 484.59 \pm 147.62$	<0.0001 <0.0001
Phosphate Iron	$\begin{array}{c} 1121.26 \pm 399.43 \\ 11.85 \pm 4.29 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1426.54 \pm 194.93 \\ 14.29 \pm 3.47 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 652.43 \pm 167.82 \\ 6.88 \pm 2.69 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1843.68 \pm 371.05 \\ 17.27 \pm 2.84 \end{array}$	<0.0001 <0.0001	$\begin{array}{c} 1009.82 \pm 128.86 \\ 11.66 \pm 2.79 \end{array}$	<0.0001 <0.0001
Sodium Potassium	3014.86 ± 1237.48 2992.26 ± 1120.88	<0.0001 <0.0001	3696.46 ± 1144.95 3826.90 ± 913.52	<0.0001 <0.0001	$\begin{array}{c} 2091.83 \pm 1133.90 \\ 1778.61 \pm 481.10 \end{array}$	<0.0001 <0.0001	$\begin{array}{r} 4232.85 \pm 923.76 \\ 4501.70 \pm 1241.59 \end{array}$	<0.0001 <0.0001	$\begin{array}{r} 2785.04 \pm 917.79 \\ 2749.64 \pm 505.69 \end{array}$	<0.0001 <0.0001
Vitamin A Carotene	389.98 ± 214.61 3135.18 ± 1994.79	<0.0001 <0.0001	508.49 ± 193.15 3594.08 ± 1749.21	<0.0001 <0.0001	205.77 ± 119.16 1704.95 ± 1180.96	<0.0001 <0.0001	633.34 ± 230.26 5729.98 ± 2428.39	<0.0001	355.54 ± 163.11 3023.99 ± 1706.27	<0.0001 <0.0001
Retinol Vitamin B1	127.94 ± 139.69 1 35 ± 0 70	<0.0001	208.98 ± 206.79 1 58 ± 0.47	<0.0001	63.69 ± 51.62 0.77 ± 0.34	<0.0001	155.85 ± 98.71 2 45 + 1 22	<0.0001	101.70 ± 87.37 1.28 ± 0.42	<0.0001
Vitamin B2 Niacin	1.59 ± 0.68 14.59 ± 7.01	<0.0001	2.17 ± 0.42 16 59 \pm 5 33	<0.0001	0.82 ± 0.25 8.61 ± 3.84	<0.0001	2.72 ± 0.28 29.21 + 6.70	<0.0001	1.20 ± 0.12 1.36 ± 0.31 13.28 ± 3.78	<0.0001
Vitamin C	74.75 ± 69.55	<0.0001	10.09 ± 0.00 113.04 ± 78.15	<0.0001	37.36 ± 22.52	<0.0001	110.96 ± 71.90	<0.0001	61.16 ± 63.35	<0.0001

Table 2. Nutrient intake of participants by phenotypes.

⁽¹⁾ Cluster 1: Diverse Nutrient Intake; ⁽²⁾ Cluster 2: Nutrient-Deficient; ⁽³⁾ Cluster 3: High-Energy & Nutrient-Rich; ⁽⁴⁾ Cluster 4: Balanced Intake.

3.5. The Relationship between Various Nutrient Intake Phenotypes and Associated Risk Factors

A distinct association was observed between nutrient intake phenotypes and the levels of various heavy metals (Table 3). The concentration of Pb (μ g/dL) showed a significantly high odds ratio in Cluster 3 and 4, with Cluster 3 having an astounding OR of 47.081 (95% CI: 23.867–92.874). This suggests that individuals in the "High-Energy & Nutrient-Rich" cluster may have had higher exposure or retention of lead. Likewise, Hg (μ g/L) levels were significantly elevated across all clusters, with the highest being in Cluster 3 at OR: 1.755 (95% CI: 1.567–1.964). Interestingly, Ni concentrations revealed an OR as low as 0.034 in Cluster 4, pointing to a decreased risk.

Heavy drinking and current smoking patterns were predominantly significant risk factors across all nutrient intake phenotypes. For heavy drinking, individuals in the "High-Energy & Nutrient-Rich" cluster (Cluster 3) displayed the least association, with an OR of 0.786. Conversely, individuals in the same cluster who ate out once a day or more were twice as likely, indicating a positive relationship (OR of 2.085). Furthermore, skipping breakfast appeared to be a strong determinant in the "Diverse Nutrient Intake" cluster (Cluster 1), showing a staggering OR of 19.105 for eating breakfast 3–4 times a week.

Household income levels and education seemed to influence nutrient phenotypes. Specifically, those with a college or higher education showed an OR of 0.528 in total, emphasizing the reduced risk among the educated. In terms of economic activity, individuals in the Nutrient-Deficient cluster were over three times more likely to be economically active than the reference group (OR of 2.863).

Physical activity played a pivotal role in shaping nutrient intake. Moderate intensity physical activity showcased protective effects, especially in Cluster 2 and Cluster 4. Walking less than 5 days a week was linked to an increased risk in Cluster 4 (OR: 1.531, 95% CI: 1.317–1.780), emphasizing the importance of regular physical activity for balanced nutrient intake.

									Nutr	ient Inta	ake Phen	otypes								
¥7		Т	otal			Clus	ter 1 ⁽¹⁾			Clus	ter 2 ⁽²⁾			Clus	ter 3 ⁽³⁾			Clus	ter 4 ⁽⁴⁾	
variables		(n =	= 5719)			(n =	: 1581)			(n =	= 1229)			(n =	= 501)			(n =	2408)	
	0	R (95% (CI)	р	O	R (95% C	CI)	р	0	R (95% (CI)	р	0	R (95% C	CI)	р	0	R (95% C	CI)	p
Pb (µg/dL)	3.388	3.140	3.657	< 0.0001	2.053	1.826	2.307	< 0.0001	3.171	2.565	3.919	< 0.0001	47.081	23.867	92.874	< 0.0001	6.047	5.294	6.907	< 0.0001
Hg (µg/L)	1.23	1.203	1.257	< 0.0001	1.144	1.104	1.185	< 0.0001	0.867	0.819	0.918	< 0.0001	1.755	1.567	1.964	< 0.0001	1.464	1.407	1.523	< 0.0001
Cd (µg/L)	1.03	0.948	1.119	0.4875	2.149	1.867	2.474	< 0.0001	1.388	1.099	1.752	0.0059	1.705	1.249	2.327	0.0008	0.56	0.483	0.650	< 0.0001
Ni (µg/L)	0.087	0.046	0.165	< 0.0001	0.734	0.263	2.048	0.5546		-		< 0.0001		-		< 0.0001	0.034	0.012	0.100	< 0.0001
Heavy drinking				< 0.0001				< 0.0001				< 0.0001				0.2614				< 0.0001
Yes	2.680	2.306	3.116		2.307	1.618	3.288		2.771	1.878	4.087		0.786	0.516	1.196		3.673	2.934	4.597	
No		1				1				1				1				1		
Current smoking				< 0.0001				0.0939												< 0.0001
Yes	2.367	2.043	2.742		1.238	0.964	1.590			-				-			2.712	2.213	3.325	
No		1				1				1				1				1		
Eating out				< 0.0001				< 0.0001				< 0.0001				0.3581				< 0.0001
$\geq 1 \text{ time/d}$	0.349	0.303	0.402		0.377	0.292	0.486		0.068	0.046	0.101		2.085	1.263	3.443		0.12	0.089	0.162	
≥ 1 time/w	0.393	0.346	0.445		0.175	0.137	0.224		1.312	0.968	1.777		1.859	0.953	3.626		0.213	0.172	0.262	
<1 time/w		1				1				1				1				1		
Eating breakfast				0.0354				< 0.0001				< 0.0001				0.9332				< 0.0001
5–7 times/w		1				1				1				1				1		
3–4 times/w	1.151	0.960	1.380		2.73	2.034	3.664		19.105	9.568	38.15		0.255	0.128	0.508		0.532	0.385	0.737	
1–2 times/w	0.329	0.284	0.380		0.586	0.325	1.058		0.39	0.303	0.502			-			0.446	0.356	0.558	
0 times/w	0.625	0.550	0.710		19.905	14.52	27.288		0.079	0.053	0.117		0.461	0.317	0.671		0.37	0.298	0.459	
Diet therapy				0.0472				< 0.0001				< 0.0001				0.7242				0.0097
Yes	1.112	1.001	1.234		0.415	0.342	0.503		3.589	2.515	5.123		1.06	0.767	1.465		1.238	1.053	1.456	
No		1				1				1				1				1		
Eating dietary supplements in a year				< 0.0001				< 0.0001				0.0247				< 0.0001				< 0.0001
Yes	1.486	1.351	1.634		0.266	0.219	0.323		0.782	0.632	0.969		3.684	2.647	5.127		3.921	3.254	4.724	
No		1				1				1				1				1		
Self-reported health status																				
Good		1		0.7804		1		< 0.0001		1		0.3500		1				1		0.0008
Average or poor	1.014	0.921	1.117		0.502	0.414	0.609		0.883	0.679	1.147			-			1.354	1.135	1.616	

Table 3. Odds ratio of nutrient intake phenotypes and risk factors of participants.

Table 3. Cont.

									Nuti	ient Inta	ake Phen	otypes								
Variables		Т	otal			Clus	ter 1 ⁽¹⁾			Clus	ster 2 ⁽²⁾			Clus	ster 3 ⁽³⁾			Clus	ter 4 ⁽⁴⁾	
variables		(n =	= 5719)			(n =	= 1581)			(n =	= 1229)			(n	= 501)			(n =	: 2408)	
	0	R (95% (CI)	р	0	R (95% (CI)	р	0	R (95% (CI)	р	0	R (95% (CI)	р	0	R (95% (CI)	р
Education level				< 0.0001				< 0.0001				< 0.0001				0.0038				< 0.0001
High school or lower		1				1				1				1				1		
College or higher	0.528	0.48	0.581		0.294	0.237	0.364		0.118	0.092	0.151		1.596	1.164	2.190		0.619	0.534	0.717	
Household income level								< 0.0001				< 0.0001								< 0.0001
Low or mid-low		1		< 0.0001		1				1				1				1		
Mid-high	0.447	0.399	0.502		0.469	0.371	0.592		0.048	0.035	0.065			-			0.380	0.312	0.464	
High	0.992	0.879	1.119		0.172	0.137	0.218		0.094	0.068	0.131			-			1.776	1.454	2.170	
Economic activity				0.1310				< 0.0001				< 0.0001				0.2614				< 0.0001
Yes	1.078	0.978	1.189		3.312	2.681	4.092		2.863	2.305	3.556		0.786	0.516	1.196		0.712	0.612	0.828	
No		1				1				1				1				1		
Stress awareness				0.0274				0.6566				< 0.0001								< 0.0001
Low		1				1				1				1				1		
High	1.133	1.014	1.266		1.066	0.803	1.416		0.520	0.411	0.657			-			2.847	2.409	3.365	
Walking				0.0225				< 0.0001				< 0.0001								< 0.0001
<5 days	0.894	0.811	0.984		0.222	0.175	0.282		1.777	1.427	2.212			-			1.531	1.317	1.780	
≥5 days		1				1				1				1				1		
Moderate intensity physical activity				0.0092				< 0.0001				< 0.0001								< 0.0001
Yes	0.872	0.786	0.967		2.144	1.756	2.619		0.295	0.201	0.433			-			0.736	0.633	0.856	
No		1				1				1				1				1		
Physical activity				0.0214				< 0.0001				< 0.0001								< 0.0001
<5 days, 30 min		1				1				1				1				1		
\geq 5 days, 30 min	1.118	1.017	1.229		3.813	3.083	4.716		0.433	0.343	0.545			-			0.681	0.587	0.790	

⁽¹⁾ Cluster 1: Diverse Nutrient Intake; ⁽²⁾ Cluster 2: Nutrient-Deficient; ⁽³⁾ Cluster 3: High-Energy & Nutrient-Rich; ⁽⁴⁾ Cluster 4: Balanced Intake.

4. Discussion

4.1. Nutrient Level Cluster Analysis

This endeavor to demystify nutrient intake and its ramifications on health outcomes required a nuanced approach, given the intricacy of the data spanning 25 diverse nutrient categories. This led us to harness the powers of unsupervised machine learning, precisely, the K-means clustering algorithm [39], which adeptly categorized the data into four discernible nutritional profiles (Figure 3 and Supplementary Figures S1 and S2). Cluster 1 radiated nutritional affluence, endorsing an eclectic array of nutrients. This cohort manifested enhanced levels of energy, fat, saturated fatty acids, cholesterol, dietary fiber, calcium, Vitamin A, and Vitamin C, evocative of a diet dense in fruits, vegetables, dairy, and meats. The salutary effects of such a diverse nutrient palette were echoed by English et al. [40], elucidating the multitude of health boons accompanying such nutritional heterogeneity. Cluster 2 projected nutritional austerity, with conspicuous deficiencies spanning almost all nutrient spectrums. The alarming paucity of energy, fats, carbohydrates, proteins, and an array of vitamins and minerals implies potential malnourishment or a diet predominantly reliant on nutritionally sterile foods. A sustained adherence to such a diet paves the way for a gamut of health pitfalls, a sentiment mirrored by Wang et al. [41]. In Cluster 3, the nutrition profile exudes abundance. Energy, carbohydrates, proteins, varied fatty acids, and a slew of vitamins and minerals flourish in this cluster. While the richness in nutrients promises a slew of health merits, the surging energy content casts shadows of escalating caloric intake. Such surpluses, if not counterbalanced with commensurate physical exertion, can brew health adversities, a notion championed by James Stubbs et al. [42]. Emanating nutritional equilibrium, Cluster 4 shows a balanced act across all nutrients. Unlike its counterparts that exhibited stark nutrient polarities, this group harmoniously aligns moderate nutrient intakes, potentially hinting at a health-optimized diet. The secret lies in striking a balance between portion sizes and ensuring a spectrum of dietary ingredients. Miller et al. [43] accentuate the myriad health dividends sprouting from such balanced indulgences. In essence, these clusters accentuate the complex situation of nutritional intake and underscore the paramountcy of dietary choices in sculpting health destinies.



Figure 3. Cluster visualization. This figure showcases the results using the K-means clustering algorithm, which adeptly categorized the data into four discernible nutritional profiles.

4.2. Analysis of Risk Factors Related to Metabolic Syndrome Using Machine Learning

This deep dive into the nexus among heavy metal exposures, environmental variables, and their repercussions on metabolic syndrome crystallized the data into four distinct nutrient intake clusters. The PyCaret 3.1.0 package, epitomizing automated machine learning, streamlined the methodological strides in this study. Among the vast spectrum of 14 classifiers, we gravitated towards five due to their stellar performance (Figure 4 and Supplementary Figure S3). Our choice found its resonance in the seminal work by Kelleher et al. [16], which advocates for uncompromising accuracy and reliability in machine learning tools, particularly when broaching critical health dimensions. Housing participants with a rich and diverse nutrient intake, Cluster 1 registered elevated levels of Cadmium (Cd) and Mercury (Hg). Mehouel and Fowler [44] had earlier walked this path, illuminating the potential dietary origins of such metal exposures. This underscores that while nutrient diversity champions health, it might unwittingly invite heavy metals, thereby mingling benefits with risks. Hallmarked by subpar nutrient intake, Cluster 2 interestingly flagged frequent dining out as a significant behavioral vector. This propensity towards external food sources possibly spells compromised food choices, echoing Chen et al. [45], who espoused the link between certain dining behaviors and nutrient paucity. The recurrent consumption of outside food might inadvertently affect the nutrient equilibrium. The participants in Cluster 3 enjoy energy-laden yet nutrient-rich diets. Their intriguing association with Nickel (Ni) might suggest their exposure to specific foods or environments [46,47]. Aslam et al. [48] have shed light on this interplay, where calorie-rich foods can sometimes be clandestine carriers of heavy metals. This underscores the imperativeness of scrupulous dietary choices even within the realms of nutrient-dense diets. However, the exemplar of dietary equilibrium, Cluster 4, displayed conspicuous Lead (Pb) and Mercury (Hg) levels. Noger-Huet et al. [49] extolled balanced diets without risking potential metal exposures. This observation pivots attention towards the roots of so-called "balanced" diets, raising the question as to whether they inadvertently double as conduits for heavy metal infiltration. These revelations reinforce those diets while being health compasses, and are also cryptic chroniclers of environmental and lifestyle narratives. Each bite not only nourishes but also narrates tales of its origin, travel, and exposures. This encourages a 360-degree purview of diets, embracing their advantages and disadvantages alike. Furthermore, our mathematical modeling, elucidating the significance of classifier outcomes, unraveled a complex set of circumstances binding health, lifestyle markers, and metabolic syndrome. It accentuates the imperativeness of a panoramic view, incorporating both prevention and intervention, in our health journeys. This investigation, in essence, is a clarion call, beckoning a deeper dive into the intriguing interplay of diet, environment, and metabolic health.

Cluster 1	Diverse Nutrient Intake	LGBM	RF	ET	DT		GBC	SUM	Cluster 2	Nutrient-Deficient	LGBM	RF	ET	D		GBC	SUM
	1 HM_Cd	0.68	1.0) (1.00	1.00	1.00	4.68		1 dining_out_frequency	0.65	(0.93	1.00	0.79	1.00	4.37
	2 HM_Hg	0.76	0.6	2 (0.38	0.51	0.46	2.74		2 house_income	0.59		1.00	0.78	1.00	0.82	4.18
	3 HM_Pb	0.80	0.5	7 (0.37	0.35	0.33	2.43		3 economic_activity	0.53	(0.61	0.74	0.50	0.73	3.11
	4 HM_Ni	0.53	0.7		0.62	0.24	0.30	2.40		4 education_level	1.00	(0.57	0.72	0.13	0.17	2.58
	5 education_level	1.00	0.2	9 1	0.44	0.00	0.25	1.98		5 HM_Ni	0.24	(0.64	0.44	0.00	0.63	1.95
	6 house_income	0.88	0.1	9 1	0.34	0.19	0.01	1.60		6 HM_Pb	0.41	(0.52	0.35	0.35	0.09	1.72
	7 breakfast_frequency	0.27	0.2	5 (0.30	0.55	0.02	1.40		7 HM_Cd	0.18	(0.65	0.48	0.00	0.14	1.45
	8 economic_activity	0.58	0.0	9 1	0.29	0.00	0.00	0.96		8 walking_day	0.07	(0.29	0.33	0.26	0.27	1.22
	9 walking_day	0.20	0.2	2 (0.35	0.00	0.11	0.89		9 HM_Hg	0.23	(0.32	0.29	0.10	0.11	1.06
	10 dining_out_frequency	0.13	0.2	7 (0.43	0.00	0.03	0.86		10 breakfast_frequency	0.16	(0.24	0.40	0.00	0.02	0.83
	11 physical_activity	0.14	0.2) (0.39	0.00	0.04	0.77		11 dietary_supple	0.14	(0.05	0.26	0.00	0.01	0.46
	12 health_awareness	0.20	0.1	5 (0.22	0.00	0.00	0.57		12 health_awareness	0.06	0	0.11	0.21	0.00	0.04	0.44
	13 dietary_supple	0.18	0.1	3 (0.21	0.00	0.00	0.53		13 drink_high	0.15	(0.08	0.17	0.00	0.02	0.42
	14 dietary_therapy	0.05	0.1	3 (0.24	0.00	0.11	0.53		14 walking_rate	0.15	(0.04	0.16	0.00	0.06	0.42
	15 walking_rate	0.01	0.1) (0.25	0.00	0.01	0.36		15 stress_perception	0.13	(0.06	0.15	0.00	0.00	0.33
	16 smoking	0.00	0.0	3 (0.12	0.00	0.00	0.14		16 dietary_therapy	0.03	(0.06	0.13	0.00	0.07	0.29
	17 stress_perception	0.01	0.0	3 1	0.09	0.00	0.00	0.13		17 physical_activity	0.03	(0.03	0.08	0.00	0.00	0.14
	18 drink_high	0.00	0.0	1	0.09	0.00	0.00	0.10		18 smoking	0.00	(0.04	0.04	0.00	0.02	0.10
	19 residence	0.00	0.0) (0.00	0.00	0.00	0.00		19 residence	0.00	(0.00	0.00	0.00	0.00	0.00
Classian 2	High-Energy & Nutrient-Rich	LGBM	RE	FT	DT		GBC	SUM		Balanced Intake	LGBM	RF	FT	D	r	GBC	SUM
Cluster 3		0.21	0.8		148	1.00	1.00	3 5 2	Cluster 4	1 HM Pb	0.72		1.00	1.00	1.00	1.00	472
	2 dietary supple	1.00	0.5	,	1.00	0.00	0.00	2.57		2 HM Hg	0.65		0.89	0.57	0.76	0.81	3.68
	2 health awareness	0.10	0.7		1.00	0.00	0.62	2.20		3 dining out frequency	0.57		0.50	0.61	0.96	0.41	3.05
	A education level	0.15	0.7		0.70	0.00	0.02	2.30		4 HM Cd	0.55		0.56	0.55	0.51	0.32	2.50
	F dising out fragueses	0.55	1.0		0.40	0.91	0.03	2.25		5 house income	0.68		0.43	0.66	0.00	0.23	2.00
	S dining_out_irequency	0.30	1.0	<u> </u>	0.00	0.00	0.01	2.23		6 HM Ni	0.29		0.46	0.51	0.38	0.04	1.68
	6 HIVI_Hg	0.20	0.4		0.52	0.00	0.87	2.04		7 education level	0.87		0.24	0.48	0.00	0.06	1.65
	7 HIM_PD	0.20	0.0	1	0.55	0.00	0.55	1.91		8 economic activity	1.00	E 1	0.11	0.30	0.00	0.02	1.43
	8 breakfast_frequency	0.33	0.5		J.08	0.00	0.08	1.04		9 stress_perception	0.16	i 1	0.21	0.50	0.38	0.15	1.40
	9 house_income	0.64	0.20) (0.29	0.44	0.05	1.62		10 health_awareness	0.07		0.20	0.42	0.51	0.11	1.32
	10 walking day	0.02	0.2) (0.65	0.00	0.49	1.44		11 breakfast_frequency	0.33		0.20	0.35	0.25	0.01	1.13
	To walking_day	0.02						4 3 3								0.19	1.00
	11 dietary_therapy	0.03	0.24	4 (0.30	0.57	0.17	1.32		12 dietary_therapy	0.24	1	0.20	0.38	0.00	0.15	0.00
	11 dietary_therapy 12 HM_Cd	0.03	0.24	1 (2 (0.30 0.57	0.57 0.00	0.17	1.32		12 dietary_therapy 13 walking_day	0.24		0.20 0.28	0.38 0.38	0.00	0.01	0.89
	11 dietary_therapy 12 HM_Cd 13 walking_rate	0.02 0.03 0.01 0.00	0.24	1 (2 (1 (0.30 0.57 0.20	0.57 0.00 0.00	0.17 0.00 0.10	1.32		12 dietary_therapy 13 walking_day 14 dietary_supple	0.24		0.20 0.28 0.11	0.38 0.38 0.34	0.00 0.00 0.00	0.01	0.89
	11 dietary_therapy 12 HM_Cd 13 walking_rate 14 drink_high	0.02 0.03 0.01 0.00 0.00	0.24	4 (2 (1 (5 (0.30 0.57 0.20 0.17	0.57 0.00 0.00 0.00	0.17 0.00 0.10 0.00	1.32 1.00 0.51 0.33		12 dietary_therapy 13 walking_day 14 dietary_supple 15 physical_activity	0.24 0.22 0.33 0.13		0.20 0.28 0.11 0.08	0.38 0.38 0.34 0.27	0.00 0.00 0.00 0.00	0.01 0.00 0.01	0.89
	11 dietary_therapy 12 HM_Cd 13 walking_rate 14 drink_high 15 physical_activity	0.02 0.03 0.01 0.00 0.00 0.00	0.24	4 0 2 0 1 0 5 0 1 0	0.30 0.57 0.20 0.17 0.16	0.57 0.00 0.00 0.00 0.00	0.17 0.00 0.10 0.00 0.00	1.32 1.00 0.51 0.33 0.18		12 dietary_therapy 13 walking_day 14 dietary_supple 15 physical_activity 16 walking_rate	0.24 0.22 0.33 0.13 0.00		0.20 0.28 0.11 0.08 0.10	0.38 0.38 0.34 0.27 0.30	0.00 0.00 0.00 0.00 0.00	0.01 0.00 0.01 0.00	0.89 0.77 0.48 0.41
	11 dietary, therapy 12 HM_Cd 13 walking_rate 14 drink_high 15 physical_activity 16 stress_perception	0.02 0.03 0.01 0.00 0.00 0.00 0.00	0.24 0.42 0.22 0.10 0.00 0.00	4 0 2 0 1 0 5 0 1 0 3 0	0.30 0.57 0.20 0.17 0.16 0.08	0.57 0.00 0.00 0.00 0.00 0.00	0.17 0.00 0.10 0.00 0.00 0.00	1.32 1.00 0.51 0.33 0.18 0.11		12 dietary_therapy 13 walking_day 14 dietary_supple 15 physical_activity 16 walking_rate 17 smoking	0.24 0.22 0.33 0.13 0.00 0.05		0.20 0.28 0.11 0.08 0.10 0.07	0.38 0.38 0.34 0.27 0.30 0.22	0.00 0.00 0.00 0.00 0.00 0.00	0.01 0.00 0.01 0.00 0.03	0.89 0.77 0.48 0.41 0.37
	10 Haiking Law 11 dietary, therapy 12 HM/Cd 13 walking, rate 14 dirink, high 15 physical, activity 16 stress, perception 17 smoking	0.02 0.03 0.01 0.00 0.00 0.00 0.00 0.00	0.24 0.44 0.27 0.11 0.07 0.07 0.07 0.07	4 0 2 0 1 0 5 0 1 0 3 0 0 0	0.30 0.57 0.20 0.17 0.16 0.08 0.00	0.57 0.00 0.00 0.00 0.00 0.00 0.00	0.17 0.00 0.10 0.00 0.00 0.00 0.00	1.32 1.00 0.51 0.33 0.18 0.11 0.00		12 dietary_therapy 13 walking_day 14 dietary_supple 15 physical_activity 16 walking_rate 17 smoking 18 drink_high	0.24 0.22 0.33 0.13 0.00 0.05 0.00		0.20 0.28 0.11 0.08 0.10 0.07 0.07	0.38 0.38 0.34 0.27 0.30 0.22 0.16	0.00 0.00 0.00 0.00 0.00 0.00 0.00	0.01 0.00 0.01 0.00 0.03 0.03	0.89 0.77 0.48 0.41 0.37 0.23

Figure 4. In-depth analysis of the association between heavy metal exposure, environmental variables and their impact on metabolic syndrome. This figure sorts the data into four distinct nutrient intake clusters using the PyCaret 3.1.0 package, an exemplar of automated machine learning. In the figure, different colors represent varying data ranges: red indicates values greater than or equal to 0.7; green shows values between 0.3 and less than 0.7; red (SUM) denotes summed values equal to or exceeding 3; and green (SUM) refers to summed values ranging from 2 to less than 3.

4.3. Nutrient and Lifestyle Clustering

The meticulous establishment of dietary clusters-"Diverse Nutrient Intake", "Nutrient-Deficient", "High-Energy & Nutrient-Rich", and "Balanced Intake"-unmasked the nuanced interplay between dietary patterns and lifestyle choices. "Diverse Nutrient Intake" mirrors a broad nutritional spectrum, but with potential exposure risks—a sentiment shared by Li et al. [50]. "Nutrient-Deficient" underscores the gaps many face, with their shadowed health ramifications highlighted by Miller et al. [51]. "High-Energy & Nutrient-Rich" presents an opulence of nutrients but with the hidden danger of excess, resonating with the findings of Lachat et al. [52]. Lastly, "Balanced Intake", celebrated for its moderation, raises questions about its sources and potential contaminants—an insight explored by Mozaffarian and Rimm [53]. These clusters, enriched by the pioneering work of Algur et al. [54], serve not only as dietary compartments but also as illuminative insights into the multifaceted interplay of diet, lifestyle, and health. As elucidated, it is not just about the nutrient profile but the broader context, emphasizing that nutrition's realm is governed by a careful balance and an understanding of both its gifts and potential pitfalls. Furthermore, the pivotal study by Lind et al. [55] has underscored the potency of such advanced clustering techniques. Their rigorous research accentuates the invaluable role of these techniques in unveiling the complexities of nutrient profiles and their intertwined relationships with metabolic outcomes. Drawing from their insights, it becomes evident that a nuanced understanding of these clusters can be paramount in the development and implementation of precise public health strategies, especially when addressing multifaceted challenges related to metabolic syndrome and associated pathologies.

4.4. Underlying Themes in Clusters

The undeniable prominence of heavy metals in all nutritional clusters underscores an omnipresent concern in contemporary nutrition. Whether one's dietary inclinations lean

toward "Diverse Nutrient Intake" or veer toward the "Nutrient-Deficient", heavy metals have manifested themselves as unvielding components in these dietary categorizations. This all-encompassing significance poses profound questions about the state of our global food chains and offers a stark reminder of the imperatives of monitoring and mitigating heavy metal exposures. The intricate interplay between dietary choices and socio-economic contours emerges as another compelling theme. Specifically, education and income-two crucial pillars of socio-economic status-appear to wield substantial influence over dietary patterns. This correlation implies that individuals' nutritional choices and the associated risks or benefits might be less about individual discretion and more about broader socioeconomic determinants. The frequency of dining out, another lifestyle determinant, surfaces as a critical variable in this narrative. Martins et al. [7] delve deep into this very paradigm, shedding light on the intertwined nature of socio-economic factors, lifestyle choices, and the risk of heavy metal exposures. Such intricate relationships reiterate the necessity for a holistic perspective on nutrition—one that goes beyond mere dietary components and investigates the broader socio-economic and environmental circumstances that shape our food choices.

4.5. Implications for Metabolic Syndrome

In our attempt to understand the complexities surrounding metabolic syndrome, the capabilities of machine learning models have proven invaluable. By affording us the ability to parse vast amounts of data and identify intricate patterns, these models have brought forth distinct insights, particularly concerning the omnipresence of certain heavy metals across various nutritional profiles. Whether one's diet is teeming with nutrients or is conspicuously deficient, the pervasiveness of these metals remains a constant. This ubiquity not only points to the possible health implications of heavy metal exposure but also raises pressing concerns about our larger food supply chain and environment. Such revelations beckon an imperative: the need for an exhaustive investigation into where our foods come from, how they are processed, and what external factors might be contaminating them. This holistic understanding is paramount to framing preventive and therapeutic measures against metabolic syndrome effectively. Deng et al. [56] accentuate this very sentiment. By underscoring the potential of machine learning in elucidating dietary risks for metabolic syndrome, they also touch upon the intertwined nature of nutrition, environment, and the broader societal structures that impact health outcomes. The onus, thus, lies not just in individual dietary choices but in a collective, comprehensive approach to nutrition-one that incorporates technological prowess with traditional nutritional wisdom to ensure holistic health outcomes.

4.6. Limitations

In the analysis of the Eighth KNHANES Data, several limitations are acknowledged. One of the key limitations of this study is that while it provides interesting data on the effects of high versus low exposure to metals, it does not establish specific cut-off points for identifying a risk threshold for heavy metal exposure. This highlights the need for future research to define precise risk thresholds that can guide public health recommendations and interventions. Additionally, the study contends with recall bias due to reliance on participants' self-reported dietary intakes and exposure to environmental factors. The cross-sectional nature of the study captures only a snapshot in time, making it difficult to infer causality among nutrient intake, heavy metal exposure, and metabolic syndrome risk. There is also potential for overfitting in the machine learning models used for analysis, which might limit their generalizability to other populations or datasets. Moreover, the focus on heavy metals might not encompass all influential environmental factors impacting metabolic syndrome. Finally, significant inter-individual variability, influenced by factors such as genetics and gut microbiota, means that metabolic responses to both nutrients and heavy metals can substantially differ among individuals. Acknowledging these limitations is crucial for accurately assessing the health risks associated with varying levels of heavy metal exposure in the diet and for guiding the interpretation of the study's insights.

4.7. Recommendations and Future Directions

This research, utilizing machine learning in conjunction with epidemiological data, offers pivotal insights for healthcare professionals and policymakers. Delving into these intricate patterns presents an irrefutable need for further empirical exploration. Healthcare practitioners can harness these insights to devise personalized nutritional and lifestyle interventions, ensuring that strategies are more attuned to individual health profiles and needs. Peña-Jorquera et al. [57] underscore the importance of such individualized interventions, arguing for a future in public health that leans heavily on data-informed, tailored approaches.

On the policy front, the consistent presence of heavy metals across different nutritional clusters signals a broader systemic issue extending beyond mere dietary choices. Policymakers, informed by such findings, face a pressing mandate to draft regulations that address these environmental contaminants in our food supply. As this study intertwines advanced machine learning techniques with public health revelations, it not only deepens our current understanding but also sets the stage for future research and policy direction. As we navigate this complex interplay of data, nutrition, and health, it is evident that integrated, multi-disciplinary research approaches are paramount in our pursuit of safeguarding public health.

5. Conclusions

In conclusion, our study has provided important insights into the relationships among nutrient intake, environmental exposure to heavy metals, and the risk of metabolic syndrome. The consistent presence of heavy metals across different nutritional clusters indicates a systemic issue that transcends individual dietary choices, pointing to a broader concern within the food supply chain. Our findings highlight the critical need for policymakers to consider regulations that address environmental contaminants in our food supply. Additionally, this study underscores the value of integrating advanced machine learning techniques with epidemiological data to uncover complex relationships in public health research. As we move forward, it is clear that an integrated, multi-disciplinary approach to research is essential in our continued efforts to safeguard public health. This study not only deepens our current understanding of these complex interplays but also sets the stage for future research and policy directions in the field of nutrition and environmental health.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/nu16050724/s1, Figure S1: Numerical Significance Between Clusters Related to Nutrient Variables. This figure provides visual confirmation through a heatmap; Figure S2: Mean Scaled Nutrient Values Across Clusters; Figure S3: Training of Machine Learning Models. The Importance Score (IS), representing the importance of the factors calculated by the model by ISMM, was determined.

Author Contributions: S.J. was responsible for conceptualization, data curation, formal analysis, and writing of the original draft. Y.-J.C. contributed to writing, review, and editing, as well as supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval was not required as KNHANES provides anonymous, secondary data that are publicly available for scientific use.

Informed Consent Statement: Written informed consent was obtained from all participants involved in the KNHANES.

Data Availability Statement: The datasets for this study are available in the Korea Centers for Disease Control and Prevention database, accessible via the following URL: https://knhanes.kdca.go.kr/knhanes/sub03/sub03_02_05.do, accessed on 7 December 2023. Individuals, including international

researchers who sign up for membership, can utilize the raw data from this website. However, it is important to note that both the data access process and the user manual are provided in Korean.

Acknowledgments: The authors extend their gratitude to the Korea Ministry of Health and Welfare and the Korea Disease Control and Prevention Agency for granting access to the database for the current study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

CI	Confidence interval
IS	Importance scores
ISMM	Importance Score Mathematical Model
KNHANES	Korea National Health and Nutritional Examination Survey
OR	Odds ratio
SD	Standard deviations

References

- Li, W.; Chen, D.; Peng, Y.; Lu, Z.; Kwan, M.P.; Tse, L.A. Association Between Metabolic Syndrome and Mortality: Prospective Cohort Study. *JMIR Public Health Surveill.* 2023, 9, e44073. [CrossRef]
- 2. Kumar, S.; Senapati, S.; Bhattacharya, N.; Bhattacharya, A.; Maurya, S.K.; Husain, H.; Bhatti, J.S.; Pandey, A.K. Mechanism and recent updates on insulin-related disorders. *World J. Clin. Cases* **2023**, *11*, 5840–5856. [CrossRef] [PubMed]
- Bovolini, A.; Garcia, J.; Andrade, M.A.; Duarte, J.A. Metabolic Syndrome Pathophysiology and Predisposing Factors. *Int. J. Sports Med.* 2021, 42, 199–214. [CrossRef] [PubMed]
- Das, D.; Shruthi, N.R.; Banerjee, A.; Jothimani, G.; Duttaroy, A.K.; Pathak, S. Endothelial dysfunction, platelet hyperactivity, hypertension, and the metabolic syndrome: Molecular insights and combating strategies. *Front. Nutr.* 2023, 10, 1221438. [CrossRef] [PubMed]
- 5. Adil, S.O.; Islam, M.A.; Musa, K.I.; Shafique, K. Prevalence of Metabolic Syndrome among Apparently Healthy Adult Population in Pakistan: A Systematic Review and Meta-Analysis. *Healthcare* **2023**, *11*, 531. [CrossRef] [PubMed]
- Ye, Q.; Wang, Z.; Deng, T.; Lou, Q.; Wu, H.; Tang, W.; Xu, F. Association of socioeconomic status with metabolic syndrome and its components among adult population: A community-based cross-sectional study in Nanjing Municipality of China. *BMJ Open* 2023, 13, e074059. [CrossRef] [PubMed]
- Martins, A.C.; Ferrer, B.; Tinkov, A.A.; Caito, S.; Deza-Ponzio, R.; Skalny, A.V.; Bowman, A.B.; Aschner, M. Association between Heavy Metals, Metalloids and Metabolic Syndrome: New Insights and Approaches. *Toxics* 2023, *11*, 670. [CrossRef] [PubMed]
- Li, Y.; Gui, J.; Liu, H.; Guo, L.L.; Li, J.; Lei, Y.; Li, X.; Sun, L.; Yang, L.; Yuan, T.; et al. Predicting metabolic syndrome by obesityand lipid-related indices in mid-aged and elderly Chinese: A population-based cross-sectional study. *Front. Endocrinol.* 2023, 14, 1201132. [CrossRef] [PubMed]
- Eom, S.Y.; Choi, S.H.; Ahn, S.J.; Kim, D.K.; Kim, D.W.; Lim, J.A.; Choi, B.S.; Shin, H.J.; Yun, S.W.; Yoon, H.J.; et al. Reference levels of blood mercury and association with metabolic syndrome in Korean adults. *Int. Arch. Occup. Environ. Health* 2014, 87, 501–513. [CrossRef]
- Kwon, J.A.; Kim, B.; Kim, E.; Kwon, K. Interaction between blood cadmium and lead concentration and physical activity on hypertension from the Korean national health and nutrition examination survey in 2008–2013. BMC Public Health 2023, 23, 703. [CrossRef]
- 11. Fu, Z.; Xi, S. The effects of heavy metals on human metabolism. Toxicol. Mech. Methods 2020, 30, 167–176. [CrossRef] [PubMed]
- Rehman, K.; Fatima, F.; Waheed, I.; Akash, M.S.H. Prevalence of exposure of heavy metals and their impact on health consequences. J. Cell Biochem. 2018, 119, 157–184. [CrossRef] [PubMed]
- 13. Popkin, B.M.; Adair, L.S.; Ng, S.W. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr. Rev.* **2012**, *70*, 3–21. [CrossRef] [PubMed]
- 14. Subedi, Y.P.; Marais, D.; Newlands, D. Where is Nepal in the nutrition transition? *Asia Pac. J. Clin. Nutr.* **2017**, *26*, 358–367. [PubMed]
- Biesbroek, S.; Kok, F.J.; Tufford, A.R.; Bloem, M.W.; Darmon, N.; Drewnowski, A.; Fan, S.; Fanzo, J.; Gordon, L.J.; Hu, F.B.; et al. Toward healthy and sustainable diets for the 21st century: Importance of sociocultural and economic considerations. *Proc. Natl. Acad. Sci. USA* 2023, 120, e2219272120. [CrossRef] [PubMed]
- 16. Kelleher, J.D.; Namee, B.M.; D'Arcy, A. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies; MIT Press: Cambridge, MA, USA, 2015.
- 17. Shin, H.; Shim, S.; Oh, S. Machine learning-based predictive model for prevention of metabolic syndrome. *PLoS ONE* **2023**, *18*, e0286635. [CrossRef] [PubMed]

- Kim, H.; Heo, J.H.; Lim, D.H.; Kim, Y. Development of a Metabolic Syndrome Classification and Prediction Model for Koreans Using Deep Learning Technology: The Korea National Health and Nutrition Examination Survey (KNHANES) (2013–2018). *Clin. Nutr. Res.* 2023, *12*, 138–153. [CrossRef]
- Kim, J.; Mun, S.; Lee, S.; Jeong, K.; Baek, Y. Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea. *BMC Public Health* 2022, 22, 664. [CrossRef]
- 20. Kweon, S.; Kim, Y.; Jang, M.J.; Kim, Y.; Kim, K.; Choi, S.; Chun, C.; Khang, Y.H.; Oh, K. Data resource profile: The Korea National Health and Nutrition Examination Survey (KNHANES). *Int. J. Epidemiol.* **2014**, *43*, 69–77. [CrossRef]
- Oh, K.; Kim, Y.; Kweon, S.; Kim, S.; Yun, S.; Park, S.; Lee, Y.K.; Kim, Y.; Park, O.; Jeong, E.K. Korea National Health and Nutrition Examination Survey, 20th anniversary: Accomplishments and future directions. *Epidemiol. Health* 2021, 43, e2021025. [CrossRef]
- 22. Moon, S.S. Additive effect of heavy metals on metabolic syndrome in the Korean population: The Korea National Health and Nutrition Examination Survey (KNHANES) 2009–2010. *Endocrine* **2014**, *46*, 263–271. [CrossRef] [PubMed]
- Chung, H.K.; Park, J.Y.; Cho, Y.; Shin, M.J. Contribution of dietary patterns to blood heavy metal concentrations in Korean adults: Findings from the Fifth Korea National Health and Nutrition Examination Survey 2010. Food Chem. Toxicol. 2013, 62, 645–652. [CrossRef]
- 24. Little, R.J.; Rubin, D.B. Statistical Analysis with Missing Data; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2019.
- 25. Iglewicz, B.; Hoaglin, D.C. How to Detect and Handle Outliers; A.S.Q.C. Quality Press: Milwaukee, WI, USA, 1993; Volume 16.
- 26. Hansen, J.; Ahern, S.; Earnest, A. Evaluations of statistical methods for outlier detection when benchmarking in clinical registries: A systematic review. *BMJ Open* **2023**, *13*, e069130. [CrossRef]
- Damigos, G.; Zacharaki, E.I.; Zerva, N.; Pavlopoulos, A.; Chatzikyrkou, K.; Koumenti, A.; Moustakas, K.; Pantos, C.; Mourouzis, I.; Lourbopoulos, A. Machine learning based analysis of stroke lesions on mouse tissue sections. *J. Cereb. Blood Flow Metab.* 2022, 42, 1463–1477. [CrossRef]
- 28. Tangni, E.K.; Huybrechts, B.; Masquelier, J.; Van Hoeck, E. Organisation of Multi-Mycotoxin Proficiency Tests: Evaluation of the Performances of the Laboratories Using the Triple A Rating Approach. *Toxins* **2021**, *13*, 591. [CrossRef]
- Alabrah, A. An Improved CCF Detector to Handle the Problem of Class Imbalance with Outlier Normalization Using IQR Method. Sensors 2023, 23, 4406. [CrossRef] [PubMed]
- Garlits, J.; McAfee, S.; Taylor, J.A.; Shum, E.; Yang, Q.; Nunez, E.; Kameron, K.; Fenech, K.; Rodriguez, J.; Torri, A.; et al. Statistical Approaches for Establishing Appropriate Immunogenicity Assay Cut Points: Impact of Sample Distribution, Sample Size, and Outlier Removal. AAPS J. 2023, 25, 37. [CrossRef]
- 31. Ajuwon, B.I.; Roper, K.; Richardson, A.; Lidbury, B.A. Routine blood test markers for predicting liver disease post HBV infection: Precision pathology and pattern recognition. *Diagnosis* 2023, *10*, 337–347. [CrossRef] [PubMed]
- 32. Holzinger, A.; Saranti, A.; Angerschmid, A.; Finzel, B.; Schmid, U.; Mueller, H. Toward human-level concept learning: Pattern benchmarking for AI algorithms. *Patterns* **2023**, *4*, 100788. [CrossRef]
- 33. Dalmaijer, E.S.; Nord, C.L.; Astle, D.E. Statistical power for cluster analysis. BMC Bioinform. 2022, 23, 205. [CrossRef]
- 34. Rennie, J.P.; Zhang, M.; Hawkins, E.; Bathelt, J.; Astle, D.E. Mapping differential responses to cognitive training using machine learning. *Dev. Sci.* 2020, 23, e12868. [CrossRef]
- 35. Cui, Y.; Zhou, H.L.; Wei, M.H.; Song, W.J.; Di, D.S.; Zhang, R.Y.; Wei, S.; Liu, J.A.; Wang, Q. Multiple vitamin co-exposure and mortality risk: A prospective study. *Clin. Nutr.* 2022, *41*, 337–347. [CrossRef]
- Andrade, V.M.B.; de Santana, M.L.P.; Fukutani, K.F.; Queiroz, A.T.L.; Arriaga, M.B.; Conceição-Machado, M.E.P.; Silva, R.C.R.; Andrade, B.B. Multidimensional Analysis of Food Consumption Reveals a Unique Dietary Profile Associated with Overweight and Obesity in Adolescents. *Nutrients* 2019, *11*, 1946. [CrossRef]
- Ahn, S. Building and analyzing machine learning-based warfarin dose prediction models using scikit-learn. *Transl. Clin. Pharmacol.* 2022, 30, 172–181. [CrossRef]
- Lee, J.S.; Lee, S.K. Identification of Risk Groups for and Factors Affecting Metabolic Syndrome in South Korean Single-Person Households Using Latent Class Analysis and Machine Learning Techniques: Secondary Analysis Study. *JMIR Form Res.* 2023, 7, e42756. [CrossRef] [PubMed]
- 39. Jain, A.K. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 2010, 31, 651–666. [CrossRef]
- English, L.K.; Ard, J.D.; Bailey, R.L.; Bates, M.; Bazzano, L.A.; Boushey, C.J.; Brown, C.; Butera, G.; Callahan, E.H.; de Jesus, J.; et al. Evaluation of Dietary Patterns and All-Cause Mortality: A Systematic Review. *JAMA Netw. Open* 2021, 4, e2122277. [CrossRef] [PubMed]
- Wang, D.D.; Li, Y.; Afshin, A.; Springmann, M.; Mozaffarian, D.; Stampfer, M.J.; Hu, F.B.; Murray, C.J.L.; Willett, W.C. Global Improvement in Dietary Quality Could Lead to Substantial Reduction in Premature Death. *J. Nutr.* 2019, 149, 1065–1074. [CrossRef] [PubMed]
- 42. Stubbs, R.J.; Horgan, G.; Robinson, E.; Hopkins, M.; Dakin, C.; Finlayson, G. Diet composition and energy intake in humans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2023**, *378*, 20220449. [CrossRef] [PubMed]
- 43. Miller, G.D.; Ragalie-Carr, J.; Torres-Gonzalez, M. Perspective: Seeing the Forest Through the Trees: The Importance of Food Matrix in Diet Quality and Human Health. *Adv. Nutr.* **2023**, *14*, 363–365. [CrossRef] [PubMed]

- 44. Mehouel, F.; Fowler, S.W. Review of the toxic trace elements arsenic, cadmium, lead and mercury in seafood species from Algeria and contiguous waters in the Southwestern Mediterranean Sea. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 3288–3301. [CrossRef] [PubMed]
- 45. Chen, M.; Xu, X.; Liu, Y.; Yao, Y.; Zhang, P.; Liu, J.; Zhang, Q.; Li, R.; Li, H.; Liu, Y.; et al. Association of eating habits with health perception and diseases among Chinese physicians: A cross-sectional study. *Front. Nutr.* **2023**, *10*, 1226672. [CrossRef] [PubMed]
- 46. Rakib, M.R.J.; Jolly, Y.N.; Enyoh, C.E.; Khandaker, M.U.; Hossain, M.B.; Akther, S.; Alsubaie, A.; Almalki, A.S.A.; Bradley, D.A. Levels and health risk assessment of heavy metals in dried fish consumed in Bangladesh. *Sci. Rep.* 2021, *11*, 14642. [CrossRef] [PubMed]
- Sobihah, N.N.; Ahmad Zaharin, A.; Khairul Nizam, M.; Ley Juen, L.; Kyoung-Woong, K. Bioaccumulation of heavy metals in maricultured fish, *Lates calcarifer* (Barramudi), *Lutjanus campechanus* (red snapper) and *Lutjanus griseus* (grey snapper). *Chemosphere* 2018, 197, 318–324. [CrossRef] [PubMed]
- Aslam, M.; Aslam, A.; Sheraz, M.; Ali, B.; Ulhassan, Z.; Najeeb, U.; Zhou, W.; Gill, R.A. Lead Toxicity in Cereals: Mechanistic Insight Into Toxicity, Mode of Action, and Management. *Front. Plant Sci.* 2021, 11, 587785. [CrossRef]
- Noger-Huet, É.; Vagner, M.; Le Grand, F.; Graziano, N.; Bideau, A.; Brault-Favrou, M.; Churlaud, C.; Bustamante, P.; Lacoue-Labarthe, T. Risk and benefit assessment of seafood consumption harvested from the Pertuis Charentais region of France. *Environ. Pollut.* 2022, 292, 118388. [CrossRef]
- 50. Li, T.; Yu, L.; Yang, Z.; Shen, P.; Lin, H.; Shui, L.; Tang, M.; Jin, M.; Chen, K.; Wang, J. Associations of Diet Quality and Heavy Metals with Obesity in Adults: A Cross-Sectional Study from National Health and Nutrition Examination Survey (NHANES). *Nutrients* **2022**, *14*, 4038. [CrossRef]
- Miller, V.; Webb, P.; Cudhea, F.; Shi, P.; Zhang, J.; Reedy, J.; Erndt-Marino, J.; Coates, J.; Mozaffarian, D. Global dietary quality in 185 countries from 1990 to 2018 show wide differences by nation, age, education, and urbanicity. *Nat. Food.* 2022, 3, 694–702. [CrossRef]
- 52. Lachat, C.; Nago, E.; Verstraeten, R.; Roberfroid, D.; Van Camp, J.; Kolsteren, P. Eating out of home and its association with dietary intake: A systematic review of the evidence. *Obes. Rev.* **2012**, *13*, 329–346. [CrossRef]
- 53. Mozaffarian, D.; Rimm, E.B. Fish intake, contaminants, and human health: Evaluating the risks and the benefits. *JAMA* 2006, 296, 1885–1899. [CrossRef]
- 54. Algur, Y.; Rummo, P.E.; McAlexander, T.P.; De Silva, S.S.A.; Lovasi, G.S.; Judd, S.E.; Ryan, V.; Malla, G.; Koyama, A.K.; Lee, D.C.; et al. Assessing the association between food environment and dietary inflammation by community type: A cross-sectional REGARDS study. *Int. J. Health Geogr.* **2023**, *22*, 24. [CrossRef] [PubMed]
- 55. Lind, L.; Sundström, J.; Elmståhl, S.; Dekkers, K.F.; Smith, J.G.; Engström, G.; Fall, T.; Ärnlöv, J. The metabolomic profile associated with clustering of cardiovascular risk factors-A multi-sample evaluation. *PLoS ONE* **2022**, *17*, e0274701. [CrossRef] [PubMed]
- Deng, X.; Qiu, L.; Sun, X.; Li, H.; Chen, Z.; Huang, M.; Hu, F.; Zhang, Z. Early prediction of body composition parameters on metabolically unhealthy in the Chinese population via advanced machine learning. *Front. Endocrinol.* 2023, 14, 1228300. [CrossRef] [PubMed]
- 57. Peña-Jorquera, H.; Cid-Jofré, V.; Landaeta-Díaz, L.; Petermann-Rocha, F.; Martorell, M.; Zbinden-Foncea, H.; Ferrari, G.; Jorquera-Aguilera, C.; Cristi-Montero, C. Plant-Based Nutrition: Exploring Health Benefits for Atherosclerosis, Chronic Diseases, and Metabolic Syndrome—A Comprehensive Review. *Nutrients* **2023**, *15*, 3244. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.