The Journal of Nutrition 156 (2026) 101743

American Society for Nutrition
Excellence in Nutrition Research and Practice

JN THE JOURNAL OF NUTRITION

journal homepage: https://jn.nutrition.org/

Community and International Nutrition

Who Is at High Risk of Poor Diet Quality? A Precision Public Health Nutrition Approach using Machine Learning

Rosanne Blanchet [1,2,*], Natalie Doan [3,4], Sara Nejatinamini [5], Camilo E Valderrama [4,6], David Hammond [3], Lana Vanderlee [7], Dana Lee Olstad [4]

[1] *Department of Social and Preventive Medicine, School of Public Health, University of Montreal, Montreal, QC, Canada;* [2] *Public Health Research Centre (Centre de recherche en santé publique; CReSP), Montreal, QC, Canada;* [3] *School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada;* [4] *Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada;* [5] *Alberta's Tomorrow Project, Cancer Research & Business Partnerships, Cancer Care Alberta, Calgary, AB, Canada;* [6] *Department of Applied Computer Science, University of Winnipeg, Winnipeg, MB, Canada;* [7] *School of Nutrition, NUTRISS Centre, Universite Laval, Quebec, QC, Canada*

A B S T R A C T

Background: Suboptimal diet quality contributes to poor health. Numerous factors at all levels of the socioecological model intersect to influence individuals' diet quality. A precision public health nutrition framework can be used to understand how these factors jointly shape diet quality across several subgroups in the population. This is, however, very challenging to operationalize.
Objectives: The purpose of this study was to assess whether machine learning can be leveraged to operationalize a precision public health nutrition approach by more precisely identifying subgroups with the highest proportion of adults with poor diet quality and the most important predictors of diet quality.
Methods: We conducted a secondary analysis of cross-sectional data from the 2018 and 2019 International Food Policy Study in Canada (n = 5093). A total of 42 candidate predictors within 4 domains (sociodemographic characteristics and socioeconomic position; food policies and environments; food literacy; health-related practices and indicators) were used. The Healthy Eating Index-2015 (HEI-2015) was used to assess diet quality; tertiles of HEI-2015 scores were defined as the outcome. Conditional inference tree (CIT) and conditional random forest (CRF) analyses were conducted.
Results: The CIT partitioned the sample into 7 subgroups with different proportions of adults with lower, moderate, or higher diet quality using 6 predictors. The probability of lower diet quality ranged from 16.9% to 49.9% across subgroups. The subgroup with the highest proportion of adults with lower diet quality was characterized by individuals confident in using $\leq$4 cooking techniques. Based on the CRF model and conditional variable importance, the 5 most important predictors of diet quality were frequency of food label use, confidence in using cooking techniques, perceived general health, household food insecurity status, and health literacy.
Conclusions: This study provides evidence that machine learning approaches can be leveraged to operationalize a precision public health nutrition approach.

Keywords: diet quality, machine learning, conditional inference tree, conditional random forest, precision public health, International Food Policy Study, Healthy Eating Index

Introduction

The majority of adults in high-income countries have suboptimal diet quality, which substantially increases their risk of developing poor physical and mental health [1,2]. However, some individuals have poorer diet quality than others. It is widely acknowledged that an array of factors at multiple levels of influence intersect in complex ways to shape individuals' dietary patterns and hence their diet quality [3–5]. In this manuscript, we use the term "intersection" in a conceptual sense to refer to the co-occurrence and mutual influence of social, economic, and behavioral determinants in people's lived

Abbreviations: ASA24-Canada, Automated Self-Administered 24-Hour Dietary Assessment Tool for Canada; CIT, conditional inference tree; CRF, conditional random forest; HEI-2015, Healthy Eating Index-2015; IFPS, International Food Policy Study; SEP, socioeconomic position.

This article is part of a special issue entitled: AI, ML, and Precision Nutrition published in The Journal of Nutrition.

* Corresponding author. *E-mail address:* rosanne.blanchet@umontreal.ca (R. Blanchet).

https://doi.org/10.1016/j.tjnut.2026.101743
Received 15 February 2026; Received in revised form 9 July 2026; Accepted 16 July 2026; Available online 20 July 2026
0022-3166/

experiences [6]. This differs from "interaction," which is a statistical term describing how the effect of one variable on an outcome depends on the level of another. Accordingly, when we state that factors "intersect," we refer to their combined presence and co-occurrence, not to statistical interactions or correlations among predictors.

Although a large number of studies have identified predictors of diet quality, such as income, education, gender, food preferences, nutritional knowledge, and food environments [7–14], only a small number of studies have examined how some factors (usually 2–3) statistically interact to shape dietary patterns [15–20]. Reviews have synthesized these empirical findings [5, 21–23] and constructed system maps of these factors that jointly shape dietary patterns [24]. This literature is useful for identifying specific factors that shape diet quality, along with a small number of interactions among them, but it provides limited insight into how numerous factors at various socioecological levels intersect simultaneously to shape diet quality among specific population subgroups [3].

A precision public health framework can be used to understand how factors intersect to shape individuals' dietary patterns, thereby identifying more precise population subgroups at risk of poor diet quality [25]. Precision public health has been defined as an approach that "investigates how multiple dimensions of social position interact to confer health risk differently for precisely defined population subgroups according to the social contexts in which they are embedded, while considering relevant biological and behavioral factors" [25]. Applied to the realm of nutrition, the purpose of such an approach is not to problematize individuals themselves, but to more precisely identify the contextual origins of their patterns of consumption to develop more precisely tailored interventions that are specific enough to be useful in improving their diet quality and reducing dietary inequities.

A precision public health nutrition approach is, however, very challenging to operationalize using conventional statistical techniques [26–28]. For example, parametric regression-based models (a statistical learning method used for causal or etiologic inference) cannot practically accommodate interactions among many variables simultaneously, are limited in their ability to handle multimodal data, require a priori hypotheses about how variables relate to one another, and rely on strict linearity assumptions that may not be met [27,29–32]. Conducting large numbers of statistical tests also substantially increases the chance of false positive results [26,30,32]. Moreover, their results can be very difficult to interpret [e.g., 26, 30,32]. In this respect, supervised machine learning may offer key advantages in attempts to more precisely identify population subgroups at higher risk of poor diet quality. This is because these machine learning approaches rely on the ability of algorithms to learn from patterns in the data, without defining a priori hypotheses, to best predict an outcome [28,33]. Additionally, many tree-based machine learning approaches can handle different types of variables (e.g., continuous, ordinal, and nominal) and yield highly interpretable outcomes [28,34]. This interpretability is a key advantage compared with unsupervised cluster analyses, another conventional approach. For example, tree-based machine learning approaches and related ensemble methods produce inherently hierarchical structures that clarify the relative importance of predictors. This feature is absent from cluster analyses and greatly enhances interpretability when many variables are involved [34,35]. Furthermore, these supervised machine learning approaches can directly identify combinations of variables that best predict an outcome [34], whereas unsupervised clustering methods do not use the outcome variable to guide classification, making them conceptually distinct [36]. In this way, supervised machine learning may provide a novel and useful means of operationalizing a precision public health approach that seeks to more precisely identify population subgroups with a high proportion of adults with poor diet quality [28].

The purpose of this study was to assess whether machine learning can be leveraged to operationalize a precision public health nutrition approach. Specific objectives were to more precisely identify subgroups with the highest proportion of adults with lower diet quality using a large number of predictors spanning all socioecological levels in a data-driven way and to identify the most important predictors of adults' diet quality. As such, it is important to note that our focus was on identifying predictive associations rather than on causal inference.

Methods

Study ethics

The current secondary analysis was approved by the University of Calgary Research Ethics Board (REB21-1307), and the full International Food Policy Study (IFPS) was approved by the University of Waterloo Research Ethics Committee (ORE #21460). All participants provided informed consent prior to participating and were compensated in accordance with the panel's customary incentive structure.

Study design and participants

This study entailed a secondary analysis of cross-sectional data collected from the 2018 and 2019 IFPS. The IFPS is an annual online survey of adults aged ≥18 y living in Canada, Australia, the United Kingdom, Mexico, and the United States. Participants living in Canada were invited to complete one 24-h dietary recall. Participants were randomly drawn from the Nielsen Consumer Insights Global Panel in Canada to approximate the population distribution of sex, age, education, and language. Full details of the IFPS are available elsewhere [37].

Eligible adults received an e-mail invitation containing information about the study and a unique survey access link. Surveys were conducted in English or French. The median time taken to complete the survey was 40 min. Figure 1 displays the participant flow chart. After removing surveys with poor data quality (as identified by data quality checks, such as not being able to correctly name the current month or completing the survey in <15 min), a total of 4397 and 4107 adults participated in 2018 and 2019, with a cooperation rate of 67.1% and 60.7%, respectively [38]. Participants who did not complete a 24-h dietary recall (n = 2597), had implausible weight or height (reported height <0.91 m or >2.13 m, or weight <20.41 kg; n = 31) [39], or who reported zero energy intake (n = 30) were excluded. The 2018 and 2019 samples were combined, and the 2019 data from participants who participated at both time points were removed (n = 187). This left a total of 5659 participants in the initial analytical sample.

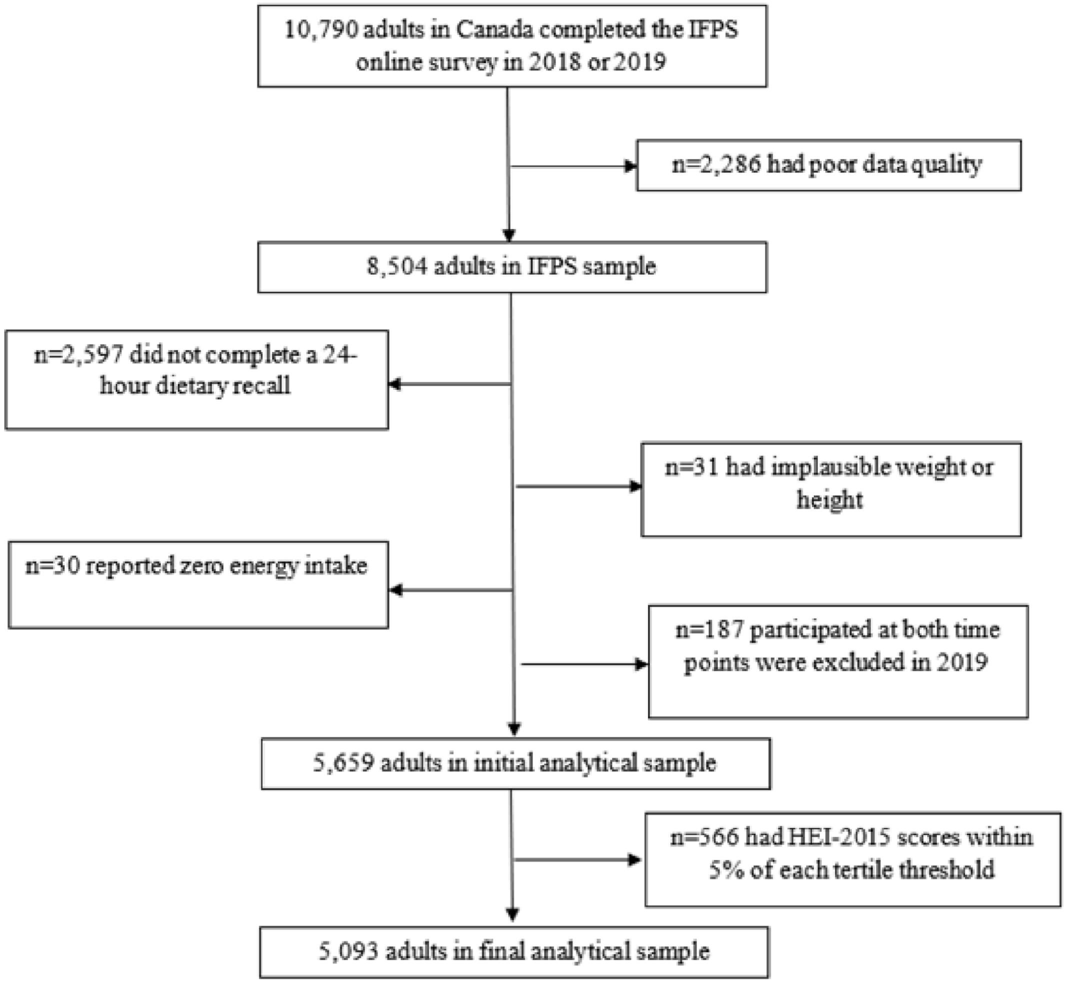

FIGURE 1. International Food Policy Study (IFPS) 2018 and 2019 flow chart. HEI-2015, Healthy Eating Index-2015.

Data collection

The majority of IFPS questionnaire items were drawn or adapted from previous national surveys or other research studies [37]. The questionnaire was pretested with 50 young adults in Canada using cognitive interviewing methodology. The surveys are available at www.foodpolicystudy.com.

Candidate predictors

We selected all the variables that were included in the IFPS that did not directly assess dietary patterns (e.g., frequency of fast-food consumption) and that displayed variance (e.g., shopping for food at a grocery store was not selected, as nearly everyone reported shopping for food in grocery stores). A total of 42 candidate predictors were included in our models. Details on all candidate predictors, including questionnaire items and how they were operationalized, are described in Supplemental Table 1.

Candidate predictors were classified into the following 4 domains to enhance comprehension and interpretability: sociodemographic characteristics and socioeconomic position (SEP); food policies and environments; food literacy; and health-related practices and indicators (Table 1). Notably, the factors included within these domains spanned all levels of the socio-ecological model.

Outcome of interest: Healthy Eating Index-2015 scores

Participants self-reported all food and beverages consumed the day prior to the survey using the validated Automated Self-Administered 24-Hour Dietary Assessment Tool for Canada (ASA24-Canada-2016 was used in 2018 and ASA24-Canada-2018 was used in 2019) [40–44]. The ASA24 is a web-based tool that guides participants through completing a 24-h dietary recall using a modified version of the 5-step Automated Multiple Pass Method [45,46]. ASA24-Canada was adapted from the ASA24 to reflect the Canadian food supply [41]. Dietary intake data were coded and linked to the Canadian Nutrient File [47] (2015 version), the Food and Nutrient Database for Dietary Studies [48], and the United States Department of Agriculture's Food Patterns Equivalents Database [49] to estimate intakes of nutrients, food groups, and other dietary components required to calculate Healthy Eating Index-2015 (HEI-2015) scores using the National Cancer Institute simple algorithm [50]. HEI-2015 scores were calculated based on intakes of adequacy components (i.e., foods to eat more of, including total fruits, whole fruits, total vegetables, greens and beans, whole grains, dairy, total protein foods, seafood and plant proteins, and fatty acids) and moderation components (i.e., foods to eat less of, including refined grains, sodium, added sugars, and saturated fats). Scores can range

TABLE 1
Distribution of candidate predictors among adults who participated in the 2018 or 2019 International Food Policy Study (n = 5093)

Candidate predictors	*n*	Mean (SD)/ percent
Sociodemographic characteristics and socioeconomic position		
Age	5093	46.7 (16.0)
Gender		
Man	2421	47.5
Woman	2621	51.5
Other identities	51	1.0
Living with child(ren) under the age of 18		
Yes	1269	24.9
No	3824	75.1
Educational attainment		
Less than high school diploma	209	4.1
High school diploma or equivalent	1031	20.2
Trade certificate or diploma from a technical/vocational school or apprenticeship training, diploma or certificate from community college or CEGEP (other than trades certificates or diplomas), some university, or university certificate/ diploma below the bachelor's level	1967	38.6
Bachelor's degree (e.g., BA, BSc)	1256	24.7
University degree above the bachelor's level (e.g., master's, professional school, doctorate)	630	12.4
Occupation		
Employed	2723	53.5
Retired	1075	21.1
Student	273	5.4
Unemployed/looking for work	217	4.3
Household work/caregiving	467	9.2
Long term illness	207	4.1
Other	131	2.6
Race/ethnicity		
White	3616	71.0
East/Southeast Asian	452	8.9
South Asian	145	2.8
Black	101	2.0
Indigenous	110	2.2
Latino/American	57	1.1
Arab/Middle Eastern	74	1.5
Mixed/other	538	10.6
Born in Canada		
Yes	4182	82.1
No	911	17.9
Province of residence		
Alberta	519	10.2
British Columbia	632	12.4
Manitoba	210	4.1
New Brunswick	122	2.4
Newfoundland and Labrador	60	1.2
Nova Scotia	167	3.3
Ontario	1922	37.7
Prince Edward Island	23	0.5
Quebec	1307	25.7
Saskatchewan	131	2.6
Urban/rural residence		
Urban	4388	86.2
Rural	705	13.8
Language		
English	4140	81.3
French	953	18.7
Perceived income adequacy		
Very difficult	390	7.7
Difficult	938	18.4
Neither easy nor difficult	1858	36.5
Easy	1238	24.3
Very easy	669	13.1
Household food insecurity status[1]		
Food secure	3192	62.7
Marginally food insecure	458	9.0
Moderately food insecure	823	16.2
Severely food insecure	620	12.2
Subjective social status[1]	5093	5.6 (1.9)
Food policies and environments		
Frequency of noticing food labels		
Never	65	1.3
Rarely	148	2.9
Sometimes	467	9.2
Often	1273	25
All the time	3140	61.7
Frequency of using food labels		
Never	282	5.5
Rarely	651	12.8
Sometimes	1500	29.5
Often	1629	32.0
All the time	1031	20.2
Understand food labels		
Very hard to understand	103	2.0
Hard to understand	384	7.5
Neither hard nor easy	1029	20.2
Easy to understand	2360	46.3
Very easy to understand	1217	23.9
Noticed nutrition information in restaurants		
Yes	1162	22.8
No	3931	77.2
Used nutrition information in restaurants		
Did not notice or nutrition information did not influence what I ordered	4551	89.4
Yes, nutrition information influenced what I ordered	542	10.6
Looked at Canada's Food Guide		
In the last month	371	7.3
In the last 6 mo	841	16.5
In the last year	749	14.7
More than a year ago	1968	38.6
Never	1164	22.9
Obtained nutrition information from government or policy-related source of information[1]	5093	0.9 (1.1)
Obtained nutrition information from industry or media[1]	5093	0.5 (0.9)
Obtained nutrition information from health professional or alternative health practitioner[1]	5093	0.4 (0.7)

(continued on next page)

TABLE 1 (continued)

Candidate predictors	n	Mean (SD)/percent
Obtained nutrition information from other sources[1]	5093	0.6 (0.8)
Trust messages from health experts on sugary drinks		
Strongly agree	984	19.3
Agree	2640	51.8
Neither agree nor disagree	1136	22.3
Disagree	237	4.7
Strongly disagree	96	1.9
Trust messages from the food and beverage industry on sugary drinks		
Strongly agree	248	4.9
Agree	809	15.9
Neither agree nor disagree	1712	33.6
Disagree	1502	29.5
Strongly disagree	822	16.1
Noticed educational messages about healthy eating		
Yes	928	18.2
No	4165	81.8
Ease of finding nutrition information in grocery stores		
Very hard to find	233	4.6
Hard to find	960	18.8
Neither hard nor easy	1851	36.3
Easy to find	1699	33.4
Very easy to find	350	6.9
Noticed unhealthy food marketing[1]	5093	2.2 (1.6)
Food literacy		
Health literacy		
Poor literacy level	767	15.1
Medium literacy level	1012	19.9
High literacy level	3314	65.1
Perceived nutrition knowledge		
Not at all knowledgeable	187	3.7
A little knowledgeable	1440	28.3
Somewhat knowledgeable	2232	43.8
Very knowledgeable	1063	20.9
Extremely knowledgeable	171	3.4
Doctor advised about healthy eating[1]	5093	0.9 (1.2)
Confidence in using cooking techniques[1]	5093	7.2 (2.6)
Ability to prepare different foods[1]	5093	10.3 (2.7)
Perceived cooking skills		
Poor	236	4.6
Fair	826	16.2
Good	1930	37.9
Very good	1589	31.2
Excellent	512	10.1
Main household food shopper		
Yes	3611	70.9
No	684	13.4
Share equally with other(s)	798	15.7
Frequency of preparing main meals		
Never	125	2.5
Only for special occasions	229	4.5
Less than once a week	305	6.0
1 or 2 days a week	586	11.5
Some days (3–4 a week)	991	19.5
Most days (5–6 a week)	1591	31.2
Every day	1266	24.9
Health-related practices and indicators		
Perceived general health		
Poor	252	4.9
Fair	1192	23.4
Good	2405	47.2
Very good	1053	20.7
Excellent	191	3.8
Perceived mental health		
Poor	338	6.6
Fair	839	16.5
Good	1665	32.7
Very good	1498	29.4
Excellent	753	14.8
Stress level		
Not at all stressful	425	8.3
Not very stressful	1442	28.3
A bit stressful	2174	42.7
Very stressful	790	15.5
Extremely stressful	262	5.1
BMI, kg/m^2	5093	27.1 (6.3)
Tobacco use		
No	4060	79.7
Yes, occasionally	303	5.9
Yes, every day	730	14.3
Marijuana use		
I have never used marijuana	3043	59.7
I have used marijuana but not in the last 12 mo	925	18.2
Less than once a month	339	6.7
Once a month	98	1.9
2 or 3 times a month	123	2.4
Once a week	75	1.5
2 or 3 times a week	103	2.0
4 to 6 times a week	90	1.8
Every day	297	5.8

Values are unweighted estimates derived after imputing missing data.
[1] See Supplemental Table 1 for detailed item descriptions and scoring procedures.

from 0 to 100, with a higher score indicating a higher diet quality [50,51]. The HEI-2015 was used in preference to Canadian diet quality indices, as the Healthy Eating Food Index-2019 had not been released at the time of our analyses [52].

Analytical methods

Descriptive characteristics of the sample were assessed using proportions and mean ± SD. HEI-2015 scores were split into tertiles to avoid assuming linearity, enhance interpretability, and capture threshold effects, whereby changes in predictors may have different implications at lower, moderate, and higher diet quality. Given that HEI-2015 scores were normally distributed (with most scores centered around the middle of the distribution), treating the variable as continuous resulted in poor predictive sensitivity for individuals at the extremes of lower and higher diet quality. Modeling 3 distinct levels of diet quality, as opposed to modeling diet quality as a discrete outcome, improved the stability and interpretability of our models. To ensure sufficient variance between subgroups and to reduce misclassification that may arise from boundary instability when converting a continuous score into discrete classes, participants whose scores fell within 5% of each threshold were excluded. As a result, participants whose HEI-2015 scores fell within the 0 to 30th, 35th to 65th, and 70th to 100th percentiles were classified as having lower, moderate, and higher diet quality, respectively, and the others were excluded ($n = 566$), resulting in a final sample size of 5093 participants.

We used 2 complementary analytical approaches to enhance rigor. First, we used conditional inference trees (CITs) [53,54], a data mining approach that can identify combinations of variables that best predict an outcome [32,33], to identify subgroups of adults with lower, moderate, and higher diet quality. CIT analyses result in inverted "tree" structures depicting predictors, thresholds, and distinct subgroups of participants who differ on the outcome of interest [54,55], facilitating interpretation. Second, we used conditional random forests (CRFs) to identify the most important predictors of adults' diet quality. CRFs provide a more generalizable assessment of the relative importance of predictors than CITs, as they are less prone to overfitting [54,56,57]. CRFs also have higher classification accuracy than traditional random forests in the presence of highly correlated predictors, such as socioeconomic characteristics [58]. Together, CITs and CRFs can be used to explore how factors intersect in complex, nonlinear ways to shape individuals' dietary patterns, which is essential for operationalizing a precision public health nutrition approach.

The CIT method embeds a data-driven statistical testing framework in a recursive partitioning algorithm for model building [59]. The algorithm randomly selects a subset of predictors, identifies the predictor that best differentiates between the outcomes of interest (i.e., bottom, middle, and top HEI-2015 tertiles), and finds the optimal threshold for this predictor to split the data into 2 datasets. Simply put, CIT makes binary splits, and therefore aggregates the categories, to best predict the outcome. The algorithm repeats this process until no further splits can be made, given the model's parameters. The 42 candidate predictors within the 4 domains described in Table 1 were entered into the algorithm. The minimum number of observations in each terminal node (subgroup) was set to 250 to ensure public health relevance [60]. We set the number of variables considered at each split to 14 (one-third of all candidate predictors), a choice that increases diversity across trees while reducing the influence of any highly predictive variables. Accordingly, at each node, the algorithm randomly selected a subset of 14 predictors, evaluated their ability to separate the 3 tertiles, and chose the most discriminating feature for the split. These parameters were set a priori based on interpretability and methodological considerations rather than optimized through cross-validation.

The CRF was conducted using the same 42 candidate predictors used for the CIT analysis. CRF uses recursive algorithms to derive an ensemble of CITs grown using bootstrap subsamples drawn from the data and averages their predictions. For each bootstrap sample, the CRF algorithm builds a CIT using a random selection of predictors. The bootstrap sampling and random selection of predictors at each split allow a diverse forest of CITs to be grown, which enhances generalizability [54,56,57]. The CRF analysis was created based on 500 CITs using the same parameters as in the CIT analysis (Supplemental Table 2).

Conditional predictor importance scores represent the relative importance of each main predictor in classifying participants into distinct subgroups with different probabilities of the outcome of interest. Scores were assessed by comparing the difference between the out-of-bag error resulting from a data set obtained through random permutations of each predictor and the out-of-bag error resulting from the original data set [61]. In other words, the conditional importance of each predictor was calculated by recalculating the overall error rate after the values of that predictor were randomly permuted, which mimics the absence of that predictor [62]. Interaction effects are not explicitly ranked, although CIT and CRF inherently account for interactions through their hierarchical splitting and ensemble structure.

For each candidate predictor, refusals and "don't know" responses were treated as missing values. The percentage of missing values for candidate predictors was <7.7%. Missing values were imputed using random forest imputation, which has been shown to provide more accurate imputations than k-nearest neighbor imputation or multivariate imputation using chained equations [63]. The outcome of interest (HEI-2015 scores) had no missing values. All analyses were performed using the "party" and "partykit" packages in R [53] (version 4.1.1; R Foundation for Statistical Computing) using the parameters detailed in Supplemental Table 2.

The CIT and CRF models were trained using 90% of the dataset and tested using 10% of the dataset to evaluate the performance of the models on new datasets. Model performance on the training and test datasets was compared to assess whether the models were overfitting the training data or were generalizing effectively when introduced to new data.

Results

Descriptive statistics

Participants had a mean age of 46.7 ± 16.0 y (Table 1). Approximately half of the participants identified as female (51.5%) and 1.0% identified as gender diverse. Most participants identified as White (71.0%) and were born in Canada (82.1%). Almost two-thirds (62.7%) of participants lived in food-secure households, 9.0% in marginally food-insecure households, 16.2% in moderately food-insecure households, and 12.2% in severely food-insecure households.

The mean HEI-2015 score was 51.5 ± 15.3 (range: 10.0–97.7 of 100). Participants who were classified in the bottom, middle, and top tertiles had mean HEI-2015 scores of 34.6 ± 6.0, 50.7 ± 3.4, and 69.1 ± 7.5, respectively (Table 2).

CIT analysis

The CIT presented in Figure 2 displays 7 subgroups with different conditional probabilities of having HEI-2015 scores in

TABLE 2
Distribution of Healthy Eating Index-2015 scores and tertiles (n = 5093)

HEI-2015 tertile	*n* (%)	HEI-2015 scores mean (SD)
Bottom	1698 (33.3)	34.6 (6.0)
Middle	1697 (33.3)	50.7 (3.4)
Top	1698 (33.3)	69.1 (7.5)
Total	5093 (100)	51.5 (15.3)

Values are unweighted estimates derived after imputing missing data.
Abbreviation: HEI-2015, Healthy Eating Index-2015.

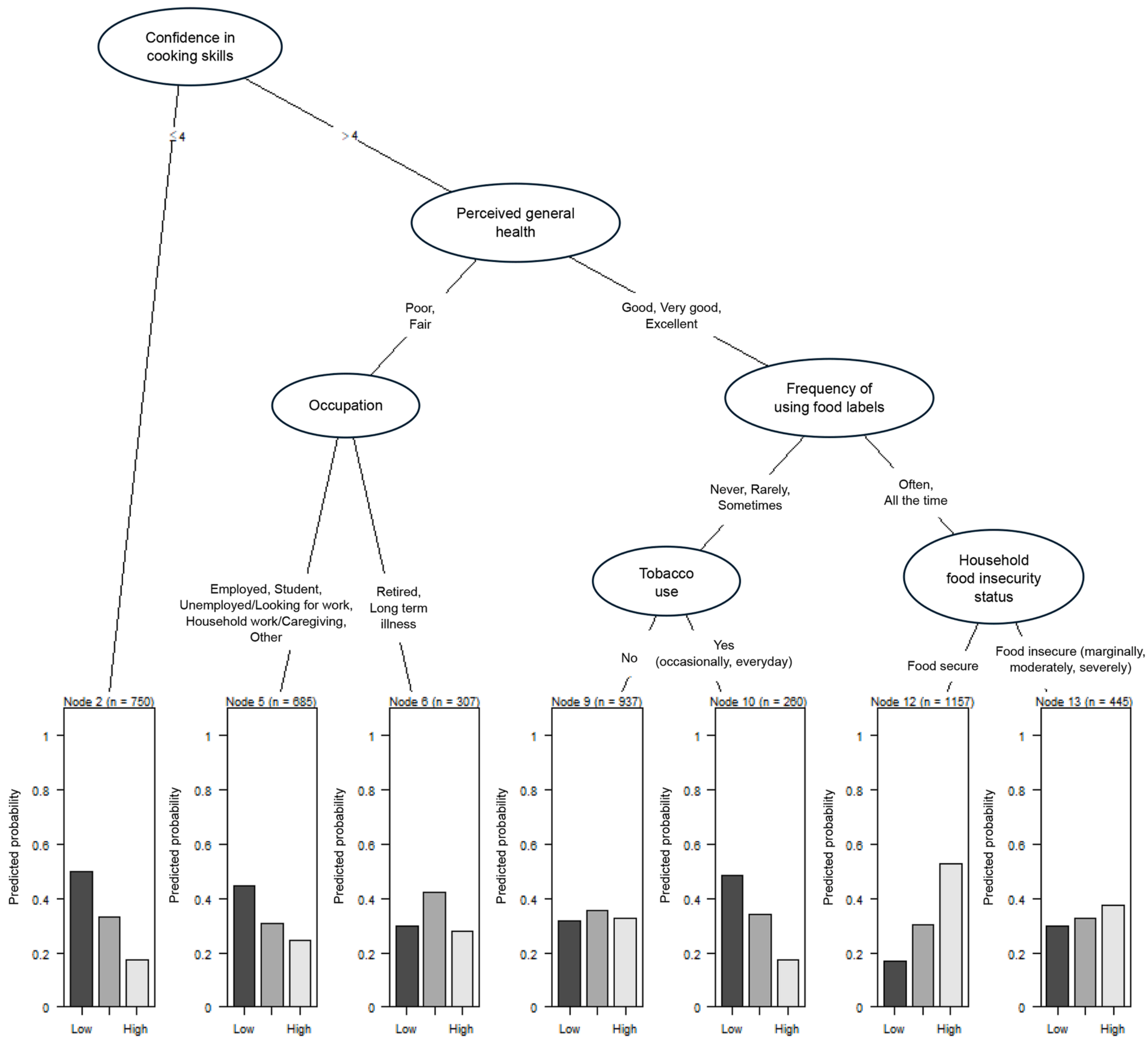

FIGURE 2. Conditional inference tree predicting Healthy Eating Index-2015 tertiles among adults who participated in the 2018 or 2019 International Food Policy Study ($n = 5093$).

the bottom, middle, and top tertiles (see Supplemental Table 3 for specific probability values) due to distinctive combinations of 6 predictors across all 4 domains (sociodemographic characteristics and SEP; food policies and environments; food literacy; and health-related practices and indicators), namely confidence in using cooking techniques, perception of general health, frequency of food label use, tobacco use, occupation, and food insecurity status. Two subgroups had distinctly higher probabilities of having an HEI-2015 score in the bottom tertile. This probability was highest (49.9%) among individuals who were confident in using ≤4 cooking techniques. It was second highest (48.5%) among individuals who were confident in using >4 cooking techniques, perceived their general health as good/very good/excellent, reported never/rarely/sometimes using food labels, and used tobacco occasionally/every day. On the other end of the spectrum, one subgroup had a higher probability of having an HEI-2015 score in the top tertile (52.7%). This subgroup was composed of individuals who were confident in using >4 cooking techniques, perceived their general health as good/very good/excellent, reported often/always using food labels, and lived in food-secure households.

CRF analysis

In total, 20 predictors from the 4 domains were selected by the CRF as the most important predictors of having lower, moderate, or higher HEI-2015 scores (Table 3). The 10 most important predictors, from most to least relative importance, were as follows: frequency of using food labels, confidence in using cooking techniques, perceived general health, household food insecurity status, health literacy, educational attainment, age, frequency of noticing food labels, tobacco use, and whether individuals obtained nutrition information from a government or policy-related source of information.

Model performance

The CIT achieved an overall classification accuracy of 45.1% and 43.7% on the training and test datasets, respectively,

TABLE 3
Top 20 important predictors of Healthy Eating Index-2015 tertiles derived from a conditional random forests model among adults who participated in the 2018 or 2019 International Food Policy Study (n = 5093)

Rank	Predictor	Conditional variable importance
1.	Frequency of food label use	0.0109
2.	Confidence in using cooking techniques	0.0066
3.	Perceived general health	0.0036
4.	Household food insecurity status	0.0025
5.	Health literacy	0.0024
6.	Educational attainment	0.0022
7.	Age	0.0018
8.	Tobacco use	0.0010
9.	Frequency of noticing food labels	0.0008
10.	Obtained nutrition information from government or policy-related sources of information	0.0006
11.	Perceived nutrition knowledge	0.0004
12.	Subjective social status	0.0004
13.	BMI	0.0003
14.	Trust messages from the food and beverage industry on sugary drinks	0.0002
15.	Perceived income adequacy	0.0002
16	Ability to prepare different foods	0.0001
17.	Understand food labels	0.0001
18.	Marijuana use	0.0001
19.	Occupation	0.0001
20.	Looked at Canada's Food Guide	0.0001

exceeding the expected accuracy of a uniform random classifier (~33% in a balanced 3-class setting; see Supplemental Table 4). Additionally, the CIT model obtained a fair agreement based on the κ statistic (0.2), indicating that prediction was better than that obtained by chance. The balanced accuracy (i.e., the mean of the sensitivity and specificity) of the CIT was between 50.9% and 62.3% across the 3 HEI-2015 score tertiles, with the highest accuracy for the top tertile and the lowest accuracy for the middle tertile. The minor discrepancy between the model performance on the training and test datasets suggests that the CIT model was generalizing effectively to new data.

The CRF had an out-of-bag classification accuracy of 47.8% and 49.6% on the training and test datasets, respectively, and obtained a fair agreement based on the κ statistic (0.2 and 0.3). The balanced accuracy ranged between 51.9% and 67.8% across the 3 HEI-2015 score tertiles, with comparable accuracy for the bottom and top tertiles and lower accuracy for the middle tertile. Similar to the CIT, there was no discrepancy between the training and test sets, indicating the absence of overfitting.

Discussion

This study aimed to assess whether machine learning approaches can be leveraged to operationalize a precision public health nutrition approach. Using CIT analyses, we identified 2 subgroups with a higher proportion of adults with lower diet quality. In total, 7 subgroups with different proportions of adults with lower, moderate, or higher diet quality were identified. These subgroups of adults were characterized by distinctive combinations of 6 predictors from 4 domains (sociodemographic characteristics and SEP; food policies and environments; food literacy; and health-related practices and indicators), namely confidence in using cooking techniques, perception of general health, frequency of food label use, tobacco use, occupation, and food insecurity status. We also identified the 20 most important predictors of having lower, moderate, or higher diet quality among adults in Canada using CRF analyses. The top 5 predictors included frequency of food label use, confidence in using cooking techniques, perceived general health, household food insecurity status, and health literacy. Promisingly, 4 of the 6 predictors in the CIT analysis were among the 5 most important predictors in the CRF analysis, and all were among the 20 most important predictors. Consistent findings from the 2 models support the validity of the CIT model [64,65]. Furthermore, the balanced prediction accuracies for CIT and CRF suggest that the models captured patterns that predict the response variable. This interpretation is consistent with the Cohen's κ values, which reflect modest agreement beyond that expected by class prevalence alone.

From a precision public health nutrition perspective, 2 subgroups stood out as having a particularly high risk of lower diet quality. First, the subgroup with the highest proportion of individuals with lower diet quality consisted of individuals who were confident in using ≤4 cooking techniques. Cooking skills are also the second most important predictor of diet quality identified by the CRF model. These results are not surprising because low cooking skills are associated with lower quality dietary patterns [66]. Lower cooking skills have been associated with lower consumption of vegetables, fruit, and homemade meals, as well as with an increased likelihood of eating in restaurants or consuming preprepared foods, many of which are energy-dense and nutrient-poor [67,68].

The subgroup with the second-highest proportion of individuals with lower diet quality was composed of people who were confident in using >4 cooking techniques, perceived themselves as healthy, rarely used food labels, and who smoked. These findings highlight the relevance of examining how cooking skills intersect with other factors to shape individuals' patterns of consumption. As noted above, although poor cooking skills were sufficient on their own to yield lower diet quality, results among this subgroup demonstrate that good cooking skills were not sufficient on their own to yield high diet quality, particularly among those who perceived themselves as healthy, rarely used food labels, and who smoked. Interestingly, some evidence suggests that individuals who consider themselves healthy but who also smoke exhibit fewer health-seeking behaviors [69], a pattern that may extend to healthy eating, which is itself one dimension of health-seeking behavior. Frequency of label use was the third split in the CIT model; it was also the most important predictor of diet quality identified by the CRF model. Food label use has been associated with diet quality [70]. Recognizing the positive impact of food labeling policies on eating patterns, an important area of public health nutrition advocacy in the past years in Canada, has focused on strengthening them [71]. The present study highlights the importance of strong food policies such as easy-to-understand front-of-package labeling. Yet, it provides a more nuanced understanding by suggesting that food label use (a food policy) intersects with cooking skills (an indicator of food literacy) and household food insecurity status (an indicator of material deprivation) among individuals with poorer diet quality. This aligns with existing evidence showing that food policy alone is insufficient to improve diet quality when poor cooking skills or economic constraints persist [20].

It is also instructive to examine the subgroups with a higher proportion of individuals with higher diet quality to identify population subgroups that are a lower priority for intervention. The subgroup with the highest proportion of individuals with a higher diet quality was those who were confident in using >4 cooking techniques, perceived themselves as healthy, often used food labels, and lived in food-secure households. These findings were unsurprising, as evidence indicates that each of these factors is individually associated with healthier dietary patterns [11,72–76]. As such, it seems logical that adults situated at the intersection of these predictors would face fewer barriers to healthy eating. Simply put, they would have more tools (skills, knowledge, and economic resources) to assist them in maintaining a healthier dietary pattern despite pervasive unhealthy food environments.

Consistent with Olstad and McIntyre's (2019) [25] precision public health framework and the literature [10,13,19,77–79], we expected that SEP would play a central role in shaping diet quality, such that SEP indicators would be in the top splits of the CIT. However, this was not the case, as the CIT model did not make any splits based on traditional indicators of SEP, such as perceived income adequacy and educational attainment. These results are coherent with findings from Doan et al. [80], who modeled only traditional SEP indicators that best predicted lower and higher diet quality and found high prediction errors. Yet, household food insecurity status, a potent indicator of material deprivation that is strongly associated with poor diet quality [72,73,81], was one of the predictors identified by the CIT model and the fourth most important predictor in the CRF model. The CRF model also identified indicators of SEP among the most important predictors of diet quality, including perceived income adequacy, educational attainment, subjective social status, and occupation. This is coherent with the literature. For example, educational attainment was the most important predictor of diet quality identified by Doan et al. [80] and has been identified as a super determinant of diet quality by Olstad and McIntyre [82].

Taken together, findings from this study may help inform public health strategies that combine population-level policies to improve food policy and environments (e.g., food labeling regulations) with skill-building programs (e.g., cooking classes and nutrition education) and economic supports (e.g., healthy food subsidies), particularly for subgroups at risk of poor diet quality. This socioecological, multilevel perspective aligns with a precision public health nutrition approach.

Research and policy implications

From a research perspective, our use of machine learning models enabled us to uncover, to our knowledge, novel insights pertaining to groups at higher risk of lower diet quality that would have been difficult or impossible to discern using conventional parametric regression approaches. For example, parametric regression-based models would have required us to specify numerous higher-order interaction terms to approximate the complex, nonlinear relationships captured by our machine learning models. Such models rapidly become implausible to interpret and require very large sample sizes to achieve adequate statistical power, particularly when variables include multiple categories or rare subgroups [26,30,32]. These limitations underscore the value of leveraging similar machine learning approaches in future studies to assess whether comparable subgroup patterns emerge in other contexts [34]. Furthermore, qualitative research with adults situated at the intersection of predictors identified by machine learning approaches may help to better understand the complex array of interconnected factors that shape individuals' dietary patterns. From a policy perspective, precisely identifying population subgroups with a high proportion of adults with lower diet quality across contexts within a precision public health nutrition framework may help inform more precisely tailored interventions that are specific enough to be useful in improving diet quality and reducing dietary inequities [25].

Strengths and limitations

To our knowledge, this is the first study to use CIT and CRF analyses to identify subgroups with a high proportion of adults with lower diet quality using a large number of predictors from all levels of the socioecological model and to assess the relative importance of predictors in shaping diet quality. Two complementary analytical approaches were used to enhance rigor. CIT is a nonparametric and statistically conservative approach that is easy to interpret and avoids multiple shortcomings of conventional parametric regression-based approaches, such as assumptions of linear relationships between variables and multicollinearity causing unreliable results [34,83,84]. CRF is less prone to overfitting and has higher generalizability than CIT, which allowed us to assess the relative importance of predictors [54,56,57]. Although CRF models quantify the relative importance of predictors, conditional variable importance measures do not provide estimates of effect size or direction, or confidence intervals for rankings, nor do they explicitly rank interaction effects. This study did not incorporate a statistical comparative method, such as cluster analysis or a parametric regression-based model. These approaches are not directly comparable to the CIT and CRF models used in the current study, and we did not wish to incorporate a third, conceptually distinct method that did not align with our aim of examining how multiple predictors intersect to shape diet quality.

This study included a large sample of individuals living across the 10 Canadian provinces. Respondents were recruited using non–probability-based sampling, and CIT and CRF analyses cannot accommodate sampling weights, which can influence model performance, particularly for groups that are underrepresented in the data. Thus, the results are not nationally representative. The non–probability-based sampling may also help explain why the rate of household food insecurity in our sample was higher than national estimates for the same period [85]; however, most Canadian surveys are believed to underestimate food insecurity rates [86]. The sociodemographic profile of participants was broadly similar to national data. For example, a similar proportion identified as White (71.1% in our sample vs. 72.8% in the 2016 Census) [87]. However, the sample included fewer immigrants than reported in the Census (17.9% vs. 21.9%) [87].

Categorizing continuous diet quality scores into tertiles rather than modeling them as a continuous outcome may have introduced misclassification bias, especially for individuals near the thresholds. To reduce this potential bias, we excluded individuals whose values were close to each threshold; however, this may have introduced selection bias. As with any statistical or machine learning model, it is possible that we did not include all important predictors or that the predictors identified in the current study may simply be proxies for other unmeasured factors. Finally, we emphasize that this study was designed to identify predictive associations and diet quality-related profiles, rather than to establish causal relationships or support etiologic inference.

Conclusions

In conclusion, this study provides evidence that machine learning approaches can be leveraged to operationalize a precision public health approach. We more precisely identified subgroups with a high proportion of adults with lower diet quality based on a large number of predictors spanning all levels of the socioecological model and identified the most important predictors shaping adults' diet quality. Findings suggest that a socioecological, multilevel approach combining population-level policies to improve food environments with skill-building programs and economic supports tailored to the specific needs of subgroups at higher risk of poor diet quality may help inform strategies to address the social processes leading to poor diet quality. Given that diet quality is strongly associated with morbidity and mortality, more precisely identifying subgroups with a higher risk of lower diet quality and intervening to improve their diet quality is a priority in efforts to improve population health and reduce dietary and health inequities.

Author contributions

The authors' responsibilities were as follows – RB, ND, SN, CEV, DH, LV, DLO: designed the current research; DH, LV: conducted the IFPS; ND, SN, CEV: performed statistical analysis; RB, DLO: wrote the paper; RB, DLO: are primarily responsible for the final content; RB, ND, SN, CEV, DH, LV, DLO: interpreted the data and critically edited the manuscript; and all authors: read and approved the final manuscript.

Conflict of interest

The authors report no conflicts of interest.

Funding

This study did not receive any specific funding. RB was funded by a Canadian Institutes of Health Research Fellowship and currently holds a FRQS Junior 1 Investigator Award (https://doi.org/10.69777/331026). LV is supported by a Canada Research Chair in Healthy Food Policy. The International Food Policy Study was funded by a Canadian Institutes of Health Research (CIHR) Project Grant (PJT-162167), with additional support from an International Health Grant, the Public Health Agency of Canada (PHAC), and a CIHR-PHAC Applied Public Health Chair (DH).

Data availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions. The analytical code used in this study is available on request from the corresponding author.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Copilot in order to improve grammar and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tjnut.2026.101743.

References

[1] Institute for Health Metrics and Evaluation (IHME), Country profile: Canada, Top 10 Risks Contributing to DALYS in 2016 and Percent Change, 2005–2016, All Ages [Internet], Available from: http://www.healthdata.org/canada.
[2] GBD 2017 Diet Collaborators, Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet 393 (10184) (2019) 1958–1972, https://doi.org/10.1016/s0140-6736(19)30041-8.
[3] M. Story, K.M. Kaphingst, R. Robinson-O'Brien, K. Glanz, Creating healthy food and eating environments: policy and environmental approaches, Annu. Rev. Public Health 29 (1) (2008) 253–272, https://doi.org/10.1146/annurev.publhealth.29.020907.090926.
[4] F.M. Stok, S. Hoffmann, D. Volkert, H. Boeing, R. Ensenauer, M. Stelmach-Mardas, et al., The DONE framework: creation, evaluation, and updating of an interdisciplinary, dynamic framework 2.0 of determinants of nutrition and eating, PLoS One 12 (2) (2017) e0171077, https://doi.org/10.1371/journal.pone.0171077.
[5] K.D. Raine, Determinants of healthy eating in Canada: an overview and synthesis, Can, J. Public Health. 96 (Suppl 3) (2005) S8–S15, https://doi.org/10.1007/bf03405195.
[6] G.R. Bauer, S.M. Churchill, M. Mahendran, C. Walwyn, D. Lizotte, A. A. Villa-Rueda, Intersectionality in quantitative research: a systematic review of its emergence and applications of theory and methods, SSM Popul. Health 14 (2021) 100798, https://doi.org/10.1016/j.ssmph.2021.100798.
[7] R. An, Effectiveness of subsidies in promoting healthy food purchases and consumption: a review of field experiments, Public Health Nutr 16 (7) (2013) 1215–1228, https://doi.org/10.1017/S1368980012004715.
[8] A. Avery, C. Anderson, F. McCullough, Associations between children's diet quality and watching television during meal or snack consumption: a systematic review, Matern. Child Nutr. 13 (4) (2017) e12428, https://doi.org/10.1111/mcn.12428.
[9] Y.L. Chu, K.E. Storey, P.J. Veugelers, Involvement in meal preparation at home is associated with better diet quality among Canadian children, J. Nutr. Educ. Behav. 46 (4) (2014) 304–308, https://doi.org/10.1016/j.jneb.2013.10.003.
[10] D. Garriguet, Diet quality in Canada, Health Rep 20 (3) (2009) 41–52.
[11] N.I. Larson, M. Story, M.E. Eisenberg, D. Neumark-Sztainer, Food preparation and purchasing roles among adolescents: associations with sociodemographic characteristics and diet quality, J. Am. Diet. Assoc. 106 (2) (2006) 211–218, https://doi.org/10.1016/j.jada.2005.10.029.
[12] H. Patrick, T.A. Nicklas, A review of family and social determinants of children's eating patterns and diet quality, J. Am. Coll. Nutr. 24 (2) (2005) 83–92, https://doi.org/10.1080/07315724.2005.10719448.
[13] E.L. Faught, L. McLaren, S.I. Kirkpatrick, D. Hammond, L.M. Minaker, K.D. Raine, et al., Socioeconomic disadvantage across the life course is associated with diet quality in young adulthood, Nutrients 11 (2) (2019) 242, https://doi.org/10.3390/nu11020242.
[14] R. Blanchet, N. Willows, S. Johnson, Okanagan Nation Salmon Reintroduction Initiatives, M. Batal, Traditional food, health, and diet quality in Syilx Okanagan adults in British Columbia, Canada, Nutrients 12 (4) (2020) 927, https://doi.org/10.3390/nu12040927.
[15] C.R. Singleton, M. Winkler, B. Houghtaling, O.S. Adeyemi, A.M. Roehll, J.J. Pionke, et al., Understanding the intersection of race/ethnicity, socioeconomic status, and geographic location: a scoping review of U.

S. consumer food purchasing, Int. J. Environ. Res. Public Health. 17 (20) (2020) 7677, https://doi.org/10.3390/ijerph17207677.

[16] R. Monge-Rojas, V. Smith-Castro, T. O'Connor, U. Colón-Ramos, B. R. Fernández, Intersectionality between parenting styles, area of residence and gender on food group consumption among Costa Rican adolescents, Appetite 166 (105443) (2021) 105443, https://doi.org/10.1016/j.appet.2021.105443.

[17] Y. Zhu, M.J. Duan, I.J. Riphagen, I. Minovic, J.O. Mierau, J.J. Carrero, et al., Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis, Int. J. Epidemiol. 50 (6) (2022) 1959–1969, https://doi.org/10.1093/ije/dyab079.

[18] K. Gilham, Q. Gu, T.J.B. Dummer, J.J. Spinelli, R.A. Murphy, Diet quality and neighborhood environment in the Atlantic Partnership for Tomorrow's Health project, Nutrients 12 (10) (2020) 3217, https://doi.org/10.3390/nu12103217.

[19] S.H. Hosseini, M. Farag, S.Z. Hosseini, H. Vatanparast, Behavioral factors are perhaps more important than income in determining diet quality in Canada, SSM Popul, Health 17 (2022) 101001, https://doi.org/10.1016/j.ssmph.2021.101001.

[20] U. Klink, T. Intemann, L.H. Bogl, L. Lissner, W. Gwozdz, S. De Henauw, et al., Consumer attitudes towards dietary behaviors: a mediator between socioeconomic status and diet quality in European adults, Eur. J. Nutr. 64 (3) (2025) 127, https://doi.org/10.1007/s00394-025-03645-6.

[21] N.D. Willows, Determinants of healthy eating in Aboriginal peoples in Canada: the current state of knowledge and research gaps, Can. J. Public Health. 96 (Suppl 3) (2005) S36–S41, https://doi.org/10.1007/BF03405199.

[22] J.P. Taylor, S. Evers, M. McKenna, Determinants of healthy eating in children and youth, Can. J. Public Health. 96 (Suppl 3) (2005) S22–S29, https://doi.org/10.1007/BF03405197.

[23] E.M. Power, Determinants of healthy eating among low-income Canadians, Can. J. Public Health. 96 (Suppl 3) (2005) S42–S48, https://doi.org/10.1007/BF03405200.

[24] A.D.M. Sawyer, F. van Lenthe, C.B.M. Kamphuis, L. Terragni, G. Roos, M.P. Poelman, et al., Dynamics of the complex food environment underlying dietary intake in low-income groups: a systems map of associations extracted from a systematic umbrella literature review, Int. J. Behav. Nutr. Phys. Act. 18 (1) (2021) 96, https://doi.org/10.1186/s12966-021-01164-1.

[25] D.L. Olstad, L. McIntyre, Reconceptualising precision public health, BMJ Open 9 (9) (2019) e030279, https://doi.org/10.1136/bmjopen-2019-030279.

[26] J. Merlo, P. Wagner, G. Leckie, A simple multilevel approach for analysing geographical inequalities in public health reports: the case of municipality differences in obesity, Health Place 58 (2019) 102145, https://doi.org/10.1016/j.healthplace.2019.102145.

[27] G.R. Bauer, Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity, Soc. Sci. Med. 110 (2014) 10–17, https://doi.org/10.1016/j.socscimed.2014.03.022.

[28] S.C. Lemon, J. Roy, M.A. Clark, P.D. Friedmann, W. Rakowski, Classification and regression tree analysis in public health: methodological review and comparison with logistic regression, Ann. Behav. Med. 26 (3) (2003) 172–181, https://doi.org/10.1207/s15324796abm2603_02.

[29] L. Bowleg, When Black + lesbian + woman /= Black lesbian woman: the methodological challenges of qualitative and quantitative intersectionality research, Sex Roles 59 (5) (2008) 312–325, https://doi.org/10.1007/s11199-008-9400-z.

[30] C.R. Evans, Adding interactions to models of intersectional health inequalities: comparing multilevel and conventional methods, Soc. Sci. Med. 221 (2019) 95–105, https://doi.org/10.1016/j.socscimed.2018.11.036.

[31] L. Bowleg, The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health, Am. J. Public Health. 102 (7) (2012) 1267–1273, https://doi.org/10.2105/AJPH.2012.300750.

[32] C.R. Evans, D.R. Williams, J.P. Onnela, S.V. Subramanian, A multilevel approach to modeling health inequalities at the intersection of multiple social identities, Soc. Sci. Med. 203 (2018) 64–73, https://doi.org/10.1016/j.socscimed.2017.11.011.

[33] L.M. Stevens, B.J. Mortazavi, R.C. Deo, L. Curtis, D.P. Kao, Recommendations for reporting machine learning analyses in clinical research, Circ. Cardiovasc. Qual. Outcomes. 13 (10) (2020) e006556, https://doi.org/10.1161/CIRCOUTCOMES.120.006556.

[34] C. Strobl, J. Malley, G. Tutz, An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests, Psychol. Methods 14 (4) (2009) 323–348, https://doi.org/10.1037/a0016973.

[35] A. Mizumoto, Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests, Lang. Learn. 73 (1) (2023) 161–196, https://doi.org/10.1111/lang.12518.

[36] E.S. Dalmaijer, C.L. Nord, D.E. Astle, Statistical power for cluster analysis, BMC Bioinform 23 (1) (2022) 205, https://doi.org/10.1186/s12859-022-04675-1.

[37] D. Hammond, International food policy study 2020 [Internet], Available from: http://foodpolicystudy.com/about/.

[38] The American Association for Public Opinion Research, Standard definitions: final dispositions of case codes and outcome rates for surveys [Internet], 9th edition, AAPOR, 2016. Available from: https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf.

[39] Y. Miller, Recommendations for the truncation of body mass index in population data, Report No. CPAH03-0005, NSW Centre for Physical Activity and Health, 2003.

[40] A.F. Subar, S.I. Kirkpatrick, B. Mittl, T.P. Zimmerman, F.E. Thompson, C. Bingley, et al., 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 (8) (2012) 1134–1137, https://doi.org/10.1016/j.jand.2012.04.016.

[41] ASA24 Canada, ASA24-Canada 2020 [Internet], Available from: https://epi.grants.cancer.gov/asa24/respondent/asa24-canada-2018.html.

[42] S.I. Kirkpatrick, A.F. Subar, D. Douglass, T.P. Zimmerman, F. E. Thompson, L.L. Kahle, et al., Performance of the automated self-administered 24-hour recall relative to a measure of true intakes and to an interviewer-administered 24-h recall, Am. J. Clin. Nutr. 100 (1) (2014) 233–240, https://doi.org/10.3945/ajcn.114.083238.

[43] Y. Park, K.W. Dodd, V. Kipnis, F.E. Thompson, N. Potischman, D. A. Schoeller, et al., Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers, Am. J. Clin. Nutr. 107 (1) (2018) 80–93, https://doi.org/10.1093/ajcn/nqx002.

[44] S.I. Kirkpatrick, P.M. Guenther, D. Douglass, T. Zimmerman, L. L. Kahle, A. Atoloye, et al., The provision of assistance does not substantially impact the accuracy of 24-hour dietary recalls completed using the automated self-administered 24-h dietary assessment tool among women with low incomes, J. Nutr. 149 (1) (2019) 114–122, https://doi.org/10.1093/jn/nxy207.

[45] J.M. Conway, L.A. Ingwersen, B.T. Vinyard, A.J. Moshfegh, Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women, Am. J. Clin. Nutr. 77 (5) (2003) 1171–1178, https://doi.org/10.1093/ajcn/77.5.1171.

[46] A.J. Moshfegh, D.G. Rhodes, D.J. Baer, T. Murayi, J.C. Clemens, W. V. Rumpler, et al., The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes, Am. J. Clin. Nutr. 88 (2) (2008) 324–332, https://doi.org/10.1093/ajcn/88.2.324.

[47] Health Canada, Canadian nutrient file (CNF), 13th Edition, updated 2015 [Internet], Available from: https://aliments-nutrition.canada.ca/cnf-fce/newSearch.

[48] United States Department of Agriculture, Food and nutrient database for dietary studies [Internet], Available from, https://data.nal.usda.gov/dataset/food-and-nutrient-database-dietary-studies-fndds.

[49] United States Department of Agriculture, Food patterns equivalents database [Internet]. Available from: https://data.nal.usda.gov/dataset/food-patterns-equivalents-database-fped.

[50] National Cancer Institute, The Healthy Eating Index SAS code 2020 [Internet]. Available from: https://epi.grants.cancer.gov/hei/sas-code.html.

[51] S.M. Krebs-Smith, T.E. Pannucci, A.F. Subar, S.I. Kirkpatrick, J. L. Lerman, J.A. Tooze, et al., Update of the Healthy Eating Index: HEI-2015, J. Acad. Nutr. Diet. 118 (9) (2018) 1591–1602, https://doi.org/10.1016/j.jand.2018.05.021.

[52] D. Brassard, L.A. Elvidge Munene, S. St Pierre, P.M. Guenther, S. I. Kirkpatrick, J. Slater, et al., Development of the Healthy Eating Food

Index (HEFI)-2019 measuring adherence to Canada's Food Guide 2019 recommendations on healthy food choices, Appl. Physiol. Nutr. Metab. 47 (5) (2022) 595–610, https://doi.org/10.1139/apnm-2021-0415.
[53] C. Strobl, T. Hothorn, A. Zeileis, Party on! A new, conditional variable-importance measure for random forests available in the Party package, R J, 1–2, 2009, pp. 14–17.
[54] T. Hothorn, K. Hornik, A. Zeileis, Unbiased recursive partitioning: a conditional inference framework, J. Comput. Graph. Stat. 15 (3) (2006) 651–674, https://doi.org/10.1198/106186006X133933.
[55] T. Hothorn, A. Zeileis, Partykit: a modular toolkit for recursive partytioning in R, J. Mach. Learn. Res. 16 (118) (2015) 3905–3909.
[56] T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, M.J. van der Laan, Survival ensembles, Biostatistics 7 (3) (2006) 355–373, https://doi.org/10.1093/biostatistics/kxj011.
[57] T. Hothorn, B. Lausen, A. Benner, M. Radespiel-Tröger, Bagging survival trees, Stat, Med. 23 (1) (2004) 77–91, https://doi.org/10.1002/sim.1593.
[58] N. Levshina, Conditional inference trees and random forests, in: M. Paquot, S.T. Gries (Eds.), A Practical Handbook of Corpus Linguistics, Springer International Publishing, Cham, Switzerland, 2020, pp. 611–643.
[59] T. Hothorn, K. Hornik, A. Zeileis, ctree: conditional inference trees [Internet]. Available from: 2016, https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf (Accessed October 2024).
[60] P. Svefors, O. Sysoev, E.C. Ekstrom, L.A. Persson, S.E. Arifeen, R. T. Naved, et al., Relative importance of prenatal and postnatal determinants of stunting: data mining approaches to the MINIMat cohort, Bangladesh, BMJ Open 9 (8) (2019) e025154, https://doi.org/10.1136/bmjopen-2018-025154.
[61] C. Strobl, A.L. Boulesteix, T. Kneib, T. Augustin, A. Zeileis, Conditional variable importance for random forests, BMC Bioinform 9 (1) (2008) 307, https://doi.org/10.1186/1471-2105-9-307.
[62] A. Liaw, M. Wiener, Classification and regression by randomForest, R News 2 (2002) 18–22.
[63] D. Stekhoven, P. Buhlmann, MissForest–non-parametric missing value imputation for mixed-type data, Bioinformatics 28 (1) (2012) 112–118, https://doi.org/10.1093/bioinformatics/btr597.
[64] C. Källestål, E. Blandon Zelaya, R. Peña, W. Peréz, M. Contreras, L. Å. Persson, et al., Predicting poverty. Data mining approaches to the health and demographic surveillance system in Cuatro Santos, Nicaragua, Int. J. Equity Health. 18 (1) (2019) 165, https://doi.org/10.1186/s12939-019-1054-7.
[65] M.R. Winkler, S. Telke, E.Q. Ahonen, M.M. Crane, S.M. Mason, D. Neumark-Sztainer, Constrained choices: combined influences of work, social circumstances, and social location on time-dependent health behaviors, SSM Popul, Health 11 (2020) 100562, https://doi.org/10.1016/j.ssmph.2020.100562.
[66] C. Hartmann, S. Dohle, M. Siegrist, Importance of cooking skills for balanced food choices, Appetite 65 (2013) 125–131, https://doi.org/10.1016/j.appet.2013.01.016.
[67] J.A. Watanabe, J.A. Nieto, T. Suarez-Diéguez, M. Silva, Influence of culinary skills on ultraprocessed food consumption and Mediterranean diet adherence: an integrative review, Nutrition 121 (2024) 112354, https://doi.org/10.1016/j.nut.2024.112354.
[68] C. Hermosa-Bosano, J.F. López-Gil, Low self-perceived cooking skills are linked to greater ultra-processed food consumption among adolescents: the EHDLA study, Nutrients 17 (7) (2025) 1168, https://doi.org/10.3390/nu17071168.
[69] M.A.M. Ismail, M.N. Shariff, M.S. Nurumal, R.A. Rahman, H.H. M. Mokhtar, Smokers' perception of their health status and health-seeking behaviour: a narrative review, Int. J. Care Sch. 5 (3) (2022) 38–43, https://doi.org/10.31436/ijcs.v5i3.261.
[70] S. Campos, J. Doxey, D. Hammond, Nutrition labels on pre-packaged foods: a systematic review, Public Health Nutr 14 (8) (2011) 1496–1506, https://doi.org/10.1017/S1368980010003290.
[71] K.D. Raine, A.R. Ferdinands, K. Atkey, E. Hobin, B. Jeffery, C.I. J. Nykiforuk, et al., Policy recommendations for front-of-package, shelf, and menu labelling in Canada: moving towards consensus, Can. J. Public Health. 108 (4) (2017) 409–413, https://doi.org/10.17269/cjph.108.6076.
[72] K.L. Hanson, L.M. Connor, Food insecurity and dietary quality in US adults and children: a systematic review, Am. J. Clin. Nutr. 100 (2) (2014) 684–692, https://doi.org/10.3945/ajcn.114.084525.
[73] S.I. Kirkpatrick, V. Tarasuk, Food insecurity is associated with nutrient inadequacies among Canadian adults and adolescents, J. Nutr. 138 (3) (2008) 604–612, https://doi.org/10.1093/jn/138.3.604.
[74] J. Farahbakhsh, M. Hanbazaza, G.D.C. Ball, A.P. Farmer, K. Maximova, N.D. Willows, Food insecure student clients of a university-based food bank have compromised health, dietary intake and academic quality, Nutr. Diet. 74 (1) (2017) 67–73, https://doi.org/10.1111/1747-0080.12307.
[75] F. Xu, S.A. Cohen, I.E. Lofgren, G.W. Greene, M.J. Delmonico, M. L. Greaney, Relationship between diet quality, physical activity and health-related quality of life in older adults: findings from 2007-2014 National Health and Nutrition Examination Survey, J. Nutr. Health Aging 22 (9) (2018) 1072–1079, https://doi.org/10.1007/s12603-018-1050-4.
[76] Government of Canada, Improving Cooking and Food Preparation Skills A Synthesis Paper of the Evidence to Inform Program and Policy Development, Government of Canada, Ottawa, Canada, 2010. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/fn-an/alt_formats/pdf/nutrition/child-enfant/cfps-acc-synthes-eng.pdf.
[77] B. Shatenstein, L. Gauvin, H. Keller, L. Richard, P. Gaudreau, F. Giroux, et al., Baseline determinants of global diet quality in older men and women from the NuAge cohort, J. Nutr. Health Aging 17 (5) (2013) 419–425, https://doi.org/10.1007/s12603-012-0436-y.
[78] M. Nardocci, B.S. Leclerc, M.-L. Louzada, C.A. Monteiro, M. Batal, J.-C. Moubarac, Consumption of ultra-processed foods and obesity in Canada, Can, J. Public Health. 110 (1) (2019) 4–14, https://doi.org/10.17269/s41997-018-0130-x.
[79] D.L. Olstad, S. Nejatinamini, C. Victorino, S.I. Kirkpatrick, L. M. Minaker, L. McLaren, Socioeconomic inequities in diet quality among a nationally representative sample of adults living in Canada: an analysis of trends between 2004 and 2015, Am. J. Clin. Nutr. 114 (5) (2021) 1814–1829, https://doi.org/10.1093/ajcn/nqab249.
[80] N. Doan, M.J. Cooke, M.P. Wallace, E. Neiterman, D.L. Olstad, Intersections of educational attainment, Indigenous identity, and race/ethnicity best predicted diet quality among adults in Canada: a conditional random forests analysis, J. Acad. Nutr. Diet. 126 (2) (2026) 156207, https://doi.org/10.1016/j.jand.2025.09.009.
[81] J. Hutchinson, V. Tarasuk, The relationship between diet quality and the severity of household food insecurity in Canada, Public Health Nutr 25 (4) (2022) 1013–1026, https://doi.org/10.1017/s1368980021004031.
[82] D.L. Olstad, L. McIntyre, Educational attainment as a super determinant of diet quality and dietary inequities, Adv. Nutr. 16 (9) (2025) 100482, https://doi.org/10.1016/j.advnut.2025.100482.
[83] S. Nayak, A. Hubbard, S. Sidney, S.L. Syme, A recursive partitioning approach to investigating correlates of self-rated health: the CARDIA study, SSM Popul, Health 4 (2018) 178–188, https://doi.org/10.1016/j.ssmph.2017.12.002.
[84] A. Venkatasubramaniam, J. Wolfson, N. Mitchell, T. Barnes, M. JaKa, S. French, Decision trees in epidemiological research, Emerg. Themes Epidemiol. 14 (1) (2017) 11, https://doi.org/10.1186/s12982-017-0064-4.
[85] V. Tarasuk, A. Mitchel, Household food insecurity in Canada, 2017-18. Toronto: Research to identify policy options to reduce food insecurity, PROOF Food Insecurity Policy Research, Toronto, Canada, 2020, https://proof.utoronto.ca/wp-content/uploads/2020/03/Household-Food-Insecurity-in-Canada-2017-2018-Full-Reportpdf.pdf.
[86] T. Li, A.A. Fafard St-Germain, V. Tarasuk, Household food insecurity in Canada, 2022. Toronto: Research to identify policy options to reduce food insecurity, PROOF Food Insecurity Policy Research, Toronto, Canada, 2023, https://proof.utoronto.ca/wp-content/uploads/2023/11/Household-Food-Insecurity-in-Canada-2022-PROOF.pdf.
[87] Statistics Canada, Census profile, 2016 census. Report No.: Catalogue no. 98-316-X2016001, Government of Canada, Ottawa, Canada, 2017.