Prev. Nutr. Food Sci. 2024;29(2):199-209
https://doi.org/10.3746/pnf.2024.29.2.199
ISSN 2287-8602

Health Star Rating of Nonalcoholic, Packaged, and Ready-to-Drink Beverages in Türkiye: A Decision Tree Model Study

Aylin Bayındır Gümüş[1], Murat Açık[2], and Sevinç Eşer Durmaz[3]

[1]First and Emergency Aid Program, Vocational School of Health Services and [3]Department of Nutrition and Dietetics, Faculty of Health Sciences, Kırıkkale University, Kırıkkale 71450, Türkiye
[2]Department of Nutrition and Dietetics, Faculty of Health Sciences, Fırat University, Elazığ 23200, Türkiye

ABSTRACT: This study aimed to compare the nutritional quality of beverages sold in Türkiye according to their labeling profiles. A total of 304 nonalcoholic beverages sold in supermarkets and online markets with the highest market capacity in Türkiye were included. Milk and dairy products, sports drinks, and beverages for children were excluded. The health star rating (HSR) was used to assess the nutritional quality of beverages. The nutritional quality of beverages was evaluated using a decision tree model according to the HSR score based on the variables presented on the beverage label. Moreover, confusion matrix tests were used to test the model's accuracy. The mean HSR score of beverages was 2.6±1.9, of which 30.2% were in the healthy category (HSR≥3.5). Fermented and 100% fruit juice beverages had the highest mean HSR scores. According to the decision tree model of the training set, the predictors of HSR quality score, in order of importance, were as follows: added sugar (46%), sweetener (28%), additives (19%), fructose-glucose syrup (4%), and caffeine (3%). In the test set, the accuracy rate and F1 score were 0.90 and 0.82, respectively, suggesting that the prediction performance of our model had the perfect fit. According to the HSR classification, most beverages were found to be unhealthy. Thus, they increase the risk of the development of obesity and other diseases because of their easy consumption. The decision tree learning algorithm could guide the population to choose healthy beverages based on their labeling information.

Keywords: beverages, food quality, nutrient intake, nutritive value, sugar-sweetened beverages

INTRODUCTION

Diet-related noncommunicable diseases (NCDs), which are also known as chronic diseases (e.g., obesity, tip 2 diabetes, cardiovascular diseases), tend to be of long duration and are major contributors to disease burden in high- and middle-income countries (WHO, 2023). These diseases are triggered by many environmental factors, especially unhealthy lifestyle habits and obesogenic environments. Because of the increased consumption of packaged and processed foods, the incidence of chronic diseases is increasing (Forouzanfar et al., 2016). The Türkiye Nutrition and Health Survey (TNHS) 2017 revealed that the prevalence of any chronic disease in individuals aged 19 years and over was 44% (Republic of Turkey Ministry of Health, 2019). The increasing accessibility of packaged products also leads to poor diet quality, which increases the risk of obesity and other diet-related NCDs (Crino et al., 2018; Vergeer et al., 2020). According to the Türkiye Ministry of Health, the mean daily consumption of water/mineralized water/soda, black tea, coffee, and soft drinks was 1,169.9±819.98, 416.4±403.37, 26.2±79.49, and 1,721.8 ±922.94 mL, respectively (Republic of Turkey Ministry of Health, 2019). Thus, making packaged beverages healthier will significantly reduce the burden of obesity and diet-related diseases nationally and globally by reducing the consumption of nutrients of concern (Dunford et al., 2019).

Nutrient profiling (NP) is used to classify or rank foods and beverages according to their nutrient composition for disease prevention and health promotion. Thus, it can be used to determine whether foods and beverages are healthy (Rayner et al., 2013). NP is designed to characterize individual foods, not diets. However, NP models are widely used to support policies designed to improve the overall nutritional quality of food supplies. The World Health Organization (WHO) has considered NP as a useful method that can be used in combination with other interventions to improve the overall nutritional quality of diets (WHO, 2022). However, there is no international

Received 14 February 2024; Revised 25 March 2024; Accepted 28 March 2024; Published online 30 June 2024
Correspondence to Sevinç Eşer Durmaz, E-mail: sevincceser@gmail.com

consensus regarding the superiority of a particular NP model. The latest WHO catalog lists more than 65 current NP models. According to a recent literature review, 78 models have been introduced in the last decade (Labonté et al., 2018). The Australasian health star rating (HSR) system is one of the most widely used nutrient profile models to assess the healthfulness of packaged food and beverage products (Dunford et al., 2019).

Providing consumers with tools to evaluate the nutritional quality of beverages represents an important public health initiative. Because of its summary indicator, HSR can be used to quickly compare similar products, which is expected to be particularly useful in the grocery shopping environment (Pelly et al., 2020). The HSR system is a useful tool for communicating health and nutrition messages to consumers. In short, it aims to provide appropriate, relevant, and easily understood nutrition information and/or guidance on food packs to help consumers make informed food purchases and healthier dietary choices (Jones et al., 2018). According to TNHS-2017, approximately 4.9% and 6.2% of individuals aged 19～64 years in Türkiye consume ready-to-drink fruit juices and carbonated beverages every day, respectively. In addition, the daily consumption of water and other beverages was higher in TNHS-2017 than in TNHS-2010 (Republic of Turkey Ministry of Health, 2019). Increasing the frequency and amount of consumption of ready-to-drink beverages other than water increases the energy intake of individuals and the risk of obesity and chronic diseases (Guzman-Vilca et al., 2022). Considering the contribution of packaged ready-to-drink beverages to energy intake in the Turkish diet and the adverse health consequences associated with their frequent consumption, the representation of the HSR system in beverages needs to be investigated. Therefore, in the present study, we aimed to analyze the nutrient content of beverage labels sold in Türkiye using HSR. Moreover, after categorizing the information on the food label, we aimed to predict the nutritional quality of beverages according to HSR scores by performing machine learning decision tree model analyses.

MATERIALS AND METHODS

Study type

In this market research study, 12 supermarkets and online markets with the highest market capacity that offered beverages in Türkiye were included. The beverages were examined between February and May 2023.

Type and classification of beverages

Nonalcoholic, packaged, and ready-to-drink beverages were included. Meanwhile, sports drinks, milk, and dairy products (kefir and ayran), plant-based milk, sparkling water, nonenergy drinks, and beverages for children were excluded. Beverages were categorized as fruit beverages [100% fruit juices (100% fruit content), fruit nectars (fruit content 25%～99%), fruity beverages (fruit content 10%～24%), aromatic beverages (fruit content<10%), and lemonades], iced teas, cold coffees, carbonated beverages, fermented beverages, and energy drinks. The labels were used to obtain the energy (kcal), carbohydrate (g), sugar (g), protein (g), total fat (g), saturated fatty acid (g), dietary fiber (g), and salt (g) contents in 100 mL beverages. The ingredient sections were examined, and the presence of added sugar, sugar alcohol, artificial sweeteners (AS), fructose-glucose syrup, E-coded additives (EA), and additional vitamins/minerals were recorded. Only one beverage that had the same ingredients in different packaging sizes and that was available in more than one market was included.

Data collection tool

The nutritional quality of beverage products was primarily assessed and ranked based on the HSR system, an interpretive front-of-package labeling system designed to help consumers select healthier beverages (Jones et al., 2018). This system is based on the NP scoring criterion developed by Food Standards Australia New Zealand for regulating food health claims in Australia and New Zealand. The HSR has been externally validated and applied worldwide (Jones et al., 2018).

HSR was calculated following the methods described in the “Guide for Industry to the Health Star Rating Calculator” (Australian Government Department of Health and Aged Care, 2020). In summary, the types of beverages were categorized into one of two HSR categories (i.e., nondairy, or dairy beverages). Then, the HSR score was calculated based on the energy and total sugar in 100 mL of nondairy beverages. The HSR levels of nondairy beverages were calculated based on the energy, saturated fat, total sugar, and sodium content per 100 mL. The HSR was calculated by (1) assigning baseline points for energy, saturated fat, total sugars, and sodium content per 100 mL; (2) calculating the overall score by subtracting modifying points from baseline points, with a lower score reflecting a more nutritious beverage or food product; and (3) assigning an HSR (from 0.5 to 5.0 stars in half-star increments) according to the overall score using the defined scoring matrix (Australian Government Department of Health and Aged Care, 2020). The higher the HSR score, the healthier the beverage or food product. Products with an HSR of ≥3.5 were considered “healthy” based on work by the New South Wales Ministry of Health in Australia. Beverages with an HSR of <3.5 were regarded as “unhealthy” as they were not compatible with healthy dietary guidelines (Dunford et al., 2015).

Statistical analysis and machine learning model (decision trees)

All statistical analyses were conducted using JASP Statistical Software (version 0.18.2, https://jasp-stats.org). The beverages were divided into two categories according to the HSR classification (unhealthy beverages for an HSR of <3.5 and healthy beverages for an HSR of ≥3.5) and 11 groups according to beverage categories [fruit beverages, 100% fruit juices, fruit nectars (25%∼99%), fruity beverages (10%∼24%), aromatic beverages (<10%), lemonades, iced teas, cold coffees, carbonated beverages, fermented beverages, energy drinks, and others (vegetable juice, herbal {detox} or mixed beverages)]. In addition, the beverages were classified as categorical variables according to the presence of AS, EA, added sugar, fructose-glucose syrup, and sugary alcohol. The energy, carbohydrate, sugar, protein, and fat contents of beverages were calculated per 100 mL. During the comparison of nutrients between groups, beverages that did not contain these nutrients were excluded from the analyses. Categorical data are shown as numbers and percentages, whereas numerical variables are shown as mean±standard deviation, and median (interquartile range, IQR). For comparisons of nominal data between HSR groups and beverage types, Pearson's chi-square test was used if the expected frequency was <5% in <25% of cells. Moreover, the Mann-Whitney U-test was performed to compare numerical data (some nutrients) according to HSR groups and beverage types (Van Belle et al., 2004).

The decision tree model of machine learning was created using the open-source software R studio (version 3.6.3, http://www.rproject.org). The classification and regression trees (CART) algorithm were used to construct a decision tree model to determine healthy and unhealthy beverages according to the HSR classification (classification tree, minimum split=20, maximum depth=5, and minimum bucket=7). The CART algorithm was selected because it can deal with multiclass outcomes, reduce feature numbers (through tree pruning), and automatically define thresholds for continuous variables. Furthermore, the decision tree is more appropriate and can be interpreted more easily compared with other machine learning algorithms because the predictors include categorical data. First, we divided the original data into training and testing data according to 8:2, indicating that 80% (n=243) of the original dataset was used for training the decision tree model, whereas the remaining 20% (n=61) was used for model testing. Then, we used the training and testing data to train the decision tree model and to verify the model, respectively. To optimize the decision tree model, a complexity parameter was used for tree pruning to control the tree size. The optimum complexity parameter was found based on the highest accuracy and kappa value to determine the optimal model (i.e., obtain a lower relative error and smaller size of tree). The values were 0.85 for accuracy, 0.72 for kappa, and 0.028 for the complexity parameter. Based on these values, the maximum depth was determined as 5. After optimal model selection, we generated a confusion matrix to test the model's performance, and we calculated the model's accuracy to validate the classification performance. To evaluate the prediction of the decision tree model, several indicators were used, including accuracy, no information rate (NIR), kappa value, McNemar's test value, sensitivity, specificity, positive prediction value, negative prediction value, precision recall, and F1 score (Jijo and Abdulazeez, 2021; Greener et al., 2022).

RESULTS

Characteristics of beverages

The number of products, mean and median values, and HSR classification for each beverage type are presented in Table 1. A total of 304 beverages were included, of which 85.2% (n=259) were produced by domestic companies. Carbonated and fruit beverages had the highest proportion of domestic products. All fermented and other types of beverages were imported. The median HSR scores of classes of beverages ranged from 1.0 to 7.0. Fermented beverages and 100% fruit juices had the highest HSR score (median=7.0 and 4.0, respectively), whereas other fruit beverages, energy drinks, and cold coffees had the lowest HSR score (all median=1.0). In addition, fermented beverages and 100% fruit juices had the highest proportion of products with an HSR of ≥3.5 (80.0% and 57.5%, respectively), whereas fruit beverages with a fruit content of 10%∼24% and fruit nectars had the lowest (0.0% and 7.0%, respectively).

HSR classification of beverages

The distribution of origin, added sugar, sugar alcohol, and fructose-glucose syrup status according to HSR classification among beverage types is shown in Table 2. As can be observed, 48% (n=12) of cold coffees were imported; among them, 25% (n=3) were in the HSR≥3.5 group. Only 7.7% (n=1) of local cold coffees were in the HSR≥3.5 group. For carbonated beverages, 44.1% had an HSR of ≥3.5; among them, 11.8% had added sugar. The proportional distribution of HSR≥3.5 was statistically lower in carbonated beverages with added sugar than in those without added sugar (x^2 test value=28.870 and P<0.001). The proportion of iced teas containing fructose-glucose syrup was 53.8% (n=7) in the HSR<3.5 group and 46.2% (n=6) in the HSR≥3.5 group. Based on chi-square analysis, the distribution of fructose-glucose syrup was similar between HSR groups (x^2=0.304 and P=0.581). However, the proportion of carbonated

Table 1. Number of brands and products by classifying the beverages included in the study

Type of beverage	Product	Domestic company	HSR	HSR	HSR≥3.5
Fruit beverage					
100% fruit juice	40	40 (100)	3.7±0.7	4.0 (3.0~4.0)	23 (57.5)
Fruit nectar (25%~99%)	43	41 (95.3)	1.6±1.0	1.0 (1.0~2.0)	3 (7.0)
Fruity beverage (10%~24%)	41	36 (87.8)	1.3±0.5	1.0 (1.0~1.5)	-
Aromatic beverage (<10%)	17	17 (100)	2.4±2.0	1.0 (1.0~4.0)	5 (29.4)
Lemonade	15	15 (100)	3.0±2.7	1.0 (1.0~7.0)	6 (40.0)
Iced tea	27	26 (96.3)	3.1±1.7	3.0 (2.0~4.0)	11 (40.7)
Cold coffee	25	13 (52.0)	2.0±1.6	1.0 (1.0~3.0)	4 (16.0)
Carbonated beverage	68	68 (100)	3.2±2.4	2.0 (1.0~6.0)	30 (44.1)
Fermented beverage	5	-	5.6±2.1	7.0 (3.5~7.0)	4 (80.0)
Energy beverage	18	3 (16.7)	2.2±2.1	1.0 (1.0~4.0)	5 (27.8)
Others[1)]	5	-	2.6±1.9	3.0 (1.0~3.5)	1 (20.0)
In total	304	259 (85.2)	2.6±1.9	2.0 (1.0~4.0)	92 (30.3)

Values are presented as number only, number (%), mean±SD, or median (interquartile range).
HSR, health star rating.
[1)]Vegetable juice, herbal (detox), or mixed beverages.

beverages containing fructose-glucose syrup was statistically lower in the HSR≥3.5 group than in HSR<3.5 group (P=0.002). Only 2.3% (n=7) of all beverages contained sugar alcohol.

The proportions of AS, additives, and vitamin fortification by beverage type and HSR category are shown in Table 3. As can be observed, the main beverages that contained AS included carbonated beverages, iced teas, and fruit beverages. The proportional distribution of HSR≥3.5 was higher in carbonated beverages containing AS than in those without AS (x^2=28.354 and P<0.001). Except for 100% fruit juices (n=8) and others (n=1), most beverage categories contained additives. The proportional distribution of the HSR≥3.5 group was similar in 100% fruit juices with food additives than in those without food additives (x^2=0.102 and P=0.749). Approximately 70.0% of vitamin/mineral-enriched 100% fruit juices had an HSR of ≥3.5, whereas 53.3% of nonenriched 100% fruit juices had an HSR of ≥3.5 (x^2=0.853 and P=0.356). Although the proportion of HSR≥3.5 was higher in vitamin/mineral-enriched iced teas (46.7%) and carbonated beverages (45.7%) than in the nonenriched group, this difference was not statistically significant (x^2=0.491 and 0.075, P=0.484 and 0.785, respectively).

Energy and nutrient contents of beverages based on HSR classification

The mean, median, and IQR of energy, some nutrients, and added sugar for all beverages according to the HSR classification are shown in Table 4. The median energy value of the HSR<3.5 group [median (IQR)=47.0 (42.0~53.0)] was higher than that of the HSR≥3.5 group [median (IQR)=19.0 (3.3~40.8)]. Furthermore, carbohydrates and added sugars were statistically lower in the HSR≥3.5 group than in the HSR<3.5 group (P<0.001). The protein, total fat, fiber, and sodium contents of beverages were very low. However, the median values of variables were similar between groups (P>0.05), except for total fat (P<0.05).

The median values of energy, carbohydrate, and added sugar were compared between HSR groups according to beverage type (Table 5). In all beverage subcategories, the energy value of the HSR<3.5 group was higher than that of the HSR≥3.5 group, except for fruit nectars (P<0.01). Lemonade and carbonated beverages had the highest energy differences between groups (median difference=48.7 and 41.2 kcal/100 mL, respectively). Excluding fruit nectars, the carbohydrate content of the HSR<3.5 group was higher than that of the HSR≥3.5 group in all beverage subgroups. While 100% fruit juices did not contain added sugar, the added sugar content of the HSR ≥3.5 group was statistically lower than that of the HSR <3.5 group in aromatic beverages, lemonade, iced teas, carbonated beverages, and energy drinks (P<0.001). Aromatic beverages and lemonades had the highest median difference in the amount of added sugar (median difference=12.0 and 11.4 g/100 mL, respectively).

The proportions of beverages containing more than 5 g of added sugar and total sugar per 100 mL were 184 (60.5%) and 236 (77.6%), respectively. While 97.8% of healthy beverages contained less than 5 g of added sugar, only 14.2% of unhealthy beverages contained less than 5 g of added sugar. Approximately 98.6% (n=209) of beverages in the HSR<3.5 group had more than 5 g of total sugars, whereas 29.3% (n=27) of beverages in the HSR ≥3.5 group had more than 5 g of total sugars. The proportion of added and total sugars was lower in healthy beverages than in unhealthy beverages, and the difference was significant (P<0.001). In addition, 93.1% of beverages were in the very-low energy density (ED) category, indicating that the ED of healthy beverages was lower than that of unhealthy beverages (P<0.001) (Table 6).

Table 2. Distribution of simple sugars and origin in the type of beverage based on HSR classification

Type of beverage	Origin		Added sugar		Fructose-glucose syrup		Sugary alcohol	
	Domestic	Import	Yes	No	Yes	No	Yes	No
Fruit beverage								
100% fruit juice								
HSR<3.5	17 (42.5)	–	–	17 (42.5)	–	17 (42.5)	–	17 (42.5)
HSR≥3.5	23 (57.5)	–	–	23 (57.5)	–	23 (57.5)	–	23 (57.5)
Fruit nectar								
HSR<3.5	39 (95.1)	1 (50.0)	36 (100)	4 (57.1)	31 (96.9)	9 (81.8)	–	40 (93.0)
HSR≥3.5	2 (4.9)	1 (50.0)	–	3 (42.9)	1 (3.1)	2 (18.2)	–	3 (7.0)
Fruity beverage								
HSR<3.5	36 (100)	5 (100)	40 (100)	1 (100)	19 (100)	22 (100)	4 (100)	37 (100)
HSR≥3.5	–	–	–	–	–	–	–	–
Aromatic beverage								
HSR<3.5	12 (70.6)	–	12 (85.7)	–	3 (75.0)	9 (69.2)	–	12 (70.6)
HSR≥3.5	5 (29.4)	–	2 (14.3)	3 (100)	1 (25.0)	4 (30.8)	–	5 (29.4)
Lemonade								
HSR<3.5	9 (60.0)	–	9 (81.8)	–	–	9 (60.0)	1 (50.0)	8 (61.5)
HSR≥3.5	6 (40.0)	–	2 (18.2)	4 (100)	–	6 (40.0)	1 (50.0)	5 (38.5)
Iced tea								
HSR<3.5	16 (61.5)	–	16 (66.7)	–	7 (53.8)	9 (64.3)	–	16 (59.3)
HSR≥3.5	10 (38.5)	1 (100)	8 (33.3)	3 (100)	6 (46.2)	5 (35.7)	–	11 (40.7)
Cold coffee								
HSR<3.5	12 (92.3)	9 (75.0)	20 (100)	1 (20.0)	–	21 (84.0)	–	21 (84.0)
HSR≥3.5	1 (7.7)	3 (25.0)	–	4 (80.0)	–	4 (16.0)	–	4 (16.0)
Carbonated beverage								
HSR<3.5	38 (55.9)	–	30 (88.2)	8 (23.5)	22 (78.6)	16 (40.0)	1 (100)	37 (55.2)
HSR≥3.5	30 (44.1)	–	4 (11.8)	26 (76.5)	6 (21.4)	24 (60.0)	–	30 (44.8)
Fermented beverage								
HSR<3.5	–	1 (20.0)	1 (100)	–	–	1 (20.0)	–	1 (20.0)
HSR≥3.5	–	4 (80.0)	–	4 (100)	–	4 (80.0)	–	4 (80.0)
Energy beverage								
HSR<3.5	2 (66.7)	11 (73.3)	13 (86.7)	–	4 (66.7)	9 (75.0)	–	13 (72.2)
HSR≥3.5	1 (33.3)	4 (26.7)	2 (13.3)	3 (100)	2 (33.3)	3 (25.0)	–	5 (27.8)
Others[1)]								
HSR<3.5	–	4 (80.0)	1 (100)	3 (75.0)	1 (100)	3 (75.0)	–	4 (80.0)
HSR≥3.5	–	1 (20.0)	–	1 (25.0)	–	1 (25.0)	–	1 (20.0)

Values are presented as number (%).
Pearson's chi-square test was performed between added sugar status and carbonated beverages, fructose-glucose syrup status and iced teas, and fructose-glucose syrup status and carbonated beverages.
HSR, health star rating.
[1)]Vegetable juice, herbal (detox), or mixed beverages.

Decision tree model

Fig. 1 shows the HSR quality score with a decision tree based on the sugar, sweetener, additive, vitamin/mineral, and caffeine status of beverages. The predictors of HSR quality score, in order of importance, were as follows: added sugar (46%), sweetener (28%), additives (19%), fructose-glucose syrup (4%), and caffeine (3%). In the first branch, the status of added sugar in beverages was analyzed. The training set data contained 243 beverages, of which 69% were unhealthy beverages (HSR<3.5). If beverages contained added sugar and no sweetener, they were predicted to be in the unhealthy beverage category with a prediction rate of 99%. In the other branch, if the beverages with sweeteners did not contain fructose-glucose syrup, 71.4% of beverages were in the healthy category (HSR≥3.5). If beverages did not contain added sugar, the tree again asked whether there was a sweetener. If there was no sweetener, the beverage was predicted to be in the healthy category with a probability of 90%. When beverages contained sweeteners and fructose-glucose syrup, they were classified as unhealthy with a prediction rate of 85.7%.

We constructed a confusion matrix to show the prediction efficacy of the decision tree in more detail (Table 7). According to the HSR classification of beverages in the test set, the accuracy rate was 90%. Moreover, the accuracy of the established model was satisfactory as the *P*-value of accuracy> NIR (P<0.001) was statistically significant. The prediction performance was 93% (specificity) for unhealthy beverages and 82% (sensitivity) for healthy

Table 3. Distribution of sweeteners, additives, and vitamin/mineral enrichment in beverages based on HSR classification

Type of beverage	Artificial sweetener		E-coded additive		Enriched with vitamin/mineral	
	Yes	No	Yes	No	Yes	No
Fruit beverage						
100% fruit juice						
HSR<3.5	–	17 (42.5)	3 (37.5)	14 (43.8)	3 (30.0)	14 (46.7)
HSR≥3.5	–	23 (57.5)	5 (62.5)	18 (56.2)	7 (70.0)	16 (53.3)
Fruit nectar						
HSR<3.5	4 (100)	36 (92.3)	38 (97.4)	2 (50.0)	21 (91.3)	19 (95.0)
HSR≥3.5	–	3 (7.7)	1 (2.6)	2 (50.0)	2 (8.7)	1 (5.0)
Fruity beverage						
HSR<3.5	10 (100)	31 (100)	41 (100)	–	29 (100)	12 (100)
HSR≥3.5	–	–	–	–	–	–
Aromatic beverage						
HSR<3.5	–	12 (80.0)	12 (80.0)	–	6 (75.0)	6 (66.7)
HSR≥3.5	2 (100)	3 (20.0)	3 (20.0)	2 (100)	2 (25.0)	3 (33.3)
Lemonade						
HSR<3.5	–	9 (90.0)	6 (50.0)	3 (100)	6 (50.0)	3 (100)
HSR≥3.5	5 (100)	1 (10.0)	6 (50.0)	–	6 (50.0)	–
Iced tea						
HSR<3.5	2 (18.2)	14 (87.5)	16 (66.7)	–	8 (53.3)	8 (66.7)
HSR≥3.5	9 (81.8)	2 (12.5)	8 (33.3)	3 (100)	7 (46.7)	4 (33.3)
Cold coffee						
HSR<3.5	–	21 (91.3)	19 (86.4)	2 (66.7)	–	21 (84.0)
HSR≥3.5	2 (100)	2 (8.7)	3 (13.6)	1 (33.3)	–	4 (16.0)
Carbonated beverage						
HSR<3.5	7 (21.9)	31 (86.1)	38 (59.4)	–	19 (54.3)	19 (57.6)
HSR≥3.5	25 (78.1)	5 (13.9)	26 (40.6)	4 (100)	16 (45.7)	14 (42.4)
Fermented beverage						
HSR<3.5	–	1 (20.0)	–	1 (50.0)	–	1 (20.0)
HSR≥3.5	–	4 (80.0)	3 (100)	1 (50.0)	–	4 (80.0)
Energy beverage						
HSR<3.5	2 (28.6)	11 (100)	9 (64.3)	4 (100)	11 (68.8)	2 (100)
HSR≥3.5	5 (71.4)	–	5 (35.7)	–	5 (31.2)	–
Others[1)]						
HSR<3.5	1 (100)	3 (75.0)	1 (100)	3 (75.0)	1 (50.0)	3 (100)
HSR≥3.5	–	1 (25.0)	–	1 (25.0)	1 (50.0)	–

Values are presented as number (%).
Pearson's chi-square test was performed between sweetener status and carbonated beverages, E-coded additive and 100% fruit juices, enriched with vitamin/mineral and 100% fruit juices, enriched with vitamin/mineral and iced teas, and enriched with vitamin/mineral and carbonated beverages.
HSR, health star rating.
[1)]Vegetable juice, herbal (detox), or mixed beverages.

beverages. Furthermore, the F1 score was 0.82, which provides information about the model's precision and recall. Hence, we can say that the prediction performance of our model is good based on the results of the confusion matrix.

DISCUSSION

This study provides a pioneering comprehensive examination of the nutritional quality of nonalcoholic and ready-to-drink beverages offered in Turkish markets. According to the Türkiye Nutrition Guide, beverages are expected to contribute to an adequate and balanced diet. Moreover, the selection of healthy beverages is an important step in maintaining a healthy weight and preventing obesity (Türkiye Nutrition Guide, 2022). According to studies conducted in Türkiye, the consumption of beverages is at a high level. A previous study showed that approximately 35% of the total daily liquid intake of 3,411 adults comprised beverages other than water, and a significant number of these beverages were carbonated soft beverages and commercial fruit juice (Nergiz-Unal et al., 2017). These findings show the necessity to determine healthy and unhealthy packaged food and beverages to provide a clear comprehensive summary of their nutritional information, which is defined as NP (WHO, 2011). The HSR is a system that uses a NP algorithm to evaluate the over-

Table 4. Mean and median values of energy, dietary fiber, and sodium per 100 mL of beverages based on HSR classification

Energy and nutrient	HSR<3.5	HSR≥3.5	Total	*P*-value[1)]	U-test
Energy (kcal)	n=212	n=92	n=304	<0.001**	2,495.0
Mean±SD	47.9±14.0	20.7±18.0	39.6±19.7		
Median (IQR)	47.0 (42.0~53.0)	19.0 (3.3~40.8)	45.0 (29.2~51.8)		
Carbohydrate (g)	n=212	n=92	n=304	<0.001**	2,435.0
Mean±SD	11.1±3.2	4.6±4.2	9.1±4.6		
Median (IQR)	11.0 (9.1~12.5)	3.9 (0.5~8.2)	10.4 (6.3~12.0)		
Sugar (g)	n=212	n=92	n=304	<0.001**	1,987.0
Mean±SD	8.1±4.2	1.2±1.8	6.0±4.9		
Median (IQR)	8.5 (6.2~11.0)	0.0 (0.0~2.5)	6.4 (0.1~9.9)		
Protein (g)	n=58	n=37	n=95	0.576	100.0
Mean±SD	1.0±0.9	0.7±0.6	0.9±0.8		
Median (IQR)	0.5 (0.2~1.4)	0.5 (0.0~0.9)	0.5 (0.2~1.2)		
Total fat (g)	n=33	n=17	n=50	0.011*	157.0
Mean±SD	0.9±0.7	0.5±0.6	0.8±0.7		
Median (IQR)	0.8 (0.5~1.2)	0.5 (0.1~0.5)	0.5 (0.2~1.1)		
Fiber (g)	n=29	n=19	n=48	0.405	236.5
Mean±SD	0.7±0.6	0.5±0.3	0.6±0.5		
Median (IQR)	0.7 (0.1~1.0)	0.5 (0.2~0.6)	0.5 (0.1~0.9)		
Sodium (g)	n=56	n=25	n=81	0.161	836.0
Mean±SD	0.09±0.1	0.3±0.5	0.2±0.3		
Median (IQR)	0.05 (0.01~0.10)	0.07 (0.03~0.20)	0.06 (0.01~0.10)		

[1)]Mann-Whitney U-test.
*$P<0.05$, **$P<0.001$.
HSR, health star rating; IQR, interquartile range.

all nutritional quality of packaged foods and beverages (Vergeer et al., 2020).

The mean HSR score of beverages examined in this study was 2.6 stars, indicating that these beverages are unhealthy. Moreover, only 30.2% of beverages had an HSR of ≥3.5. In a recent study conducted in Türkiye, the results were different. According to the study 92.6% of 191 nonalcoholic beverages were evaluated as "healthy," and the mean HSR was 3.5 stars (Bayram and Ozturkcan, 2021). This is thought to be because the number of beverages included in the studies is different and the variety and number of these products in the markets are rapidly increasing. Conversely, in a previous report on the nutritional profile of food and beverages marketed by the 21 largest global companies in nine countries, the mean HSR of beverages was 2.5, and two companies had the highest mean HSR; this is because they marketed 100% fruit juices, bottled waters, or dairy-based beverages, and 32% of overall beverages were considered as "healthy" (Dunford and Taylor, 2018). In the present study, the first two beverages that had the highest HSR score included fermented beverages (5.6 stars) and 100% fruit juices (3.7 stars). Vegetable juices and dairy beverages that were included in this study could not be evaluated because they were very limited in the market. Fermented beverages have the highest HSR score because they are made from vegetables. It is stated that for nondairy beverages with a fruit, vegetable, nut, or legume (FVNL) content ≥40%, FVNL is the main driver of the HSR. However, the loss of positive nutrients that may occur during the production of fruit juices should also be considered. In addition, fruits should mostly be consumed fresh and raw because of the lower fiber content and acidity of fruit juice, increasing the risk of dental erosion (HSR Technical Advisory Group, 2018).

Evidence suggests that the consumption of sugar-sweetened beverages (SSBs) is associated with weight gain, obesity, risk of coronary heart disease, and type 2 diabetes (Malik et al., 2006; Keller and Bucher Della Torre, 2015; Malik and Hu, 2019). In the present study, all sugar-sweetened fruit nectars, fruit beverages, and cold coffee were "unhealthy." Reducing the consumption of these beverages and raising consumer awareness are essential to prevent the development of NCDs and obesity. In a previous study conducted in Türkiye, university students were found to consume an average of 11.34 g/d of sugar through only SSBs (Meric et al., 2021). Although there is usually information about the amount of sugar on the front label of SSBs, there should be remarkable icons, including star rating, and consumers should be informed. By contrast, many SSBs are made with large amounts of high-fructose corn syrup (HFCS). Compared with glucose, fructose is metabolized differently in the body, and the excessive consumption of fructose has been shown to increase the risk of dyslipidemia, nonalcoholic fatty liver disease, increased abdominal fat, and decreased insulin sensitivity (Clifford and Maloney, 2016). Similarly, 96.9% of fruit nectars with added HFCS and all fruit

Table 5. Median and IQR values of energy and some nutrients based on beverage and HSR classification

Energy and nutrient	N	Energy	Carbohydrate	Sugar	Added sugar
100% fruit juice					
HSR<3.5	17	57.0 (54.5~60.0)	14.0 (13.0~15.0)	13.0 (10.0~14.0)	–
HSR≥3.5	23	47.0 (42.0~49.0)	11.0 (9.3~12.0)	9.0 (8.0~11.0)	–
P-value[1)]		<0.001***	<0.001***	<0.001***	–
Fruit nectar					
HSR<3.5	40	52.1 (46.4~55.7)	12.0 (11.2~13.4)	11.6 (10.4~13.1)	6.5 (5.5~7.7)
HSR≥3.5	3	45.0 (28.0~49.0)	10.4 (6.2~12.0)	6.3 (6.1~9.1)	–
P-value		0.058	0.083	0.004**	–
Aromatic beverage					
HSR<3.5	12	50.0 (48.5~52.5)	12.4 (11.5~12.9)	12.2 (11.5~12.4)	12.1 (11.0~12.6)
HSR≥3.5	5	16.0 (4.6~36.8)	4.0 (0.3~8.5)	3.4 (0.0~6.8)	0.0 (0.0~2.3)
P-value		0.004**	0.004**	0.001**	<0.001***
Lemonade					
HSR<3.5	9	53.0 (46.2~55.0)	13.0 (11.4~13.2)	11.5 (10.3~12.9)	11.5 (9.6~13.0)
HSR≥3.5	6	4.0 (4.0~20.0)	0.6 (0.4~4.6)	0.5 (0.3~4.6)	0.0 (0.0~4.5)
P-value		<0.001***	0.001**	<0.001***	<0.001***
Iced tea					
HSR<3.5	16	34.5 (29.0~37.2)	8.4 (6.9~9.0)	8.3 (6.9~8.9)	8.4 (6.9~9.0)
HSR≥3.5	11	19.0 (4.0~20.0)	4.6 (1.0~4.8)	4.5 (0.9~4.8)	4.6 (0.0~4.8)
P-value		<0.001***	<0.001***	<0.001***	<0.001***
Cold coffee					
HSR<3.5	21	53.0 (48.5~64.5)	8.8 (8.1~9.4)	8.4 (7.7~8.6)	5.8 (4.2~6.4)
HSR≥3.5	4	38.5 (9.5~48.0)	3.9 (1.0~5.9)	3.5 (0.9~5.8)	–
P-value		0.006**	0.008**	0.003**	–
Carbonated beverage					
HSR<3.5	38	43.5 (37.5~45.0)	10.6 (9.2~11.0)	9.6 (8.8~10.5)	9.6 (8.8~10.5)
HSR≥3.5	30	2.2 (1.0~13.2)	0.5 (0.0~3.1)	0.0 (0.0~3.1)	0.0 (0.0~3.1)
P-value		<0.001***	<0.001***	<0.001***	<0.001***
Energy beverage					
HSR<3.5	13	45.0 (38.0~46.0)	11.0 (9.3~11.0)	11.0 (8.9~11.0)	11.0 (8.9~11.0)
HSR≥3.5	5	11.0 (2.5~29.0)	0.9 (0.5~2.7)	0.0 (0.0~2.5)	0.0 (0.0~2.5)
P-value		<0.001***	<0.001***	<0.001***	<0.001***

Values are presented as median (IQR).
Energy is shown in kcal, while other nutrients are shown in grams.
IQR, interquartile range; HSR, health star rating.
[1)]Mann-Whitney U-test.
***P*<0.01, ****P*<0.001.

Table 6. Energy density and health code distribution of beverages according to HSR classification

	Total (n=304)	HSR<3.5 (n=212)	HSR≥3.5 (n=92)	*P*-value[1)]
Minimum 5 g of added sugar per 100 mL				<0.001***
Yes	184 (60.5)	182 (85.8)	2 (2.2)	
No	120 (39.5)	30 (14.2)	90 (97.8)	
Minimum 5 g of total sugar per 100 mL				<0.001***
Yes	236 (77.6)	209 (98.6)	27 (29.3)	
No	68 (22.4)	3 (1.4)	65 (70.7)	
Energy density				<0.001***
Very-low (<0.6 kcal/mL)	283 (93.1)	191 (90.1)	92 (100)	
Low (0.6~1.5 kcal/mL)	21 (6.9)	21 (9.9)	-	

Values are presented as number (%).
HSR, health star rating.
[1)]Pearson's chi-square test.
****P*<0.001.

beverages with added HFCS had <3.5 stars.

While HSR considers many ingredients of beverages, it does not evaluate AS, EA, and added micronutrients (Vergeer et al., 2020). All fruit nectars and fruity beverages containing AS are considered "unhealthy," whereas all lemonades and cold coffees containing AS are con-

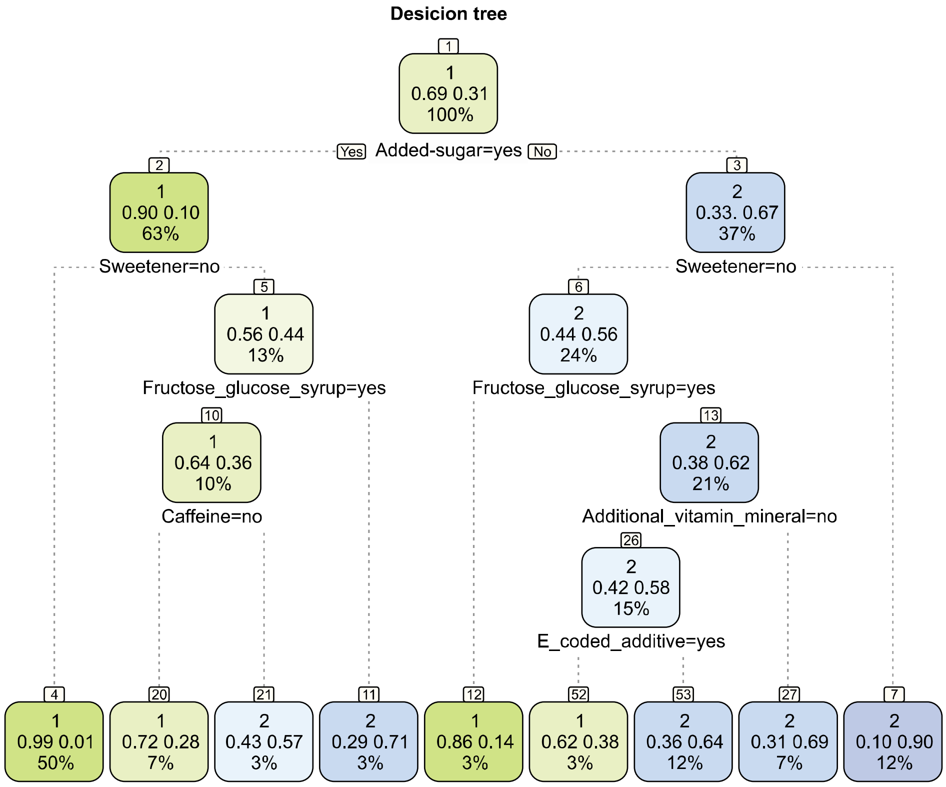

Fig. 1. Decision tree for predicting health star rating quality score. A total of 304 nonalcoholic ready-to-drink beverages were included in the study. Data from all beverages were analyzed. The decision tree model shows that beverages are predicted to be healthy or unhealthy depending on whether they contain sweeteners, fructose-glucose syrup, caffeine, additives, and vitamins and minerals.

Table 7. Confusion matrix table of the decision tree model

Predicted value	Actual value	
	HSR<3.5	HSR≥3.5
HSR<3.5	41	3
HSR≥3.5	3	14
Outcomes of statistics		
Accuracy (95% confidence interval)	0.90 (0.79~0.96)	
NIR	0.72	
P-value (Acc>NIR)	<0.001	
Kappa	0.75	
McNemar's test *P*-value	>0.999	
Sensitivity	0.82	
Specificity	0.93	
Positive predicted value	0.82	
Negative predicted value	0.93	
Prevalence	0.27	
Detection rate	0.22	
Balanced accuracy	0.87	
Precision	0.82	
Recall	0.82	
F1 score	0.82	

Confusion matrix, accuracy (95% confidence interval), kappa and McNemar's test of the decision tree in the testing (validation) group.
HSR, health star rating; NIR, no information rate; Acc, accuracy.

sidered healthy. Accordingly, some beverages containing EA or micronutrient(s) can be "healthy" or "unhealthy." In other words, there does not seem to be a relationship between these contents and the HSR of beverages. In examining HSR components, healthy beverages (≥3.5 stars) contained lower median energy, carbohydrate, sugar, and total fat contents compared with unhealthy beverages. The reason why the protein, fiber, and sodium contents of healthy and unhealthy beverages were not different may be because both types of beverages are not a source of protein (dairy beverages were excluded), fiber, and sodium. Although the energy, carbohydrate, sugar, and added sugar contents varied among beverage categories, there was a similar situation in every category. The ED of food and beverages is remarkable beyond the total energy. Food can be divided into four ED categories: very low ED (less than 0.6 kcal/g), low ED (0.6~1.5 kcal/g), medium ED (1.6~3.9 kcal/g), and high ED (4.0~9.0 kcal/g) (Rolls, 2017). According to this classification, all healthy beverages and 90.1% of unhealthy beverages in the present study had very low ED. Most unhealthy beverages had a very-low ED probably because of their high water content. This also explains why no beverages were classified as medium and high ED. However, this difference was found to be statistically significant.

In the decision tree created with various variables, variables other than those evaluated in HSR also affected their nutritional quality. According to HSR, added sugar, sweeteners, additives, fructose-glucose syrup, and caffeine were the strongest predictors of healthy beverages. Machine learning has better performance than the traditional statistical analysis approach. However, each decision tree machine learning algorithm has its strengths and weaknesses. Analyzing categorical data through a decision tree provides a more robust result that can be more easily interpreted by people. Various supervised machine learning models based on decision trees have been successfully applied to medical data for the accurate prediction of several clinical conditions, including metabolic syndrome (Yu et al., 2020), myocardial infarction (Baxt, 1991), cancer (Salama et al., 2012), and others (Joshi et al., 2010; Pei et al., 2020). However, there is limited research that will provide information about the health claims of foods and beverages by applying machine learning models to nutrition labels. The decision tree constructed in the present study provides information about the nutritional quality of the most frequently consumed beverages based only on the presence or absence of data in the label information, regardless of their quantity. One of the main reasons why added sugars and sweeteners in beverages were the most important determinants is that these components are included in the HSR score calculation. However, the resulting decision tree model only draws this conclusion based on simple statements such as “present” or “absent” while calculating the HSR score with a series of mathematical calculations. In this context, such simplified machine learning can be analyzed by people. Moreover, the high accuracy, precision, and F1 score in the cross-validation step show that this model has good predictive capability.

In conclusion, many ratings can determine whether food and beverages are healthy or unhealthy. According to HSR, most beverages examined in the present study were unhealthy. Because of the easy consumption of beverages, added sugar is a risk factor for the development of obesity and other diseases. Therefore, consumer awareness needs to be raised. The presence of star rating icons on beverages can guide consumers in correctly choosing a healthy beverage. To the best of our knowledge, this study is the first to apply machine learning algorithms to determine whether beverages are healthy based on nutrition label information. We found that decision tree learning algorithms identified healthy beverages based on simple information (e.g., added sugar and AS) provided on the label of beverages. Thus, these algorithms can play a role in reducing the risk of NCDs by providing a simple and effective approach to reducing the consumption of unhealthy beverages. However, further research is needed to validate our results by applying machine learning model analysis to access the label information of beverages that are much more prevalent in the market.

ACKNOWLEDGEMENTS

We thank the supermarket staff for their assistance during the study’s data collection.

FUNDING

None.

AUTHOR DISCLOSURE STATEMENT

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Concept and design: ABG, SED, MA. Analysis and interpretation: ABG, SED, MA. Data collection: ABG, SED. Writing the article: ABG, MA, SED. Critical revision the article: ABG, MA, SED. Final approval the article: ABG, MA, SED. Statistical analysis: MA. Overall responsibility: SED.

REFERENCES

Australian Government Department of Health and Aged Care. Health Star Rating system calculator and style guide. Australian Government Department of Health and Aged Care. 2020.

Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med. 1991. 115:843-848.

Bayram HM, Ozturkcan A. Nutrition quality of the Türkiye packaged foods and beverages: A comparison of two nutrient profile models. J Food Prod Mark. 2021. 27:255-265.

Clifford J, Maloney K. Sugar-sweetened beverages. 2016 [cited 2023 Nov 23]. Available from: https://extension.colostate.edu/topic-areas/nutrition-food-safety-health/9286-2/

Crino M, Sacks G, Dunford E, Trieu K, Webster J, Vandevijvere S, et al. Measuring the healthiness of the packaged food supply in Australia. Nutrients. 2018. 10:702.

Dunford E, Cobcroft M, Thomas M, Wu J. Technical report: Alignment of NSW healthy food provision policy with the Health Star Rating system. NSW Ministry of Health. 2015.

Dunford E, Taylor F. Report on the comparative nutritional profile of food and beverage products marketed by the 21 largest global companies in 9 countries. The George Institute for Global Health. 2018.

Dunford EK, Ni Mhurchu C, Huang L, Vandevijvere S, Swinburn B, Pravst I, et al. A comparison of the healthiness of packaged foods and beverages from 12 countries using the Health Star Rating nutrient profiling system, 2013-2018. Obes Rev. 2019. 20:107-115.

Forouzanfar MH, Afshin A, Alexander LT, Anderson HR, Bhutta ZA, Biryukov S, et al.; GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016. 388: 1659-1724.

Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022. 23:40-55.

Guzman-Vilca WC, Yovera-Juarez EA, Tarazona-Meza C, García-Larsen V, Carrillo-Larco RM. Sugar-sweetened beverage consumption in adults: Evidence from a national health survey in Peru. Nutrients. 2022. 14:582.

HSR Technical Advisory Group (TAG). Five year review of the

Health Star Rating (HSR) system: Non-dairy beverages. 2018 [cited 2023 Dec 2]. Available from: https://www.mpconsulting.com.au/wp-content/uploads/2018/10/Non-dairy-beverages.pdf

Jijo BT, Abdulazeez AM. Classification based on decision tree algorithm for machine learning. J Appl Sci Technol Trends. 2021. 2:20-28.

Jones A, Shahid M, Neal B. Uptake of Australia's Health Star Rating system. Nutrients. 2018. 10:997.

Joshi S, Shenoy D, Simha GG V, Rrashmi PL, Venugopal KR, Patnaik LM. Classification of Alzheimer's disease and Parkinson's disease by using machine learning and neural network methods. 2010 Second International Conference on Machine Learning and Computing. 2010. Feb 9-11. p. 218-222.

Keller A, Bucher Della Torre S. Sugar-sweetened beverages and obesity among children and adolescents: A review of systematic literature reviews. Child Obes. 2015. 11:338-346.

Labonté MÈ, Poon T, Gladanac B, Ahmed M, Franco-Arellano B, Rayner M, et al. Nutrient profile models with applications in government-led nutrition policies aimed at health promotion and noncommunicable disease prevention: A systematic review. Adv Nutr. 2018. 9:741-788.

Malik VS, Hu FB. Sugar-sweetened beverages and cardiometabolic health: An update of the evidence. Nutrients. 2019. 11: 1840.

Malik VS, Schulze MB, Hu FB. Intake of sugar-sweetened beverages and weight gain: a systematic review. Am J Clin Nutr. 2006. 84:274-288.

Meric ÇS, Ayhan NY, Yilmaz HÖ. Evaluation of added sugar and sugar-sweetened beverage consumption by university students. Kesmas: J Kesehat masy Nas. 2021. 16:9-15.

Nergiz-Unal R, Akal Yildiz E, Samur G, Besler HT, Rakicioğlu N. Trends in fluid consumption and beverage choices among adults reveal preferences for ayran and black tea in central Turkey. Nutr Diet. 2017. 74:74-81.

Pei D, Yang T, Zhang C. Estimation of diabetes in a high-risk adult Chinese population using J48 decision tree model. Diabetes Metab Syndr Obes. 2020. 13:4621-4630.

Pelly FE, Swanepoel L, Rinella J, Cooper S. Consumers' perceptions of the Australian Health Star Rating labelling scheme. Nutrients. 2020. 12:704.

Rayner M, Scarborough P, Kaur A. Nutrient profiling and the regulation of marketing to children. Possibilities and pitfalls. Appetite. 2013. 62:232-235.

Republic of Turkey Ministry of Health. Turkey nutrition and health survey 2017. Republic of Turkey Ministry of Health. 2019.

Rolls BJ. Dietary energy density: Applying behavioural science to weight management. Nutr Bull. 2017. 42:246-253.

Salama GI, Abdelhalim MB, Zeid MA. Breast cancer diagnosis on three different datasets using multi-classifiers. Int J Comput Inf Technol. 2012. 1:36-43.

Van Belle G, Fisher LD, Heagerty PJ, Lumley T. Biostatistics: A methodology for the health sciences. John Wiley & Sons. 2004.

Vergeer L, Vanderlee L, Ahmed M, Franco-Arellano B, Mulligan C, Dickinson K, et al. A comparison of the nutritional quality of products offered by the top packaged food and beverage companies in Canada. BMC Public Health. 2020. 20:650.

WHO (World Health Organization). Noncommunicable diseases. 2023 [cited 2023 Nov 12]. Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

WHO. Nutrient profiling: Report of a WHO/IASO technical meeting London, United Kingdom 4-6 October 2010. World Health Organization. 2011.

WHO. Use of nutrient profile models for nutrition and health policies: Meeting report on the use of nutrient profile models in the WHO European region, September 2021. World Health Organization. 2022.

Türkiye Nutrition Guide. 2022 [cited 2024 May 5]. Available from: https://hsgm.saglik.gov.tr/tr/web-uygulamalarimiz/357.html

Yu CS, Lin YJ, Lin CH, Wang ST, Lin SY, Lin SH, et al. Predicting metabolic syndrome with machine learning models using a decision tree algorithm: Retrospective cohort study. JMIR Med Inform. 2020. 8:e17110. https://doi.org/10.2196/17110