Current Research in Food Science 6 (2023) 100522

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Recommendations for validating hierarchical clustering in consumer sensory projects

Attila Gere

Hungarian University of Agriculture and Life Sciences, Institute of Food Science and Technology, Department of Postharvest, Supply Chain, Commerce and Sensory Science, Villányi út, 29-43, H-1118, Budapest, Hungary

ARTICLE INFO

Handling Editor: Dr. Maria Corradini

Keywords:
Hierarchical clustering
Cluster numbers
Distance measure
Sensory

ABSTRACT

Choosing the proper hierarchical clustering algorithm and number of clusters is always a key question in consumer sensory projects. In many cases, researchers do not publish any reason why it was chosen a given distance measure and linkage rule along with cluster numbers. The reason behind this could be that different cluster validation and comparison techniques give contradictory results in most cases. A complex evaluation to define the proper clustering might be time-consuming and tedious. The paper introduces the clustering of three sensory data sets using different distance metrics and linkage rules for different numbers of clusters. The results of the validation methods deviate, suggesting that clustering depends heavily on the data set in question. Although Euclidean distance, Ward's method seems a safe choice, testing, and validation of different clustering combinations is strongly suggested.

1. Introduction

Cluster analysis is a widely used unsupervised pattern recognition technique, which is able to uncover underlying patterns of data sets. In food sensory analysis, cluster analysis usually used to group sensory assessors (usually consumers) based on their sensory evaluations of different food products (*e.g.*: wholegrain buckwheat enriched pasta (Škrobot et al., 2022), cracker-type (Araújo et al., 2021) or gluten-free biscuits (Di Cairano et al., 2022), fried sweet potato (Dery et al., 2021), *etc.*) as well as to group food products based on their sensory attributes (see *e.g.* (Sridhar and Charles, 2021):), however, the earlier is used more widely. As cluster analysis gives an important information for sensory scientists, it is widely used. A Scopus search with the terms "sensory AND cluster" within Agricultural and Biological Sciences (limited to journals related to food sciences) gives more than 1100 hits for the time range 2000–2023, from which more than 100 comes from 2022 to 2023. The literature shows that, apart from a few examples, authors usually use agglomerative hierarchical clustering, even though there is a vast number of clustering methods available. Within agglomerative hierarchical clustering, there are also plenty options to choose from, namely the distance measures and linkage rules that make the life of a sensory scientist even more complicated when it comes to clustering. In the literature, there have been some attempts to highlight the importance of clustering in sensometrics (Yenket and Chambers, 2017), the experience shows that the majority of the papers dealing with clustering of consumer sensory data uses Euclidean distance and Ward's method, usually without the presentation of any cluster validation. Therefore, the presented paper aims to give recommendations and an R script to commonly used methods that are able to provide essential information about the clustering used in order to support food scientists in the validation process.

The paper is structured as follows. Section 1 provides a brief overview of sensory analysis and introduces different cluster analysis approaches with a special focus on agglomerative hierarchical clustering. Section 2 outlines multiple approaches employed to compare and validate clustering solutions. Section 3 introduces the case studies and the R packages used during the data analysis. Section 4 presents the obtained results by validation approaches. Section 5 provides conclusions from the study and recommendations for future research.

1.1. Agglomerative hierarchical clustering

Cluster analysis, however, can be completed with a set of different statistical techniques ranging from connectivity-based (*e.g.*, hierarchical clustering (Hastie et al., 2022)) through centroid-based (*e.g.*, k-means clustering (Palczak et al., 2020)) to density-based (*e.g.*, DBSCAN (Lu et al., 2020)) methods (Lee and Yang, 2009; Saxena et al., 2017). Although these methods rely on significantly different algorithms, their

E-mail addresses: gereattilaphd@gmail.com, gere.attila@uni-mate.hu.

https://doi.org/10.1016/j.crfs.2023.100522
Received 2 February 2023; Received in revised form 4 May 2023; Accepted 18 May 2023
Available online 19 May 2023

aim is the same: finding characteristic patterns in the data set that were previously unknown. Due to the popularity of connectivity- and centroid-based methods, they are available in nearly all commonly used statistical software (*e.g.,* XL-Stat, SPSS, StatSoft Statistica, R-project, *etc.*). Although the different approaches have their pros and cons, the current paper does not aim to compare these, rather, the focus will be taken on the most popular one, agglomerative hierarchical clustering (AHC). Although there is no available literature on the driving forces behind the choices, an undoubtable advantage of AHC is its simplicity and transparency. One of the most frequently used output of AHC, the dendrogram, gives an easy-to-understand visual representation of the clustering, that significantly enhances its understandability. Other, more sophisticated clustering methods usually require a deeper background in statistics that might discourage researchers.

Agglomerative hierarchical clustering requires three important points to be addressed during the data analysis process (James et al., 2021).

1. Choosing the distance metric
2. Choosing the linkage measure
3. Defining the number of clusters

The most popular answers to these questions might be Euclidean distance, Ward's method, and visual observation of the dendrograms, however, there are multiple tools that can support or even oppose these choices. The use of cluster validation tools becomes even more important when the number of distance metrics, linkage measures and tools used to define the optimal number of clusters is considered.

1.2. Distance metrics

Distance metrics define the pairwise distances between the objects (Gareth et al., 2013). However, the list of distance metrics for calculating of the distances between the objects is rather long. For example, XL-Stat lists more than 20 distance metrics, SPSS lists 8 of them, while the dist function from R-project stats package (R Core Team, 2022) lists 6 of them, just to list the three possibly most used software by average users. Although the number of options differs, some distance metrics are usually listed by any software, *e.g.,* Euclidean, Chebyshev (or Maximum) and Manhattan (or city-block) distance. Let us define two points in the two-dimensional space as A (x1, x2) and B (y1, y2), the Euclidean distance of two points of the length of a line segment between them (using the Pythagorean theorem):

$$d = \sqrt{\left[(x_2 - x_1)^2 + (y_2 - xy_1)^2\right]}$$

while the Chebyshev distance between two vectors is the greatest of their differences:

$$d = max(|x_1 - x_2|, |y_1 - y_2|)$$

while the absolute distance between two vectors is given by Manhattan distance (sum of two legs in a right triangle):

$$d = |x_1 - x_2| + |y_1 - y_2|$$

Although Euclidean distance is one of the most common distance metrics, it must be highlighted that it is sensitive to differently scaled variables. Usually, consumer sensory data uses one scale (hedonic or preference scales). However, when other variables or scales are involved, normalization is needed before using this distance measure. Additionally, Euclidean distance works well in low-dimensional data, which is more common in consumer sensory data sets. As Chebyshev distance can also be defined as the maximum distance between the objects, it is also called as chessboard distance as Chebyshev distance equals to the minimum number of moves a king needs to go from one square to the other. Although the distance is widely available in many statistical software, it is more popular in signal processing (*e.g.* for neuro imaging (Omidvarnia et al., 2021)) or with spatial data (*e.g.* logistics (Baykasoğlu and Subulan, 2016)). Manhattan distance works well with binary and/or discrete variables or in situations where no straight lines between objects exist. Since the distance is defined as the path along gridlines, Manhattan distances are usually greater than that of obtained using Euclidean distance (Fig. 1). When choosing distance metrics for cluster analysis, one should consider the sensitivity of the given metric to outliers, as *e.g.,* Euclidean and Chebyshev distances are more sensitive for outliers than Manhattan distance. For further information and list of possible distances used in different scientific fields, see ref. (Deza and Deza, 2013), while Table 1 of Shirkhorshidi et al. (2015) introduces the advantages and disadvantages of the different distance measures.

1.3. Linkage methods

The second important issue is the definition of the linkage method, *e.g.*, how to link the objects. However, this decision needs several questions to be answered as it is with the distance metrics, namely, there are vast number of options for linking the objects. In order to follow the previous examples, XL-Stat lists 6 linkage methods, SPSS lists 8 of them, while the hclust function from R-project stats package lists 7 of them. Out of a few exceptions, the distance metrics and linkage methods can be combined, creating a high number of possible clustering combinations. Focusing only on the most widely used linkage methods (*e.g.,* Ward's method, single and complete linkage *etc.*) the number of combinations decreases significantly but it might be still high.

During the process, the algorithm takes all objects as a separate cluster and, using a linkage method, links them until all clusters are grouped into one single cluster (Fig. 2). Calculation of linkages (L) between two clusters (C1 and C2) for points *j* and *k* is done differently by the linkage methods (Kassambara, 2017). For example, single linkage links two clusters with the closest minimum distance as:

$$L(C1, C2) = min(D(j,k)), j \in C1, k \in C2$$

while complete linkage links two clusters with the closest maximum distance:

$$L(C1, C2) = max(D(j,k)), j \in C1, k \in C2$$

while average linkage links to clusters based on their lowest average distances:

$$L(C1, C2) = \frac{1}{n_{C1} + n_{C2}} \sum_{j=1}^{n_{C1}} \sum_{k=1}^{n_{C1}} D(j,k), j \in C1, k \in C2$$

where n_{C1} and n_{C2} are the number of data points in *C1* and *C2*, respectively,

While centroid linkage links two clusters with the lowest centroid distances:

$$L(C1, C2) = \|c_{C1} - c_{C2}\|,$$

where c_{C1} and c_{C2} are the centroids of clusters *C1* and *C2*, respectively,

While Ward's method links two clusters having the lowest sum of squares values (for further details on the formula, see (Zaki and Meira, 2020)):

$$L(C1, C2) = \frac{n_{C1} n_{C2}}{n_{C1} + n_{C2}} \|c_{C1} - c_{C2}\|^2$$

2. Comparison of clustering combinations

2.1. Visual assessment of dendrograms

Dendrograms generally follow a tree-like structure. The most substantial parts are the leaves, representing the objects to be clustered.

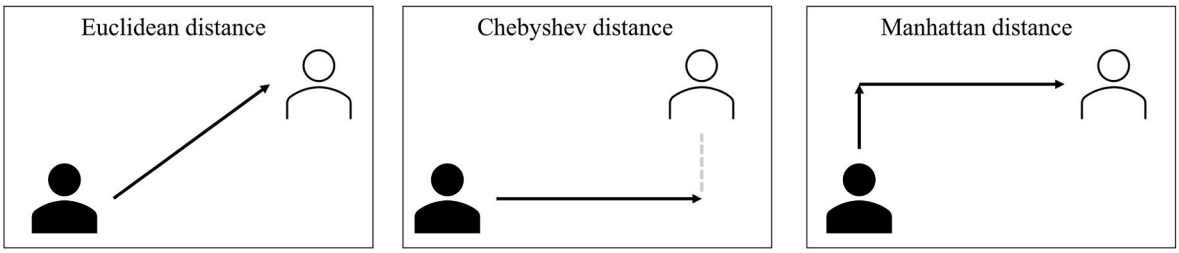

Fig. 1. Visual representation of the Euclidean, Manhattan and Chebyshev distances.

Table 1
Cophenetic correlation coefficients of the 15 dendrograms. Highest values within distance metrics are highlighted with bold.

Data set	Linkage method	Euclidean distance	Chebyshev distance	Manhattan distance
#1	Ward's method	0.496	0.414	0.397
	Single linkage	0.582	0.541	0.583
	Complete linkage	0.516	0.513	0.459
	Average linkage	**0.687**	**0.625**	**0.673**
	Centroid linkage	0.591	0.561	0.576
#2	Ward's method	0.509	0.543	0.510
	Single linkage	0.684	0.593	0.626
	Complete linkage	0.588	0.607	0.559
	Average linkage	**0.736**	**0.714**	**0.691**
	Centroid linkage	0.667	0.644	0.594
#3	Ward's method	0.426	0.351	0.406
	Single linkage	0.513	0.477	0.409
	Complete linkage	0.468	0.406	0.516
	Average linkage	**0.625**	**0.589**	**0.644**
	Centroid linkage	0.502	0.494	0.543

First, the leaves fuse into branches, represented by horizontal lines, then, the smaller branches fuse into larger ones up until all the branches are fused. Naturally, leaves fused an early step are more similar to each other than the ones fused later (higher on the tree). The height of the fusion expresses the similarity of the branches (James et al., 2021). As visual information is processed quickly, interpretation and comparison of dendrograms are quick and patterns are seen immediately. Definition of the number of clusters can be done by looking for the highest fusion steps between branches, where we cut the dendrograms. In some cases, finding this cut is easy, but sometimes it becomes tricky when the heights of the fusions are similar. In such cases other methods are required.

2.2. Cophenetic correlation coefficient

Cophenetic correlation coefficients are used to assess clustering solutions as it determines, how a given dendrogram preserves the pairwise distances of the original distance matrix (Saraçli et al., 2013). Cophenetic correlation coefficients are calculated between the original distance matrix and the cophenetic distance matrix, where in the original distances of the objects are replaced by the computed distances between their clusters at the time of these clusters' fuse. Cophenetic correlation coefficients are computed for the clustering solutions individually, and higher values mean a better clustering. Using the cophenetic correlation coefficients, the different clustering solutions can be compared easily and quickly, however, the method is sensitive to outliers (Sokal and Rohlf, 1962).

Let us suppose that the original data $\{X_i\}$ have been clustered to generate a dendrogram $\{T_i\}$ and define the following distance measures $x(i,j)$ as the Euclidean distance between the ith and jth points and $t(i,j)$ as the dendrogramatic distance between T_i and T_j model points, while $\overline{x}$ and $\overline{t}$ the averages of points $x(i,j)$ and $t(i,j)$. The cophenetic correlation

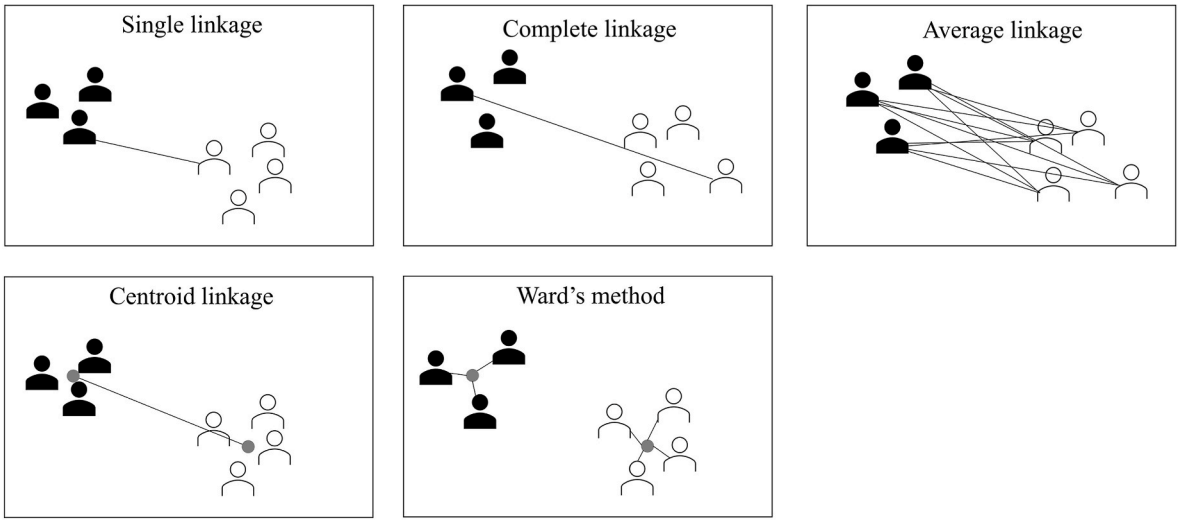

Fig. 2. Visual representation of single, complete, average, centroid linkages and Ward's method.

coefficient CPH is given by (Saraçli et al., 2013; Sokal and Rohlf, 1962):

$$CPH = \frac{\sum_{i<j}[x(i,j)-\bar{x}][t(i,j)-\bar{t}]}{\sqrt{\sum_{i<j}x(i,j)-\bar{x}^2\sum_{i<j}t(i,j)-\bar{t}^2}}$$

2.3. Cluster validity indices

When it comes to the definition of the number of clusters present in the data set, cophenetic correlation coefficient cannot be of any help as it evaluates the dendrogram. The number of clusters, therefore, is defined by the visual inspection of the dendrogram. Mathematical evidence to validate the clustering could be the use of cluster validity indices. Interestingly, researchers have limited power to choose cluster validity indices to compare the different cluster numbers. In spite of the high number of different cluster validity indices, statistical software packages do not make them widely available for the researchers.

Two of the most widely used cluster validity indices are the Silhouette index and Rand index. Silhouette index measures how similar an object is to its cluster compared to other clusters (Rousseeuw, 1987), while Rand index determines the similarity of two clusterings (Rand, 1971). While Silhouette index can take up a value between −1 and 1 (where 1 means that the object is well clustered), Rand index goes from 0 to 1, where 1 means that two clustering group the object similarly. There are numerous other options to calculate cluster validity measures. One of the most versatile packages in R-project is called NbClust, which provides 30 cluster validity indices to compare their results on the same clustering but different cluster numbers. The package then uses a voting system to define the optimal number of clusters (*e.g.*, offers the number of clusters that proved to be best by the highest number of indices). For further information on the comparison of different cluster validity indices, see refs (Hämäläinen et al., 2017) and (Arbelaitz et al., 2013).

Based on (Charrad et al., 2014), Silhouette index is defined as follows:

$$Silhouette = \frac{\sum_{i=1}^{n} S(i)}{n}, Silhouette \in [-1, 1]$$

where.

$S(i) = \frac{b(i)-a(i)}{max\{a(i):b(i)\}}$,

$a(i) = \frac{\sum_{j\in\{C_r\setminus i\}} d_{i,j}}{n_r - 1}$ is the average dissimilarity of the *i*th object to all other objects within cluster C_r

$b(i) = max_{s\neq r}\{d_{iC_s}\},.$

$d_{iC_s} = \frac{\sum_{j\in C_s} d_{ij}}{n_s}$ is the average dissimilarity of the *i*th object to all other objects withing cluster C_s

2.4. Profile plots

A profile plot is usually a line plot presenting the average values of each variable for the different clusters. The profile plot, therefore, gives a graphical representation of the created groups, highlighting those variables that play a key role in clustering. Further evaluation should use *t*-test or analysis of variance (depending on the number of clusters) to look for significant differences between the created clusters variable-wise. Clustering solutions showing the highest number of significant variables should be considered as better.

2.5. Stability analysis

Stability analysis helps to determine the number of clusters to be used. It supposes that a clustering is good if the results are similar when the same clustering is run on several subsamples of the original data set. Therefore, stability analysis involves the generation of subsamples of the original data set and the subsamples are clustered with the same clustering into *k* number of clusters (Leisch, 2016). By calculating the pairwise distances between these clusterings, the instability of the clustering can be quantified. Instability is essentially the mean distances between the clusterings. As the instability is calculated to all *k* number of clusters, the cluster number showing the lowest instability (or highest stability) should be chosen (Ullmann et al., 2022; von Luxburg, 2010). There are multiple R packages available to test the stabilities of clusterings. The package clValid, for example, computes four different measures. The clValid function of the package clusters the full data set into *k* number of clusters, then it does a next clustering on a data set after removing one variable (each variable is removed once). The average proportion of non-overlap (APN) gives the average proportion of cases not placed in the same cluster (full data set vs. cropped data set), while the average distance (AD) computes the average distance between cases placed in the same cluster. The average distance between means (ADM) measure gives the average distance between cluster centroids for cases placed in the same cluster. The figure of merit (FOM) measures the average intra-cluster variance of the cases in the deleted column, where the clustering is based on the remaining (undeleted) samples. FOM estimates the mean error using predictions based on the cluster averages (Brock et al., 2008).

2.6. Correlation analysis

An interesting approach is to calculate the Pearson correlation coefficient between the objects and the cluster mean scores. The approach looks for individuals who have a predefined, *e.g.:* less than 0.6, correlation coefficient and regroups them into other clusters. If there are no clusters, in which their correlation coefficient reaches the predefined limit, the object is labelled as “unclassified” (Ramsey et al., 2021; Yang et al., 2019). This approach enables the researchers to create homogeneous clusters, where members are closely related to each other. However, if the clustering was less efficient, the number of objects in the unclassified group goes up. Additionally, the limit of correlation coefficient should be chosen according to the number of objects within the clusters as with small number of objects higher correlation coefficients might not be significant.

2.7. Validation datasets

Although less relevant to consumer sensory science, an important validation strategy is the application of validation data sets. Validation data set refers to an independently collected data set that is similar to the data set we are working with. For example, a validation data set can be the consumer sensory evaluation of the same samples recorded at a different time or with a different consumer panel. If there is no such data is available and the collected data enables, splitting the data into 50/50 split ratio is suggested (Ullmann et al., 2022). Running the same clustering on the validation data set is expected to provide similar results compared to the original data set determined by a partition similarity index, for example adjusted Rand or Jaccard index (Meila, 2016).

3. Materials and methods

3.1. Case studies

Three case studies were used, and the following criteria were set: i) the data set should be published, ii) the data sets should contain different number of consumers (rows) and products (variables), iii) the data set should contain preference data registered on a 9 or 10-point hedonic scale. Case study #1 contains sensory preference data of six flavored kefir products that were rated by 59 consumer assessors on a 9-point hedonic scale. Further information of their formulations and analyses other than cluster analysis can be found in the original publication

(Gere et al., 2014). Case study #2 was introduced by (Rothman and Parker, 2009) and includes the preference scores of 102 consumers on 5 snack products. For further information on the data set see: (Gere et al., 2017; Rothman and Parker, 2009). Case study #3 is the strawberry preference sensory data set published by (Ares and Jaeger, 2013). Preference data of 107 participants were recorded on six products.

All three cases, the input matrix lists the consumers in rows and the products in columns. Only the preference scores were used during the clustering. Standardization of the input data matrix is usually not required in consumer sensory clustering as the consumers use the same hedonic scale to express their preferences. In cases, when other variables are included in the clustering, standardization of the variables is needed. As the presented case studies use only preference scores, no standardization was done. Further information on standardization of variables in cluster analysis is provided by (Milligan and Cooper, 1988).

All three data sets were screened for missing data and for cases with 0 standard deviation (*e.g.* for consumers who rated all the products the same). Where missing data or 0 standard deviation was found, the case has been eliminated from further analysis.

In case study #1, two participants (coded as 12 and 60) rated all six products as 10 (maximum value on the hedonic scale). Their ratings are unique in a way that all other participants rated the product differently, *e.g.,* stated different levels of preference towards the products. The reason(s) behind these ratings are not discussed here as these might be

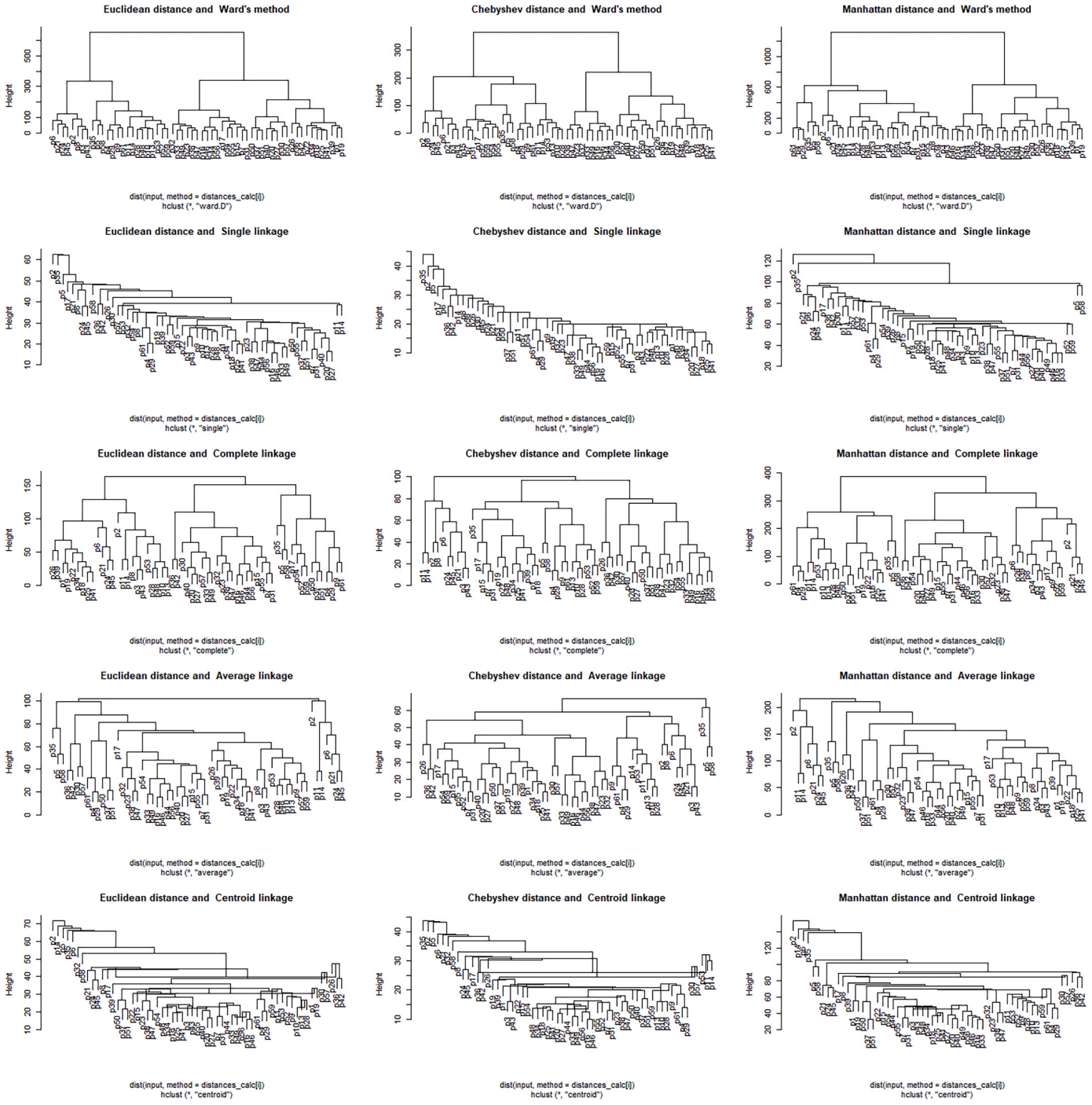

Fig. 3. Dendrograms of the flavored kefir data set (data set #1) created by the combinations of the three distances (Euclidean, Chebyshev and Manhattan distances) and the five linkage rules (Ward's method, single, complete, and centroid linkages).

caused by problems during data recording or data transformation, *etc.* Therefore, participants 12 and 60 have been excluded from the further data analysis. In case study #2, participant p91 rated all products as maximum, therefore, has been eliminated. There was also one participant in case study #3 who was eliminated because of providing the maximum value for all products (p66).

3.2. Data analysis

Distance metrics were calculated using dist, while linkage measures and clustering was calculated using the hclust function of stats package (ver. 4.0.2), that is part of R-project (version 4.0.2) (R Core Team, 2022). Cophenetic correlation coefficient was calculated using dendextend R-package (Galili, 2015), Silhouette indices were calculated using cluster package (ver. 2.1.0) (Maechler et al., 2019), comparison of multiple cluster validation indices was done using NbClust package (Charrad et al., 2014), while analysis of variance was run using the aov function of stats package. Cluster validation was computed using clValid package (Brock et al., 2008). The used R code is provided in the supplementary material.

4. Results

4.1. Visual assessment of dendrograms

One of the first and most widely used methods to compare and evaluate agglomerative hierarchical clustering is to visualize them using dendrograms. Dendrograms present the cases and their relationships, therefore, the distances of the cases are shown. Vertical lines express distances, while horizontal lines represent agglomerations, *e.g.,* when two cases are clustered.

Three distance metrics and five linkage methods, altogether 15 dendrograms, were created. In order to make it easier to compare the different dendrograms, the ones using the same distances are presented in columns, while linkages are ordered in rows. This arrangement of dendrograms enables us to compare the distances linkage-wise as well to compare the linkages distance-wise.

Fig. 3 presents 15 dendrograms of case study #1. It becomes immediately obvious that linkages have a significantly higher impact on the structures of the dendrograms than distance metrics, as each linkage shows a characteristic structure regardless of the distance metric used. Dendrograms using Ward's method and complete linkage seems to be the most appealing as these dendrograms are symmetric and seem to have evenly distributed clusters. Single and centroid linkages provide a less useful solution, as it would be difficult to find reasonable clusters.

Dendrograms of case studies #2 and #3 are found in the supplementary material as figures S1 and S2, respectively.

Although the two other data sets contain different number of consumers, the structures of the obtained dendrograms are similar. In both cases, the linkage rules determine the structure of the data set. Again, Ward's method and complete linkage were the two that provided the most promising dendrograms, *e.g.*, the ones that could be used for further characterization of consumer groups. Based on the dendrograms, a reasonable choice could be a clustering having Ward's method or complete linkage.

4.2. Cophenetic correlation coefficient

Cophenetic correlation coefficient measures how faithfully a dendrogram preserves the pairwise distances between the original unmodeled data points (Rohlf and Fisher, 1968). The cophenetic correlation coefficient, therefore, helps us to determine, how much did the dendrogram preserve the structure of the original data set. A higher cophenetic correlation coefficient means higher similarity between the two.

Table 1 presents the cophenetic correlation coefficients of the dendrograms of all three data sets introduced by Fig. 1, Figs. S1 and S2. The highest values have been highlighted distance-wise. All cases, average linkage received the closest value to 1, therefore, cophenetic correlation coefficient suggests using the average linkage. Linkage-wise comparison suggests that Euclidean and Manhattan distances should be used, depending on the data set.

Based on the results of cophenetic correlation coefficients, the suggested clustering should be a clustering having average linkage and preferably Euclidean distance for data sets #1 and #2, while Manhattan distance and average linkage for data set #3.

4.3. Clustering indices

Until this point, the analysis focused merely on the dendrograms and the different agglomerative hierarchical clusterings using different distance measures and linkage rules. However, one of the main aim of such analyses is to define distinct clusters, whose members have significantly different preferences. With other words, we are looking for clusters whose members prefer the same products similarly and this preference is different among clusters.

Definition of clusters is relatively easy: to draw a horizontal line on the dendrogram and count the vertical lines matching it. Participants connecting to the matching vertical lines will now belong to the same cluster. However, definition of where to draw the horizontal line might be surprisingly complicated to find.

A rule of thumb would be to cut the dendrogram where the heights of the vertical lines are the longest. In our case, this would be obvious with clustering using Ward's method, but it becomes more challenging at dendrograms created with average linkages (see Fig. 2, S1 and S2).

Cluster validity indices have been invented to help us determining the best number of clusters. In the followings, results of the three distance metrics with Ward's method, complete and average linkages will be compared, as these solutions seem to be the most promising based on the visual inspection of the dendrograms and the cophenetic correlation coefficients.

Table 2 presents the average Silhouette indices obtained for the nine clustering and their 2, 3, 4 and 5 cluster numbers for all three data sets. Silhouette index is widely used to determine the optimal cluster number for the same clustering; however, when the same data set is used (*e.g.,* the number of variables is the same), the obtained indices can be used to compare different clustering algorithms, as well. For all data sets, clustering with Manhattan distance and average linkage gave the highest index value for two clusters. It can be seen, however, that in the most cases, two clusters for all three data sets.

The absolute highest index value was obtained again for average linkages. Interestingly, not only the first but the first three highest row-wise index values were obtained with average linkage. The only exception is case study #3, where the Silhouette index of Manhattan distance and complete linkage is within the first three highest.

Multiple questions might arise, as we see the results introduced by Table 2. One of the most important one would be the arbitrary selection of Silhouette index. Working with one clustering index, might provide biased results. For example, the Silhouette index might have indicated the average linkage the best due to the similarity in their concepts. There are numerous other indices available, from which researchers can chose from, each having their pros and cons, making the analysis ambiguous.

A good compromise would be computing multiple clustering indices and use their consensus. One such approach is the NbClust R-package. Details on how the optimal number of clusters is determined by the indices in NbClust is presented in Table 2 of Charrad et al. (2014). Results of the NbClust comparison is shown on the leftmost column of Table 2. Comparisons were done based on the results of 23 indices and the cluster number found to be the best the most times is chosen as suggested. Accepting the default settings of the function, clusters between 2 and 15 were compared. The results are in correlation with those obtained using only the Silhouette index, namely two clusters were

Table 2
Average Silhouette widths of the nine clustering with 2, 3, 4 and 5 cluster numbers for the three data sets. Higher average Silhouette widths mean better clustering. Highest values are highlighted with bold row-wise. The last column gives the result of NBClust r-package, that uses the consensus of multiple cluster validity indices.

Data set		Number of clusters				NBClust
		2	3	4	5	
#1	Euclidean distance Ward's method	0.227	0.163	0.199	0.186	2
	Chebyshev distance Ward's method	0.178	0.129	0.162	0.187	2
	Manhattan distance Ward's method	0.194	0.13	0.14	0.159	2
	Euclidean distance average linkage	0.254	0.214	0.207	0.131	2
	Chebyshev distance average linkage	0.232	0.201	0.212	0.208	2
	Manhattan distance average linkage	0.270	0.248	0.206	0.159	8
	Euclidean distance complete linkage	0.186	0.210	0.217	0.193	3
	Chebyshev distance complete linkage	0.203	0.153	0.158	0.147	3
	Manhattan distance complete linkage	0.151	0.179	0.148	0.164	3
#2	Euclidean distance Ward's method	0.252	0.252	0.268	0.23	2
	Chebyshev distance Ward's method	0.272	0.178	0.195	0.208	2
	Manhattan distance Ward's method	0.293	0.257	0.208	0.207	2
	Euclidean distance average linkage	0.309	0.276	0.244	0.148	2
	Chebyshev distance average linkage	0.300	0.205	0.261	0.262	4
	Manhattan distance average linkage	0.332	0.27	0.253	0.26	2
	Euclidean distance complete linkage	0.252	0.256	0.255	0.276	2
	Chebyshev distance complete linkage	0.212	0.265	0.268	0.128	3
	Manhattan distance complete linkage	0.281	0.27	0.206	0.209	2
#3	Euclidean distance Ward's method	0.180	0.132	0.145	0.16	2
	Chebyshev distance Ward's method	0.120	0.138	0.140	0.161	2
	Manhattan distance Ward's method	0.189	0.141	0.170	0.185	2
	Euclidean distance average linkage	0.254	0.150	0.162	0.143	2
	Chebyshev distance average linkage	0.214	0.171	0.108	0.109	2
	Manhattan distance average linkage	0.281	0.241	0.186	0.163	2
	Euclidean distance complete linkage	0.175	0.146	0.117	0.122	2
	Chebyshev distance complete linkage	0.093	0.088	0.064	0.026	2
	Manhattan distance complete linkage	0.236	0.160	0.142	0.152	3

found to be optimal by the most. In case of case study #1, Manhattan distance and average linkage suggests 8 clusters, however, keeping eight clusters with a data set of 59 participants produces fragmented clusters. Another interesting observation that linkages have again higher influence on the results as three clusters were only suggested in the case of complete linkage only.

4.4. Cluster sizes

An important characteristic of clustering is the size of the created clusters, *e.g.,* the cases within a cluster. Table 3 contains the size of clusters for all nine clustering with two and three clusters for case study #1. Table 3 shows that average linkage results unevenly distributed clusters, when the dendrogram is cut at two or three clusters. Clusterings done using Ward' method or complete linkage (except with Chebyshev distance), however, provides balanced clusters.

Cluster sizes cannot tell anything about the members of the clusters, *e.g.,* how similar (or different) are the clusters based on the variables we want to separate them. With other words, it would be beneficial to create clusters whose members show significantly different preferences towards the products. Knowing such information is key for food product development, when the aim is to target different sensory clusters present on the market.

Comparing the clusters by variables was done using analysis of variance. Results of data set #1 are presented in the last six columns of Table 3, where bold indicates significant differences between cluster preferences. Naturally, we expect clusters that have significant differences (or clusters that have the most significant differences along the variables). The most significant differences between clusters were identified by Ward's method for the two cluster solution (5–6 significant variables out of 6), while complete linkage found the most significant differences in the case of three clusters (4 significant variables out of 6).

One-way ANOVAs for case studies #2 and #3 are presented by Tables S1 and S2. Regarding cluster sizes, the pattern is similar to the ones observed in Table 3: Clusterings done using Ward' method or complete linkage (except with Chebyshev distance), provides balanced clusters. For two clusters, there are more significant differences, while three clusters show slightly less significantly different variables.

In order to test the effect of distances, linkages and cluster numbers on F-values, four-way factorial ANOVA was run with distance metrics (Euclidean, Chebyshev and Manhattan), linkage rules (Ward's, average and complete linkages), cluster numbers (between 2 and 6) and products as factors (Tables S3–S5). Product factor was significant in all cases, while other factors proved to be dataset dependent. For case study #1, factor linkage was significant with complete linkage providing significantly higher F-values compared to the other two. For case study #2, distance was the second significant factor. The highest F-values were registered when Euclidean distance was used, however, Tukey *post hoc* test indicated that no significant difference between Euclidean and Manhattan distances (and between Manhattan and Chebyshev) exists, while a significant difference was found between Euclidean and Chebyshev distances. The most significant factors were found with case study #3, where linkage and cluster number were also significant, not just the product factor. The use of Ward's method provided significantly higher F-values compared to the other two. Although the factor cluster number was significant, Tukey *post hoc* test indicated significant differences between two clusters and 6 clusters, all the in between clusters showed no significant differences.

4.5. Stability analysis

Stability measures run on eight clustering combinations between cluster numbers 2 and 6 provided extreme results, namely most of the times either the lowest or the largest cluster number was identified as best in the case of Euclidean distance (Table 4). Manhattan distance, especially with single and complete linkage, suggested 3 clusters for APM and ADM. AD and FOM values suggested consistently 6 clusters. Interestingly, AD suggested 6 clusters for all the combinations. After raising the range of cluster numbers, the pattern showed no differences. The reason behind the results could be the size of the data set, as data set #1 and #3 contains six variables, while data set #2 has only five. These stability measures compare the results from clustering based on the complete data set to clustering based on removing each column, one at a time. As sensory preference studies usually don't involve a high number of samples, therefore the number of variables is rather low.

Table 3
Cluster sizes denoted by C1, C2 and C3 for the nine clustering combinations cut at two and three clusters. The last six columns show the results of analysis of variance. Boldface means significant difference between clusters based on the variables.

	C1	C2	C3	AppleA	AppleB	CurrantA	CurrantB	GrapeA	GrapeB
Two clusters									
Euclidean distance Ward's method	35	24	–	<0.001	<0.001	0.004	0.162	<0.001	<0.001
Euclidean distance average linkage	52	7	–	<0.001	<0.001	0.561	0.657	0.394	0.381
Euclidean distance complete linkage	24	35	–	<0.001	<0.001	0.024	0.097	0.019	0.53
Chebyshev distance Ward's method	30	29	–	<0.001	<0.001	<0.001	0.008	<0.001	<0.001
Chebyshev distance average linkage	56	3	–	0.013	0.006	0.004	0.049	0.311	0.282
Chebyshev distance complete linkage	49	10	–	<0.001	<0.001	0.513	0.819	0.508	0.835
Manhattan distance Ward's method	27	32	–	<0.001	<0.001	0.001	0.03	<0.001	<0.001
Manhattan distance average linkage	52	7	–	<0.001	<0.001	0.561	0.657	0.394	0.381
Manhattan distance complete linkage	22	37	–	0.752	0.161	0.293	0.211	<0.001	<0.001
Three clusters									
Euclidean distance Ward's method	19	24	16	0.816	0.405	<0.001	<0.001	0.028	0.063
Euclidean distance average linkage	52	1	6	<0.001	<0.001	0.334	0.425	0.232	0.25
Euclidean distance complete linkage	24	14	21	<0.001	<0.001	0.156	0.252	<0.001	0.005
Chebyshev distance Ward's method	20	29	10	0.753	0.374	<0.001	<0.001	0.054	0.034
Chebyshev distance average linkage	56	2	1	0.015	0.009	0.01	0.055	0.176	0.75
Chebyshev distance complete linkage	26	10	23	0.029	<0.001	0.096	0.713	<0.001	<0.001
Manhattan distance Ward's method	18	32	9	0.856	0.09	<0.001	<0.001	0.162	0.049
Manhattan distance average linkage	49	7	3	0.358	<0.001	0.047	0.054	0.166	0.147
Manhattan distance complete linkage	22	15	22	0.039	0.002	0.676	0.275	<0.001	<0.001

Table 4
Cluster stability values of eight clustering combinations run between cluster numbers 2 and 6. The best cluster identified by the stability measures are presented.

Data set	Clustering	APN	AD	ADM	FOM
#1	Euclidean distance Ward's method	2	6	6	6
	Euclidean distance single linkage	2	6	2	6
	Euclidean distance complete linkage	2	6	2	6
	Euclidean distance average linkage	2	6	2	6
	Manhattan distance Ward's method	2	6	2	5
	Manhattan distance single linkage	3	6	3	6
	Manhattan distance complete linkage	3	6	3	6
	Manhattan distance average linkage	2	6	2	6
#2	Euclidean distance Ward's method	2	6	2	5
	Euclidean distance single linkage	4	6	2	6
	Euclidean distance complete linkage	2	6	5	6
	Euclidean distance average linkage	2	6	2	5
	Manhattan distance Ward's method	2	6	2	6
	Manhattan distance single linkage	2	6	3	6
	Manhattan distance complete linkage	3	6	3	6
	Manhattan distance average linkage	2	6	2	6
#3	Euclidean distance Ward's method	2	6	5	6
	Euclidean distance single linkage	2	6	2	6
	Euclidean distance complete linkage	2	6	2	6
	Euclidean distance average linkage	2	6	2	6
	Manhattan distance Ward's method	2	6	2	6
	Manhattan distance single linkage	2	6	2	6
	Manhattan distance complete linkage	2	6	2	6
	Manhattan distance average linkage	2	6	3	6

APN: average proportion of non-overlap, AD: average distance, ADM: average distance between means, FOM: figure of merit.

4.6. Correlation analysis

The correlation-based method assumes that the number of clusters have already been defined. Once the number of clusters have been identified, however, this approach is a useful tool for rearranging cluster members to create more homogenous clusters.

Pearson correlation coefficients were calculated between the individual evaluations and cluster centroids. For presentation purposes, the results of Euclidean distance and Ward's method with three clusters will be presented on case study #1. The original clusters consisted of 14, 27 and 18 members, respectively (Table 3). Rearrangement should be done as if the correlation coefficient of the individual's liking and the cluster centroids is higher than 0.6, then it should be placed in that given cluster. If there is no cluster where the individual can be placed (*e.g.,* there are no clusters that have an $r > 0.6$ with the individual), then the individual should be clustered as "other". Ten individuals have been clustered into the "other" cluster, while the new clusters consisted of 10, 22 and 17 members, respectively. Three participants were reclustered from cluster 2 to cluster 1, while other two participants were placed from cluster 2 to cluster 1. The other 5 participants had a correlation coefficient lower than 0.6 with any of the clusters, however, their highest correlation coefficient belonged to their original cluster. For example, participant #1 had the following correlation coefficients with the three clusters: 0.27, −0.35, −0.68, while the original cluster was cluster 1.

5. Conclusions

It has been shown that there are no golden standards for cluster analysis and the validation of cluster analysis as its results depend heavily on the data set. There are multiple validation methods available, that might provide different results. The presented paper is the first that compares different clusterings and validation tools on three consumer sensory data sets. Visual comparison of dendrograms suggest that Ward's method and complete linkage should be used regardless of the distance metric. The provided dendrograms are appealing to the eye and suggest well distributed clusters. Cophenetic correlation coefficients, however, suggest average linkage, again, regardless of the distance metric used. The highest *Silhouette indices* were obtained again for average linkages, suggesting that two clusters should be formed (*e.g.,* to cut the dendrogram at two clusters). As NbClust compares possible cluster numbers, it suggests, again, the use of two clusters. The most balanced distributions of two and three clusters were obtained with Ward's method, regardless of the distance used. Additionally, clusters that showed the most significantly different likings between products were obtained with Ward's method and two clusters. To support these, factorial analysis of variance showed, that the clustering results depend on the data set. Stability analysis showed extreme results, the measures suggested either the lowest or the highest number of clusters chosen.

The presented work is far from complete, as there are multiple other tools that can be used for cluster validations. Such tools include *e.g.* cluster cohesion measured by the sum of squares (WSS) within cluster and about cluster separation measured by the sum of squares between clusters (BSS) or similarity matrices as visual tools. Further research should focus on the combination of these tools to help researchers validating their clusterings. Additionally, other data sets should also be

involved to extend the validity of the results. Involving more data sets and more validation tools will enable us to compare the validation methods and give a workflow on which validation techniques are the most robust.

Although the validation techniques provide unambiguous results, it would be advantageous to use at least one validation method, at least the visual inspection of dendrograms, if no other tools are available. Among the listed techniques, NbClust would be a reasonable choice as it uses multiple clustering indices that measure different aspects of the clusterings and the final cluster number is suggested based on a voting scheme. Additionally, it is easy to run and interpret the results. Although Euclidean distance, Ward's method seems a safe and reasonable choice, testing, and validation of different clustering combinations is strongly suggested.

CRediT authorship contribution statement

Attila Gere: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The author thanks the support of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and the support of the National Research, Development, and Innovation Office of Hungary (OTKA, contracts No FK 137577 and K 134260).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2023.100522.

References

Araújo, C.I.A., Sant'Anna, L.J., da Silva Moreira, E., de Paula, M.C., Della Lucia, S.M., de Carvalho, R.V., Saraiva, S.H., Lima, R.M., Lima Filho, T., 2021. How much can sodium chloride be substituted for potassium chloride without affecting the sensory acceptance of cracker-type biscuits? Food Res. Int. 150, 110798 https://doi.org/10.1016/j.foodres.2021.110798.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recogn. 46 (1), 243–256. https://doi.org/10.1016/j.patcog.2012.07.021.

Ares, G., Jaeger, S.R., 2013. Check-all-that-apply questions: influence of attribute order on sensory product characterization. Food Qual. Prefer. 28 (1), 141–153. https://doi.org/10.1016/j.foodqual.2012.08.016.

Baykasoğlu, A., Subulan, K., 2016. A multi-objective sustainable load planning model for intermodal transportation networks with a real-life application. Transport. Res. E Logist. Transport. Rev. 95, 207–247. https://doi.org/10.1016/j.tre.2016.09.011.

Brock, G., Pihur, V., Datta, S., Datta, S., 2008. clValid: an R package for cluster validation. J. Stat. Software 25 (4), 1–22. https://doi.org/10.18637/jss.v025.i04.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: an R package for determining the relevant number of clusters in a data set. J. Stat. Software 61 (6), 1–36. https://doi.org/10.18637/jss.v061.i06.

Dery, E.K., Carey, E.E., Ssali, R.T., Low, J.W., Johanningsmeier, S.D., Oduro, I., Boakye, A., Omodamiro, R.M., Yusuf, H.L., 2021. Sensory characteristics and consumer segmentation of fried sweetpotato for expanded markets in Africa. International Journal of Food Science \& Technology 56 (3), 1419–1431. https://doi.org/10.1111/ijfs.14847.

Deza, M.M., Deza, E., 2013. Encyclopedia of Distances, first ed. Springer-Verlag. https://doi.org/10.1007/978-3-642-30958-8.

Di Cairano, M., Condelli, N., Cela, N., Sportiello, L., Caruso, M.C., Galgano, F., 2022. Formulation of gluten-free biscuits with reduced glycaemic index: focus on in vitro glucose release, physical and sensory properties. LWT 154, 112654. https://doi.org/10.1016/j.lwt.2021.112654.

Galili, T., 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31 (22), 3718–3720. https://doi.org/10.1093/bioinformatics/btv428.

Gareth, J., Witten, D., Hastie, T., Tibshirani, R., 2013. Classification. In: James, G., Witten, D., Hastie, T., Tibshirani, R. (Eds.), An Introduction to Statistical Learning with Applications in R. Springer-Verlag New York, pp. 127–173. https://doi.org/10.1007/978-1-4614-7138-7.

Gere, A., Kovács, S., Pásztor-Huszár, K., Kókai, Z., Sipos, L., 2014. Comparison of preference mapping methods: a case study on flavored kefirs. Journal of Chemometrics 28 (4), 293–300. https://doi.org/10.1002/cem.2594.

Gere, A., Sipos, L., Kovács, S., Kókai, Z., Héberger, K., 2017. Which just-about-right feature should be changed if evaluations deviate? A case study using sum of ranking differences. Chemometrics and Intelligent Laboratory Systems 161, 130–135. https://doi.org/10.1016/j.chemolab.2016.12.007.

Hämäläinen, J., Jauhiainen, S., Kärkkäinen, T., 2017. Comparison of internal clustering validation indices for prototype-based clustering. Algorithms 10 (3). https://doi.org/10.3390/a10030105.

Hastie, M., Torrico, D.D., Hepworth, G., Jacob, R., Ha, M., Polkinghorne, R., Warner, R. D., 2022. Combining hierarchical clustering and preference mapping differentiates consumer preference for dry aged mutton. Meat Science 192, 108890. https://doi.org/10.1016/j.meatsci.2022.108890.

James, Gareth, Witten, D., Hastie, T., Tibshirani, R., 2021. Unsupervised learning. In: An Introduction to Statistical Learning with Applications in R, second ed. Springer, pp. 521–529.

Kassambara, A., 2017. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, 1st ed. STHDA.

Lee, I., Yang, J., 2009. Common clustering algorithms. In: Brown, S.D., Taule, R., Walczak, B. (Eds.), Comprehensive Chemometrics, first ed., Vol. 2. Elsevier B.V, pp. 577–618. https://doi.org/10.1016/B978-044452701-1.00064-8.

Leisch, F., 2016. Resampling methods for exploring cluster stability. In: Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.), Handbook of Cluster Analysis, first ed. Chapman and Hall/CRC, pp. 637–652.

Lu, N.-V., Tansuchat, R., Yuizono, T., Huynh, V.-N., 2020. Incorporating active learning into machine learning techniques for sensory evaluation of food. International Journal of Computational Intelligence Systems 13 (1), 655–662. https://doi.org/10.2991/ijcis.d.200525.001.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2019. *Cluster: Cluster Analysis Basics And Extensions. R Package* (2.1.0.). CRAN.

Meila, M., 2016. Criteria for comparing clusterings. In: Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.), Handbook of Cluster Analysis, first ed. Chapman and Hall/CRC, pp. 619–636.

Milligan, G.W., Cooper, M.C., 1988. A study of standardization of variables in cluster analysis. Journal of Classification 5 (2), 181–204. https://doi.org/10.1007/BF01897163.

Omidvarnia, A., Zalesky, A., Mansour, L.S., Van De Ville, D., Jackson, G.D., Pedersen, M., 2021. Temporal complexity of fMRI is reproducible and correlates with higher order cognition. NeuroImage 230, 117760. https://doi.org/10.1016/j.neuroimage.2021.117760.

Palczak, J., Blumenthal, D., Delarue, J., 2020. Influence of sensory complexity on preferences for novel gourmet dairy desserts. Does Berlyne's theory apply to desserts? Food Quality and Preference 84, 103957. https://doi.org/10.1016/j.foodqual.2020.103957.

R Core Team, 2022. R: a language and environment for statistical computing. https://www.r-project.org/.

Ramsey, I., Yang, Q., Fisk, I., Ford, R., 2021. Understanding the sensory and physicochemical differences between commercially produced non-alcoholic lagers, and their influence on consumer liking. Food Chemistry: X (9), 100114 https://doi.org/10.1016/j.fochx.2021.100114. June 2020.

Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66 (336), 846–850. https://doi.org/10.1080/01621459.1971.10482356.

Rohlf, F.J., Fisher, D.R., 1968. Tests for hierarchical structure in random data sets. Systematic Zoology 17 (4), 407–412. https://doi.org/10.2307/2412038.

Rothman, L., Parker, M., 2009. Just-about-right (JAR) scales: design, usage, benefits, and risks. In: Rothman, L., Parker, M. (Eds.), ASTM Standard. ASTM International. https://doi.org/10.1520/MNL63-EB.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Saraçli, S., Doğan, N., Doğan, İ., 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of Inequalities and Applications 2013 (1), 203. https://doi.org/10.1186/1029-242X-2013-203.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.-T., 2017. A review of clustering techniques and developments. Neurocomputing 267, 664–681. https://doi.org/10.1016/j.neucom.2017.06.053.

Shirkhorshidi, A.S., Aghabozorgi, S., Wah, T.Y., 2015. A comparison study on similarity and dissimilarity measures in clustering continuous data. PLOS ONE 10 (12), 1–20. https://doi.org/10.1371/journal.pone.0144059.

Škrobot, D., Pezo, L., Tomić, J., Pestorić, M., Sakač, M., Mandić, A., 2022. Insights into sensory and hedonic perception of wholegrain buckwheat enriched pasta. LWT 153, 112528. https://doi.org/10.1016/j.lwt.2021.112528.

Sokal, R.R., Rohlf, F.J., 1962. The comparison of dendrograms by objective methods. Taxon 11 (2), 33–40. https://doi.org/10.2307/1217208.
Sridhar, K., Charles, A.L., 2021. Proximate, functional, and sensory properties of Kyoho grape (Vitis labruscana) skin herbal infusions: potential as sustainable novel functional beverages. LWT 152, 112289. https://doi.org/10.1016/j.lwt.2021.112289.
Ullmann, T., Hennig, C., Boulesteix, A.-L., 2022. Validation of cluster analysis results on validation data: a systematic framework. WIREs Data Mining and Knowledge Discovery 12 (3), e1444. https://doi.org/10.1002/widm.1444.
von Luxburg, U., 2010. Clustering stability: an overview. Foundations and Trends® in Machine Learning 2 (3), 235–274. https://doi.org/10.1561/2200000008.
Yang, Q., Kraft, M., Shen, Y., MacFie, H., Ford, R., 2019. Sweet Liking Status and PROP Taster Status impact emotional response to sweetened beverage. Food Quality and Preference 75 (November 2018), 133–144. https://doi.org/10.1016/j.foodqual.2019.02.016.
Yenket, R., Chambers, E., 2017. Influence of cluster analysis procedures on variation explained and consumer orientation in internal and external preference maps. Journal of Sensory Studies 32 (5), 1–11. https://doi.org/10.1111/joss.12296.
Zaki, M.J., Meira, W.J., 2020. Hierarchical clustering. In: Data Mining and Machine Learning: Fundamental Concepts and Algorithms, second ed. Cambridge University Press, pp. 368–369.