Proceeding Paper

Prediction of Aroma Partitioning Using Machine Learning †

Marvin Anker [1], Christian Krupitzer [1], Yanyan Zhang [2] and Christine Borsum [1,3,*]

1 Department of Food Informatics and Computational Science Hub, University of Hohenheim, 70599 Stuttgart, Germany; marvin.anker@uni-hohenheim.de (M.A.); christian.krupitzer@uni-hohenheim.de (C.K.)
2 Department of Flavor Chemistry, University of Hohenheim, 70599 Stuttgart, Germany; yanyan.zhang@uni-hohenheim.de (Y.Z.)
3 Department of Process Engineering (Essential Oils, Natural Cosmetics), University of Applied Sciences Kempten, 87406 Kempten, Germany; christine.borsum@hs-kempten.de (C.B.)
* Correspondence: christine.borsum@uni-hohenheim.de; Tel.: +49-173-7310-019
† Presented at the 2nd International Electronic Conference on Processes: Process Engineering—Current State and Future Trends (ECP 2023), 17–31 May 2023; Available online: https://ecp2023.sciforum.net/.

Abstract: Intensive research in the field over the past decades highlighted the complexity of aroma partition. Still, no general model for predicting aroma matrix interactions could be described. The vision outlined here is to discover the blueprint for the prediction of aroma partitioning behavior in complex foods by using machine learning techniques. Therefore, known physical relationships governing aroma release are combined with machine learning to predict the K_{mg} value of aroma compounds in foods of different compositions. The approach will be optimized on a data set of a specific food product. Afterward, the model should be transferred using explainable artificial intelligence (XAI) to a different food category to validate its applicability. Furthermore, we can transfer our approach to other relevant questions in the food field such as aroma quantification, extraction processes, or food spoilage.

Keywords: aroma release; food reformulation; machine learning; explainable artificial intelligence

1. Introduction

Two major challenges of modern societies are nutrition-related diseases and climate change. Both topics are creating forces to change the composition of our food. On the one hand, fat, salt, or sugar content must be reduced to lower the health risks and costs associated with cardiovascular diseases, obesity, and diabetes. On the other hand, protein sources are moving from animal to plant origin, driven by concerns about animal welfare as well as the environmental effects of raw material production. However, nutritional habits are only changed long-term if the alternative product is matching the original's sensory properties. This is why understanding the effect of compositional changes on aroma perception is highly relevant to tackle the challenges of today's world.

Extensive research has been conducted on the topic of aroma release to explain the perception formed in the brain. Aroma perception is a complex phenomenon, as it depends on physiological parameters showing large inter-individual differences (e.g., saliva, breathing) [1], and it shows cross-modalities to our other sensory inputs, i.e., texture and taste. However, from the food perspective aroma, the release mainly depends on the interactions between the aroma compound and the ingredients of the food (fat, carbohydrates, proteins, etc.) in a defined food environment (pH, temperature). The strength of these interactions can be quantified by the partition coefficient K_{mg}, defined as the quotient between the flavor concentration in the food, and the concentration in the headspace above the food [1]. The K_{mg} value is determined in equilibrium; thus, it describes the thermodynamic end state. However, it also determines the kinetics of aroma release, as the release rate is higher for aroma compounds showing weak interactions with the matrix.

When the composition of a food is changed, e.g., to decrease the fat content, the aroma profile is changed completely, because every aroma compound shows different interactions with the fat phase. To compensate for this change, it is crucial to know the K_{mg} value of this compound in the new food composition. In combination with the aroma compound concentration in the food, the concentration of aroma, which could be released during oral processing, can be estimated. This is why it is crucial for the acceptance of reformulated food products to predict the change in aroma partition caused by the change in composition.

In food technology, often, the basic physical principles are known, but the complexity in the composition of the matrices is forcing research to work very empirically. Machine learning could close this gap since it could combine background knowledge with large empirical data sets to build prediction models. Understanding these models enables the transfer of learnings into the development of healthy and sustainable foods.

In this vision paper, we describe our concept for a machine learning-based analysis of the K_{mg} value and its composition. Based on such an approach, it would be feasible to model the composition of the K_{mg} value in more detail and analyze how changing one factor (e.g., amount of fat, sugar content, or protein type) would influence the sensory perception of the food. We decide to target an explainable artificial intelligence (XAI) approach, as such XAI approaches can explain the results of the machine learning analysis.

The remainder of this paper is structured as follows. Next, we describe relevant concepts in Section 2: aroma–matrix interaction, prediction of aroma–matrix interactions, and machine learning. After, Section 3 introduces our concept for XAI-based aroma prediction. Finally, Section 4 summarizes the paper and presents relevant research challenges.

2. Background

2.1. Aroma–Matrix Interaction

Aroma compounds can interact with food ingredients in different ways: interactions such as hydrogen bonds, electrostatic interactions, van-der-Waals, or hydrophobic interactions being the most common ones, will be the focus of this project, as they determine the aroma perception of food [2]. Figure 1 illustrates the different processes involved during the aroma release from a food:

1. The aroma partition coefficient describes the state in closed packaging; therefore, the first orthonasal impression of the food is determined by the aroma concentration in the headspace c_g.
2. During oral processing, the food is cooled down or warmed up to a physiological temperature, mixed with saliva, and mechanically processed [3].
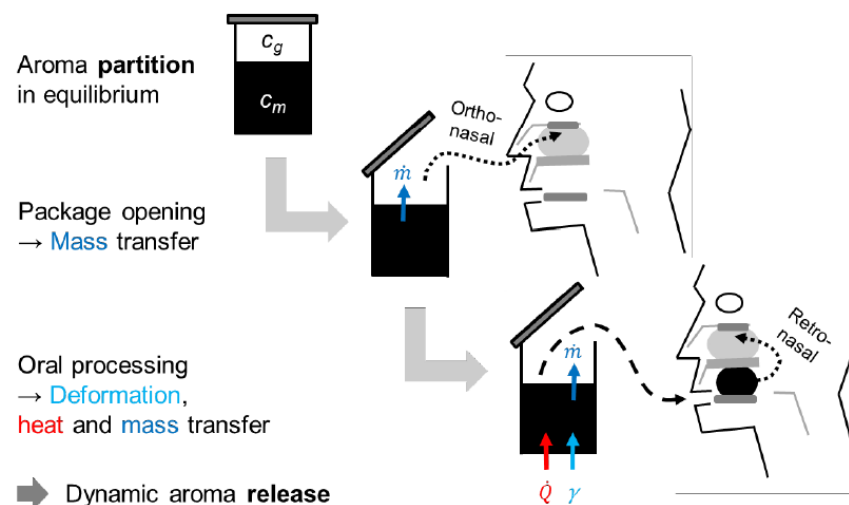

Figure 1. Physical processes influencing aroma perception of food.

These processes lead to a fast release of the reversibly bound aroma compounds in the food (c_m), leading to retronasal aroma perception. This is why knowing real concentrations in matrix and headspace, calculated from the K_{mg} value, is the basis for the development of reformulated foods with a similar aroma perception to the original.

2.2. Prediction of Aroma–Matrix Interactions

Aroma–matrix interactions depend on the chemical properties of the aroma compound on the one hand and on the composition and processing of the food on the other. In a nutshell, all prediction models are built up in a similar way. First, the dominating mechanism of interaction with the studied food compound is determined, e.g., with hydrogen binding or hydrophobic interactions. Second, a coefficient needs to be found that quantifies the ability of the aroma compound for these interactions. In the case of hydrophobic interactions, this is often the *log P* value [4], and the logarithmic partition ratio between octanol (c_o) and water (c_w) (see Equation (1)). Moreover, the chain length of the molecule is a relevant parameter for hydrophobicity [5].

$$logP = log(c_o/c_w) \tag{1}$$

Third, the method of partial least square regression (PLSR) is used to find the correlation function to link the coefficients with the correct weights to the output, the K_{mg} value. This method is called the quantitative structure property or activity relationship (QSPR or QSAR) [6,7], as it is using the information of the chemical structure to predict a coefficient of functionality, in this case, the aroma–matrix interaction.

The influence of food ingredients on the K_{mg} value has also been extensively studied [2]; however, most studies were conducted focusing on one ingredient, for example, beta-lactoglobulin [4,8]. However, this knowledge is only a basis for understanding aroma–matrix interactions of complex food. Reformulating foods need a model able to describe and compare aroma partition in real foods containing lipids, proteins, carbohydrates, salts, and water. Fat, for example, can bind much more aroma compounds than proteins [9,10]. Additionally, the inclusion of a gas phase also significantly changes aroma binding in foods [11]. This is why the number of parameters influencing aroma matrix interactions in real foods is larger than the simplified models described in scientific literature.

In addition to the compositional complexity, food processing also influences aroma–matrix interactions. Heating steps have an influence on protein conformation, thus influencing possible binding sites of aroma compounds [12,13]. Microbiological fermentation steps are also relevant, as they often decrease the pH value. If electrostatic interactions bind aroma compounds, they will be affected by changes in pH [14]. Additionally, the protein conformation depends on pH if denaturation takes place. This process can also be caused by changes in salt content [15].

2.3. Machine Learning

Machine learning methods are used to find and describe relationships between different attributes in a large data set to predict, classify, or forecast one or more output parameters. In the food domain, big data methods are already being applied in several fields, such as agricultural production, product innovation, food quality, and food safety. For example, IoT and big data analysis in agriculture can decrease the usage of herbicides by crop and weed imaging [16]. Food safety can be increased by traceability through blockchain technology to modernize the supply chain [17], and food waste can be reduced by using intelligent packaging indicating the degree of freshness [18].

Machine learning can hold the key to determining the most influential factors and their dependencies for a complex process such as aroma release, which is influenced by various parameters from the aroma and food side. The physical models predicting aroma release presented so far focus on specific protein–aroma interactions, but real food systems are far more complex. Predictions using more parameters were tried with multiple linear modeling [19]. However, not all aroma compounds could be described successfully due to

their non-linear behavior. It was demonstrated that combining PLS and artificial neural networks (ANN) improved the prediction accuracy of the consumer liking of green tea beverages [20].

3. Approach

The approach's first major step is identifying and evaluating suited input parameters. The selection of the input parameters is based on the known physical laws of flavor release. Then, the data sets should be firstly verified by the reproduction of selected experiments. This is important because the accuracy of the machine learning model relies heavily on raw data quality. As an example, even though the K_{mg} is determined in most cases with the phase ratio variation (PRV) method, differences in the results are expected. This is due to different analytical settings, e.g., sensitivity of the used detector or sampling method. Since the K_{mg} value is determined using the slope of a linear fit and the lines intersect, it is very susceptible to deviations.

After finding and evaluating the input parameters, an algorithm has to be selected and optimized until the K_{mg} values for all aroma compounds are successfully predicted. The "No Free Lunch Theorem" states that no universal best machine learning method exists. Therefore, a machine learning algorithm must fit the data, which is why different methods of machine learning for prediction will be compared, including decision trees (random forest and eXtreme gradient boosting), support vector machines, k-nearest neighbor algorithms, and ANNs. Since different machine learning models may be needed for different patterns in the data, respectively, for describing the ratios for different flavors, the integration mechanisms of adaptive software are also useful, e.g., to implement a recommendation service to support the selection of the best algorithm as well as its parameter configuration (hyper-parameter tuning). The validation of the machine learning method is performed continuously while the method is selected and optimized. Accordingly, the machine learning algorithm and optimization selection are not sequential but iterative.

Usually, the optimization goal for machine learning models is the accuracy of prediction. In our setting, we also aim for transparency, as we need to understand the models for not only getting a prediction of the K_{mg} value but also to explain changes in the value when adjusting the composition of the food.
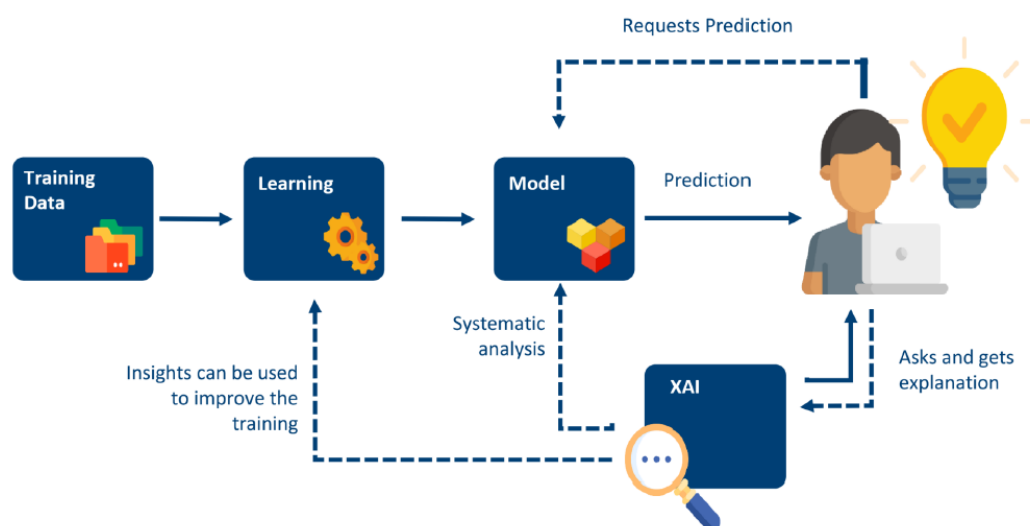

Figure 2. Process of machine learning-based prediction with an additional XAI component to explain the results to the users.

One of the disadvantages of machine learning, especially using Deep learning procedures with ANNs, is the missing transparency of the prediction and the mistrust derived from it; such machine learning models are often referred to as "black boxes". However,

with the help of explainable artificial intelligence (XAI), there is an opportunity to learn from the models [21,22]. For example, physicochemical relationships could be inferred from data-driven models, e.g., the influence of the functional group exceeds that of the chain length. Figure 2 illustrates how XAI can be included in a machine learning prediction to increase the understanding of the results from the model. Therefore, the normal process for machine learning is complemented by an XAI component. This XAI component either extracts the explanation directly from the learned model if the model is transparent (e.g., this is valid for decision trees), or the component learns how the model comes to a prediction, e.g., by integrating the existing scientific models for the determination of the K_{mg} value.

In addition, performance as computational time is an important parameter for the machine learning process. Validation of the machine learning method will be performed continuously, as the method is selected, optimized, and transferred using state-of-the-art machine learning evaluation procedures and metrics. We will focus on the process of learning the model rather than only learning one specific model for a specific food. Hence, the process can be re-used with other products if the relevant K_{mg} values for learning are available along with the receipts of that food.

4. Discussion and Outlook

There are several remaining challenges that we need to solve for a functioning approach. First, our approach requires training data, i.e., measurements of the K_{mg} values for a specific food composition. For milk, there are databases available; hence, we have started to focus on this category of products. Second, we need to identify the best-fitting machine learning algorithms. According to the "No free lunch" theory from optimization sciences, which is also valid for machine learning approaches, no single optimization/machine learning algorithm is superior in all settings. In machine learning, the choice of the best machine learning algorithm is influenced by the data patterns. Third, if the best algorithm lacks interpretability, i.e., the model is not explainable, we must implement the XAI component. There are approaches available that fully focus on the data used for learning and are process-independent, i.e., independent of the specific machine learning algorithm. However, we plan to implement an XAI approach that integrates the existing scientific models for predicting the K_{mg} value, as those models enable an additional validation of the machine learning process.

The novel approach envisioned in this paper can be transferred to new food systems, e.g., plant-based food products, which serve as alternatives to milk from an animal source or meat. The results of the XAI component will decrease the time and complexity of developing the machine learning model to predict aroma partition in plant-based products. This is how our research can contribute to developing more ecologically sustainable food products. Our approach could not only be transferred to the prediction of aroma partition in other food products but also much wider transfered into the food field using data science approaches [23]. The knowledge of aroma partition is also highly relevant in aroma analysis, as quantification is often performed in the headspace of an equilibrated product, e.g., via solid phase microextraction (SPME). The model established in this project could be used to calculate the aroma concentration in the food via headspace analysis, enabling aroma quantification in food without extraction of the matrix. We also see the potential here, to model the relations as a digital food twin [24,25]. Taking the approach a step further to food process engineering, substance transfer during extraction processes could be predicted; either the goal could be an increase in valuable substances, e.g., essential oils, or the decrease in unwanted substances, e.g., furocoumarins in citrus oils. As a third idea relevant to food safety, the approach could be used to predict microbial spoilage in complex food systems if composition, processing, and environmental data from storage were included in the model.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial neural network
PLS	Partial least square
PLSR	Partial least square regression
PRV	Phase ratio variation
QSAR	Quantitive activity relationship
QSPR	Quantitative structure property
SPME	Solid phase microextraction
XAI	Explainable artificial intelligence

References

1. Guichard, E.; Salles, C. *Flavor: From Food to Behaviours, Wellbeing and Health*; Guichard, E., Salles, C., Voilley, A., Eds.; Woodhead Publishing, Sawston, UK: **2016**; pp. 3–22. https://www.sciencedirect.com/book/9780081002957
2. Guichard, E. Interactions between flavor compounds and food ingredients and their influence on flavor perception. *Food Rev. Int.* **2002**, *18*, 49–70. https://doi.org/10.1081/FRI-120003417
3. Chen, J. Food oral processing—A review. *Food Hydrocoll.* **2009**, *23*, 1–25. https://doi.org/10.1016/j.foodhyd.2007.11.013
4. Tromelin, A.; Guichard, E. Interaction between flavour compounds and β-lactoglobulin: Approach by NMR and 2D/3D-QSAR studies of ligands. *Flavour Fragr. J.* **2006**, *21*, 13–24. https://doi.org/10.1002/ffj.169
5. Tan, Y.; Siebert, K.J. Modeling bovine serum albumin binding of flavor compounds (alcohols, aldehydes, esters, and ketones) as a function of molecular properties. *J. Food Sci.* **2008**, *73*, S56–S63. https://doi.org/10.1111/j.1750-3841.2007.00591.x
6. Friel, E.N.; Linforth, R.S.T.; Taylor, A.J. An empirical model to predict the headspace concentration of volatile compounds above solutions containing sucrose. *Food Chem.* **2000**, *71*, 309–317. https://doi.org/10.1016/S0308-8146(00)00190-4
7. Tromelin, A. Prediction of perception using structure–activity models. In *Flavor*; Guichard, E., Salles, C., Voilley, A., Eds.; Woodhead Publishing, Sawston, UK: **2016**; pp. 181–200. https://www.sciencedirect.com/book/9780081002957
8. Andriot, I.; Harrison, M.; Fournier, N.; Guichard, E. Interactions between methyl ketones and β-lactoglobulin: Sensory analysis, headspace analysis, and mathematical modeling. *J. Agric. Food Chem.* **2000**, *48*, 4246–4251. https://doi.org/10.1021/jf991261z
9. Guichard, E. Flavour retention and release from protein solutions. *Biotechnol. Adv.* **2006**, *24*, 226–229. https://doi.org/10.1016/j.biotechadv.2005.11.003
10. Heilig, A.; Heimpel, K.; Sonne, A.; Schieberle, P.; Hinrichs, J. An approach to adapt aroma in fat-free yoghurt systems: Modelling and transfer to pilot scale. *Int. Dairy J.* **2016**, *56*, 101–107. https://doi.org/10.1016/j.idairyj.2016.01.011
11. Thomas, C.F.; Ritter, J.; Mayer, N.; Nedele, A.K.; Zhang, Y.; Hinrichs, J. What a difference a gas makes: Effect of foaming on dynamic aroma release and perception of a model dairy matrix. *Food Chem.* **2022**, *378*, 131956. https://doi.org/10.1016/j.foodchem.2021.131956
12. Wang, K.; Arntfield, S.D. Binding of selected volatile flavour mixture to salt-extracted canola and pea proteins and effect of heat treatment on flavour binding. *Food Hydrocoll.* **2015**, *43*, 410–417. http://dx.doi.org/10.1016/j.foodhyd.2014.06.011
13. Guo, J.; He, Z.; Wu, S.; Zeng, M.; Chen, J. Binding of aroma compounds with soy protein isolate in aqueous model: Effect of preheat treatment of soy protein isolate. *Food Chem.* **2019**, *290*, 16–23. https://doi.org/10.1016/j.foodchem.2019.03.126
14. Guo, J.; He, Z.; Wu, S.; Zeng, M.; Chen, J. Binding of aromatic compounds with soy protein isolate in an aqueous model: Effect of pH. *J. Food Biochem.* **2019**, *43*, e12817. https://doi.org/10.1111/jfbc.12817
15. Ammari, A.; Schroen, K. Flavor Retention and Release from Beverages: A Kinetic and Thermodynamic Perspective. *J. Agric. Food Chem.* **2018**, *66*, 9869–9881. https://doi.org/10.1021/acs.jafc.8b04459
16. Misra, N.N.; Dixit, Y.; Al-Mallahi, A.; Bhullar, M.S.; Upadhyay, R.; Martynenko, A. IoT, big data and artificial intelligence in agriculture and food industry. *IEEE Internet Things J.* **2020**, *9*, 6305–6324. https://doi.org/10.1109/JIOT.2020.2998584
17. Cruz Introini, S.; Boza, A.; Alemany, M.M.E. Traceability in the Food Supply Chain: Review of the literature from a technological perspective. *Dir. Organ.* **2018**, *64*, 50–55. http://hdl.handle.net/10251/121089
18. Müller, P.; Schmid, M. Intelligent packaging in the food sector: A brief overview. *Foods* **2019**, *8*, 16. https://doi.org/10.3390/foods8010016

19. Heilig, A.; Sonne, A.; Schieberle, P.; Hinrichs, J. Determination of aroma compound partition coefficients in aqueous, polysaccharide, and dairy matrices using the phase ratio variation method: A review and modeling approach. *J. Agric. Food Chem.* **2016**, *64*, 4450–4470. https://doi.org/10.1021/acs.jafc.6b01482

20. Yu, P.; Low, M.Y.; Zhou, W. Development of a partial least squares-artificial neural network (PLS-ANN) hybrid model for the prediction of consumer liking scores of ready-to-drink green tea beverages. *Food Res. Int.* **2017**, *103*, 68–75. https://doi.org/10.1016/j.foodres.2017.10.015

21. Adadi A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

22. Burkardt, M.; Huber, M.F. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Intell. Res.* **2021**, *70*, 245–317. http://dx.doi.org/10.1613/jair.1.12228qslack

23. Krupitzer, C.; Stein, A. Food Informatics—Review of the Current State-of-the-Art, Revised Definition, and Classification into the Research Landscape. *Foods* **2021**, *10(11)*, 2889. https://doi.org/10.3390/foods10112889

24. Henrichs, E.; Noack, T.; Pinzon Piedrahita, A.M.; Salem, M.A.; Stolz, J.; Krupitzer C. Can a Byte Improve Our Bite? An Analysis of Digital Twins in the Food Industry. *Sensors* **2021**, *22(1)*, 115. https://doi.org/10.3390/s22010115

25. Krupitzer, C.; Noack, T.; Borsum, C. Digital Food Twins Combining Data Science and Food Science: System Model, Applications, and Challenges. *Processes* **2022**, *10(9)*, 1781. https://doi.org/10.3390/pr10091781