

## Article

# Digital Food Twins Combining Data Science and Food Science: System Model, Applications, and Challenges <sup>†</sup>

Christian Krupitzer \* , Tanja Noack  and Christine Borsum

Food Informatics Department &amp; Computational Science Lab, University of Hohenheim, 70599 Stuttgart, Germany

\* Correspondence: christian.krupitzer@uni-hohenheim.de; Tel.: +49-711-459-23664

† Extended version of a paper presented at the 1st International Electronic Conference on Processes: Processes System Innovation, Online, 17–31 May 2022. Please find the conference publication here: Krupitzer, C.; Noack, T.: DigiFoodTwin: Digital Biophysical Twins Combined with Machine Learning for Optimizing Food Processing. *Eng. Proc.* **2022**, *19*(1), 42. <https://doi.org/10.3390/ECP2022-12623>.

**Abstract:** The production of food is highly complex due to the various chemo-physical and biological processes that must be controlled for transforming ingredients into final products. Further, production processes must be adapted to the variability of the ingredients, e.g., due to seasonal fluctuations of raw material quality. Digital twins are known from Industry 4.0 as a method to model, simulate, and optimize processes. In this vision paper, we describe the concept of a digital food twin. Due to the variability of the raw materials, such a digital twin has to take into account not only the processing steps but also the chemical, physical, or microbiological properties that change the food independently from the processing. We propose a hybrid modeling approach, which integrates the traditional approach of food process modeling and simulation of the bio-chemical and physical properties with a data-driven approach based on the application of machine learning. This work presents a conceptual framework for our digital twin concept based on explainable artificial intelligence and wearable technology. We discuss the potential in four case studies and derive open research challenges.



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

Academic Editor: Jean-Pierre Corriou

Received: 27 July 2022

Accepted: 23 August 2022

Published: 5 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** digital twin; food processing; Industry 4.0; artificial intelligence; machine learning; self-aware computing systems

## 1. Introduction

The term Industry 4.0 refers to current technological changes in the environment of industrial production enabled by advances in information technology. The focus of Industry 4.0 is the smart factory, i.e., the connection of cyber-physical production systems with Internet of Things (IoT) technology as well as intelligent data analysis. A core element of Industry 4.0 is the digital twin: a virtual model of a product, the machines, or the production process created with data collected by sensors that enables simulations or real-time analyses of the production status [1]. As a digital twin integrates real-time data, it provides a detailed model that can support decision making through simulation.

The use of digital twins seems beneficial in food processing for various reasons. The Corona pandemic demonstrated the vulnerability of food supply chains and thus the need for higher resilience. To ensure the supply of food, production processes must allow a high flexibility and adaptivity which requires traceability. The survey “Die Ernährung 4.0-Status Quo, Chancen und Herausforderungen” (Nutrition 4.0-Status Quo, Opportunities and Challenges) conducted by the digital association Bitkom and the Federation of German Food and Drink Industries showed that 70% of the more than 300 companies surveyed in the food industry consider end-to-end traceability from the origin of the goods to the customer to be an important scenario for the current decade [2]. Various types of sensors exist to support this. However, the potential is far from being exploited. Furthermore, product quality is influenced by different quality levels of input materials. Especially in case of seasonal fluctuations of this raw material quality, an adjustment of parameters in

the production process is essential. Introducing new products that are related to existing ones is also a challenge in food processing. Introduction processes of new products could be simplified by a digital twin of already existing products. The digital twin is able to learn the correct process parameters for production, and is used as knowledge foundation within a self-adaptive system [3]. All those application scenarios show the potential of digital twins in the food supply chain and their huge potentials, e.g., in determining food quality, traceability, or designing personalized foods.

In a recent literature review of the state-of-the-art of digital twins in agriculture and the food industry, we identified the following issues [4]:

(1) The survey revealed that the application of digital twins mainly either targets the production (agricultural production) or the food processing stage.

(2) Nearly all applications are used for monitoring and many for prediction. However, only a small amount focuses on the integration in systems for autonomous control or providing recommendations to humans.

(3) The main challenges of implementing digital twins are combining multidisciplinary knowledge and providing enough data as a digital twin of food production has additional specific requirements compared to digital twins of the production of material goods.

(4) Due to the variability of raw materials, the digit twins cannot be based only on the processing steps, but must also take into account the chemical, physical or (micro)biological properties of the food.

This vision paper aims to provide a concept that complements the typical, retrospective analysis of machine and process data with short-term (detection of potential problems), and medium-term data analysis approaches (planning and optimization) as well as product-related analysis for constructing digital food twins. While Industry 4.0 approaches often focus on the analysis of machine data, this paper describes a product-related data analysis as well. Therefore, we propose the integration of existing models and simulations from food science (*white-box* modeling approach) in combination with data-driven machine learning-based data analysis (*black-box* modeling approach) to achieve a hybrid approach for modeling digital food twins. We discuss the application potential of such a digital food twin in four case studies: (i) proactive decision making of adaptation in the food production, (ii) food reformulation and food product development, (iii) improving the scale-up of prototypes to the production environment, and (iv) supporting the resilient food supply with detection of critical events and tracking the current state of food at any time. Further, we derive from those case studies research challenges in the area of digital food twins. Solving those research challenges can be the foundation for an adaptive system that is able to control the process, autonomously react to changes, and continuously improve its performance through learning. Consequently, such a concept helps to better (i) understand the behavior of a food production process, (ii) predict critical situations, and (iii) determine a new plan.

The remainder of the paper is structured as follows. Next, Section 2 describes related concepts and current approaches in literature. Afterwards, Section 3 presents our concept for a digital food twin derived from a hybrid approach that combines data-driven machine learning analysis with traditional chemo-physical modeling and simulation. Afterward, Section 4 discusses the potential of the concept in different case studies. Then, Section 5 derives research challenges for the implementation of our concept out of the case studies. Finally, Section 6 concludes this paper.

## 2. Background and Related Work

This section presents several approaches and concepts that we identified in the literature and which are relevant to the field of digital twins for the food processing industry.

**Smart factory in the food industry:** Current approaches in Industry 4.0 focus on intelligent collection of data with technology from the IoT and its analysis with machine learning algorithms [5]. This includes a variety of data sources, including raw material data, machine data, or customer data (e.g., information about sales or complaints). In particular,

production planning can be optimized with machine learning in this context [6], e.g., using genetic algorithms for optimizing the order of production steps or integrating picture recognition for quality control. Another use case is predictive maintenance of machines [7,8]. However, the focus is primarily on the view of the process and the machines. Internal processes in the food industry are not included and the view of the product is limited to identifying products with bar or QR codes [4]. Proactive adaptation improves system performance as it forecasts adaptation concerns (e.g., through identification of patterns in historical data) and reacts either by preparing an adaptation or adapting [9]. Real-time data of production sites would help to realize proactive adaptation and dynamic adjustment when a disruption takes place.

**Modeling and simulation of food properties:** Food science often proposes to apply modeling or simulation to describe the food properties and characteristics. In the following, we present some examples to indicate the bandwidth of available applications. Myhan, Białobrzewski, and Markowski [10] developed a mathematical model describing the rheological properties of food materials. The authors of [11] reviewed mechanistic and empirical approaches to explain and predict the effect of food matrix on chemical reactivity. Van Boekel discussed the possibility to describe aspects of food, such as color, nutrient content, and safety, in a quantitative way via mathematical models [12]. Further, numerical simulations are applied in the food sector to simulate product or process characteristics. Hartmann analyzed thermodynamic and fluid-dynamic effects of high-pressure treatment by means of numerical simulation [13]. Abdul Ghani et al. simulated natural convection heating within a can of liquid food during sterilization by solving the governing equations for continuity, momentum and energy conservation for an axisymmetric case [14]. The authors of [15] studied the heat transfer into an aerated food matrix using the finite element method. The advantage of such modeling or simulation-based approaches is the white-box approach, i.e., the relations of the different variables can be extracted from and explained using the models. However, due to the high complexity of the modeled aspects, these approaches always require abstractions from different aspects. This can limit their applicability in productive settings. Further, those approaches require specific knowledge of both aspects: the modeling/simulation technique as well as domain knowledge of the food properties.

**Digital twins in the food sector:** Digital twins can be classified in six types—(i) imaginary that simulate reference objects, (ii) digital twins that monitor in real time the state and behavior of an object, (iii) predictive ones that forecast future states and behavior of an object, (iv) prescriptive ones, (v) autonomous digital twins (using artificial intelligence), and (vi) recollection digital twins with historical data [16]. However, there are still few concepts for digital twins specialized for food processing. Further, in a recent review [4], we showed that agri-food digital twins are limited to specific aspects (e.g., animal monitoring, crop management, or hydroponics) rather than generically applicable throughout the value chain. Approaches focus mainly on monitoring and some does support prediction. However, only a small amount focuses on integrating autonomous control. The survey further showed that implementing digital food twins requires multidisciplinary knowledge, especially on how to model the bio-physical processes that change properties of food due to the variability of raw materials. This is so far insufficiently integrated in literature. Combining the strengths of numerical simulation with data driven methods has been shown to predict sensory perception of complex food systems like yoghurt [17]. Therefore, hybrid modeling that complements existing food models and simulations with a data-driven perspective seems to be beneficial—which this vision paper targets. Most closely related to those demands, the *smartFoodTechnologyOWL* initiative investigates the transferability of the digital twin concept to food processing [18]. The focus is on mapping the process for better control of cyber-physical production systems. In order to make quality control of food safer and more efficient, their goal is to continuously generate a “virtual image” of the product during production. Other projects focus on the integration of physical models to better predict the changes to the food through its processing. In [19],

the authors describe the integration of physical, biochemical, and microbiological processes. However, this type of digital twin often lacks the data-driven perspective of the processes and [19] propose to include real-time coupling of sensor data with the digital twin. That would help to foresee problems and proactively react to them. However, the focus is not on adapting the production process based on the gained information nor on processing the data for predicting critical events. Digital twins are used in production for monitoring a production process [20]. Autonomous systems can respond to changes in state during ongoing operation while digital twins can integrate a variety of data like environment data, operational data, and process data [20,21]. Today food process modeling has mostly pure design optimization and costs targets, but there is a great potential in reducing inter-product variability, achieving higher transparency, and reducing use of resources [22].

**Sensors and indicators:** With the help of *indicators*, the presence or absence of a substance, reactions between different substances, or the concentration of a particular substance can be detected. Indicators show the analysis results by direct changes (usually different color intensities) and are placed inside or outside the packaging. Different types of indicators exist. Most common types are time-temperature indicators that show that critical temperatures have been reached; freshness indicators that monitor the quality of food products based on microbiologically motivated or chemical changes in the products; and gas indicators that detect changes in the atmosphere of the package. In contrast to immutable indicators that cannot be reused once they changed their state, *sensors* that are either integrated into the food packaging or in the environment can detect temperature, humidity, pressure on food or vibrations (accelerometers). Specific sensors such as gas sensors or biosensors measure the concentration of certain gases such as carbon dioxide ( $CO_2$ ) or hydro-sulfuric acid, which allow conclusions to be drawn about perishability.  $CO_2$  concentration can be measured using non-dispersive infrared (NDIR) sensors or chemical sensors; infrared sensors as well as electrochemical, ultrasonic, and laser technologies are used to detect the oxygen concentration. Another type of sensors are biosensors based on receivers made of biological materials such as enzymes, antigens, hormones, or nucleic acids. These are used, for example, to identify pathogens such as salmonella, *E.coli*, or listeria. The overview in [23] describes the state-of-the-art in sensor and indicator types. Especially sensors facilitate real-time data collection which supports building a digital twin.

**Contribution:** In the case of the food supply chain, a detailed model of the supply chain, which integrates real-time data to predict supply chain dynamics, can be a promising concept to respond to unexpected events in the whole supply chain including field, factory, retailer, and consumer. The goal of this vision paper is to describe how to create a digital food twin that can be used to track the current state of production at any time, but also trace the food through the food supply chain. While Industry 4.0 approaches often focus on the analysis of machine data, this project aims at also including a product-related data analysis (e.g., the effects of pressure exerted by machines). The main contribution is to provide a framework which describes how to include a data-driven perspective into the traditional approach of modeling food properties within a digital food twin. The sensor measurements are complemented by the result of machine learning methods, continuous simulation, and critical event prediction based on forecasting.

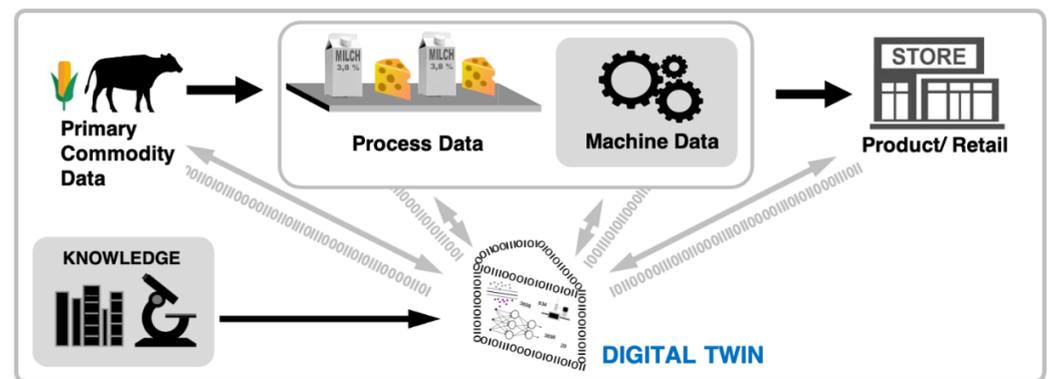
### 3. System Model of a Hybrid Digital Food Twin

This paper presents and discusses a concept that complements the typical modeling or simulation to analyze the processing of food with a data-driven perspective to fill the gap resulting from the abstraction of the models to create a hybrid digital food twin for achieving a real-time, predictive decision making of adaptation in the food supply chain. Consequently, such a concept helps to better (i) understand the behavior of a supply chain, (ii) predict critical situations, and (iii) determine how to adapt the processes.

### 3.1. Data Sources

With the help of machine learning and artificial intelligence, the digital twin is generated from production data and additional data sources (e.g., scientific models, process data, or raw material data) to ensure the traceability of the production and the food status, but also to enable the simulation of the variability of the food in the process operation.

Figure 1 shows our concept of the digital twin. In the figure and the following, we focus on the example of a dairy product (e.g., cheese). The digital twin gets its data from the production site (i.e., sensor, machine, and processing data, e.g., temperature, pressure or pH value) and also integrates raw material data, complaints, and knowledge from experts (e.g., about the handling of production issues). Using different simulation methods based on chemo-physical models and numerical simulations from food science, the digital twin provides information about the actual food processing and supports real-time feedback to the food process operation. Additionally, those simulations based on scientific models help to predict how the product will be changed through the processes as well as the conditions. This information can be used to generate forecasts on how the process steps might influence the quality of the product. Accordingly, the digital twin is suitable for retrospective but also predictive analytics of the process and the quality of the product.



**Figure 1.** The digital food twin which integrates the data from various sources.

We want to illustrate the potential on the example of yogurt fermentation. Applying a traditional digital twin concept as known from Industry 4.0 would be not feasible. Those concepts use process data (mainly from machines) to control and describe the production processes. Without actions from the machines, the product's state will not change. However, for yogurt fermentation, the process is mainly based on resting after inoculation with starter cultures. Hence, the process data cannot describe sufficiently the process as the process happens within the product. When we complement the sparse process information with known models from science to describe the behavior of the bacteria, we get a more accurate picture. But also the model itself would not be sufficient after inoculation with starter cultures, as the model abstracts and each batch of starter culture also has its variations, similar to the milk which properties differ over the year (due to different feeding). Accordingly, a mixture of both is important: The model to understand how the transformation from milk to yogurt works as well as the available data to adjust the model parameters. This is what we want to target with our digital food twin concept.

### 3.2. Data Interpretation

The common approach for extracting information from large datasets would be to rely on deep learning procedures. Those procedures provide a high degree of autonomy in the learning process and are very powerful. The inner workings are thus hidden from the user; the user receives a black-box. However, the resulting models are highly complex and difficult to interpret. This opacity means that intelligent systems cannot be sufficiently validated before they are used productively. This makes it infeasible to combine the resulting prediction with the existing models from food science. Hence, we rely on machine

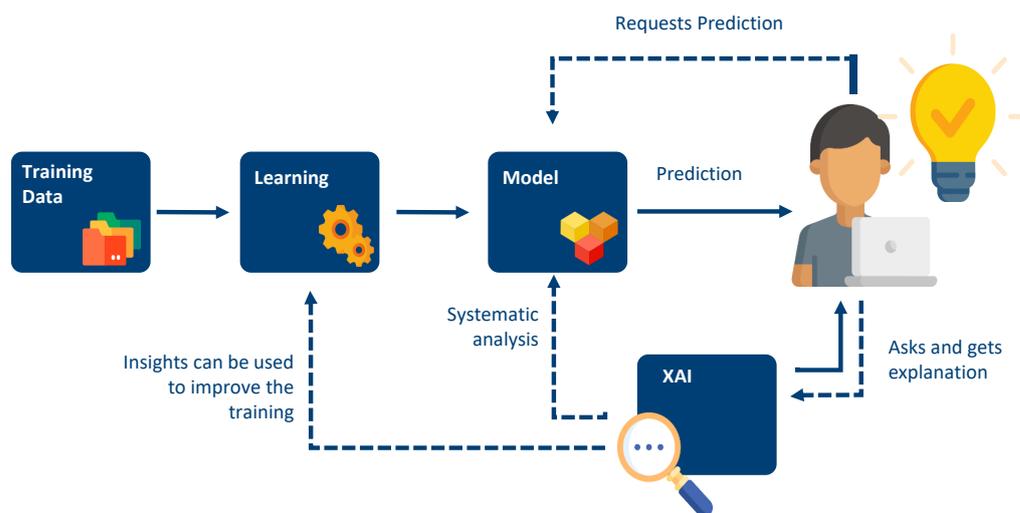
learning procedures, especially from the field of explainable artificial intelligence (XAI), for constructing the digital twin.

XAI describes the systematic explanation and validation of machine learning with the involvement of the users. XAI deals with methods and algorithms that provide humans with explanations as to why a decision was made. Thus, the user is integrated into the machine learning process and can actively contribute to the quality of the system through his/her cognitive abilities, such as those of generalization. XAI bridges the gap between the enormous potential of artificial intelligence and their inevitable risks w.r.t. the missing explainability of the results.

Such XAI approaches help to transform the sensor data into a digital twin model, which can be used for simulation. Further, in contrast to approaches based on artificial neural networks (e.g., deep learning), those XAI models are explainable and humans are able to understand and adjust them. This simplifies the integration of expert knowledge in the learning process. Two approaches for XAI are feasible:

(1) Some machine learning techniques are inherently explainable, e.g., decision trees or random forest. However, those might come to limitations for large data sets and do not support automatic feature extraction as it is the case for deep learning methods.

(2) For non-explainable approaches, such as deep neural networks, the idea is to have a second component called *XAI component*, which tries to use models to explain the results (see Figure 2).



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

We target the second approach as the XAI component can be based on the mentioned models and simulations from food science. Further, the second approach relies on deep learning procedures which have more performance than other machine learning techniques. The first category is preferable if techniques such as random forest can be used as those techniques return explainable models. In general, the decision depends on the use case, the available data set (only large datasets are usable with deep learning techniques), the performance of the techniques with explainable models (such as random forest) from category 1, and if it is feasible to implement the XAI component.

### 3.3. Model Selection

A difficulty lies in the generation of a machine learning workflow, that supports automatically the data pre-processing, choice of a machine learning technique, and learning of machine learning models. Especially the choice of the machine learning technique is important, as the data pattern influence the performance of the technique for learning. Based on the “*No-Free-Lunch Theorem*” [24] from 1997, stating that there is no optimization algorithm best suited for all scenarios, an analogy can be drawn to the domain of machine

learning: there is no machine learning method that performs best for all types of data. Recommendation systems can help to choose the best fitting technique and configure its parameters, i.e., hyper-parameter tuning, based on the characteristics of the data and the data pattern. We plan to adapt and integrate previous work on time series forecasting recommendation systems [9].

One important step for the data-driven analysis using machine learning is the pre-processing of the data, i.e., the generation of features for learning. Usually, this activity requires manual effort and domain expertise. Deep learning algorithms perform this feature extraction automatically, however, have the disadvantage of reduced explainability. Log-based prediction approaches support the structured, automatic extraction of features from log data. Such approaches use historical event-log data to train machine learning algorithms [25]. With log-based prediction, the required steps can be automatized, but are clearly defined and, hence, support the explainability of the data pre-processing. Gutsch et al. presented an approach for predictive maintenance which integrates log-based prediction [25]. Lopez et al. use historical log data emitted by production machines from several industrial factories to predict sufficiently early the occurrence of a critical malfunction in machines [26]. We plan to integrate a similar log-based prediction for automatizing the extraction of relevant data patterns and features from historical data, especially as such an approach might enhance transferability of the digital twin between different food categories.

### 3.4. Integrating Simulation

While Industry 4.0 approaches for digital twins often focus on the analysis of machine data, our concept also integrates the simulation of the internal product states. Still, we so far focus on integration of process data, e.g., collected with in-line sensors or sensors integrated into processing machines. This could be extended by an approach for data collection using replicas of food products created with a 3D printer. Those replicas can be equipped with sensors to allow data collection from the point of the products for drawing conclusions about the processing steps and their influence on the products. Hence, using IoT technology, especially smart miniaturized sensors, and the use of 3D printing, there is an opportunity for collecting intra-process data from the products' point of view. There are two examples for such an approach. The nPotato [27] is an artificial potato equipped with sensors that is laid out in the field while harvesting. The data is analyzed to detect whether the harvester is set correctly. The artificial mango [28] enables an improved thermal profiling during the complete transport in the supply chain. The combination of the proposed food modeling/simulation with data analytics of process data using machine learning (as external process perspective) as well as the integration of product data from those artificial food replicas (as internal product perspective) would support a very precise view on the food items and the process.

## 4. Case Study Analysis

The previous section described our approach for a hybrid digital twin which combines traditional modeling/simulation of food properties (white-box approach) with machine learning enabled analysis for integration specific information (black-box approach). Such an approach can rely on the power of data analysis for filling gaps due to model abstraction for optimizing the parametrization of the models.

This section presents four case studies which show the potential of the digital twin concept. First, Section 4.1 describes the application of the digital twin to support proactive, adaptive food processing. Second, Section 4.2 targets the product development and explains the use of the digital twin for food reformulation. Third, Section 4.3 combines the process and product perspectives and discusses the usefulness of the digital twin to support the scale-up from demonstration to productive systems and also for shelf-life prediction. Last, Section 4.4 focuses the food supply chain and presents how to improve the traceability of food with the digital twin.



are used by the reasoning process to determine the next actions that the system should take to achieve these goals. The generated models can be complemented by other models, which, e.g., describe biological, physical, or chemical relations that influence the food. These actions can affect the behavior of the system and have an impact on the environment as well. The LRA-M loop is adapted as we want to include knowledge provided by the previous introduced digital twin into the framework. Thereby, the knowledge provided by the digital twin is not only a simple knowledge database but processed data which is generated using critical event prediction or different machine learning approaches. The main goal is that the SeAC system adjusts the parameters autonomously or provides recommendations to the user on how to change machine configurations.

#### 4.2. Food Reformulation and Food Product Development

Two of the 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 needs to 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 of animal welfare as well as 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.

Aroma perception is a complex phenomenon, as it depends on physiological parameters showing large inter-individual differences (e.g., saliva, breathing) and it shows cross-modalities to our other sensory inputs, i.e., texture and taste [30,31]. However, from the food perspective, aroma release mainly depends on the interactions between the aroma compound and the ingredients of the food (fat, carbohydrates, proteins, etc.). 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.

Due to the intensive research in the field, a large amount of data for different aroma substances in various food matrices is available. Hence, it would be possible to combine the models describing the known physical relationships governing aroma release with machine learning to generate a digital twin for predicting the  $K_{mg}$  value of aroma compounds in foods of different composition. With the data-driven approach, it would be feasible to generate the digital twin for a specific food category combining machine learning and scientific models for a specific food category, e.g., dairy products. This product category shows many dimensions of variation besides composition (e.g., protein conformation, pH). Afterwards, transfer of the model to products based on plants like soy or legumes is possible, as the data-driven component can adjust the digital twin to the specifics of the other food product.

#### 4.3. Scale-Up and Shelf-Life Prediction

The digital twin can help to decrease the time to market for new products and support the scale-up of the production of new products. It is possible to use the digital twin of a product with similar properties or a similar food matrix, adjust this digital twin, and use it as base to learn the required adjustments in the production process (e.g., new configurations of machines) for fastening the scale-up of new products. Similarly, the digital twin can also support deriving necessary changes of the process parameters, i.e., machine configuration, as a result to the variability of the ingredients' quality or composition.

Additionally, it is feasible to use the digital twin's information for determining the potential shelf-life of a new product based on the observations for similar products and the adjustments of a corresponding existing digital twin for the new product. Also the digital twin models the chemo-physical and microbiological properties of the food. Those relations can be used to simulate the food's perishability. Further, it is also feasible to

integrate a digital twin of the packaging and, by combining this with the digital twin of the food, to determine the interactions between the packaging and the perishability of the food. For example, see the Fraunhofer project on simulation of shelf life [32].

#### 4.4. Resilient Food Supply and Traceability

The food supply chain integrates all process steps and supports a continuous tracking of the food throughout the production process. The digital twin concept could also support various functionalities of the food supply chain. Especially the possibility to simulate various aspects and, through that, predict critical situation in advance (e.g., cold chain violations), help to proactively react, adapt the process, and improve the resilience. This work presents the underlying concept that shows how processed data (e.g., raw material, machine data, etc.) is used as input for the manufacturing site to adapt production processes based on predicting critical situations. Especially the data-driven machine learning component helps to provide a real-time analysis of the food items as it integrates the sensor values. Smart packaging concepts [23] that add sensor technology into packaging can support the data collection.

The digital twin helps not only to model the information of the current condition of the food, but also to improve its traceability. This can be complemented with Blockchain technology [33]. Blockchains offer a distributed data management solution which might be beneficial as those reduce data duplication and increase robustness of the data access. Several authors [34,35] propose to integrate the Blockchain for traceability purposes, as the complete documentation of the origin of ingredients and food is highly important and often a legal obligation.

## 5. Discussion

One of the major challenges of implementing digital twins is the lack of a general method, which describes how to gather the information from the physical to the virtual object [20,22,36]. Koulouris et al. state that the specific characteristics of the food sector and high-value product industries, such as specialized equipment, component complexity, and high-quality standards, are responsible for the delay in the adoption of real-time process simulation for design and modeling [37]. Thus, the individual projects for implementing a digital twin lead to higher investment costs due to the diversity of approaches and, therefore, are particularly challenging in smaller companies and poorer countries [20,22,38]. In the following, we discuss particular research challenges for our vision of hybrid digital twins that we derived from analyzing the described case studies from the previous section.

**Complexity of food:** The complexity and variability of raw materials and their properties used to create food products, the rigidity in the process, and the limited shelf-life not only of food raw materials but also the products made of it are limiting the application [22,37]. Further, plants, processes, and knowledge are continuously changing environments, forcing the related digital twins to improve permanently [39]. The complexity of the food items also require the support for customizing the digital twins.

**Absence of physicochemical models:** The absence of good physicochemical data is another major impediment to the use of modeling and simulation tools [37]. For instance, food processing faces a wide range of foods with insufficiently described properties, hard to calculate or even to predict, such as molecular weight, pH, or water activity, and not so well understood thermodynamics. Furthermore, the kinetics of biological and chemical processes need to be understood and made calculable as physics-based models [22]. This effect is intensified by production mixes, technology variability, and the unpredictability of the physical solution [40], resulting in complex integration of different modeling methods [22]. However, process models can already be incorporated to estimate the energy and material requirements and expected process yield during the food processing [37].

**Explainability of the data analysis:** We rely on machine learning algorithms for the analysis of the data. As mentioned above, we have here the tradeoff between performance of the learning and the explainability. We plan the integration of an XAI component based

on the physicochemical models. However, the field of model-based XAI is a rather new field [41]. Hence, significant contributions in this field are necessary for the combination of the models and data analysis.

**Data fusion:** The combination of the proposed food modeling/simulation with data analytics of process data using machine learning (as external process perspective) as well as the integration of product data from those artificial food replicas (as internal product perspective) would support a very precise view on the food items and the process. This often results in many challenges as conflicting data, corrupted data, or incompatible data.

**Complexity of data integration:** Depending on the complex integration of different methods (modeling, simulation, machine learning) in the digital twin application, the maturity of prescriptive analytic techniques might become a risk due to unreliability, thus a barrier for implementing a digital twin [42]. Here, data security and validation need to be considered [43]. Especially the validation of the captured data is an important aspect, as sensor values are always uncertain to some degree.

**Validation of Data and Digital Twin Models:** One critical issue will be the validation of the used data. Following the famous “garbage in, garbage out” principle for machine learning, which postulates that the quality of learned models will be low if the data quality is low, the validation of data is an important step. Data quality might be influenced especially by sensor faults or missing data, such as parameter changes through manual operation. Hence, the validation of data is an important aspect to guarantee the validity of the models. We plan to integrate automatic feature engineering, such as known from AutoML [44], to address those issues. Further, the quality of the digital twin models is an important aspect. On the one hand, this requires an approach to estimate the quality of the digital twin models. On the other hand, even if the quality is satisfiable at a specific point of time, this quality might be reduced over time, for example, in case of context drift, i.e., the pattern of the data changes so that the learned models does not capture those patterns anymore. Accordingly, besides a metric to estimate the quality of the digital twin models, also a process for continuous learning on-the-fly is important to achieve a high validity for the digital twin models.

**Missing technological infrastructure:** Another challenge is that only by advancing sensor, communication, and data processing technologies, real-time interaction between actual and virtual twins can be achieved [40]. The systems themselves have to enable the implementation of digital twins, i.e., their properties must be known or observable, as well as they have to provide high-quality data [45].

**Missing knowledge of employees:** The required expertise of knowledge becomes a real challenge for project teams [38]. In order to address the requirements resulting from the key elements, multidisciplinary knowledge is required [46]. This includes expert, plant, machine, and product knowledge [43]. Additionally, the ICT infrastructure, as well as their establishment and organization, play important roles [42,43].

**Integration into the Supply Chain:** One important goal of the digital twin approach is the support of the various stakeholders of and activities in the supply chain. The challenge hereby is many the variety of the different used systems and how those can be integrated for the digital approach. The standardization of ERP systems such as SAP can support this. However, the presence of such systems in the food industry is smaller than in comparable industries. One reason is the specific needs of the industry. Hence, specific types of such ERP systems emerged recently. Addressing this challenge of a holistic view on the supply chain helps to apply the digital twin concept to simulate and evaluate different steps in the supply chain and to optimize them. This can also support the resilience of the whole supply chain.

## 6. Conclusions

In this paper, we discussed the idea of using biophysical digital twins—composed of data from the process (collected by sensors) and raw material of the products combined with scientific models from food science—to capture and simulate the state of a food

product and process during food processing. Such a hybrid digital twin would have several benefits, especially it can be the base for reasoning on process adjustment and adaptations, but also facilitates the product development. This paper described the idea of integrating XAI procedures to improve the construction of the digital twins and integrating expert knowledge—transferring the black-box of machine learning to a white-box through the combination with scientific model/simulations. Further, the paper describes how to apply the digital twin concept in several case studies: (i) SeAC systems can support adaptive food processing, (ii) product development and food reformulation, (iii) support the scale-up and shelf-life prediction, and (iv) traceability in the food supply chain.

In our research group, we made the first steps towards our vision. Obviously, there are several challenges we still have to tackle. This includes a general applicable model for describing the properties of the digital twin which can be applied to different categories of food products. Further, we currently build the digital twins manually. We are working on solutions that automate the construction of digital twins as well as the analysis of the modeled food similar to solutions from the area of machine learning, e.g., AutoML [44] or based on our previous works [8,9]. Additionally, we already have several previous works for systems that can adapt the process and support adaptability [3,8,9,47]. We are currently working on integrating and adjusting them for food processing.

**Author Contributions:** Conceptualization, C.K. and T.N.; methodology, C.K. and T.N.; formal analysis, C.K. and T.N.; writing—original draft preparation, T.N.; writing—review and editing, C.K., C.B. and T.N.; visualization, T.N. and C.K.; supervision, C.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

ICT	information and communication technology
IoT	internet of things
LRA-M	learn-reason-action-model
SeAC	self-aware computing
XAI	explainable artificial intelligence

### References

1. Stark, R.; Damerou, T. Digital Twin. In *CIRP Encyclopedia of Production Engineering*; Chatti, S., Tolio, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1–8. [CrossRef]
2. Rohleder, B.; Minhoff, C. Die Ernährung 4.0-Status Quo, Chancen und Herausforderungen. 2019. Available online: <https://www.bitkom.org/sites/default/files/2019-03/Bitkom-Charts%20190326%20Digitalisierung%20der%20Ern%C3%A4hrungsindustrie.pdf> (accessed on 25 July 2022).
3. Krupitzer, C.; Roth, F.M.; VanSyckel, S.; Schiele, G.; Becker, C. A Survey on Engineering Approaches for Self-adaptive Systems. *Pervasive Mob. Comput.* **2015**, *17*, 184–206. [CrossRef]
4. 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* **2022**, *22*, 115. [CrossRef]
5. Cadavid, J.P.U.; Lamouri, S.; Grabot, B.; Pellerin, R.; Fortin, A. Machine learning applied in production planning and control: A state-of-the-art in the era of industry 4.0. *J. Intell. Manuf.* **2020**, 1–28. [CrossRef]
6. Cioffi, R.; Travaglion, M.; Piscitelli, G.; Petrillo, A.; De Felice, F. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability* **2020**, *12*, 492. [CrossRef]
7. Krupitzer, C.; Wagenhals, T.; Züfle, M.; Lesch, V.; Schäfer, D.; Mozaffarin, A.; Edinger, J.; Becker, C.; Kounev, S. A survey on predictive maintenance for industry 4.0. *arXiv* **2020**, arXiv:2002.08224.
8. Züfle, M.; Moog, F.; Lesch, V.; Krupitzer, C.; Kounev, S. A machine learning-based workflow for automatic detection of anomalies in machine tools. *ISA Trans.* **2022**, *125*, 445–458. [CrossRef]

9. Zuefle, M.; Bauer, A.; Lesch, V.; Krupitzer, C.; Herbst, N.; Kounev, S.; Curtef, V. Autonomic Forecasting Method Selection: Examination and Ways Ahead. In Proceedings of the 2019 IEEE International Conference on Autonomic Computing (ICAC), Umea, Sweden, 16–20 June 2019; pp. 167–176. [CrossRef]
10. Myhan, R.; Białobrzewski, I.; Markowski, M. An approach to modeling the rheological properties of food materials. *J. Food Eng.* **2012**, *111*, 351–359. [CrossRef]
11. Capuano, E.; Oliviero, T.; van Boekel, M.A. Modeling food matrix effects on chemical reactivity: Challenges and perspectives. *Crit. Rev. Food Sci. Nutr.* **2018**, *58*, 2814–2828. [CrossRef]
12. Van Boekel, M.A. Kinetic Modeling of Food Quality: A Critical Review. *Compr. Rev. Food Sci. Food Saf.* **2008**, *7*, 144–158. [CrossRef]
13. Hartmann, C. Numerical simulation of thermodynamic and fluid-dynamic processes during the high-pressure treatment of fluid food systems. *Innov. Food Sci. Emerg. Technol.* **2002**, *3*, 11–18. [CrossRef]
14. Abdul Ghani, A.; Farid, M.; Chen, X.; Richards, P. Numerical simulation of natural convection heating of canned food by computational fluid dynamics. *J. Food Eng.* **1999**, *41*, 55–64. [CrossRef]
15. Borsum, C.; del, K.; Stefan, T.; Schmidt, C.; Hinrichs, J. Design of in vitro model to study oral aroma release: Experimental study and numeric simulation of heat transfer in a foamed dairy matrix. *J. Food Eng.* **2020**, *278*, 109940. [CrossRef]
16. Verdouw, C.; Tekinerdogan, B.; Beulens, A.; Wolfert, S. Digital twins in smart farming. *Agric. Syst.* **2021**, *189*, 103046. [CrossRef]
17. Rauh, C.; Singh, J.; Nagel, M.; Delgado, A. Objective analysis and prediction of texture perception of yoghurt by hybrid neuro-numerical methods. *Int. Dairy J.* **2012**, *26*, 2–14. [CrossRef]
18. Technische Hochschule Ostwestfalen-Lippe. Smartfoodtechnology OWL—Projects. Available online: <https://www.sft-owl.de/en/projects/> (accessed on 25 July 2022).
19. Defraeye, T.; Shrivastava, C.; Berry, T.; Verboven, P.; Onwude, D.; Schudel, S.; Bühlmann, A.; Cronje, P.; Rossi, R.M. Digital twins are coming: Will we need them in supply chains of fresh horticultural produce? *Trends Food Sci. Technol.* **2021**, *119*, 245–258. [CrossRef]
20. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Informatics* **2019**, *15*, 2405–2415. [CrossRef]
21. Rosen, R.; Von Wichert, G.; Lo, G.; Bettenhausen, K.D. About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine* **2015**, *48*, 567–572. [CrossRef]
22. Verboven, P.; Defraeye, T.; Datta, A.K.; Nicolai, B. Digital twins of food process operations: The next step for food process models? *Curr. Opin. Food Sci.* **2020**, *35*, 79–87. [CrossRef]
23. Müller, P.; Schmid, M. Intelligent Packaging in the Food Sector: A Brief Overview. *Foods* **2019**, *8*, 16. [CrossRef]
24. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
25. Gutschi, C.; Furian, N.; Suschnigg, J.; Neubacher, D.; Voessner, S. Log-based predictive maintenance in discrete parts manufacturing. *Procedia CIRP* **2019**, *79*, 528–533. [CrossRef]
26. Lopez, M.; Beurton-Aimar, M.; Diallo, G.; Maabout, S. A simple yet effective approach for log based critical errors prediction. *Comput. Ind.* **2022**, *137*, 103605. [CrossRef]
27. Maaß, W.; Pier, M.; Moser, B. Smart Services in der Landwirtschaft. In *Service Engineering: Von Dienstleistungen zu digitalen Service-Systemen*; Meyer, K., Klingner, S., Zinke, C., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2018; pp. 167–181. [CrossRef]
28. Defraeye, T.; Wu, W.; Prawiranto, K.; Fortunato, G.; Kemp, S.; Hartmann, S.; Cronje, P.; Verboven, P.; Nicolai, B. Artificial fruit for monitoring the thermal history of horticultural produce in the cold chain. *J. Food Eng.* **2017**, *215*, 51–60. [CrossRef]
29. Kounev, S.; Kephart, J.; Milenkoski, A.; Zhu, X. *Self-Aware Computing Systems*; Springer International Publishing: Cham, Switzerland, 2017.
30. Guichard, E. Interactions between flavor compounds and food ingredients and their influence on flavor perception. *Food Rev. Int.* **2002**, *18*, 49–70. [CrossRef]
31. Chen, J. Food oral processing—A review. *Food Hydrocoll.* **2009**, *23*, 1–25. [CrossRef]
32. Fraunhofer IVV. Simulation of Shelf Life—Shelf Life Prediction and the Sizing of Packaging. Available online: <https://www.ivv.fraunhofer.de/en/packaging/modeling-shelf-life.html> (accessed on 24 July 2022).
33. 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*, 2889. [CrossRef]
34. Tian, F. A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things. In Proceedings of the ICSSM 2017—International Conference on Service Systems and Service Management, Dalian, China, 16–18 June 2017; pp. 1–6. [CrossRef]
35. Mondal, S.; Wijewardena, K.P.; Karuppuswami, S.; Kriti, N.; Kumar, D.; Chahal, P. Blockchain Inspired RFID-Based Information Architecture for Food Supply Chain. *IEEE IoTJ* **2019**, *6*, 5803–5813. [CrossRef]
36. Martínez, G.S.; Sierla, S.; Karhela, T.; Vyatkin, V. Automatic Generation of a Simulation-Based Digital Twin of an Industrial Process Plant. In Proceedings of the IECON 2018—44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 21–23 October 2018; pp. 3084–3089. [CrossRef]
37. Koulouris, A.; Misailidis, N.; Petrides, D. Applications of process and digital twin models for production simulation and scheduling in the manufacturing of food ingredients and products. *Food Bioprod. Process.* **2021**, *126*, 317–333. [CrossRef]

38. Dittmann, S.; Zhang, P.; Glodde, A.; Dietrich, F. Towards a scalable implementation of digital twins—A generic method to acquire shopfloor data. *Procedia CIRP* **2021**, *96*, 157–162. [[CrossRef](#)]
39. Werner, R.; Beugholt, A.; Takacs, R.; Geier, D.; Becker, T.; Sollacher, R.; Mauermann, M.; Weißenberg, N.; Roest, M.; Istaitih, J. Standardized digitalization of an existing pudding production: Integration of a digital twin management system. *Int. Dairy Mag.* **2020**, *5*, 22–25.
40. Barni, A.; Fontana, A.; Menato, S.; Sorlini, M.; Canetta, L. Exploiting the Digital Twin in the Assessment and Optimization of Sustainability Performances. In Proceedings of the 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25–27 September 2018; pp. 706–713. [[CrossRef](#)]
41. Došilović, F.K.; Brčić, M.; Hlupić, N. Explainable artificial intelligence: A survey. In Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 21–25 May 2018; pp. 210–215. [[CrossRef](#)]
42. Neto, A.A.; Deschamps, F.; Da Silva, E.R.; Lima, E.P. Digital twins in manufacturing: An assessment of drivers, enablers and barriers to implementation. *Procedia CIRP* **2020**, *93*, 210–215. [[CrossRef](#)]
43. Werner, R.; Takacs, R.; Geier, D.; Becker, T.; Weißenberg, N.; Haße, H.; Sollacher, R.; Thalhoffer, M.; Schumm, B.; Steinke, I. The Challenge of Implementing Digital Twins in Operating Value Chains. In *Digital Twins: Applications to the Design and Optimization of Bioprocesses*; Herwig, C., Pörtner, R., Möller, J., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 127–166. [[CrossRef](#)]
44. Hutter, F.; Kotthoff, L.; Vanschoren, J. *Automated Machine Learning: Methods, Systems, Challenges*; Springer: Berlin/Heidelberg, Germany, 2019.
45. Qiao, Q.; Wang, J.; Ye, L.; Gao, R.X. Digital Twin for Machining Tool Condition Prediction. *Procedia CIRP* **2019**, *81*, 1388–1393. [[CrossRef](#)]
46. Qi, Q.; Tao, F.; Hu, T.; Anwer, N.; Liu, A.; Wei, Y.; Wang, L.; Nee, A. Enabling technologies and tools for digital twin. *J. Manuf. Syst.* **2021**, *58*, 3–21. [[CrossRef](#)]
47. Krupitzer, C.; Szttyler, T.; Edinger, J.; Breitbach, M.; Stuckenschmidt, H.; Becker, C. Beyond position-awareness—Extending a self-adaptive fall detection system. *Pervasive Mob. Comput.* **2019**, *58*, 101026. [[CrossRef](#)]