Food Engineering Reviews (2025) 17:505–531
https://doi.org/10.1007/s12393-025-09420-x

# Machine Learning in Cheese-Making: Methods, Applications, and the Future

Daniel Pardo[1,2] · Manuel Castillo[2] · Mehmet Oguz Mulayim[1,2] · Jesus Cerquides[1]

Received: 19 February 2025 / Accepted: 30 June 2025 / Published online: 11 July 2025


**Abstract**
Cheese-making is a complex process involving numerous stages, with multiple factors contributing and complex interactions occurring among the physicochemical elements involved. Understanding the process and optimizing its stages has attracted the attention of numerous investigations. In recent years, Machine Learning (ML) has established itself as one of the most advanced tools for data analysis and modeling thanks to its ability to capture complex and non-linear patterns. In the area of food science and engineering, these algorithms have started to be used as an alternative to more traditional statistical and mathematical prediction models. This paper explores the main research on ML applied to the study of cheese, from its production stages (i.e., fermentation or coagulation process) to the final product (i.e., detection of adulterations or food fraud). Particularly, we review 42 papers published between January 2014 and January 2025, with the aim of identifying common approaches. First, we present an explanation of the main concepts required to bring these approaches closer to researchers who are not experienced in applying ML. Then, we analyze the selected publications to detail the tasks of interest and the algorithms proposed to solve them. Finally, we detect gaps and opportunities to incorporate ML into future cheese research.

**Keywords** Artificial intelligence · Cheese · Cheese industry · Cheese production · Dairy · Machine learning

## Introduction

Milk is a food present in the regular diet of many people due to its high nutritional value. For example, it is an important source of protein, as it provides around 32 g of protein per liter. In addition, its nutritional composition is a protective factor for the prevention of various diseases. In general, dairy products are considered balanced and nutritious foods [97].

For the next decade, milk production (roughly 81% cow, 15% buffalo, and 4% for goat, sheep, and camels combined) is expected to grow by 1.6%, reaching 1.085 Mt in 2033, faster than most other main agricultural commodities [95].

Milk can be transformed into derived products for consumption. Globally, most dairy production is consumed in the form of fresh, unprocessed dairy, with only around 30% of milk being processed to create derived products. However, these global figures are skewed by the high volume of consumption in Asian countries, where fresh dairy predominates. In contrast, the opposite trend is observed in the European Union and the United States, where approximately 60% of dairy consumption consists of processed products. Among them, cheese is the most important due to its direct consumption and as an ingredient in processed foods, mainly in Europe and North America [95].

Therefore, forecasts for increased consumption and production of milk and cheese represent a challenge for producers. Raw milk is highly variable, and although processes can be applied to standardize it, the quality of the cheese is largely dependent on milk with good chemical functionality and microbiological properties [42, 43].

Since the properties of cheese depend in part on the physicochemical characteristics of the milk used, it is common to standardize its composition to the desired levels depending on the cheese variety. Centrifugal separation is commonly used to correct fat levels, and, recently, ultrafiltration has been used to adjust protein levels [42, 73].

However, ensuring the quality of cheese requires control throughout the production process. Modern technology has introduced new ways to more accurately monitor, automate,

✉ Daniel Pardo
daniel.pardo@iiia.csic.es

[1] Artificial Intelligence Research Institute, IIIA-CSIC, Cerdanyola del Vallès 08193, Barcelona, Spain

[2] Universitat Autònoma de Barcelona, UAB, Cerdanyola del Vallès 08193, Barcelona, Spain

and control cheese production stages. For example, it is common to implement Process Analytical Technology (PAT) [41] to continuously monitor the most relevant parameters, which increases productivity. PAT is a framework for developing systems that enable real-time analysis and control of critical processing variables to ensure consistent quality and efficiency. In addition, in recent years, special emphasis has been placed on the ripening phase with new proposals [125].

However, in the artisanal industry, traditional control is maintained during cheese production based on subjective manual and visual techniques that require experienced operators, which may produce significant differences in performance depending on the operator. Therefore, two significant challenges arise, namely: (i) the difficulty of standardizing the process and the final product due to the high variability of the raw material and (ii) the subjectivity of the process [62].

In the context of Industry 4.0, the cheese production industry can significantly improve and optimize production processes. The conversion of real-world data acquired by physical sensors into digital data enables the creation of smart models or devices through data analytics and Machine Learning (ML). The potential of these advances is an improvement in manufacturing productivity and efficiency or waste reduction. Therefore, digitalization emerges as a relevant opportunity for the cheese industry [49].

The idea of ML is to use general models or algorithms that can solve specific tasks by optimizing the values of the model parameters for each problem. The parameter optimization is known as the learning process, and training data is used for this. The goal is to create ML models with the capacity and ability to perform inference and learn from samples, which allows them to make predictions or descriptions [5]. A *sample* is a single data point or observation used to train or evaluate a model.

The versatility of ML algorithms allows them to learn and solve problems in any field as long as they can be modeled appropriately. For this reason, ML algorithms are currently used in practically all industries, including food science and the dairy industry. This review aims to present, from a didactic perspective, the most relevant works in the cheese industry in which ML is used and to introduce the general approaches. In this way, researchers in food science in general, but especially those whose work focuses on cheese, will be able to learn about the different tools and possibilities that may be of interest for further exploration or integration into their work.

The research is organized in a pragmatic way according to the stage of cheese production to find out in which processes or problems ML has been applied. In addition, the general results and the impact of applying these methods are presented, so that researchers can learn about the potential benefits that they can bring. The main contributions of the paper are:

- **Introduction to ML applied to the cheese industry and food science**, providing a clear conceptual basis for researchers without prior experience in the field, enabling its effective application in research
- **Classification of ML studies according to the main stages of the cheese production process**, allowing for a comprehensive and practical perspective of the state of the art, unlike previous reviews that focus on isolated aspects or broader food science applications.
- **Identification of gaps in the literature and methodological variability**, highlighting the lack of applications validated in industrial settings
- **Proposal for future lines of research in ML for cheesemaking**, highlighting key areas where its implementation can drive significant improvements

The structure of the article is divided as follows. First, in Section "Fundamental Machine Learning Concepts", we explain the key concepts in ML so that someone without prior knowledge can approach and understand the basic ideas. Next, in Section "Literature Review Methodology", we detail how the systematic search for the articles analyzed in this review was conducted. Then, in Section "Literature Landscape", we examine the selected articles, presenting their main application areas and the algorithms used while also analyzing challenges and future perspectives. Finally, in Section "Discussion", we highlight the main patterns and findings, and we provide a summary of the article's general ideas in Section "Conclusions".

## Fundamental Machine Learning Concepts

Although the concepts of Artificial Intelligence (AI) and ML are closely linked and sometimes used interchangeably, the reality is that they are not the same. AI is the discipline that studies the creation of models to solve problems that require human intelligence, while ML is a subdiscipline, that is, a way of creating AI models [52].

Some problems can be described with formal or mathematical rules, while others are difficult to describe. Although humans can intuitively solve tasks in their daily lives, such as seeing, hearing, or speaking, these problems are the most difficult for computers because their formal definition is very complex. ML algorithms aim to solve these problems by allowing models to acquire their own knowledge by extracting patterns from raw data. Currently, ML is considered to be the only approach to building AI models that work in complex real-world environments [52].

In so-called shallow ML models, such as decision trees or Support Vector Machines (SVM), it is first necessary to extract features from the data and then pass them to the model. However, for complex problems, it is generally difficult to know which features are the most representative of the dataset. An alternative, known as representation learning, is to let the model learn by itself to transform the raw data into representative features. Some well-known examples of this type of approach are Principal Component Analysis (PCA) or autoencoders. Finally, Deep Learning (DL), another key concept, is a particular type of representation learning based essentially on Artificial Neural Network (ANN) with many layers that can learn complex patterns from large datasets. Figure 1 shows the relationship between the different concepts [52].

The term ML was first used in 1959 by Arthur Samuel. Although he did not give a formal definition, he did describe some key concepts. For example, in his work, he anticipated that "a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program [...] in a remarkably short period of time". He also added that "the principles of ML verified by these experiments are applicable to many other situations" [102].

One of the first and most popular formal definitions was the one proposed by Tom Mitchell in 1997: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." Therefore, the key concepts are the task (T), the performance measure (P), and the data (or experience, E) [90]. In general terms, it can be said that "ML is programming computers to optimize a performance criterion using example data or past experience" [5].

In recent years, investment and research in AI has grown at an accelerated level. In terms of publications, it has gone from 88,000 in 2010 to almost triple in 2022, reaching 240,000. Meanwhile, the number of patents registered in 2022 is more than 30 times higher than in 2010. In this sense, the growth in the number of AI patents is accelerated, with 5,136 in 2017, 8,144 in 2018, 15,783 in 2019, 23,373 in 2020, 38,268 in 2021, and 62,264 in 2022 [85].

In the same report, the authors group private investments in 2023 into 25 categories, out of which *AI infrastructure/research/governance* stands out with the highest private investment, which increased 22 times compared to 2022. In this regard, the *Manufacturing* category ranks 14th in terms of highest investment and is one of the few that saw an increase in terms of investment compared to the previous year. In contrast, the *Drones* and *Medical and healthcare* categories stand out negatively as they have significantly reduced investment to less than half compared to 2022.

To expand general knowledge and concepts of ML, an excellent option is the introductory book by Alpaydin [5]. It presents the paradigms, the main tasks, and the most important ML algorithms. It introduces the concepts of ANN and Multi Layer Perceptron (MLP), a type of ANN with at least one hidden layer and with all layers fully connected, as a universal approximator. For a brief basic introduction to ANN, see Krogh [69]. For a deeper focus on DL, the article by LeCun et al. [71] is a key reference in the field, concisely presenting the most relevant aspects and accessible to a non-specialized audience. In addition, for more exhaustive and detailed knowledge, the work of Goodfellow et al. [52] is an essential reference guide. Furthermore, Bishop and Bishop [16] presents an updated perspective, incorporating the latest advances in DL.

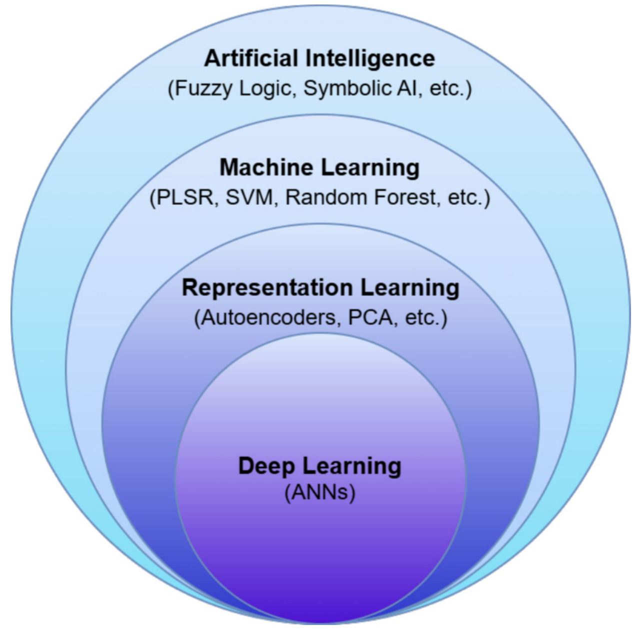


**Fig. 1** Artificial Intelligence categories

## Machine Learning Life Cycle

The general process of developing a ML model goes from the conceptual phase and data acquisition to evaluation or deployment. To create this section, the workflow described by Géron [54] is used as a reference with slight adjustments to differentiate some relevant steps clearly. Figure 2 shows a diagram of the main steps detailed below.

### Problem Statement

The first task is to precisely determine the problem and the objectives to be addressed. This will allow the definition of the problem domain, the current context, and the most appropriate performance measure [54].

### Data Collection

Data collection takes up a large portion of the time in the ML process and constitutes one of the bottlenecks. For models to learn to generalize, they need to be trained with quality data, i.e., data representative of the application domain, making

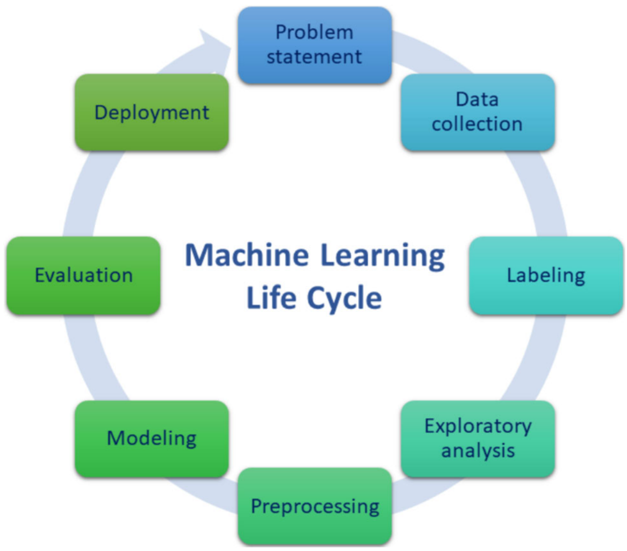


**Fig. 2** Machine learning life cycle

collection and preprocessing critical stages. In addition, the most advanced DL models require a very large amount of data for training. There are three main approaches to obtaining a dataset: data generation, data discovery, or data augmentation [101]:

- **Data Discovery:** It consists of searching for and finding existing datasets that are suitable for our problem and can be used to train a model. There are collaborative systems or specialized platforms, such as Kaggle[1] or Hugging Face Datasets[2], that store, share, and facilitate access to curated datasets for a wide range of ML tasks. Many of the main high-quality datasets widely used to train ML models are available in public repositories and are freely accessible.
- **Data generation:** When there is no existing dataset available or suitable for the task to be solved, data must be collected. Real data can be captured through various analysis techniques and experimental measurements, different types of sensors, or computer records. However, when real data is unavailable, an alternative is synthetic data generation, which consists of creating artificial but realistic data that mimics real-world patterns. Some of the most advanced techniques use ML algorithms such as Generative Adversarial Networks (GANs) or diffusion models to generate data.
- **Data augmentation:** This approach relies on an existing dataset to generate new samples. This can be achieved by integrating data from different sources to create a new homogeneous dataset or by applying small modifications to the samples. For example, in image tasks, simple transformations such as rotation, translation, flipping, or adding noise to images are commonly used to augment the dataset. These modifications to the data or specific classes produce similar new samples based on the originals, preserving their characteristics.

[1] https://www.kaggle.com, accessed May 2025

[2] https://huggingface.co/datasets, accessed May 2025

### Labeling

*Labels* are the target values or annotations assigned to each data sample, representing the objective that the algorithm learns to predict during training. Labeling the samples is a critical aspect that determines the ML algorithms and approaches that can be used, as not all problems have labeled data. The labeling process depends on the type of problem to be solved, such as classification, where each sample belongs to a category (e.g., cheese varieties such as Cheddar or Brie), or regression, where the goal is to predict a continuous attribute (e.g., pH value). This process can be done in parallel during data acquisition or afterward [101]. An *attribute* is a variable or a property that describes a characteristic of a sample. This includes also the target variable and the features. A *feature* is an input variable of a sample and is used by the model to make predictions. In DL, this term can also refer to the abstract internal representations that ANNs learn automatically.

### Exploratory Analysis

Data can be in various formats, such as tabular form, images, audio, video, etc. An initial approach involves performing an exploratory analysis to obtain a general description and overview of the data, such as the number of samples and attributes, the type of each attribute, or the number of missing values. In this phase, potential correlations between variables must be examined, and basic descriptive statistics such as the mean, standard deviation, percentiles, or minimum and maximum values in each case must be known. Data visualization is another essential exploratory technique. This includes everything from visualizing different representative samples of the dataset to showing the distribution of the data or the relationships between the variables using appropriate graphics. This preliminary information helps to understand some patterns in the data, and it is crucial for subsequent analyses, as it determines whether it is necessary to apply standardization techniques to avoid bias, to treat missing values, or to detect if the dataset is unbalanced, among other things. In this regard, an unbalanced dataset is one where some classes are underrepresented, potentially leading to biased predictions or poor generalization [54].

### Preprocessing

In this stage, the data must be prepared for the ML algorithms. It should first be divided into training, validation, and test subsets: the training set is used to fit the model, the validation set helps tune model hyperparameters, and the test set evaluates the final model's performance on unseen data. The data preprocessing operations must be performed separately and independently for each subset. Next, standard data cleaning tasks such as handling missing values by imputing (filling in missing data) or removing them, treating outliers, converting variable types, and scaling features must be performed. It is also possible to create (from existing variables) or delete variables, reduce the dimensionality of the data, or convert labels to a one-hot encoding representation (which represents each categorical variable as a binary vector) [54, 101].

### Modeling

The next step is to train the ML algorithm using the training set and tune the hyperparameters of the selected ML model using the validation set, taking into account the validation metric defined according to the objectives. *Hyperparameters* are adjustable settings that influence how the model learns, such as the regularization term, the learning rate, or the number of hidden layers in a ANN. Each algorithm has its own specific hyperparameters that must be manually predefined, while other parameters, such as weights and biases, are automatically learned from the data during training.

Additionally, the risk of overfitting and underfitting should always be considered during the training process. Overfitting occurs when a model fits the training data too precisely, including noise, leading to memorization rather than generalization. This results in high performance on the validation set but poor performance on unseen test data. In contrast, underfitting happens when a model is too simple to capture the underlying patterns in the data, resulting in poor performance overall. Depending on the amount of data, the cross-validation method can be used. Cross-validation is a statistical technique for robust evaluation when there is no independent test set or the dataset is too small for a standard train-test split. The most widely used method is *k-fold cross-validation*, which partitions the data into $k$ subsets and performs $k$ independent training iterations of the model. In each iteration one different subset is used for validation and the remaining *k-1* subsets are used for training. For further technical details, it is recommended to consult the work of James et al. [60].

Once the model hyperparameters are set, they cannot be adjusted using the test data. Some ML algorithms require prior feature extraction, as they use this information as input, while DL approaches can handle this internally and process raw data directly. Testing different algorithms and hyperparameter configurations is essential. Transfer learning techniques can also be used, which consist of using a model previously trained on a large dataset and fine-tuning it for a specific problem. This approach leverages learned features in the larger dataset, reducing training time and improving performance, especially when only a limited amount of data is available. After identifying promising models, further fine-tuning should be performed on each of them to optimize their hyperparameters. This hyperparameter tuning can be done exhaustively (evaluating all possible combinations of the defined hyperparameters), randomly (selecting random configurations), or using optimization approaches (such as those based on Bayesian probability) [54, 61].

### Evaluation

To verify that the model's performance is good and generalizes appropriately, the test set is used to perform a final evaluation. In this section, key performance metrics are expected to be represented, e.g., the confusion matrix based on metrics such as Accuracy or F1-Score, and the Receiver Operating Characteristic (ROC) curve, which are commonly used in classification tasks. On the other hand, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are typically used in regression. This also involves conducting a critical analysis of the model's results to identify its strengths and areas for improvement, so the analysis should be both quantitative and, where appropriate (such as in text or image analysis), qualitative. [54].

In addition to traditional evaluation metrics, *explainability* focuses on assessing the comprehensibility of ML models, incorporating both quantitative and qualitative elements. Explainability refers to a set of techniques and algorithms designed to make the predictions of ML models understandable to humans, especially black-box models such as ANNs. The motivation for explainability is multifaceted: it includes understanding decision-making mechanisms, identifying which features influence predictions, diagnosing model errors, detecting potential biases, and enabling model improvements through domain knowledge. The survey by Burkart and Huber [22] offers an extensive and comprehensive review that serves as a key reference for exploring this field in depth.

### Deployment

Finally, the model can be deployed in production and should be monitored to check its performance and review any possible drops when it faces more complex real-world problems. In this sense, the process is incremental and always subject to changes and improvements at the different stages described [54].

These stages are similar to those proposed in the Cross-Industry Standard Process for Data Mining (CRISP-DM) model, which describes the standard methodology process for data mining and data analysis projects. The approach taken in the CRISP-DM model is business-oriented, with special emphasis on business understanding, but with all the same critical stages described above [114].

## Machine Learning Paradigms and Tasks

ML algorithms can be categorized into four paradigms based on their scope and applicability: supervised, unsupervised, semi-supervised, and reinforcement learning [103]. Each paradigm is associated with specific ML tasks. Figure 3 shows a summary of the paradigms, their corresponding tasks, and representative algorithms.

### Supervised Learning

Conceptually, it consists of learning a function that maps an input to an output and is the most widely used approach to solve ML problems. In this learning category, we need the samples to be an input-output pair, where the output is the target variable of the learning process. Therefore, one of the fundamental characteristics is that the samples are labeled with the target variable, and the objective of the learning is to infer the function from the sample data. Among the most common tasks of supervised learning, we find classification problems in which the objective is to predict a category or discrete variable, regression that fits the data to predict a continuous value, anomaly detection, or dimensionality reduction to project the data in a lower dimension. There are different variants for each type of task [71, 103].

Classification problems can be based on the approach (flat, hierarchical) or the labels (binary, multilabel, multiclass). As for the approach, a flat classification would be the distinction between independent categories such as cheese varieties (Camembert, Cheddar) while hierarchical classification implies breaking down the problem and structuring the categories in hierarchical levels, for example, first classifying between hard and soft cheese and once this is done detecting the cheese variety. Some algorithms originally designed for classification or regression (such as SVM) can be tuned to anomaly detection (One-Class), while others are specific to this task, such as Isolation Forest (IF), which is inherently an unsupervised algorithm. As for the labels, they are binary when the dataset only has two classes, multiclass when there are more than two classes, and multilabel when a sample can belong to more than one class at a time. In this last case, if for example, we want to classify cheeses according to the type of milk, cheeses such as Parmigiano Reggiano made only from cow's milk will have only one label while cheeses such as Cabrales that use a mixture of milks will have more than one label.

In regression, models can be based on the type of relationship (linear, non-linear), the number of independent variables (simple, multiple), or the number of target variables (univariate, multivariate). Regarding the relationship, linear

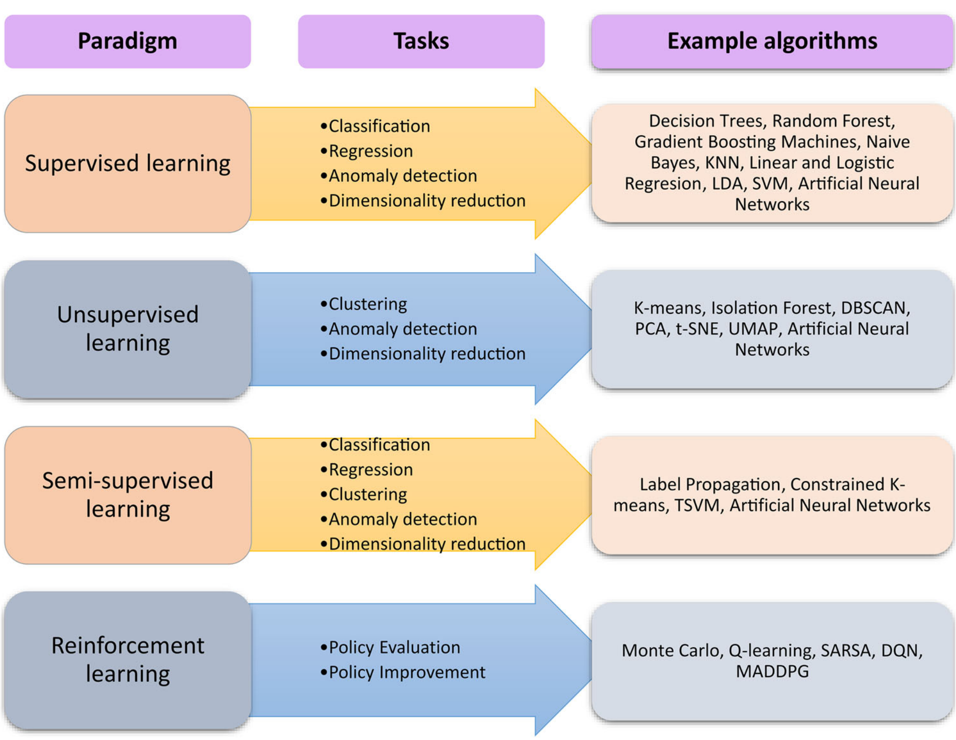


**Fig. 3** Machine learning paradigms, tasks, and algorithms

regression occurs when the input variable and the target variable have a direct relationship. This type of relationship is found in simple problems, such as the increase in the price of a cheese of the same variety based on its weight. Non-linear regression models more complex relationships that do not follow a straight line, as is the case in most cheese production processes (physicochemical changes, texture development, ripening time). If only one independent variable is used, it is a simple regression model, and multiple regression if there is more than one. Regarding the number of target variables, if we want to predict only one characteristic such as fat content it would be a univariate regression. However, if we want to predict more than one characteristic at the same time such as fat and protein content we would refer to this as multivariate regression.

In addition, the nature of the problem and the data must be taken into account since there may be independent, or with sequential or spatial dependence. For example, measurements of a cheese, such as its fat, protein, and moisture content, are independent attributes (although there may be a correlation or dependency between some of them). On the other hand, there is temporal dependence when measurements are taken over a period of time. For example, by monitoring changes in the cheese's moisture content we can obtain a set of measurements that are part of the same sample in the form of a time series. Similarly, images have a spatial relationship since they are evaluated at the pixel level, and it is important to consider the relationship of each pixel with its neighbors to capture complex patterns [71, 103].

#### Unsupervised Learning

These are algorithms used when there are no labels available in the samples, and therefore, we do not know what the correct expected output is. Learning is a data-driven process, so interesting patterns and structures are learned automatically from the data itself. They are used for clustering tasks where we want to group samples based on some similarity criteria or dimensionality reduction in cases where there are no labels available. A classic clustering algorithm is k-means [9], which allows samples to be grouped based on their similarity without knowing the label. For example, we could group cheeses with similar characteristics without really knowing the variety of each one. As for dimensionality reduction, the PCA algorithm is widely used in food science and is a good tool for visualization and preliminary analysis of data [1]. Unsupervised learning is also used in some neural network architectures, such as autoencoders, to extract features without labels in some layers, being especially useful in small datasets to avoid overfitting [71, 82, 103].

#### Semi-supervised Learning

In some real-world contexts, it is difficult to have data completely labeled. To solve this problem, semi-supervised algorithms combine ideas from both supervised and unsupervised methods, being trained on a dataset that includes both labeled and unlabeled samples. However, it is always desirable to label all data if possible and use supervised algorithms. They are mainly used to solve the same tasks as in supervised learning, and some areas of application stand out, such as labeling data or automated text translation (machine translation) [103].

#### Reinforcement Learning

This paradigm has a different approach from the previous ones and is based on the idea of an intelligent agent, an entity that interacts with a predefined environment by taking actions and receiving feedback. Through these actions, the agent (or agents) can automatically learn the optimal behavior within a given context, making it an environment-driven approach. The learning of the models is guided by a reward or penalty system based on the agent's decisions, with the main objective of maximizing the accumulated reward. Some areas of application are robotics, autonomous driving, and problems modeled by game theory. However, it is better not to use this type of ML for basic problems. In the context of cheese production, finding a problem that is optimally modeled with this approach is challenging since the production processes do not readily fit into a reinforcement learning scheme [103].

## Literature Review Methodology

To conduct the search for articles, our methodology was inspired by the PRISMA 2020 statement [18]. The central research questions on which this review is based are:

(i) In which cheese production processes/problems has ML been applied?
(ii) What were the main impacts that ML methods have had on cheese production processes?

To address the research questions, for each process/problem it is necessary to ask the following accessory questions:

- How has the ML been applied to the process/problem?
- Which ML methods have been implemented and with what outcomes?
- What are the benefits of implementing ML methods?
- How can ML improve the methods currently used?

Therefore, the objective of this work was to answer these questions by carrying out a literature review focused on cheese production and the field of ML. To this end, we used the databases Scopus, Web of Science, PubMed, ScienceDirect, and IEEE Xplore. Additionally, we carried out a search in Google Scholar to complement the previous results and obtain articles that are not indexed in those databases. The search strategy was restricted to articles published between January 2014 and January 2025 to have the last decade of advances as a reference. This was consistent with the great advances in AI, largely driven by DL, that have occurred in recent years, especially since the publication of the AlexNet architecture in 2012 [68]. In addition, we required the source to be in English.

The search query was defined separately for each database since each one implements different operators and search criteria. Therefore, to perform a consistent search, the query was adjusted so that the result were comparable. In each database, the search words were defined on the "title," "abstract," and "keywords" sections. These sections had to contain at least one reference to "artificial intelligence" or "machine learning" and also the word "cheese." On the other hand, after observing irrelevant results from preliminary searches, those articles containing the words "marketing" or "cancer" were excluded from the search. In the case of Google Scholar, operators could not be restricted exclusively to "title," "abstract," and "keywords." Furthermore, Google Scholar also indexed a larger number of articles. For this reason, a more restricted query was defined since the search covered the entire body of the article. Table 1 shows the specific query for each database and the number of articles retrieved on January 11, 2025. Even though the Google Scholar query was more restricted, it still provided the largest number of articles.

Once the query results were obtained, they were stored, and the information from the sources was standardized since not all databases contained the same information. The fields that were kept, if available, were: 'Authors', 'Title', 'Year', 'Cites', 'Venue', 'Publisher', 'Type', 'Language', 'Abstract', 'Keywords', 'Index Terms', 'URL'. The articles were then filtered to ensure that all publications were from 2014 or later and in English. In the case of Google Scholar, an additional restriction was added, which was that articles older than 2023 had to have been cited at least once to ensure a minimum relevance. The resulting Google Scholar articles were manually reviewed to assess their relevance based on the title. Finally, all information sources were integrated into a single dataset that met the eligibility criteria, allowing us to filter out duplicates.

After duplicate removal, we carried out a selection process to filter the articles based mainly on the reading of the abstract and, in some cases, parts of the article. In this way, the articles eligible for inclusion in the review and those that are out of scope were identified, taking into account their relevance to the research questions. Figure 4 shows a flowchart of the process followed, illustrating the steps described and the number of publications in each case, for the databases and for Google Scholar.

The final search process included a total of 42 papers that met the specified criteria, comprising 35 articles, 2 conference papers, and 5 reviews relevant to the topic. The number of articles obtained was limited, suggesting that ML methods in cheese production research are not widespread yet, perhaps because these are still relatively distant fields of research. This should be interpreted as an opportunity for ML researchers together with food science experts to innovate in a field that is not yet developed.

In Fig. 5, the total number of articles retrieved from each database is shown, along with the number of unique articles that are only available in a single database and not in the others. It should be noted that some of the articles may be included in multiple databases simultaneously. Some sources, like Scopus or Google Scholar, stood out with a higher number of articles than others; however, this did not diminish the importance of the rest. For instance, one arti-

**Table 1** Query search of the databases

| Database | Query | Articles |
|---|---|---|
| Web Of Science | (TS=(artificial intelligence) OR TS=(machine learning)) AND TS=(cheese) NOT TS=(marketing) NOT TS=(cancer) | 72 |
| Scopus | (TITLE-ABS-KEY (("machine learning" OR "artificial intelligence")) AND TITLE-ABS-KEY (("cheese")) AND NOT TITLE-ABS-KEY ("marketing") AND NOT TITLE-ABS-KEY ("cancer")) | 119 |
| Science Direct | ("artificial intelligence" OR "machine learning") AND (cheese) NOT (marketing) NOT (cancer) | 25 |
| PubMed | (("machine learning" [Title/Abstract]) OR ("artificial intelligence" [Title/Abstract])) AND (cheese [Title/Abstract]) NOT (marketing [Title/Abstract]) NOT (cancer [Title/Abstract]) | 25 |
| IEEE Xplore | ((("Abstract":artificial intelligence) OR ("Abstract":machine learning)) AND ("Abstract":cheese)) OR ((("Author Keywords":artificial intelligence) OR ("Author Keywords":machine learning)) AND ("Author Keywords":cheese)) OR ((("Document Title":artificial intelligence) OR ("Document Title":machine learning)) AND ("Document Title":cheese)) | 5 |
| Google Scholar | "cheese production" OR "cheese making" OR "cheese manufacturing" AND "artificial intelligence" OR "machine learning" -"marketing" -"cancer" | 980 |

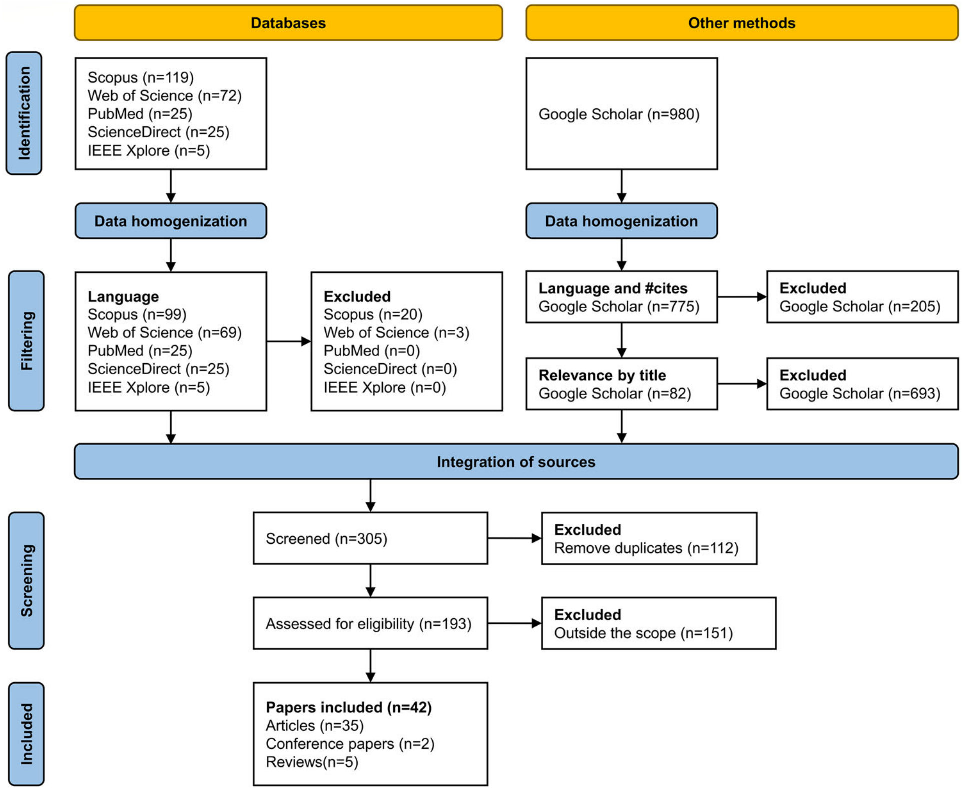


**Fig. 4** Search process diagram

cle from IEEE Xplore and one from Web of Science were only indexed in these databases, highlighting the importance of using multiple sources to be exhaustive in the search and to avoid omitting relevant results. On the other hand, Fig. 6 shows the distribution of the selected articles by year. The most notable aspect is the upward trend of ML articles related to the cheese production process, indicating a growing interest in implementing more advanced methods and integrating ML into research in this field.

The five retrieved reviews can be summarized as follows. The review by Lukinac et al. [79] focused solely on cheese quality and computer vision algorithms. It referenced various studies, explaining their objectives and the problems they aimed to address, but it did not provide specific details of

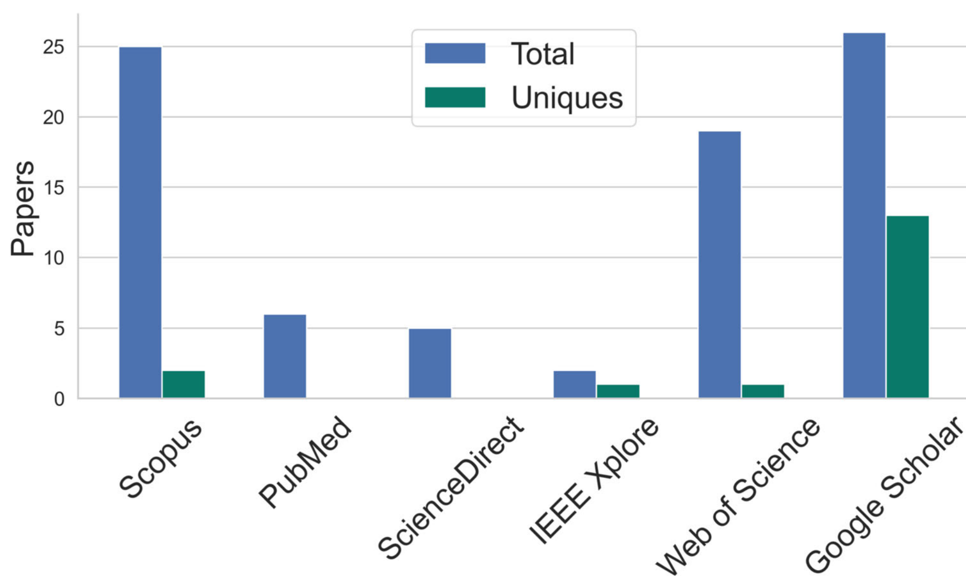


**Fig. 5** Final articles by database

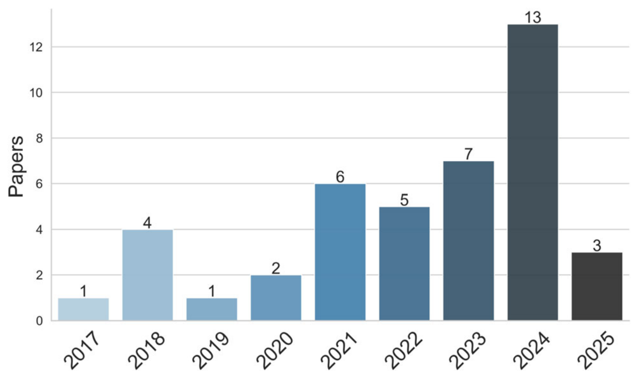


**Fig. 6** Final articles by year

the techniques employed. The main drawback of this review is that it is somewhat outdated, as it neither explains nor cites the most recent advances in computer vision. There is no mention of ML, and the most recent reference to neural networks is from 2007.

Similarly, the review by Bosakova-Ardenska [20] also focused on cheese quality and computer vision. However, some of the cited works were closer to the field of signal processing, while others focused essentially on statistical analysis without employing advanced computer vision or ML techniques. Consequently, these two reviews are limited since (i) they only assessed one relevant aspect of cheese, and (ii) they did not specifically focus on ML algorithms or advanced computer vision techniques.

At the level of technical innovation in the industry, Croguennec et al. [28] reviewed the main innovations in the use of advanced sensors and real-time monitoring in the cheese production industry. The authors considered that analysis techniques and advanced ML models could use the large amount of data generated to optimize production parameters.

The review by Freire et al. [46] aimed to present the main works applying ML to the dairy industry in general. It lists a wide variety of studies with different approaches and for different products without delving into specific details, as this was not the primary objective. Most of the cited research focused on milk studies, and the articles were grouped into general categories for all dairy products. In this regard, there were twelve references to work related to cheese compared to thirty on milk. On the other hand, the review by Singh et al. [106] focused specifically on studies that use ML methods for the milk pasteurization process.

To gain a more comprehensive overview, additional reviews in the field beyond the five retrieved through the proposed search methodology have been considered. However, most reviews, such as the one by Mavani et al. [86], focused on analyzing the application of ML in food science broadly. Along the same lines, the review by Thapa et al. [108] analyzed the use of AI in the food industry. In addition to ML references, they included expert systems or fuzzy logic models used in the production of different products. To the best of our knowledge, there is a general lack in the literature of a review that exclusively examines the main ML studies related specifically to cheese production and the different relevant stages.

## Literature Landscape

After reviewing the fundamentals of ML in Section "Fundamental Machine Learning Concepts" and detailing the methodology for article selection in Section "Literature Review Methodology", this section analyzes key studies in cheese-making that have implemented ML algorithms. Building on the previous concepts, the applied methods and approaches, as well as the main results, are analyzed with the aim of identifying common research trends.

Cheese production, encompassing every stage from the reception of raw materials to the final product, is a complex process that can last from about a day (e.g., fresh cheese such as Burgos cheese) to several years (e.g., hard cheese such as Parmesan). It involves multiple factors that determine the quality of the final product. Each type of cheese develops particular characteristics and differs in appearance, texture, flavor, or nutritional value. Some stages of production are specific to each variety, but some stages are common to the production of any type of cheese [43].

The reviewed articles can be grouped according to the stage of cheese production they focus on. The stages in which the papers are distributed are *acidification* (commonly employed for many cheeses, in some cases by acid addition or more frequently by fermentation), *coagulation* (mandatory for all cheeses), *ripening* (only for mature cheeses, not necessary for fresh cheeses), *final product assessment* (once the production process has finished), and *others* (works not focused on a specific stage of the production process). The Fig. 7 includes the stages into which the articles in this review are divided, but it is not exhaustive of the production process and does not contain other critical stages, such as whey removal, for which we have not found articles. Additionally, this figure shows the number of papers falling into each stage, with the final product assessment stage clearly emerging as the one with the larger number of works, while there are four additional studies categorized as 'Others'.

### Acidification

During the initial milk acidification step (commonly induced by the starter), which can last up to 18 hours depending on the type of cheese, the pH decreases to the level required for coagulation.

The purpose of this acidification depends on the cheese-making process: in acid coagulation, the pH drops until

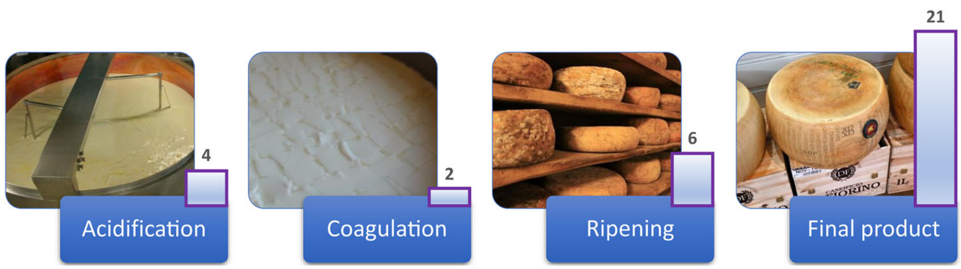


**Fig. 7** Distribution of retrieved articles by stage of the cheese-making process

gelation occurs, whereas in enzymatic coagulation often the pH is often lowered to a level that enhances enzymatic hydrolysis. This acidification can occur through bacterial fermentation, where lactic acid bacteria use lactose to grow and produce lactic acid, which causes a decrease in pH. Alternatively, it can be achieved through direct acid addition, where acids such as citric or lactic acid are added to reduce the pH without bacterial activity. When the desired pH level is reached, which varies for each variety of cheese, the process is stopped [45]. At this stage, research focuses on pH prediction and pathogen detection.

The acidification stage has a critical impact and determines the following steps in the cheese production chain. Bacterial fermentation is a complex process, and predicting the moment when the required pH is reached remains challenging. Under this line of research, we found three similar works on cream cheese aimed at predicting the pH level and thus determining the fermentation process end-time. First of all, Li et al. [74] proposed to use ML algorithms together with mechanistic models to overcome the limitations of traditional modeling and the lack of online data on critical variables. Mechanistic models are mathematical models that rely on fundamental natural laws, such as physics and biochemistry, to describe and explain the behavior of a system through its underlying mechanisms, and often predict changes over time. A mechanistic model with differential equations is used to predict changes from initial measurements in variables such as biomass, lactose, and lactic acid concentration during fermentation. These changes form a time series that served as input for a Long Short-Term Memory (LSTM) model, a type of deep neural network. The study aimed to predict the pH value during fermentation without the need for continuous measurements.

Following this idea, Ebrahimpour et al. [35] compared the model proposed by Li et al. [74] with a purely mechanistic model (to check its nonsuitability) and with a model completely based on ML. Since pH is a variable that can be measured online, Ebrahimpour et al. [35] used a Nonlinear Autoregressive Neural Network (NARNN) to predict future pH values from past pH measurements. NARNN is a type of neural network that works with time series to predict future values from past values [3].

Finally, Guo et al. [55] extended Li et al. [74] by integrating an LSTM with a more complex mechanistic model that incorporates a semi-batch fermentation process, adding the starting inoculation volume as an additional variable to improve the representation of fermentation dynamics.

Another line of research has focused on food poisoning by *Staphylococcus aureus* to detect staphylococcal enterotoxins in the fermentation stage that are potentially dangerous. In this regard, Cai et al. [23] investigated the fermentation stage of Kazak cheese, using lactic acid bacteria, and analyzed an opportunistic pathogen in dairy products, *Staphylococcus aureus*, that produces staphylococcal enterotoxins. For this purpose, strains of the pathogen were prepared and inoculated into pasteurized milk. The aim of this work was twofold. Firstly, the authors wanted to predict the maximum growth rate and lag time in fermentation based on the inoculation amount, water activity, and fermentation temperature. To do this, they first used the modified Gompertz model to fit the growth curves from these variables and other experimental measurements, such as the population of a microorganism at a specific time during fermentation. The Gompertz model is a mathematical model developed in 1825 initially to describe mortality rates. The Zwietering-modified version is an adaptation of the original idea that allows the calculation of lag-time and maximum growth rate curves in microbial growth [113]. These curves were used as the target prediction variable, and the authors created two single hidden layer ANN models to predict each of them using only the initial variables (inoculation amount, water activity, and fermentation temperature) as input. Secondly, they predicted the binary probability (present or absent) that a product is contaminated by staphylococcal enterotoxins from the same three variables (inoculation amount, water activity, and fermentation temperature) using classification models such as Logistic Regression (LR) or MLP (Table 2).

## Challenges and Future Trends

As we have seen, predicting characteristics or values in milk fermentation during cheese production, such as pH level, is challenging due to the technical difficulty of conducting online measurements. Some works [35, 55, 74] created a

**Table 2** Acidification summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| pH value | LSTM | Experimental measurements | Time series | Unknown | [74] |
| pH value | NARNN | Experimental measurements | Time series | Unknown | [35] |
| pH value | LSTM | Experimental measurements | Time series | Unknown | [55] |
| Growth rate and lag time / Pathogen contamination detection | MLP / LR | Experimental measurements | Inoculation amount, Aw, fermentation temperature | 66 | [23] |

mechanistic model and combined it with a ML algorithm to predict the pH value, but these are complex models that depend on the measurement of variables which are not always available. However, the authors identified that LSTM models can be suitable for this task by treating the data as time series. In this regard, researchers were also studying the application of LSTM in the fermentation processes of other food products, to measure pH or other relevant factors, yielding promising results. In addition, we identified other DL approaches, which we will discuss below and that can be applied to cheese production.

The fermentation of milk for the subsequent production of yogurt differs from that of cheese in the type of bacteria used, but the dynamics of the process are similar. In this sense, Bowler et al. [21] have recently used low-cost and easily applicable ultrasonic sensors combined with ML algorithms for pH measurement in yogurt fermentation. These sensors are non-invasive and allow for obtaining data in real-time. The best model created combines Convolutional Neural Network (CNN) with LSTM, which allows for predicting pH with high performance. Continuing with yogurt, Wu et al. [115] created an LSTM model to predict the performance of *Lactobacillus* fermentation, which represents a step towards efficient production within the context of Industry 4.0. These approaches show that, as identified in the studies analyzed, LSTMs are useful models to address this problem.

In a broader sense, Zhang et al. [122] used DL algorithms to predict molecular biotransformations in fermented plant products, such as fermented chili sauces based on peppers, to predict capsaicin metabolites or toxins. The best performance was obtained by a Deep Forest model, which is a novel DL approach for decision trees with the advantage that it requires less data compared to other methods such as ANN. Therefore, the application of these methodologies in the fermentation process in cheese production could be investigated, and the development of appropriate models could continue.

## Coagulation

Milk coagulation, forming a gel from liquid milk using coagulating agents, is a fundamental step in the cheese-making process. It is determined by the type of cheese, the properties and characteristics of the milk, and the acidification process, although in general, it lasts between 40 and 60 minutes for most cheeses [44]. The gel or curd must be cut at the moment when it has the appropriate rheological and microstructural properties. Therefore, determining the final coagulation time is one of the critical variables of the process and is subject to great variability. However, in most factories, this is done from a fixed, established time or by visual and subjective inspection carried out by an operator [8].

For this reason, the use of advanced computer vision techniques together with ML algorithms is a solution to improve the efficiency of the process. One proposed approach by Loddo et al. [77] is based on the detection of anomalies from RGB images of Pecorino cheese coagulation to determine the optimal cutting time. To do this, different methods trained on the non-target class only (normal curd condition, not the optimal cutting time) were tested. Firstly, methods with feature extraction using handcrafted techniques or DL features with pre-trained CNN together with ML algorithms for anomaly detection such as IF or One-Class SVM. However, the best result was obtained with One-Class Fully Convolutional Data Description Network (FCDDN), a DL model for anomaly detection that performs feature extraction internally with fully convolutional layers [75].

From the same dataset, Loddo et al. [78] developed a classification model. Two methodological approaches are proposed: space-based features, where a single image was used to extract spatial features and classify them, and spatiotemporal-based features, where a sequence of images was used to extract temporal features and merge them using autocorrelation and Structural Similarity Index Measure (SSIM) to finally classify. Similarly, feature extraction is performed using handcrafted techniques or DL with pre-trained CNN. For the classification process, Random Forest (RF) algorithm and an MLP, a type of ANN, is used. The best results were obtained with features extracted with pre-trained CNN and an MLP model as a classifier (Table 3).

### Challenges and Future Trends

The coagulation stage represents a challenge for researchers due to the complication of obtaining data. Many relevant and interdependent factors influence the coagulation process, but these are often hard to measure accurately, leading to a

**Table 3** Coagulation summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| Optimal cutting time | FCDDN | Digital camera | RGB images | 1162 | [77] |
| Optimal cutting time | Pre-trained CNN + MLP | Digital camera | RGB images | 1162 | [78] |

scarcity of high-quality data suitable for ML methodologies. As we have seen, the main variable of interest is to predict the cutting time, and other research in the literature has also focused on this problem using other approaches. Furthermore, computer vision methods are promising for addressing this problem, alongside other techniques such as spectrometry and pulse-echo. However, the studies do not use ML methods, which represents an opportunity for its application.

For example, Nicolau et al. [94] used a Near-Infrared (NIR) light backscatter sensor as a non-invasive method to monitor the process. With the data obtained, they performed statistical analysis and built linear regression models to predict the curd-cutting time. Similarly, Villaquiran et al. [112] evaluated a low-cost commercial multifiber probe as an alternative to NIR for measuring the firmness of the elastic modulus to predict the optimal cutting time. A non-linear regression model was fitted using least squares optimization, achieving high accuracy. Another alternative is to use the ultrasonic pulse-echo technique to measure the acoustic impedance and, from it, build a mechanistic model that allows the cutting time to be predicted [34]. These techniques are compatible with ML models that are likely to improve current performance. Therefore, a valuable contribution to the area would be to investigate the use of these sensors together with ML algorithms to compare the results with respect to the traditional methods developed.

Finally, the work of Everard et al. [38] is also worth mentioning, where computer vision techniques based on color are used to study the coagulation phase. This research uses both digital images and color measurements obtained by a colorimeter. However, they used four-parameter non-linear mathematical models and basic processing techniques based on thresholds. Therefore, the use of advanced DL techniques, such as those described above, but based on color, is a promising approach for future research.

While most studies have focused on predicting the optimal cutting time, ML algorithms could also be applied to real-time monitoring of key coagulation parameters such as temperature, coagulant action time, and texture. These factors influence gel formation and could be integrated into adaptive control models to optimize the process dynamically.

## Ripening

The ripening stage, which can last from weeks to several years, is key to developing a quality final product. However, determining the state of the cheese is a difficult process to monitor and has traditionally been done industrially with methods such as visual inspection by operators, grading (tasting), or weight control, which are limited in precision and subject to human error. Additional instrumental methods have also been used at laboratory levels. Using innovative and noninvasive techniques is essential for the cheese industry [66].

In research, three key problems have been addressed with ML implementations: predicting sensory characteristics, determining the ripening extent, and estimating the number of ripening weeks or days. Various techniques, including spectroscopy, computer vision, and volatile compound analysis, have been applied. One of the first approaches has been the creation of a system to predict the sensory characteristics of cheeses at the ripening stage using NIR Spectroscopy and ANN [48]. In this case, a panel of experts has been used to rate 16 sensory attributes as the target variable for the prediction. An MLP model was trained for each sensory attribute using the NIR spectrum data reduced by PCA to 3 components. However, there was variability in the performance of the models, and some characteristics, such as intensity of flavor or greasy feeling, obtained a better result. Another approach to characteristic prediction performed by Golzarijalal et al. [50] was based on data extracted on Cheddar (832 samples) and Mozzarella (327 samples) from a literature review. For both cheeses, there are seven common characteristics, and two specific ones are added for each type. The objective was to train regressors capable of predicting soluble nitrogen, and among the different algorithms, Gradient Boosting was the one that obtained the best results in both cases. However, the authors do not specify which implementation was used or how it was configured. Gradient Boosting is a method that combines multiple weak decision trees sequentially, correcting errors to improve model accuracy [14]. In addition, the importance of the variables was analyzed using SHapley Additive exPlanations (SHAP) values where storage time is the most important. SHAP values are a technique for interpreting ML models. They are based on game theory to explain each feature's contribution to a model's prediction [80].

Computer vision is one of the low-cost, noninvasive techniques that can help monitor the cheese production process. Loddo et al. [76] used digital images of Pecorino cheese wheels on different days of ripening, which are data augmented and processed to separate the background from the whole cheese. Feature extraction methods were then used, both handcrafted features and pre-trained CNN models, to train different ML classifiers to predict the ripening day. The best result was obtained with pre-trained CNN features combined with an SVM classifier, even outperforming transfer learning with fine-tuning of the last layer in pre-trained CNN models. Continuing with computer vision and images of cheese wheels, Zedda et al. [121] proposed a novel approach that develops a hierarchical classifier with two levels of classification that allows grouping of cheeses by category and ripening level (ready or not). The authors used hierarchical concatenation and selection of features extracted from different pre-trained CNNs and trained a stacked classifier (using classifiers such as RF or SVM) that was combined with a logistic regression model. This approach compared and outperformed the simple feature extraction and flat classifier methodologies. Additionally, as a novelty, they explored the use of vision transformer architectures for feature extraction, but it performed worse than CNN models.

Knowing the state of ripening is essential and has generated more research. From another perspective, High et al. [58] analyzed the use of Volatile Compound (VOC)s extracted using the Solid Phase MicroExtraction, Gas Chromatography and Mass Spectrometry (SPME-GC-MS) technique from blue cheese samples. These samples were used directly and also in subsets selected by the Variable Identification Coefficients (VIC) and Entropy-Based Feature Selection (EFS) methods to train Self-Organizing Maps (SOM). SOMs are a type of unsupervised neural network that is used primarily to reduce dimensionality while maintaining the topology of the data, but in this case, they were used in a supervised context. The best result was obtained with the combination of EFS and SOM, outperforming linear methods such as Partial Least Squares Regression (PLS-R). However, extraction of VOCs is a complex and expensive technique, so for the specific problem of predicting ripening weeks, it may not be suitable for application in the industry. Finally, Finally, Martín Tornero et al. [84] also used NIR Spectra for the identification of both the ripening week and the cheese variety between Torta del Casar (TC) (402 samples) and Queso de la Serena (QS) (569 samples). In addition to using the spectrum data directly, different dimensionality reduction techniques such as Multivariate Curve Resolution with Alternating Least-Squares (MCR-ALS), Successive Projection Algorithm (SPA), or PCA were explored. The data was used to train a binary classifier that allows the cheese variety to be differentiated and a multiclass classifier for the ripening week. For the binary problem, an ANN model obtains the best performance, while for predicting the week of ripening, the Linear Discriminant Analysis (LDA) model outperformed the Quadratic Discriminant Analysis (QDA) and ANN models (Table 4).

## Challenges and Future Trends

The ripening phase has been studied using different approaches, including data from panels, experimental measurements and volatile compounds, spectral data, or images. Research employing the most advanced ML methods utilizes transfer learning techniques based on pre-trained CNN architectures for images. Therefore, it is interesting to extend computer vision methodologies to other problems during ripening, such as determining the optimal ripening stage, monitoring the evolution of texture or identifying surface defects [87], and exploring more advanced DL alternatives. For example, Priyashantha et al. [99] used Hyperspectral Imaging (HSI) in NIR range to predict the end date of the ripening process. From 425 data samples, they trained a PLS-R model that achieved a relatively modest performance that could be potentially improved with DL techniques.

In addition, Seratlic et al. [105] reviewed Fourier Transform Infrared (FT-IR), NIR, and Nuclear Magnetic Resonance (NMR) techniques for the ripening stage of Cheddar cheese. The authors highlighted the potential use of ML in combination with these methods to overcome the limitations of conventional analytical techniques still in use.

Much research has focused on the study of VOCs in cheese. Bertuzzi et al. [15] reviewed a wide range of analytical techniques for detecting VOCs in cheese. Therefore,

**Table 4** Ripening summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| 17 sensory attributes | MLP | NIR and Sensory panel | NIR spectra | 256 | [48] |
| Soluble nitrogen value | Gradient boosting | Literature review | 7 common characteristics + 2 specific to each variety | 832 / 327 | [50] |
| Degree of ripeness (not mature, mature, too mature) | Pre-trained CNN + SVM | Digital camera | RGB images | 195 | [76] |
| Adequate ripening or not | EfficientNet-DarkNet-53 + Stacked Classifier | Digital camera | RGB images | 378 | [121] |
| Ripening week | EFS + SOM | SPME-GC-MS | 23 VOCs characteristics | 108 | [58] |
| Variety of cheese / Ripening week | MLP / LDA | NIR | NIR spectra | 402 / 569 | [84] |

applying ML approaches to analyze these datasets seems like a logical next step in this line of research. As an example of other types of food, Gonzalez Viejo et al. [51] analyzed VOCs in the case of beer. As previously mentioned, the SPME-GC-MS technique for obtaining VOCs is complex to apply, so these authors create an ANN model to quickly predict VOCs using low-cost gas sensors. This approach allowed them to predict the aroma of beer with acceptable results, although it could be potentially improved with more complex DL architectures. This study demonstrates the possibility of developing new alternative approaches based on DL that can effectively replace traditional and expensive techniques.

From another perspective, automated control systems incorporating sensors and mathematical models are already used in industrial productions [6]. Therefore, the feasibility and benefits of integrating these systems with ML could be investigated. In this way, ML could be explored as a tool for optimizing storage conditions in cheese ripening chambers, potentially assisting in adjusting temperature, humidity, and airflow based on predictive models.

## Final Product Assessment

The final product assessment stage, once the manufacturing process has finished and we have the final cheese, is the one that concentrates the greatest number of research works that have used ML methods. Four main categories of objectives are identified at this stage: Food safety, Adulteration, and fraud, Prediction of characteristics, and Classification by variety or region of origin. So, articles are grouped for each of them as follows.

### Food Safety

*Salmonella*, *Escherichia coli*, and *Listeria* are some of the most important pathogens in the dairy industry. The contamination of cheese by these bacteria has been detected and is associated with outbreaks of diseases. In the production process, these bacteria can contaminate the cheese even after the pasteurization stage, which is a challenge for the industry [96].

In this sense, Zhen Jia et al. [124] investigated the identification of *Salmonella* and *Escherichia coli* in grated cheese with high background microflora in a noninvasive way by analyzing digital images to identify color patterns generated by the Paper Chromogenic Array sensor technique. This method detects VOCs emitted by bacteria, which cause specific color changes in the sensor, allowing their identification. In this case, the color variations extracted from the RGB images were used to train an MLP neural network architecture that classifies the samples based on their contamination (Table 5).

On the other hand, the detection of contaminating fungal pathogens has also been investigated by noninvasive methods using HSI. Specifically, hyperspectral images with 204 bands are obtained, and PCA was applied as an unsupervised method for dimensionality reduction. The visualization of the three principal components allowed the identification of contaminated samples [53]

### Adulteration and Fraud

Among dairy products, cheese has the highest rate of adulteration and food fraud. Some of the main practices are insufficient ripening time, replacing ingredients with cheaper ones, mislabeling, or not following the manufacturing processes established by the Protected Designation of Origin (PDO) [2]. In addition to not complying with regulations and harming consumers with lower-quality products, fraudulent practices also generate unfair competition between producers as they alter prices in the market [29].

Using HSI from the ultraviolet to NIR range has also been analyzed by [11] to detect adulteration with corn flour in fresh cheese. The goal was to predict the starch concentration since corn flour increases its content in the cheese. From hyperspectral images of 101 bands that were segmented to define the region of interest, a PLS-R was trained that obtained a high performance and identified the 5 most important wavelengths. When using these 5 wavelengths to train a new PLS-R model, the result decreased, but the model was simpler and more efficient.

The Italian cheese Fiore Sardo PDO must be produced from raw milk without any processing. However, many industries treat the milk in some way to prolong its shelf life, which constitutes food fraud. To detect this fraud, Anedda et al. [7] investigated the use of up to 6 pre-trained CNN architectures using two different types of input: Magnetic Resonance Imaging (MRI) Relaxometry and another one from digital images. In both cases, the architecture that worked best is SqueezeNet, performing transfer learning in which the last layer was fine-tuned, but with the advantage

**Table 5** Final product assessment: Food safety summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| Pathogen identification | MLP | Paper Chromogenic Array sensor and Digital camera | 27 color shift values | 1500 | [124] |
| Pathogen detection | N / A | HSI | HSI images | Unknown | [53] |

that the MRI analysis is not invasive, whereas, for digital images, the cheese must be cut.

Similarly, Buffalo Mozzarella Cheese PDO must be produced from fresh buffalo milk. However, there are periods of shortage of this milk. Some producers adulterate the production process by freezing part of the fresh curd produced during periods of abundance and using it in fresh and frozen curd blends when needed. To identify cheese production from frozen curd, Mengucci et al. [89] proposed to use NMR Relaxometry with Carr-Purcell-Meiboom-Gill (CPMG) to analyze the structural properties of cheeses. They then processed the data to reduce its dimensionality with PCA and selected 4 components to train a QDA classifier model that allowed differentiating between four percentages of added frozen curd.

In the case of Parmigiano Reggiano cheese, a fraudulent practice is to add other cheaper cheeses, such as Ricotta, to increase the quantity of grated products. At the visual level, even if there is 40% Ricotta, it is difficult to detect fraud. In addition, proteolysis analysis does not provide a clear distinction. Fagnani et al. [39] proposed using urea PolyAcrylamide Gel Electrophoresis (PAGE) to quantify the protein and, with this data, trained classification models such as Naïve Bayes (NB) that allowed the identification of adulterated samples.

Another form of fraud in grated Parmigiano Reggiano cheese is the inclusion of rind, which affects the authenticity and quality of the cheese. Starting from metabolomic profiles (2739 initials that are reduced to 598 compounds) obtained by Vanquish Ultra-High-Pressure Liquid Chromatography (UHPLC), Becchi et al. [13] performed various exploratory analyses such as Hierarchical Cluster Analysis (HCA), Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), and ANOVA Multiblock Orthogonal Partial Least Squares (AMOPLS), but these analyses did not reveal significant metabolomic differences related to rind inclusion. Finally, with the RF algorithm, the authors were able to classify the cheese samples based on rind inclusion, although with lower performance compared to the prediction of the altimetric zone of production, which achieved better results (Table 6)

### Prediction of Characteristics

The prediction of characteristics, whether physicochemical or sensory, is important for the development of cheese sensory properties such as aroma, texture, color, or taste. These types of characteristics determine the quality or consumer acceptability and shelf life of cheeses, so it is important to control them [59].

Sánchez-González et al. [107] studied the shelf life of the vacuum-packed Cajamarca soft cheese since it is an important aspect of the commercialization of the product. To do this, they used experimental measurements such as maturation time and temperature storage together with failure probability during storage, which was calculated from modified Weibull [70] analysis based on data obtained by a sensory panel. This information was used to train ANN and predict shelf lifetime (based on panel rejection) and titratable acidity (chemical measurement). The results of the ANNs outperformed traditional mathematical methods such as the multivariate linear regression model.

On the other hand, Münch et al. [91] investigated the prediction of $CO_2$ solubility, a key factor in the shelf life of cheese. To achieve this, they used the variables fat, water, protein, salt, and temperature measures, as well as $CO_2$ solubility values, all collected from 21 references in the literature. From these variables, they trained linear, local, and ensemble methods of ML algorithms to predict the value of $CO_2$ solubility, among which the performance of RF stood out. In addition, the importance of the variables was then analyzed from the SHAP values.

Furthermore, to comply with the European labeling regulations under *Regulation (EU) No 1169/2011* [37], producers must report the level of minerals in cheeses. However, there are no fast and non-invasive methods to calculate these components. For this reason, Menevseoglu et al. [88] focused on the use of NIR and MidInfrared (MIR) Spectroscopy for feature extraction and ML algorithms for prediction. In the study, the researchers used samples of Erzincan Tulum cheese and obtained labels for prediction models using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) to determine the

**Table 6** Final product assessment: Adulteration and fraud summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| Starch concentration | PLS-R | HSI | HSI images | 44 | [11] |
| Detect raw or treated milk | Pre-trained SqueezeNet | MRI / Digital camera | MRI Relaxometry / RGB images | 113 / 25 | [7] |
| Detect frozen curd content | QDA | NMR Relaxometry and CPMG | T2 relaxometry curves | 60 | [89] |
| Detect authentic cheese content | NB | Urea PAGE | 6 protein values | 35 | [39] |
| Detect rind inclusion content | RF | Vanquish UHPLC | 598 metabolomic profiles | 164 | [13] |

real values of mineral content (Al, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, and P) and DPPH antioxidant activity, as it is also a relevant quality parameter. The results showed that feature selection with EFS together with Gaussian Process Regression (GPR) superseded methods such as PLS-R or SVM.

In terms of sensory perception analysis, Rocha et al. [100] investigated the use of emerging techniques such as Ohmic heating treatment at different voltage levels, and its impact on the sensory acceptability of Minas Frescal cheese. To do so, they used a panel of consumers who evaluated 16 sensory attributes. From the evaluations, different ML regression algorithms were used to predict the overall liking and determine both the most polite ohmic treatment and the most influential sensory attributes on acceptability. The authors highlighted the use of RF to achieve these objectives. Using another approach, Chaturvedi et al. [25] predicted, in addition to the overall acceptability, some of the key sensory characteristics such as taste, smell, and mouthfeel of Paneer cheese using fuzzy logic and an ANN model for prediction. To do this, they used 10 experimental chemical and microbiological measures as input and a panel of 25 experts to obtain the values of the sensory characteristics.

Focusing solely on aroma as a sensory characteristic, Caille et al. [24] created prediction models for processed cream cheese. They analyzed the use of VOCs obtained by in vitro mastication and HeadSpace Solid-Phase MicroExtraction (HS-SPME) followed by Gas Chromatography Time-of-Flight Mass Spectrometry (GCToF-MS) and selected 18 of them. On the other hand, the fresh cream aroma sensory descriptor was obtained through the Rate-All-That-Apply (RATA) sensory test. From these characteristics, the RF ML model was successfully trained, which outperformed simpler regression tree models.

Finally, Zlatev et al. [126] developed a multimodal approach to predict the characteristics of white brine cheeses. In this sense, three different data sources were used: 9 visible spectra indices obtained from digital images, 12 characteristics from gas sensors, and 12 ultrasonic characteristics. The information was then combined to select the most informative features and create vectors, which are reduced in dimensionality by employing Latent Variables. With this data, regression models were created to predict the values of pH, electrical conductivity, total dissolved solids, and oxidation-reduction potential that were obtained from direct experimental measurements on the cheese samples. In this case, a model was trained for each characteristic, and the performance of PLS-R over Principal Component Regression (PCRe) and a quadratic mathematical model was highlighted (Table 7).

## Classification by Variety or Region of Origin

The human olfactory system can identify complex chemical patterns such as those occurring in cheese and are characteristic of the product. Schroeder et al. [104] proposed to mimic this system with a technology that uses an array of chemical sensors based on a Single Walled Carbon NanoTube (SWCNT), each one with a different selector, to detect odors by reacting with the volatile chemical compounds present in cheese samples. The goal was, therefore, to use this information to classify samples based on the cheese variety. The samples were then heated and exposed for several minutes, allowing data to be captured in the form of time series. Better results were obtained by extracting features from the time series and training a classification algorithm such as RF rather

**Table 7** Final product assessment: Prediction of characteristics summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| Shelf lifetime and titratable acidity | MLP | Experimental measurements and Sensory panel | Maturation time, storage temperature and failure probability (F(x)) | Unknown | [107] |
| $CO_2$ Solubility | RF | Literature review | Fat, water, protein, salt and temperature measures | 258 | [91] |
| Mineral content and antioxidant activity | GPR | NIR, MIR and IPC-MS | NIR and MIR spectra | 70 | [88] |
| Overall liking | RF | Sensory panel | 16 sensory attributes | 400 | [100] |
| Color, Taste, Smell, Mouthfeel and Overall liking | MLP | Experimental measurements and Sensory panel | 10 experimental measures | Unknown | [25] |
| Aroma | RF | HS-SPME, GC-ToF-MS and Sensory panel | 18 VOCs characteristics | Unknown | [24] |
| pH, electrical conductivity, total dissolved solids, oxidation-reduction potential | PLS-R | Digital camera, Gas sensors, Ultrasonic sensors, and Experimental measurements | Feature vector of combined data | 360 | [126] |

than using them directly as input into a K-Nearest Neighbors (KNN).

VOCs have also been successfully used to differentiate between cheese-type varieties. In particular, High et al. [57] focused on 17 different blue cheese varieties and uses the SPME-GC-MS technique to obtain volatile fingerprints and compared different approaches for classification. All obtained VOCs were used as input, and feature selection techniques such as VIC and EFS were also compared to select the most important VOCs. The best result was obtained with a supervised SOM model using EFS for VOCs selection, outperforming linear methods such as Partial Least Squares Discriminant Analysis (PLS-DA).

Different varieties of Swiss cheese have also been studied by Fröhlich-Wyder et al. [47] to classify and differentiate them based on Free Volatile Carboxylic Acid (FVCA) since they are an indicator of quality. The chemical profiles of the FVCAs were extracted, and from eight features (C1, C2, C3, C4, C6, iso-C4, iso-C5, iso-C6) up to 14 different ML classification algorithms were trained, among which Extra Trees and RF stand out above the rest. In addition, SHAP values were used to evaluate the importance of these eight characteristics in the classification result, and it was determined that FVCAs C1, C3, C6, and iso-C4 are the most relevant.

The study by de Andrade et al. [32] performed a mineral analysis in the samples to predict both the variety of Brazilian artisanal cheese and its origin. With the analysis of the composition of seven minerals (major mineral: Ca, K, Mg, Na; minor mineral: Cu, Mn, Zn) from Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES), the type of cheese variety was identified among eleven different ones, obtaining the best results with the SVM algorithm that surpassed the performance of ANN and other algorithms. On the other hand, the production region of the cheese could be detected using the RF classifier that surpassed the performance of ANN and other algorithms.

Furthermore, Kamilari et al. [64] also investigated the origin of cheeses using a genetic biomarker-based approach to identify their provenance. To do this, deoxyribonucleic acid (DNA) sequences are extracted from cheese samples from three countries and grouped into Operational Taxonomic Unit (OTU) to assign a taxonomic classification. From the OTU abundance data, the RF classification algorithm was used to predict the country of origin of the samples with high accuracy. Magarelli et al. [81] applied metagenomic sequencing to obtain the taxonomic profiles of bacterial relative abundance in samples of Mozzarella di Bufala Campana PDO cheese. The objective was to differentiate the region of origin and distinguish between Salerno and Caserta. To do so, they used the data obtained to train classification models, with RF achieving the best results. In addition, they explored the use of SHAP analysis to identify the most relevant bacteria that allow for distinguishing the region to which a cheese sample belongs.

Finally, Zhang et al. [123] classified between cheese brands from different producers. For this purpose, they relied on Raman spectroscopy technology to extract features that are reduced by PCA to 74 components to train an Extreme Learning Machine (ELM model. This work also showed the importance of using data processing techniques such as normalization or denoising since they significantly improved the performance of the model. In Zlatev et al. [126], in addition to predicting sensory characteristics, the authors also trained binary classifiers to identify the manufacturer that produced the cheese, where the performance of the SVM stands out (Table 8).

## Challenges and Future Trends

Most ML applications focus on the final product assessment stage. One possible explanation is due to the easier access to data and experimental measurements. Character-

**Table 8** Final product assessment: Classification by variety or region of origin summary

| Objective | Best model | Data Source | Input | Samples | Reference |
|---|---|---|---|---|---|
| Variety of cheese | RF | SWCNT | 794 extracted features from time series | 720 | [104] |
| Variety of cheese | EFS + SOM | SPME-GC-MS | 14 VOCs characteristics | 102 | [57] |
| Variety of cheese | Extra Trees | FVCA | 8 FVCAs characteristics | 241 | [47] |
| Variety of cheese / Region of cheese | SVM / RF | IPC-OES | 7 mineral values | 402 | [32] |
| Region of cheese | RF | Metataxonomic sequencing | OTU abundance data | 49 | [64] |
| Brand of cheese | ELM | Raman spectroscopy | Raman spectra | 75 | [123] |
| Region of cheese | RF | Metagenomic sequencing | Taxonomic profiles of bacterial relative abundance | 65 | [81] |
| Manufacturer | SVM | Digital camera, Gas sensors, Ultrasonic sensors, and Experimental measurements | Feature vector of combined data | 360 | [126] |

istics prediction of the final cheeses relies on both complex data acquisition techniques, such as HS-SPME followed by GC-ToF-MS for VOCs analysis, and simpler image-based approaches. The applicability of these techniques should be evaluated, as more accessible alternatives could be more viable. However, the ML models employed in these studies are generally not highly complex or optimized, highlighting an opportunity for further advancements. In this regard, Elangovan et al. [36] used more advanced ML methods to assess the quality and freshness of meat. The authors employed ConvNet-18 and ConvNet-24 CNN architectures entirely trained on the MEAT2C and MEAT3C image databases, achieving satisfactory results. In addition, systematic dimensionality reduction of the data is not necessary, and it would be interesting to analyze the performance of advanced DL models since they can handle high-dimensional data [71] such as novel approaches based on transformer architectures [110]. In any case, for dimensionality reduction, it is possible to explore advanced non-linear alternatives to the recurrent use of PCA, such as the one proposed by Nareklishvili and Geitle [93].

Regarding minerals, for the prediction of 17 macro, trace, and environmental minerals in bovine milk, Bisutti et al. [17] used FT-IR and trained various ML prediction models. Their significant contribution was the improvement in performance achieved by combining predictions through a stacking ensemble model.

On the other hand, multimodal models that combine data from different sources to perform a common task have also been successfully developed using DL approaches [4]. For example, Wu and Jia [116] proposed a multimodal approach from a literature review of fermented dairy products, combining different omic layers (lipids, peptides, and proteins) using MLP and self-attention mechanisms.

Digital images were also used by Badaró et al. [10] to predict cheese meltability, aiming to automate and standardize the measurement of this property. Their results outperformed traditional analytical methods, such as the Schreiber test. However, the study relied on outdated computer vision techniques, suggesting that future research should explore more advanced methods for improved results. Similarly, Vega-Chinchay et al. [111] studied the bactericidal activity of *Lactobacillus acidophilus* using outdated computer vision techniques, highlighting the need for more advanced image analysis methods

Pathogen detection in cheese has been studied for many years to identify sources of contamination and propose control measures as it poses a risk to food safety [67]. The articles analyzed have focused on the identification of pathogens using non-invasive methods such as VOCs analysis and HSI. There are other numerous studies on the subject, but the use of ML is not widespread, and the number of works that integrate these techniques is limited. For instance, Torres-Vitela et al. [109] investigated the incidence of *Salmonella*, *Escherichia coli*, and *Listeria* in 200 samples of Panela and Adobera cheese. Using microbiological detection methods and statistical techniques like ANOVA, they determined the presence of at least one pathogen in 46% of the samples.

It is also common to use Polymerase Chain Reaction (PCR) to detect pathogens in cheese. Yoon et al. [120] reviewed several examples where this technique is used to detect contaminating pathogens in cheeses made from raw milk. Cremonesi et al. [27] used Droplet Digital PCR together with statistical methods to detect contaminating pathogens in fresh cheeses. However, there are no studies in the literature that explore the use of techniques like PCR with ML algorithms in cheese products. Nevertheless, research such as that of Yang et al. [119] explored improving the Droplet Digital PCR technique with DL algorithms so it can be applied.

On the other hand, in fields like medicine, various studies have shown that PCR detection methods can be improved with ML models to achieve superior performance [72] or even to create ML prediction models that can replace the need for PCR testing [12]. Therefore, these approaches and others that appear in the literature can be applied to cheese production.

Fraud in cheese production can occur in different ways, leading to a wide range of research using ML techniques and diverse approaches. Among these, imaging techniques stand out as non-invasive, cost-effective, and have demonstrated good performance. However, studies could use more sophisticated and powerful methods to improve results. In this regard, Barreto et al. [11] used HSI to train a PLS-R model instead of more advanced models. For example, Anedda et al. [7] employed transfer learning from pre-trained CNN architectures using MRI and RGB images, similar to Loddo et al. [78] and Loddo et al. [77] in the coagulation phase and Loddo et al. [76] and Zedda et al. [121] in the ripening phase. Transfer learning from pre-trained CNN architectures is also applicable to HSI [56]. For instance, it has also been successfully used by Yang et al. [118] to detect starch adulteration in minced chicken meat. The authors used HSI together with the GoogLeNet architecture pre-trained on ImageNet, which outperformed other CNNs and SVM models.

Fan et al. [40] proposed a novel approach using NIR cameras for real-time inspection of apples to identify anomalies. The authors trained and optimize a YOLOv4 architecture [19], achieving promising results. Notably, as of the date of this publication, the latest version of the YOLO architecture family is YOLOv11 [65], released in September 2024, suggesting potential for further improvements.

Furthermore, advanced algorithms can also be used to process NMR relaxometry data. In the literature, we found research lacking any ML approach, such as Małkowska-

Kowalczyk et al. [83] where they applied NMR relaxometry technique to evaluate the authenticity of different cheese varieties, relying on statistical methods like ANOVA. As an alternative, Date et al. [30] proposed a novel approach based on SVM with bootstrap using fish samples that allowed an optimized analysis of T2 relaxation curves. From the signal processing point of view, DL models such as CNN or LSTM have been used to process NMR data, predict parameters, or verify structures [26]. Finally, Deng et al. [33] provided a literature review on the use of ML and DL methods to verify authenticity and detect adulterations in food more broadly. Therefore, many of these approaches applied to the food industry could also be utilized in cheese production. However, in any case, it is crucial to analyze and discuss the necessity and feasibility of complex data acquisition methods such as NMR or MRI, considering the availability of efficient and simple alternatives or, conversely, the significant benefits and improvement offered by the use of these methods.

ML algorithms are not inherently tied to a specific category of the final product phase. What truly matters is the availability and quality of data, as well as the nature of the modeling task, which can be similar across different domains. Therefore, advanced techniques can potentially be applied to solve these problems using various data sources, including those from previous research, provided that the structure of the problem and the relevant variables are sufficiently comparable.

### Others

There have also been studies that do not focus on a single stage of the process but rather analyze multiple stages. For example, in the study by Jox et al. [63], parameters were collected from the fermentation, coagulation, and whey removal stages during the production of fresh cheese. If the data were in the form of a time series, simple statistical summaries were extracted and combined with the rest of the specific measurements. The main objective was to train a classification model to predict whether a production batch would have issues with whey turbidity and require additional cleaning. The eXtreme Gradient Boosting (XGBoost) and MLP models achieved the best results. In a different approach, Perrignon et al. [98] collected information using sensors and experimental measurements during the production of Emmental cheese. Data were obtained from the milk collection phase to brine, including standardization, coagulation, and whey removal. The objective was to predict the dry matter content at the brine stage based on these parameters. It is an approximation to combining variables from different stages of production. The best result was achieved with a RF model and, in addition, the importance of the variables and the SHAP values were analyzed.

Other types of research have not focused on a specific production stage but rather carry out a more theoretical development. For example, Munch et al. [92] proposed a method for learning a Probabilistic Relational Model. This approach represented the structural relationship in the data information (classes and attributes) to create a probabilistic model. Their objective was to extend a basic model by incorporating temporal information, thereby identifying causal relationships between key control variables of the process. The authors used data from the project funded by the European Union TrueFood and define temperature, type of milk, and starters as control parameters. They also used the measured attributes of hardening time, clotting time, and concentrations of various acids and sensory evaluations of texture and flavor. On the other hand, Yakoubi [117] explored the creation of plant-based cheese, a field of study that was generating increasing interest. In this regard, optimizing protein-ligand interactions was key to obtaining good sensory attributes. Based on features of free binding energies of protein-ligand complexes, classifiers were trained to predict docked free energy binding of protein-ligand complexes, and SVM stood out for its results.

## Discussion

The review has analyzed different ML approaches and their applicability in cheese-making. By distributing the research into specific stages of the production process, it has been observed that the final product assessment, when the cheese is finished, is the stage that concentrates a greater number of works. Specifically, this is due to the interest in detecting fraud and adulteration in cheeses since this is a problem that is increasing [2]. Moreover, at this stage, a large number of studies are also focused on evaluating the properties and quality of the cheese. Quality control should be carried out throughout the entire production process, but as has already been observed, the complexity of the process implies that research related to product quality is conducted once the cheese is made.

Other stages, such as coagulation or fermentation, present their own complexities and difficulties in creating datasets, which makes scientific progress in these stages more difficult and the development of solutions with ML. In general, cheese production and also the dairy industry has a problem with the scarcity of data, which is a major barrier to scientific and technological development in the sector [77].

Beyond the challenge of data scarcity, the real-world applicability of the developed models represents an additional barrier. A method may be successfully validated in research settings, but if its implementation requires complex procedures, its industrial applicability may be severely limited. Many studies focus on laboratory-scale experimental designs and do not integrate strategies compatible with Process Analytical Technology (PAT), which emphasizes real-time process monitoring and control. Developing ML models that align with PAT principles could lead to significant scientific impact.

In this regard, ML models have proven effective, and the results have been promising, emphasizing the need to continue exploring these methods. The incorporation of ML techniques and methods represents an opportunity for the development of future work, but it also implies complying with the requirements and methodologies exhaustively, as has been discussed in depth in Section "Fundamental Machine Learning Concepts". Although this review has focused primarily on presenting the studies rather than assessing their validity, we have observed that methodological rigor varies across works. In addition, differences in research approaches and data acquisition methods make direct comparisons difficult. It is essential to adopt a robust methodological approach to data management to ensure valid and generalizable results in future research.

To overcome these limitations, collaboration between research groups is essential. On the one hand, data and experiments must be shared whenever possible, allowing the rest of the scientific community to validate the results and propose improvements. On the other hand, food science is a multidisciplinary field that encompasses, among others, biology, chemistry, physics, engineering, nutrition, and now also areas such as artificial intelligence. This diversity means that researchers cannot have the necessary knowledge in all areas, which leads to a decline in methodological quality. Therefore, it is necessary to create research teams with experts in different areas.

Industries seek more efficient and data-driven approaches, generating growing interest in the integration of computational tools in food production. The term Computer-Aided Food Engineering (CAFE) has recently been introduced, which is defined as "the use of computer software for analysis in the design and control of food products, processes, packaging or equipment, from the molecular scale to that of an entire food plant" [31]. The authors highlight the need to implement data-driven models and advanced ML techniques to optimize production processes in the food industry. In this context, cheese production should also join the sector's trend towards digitalization and benefit from innovations by implementing these approaches.

In summary, Fig. 8 shows a general overview of the ML techniques used that have obtained the best result for each of the investigations. To improve visualization, algorithms that appear in only one or two studies are grouped together. For example, in the Others-DL category, DL algorithms such as LSTM, NARNN, and FCDDN are grouped together. In the Others-ML category, various ML algorithms, such as NB, decision tree variants, and boosting algorithms are grouped together. Meanwhile, in the Others-Classical category, we have the traditional approaches LDA, QDA, PLS-R, and LR grouped together. Overall, we observe a trend toward introducing more advanced techniques that outperform classical statistical methods.

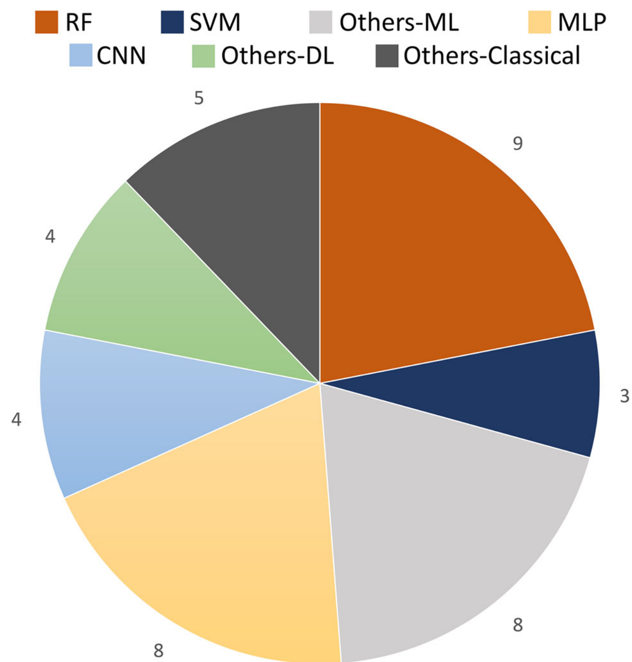


**Fig. 8** Algorithms distribution

However, the presence of classical ML algorithms remains highly significant. Additionally, MLP models from the majority of studies should be included in this group, as they are very small and simple architectures with a single hidden layer. In this case, we see that the RF algorithm stands out, which is a method that has traditionally given good results but has been surpassed more recently by the advance of DL. Regarding this, throughout the review, it has been observed that pre-trained CNN architectures have been more effective than other techniques for extracting features from data, specifically in images. However, the limitation to obtaining large sets of quality data prevents them from being developed properly and from being able to exploit their full potential. Therefore, this can also be seen as an opportunity for the sector.

Future lines of research exploring the potential uses of advanced ML and DL algorithms must be considered as a challenge within this field at an early stage, taking into account the general workflow and the importance of data. In this way, research can focus efforts on obtaining quality data and creating datasets that will lead to robust research with generalizable results that can be applied to the development processes of real industry.

## Conclusions

The context of Industry 4.0 presents an opportunity for sectors that have not yet widely implemented digitalization into their processes. In this context, the food industry is in a phase of integrating advanced methods. This review examines the use of ML techniques in cheese production, as it is a sector that particularly relies on traditional methods. Furthermore, the number of studies exploring the use of ML for cheese production has been limited.

For this reason, the article provides an introduction to the general concepts of ML so that researchers without knowledge in the field can assimilate basic concepts and have access to key references to expand their knowledge. The application of ML in cheese production has the potential to improve traditional analysis techniques. However, a transformation is required in the industry, which involves raising awareness of the importance of data as a key factor in developing quality models. In this regard, it is essential that data collection processes be thorough and conducted with the utmost rigor.

When applied appropriately, ML methods offer the potential to make a qualitative leap in industry and production processes, enhancing decision-making and optimizing operations. In addition, research also highlights the potential of using these models to gain a deeper understanding of processes by applying explainability techniques, allowing researchers to identify the most relevant specific factors for their studies. Consequently, ML can be used in many ways to analyze compositional, processing, and quality data to enhance cheese manufacturing and final product control.

In terms of research focus, most studies have focused on the final product assessment stage when the cheese is already made. This phase is relevant from the perspective of adulteration and fraud, as it allows for the detection of fraudulent practices in the industry. These practices, besides deceiving consumers, represent an unfair advantage over producers who follow proper processes. Therefore, it is precisely these producers who should be most interested in promoting ML models to help mitigate such unfair actions.

Finally, the remaining stages of production present a considerable challenge for researchers due to the complexity of the process and the difficulty of obtaining data. In this regard, the measurement of relevant variables and continuous monitoring is complicated. To address these difficulties, studies show that the use of advanced computer vision techniques using digital images is a promising option and achieves satisfactory results. Research has also been conducted using techniques such as HSI, MRI, NMR, or spectroscopy as a data source, but their implementation in industry presents greater difficulties.

### Glossary of Acronyms

| | |
|---|---|
| *AI* | Artificial Intelligence |
| *AMOPLS* | ANOVA Multiblock Orthogonal Partial Least Squares |
| *ANN* | Artificial Neural Network |
| *CNN* | Convolutional Neural Network |
| *CPMG* | Carr-Purcell-Meiboom-Gill |
| *CRISP-DM* | Cross-Industry Standard Process for Data Mining |
| *DL* | Deep Learning |
| *DNA* | deoxyribonucleic acid |
| *EFS* | Entropy-Based Feature Selection |
| *ELM* | Extreme Learning Machine |
| *FCDDN* | Fully Convolutional Data Description Network |
| *FT-IR* | Fourier Transform Infrared |
| *FVCA* | Free Volatile Carboxylic Acid |
| *GANs* | Generative Adversarial Networks |
| *GC-ToF-MS* | Gas Chromatography Time-of-Flight Mass Spectrometry |
| *GPR* | Gaussian Process Regression |
| *HCA* | Hierarchical Cluster Analysis |
| *HS-SPME* | HeadSpace Solid-Phase MicroExtraction |
| *HSI* | Hyperspectral Imaging |
| *ICP-MS* | Inductively Coupled Plasma Mass Spectrometry |
| *ICP-OES* | Inductively Coupled Plasma Optical Emission Spectrometry |
| *IF* | Isolation Forest |
| *KNN* | K-Nearest Neighbors |
| *LDA* | Linear Discriminant Analysis |
| *LR* | Logistic Regression |
| *LSTM* | Long Short-Term Memory |
| *MCR-ALS* | Multivariate Curve Resolution with Alternating Least-Squares |
| *MIR* | Mid-Infrared |
| *ML* | Machine Learning |
| *MLP* | Multi Layer Perceptron |
| *MRI* | Magnetic Resonance Imaging |
| *MSE* | Mean Squared Error |
| *NARNN* | Nonlinear Autoregressive Neural Network |
| *NB* | Naïve Bayes |
| *NIR* | Near-Infrared |
| *NMR* | Nuclear Magnetic Resonance |
| *OPLS-DA* | Orthogonal Partial Least Squares Discriminant Analysis |
| *OTU* | Operational Taxonomic Unit |
| *PAGE* | PolyAcrylamide Gel Electrophoresis |
| *PAT* | Process Analytical Technology |
| *PCA* | Principal Component Analysis |
| *PCR* | Polymerase Chain Reaction |
| *PCRe* | Principal Component Regression |
| *PDO* | Protected Designation of Origin |
| *PLS-DA* | Partial Least Squares Discriminant Analysis |
| *PLS-R* | Partial Least Squares Regression |

| | |
|---|---|
| *QDA* | Quadratic Discriminant Analysis |
| *QS* | Queso de la Serena |
| *RATA* | Rate-All-That-Apply |
| *RF* | Random Forest |
| *RMSE* | Root Mean Squared Error |
| *ROC* | Receiver Operating Characteristic |
| *SHAP* | SHapley Additive exPlanations |
| *SOM* | Self-Organizing Maps |
| *SPA* | Successive Projection Algorithm |
| *SPME-GC-MS* | Solid Phase MicroExtraction, Gas Chromatography and Mass Spectrometry |
| *SSIM* | Structural Similarity Index Measure |
| *SVM* | Support Vector Machines |
| *SWCNT* | Single Walled Carbon NanoTube |
| *TC* | Torta del Casar |
| *UHPLC* | Ultra-High-Pressure Liquid Chromatography |
| *VIC* | Variable Identification Coefficients |
| *VOC* | Volatile Compound |
| *XGBoost* | eXtreme Gradient Boosting |

**Author Contributions** Daniel Pardo: Conceptualization, Methodology, Software, Investigation, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. Manuel Castillo: Supervision, Conceptualization, Writing – review & editing. Mehmet Oguz Mulayim: Supervision, Writing – review & editing. Jesus Cerquides: Supervision, Conceptualization, Methodology, Writing – review & editing.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Daniel Pardo is funded by the AGAUR FI Joan Oró grant (2024 FI-1 00424).
Jesus Cerquides and Mehmet Oguz Mulayim are funded by European Union KHealthInAir and GUARDEN projects with No. 101057693 and 101060693. This work was supported by grant PID2022-136787NB-I00 funded by MCIN/AEI/10.13039/501100011033.
Manuel Castillo is funded by pHotoYOGURT (ACC_2023_EXP_SIA001_40_0000092; DACC, GENCAT) and OPTICmizing (2023 INNOV 00027; AGAUR, GENCAT) projects.

**Data Availability** No datasets were generated or analysed during the current study.

## Declarations

**Competing interests** The authors declare no competing interests.



## References

1. Abdi H, Williams LJ (2010) Principal component analysis. WIREs Comput Stat 2(4):433–459. https://doi.org/10.1002/wics.101
2. Abedini A, Salimi M, Mazaheri Y, Sadighara P, Alizadeh Sani M, Assadpour E, Jafari SM (2023) Assessment of cheese frauds, and relevant detection methods: a systematic review. Food Chem: X 19:100825. https://doi.org/10.1016/j.fochx.2023.100825
3. Adhikari R, Agrawal RK (2013) An introductory study on time series modeling and forecasting. arXiv:1302.6613
4. Akkus C, Chu L, Djakovic V, Jauch-Walser S, Koch P, Loss G, Marquardt C, Moldovan M, Sauter N, Schneider M, Schulte R, Urbanczyk K, Goschenhofer J, Heumann C, Hvingelby R, Schalk D, Aßenmacher M (2023) Multimodal deep learning. arXiv:2301.04856
5. Alpaydin E (2020) Introduction to machine learning, 4th edn. MIT Press. Google-Books-ID: tZnSDwAAQBAJ
6. Alvarenga N, Martins J, Caeiro J, Garcia J, Pássaro J, Coelho L, Santos MT, Lampreia C, Martins A, Dias J (2021) Applying computational fluid dynamics in the development of smart ripening rooms for traditional cheeses. Foods 10(8):1716. https://doi.org/10.3390/foods10081716
7. Anedda R, Melis R, Curti E (2021) Quality control in Fiore Sardo PDO cheese: detection of heat treatment application and production chain by MRI relaxometry and image analysis. Dairy 2(2):270–287. https://doi.org/10.3390/dairy2020023
8. Arango O, Castillo M (2018) A method for the inline measurement of milk gel firmness using an optical sensor. J Dairy Sci 101(5):3910–3917. https://doi.org/10.3168/jds.2017-13595
9. Arthur D, Vassilvitskii S (2006) k-means++: the advantages of careful seeding. Technical report, Stanford
10. Badaró AT, de Matos GV, Karaziack CB, Viotto WH, Barbin DF (2021) Automated method for determination of cheese meltability by computer vision. Food Anal Methods 14(12):2630–2641. https://doi.org/10.1007/s12161-021-02094-1
11. Barreto A, Cruz-Tirado JP, Siche R, Quevedo R (2018) Determination of starch content in adulterated fresh cheese using hyperspectral imaging. Food Biosci 21:14–19. https://doi.org/10.1016/j.fbio.2017.10.009
12. Batista AFdM, Miraglia JL, Donato THR, Filho ADPC (2020) COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. Pages: 2020.04.04.20052092
13. Becchi PP, Rocchetti G, García-Pérez P, Michelini S, Pizzamiglio V, Lucini L (2024) Untargeted metabolomics and machine learning unveil quality and authenticity interactions in grated Parmigiano Reggiano PDO cheese. Food Chem 447:138938. https://doi.org/10.1016/j.foodchem.2024.138938
14. Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. Artif Intell Rev 54:1937–1967
15. Bertuzzi AS, McSweeney PL, Rea MC, Kilcawley KN (2018) Detection of volatile compounds of cheese and their contribution to the flavor profile of surface-ripened cheese. Compr Rev Food Sci Food Saf 17(2):371–390. https://doi.org/10.1111/1541-4337.12332
16. Bishop CM, Bishop H (2023) Deep learning: foundations and concepts. Springer Nature
17. Bisutti V, Mota LFM, Giannuzzi D, Toscano A, Amalfitano N, Schiavon S, Pegolo S, Cecchinato A (2024) Infrared spectroscopy coupled with machine learning algorithms for predicting

the detailed milk mineral profile in dairy cattle. Food Chem 461:140800. https://doi.org/10.1016/j.foodchem.2024.140800
18. BMJ (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews | The BMJ
19. Bochkovskiy A, Wang CY, Liao HYM (2020) YOLOv4: optimal speed and accuracy of object detection. arXiv:2004.10934
20. Bosakova-Ardenska A (2024) Recent trends in computer vision for cheese quality evaluation. Eng Proc 60(1):12. https://doi.org/10.3390/engproc2024060012
21. Bowler A, Ozturk S, di Bari V, Glover ZJ, Watson NJ (2023) Machine learning and domain adaptation to monitor yoghurt fermentation using ultrasonic measurements. Food Control 147:109622. https://doi.org/10.1016/j.foodcont.2023.109622
22. Burkart N, Huber MF (2021) A survey on the explainability of supervised machine learning. J Artif Intell Res 70:245–317. https://doi.org/10.1613/jair.1.12228
23. Cai H, Pei S, Zhang Y, Liu R, Lu S, Li B, Dong J, Wang Q, Zhu X, Ji H (2023) Construction of a dynamic model to predict the growth of Staphylococcus aureus and the formation of enterotoxins during Kazak cheese maturation. Food Microbiol 112:104234. https://doi.org/10.1016/j.fm.2023.104234
24. Caille C, Boukraâ M, Rannou C, Villière A, Catanéo C, Lethuaut L, Lagadec-Marquez A, Bechaux J, Prost C (2023) Analysis of volatile compounds in processed cream cheese models for the prediction of “Fresh Cream” aroma perception. Molecules 28(20):7224. https://doi.org/10.3390/molecules28207224
25. Chaturvedi K, Khubber S, Singha S, Goel H, Barba FJ, Das K (2020) Prediction and qualitative analysis of sensory perceptions over temporal vectors using combination of artificial neural networks and fuzzy logic: Validation on Indian cheese (paneer). J Food Process Preserv 44(12):e14955. https://doi.org/10.1111/jfpp.14955
26. Cobas C (2020) NMR signal processing, prediction, and structure verification with machine learning techniques. Magn Reson Chem 58(6):512–519. https://doi.org/10.1002/mrc.4989
27. Cremonesi P, Cortimiglia C, Picozzi C, Minozzi G, Malvisi M, Luini M, Castiglioni B (2016) Development of a droplet digital polymerase chain reaction for rapid and simultaneous identification of common foodborne pathogens in soft cheese. Front Microbiol 7:1725. https://doi.org/10.3389/fmicb.2016.01725
28. Croguennec T, Le Loir Y, Jeantet R (2025) Breakthrough innovations in industrial cheesemaking processes. Curr Opin Food Sci 62:101267. https://doi.org/10.1016/j.cofs.2024.101267
29. Danezis GP, Tsagkaris AS, Camin F, Brusic V, Georgiou CA (2016) Food authentication: techniques, trends & emerging approaches. TrAC Trends Anal Chem 85:123–132. https://doi.org/10.1016/j.trac.2016.02.026
30. Date Y, Wei F, Tsuboi Y, Ito K, Sakata K, Kikuchi J (2021) Relaxometric learning: a pattern recognition method for T2 relaxation curves based on machine learning supported by an analytical framework. BMC Chem 15(1):13. https://doi.org/10.1186/s13065-020-00731-0
31. Datta A, Nicolaï B, Vitrac O, Verboven P, Erdogdu F, Marra F, Sarghini F, Koh C (2022) Computer-aided food engineering. Nature Food 3(11):894–904. https://doi.org/10.1038/s43016-022-00617-5
32. de Andrade BM, Margalho LP, Batista DB, Lucena IO, Kamimura BA, Balthazar CF, Brexó RP, Pia AKR, Costa RAS, Cruz AG, Granato D, Sant’Ana AS, Luna AS, de Gois JS (2022) Chemometric classification of Brazilian artisanal cheeses from different regions according to major and trace elements by ICP-OES. J Food Compos Anal 109:104519. https://doi.org/10.1016/j.jfca.2022.104519
33. Deng Z, Wang T, Zheng Y, Zhang W, Yun YH (2024) Deep learning in food authenticity: recent advances and future trends. Trends Food Sci Technol 144:104344. https://doi.org/10.1016/j.tifs.2024.104344
34. Derra M, Bakkali F, Amghar A, Sahsah H (2018) Estimation of coagulation time in cheese manufacture using an ultrasonic pulse-echo technique. J Food Eng 216:65–71. https://doi.org/10.1016/j.jfoodeng.2017.08.003
35. Ebrahimpour M, Yu W, Young B (2021) Artificial neural network modelling for cream cheese fermentation pH prediction at lab and industrial scales. Food Bioprod Process 126:81–89. https://doi.org/10.1016/j.fbp.2020.12.006
36. Elangovan P, Dhurairajan V, Nath MK, Yogarajah P, Condell J (2024) A novel approach for meat quality assessment using an ensemble of compact convolutional neural networks. Appl Sci 14(14):5979. https://doi.org/10.3390/app14145979
37. European Parliament and Council of the European Union (2011) Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers
38. Everard CD, O’Callaghan DJ, Fagan CC, O’Donnell CP, Castillo M, Payne FA (2007) Computer vision and color measurement techniques for inline monitoring of cheese curd syneresis. J Dairy Sci 90(7):3162–3170. https://doi.org/10.3168/jds.2006-872
39. Fagnani R, Damião BCM, Trentin RPS, Vanot RL (2022) Predicting adulteration of grated Parmigiano Reggiano cheese with Ricotta using electrophoresis, multivariate nonlinear regression and computational intelligence methods. Int J Dairy Technol 75(1):239–245. https://doi.org/10.1111/1471-0307.12818
40. Fan S, Liang X, Huang W, Jialong Zhang V, Pang Q, He X, Li L, Zhang C (2022) Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOV4 network. Comput Electron Agric 193:106715. https://doi.org/10.1016/j.compag.2022.106715
41. FDA U (2004) PAT-A framework for innovative pharmaceutical development, manufacturing and quality assurance. Guidance for Industry
42. Fox P, McSweeney P (2004) Cheese: an overview. Cheese: chemistry, physics and microbiology, vol 1. Elsevier, pp 1–18. https://doi.org/10.1016/S1874-558X(04)80060-5
43. Fox PF, Cogan TM (2004) Factors that affect the quality of cheese. In: Fox PF, McSweeney PLH, Cogan TM, Guinee TP (eds) Cheese: chemistry, physics and microbiology, vol 1 of General Aspects, pp 583–608. Academic Press. https://doi.org/10.1016/S1874-558X(04)80084-8
44. Fox PF, Guinee TP, Cogan TM, McSweeney PLH (2017) Enzymatic coagulation of milk. In: Fox PF, Guinee TP, Cogan TM, McSweeney PLH (eds) Fundamentals of cheese science. Springer US, Boston, MA, pp 185–229. https://doi.org/10.1007/978-1-4899-7681-9_7
45. Fox PF, Guinee TP, Cogan TM, McSweeney PLH (2017) Starter cultures. In: Fox PF, Guinee TP, Cogan TM, McSweeney PLH (eds) Fundamentals of cheese science. Springer US, Boston, MA, pp 121–183. https://doi.org/10.1007/978-1-4899-7681-9_6
46. Freire P, Freire D, Licon CC (2024) A comprehensive review of machine learning and its application to dairy products. Crit Rev Food Sci Nutr, 1–16. https://doi.org/10.1080/10408398.2024.2312537
47. Fröhlich-Wyder MT, Bachmann HP, Schmidt RS (2023) Classification of cheese varieties from Switzerland using machine learning methods: free volatile carboxylic acids. LWT 184:115095. https://doi.org/10.1016/j.lwt.2023.115095
48. García-Esteban JA, Curto B, Moreno V, González-Martín I, Revilla I, Vivar-Quintana A (2018) A digitalization strategy for quality control in food industry based on Artificial Intelligence techniques. In: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), pp 221–226. ISSN: 2378-363X

49. Ghobakhloo M (2020) Industry 4.0, digitization, and opportunities for sustainability. J Clean Prod 252:119869. https://doi.org/10.1016/j.jclepro.2019.119869
50. Golzarijalal M, Ong L, Neoh CR, Harvie DJE, Gras SL (2024) Machine learning for the prediction of proteolysis in Mozzarella and Cheddar cheese. Food Bioprod Process 144:132–144. https://doi.org/10.1016/j.fbp.2024.01.009
51. Gonzalez Viejo C, Fuentes S, Godbole A, Widdicombe B, Unnithan RR (2020) Development of a low-cost e-nose to assess aroma profiles: an artificial intelligence application to assess beer quality. Sens Actuators B: Chem 308:127688. https://doi.org/10.1016/j.snb.2020.127688
52. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press. Google-Books-ID: omivDQAAQBAJ
53. Griffin S, Magro M, Farrugia J, Falzon O, Camilleri K, Valdramidis VP (2021) Towards the development of a sterile model cheese for assessing the potential of hyperspectral imaging as a non-destructive fungal detection method. J Food Eng 306:110639. https://doi.org/10.1016/j.jfoodeng.2021.110639
54. Géron A (2022) Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts, tools, and techniques to build intelligent systems. "O'Reilly Media, Inc.". Google-Books-ID: X5ySEAAAQBAJ
55. Guo S, Yu W, Wilson DI, Young BR (2023) pH prediction for a semi-batch cream cheese fermentation using a grey-box model. Chem Prod Process Model 18(4):581–590. https://doi.org/10.1515/cppm-2021-0048
56. He X, Chen Y, Ghamisi P (2020) Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network. IEEE Trans Geosci Remote Sens 58(5):3246–3263. https://doi.org/10.1109/TGRS.2019.2951445
57. High R, Eyres GT, Bremer P, Kebede B (2021) Characterization of blue cheese volatiles using fingerprinting, self-organizing maps, and entropy-based feature selection. Food Chem 347:128955. https://doi.org/10.1016/j.foodchem.2020.128955
58. High R, Kebede B, Eyres G, Bremer P (2022) Modelling the non-linear development of Shenley Station blue cheese volatiles during ripening using untargeted volatile fingerprinting and self-organizing maps. Food Chem Adv 1:100047. https://doi.org/10.1016/j.focha.2022.100047
59. Jafarzadeh S, Salehabadi A, Mohammadi Nafchi A, Oladzadabbasabadi N, Jafari SM (2021) Cheese packaging by edible coatings and biodegradable nanocomposites; improvement in shelf life, physicochemical and sensory properties. Trends Food Sci Technol 116:218–231. https://doi.org/10.1016/j.tifs.2021.07.021
60. James G, Witten D, Hastie T, Tibshirani R, Taylor J (2023) Resampling methods. In: James G, Witten D, Hastie T, Tibshirani R, Taylor J (eds) An introduction to statistical learning: with applications in python. Springer International Publishing, Cham, pp 201–228. https://doi.org/10.1007/978-3-031-38747-0_5
61. Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31(3):685–695. https://doi.org/10.1007/s12525-021-00475-2
62. Johnson ME (2017) *A 100-year review*: cheese production and quality. J Dairy Sci 100(12):9952–9965. https://doi.org/10.3168/jds.2017-12979
63. Jox D, Borsum C, Hummel D, Hinrichs J, Krupitzer C (2025) Enhancing dairy processing with machine learning and domain knowledge: a combined analysis of offline and time series data. J Food Eng 391:112423. https://doi.org/10.1016/j.jfoodeng.2024.112423
64. Kamilari E, Tsaltas D, Stanton C, Ross RP (2022) Metataxonomic mapping of the microbial diversity of irish and eastern mediterranean cheeses. Foods 11(16):2483. https://doi.org/10.3390/foods11162483
65. Khanam R, Hussain M (2024) YOLOv11: an overview of the key architectural enhancements. arXiv:2410.17725
66. Khattab AR, Guirguis HA, Tawfik SM, Farag MA (2019) Cheese ripening: a review on modern technologies towards flavor enhancement, process acceleration and improved quality assessment. Trends Food Sci Technol 88:343–360. https://doi.org/10.1016/j.tifs.2019.03.009
67. Kousta M, Mataragas M, Skandamis P, Drosinos EH (2010) Prevalence and sources of cheese contamination with pathogens at farm and processing levels. Food Control 21(6):805–815. https://doi.org/10.1016/j.foodcont.2009.11.015
68. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, vol 25. Curran Associates, Inc
69. Krogh A (2008) What are artificial neural networks? Nat Biotechnol 26(2):195–197. https://doi.org/10.1038/nbt1386
70. Lai C, Xie M, Murthy D (2003) A modified Weibull distribution. IEEE Trans Reliab 52(1):33–37. https://doi.org/10.1109/TR.2002.805788
71. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
72. Lee Y, Kim YS, Di Lee, Jeong S, Kang GH, Jang YS, Kim W, Choi HY, Kim JG, Choi Sh (2022) The application of a deep learning system developed to reduce the time for RT-PCR in COVID-19 detection. Sci Rep 12(1):1234. https://doi.org/10.1038/s41598-022-05069-2
73. Legg AK, Carr AJ, Bennett RJ, Johnston KA (2017) Chapter 26 - general aspects of cheese technology. In: McSweeney PLH, Fox PF, Cotter PD, Everett DW (eds) Cheese, 4th edn. Academic Press, San Diego, pp 643–675. https://doi.org/10.1016/B978-0-12-417012-4.00026-0
74. Li B, Lin Y, Yu W, Wilson DI, Young BR (2021) Application of mechanistic modelling and machine learning for cream cheese fermentation pH prediction. J Chem Technol Biotechnol 96(1):125–133. https://doi.org/10.1002/jctb.6517
75. Liznerski P, Ruff L, Vandermeulen RA, Franks BJ, Kloft M, Müller KR (2021) Explainable deep one-class classification. arXiv:2007.01760
76. Loddo A, Di Ruberto C, Armano G, Manconi A (2022) Automatic monitoring cheese ripeness using computer vision and artificial intelligence. IEEE Access 10:122612–122626. https://doi.org/10.1109/ACCESS.2022.3223710
77. Loddo A, Ghiani D, Perniciano A, Zedda L, Pes B, Di Ruberto C (2024) An anomaly detection approach to determine optimal cutting time in cheese formation. Information 15(6):360. https://doi.org/10.3390/info15060360
78. Loddo A, Ruberto CD, Armano G, Manconi A (2025) Detecting coagulation time in cheese making by means of computer vision and machine learning techniques. Comput Ind 164:104173
79. Lukinac JL, Jukić MJ, Mastanjević KM, Lučan ML (2018) Application of computer vision and image analysis method in cheese-quality evaluation: a review. Ukrainian Food J 7(2):192–214. https://doi.org/10.24263/2304-974X-2018-7-2-4
80. Lundberg S, Lee SI (2017) A unified approach to interpreting model predictions. arXiv:1705.07874
81. Magarelli M, Novielli P, De Filippis F, Magliulo R, Di Bitonto P, Diacono D, Bellotti R, Tangaro S (2024) Explainable artificial intelligence and microbiome data for food geographical origin: the Mozzarella di Bufala Campana PDO Case of Study. Front Microbiol 15:1393243. https://doi.org/10.3389/fmicb.2024.1393243
82. Mahesh B (2020) Machine learning algorithms-a review. Int J Sci Res (IJSR).[Internet] 9(1):381–386
83. Małkowska-Kowalczyk M, Żulewska J, Kruk D, Mieloch A (2024) Dynamic water profile in various types of cheese analyzed

by means of nuclear magnetic resonance relaxometry. J Dairy Sci 107(10):7691–7703. https://doi.org/10.3168/jds.2024-24793
84. Martín Tornero E, Durán Merás I, Alcaraz MR, Muñoz de la Peña A, Galeano Díaz T, Goicoechea HC (2024) Applying multivariate curve resolution modelling combined with discriminant tools on near-infrared spectra for distinguishing between cheese varieties and stages of ripening. Microchem J, 111039. https://doi.org/10.1016/j.microc.2024.111039
85. Maslej N, Fattorini L, Perrault R, Parli V, Reuel A, Brynjolfsson E, Etchemendy J, Ligett K, Lyons T, Manyika J, Niebles JC, Shoham Y, Wald R, Clark J (2024) The AI index 2024 annual report. Technical report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA
86. Mavani NR, Ali JM, Othman S, Hussain MA, Hashim H, Rahman NA (2021) Application of artificial intelligence in food industry—a guideline. Food Eng Rev 14(1):134. https://doi.org/10.1007/s12393-021-09290-z
87. McSweeney PLH (2004) Biochemistry of cheese ripening. Int J Dairy Technol 57(2–3):127–144. https://doi.org/10.1111/j.1471-0307.2004.00147.x
88. Menevseoglu A, Gunes N, Ayvaz H, Sarikaya SB, Zehiroglu C (2023) Near- and mid-infrared spectroscopy combined with machine learning algorithms to determine minerals and antioxidant activity in commercial cheese. Turkish J Agri - Food Sci Technol 11(12):2435–2445. https://doi.org/10.24925/turjaf.v11i12.2435-2445.6526
89. Mengucci C, Rabiti D, Urbinati E, Picone G, Romano R, Aiello A, Ferranti P, Capozzi F (2021) Spotting frozen curd in PDO buffalo mozzarella cheese through insights on its supramolecular structure acquired by 1H TD-NMR relaxation experiments. Appl Sci 11(4):1466. https://doi.org/10.3390/app11041466
90. Mitchell TM (1997) Machine learning (Nachdr. ed.). McGraw-Hill series in Computer Science. McGraw-Hill, New York
91. Münch M, Guillard V, Gaucel S, Destercke S, Thévenot J, Buche P (2023) Composition-based statistical model for predicting CO2 solubility in modified atmosphere packaging application. J Food Eng 340:111283. https://doi.org/10.1016/j.jfoodeng.2022.111283
92. Munch M, Wuillemin PH, Dibie-Barthelemy J, Manfredotti C, Allard T, Buchin S, Guichard E (2018) Identifying control parameters in cheese fabrication process using precedence constraints. In: Soldatova L, Vanschoren J, Papadopoulos G, Ceci M (eds) 21st International Conference on Discovery Science (DS 2018), vol 11198 of Discovery Science, Limassol, Cyprus. Springer, pp 421–434
93. Nareklishvili M, Geitle M (2024) Deep ensemble transformers for dimensionality reduction. IEEE Trans Neural Netw Learn Syst, 1–12. https://doi.org/10.1109/TNNLS.2024.3357621
94. Nicolau N, Buffa M, O'Callaghan DJ, Guamis B, Castillo M (2015) Estimation of clotting and cutting times in sheep cheese manufacture using NIR light backscatter. Dairy Sci Technol 95(4):495–507. https://doi.org/10.1007/s13594-015-0232-7
95. OECD/FAO (2024) OECD-FAO Agricultural Outlook 2024-2033. OECD Publishing, Paris and Rome
96. Park JS, Ha JW (2019) X-ray irradiation inactivation of *Escherichia coli* O157:H7, *Salmonella enterica* Serovar Typhimurium, and *Listeria monocytogenes* on sliced cheese and its bactericidal mechanisms. Int J Food Microbiol 289:127–133. https://doi.org/10.1016/j.ijfoodmicro.2018.09.011
97. Pereira PC (2014) Milk nutritional composition and its role in human health. Nutrition (Burbank, Los Angeles County, Calif.) 30(6):619–627. https://doi.org/10.1016/j.nut.2013.10.011
98. Perrignon M, Emily M, Munch M, Jeantet R, Croguennec T (2025) Machine learning for predicting industrial performance: example of the dry matter content of emmental-type cheese. Int Dairy J 162:106143. https://doi.org/10.1016/j.idairyj.2024.106143
99. Priyashantha H, Höjer A, Saedén KH, Lundh A, Johansson M, Bernes G, Geladi P, Hetta M (2021) Determining the end-date of long-ripening cheese maturation using NIR hyperspectral image modelling: a feasibility study. Food Control 130:108316. https://doi.org/10.1016/j.foodcont.2021.108316
100. Rocha RS, Calvalcanti RN, Silva R, Guimarães JT, Balthazar CF, Pimentel TC, Esmerino EA, Freitas MQ, Granato D, Costa RGB, Silva MC, Cruz AG (2020) Consumer acceptance and sensory drivers of liking of Minas Frescal Minas cheese manufactured using milk subjected to ohmic heating: performance of machine learning methods. LWT 126:109342. https://doi.org/10.1016/j.lwt.2020.109342
101. Roh Y, Heo G, Whang SE (2021) A survey on data collection for machine learning: a big data - AI integration perspective. IEEE Trans Knowl Data Eng 33(4):1328–1347. https://doi.org/10.1109/TKDE.2019.2946162
102. Samuel AL (1959) Some Studies in machine learning using the game of checkers. IBM J Res Dev 3(3):210–229. https://doi.org/10.1147/rd.33.0210
103. Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SN Comp Sci 2(3):160. https://doi.org/10.1007/s42979-021-00592-x
104. Schroeder V, Evans ED, Wu YCM, Voll CCA, McDonald BR, Savagatrup S, Swager TM (2019) Chemiresistive sensor array and machine learning classification of food. ACS Sensors 4(8):2101–2108. https://doi.org/10.1021/acssensors.9b00825
105. Seratlic S, Guha B, Moore S (2024) Advances in spectroscopic methods for predicting cheddar cheese maturity: a review of FT-IR, NIR, and NMR techniques. NDT 2(4):392–416. https://doi.org/10.3390/ndt2040024
106. Singh P, Pandey S, Manik S (2024) A comprehensive review of the dairy pasteurization process using machine learning models. Food Control 164:110574. https://doi.org/10.1016/j.foodcont.2024.110574
107. Sánchez-González JA, Oblitas-Cruz JF, Sánchez-González JA, Oblitas-Cruz JF (2017) Application of Weibull analysis and artificial neural networks to predict the useful life of the vacuum packed soft cheese. Revista Facultad de Ingeniería Universidad de Antioquia (82): 53–59. https://doi.org/10.17533/udea.redin.n82a07
108. Thapa A, Nishad S, Biswas D, Roy S (2023) A comprehensive review on artificial intelligence assisted technologies in food industry. Food Biosci 56:103231. https://doi.org/10.1016/j.fbio.2023.103231
109. Torres-Vitela MR, Mendoza-Bernardo M, Castro-Rosas J, Gomez-Aldapa CA, Garay-Martinez LE, Navarro-Hidalgo V, Villarruel-López A (2012) Incidence of *Salmonella, Listeria monocytogenes, Escherichia coli* O157:H7, and staphylococcal enterotoxin in two types of mexican fresh cheeses. J Food Prot 75(1):79–84. https://doi.org/10.4315/0362-028x_jfp-11-258
110. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2023) Attention is all you need. arXiv:1706.03762
111. Vega-Chinchay EM, Jimenez-Bustamante JN, Bustamante-Bustamante F, Caro-Degollar EM (2024) Image-based analysis of the bactericidal effect of lactobacillus acidophilus application in cheeses made from raw milk. Chem Eng Trans 110:475–480. https://doi.org/10.3303/CET24110080
112. Villaquiran Z, Zamora A, Arango O, Castillo M (2024) Inline determination of the gel elastic modulus during milk coagulation using a multifiber optical probe. Food Bioprocess Technol 17(10):3149–3161. https://doi.org/10.1007/s11947-023-03294-9
113. Wang J, Guo X (2024) The Gompertz model and its applications in microbial growth and bioproduction kinetics: past, present

and future. Biotechnol Adv 72:108335. https://doi.org/10.1016/j.biotechadv.2024.108335
114. Wirth R, Hipp J (2000) CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, vol 1. Manchester, pp 29–39
115. Wu J, Wu CC, Liao CS (2021) Novel lactobacillus fermentation prediction using deep learning. In: 2021 7th International Conference on Applied System Innovation (ICASI), pp 54–57. ISSN: 2768-4156
116. Wu X, Jia W (2024) Multimodal deep learning as a next challenge in nutrition research: tailoring fermented dairy products based on cytidine diphosphate-diacylglycerol synthase-mediated lipid metabolism. Crit Rev Food Sci Nutr 64(33):12272–12283. https://doi.org/10.1080/10408398.2023.2248633
117. Yakoubi S (2024) Enhancing plant-based cheese formulation through molecular docking and dynamic simulation of tocopherol and retinol complexes with zein, soy and almond proteins via SVM-machine learning integration. Food Chem 452:139520. https://doi.org/10.1016/j.foodchem.2024.139520
118. Yang F, Sun J, Cheng J, Fu L, Wang S, Xu M (2023) Detection of starch in minced chicken meat based on hyperspectral imaging technique and transfer learning. J Food Process Eng 46(4):e14304. https://doi.org/10.1111/jfpe.14304
119. Yang H, Yu J, Jin L, Zhao Y, Gao Q, Shi C, Ye L, Li D, Yu H, Xu Y (2023) A deep learning based method for automatic analysis of high-throughput droplet digital PCR images. Analyst 148(2):239–247. https://doi.org/10.1039/D2AN01631A
120. Yoon Y, Lee S, Choi KH (2016) Microbial benefits and risks of raw milk cheese. Food Control 63:201–215. https://doi.org/10.1016/j.foodcont.2015.11.013
121. Zedda L, Perniciano A, Loddo A, Di Ruberto C (2024) Understanding cheese ripeness: an artificial intelligence-based approach for hierarchical classification. Knowl-Based Syst 295:111833. https://doi.org/10.1016/j.knosys.2024.111833
122. Zhang D, Jia C, Sun D, Gao C, Fu D, Cai P, Hu QN (2023) Data-driven prediction of molecular biotransformations in food fermentation. J Agric Food Chem 71(22):8488–8496. https://doi.org/10.1021/acs.jafc.3c01172
123. Zhang ZY, Jiang MQ, Xiong HM (2023) Optimized identification of cheese products based on Raman spectroscopy and an extreme learning machine. New J Chem 47(14):6889–6894. https://doi.org/10.1039/D2NJ06350F
124. Jia Z, Lin Z, Luo Y, Cardoso Z, Wang D, Flock G, Thompson-Witrick KA, Hengyong Y, Zhang B (2024) Enhancing pathogen identification in cheese with high background microflora using an artificial neural network-enabled paper chromogenic array sensor approach. Sens Actuators B: Chem 410:135675. https://doi.org/10.1016/j.snb.2024.135675
125. Zheng X, Shi X, Wang B (2021) A review on the general cheese processing technology, flavor biochemical pathways and the influence of yeasts in cheese. Front Microbiol 12:703284. https://doi.org/10.3389/fmicb.2021.703284
126. Zlatev Z, Spahiu T, Taneva I, Dimov M, Vasilev M (2024) Integration of sensor fusion to enhance quality assessment of white brine cheeses. Dairy 5(2):249–270. https://doi.org/10.3390/dairy5020021