information

Article

An Anomaly Detection Approach to Determine Optimal Cutting Time in Cheese Formation

Andrea Loddo *,†, Davide Ghiani †, Alessandra Perniciano, Luca Zedda, Barbara Pes and Cecilia Di Ruberto

Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy; d.ghiani6@studenti.unica.it (D.G.); alessandra.pernician@unica.it (A.P.); luca.zedda@unica.it (L.Z.); pes@unica.it (B.P.); cecilia.dir@unica.it (C.D.R.)

* Correspondence: andrea.loddo@unica.it

† These authors contributed equally to this work.

Abstract: The production of cheese, a beloved culinary delight worldwide, faces challenges in maintaining consistent product quality and operational efficiency. One crucial stage in this process is determining the precise cutting time during curd formation, which significantly impacts the quality of the cheese. Misjudging this timing can lead to the production of inferior products, harming a company's reputation and revenue. Conventional methods often fall short of accurately assessing variations in coagulation conditions due to the inherent potential for human error. To address this issue, we propose an anomaly-detection-based approach. In this approach, we treat the class representing curd formation as the anomaly to be identified. Our proposed solution involves utilizing a one-class, fully convolutional data description network, which we compared against several state-of-the-art methods to detect deviations from the standard coagulation patterns. Encouragingly, our results show F1 scores of up to 0.92, indicating the effectiveness of our approach.

Keywords: image processing; computer vision; machine learning; food industry; curd-firming time detection

check for updates

Citation: Loddo, A.; Ghiani, D.; Perniciano, A.; Zedda, L.; Pes, B.; Di Ruberto, C. An Anomaly Detection Approach to Determine Optimal Cutting Time in Cheese Formation. *Information* **2024**, *15*, 360. https://doi.org/10.3390/info15060360

Academic Editor: Zahir M. Hussain

Received: 29 April 2024
Revised: 13 June 2024
Accepted: 14 June 2024
Published: 18 June 2024

1. Introduction

Dairy products possess intrinsic qualities that enhance gastrointestinal tract health and contribute to the well-being of the human microbiome. They play a pivotal role in the food industry due to their rich content of protein, calcium, and micronutrients, all of which are crucial for maintaining bone and muscle health. Among dairy products, cheese stands out as one of the most widely consumed and versatile options globally. With its diverse array of flavors and forms, cheese holds a prominent position in culinary culture, contributing significantly to dietary diversity and enjoyment.

Cheese is a fundamental ingredient in numerous culinary recipes and is often enjoyed on its own. Consequently, evaluating its quality becomes essential for consumers and the industry [1].

In the cheese-manufacturing process, curd represents a crucial intermediate stage achieved by heating milk and introducing rennet. Rennet induces the coagulation of casein granules in the milk, resulting in the formation of curd, which settles at the bottom, accompanied by the generation of whey. However, in many cheese varieties, the natural separation of whey and curd does not occur spontaneously, necessitating the mechanical cutting of the coagulated mass into small cubes, referred to as curd grains [2].

As demonstrated by Johnson et al. [3], the coagulation process induced by rennet during cheese production, and consequently, the timing of curd cutting, significantly impacts cheese quality. Furthermore, Grundelius et al. [4] investigated the influence of parameters such as pH, rennet concentration, and curd granule size, highlighting that granule size significantly impacts curd shrinkage, particularly in the early stages of the process: smaller curd granules result in more intense whey separation, a finding confirmed by Thomann et al. [5]. The duration of cutting is known to be inversely related to the size of the granules. Therefore, to automate the process, it is essential to identify the phase where the granules start to become diffuse throughout the entire boiler [4].

Determining the cutting time is contingent upon the rheological and microstructural properties of the curd gels, which are influenced by various factors, including milk pretreatment, composition, and coagulation conditions. Consequently, the identification of the cutting time, manually performed by a diary operator, varies across different cheese varieties and profoundly affects parameters such as moisture content, yield, and the overall quality of the cheese, as well as losses in whey fat [2].

This challenge is notably accentuated in large-scale automated production facilities, where the variability in coagulation conditions, process alterations, and the potential for human errors introduce complexities in maintaining precise control over cutting times [2,6,7].

For these reasons, integrating advanced methodologies, such as computer vision (CV) techniques, become indispensable to mitigate the issues, improve the cheese-making process, enhance production efficiency, and optimize product quality.

In light of the challenging task posed by identifying the optimal cutting time in cheese production, we approach it through the lens of anomaly detection (AD). Given the abundance of images depicting the normal condition of the milk before its cutting time, we adopted an AD setup to discern anomalies within this dataset. Since an extensive presence of curd spots indicates a possible optimal cutting time [2], we considered the curd spot an anomaly, seeking to identify it amidst the distribution of normal curd images. By leveraging this approach, we aim to effectively identify deviations from the standard milk appearance, thereby facilitating the accurate determination of the curd and related cutting time.

Specifically, we propose adapting a deep learning (DL) technique belonging to the realm of one-class classification, termed the Fully Convolutional Data Description Network (FCDDN). This method employs a neural network to reconfigure the data such that normal instances are centered on a predefined focal point, while anomalous instances are situated elsewhere. Additionally, a sampling technique transforms the data into images representing a heatmap of subsampled anomalies. Pixels in this heatmap distant from the center correspond to anomalous regions within the input image. The FCDDN exclusively utilizes convolutional and pooling layers, thereby constraining the receptive field of each output pixel [8]. Moreover, we also compared our findings with more classical machine learning (ML) approaches, specifically trained with handcrafted (HC) and deep features. The latter were extracted by pre-trained convolutional neural network (CNN) architectures.

The contributions of this paper can be summarized as follows:

- Investigate the optimal cutting time: We conducted a feasibility study by introducing a novel AD-based approach to determine the optimal cutting time during curd formation in cheese production.
- Development of a one-class Fully Convolutional Data Description Network: We propose and implemented a one-class FCDDN to identify curd formation by treating it as an anomaly to verify against the milk in its usual state.
- Comparison with shallow AD methods: We compared the proposed approach with shallow learning methods to emphasize its robustness in this scenario on different sets of images.
- High accuracy in AD: The proposed approach achieved encouraging results with F1 scores of up to 0.92, demonstrating the effectiveness of the method.
- Application in the dairy industry: This work investigates if the curd-firming time identification can be achieved with an AD-based approach and, at the same, aims to provide a non-invasive, non-destructive, and technologically advanced solution.

The rest of the manuscript is organized as follows. Section 2 provides a comprehensive review of existing methodologies for analyzing milk coagulation and AD techniques. Section 3 elucidates the details regarding the dataset, feature-extraction methodologies, classifiers adopted, and evaluation measures. Section 4 delves into the experimental evaluation conducted, offering a presentation of the undertaken experiments along with the corresponding results and subsequent discussions. The concluding remarks of this study, along with insightful suggestions for potential enhancements and avenues for future research based on our findings, are given in Section 5.

2. Related Work

This section gives an overview of the existing automated methods in the dairy industry (Section 2.1) and the AD techniques (Section 2.2).

2.1. Automated Methods in Dairy Industry

The analysis of milk-related products often demands specialized acquisition techniques such as fluorescence spectroscopy, an effective spectral approach for determining the intensity of fluorescent components in cheese, and near-infrared spectroscopy, aimed at developing non-destructive assessment methods [1,9].

CV methods necessitate the utilization of cameras and controlled illumination setups to capture images of dairy products. These techniques exhibit versatility, extending beyond the assessment of optimal cutting times [10]. Combining CV methods with artificial intelligence (AI) has found application in various tasks, including the classification of cheese ripeness in entire cheese wheels [11] and the inspection and grading of cheese meltability [12].

Moreover, beyond production processes, the integration of AI in the analysis of dairy products extends to estimating the shelf life of such products [13,14].

However, recent advancements, driven by the scarcity of dairy-related data, aim to tackle challenging tasks such as detecting adulteration or identifying rare product-compromising events through the utilization of AD approaches [15,16].

While existing studies have primarily concentrated on various methodologies for monitoring the coagulation of milk and assessing cheese quality, the proposed research endeavor introduces a novel and non-invasive AD-based approach explicitly tailored for automating the detection of optimal cutting time during cheese formation, from images. Prior investigations have delved into diverse techniques, such as electrical, thermal, optical, and ultrasonic methods; however, our proposed methodology harnesses the potent capabilities of CV and explainable DL. The innovation lies in the simplicity of the setup, requiring only a camera connected to a computer, thereby offering a practical and accessible solution for the food industry. In contrast to studies that primarily address cheese quality parameters or maturation stages, our work is exclusively focused on the critical phase of cheese production. This distinction positions our methodology as a pioneering contribution to the field, addressing a crucial aspect of cheese production that has been less explored in the existing literature.

2.2. Anomaly Detection

AD, the identification of patterns or instances that do not conform to expected behavior, has garnered significant attention across various domains due to its critical importance in identifying potential threats, faults, or outliers, with several key methodologies and approaches [8,17,18]. In this subsection, we distinguish the main proposed approaches into three different categories: statistical methods (Section 2.2.1), machine learning-based (Section 2.2.2), and deep learning-based (Section 2.2.3).

2.2.1. Statistical Methods

These form the foundation of many AD techniques. One of the earliest approaches is based on statistical properties such as the mean, variance, and probability distributions. Techniques like the Z-score, Grubbs', and Dixon's Q-test utilize statistical thresholds to identify outliers [19–21]. However, these methods often assume normality and may not effectively capture complex patterns in high-dimensional data.

2.2.2. Machine Learning-Based Methods

ML techniques have become increasingly prominent in AD due to their capability to manage complex data patterns [22,23]:

Clustering-based methods and density estimation: Techniques such as k-means and DBSCAN for clustering and Gaussian Mixture Models for density estimation are commonly used for AD without requiring labeled data. These methods detect outliers by identifying deviations from normal data distributions. However, they can struggle with high-dimensional or sparse data and are sensitive to parameter settings [24,25].

Unsupervised learning techniques: When labeled data are available, modifications of classic ML algorithms are employed for AD. Notable examples include one-class SVM (OCSVM) [26] and Isolation Forest (IF) [27,28], which are adaptations of Support Vector Machines (SVMs) and Random Forest (RF), respectively.

Ensemble and hybrid approaches: Ensemble methods enhance AD performance and robustness by combining multiple algorithms. Techniques such as IF [27,28] and the Local Outlier Factor [29,30] utilize ensemble principles, aggregating results from several base learners to identify anomalies. Hybrid approaches, which integrate various AD techniques, further improve detection accuracy and reliability by leveraging the strengths of each method [31]. In industrial applications, hybrid methods involving both ML and DL techniques have been proposed. For instance, Wang et al. [32] introduced a loss switching fusion network that combines spatiotemporal descriptors, applying it as an AD approach for classifying background and foreground motions in outdoor scenes.

2.2.3. Deep Learning-Based Methods

DL techniques, particularly Artificial Neural Networks (ANNs) and their customizations provided throughout the last decade, process raw input data and independently learn relevant feature representations [33]. This capability enables ANNs to outperform traditional methods that rely on manually created rules or advanced feature-engineering techniques [34], even in the context of AD [35]. An overview of the main architectures is provided below:

Autoencoder-based architectures: Autoencoders, including Variational Autoencoders (VAEs) [36], are a popular choice for AD in CV. VAEs learn to encode input data into a compact latent representation and then reconstruct the data from this representation. Anomalies are detected based on the reconstruction error, as anomalous data typically result in higher reconstruction errors compared to normal data [37,38].

Generative Adversarial Networks (GANs): In a typical GAN setup, a generator network creates synthetic data, while a discriminator network attempts to distinguish between real and generated data. For AD, the generator learns to produce data that mimic the normal data distribution. Anomalies can then be identified based on how well the discriminator distinguishes the actual data from the generated data. High discriminator scores indicate potential anomalies, as the generated data fail to accurately represent these outliers. GANs have also been successfully applied to AD tasks [39,40].

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks: for sequential data, such as video frames or time series, RNNs and LSTM networks are particularly effective due to their ability to capture temporal dependencies [41,42]. These networks maintain a memory of previous inputs, allowing them to understand context over time. In the context of AD, RNNs and LSTMs can model the normal sequence of events or patterns. Anomalies are detected when the predicted sequence deviates significantly from the actual observed sequence [43].

Convolutional neural networks: CNNs are widely used, even in AD, for their powerful feature-extraction capabilities from image data [44]. By learning hierarchical feature representations, CNNs can detect subtle anomalies in visual data that may not be apparent to traditional methods. In AD, CNNs are often combined with other architectures, such as autoencoders or GANs, to enhance detection accuracy [45].

Attention mechanisms and transformers: Attention mechanisms and transformer models, initially proposed for natural language processing tasks, have been adapted for CV and AD. These models can focus on relevant parts of the input data, improving the detection of anomalies in complex scenes [46]. Transformers, with their self-attention layers, have shown remarkable success in modeling dependencies and identifying anomalies in high-dimensional data [47].

Self-supervised and unsupervised learning: DL methods for AD often rely on self-supervised [48,49] and unsupervised [34,50] learning approaches, where the model learns useful representations without requiring labeled data. Techniques such as contrastive learning and pretext tasks enable the model to learn discriminative features that are effective for identifying anomalies, for example in scenarios where labeled anomalous data are scarce or unavailable.

Hybrid models: Recent advancements have explored hybrid models that combine multiple DL architectures to leverage their individual strengths [50,51]. For instance, combining CNNs with LSTMs allows the model to capture both spatial and temporal features, improving the robustness [52,53]. Similarly, integrating VAEs with GANs can enhance the model's ability to generate realistic data and detect anomalies based on reconstruction errors and adversarial loss, particularly on time series data [54].

3. Materials and Methods

In this section, we first provide a description of the dataset analyzed (Section 3.1). Following this, we elaborate on the methodology employed for the experiments conducted. Section 3.2 offers a comprehensive explanation of the various features utilized in our research, categorized into two main groups: handcrafted (HC) features and deep features. The HC features were extracted from the images using established algorithms (Section 3.2.1), while the deep features were derived from the activations of convolutional neural networks (CNNs) (Section 3.2.2). Additionally, Section 3.3 details the classifiers used in our study. Finally, Section 3.4 outlines the performance measures applied to evaluate the classification results.

3.1. Dataset

The dataset was assembled by collecting images from a dairy company based in Sardinia, Italy. The image-acquisition process involved using a Nikon D750 camera (Tokyo, Japan) equipped with a CMOS sensor measuring 35.9 × 24.0 mm and a resolution of 24 megapixels. All images were in RGB format, with resolutions of 6016 × 4016 pixels.

This consisted of 12 distinct sets of images. Each one documents the coagulation process as milk transforms from its initial liquid state to the curd stage. More specifically, every set contains two distinct classes: one with images representing the normal curd condition, identified as a **non-target** (i.e., the normal instances), and one with images representing the optimal moments for cutting time, identified as the **target** (i.e., the anomaly instances).

Table 1 provides a comprehensive summary of each set, numbered from 1 to 12, including details such as the number of images and the class composition. As can be seen, the dataset has a significant class imbalance, with most images falling into the **non-target** class. Sets 1 and 2 depict the coagulation process of fresh whole sheep's milk (*Pecorino Romano*), while subsequent sets involve a blend of cow and sheep milk.

Table 1. Comprehensive dataset details: the table provides information on the 12 distinct sets of time-ordered images, denoted from 1 to 12. Each set is characterized by its number of images, representing the number of images it contains, and their distribution between the two classes.

Set	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Number of images	94	102	128	112	77	70	84	96	105	94	89	111
Non-target samples	77	90	108	89	54	45	63	72	68	59	60	77
Target samples	17	12	20	23	23	25	21	24	37	35	29	34

Figure 1 presents two sample images from set 11, highlighting the distinctions between the two classes. Additionally, we include the peak response projection of the inverse image, generated using a median filter of size 41 × 41, to comprehensively illustrate the image structure.

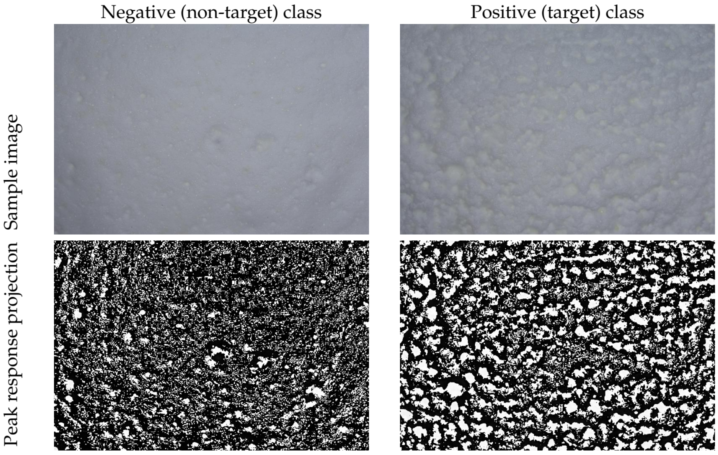

Figure 1. Sample images from set 11: on the left, a picture representing the negative (non-target) class; on the right, a picture representing the positive (target) class.

3.2. Feature Extraction

This section provides a comprehensive summary of the feature-extraction process employed to train the ML methods used in this study as a comparison against the proposed FCDDN. Section 3.2.1 illustrates the HC features, while Section 3.2.2 illustrates the deep features. This study extracted HC and deep features from the images converted to grayscale to simplify the analysis and computation process.

3.2.1. Handcrafted Features

HC features encompass diverse techniques and methodologies to extract morphological, pixel-level, and textural information from images. These features can be categorized into three primary categories: invariant moments, textural features, and color-based features [55]. Each category is briefly described below, while every parameter has been set by considering approaches in similar contexts [11].

Invariant moments: Denoted as the weighted averages of pixel intensities within an image, these are used to extract specific image properties, aiding in characterizing segmented objects. In this study, three distinct types of moments were used:

- **Chebyshev moments (CHs)**: Introduced by Mukundan and Ramakrishnan [56] and derived from Chebyshev polynomials, they were employed with both first-order (CH_1) and second-order (CH_2) moments of order 5.
- **Legendre moments (LMs)**: Initially proposed by Teague [57] and derived from Legendre orthogonal polynomials [58], they were used with second-order of order 5.
- **Zernike moments (ZMs)**: Introduced by Oujaoura et al. [59] and derived from Zernike polynomials, they were applied with order 6 and a repetition of 4.

Texture features: They focus on fine textures with different approaches. Here, the following were used:

- **Haar features (Haar)**: Consisting of adjacent rectangles with alternating positive and negative polarities, they were used in various forms, such as edge features, line features, four-rectangle features, and center-surround features. Haar features play a crucial role in cascade classifiers as part of the Viola–Jones object-detection framework [60].
- **Rotation-invariant Haralick features (HARris)**: Thirteen Haralick features [61], derived from the Gray-Level Co-occurrence Matrix (GLCM), were transformed into rotation-invariant features [62]. This transformation involved computing GLCM variations with the parameters set to $d = 1$ and angular orientations $\theta = [0°, 45°, 90°, 135°]$.
- **Local Binary Pattern (LBP)**: As described by Ojala et al. [63], the LBP characterizes texture and patterns within images. In this work, the histogram of the LBP, converted to a rotation-invariant form (LBP_ri) [64], was extracted using a neighborhood defined by a radius $r = 1$ and a number of neighbors $n = 8$.

Color features: They aim to extract color intensity information from images. In this study, these descriptors were calculated from images converted to grayscale, simplifying the analysis and computation process. More precisely, as the color histogram characterizes the global color distribution within images, seven statistical descriptors, the mean, standard deviation, smoothness, skewness, kurtosis, uniformity, and entropy, were computed from the **grayscale histogram features (Hist)** (Table 2) .

Table 2. Employed convolutional neural networks' details including reference paper, number of trainable parameters in millions, input shape, feature-extraction layer, and related feature vector size.

Ref.	Params (M)	Input Shape	Feature Layer	# of Features
AlexNet [65]	60	224 × 224	Pen. FC	4096
DarkNet-53 [66]	20.8	224 × 224	Conv53	1000
DenseNet-201 [67]	25.6	224 × 224	Avg. Pool	1920
GoogLeNet [68]	5	224 × 224	Loss3	1000
EfficientNetB0 [69]	5.3	224 × 224	Avg. Pool	1280
Inception-v3 [70]	21.8	299 × 299	Last FC	1,000
Inception-ResNet-v2 [71]	55	299 × 299	Avg. pool	1536
NasNetL [72]	88.9	331 × 331	Avg. Pool	4032
ResNet-18 [73]	11.7	224 × 224	Pool5	512
ResNet-50 [73]	26	224 × 224	Avg. Pool	1024
ResNet-101 [73]	44.6	224 × 224	Pool5	1024
VGG16 [74]	138	224 × 224	Pen. FC	4096
VGG19 [74]	144	224 × 224	Pen. FC	4096
XceptionNet [75]	22.9	299 × 299	Avg. Pool	2048

3.2.2. Deep Features

CNNs have demonstrated their effectiveness as deep feature extractors in different contexts [76–78], as well as in AD setups [79–81]. CNNs excel at capturing global features from images by processing the input through multiple convolutional filters and progressively reducing dimensionality across various architectural stages. For our experiments, we selected several architectures pre-trained on the Imagenet1k dataset [82], as presented in Table 2, along with comprehensive details regarding the selected layers for feature extraction, input size, and the number of trainable parameters for each architecture.

3.3. Classification Methods

This section presents the ML and DL classification methods employed in our analysis. As suggested in Section 2, we selected three classifiers that could cope with the data to analyze, taking into account that each set is unique and potentially requires different methods for analysis. Also, we aimed to compare three baseline methods to obtain insights into a novel dataset with limited images. For this reason, we avoided using more complex DL-based methods, such as VAEs or GANs, as we faced a relatively small-sample-size dataset.

More specifically, we selected two ML-based classifiers: the OCSVM and the IF. Finally, we chose the FCCDN from the DL-based approaches. The former two were trained with every single feature described in Section 3.2, while, being a DL approach, the latter was trained end-to-end. A brief description of these methods is now given.

3.3.1. One-Class SVM

As introduced in Section 2, this ML algorithm belongs to the family of SVMs and is primarily designed for AD in datasets where only one class, typically the majority class (normal instances), is available for training.

It aims to learn a decision boundary that encapsulates the normal instances in the feature space. This boundary is constructed in such a way that it maximizes the margin between the normal instances and the hyperplane, while minimizing the number of instances classified as outliers. Unlike traditional SVMs, which aim to find a decision boundary that separates different classes, the one-class SVM focuses solely on delineating the region of normality [83].

After training, the one-class SVM can classify new instances as either normal or anomalous based on their distance from the decision boundary. Instances within the margin defined by the support vectors are considered normal, while those outside the margin are classified as anomalies.

3.3.2. Isolation Forest

This is an ensemble-based AD algorithm that operates on the principle of isolating anomalies rather than modeling normal data points. It is particularly effective in identifying anomalies in high-dimensional datasets and is capable of handling both numerical and categorical features [84].

The main idea behind the Isolation Forest algorithm is to isolate anomalies by constructing random decision trees. Unlike traditional decision trees, which aim to partition the feature space into regions containing predominantly one class, the Isolation Forest builds trees that partition the data randomly. Specifically, each tree is constructed by recursively selecting random features and splitting the data until all instances are isolated in leaf nodes.

The anomaly score assigned to each instance is based on the average path length in the trees. Anomalies are expected to have shorter average path lengths compared to normal instances, making them distinguishable from the majority of data points. The anomaly score can be thresholded to identify outliers, with instances exceeding the threshold considered anomalies.

3.3.3. FCDD Network

This is a DL technique originally proposed by Liznerski et al. [8] for AD, particularly within the framework of one-class classification. The FCDDN employs a backbone CNN composed solely of convolutional and pooling layers devoid of fully connected layers.

At its core, the FCDDN aims to transform the input data such that normal instances are concentrated around a predetermined center while anomalous instances are situated elsewhere in the feature space. This transformation enables the model to effectively distinguish between normal and anomalous patterns in the data distribution.

One distinctive feature of the FCDDN is its utilization of a sampling method to convert data samples into images representing heatmaps of subsampled anomalies. These heatmaps visualize the anomalies present in the input data, with pixels farther from the center corresponding to anomalous regions within the original images. In the current scenario, this means a correct classification of the target class. An example is shown in Figure 2.

By employing convolutional and pooling layers exclusively, the FCDDN restricts the receptive field of each output pixel, allowing for localized feature extraction and preserving spatial information within the data. This characteristic makes the FCDDN particularly suitable for processing image data, where spatial relationships and local patterns play a crucial role in AD. In this setting, the FCDDN was trained with color images to maintain coherence with the classical CNN training scenario.

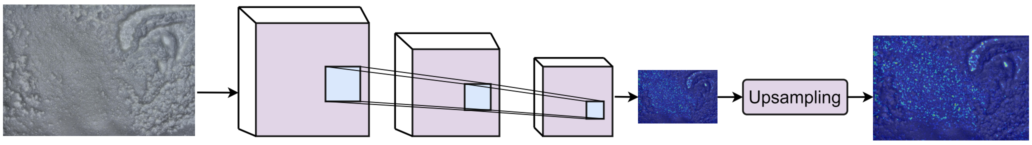

Figure 2. Schematic representation of generating full-resolution anomaly heatmaps using the FCDD approach. On the left, the original input image is taken as the input by the backbone CNN. An upsampling procedure is then applied to the scaled CNN's output, by means of a transposed Gaussian convolution, to obtain a full-size heatmap [8].

3.4. Evaluation Measures

Consider a binary classification task, where each example e is represented by a pair $\langle i, t \rangle$, with i denoting the feature values and t representing the target category. A dataset D comprises such examples. In this binary scenario, the categories are typically labeled as negativeand *positive*.

The performance evaluation of a binary classifier on dataset D involves labeling each instance as negative or positive based on the classifier's output. The evaluation is based on the following metrics:

- True negatives (TNs): instances correctly predicted as negative.
- False positives (FPs): instances incorrectly predicted as positive.
- False negatives (FNs): instances incorrectly predicted as negative.
- True positives (TPs): instances correctly predicted as positive.

These metrics lead to the following definitions:

- Precision (P): the fraction of positive instances correctly classified among all instances classified as positive:

$$P = \frac{TP}{TP + FP} \tag{1}$$

- Recall (R) (or sensitivity): measures the classifier's ability to predict the positive class against FNs (also known as the true positive rate):

$$R = \frac{TP}{TP + FN} \tag{2}$$

- F1 score (F1): the harmonic mean between precision and recall:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{3}$$

4. Experimental Results

In this section, we comprehensively explore the interpretation and implications of the results derived from our study. We structure our analysis into three distinct sections: Firstly, we present the outcomes obtained using various shallow learning classifiers with handcrafted features and deep learning-based features. Following this, we discuss the results achieved by employing an FCDDN as an AD method. However, to simplify our discussion, we report only the best-performing pairs of classifiers and features. This systematic approach provides a detailed examination of the efficacy and nuances of each method employed, highlighting valuable insights into their respective performance. Finally, we provide a global experiment result analysis.

4.1. Experimental Setup

In this research, all the selected classifiers were trained with default parameters to prevent the generalization capability of the trained models from being influenced. For the sake of brevity, we report only the results obtained with the best ML classifier based on the F1 score value obtained on the target class, which is the OCSVM. We also provide the outcomes achieved through the FCDDN.

The experiments conducted with the ML techniques have been carried out over eight HC feature categories and fourteen deep features extracted from off-the-shelf, pre-trained CNNs. Both categories are described in Section 3.2. Moreover, the classifiers were trained on the non-target class only, following the one-class classification paradigm [18].

The evaluation strategy was five-fold cross-validation. The FCDDN was composed of a ResNet-18 as the backbone. It was trained with the following splits: 50%, 10%, and 40% for training, calibration, and testing for the normal (non-target) class.Each training was repeated five times. Moreover, the training procedures were executed for a maximum of 100 epochs, with early stopping based on the calibration set performance. A batch size of 32, the Adam optimizer, and a learning rate of 0.001 were employed.

Relevant performance measures have been reported for each experiment. In particular, for both classes, we report the precision, recall, and F1 score, as described in Section 3.4.

The experiments were executed on a workstation equipped with the following hardware: an Intel(R) Core(TM) i9-8950HK @ 2.90GHz CPU, 32 GB of RAM, and an NVIDIA GTX1050 Ti GPU with 4GB of memory.

4.2. Quantitative Results

This section presents the results achieved using the best ML-based approach, the OCSVM, when trained with either HC (Section 4.2.1) or deep features (Section 4.2.2), as well as the performance of the FCDDN (Section 4.2.3). Three tables are provided to display the optimal quantitative results obtained with these approaches. Specifically, the tables present the results across different image sets from the dataset, with each set representing a distinct cheese coagulation process, which may involve varying types of milk, environmental conditions, or production batches. Table 3 highlights the best performing HC feature for each set, while Table 4 identifies the CNNs from which the most effective deep features were extracted. Finally, Table 5 shows the results obtained with the FCDDN approach.

Table 3. Results obtained with the best ML-based approaches (one-class SVM) trained on each set with HC features. Average performance over the k folds ($k = 5$) and standard deviation (the latter within round brackets) for each set and class are reported. The feature column reports the best HC feature for the specific set.

		Non-Target			Target		
Set	**Features**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
1	ZM	0.97 (0.07)	0.61 (0.08)	0.74 (0.08)	0.75 (0.02)	0.99 (0.01)	0.85 (0.02)
2	ZM	0.94 (0.03)	0.62 (0.03)	0.75 (0.03)	0.62 (0.03)	0.93 (0.02)	0.75 (0.02)
3	CH_2	0.81 (0.11)	0.76 (0.09)	0.78 (0.10)	0.75 (0.02)	0.80 (0.02)	0.77 (0.02)
4	ZM	0.88 (0.07)	0.48 (0.27)	0.62 (0.14)	0.71 (0.06)	0.95 (0.02)	0.81 (0.09)
5	Haar	1.00 (0.05)	0.11 (0.35)	0.2 (0.15)	0.71 (0.06)	1.00 (0.01)	0.83 (0.03)
6	ZM	0.40 (0.03)	0.16 (0.03)	0.21 (0.03)	0.76 (0.03)	0.94 (0.03)	0.84 (0.04)
7	ZM	1.00 (0.03)	0.38 (0.03)	0.52 (0.03)	0.73 (0.04)	1.00 (0.01)	0.84 (0.03)
8	ZM	1.00 (0.02)	0.25 (0.03)	0.40 (0.03)	0.69 (0.14)	1.00 (0.01)	0.82 (0.03)
9	Hist	0.82 (0.03)	0.70 (0.04)	0.75 (0.03)	0.90 (0.00)	0.93 (0.01)	0.91 (0.01)
10	ZM	0.69 (0.14)	0.22 (0.07)	0.30 (0.09)	0.79 (0.04)	0.98 (0.01)	0.87 (0.02)
11	Hist	1.00 (0.04)	0.30 (0.14)	0.45 (0.11)	0.78 (0.04)	1.00 (0.01)	0.87 (0.02)
12	ZM	0.40 (0.03)	0.39 (0.03)	0.39 (0.03)	0.72 (0.02)	0.72 (0.02)	0.72 (0.02)

Table 4. Results obtained with the best ML-based approaches (one-class SVM) trained on each set with deep features. Average performance over the k folds ($k = 5$) and standard deviation (the latter within round brackets) for each set and class are reported. The feature column reports the best deep feature for the specific set.

			Non-Target			**Target**	
Set	**Features**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
1	XceptionNet	1.00 (0.01)	0.47 (0.07)	0.64 (0.04)	0.67 (0.02)	1.00 (0.01)	0.81 (0.01)
2	ResNet-18	1.00 (0.01)	0.41 (0.11)	0.57 (0.09)	0.54 (0.04)	1.00 (0.01)	0.70 (0.05)
3	EfficientNetB0	1.00 (0.01)	0.56 (0.07)	0.71 (0.08)	0.68 (0.03)	1.00 (0.01)	0.81 (0.01)
4	EfficientNetB0	1.00 (0.01)	0.54 (0.04)	0.68 (0.06)	0.75 (0.05)	1.00 (0.01)	0.85 (0.01)
5	XceptionNet	0.79 (0.02)	0.68 (0.03)	0.72 (0.02)	0.86 (0.01)	0.90 (0.02)	0.87 (0.02)
6	EfficientNetB0	0.80 (0.01)	0.31 (0.14)	0.44 (0.12)	0.80 (0.02)	1.00 (0.01)	0.89 (0.02)
7	XceptionNet	1.00 (0.01)	0.27 (0.04)	0.41 (0.03)	0.70 (0.02)	1.00 (0.01)	0.82 (0.02)
8	Inception-ResNet-v2	1.00 (0.01)	0.32 (0.09)	0.48 (0.03)	0.71 (0.04)	1.00 (0.01)	0.83 (0.02)
9	XceptionNet	1.00 (0.01)	0.29 (0.02)	0.43 (0.03)	0.80 (0.01)	1.00 (0.01)	0.89 (0.02)
10	XceptionNet	1.00 (0.01)	0.50 (0.00)	0.66 (0.02)	0.86 (0.01)	1.00 (0.01)	0.92 (0.03)
11	XceptionNet	1.00 (0.01)	0.27 (0.02)	0.41 (0.01)	0.77 (0.03)	1.00 (0.01)	0.87 (0.02)
12	XceptionNet	1.00 (0.01)	0.65 (0.03)	0.78 (0.01)	0.87 (0.03)	1.00 (0.01)	0.93 (0.03)

Table 5. Results obtained with the DL approach, i.e., the FCDD network. Average performance over the k folds ($k = 5$) and standard deviation (the latter within round brackets) for each set and class are reported.

		Non-Target			**Target**	
Set	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
1	1.00 (0.00)	0.75 (0.05)	0.86 (0.03)	1.00 (0.00)	0.83 (0.03)	0.96 (0.03)
2	1.00 (0.01)	0.92 (0.01)	0.96 (0.02)	0.75 (0.02)	1.00 (0.00)	0.86 (0.01)
3	1.00 (0.01)	0.89 (0.01)	0.94 (0.02)	1.00 (0.01)	0.80 (0.02)	0.89 (0.01)
4	1.00 (0.00)	0.83 (0.03)	0.96 (0.03)	1.00 (0.00)	0.80 (0.02)	0.89 (0.01)
5	1.00 (0.01)	0.75 (0.02)	0.86 (0.03)	1.00 (0.01)	0.83 (0.03)	0.96 (0.03)
6	1.00 (0.01)	0.80 (0.02)	0.89 (0.02)	1.00 (0.01)	0.80 (0.02)	0.89 (0.01)
7	1.00 (0.01)	0.86 (0.01)	0.92 (0.02)	1.00 (0.01)	0.75 (0.02)	0.86 (0.01)
8	1.00 (0.01)	0.88 (0.03)	0.94 (0.01)	1.00 (0.01)	0.80 (0.02)	0.89 (0.01)
9	1.00 (0.01)	0.88 (0.03)	0.94 (0.01)	1.00 (0.01)	0.86 (0.01)	0.92 (0.01)
10	1.00 (0.01)	0.83 (0.03)	0.96 (0.03)	1.00 (0.01)	0.86 (0.01)	0.92 (0.01)
11	1.00 (0.01)	0.86 (0.01)	0.92 (0.01)	1.00 (0.01)	0.83 (0.03)	0.96 (0.03)
12	1.00 (0.01)	0.88 (0.03)	0.94 (0.02)	1.00 (0.01)	0.86 (0.04)	0.92 (0.01)

4.2.1. Results with ML Approaches and HC Features

The results of this approach are summarized in Table 3. In this setting, the OCSVM consistently outperformed the IF on every single set. Significant differences were observed in the F1 values for the target class, ranging from 0.72% (set 12) to 0.91 (set 9), indicating how some sets were more challenging than others. Even though the precision for the target class was not particularly high (constantly below 0.80% except for set 9), the recall scores basically confirmed the discrete results on this class.

However, the main issue with this setting lies in the classification of the non-target class, which can determine the identification of the wrong cutting time and, therefore, a sub-optimal product. Despite some high precision scores (e.g., 1.00 on sets 5, 7, 8, and 11), the recall values demonstrated several misclassification, since no one surpassed 0.76.

From a feature point of view, the invariant moments, the ZMs in particular, resulted in the best in 8 out of the 12 sets, demonstrating superior performance compared to the other HC features.

4.2.2. Results with ML Approaches and Deep Features

The summarized results are presented in Table 4. In this setting, the OCSVM outperformed the IF on 9 out of the 12 sets. In fact, we acknowledge that the IF obtained scores between 0.98 and 0.99 for both classes on sets 7, 8, and 9. However, the OCSVM obtained a higher F1 score on average in the target class. For this reason, we report the results obtained with the OCSVM.

Even in this case, the performance varied across different sets and classes. For the target class, the precision values ranged from 0.54 to 0.87, with recall scores varying between 0.90 and 1.00. Similarly, for the non-target class, the precision scores ranged from 0.79 to 1.00, with corresponding recall scores ranging from 0.27 to 0.68.

Even though the deep features showcased notable performance improvements compared to the handcrafted ones, particularly in the target class, the disastrous results obtained with the recall scores on the non-target classes almost confirmed the results already obtained with the HC features.

Considering the architectures, XceptionNet produced the best features in 7 sets out of the 12, followed by EfficientNetB0 with 3.

4.2.3. Results with FCCDN

The results obtained by the FCDDN are presented in Table 5. In general, the performance was higher than the OCSVM and IF as this approach achieved the best results for both classes, demonstrating its ability to address the issue of low recall shown in the previous approaches. It also significantly improved the result's stability across different sets and demonstrated near-perfect precision scores for each set except for the target class of set 2. The results showed an average F1 score of 0.91 for the non-target class and 0.89 for the target class, indicating an increase of 34% and 5%, respectively, compared to shallow learning approaches using deep features.

The outcomes obtained through the FCDDN are detailed in Table 5. Generally, the performance surpassed that of the OCSVM and IF algorithms, as the FCDDN achieved superior results for both classes. This underscores its efficacy in addressing the issue of low recall observed in prior methodologies. Additionally, the FCDDN enhanced the stability of the results across various sets, exhibiting near-perfect precision scores for each set except for the target class of set 2. Specifically, the average F1 scores for the non-target and target classes were 0.91 and 0.89, respectively. These findings denote an enhancement of 34% and 5%, respectively, compared to ML approaches trained with deep features.

4.3. Qualitative Results

Given the importance of relying solely on the diffusion of granules and their size to identify the optimal cutting time, as indicated in Section 1, it is desirable for the FCDDN to exploit only these indicators for making predictions. Consequently, activation maps should primarily highlight the multitude of curd grains.

This reason motivated the conduction of the qualitative assessment of the results produced by the FCDDN, also to enhance transparency. Visual explanations with the produced heatmaps are calculated. The network generated coarse localization maps that highlighted regions in the image crucial for predicting the target class.

Figure 3 shows the heatmaps obtained for some sets (4, 5, 10) for both the non-target and target classes using the proposed FCDDN. The FCDDN classified the target class by utilizing a high number of curd grains, particularly in contrast to the non-target class. This distinction is logical since the target class in the examined dataset represents the optimal cutting time, which occurs when the coagulated milk mass consists of small curd grains.

Finally, considering that the nature of the proposed system is entirely based on a non-invasive and non-destructive visual inspection through a picture acquired by a digital camera, there are no other indicators, e.g., moisture or flavor, that can determine the appropriate cutting time. However, this aspect opens the field to the introduction of further non-invasive and non-destructive sensors for future work.

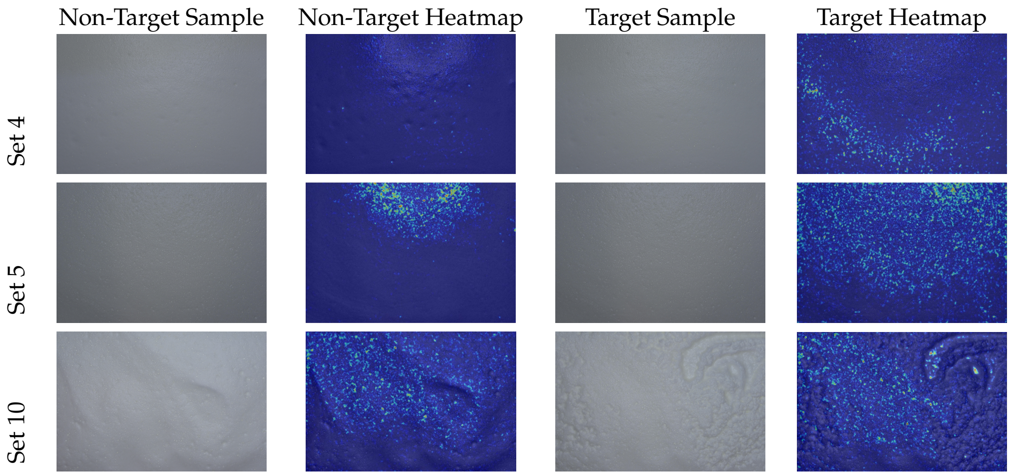

Figure 3. FCDDN heatmaps generated for some sample sets obtained for both classes.

4.4. Discussion

Based on the results, the FCDDN achieved an average F1 score of 0.91 and 0.89 for the non-target and target classes, respectively. This demonstrates its ability to accurately identify standard coagulation patterns and deviations, outperforming other evaluated methods. It also showed more stable prediction across different sets.

With ML classifiers, the deep features generally outperformed the HC features, and XceptionNet provided the best performance. Furthermore, the OCSVM outperformed the IF for all sets except three in the deep features setting. Despite the key advantage of the IF and OCSVM in their efficiency in handling high-dimensional data, the results have clearly shown that the IF struggled with datasets containing structured anomalies, which may be the case. Similarly, the OCSVM's performance on the non-target class may be sensitive to the choice of the kernel function and parameters, such as the nu parameter, which controls the trade-off between the margin size and the number of outliers.

The evaluated FCDDN provides a promising solution for automatically determining optimal cutting time. However, it is important to note that this work is a feasibility study relating to the problem under consideration. We must consider how similar datasets do not exist for the state of the art and that, to verify the generalizability of the proposed solution, additional datasets must be refined, even with synthetic data, as already in use in other contexts, from from video surveillance [85,86] to healthcare [87] and face detection [88].

Synthetic data are being used in CV tasks for object and AD. This provides a controlled environment for generating diverse data when real-world data are scarce [89]. In the context of cheese production, simulating the cheese-formation process and creating synthetic images could improve the models' generalization capabilities. However, synthetic data should closely resemble real-world conditions to ensure effective knowledge transfer to real-world scenarios. This aspect is particularly challenging in this scenario due to the physical and chemical changes that occur during cheese formation.

An additional consideration pertains to the potential necessity of re-training the system after its deployment in the dairy industry. Specifically, under the acquisition conditions outlined in this study, re-training the system is not required as the proposed pipeline is robust and not susceptible to domain shift, given that the acquisition condition is consolidated within the system pipeline. However, re-training may become essential to preserve accuracy and effectiveness under varying acquisition conditions. For instance, if the system is deployed in a dairy industry with different operational conditions, the production variables may change, potentially impacting the performance of the anomaly-detection model. Consequently, re-training the model with new data in such contexts will enable it to adapt to these changes, ensuring the detection of the optimal cutting time and maintaining the cheese quality.

In any case, with the refinement of additional datasets, real-world industrial implementation has the potential to enhance production efficiency and quality control.

5. Conclusions

This study aimed to create an automated approach for determining the optimal cutting time during cheese formation. One key finding was that deep learning-based algorithms, particularly the FCDDN, outperformed ML-based classifiers for this task. They achieved an average F1 score of 0.91 and 0.89 for the non-target and target classes, respectively. This demonstrates their ability to accurately identify standard coagulation patterns and deviations, surpassing other evaluated methods. Additionally, these algorithms showed more stable predictions across different sets.

Future research will focus on exploring and evaluating the proposed approach with more extensive and diversified potential industrial datasets, aiming to confirm the generalizability of the findings. In fact, the main further objective is the experimentation of the proposed system in the real-world scenario of the dairy industry that provided the dataset.

Also, the inclusion of data from a wide array of cheese varieties and production environments has the potential to yield valuable insights. Furthermore, integrating additional modalities, such as thermal imaging, could offer a more comprehensive understanding of the coagulation process, potentially enhancing detection performance within multimodal frameworks.

To address challenges stemming from limited labeled data availability, apart from the possible introduction of variations in factors like lighting conditions and camera angles to create synthetic data, adapting self-supervised and semi-supervised learning paradigms may prove advantageous, particularly through pre-training models on extensive unlabeled datasets. Moreover, the development of uncertainty estimation techniques and the investigation of further backbones could facilitate the calibration of predictions based on input ambiguity, thereby enhancing practical deployment in real-world scenarios.

Focusing on efficiency and deployability factors, designing embedded targeted networks tailored to considerations like model size and latency could streamline on-device implementation for integrated inspection systems. Finally, exploring domain adaptation and transfer learning approaches holds promise for generalizing models in this context.

Author Contributions: Conceptualization, A.L. and C.D.R.; methodology, A.L.; software, A.L. and D.G.; validation, A.L., D.G., A.P., L.Z., B.P. and C.D.R.; formal analysis, A.L., D.G., A.P., L.Z., B.P. and C.D.R.; investigation, A.L.; resources, A.L. and D.G.; data curation, A.L.; writing—original draft preparation, A.L., A.P., L.Z., B.P. and C.D.R.; writing—review and editing, A.L., A.P., L.Z., B.P. and C.D.R.; visualization, A.L., D.G., A.P., L.Z., B.P. and C.D.R.; supervision, B.P. and C.D.R.; project administration, C.D.R.; funding acquisition, C.D.R. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5—Call for tender No. 3277 published on 30 December 2021 by the Italian Ministry of University and Research (MUR) funded by the European Union—NextGenerationEU. Project Code ECS0000038—Project Title eINS Ecosystem of Innovation for Next Generation Sardinia—CUP F53C22000430001-Grant Assignment Decree No. 1056 adopted on 23 June 2022 by the Italian Ministry of University and Research (MUR).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The employed dataset is publicly available at the following https://figshare.com/s/140b812fcab25e7b9b8b (accessed on 28 April 2024).

Acknowledgments: We thank Massimiliano Sicilia of LAORE Sardegna and "Podda Formaggi" dairy industry for the dataset acquisition and Gianluca Dettori of "BiosAbbey" for supporting this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CV	computer vision
AD	anomaly detection
DL	deep learning
FCDDN	Fully Convolutional Data Description Network
HC	handcrafted
CNN	convolutional neural network
CH	Chebyshev moment
LM	Legendre moment
ZM	Zernike moment
Haar	Haar feature
HARri	rotation-invariant Haralick features
LBP	Local Binary Pattern
Hist	grayscale histogram feature
OCSVM	one-class SVM
IF	Isolation Forest
VAE	Variational Autoencoder
GAN	Generative Adversarial Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory

References

1. Lei, T.; Sun, D.W. Developments of Nondestructive Techniques for Evaluating Quality Attributes of Cheeses: A Review. *Trends Food Sci. Technol.* **2019**, *88*, 527–542. [CrossRef]
2. Castillo, M. *Cutting Time Prediction Methods in Cheese Making*; Taylor & Francis Group: Oxford, UK, 2006. [CrossRef]
3. Johnson, M.E.; Chen, C.M.; Jaeggi, J.J. Effect of rennet coagulation time on composition, yield, and quality of reduced-fat cheddar cheese. *J. Dairy Sci.* **2001**, *84*, 1027–1033. [CrossRef] [PubMed]
4. Grundelius, A.U.; Lodaite, K.; Östergren, K.; Paulsson, M.; Dejmek, P. Syneresis of submerged single curd grains and curd rheology. *Int. Dairy J.* **2000**, *10*, 489–496. [CrossRef]
5. Thomann, S.; Brechenmacher, A.; Hinrichs, J. Comparison of models for the kinetics of syneresis of curd grains made from goat's milk. *Milchwiss.-Milk Sci. Int.* **2006**, *61*, 407–411.
6. Gao, P.; Zhang, W.; Wei, M.; Chen, B.; Zhu, H.; Xie, N.; Pang, X.; Marie-Laure, F.; Zhang, S.; Lv, J. Analysis of the non-volatile components and volatile compounds of hydrolysates derived from unmatured cheese curd hydrolysis by different enzymes. *LWT* **2022**, *168*, 113896. [CrossRef]
7. Guinee, T.P. Effect of high-temperature treatment of milk and whey protein denaturation on the properties of rennet–curd cheese: A review. *Int. Dairy J.* **2021**, *121*, 105095. [CrossRef]
8. Liznerski, P.; Ruff, L.; Vandermeulen, R.A.; Franks, B.J.; Kloft, M.; Müller, K. Explainable Deep One-Class Classification. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, 3–7 May 2021. Available online: https://openreview.net/ (accessed on 26 April 2024).
9. Alinaghi, M.; Nilsson, D.; Singh, N.; Höjer, A.; Saedén, K.H.; Trygg, J. Near-infrared hyperspectral image analysis for monitoring the cheese-ripening process. *J. Dairy Sci.* **2023**, *106*, 7407–7418. [CrossRef]
10. Everard, C.D.; O'callaghan, D.; Fagan, C.C.; O'donnell, C.; Castillo, M.; Payne, F. Computer vision and color measurement techniques for inline monitoring of cheese curd syneresis. *J. Dairy Sci.* **2007**, *90*, 3162–3170. [CrossRef] [PubMed]
11. Loddo, A.; Di Ruberto, C.; Armano, G.; Manconi, A. Automatic Monitoring Cheese Ripeness Using Computer Vision and Artificial Intelligence. *IEEE Access* **2022**, *10*, 122612–122626. [CrossRef]
12. Badaró, A.T.; de Matos, G.V.; Karaziack, C.B.; Viotto, W.H.; Barbin, D.F. Automated method for determination of cheese meltability by computer vision. *Food Anal. Methods* **2021**, *14*, 2630–2641. [CrossRef]
13. Goyal, S.; Kumar Goyal, G. Shelflife Prediction of Processed Cheese Using Artificial Intelligence ANN Technique. *Hrvat. Časopis Prehrambenu Tehnol. Biotehnol. Nutr.* **2012**, *7*, 184–187.
14. Goyal, S.; Goyal, G. Smart artificial intelligence computerized models for shelf life prediction of processed cheese. *Int. J. Eng. Technol.* **2012**, *1*, 281–289. [CrossRef]
15. da Paixao Teixeira, J.L.; dos Santos Carames, E.T.; Baptista, D.P.; Gigante, M.L.; Pallone, J.A.L. Rapid adulteration detection of yogurt and cheese made from goat milk by vibrational spectroscopy and chemometric tools. *J. Food Compos. Anal.* **2021**, *96*, 103712. [CrossRef]
16. Vasafi, P.S.; Paquet-Durand, O.; Brettschneider, K.; Hinrichs, J.; Hitzmann, B. Anomaly detection during milk processing by autoencoder neural network based on near-infrared spectroscopy. *J. Food Eng.* **2021**, *299*, 110510. [CrossRef]

17. Li, Z.; Zhu, Y.; van Leeuwen, M. A Survey on Explainable Anomaly Detection. *ACM Trans. Knowl. Discov. Data* **2024**, *18*, 23:1–23:54. [CrossRef]
18. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, K. A Unifying Review of Deep and Shallow Anomaly Detection. *Proc. IEEE* **2021**, *109*, 756–795. [CrossRef]
19. Samariya, D.; Aryal, S.; Ting, K.M.; Ma, J. A New Effective and Efficient Measure for Outlying Aspect Mining. In Proceedings of the Web Information Systems Engineering—WISE 2020—21st International Conference, Amsterdam, The Netherlands, 20–24 October 2020; Proceedings, Part II; Lecture Notes in Computer Science; Huang, Z., Beek, W., Wang, H., Zhou, R., Zhang, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12343, pp. 463–474.
20. Nguyen, V.K.; Renault, É.; Milocco, R.H. Environment Monitoring for Anomaly Detection System Using Smartphones. *Sensors* **2019**, *19*, 3834. [CrossRef] [PubMed]
21. Violettas, G.E.; Simoglou, G.; Petridou, S.G.; Mamatas, L. A Softwarized Intrusion Detection System for the RPL-based Internet of Things networks. *Future Gener. Comput. Syst.* **2021**, *125*, 698–714. [CrossRef]
22. Papageorgiou, G.; Sarlis, V.; Tjortjis, C. Unsupervised Learning in NBA Injury Recovery: Advanced Data Mining to Decode Recovery Durations and Economic Impacts. *Information* **2024**, *15*, 61. [CrossRef]
23. Zhao, X.; Zhang, L.; Cao, Y.; Jin, K.; Hou, Y. Anomaly Detection Approach in Industrial Control Systems Based on Measurement Data. *Information* **2022**, *13*, 450. [CrossRef]
24. Li, J.; Izakian, H.; Pedrycz, W.; Jamal, I. Clustering-based anomaly detection in multivariate time series data. *Appl. Soft Comput.* **2021**, *100*, 106919. [CrossRef]
25. Falcão, F.; Zoppi, T.; Silva, C.B.V.; Santos, A.; Fonseca, B.; Ceccarelli, A.; Bondavalli, A. Quantitative comparison of unsupervised anomaly detection algorithms for intrusion detection. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC 2019, Limassol, Cyprus, 8–12 April 2019; Hung, C., Papadopoulos, G.A., Eds.; ACM: New York, NY, USA, 2019; pp. 318–327. [CrossRef]
26. Schölkopf, B.; Williamson, R.C.; Smola, A.; Shawe-Taylor, J.; Platt, J. Support Vector Method for Novelty Detection. In *Advances in Neural Information Processing Systems*; Solla, S., Leen, T., Müller, K., Eds.; MIT Press: Cambridge, MA, USA, 1999; Volume 12.
27. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422. [CrossRef]
28. Liang, J.; Liang, Q.; Wu, Z.; Chen, H.; Zhang, S.; Jiang, F. A Novel Unsupervised Deep Transfer Learning Method With Isolation Forest for Machine Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2024**, *20*, 235–246. [CrossRef]
29. Wang, W.; ShangGuan, W.; Liu, J.; Chen, J. Enhanced Fault Detection for GNSS/INS Integration Using Maximum Correntropy Filter and Local Outlier Factor. *IEEE Trans. Intell. Veh.* **2024**, *9*, 2077–2093. [CrossRef]
30. Kumar, R.H.; Bank, S.; Bharath, R.; Sumati, S.; Ramanarayanan, C.P. A Local Outlier Factor-Based Automated Anomaly Event Detection of Vessels for Maritime Surveillance. *Int. J. Perform. Eng.* **2023**, *19*, 711. [CrossRef]
31. Siddiqui, M.A.; Stokes, J.W.; Seifert, C.; Argyle, E.; McCann, R.; Neil, J.; Carroll, J. Detecting Cyber Attacks Using Anomaly Detection with Explanations and Expert Feedback. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2872–2876. [CrossRef]
32. Wang, L.; Huynh, D.Q.; Mansour, M.R. Loss Switching Fusion with Similarity Search for Video Classification. In Proceedings of the 2019 IEEE International Conference on Image Processing, ICIP 2019, Taipei, Taiwan, 22–25 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 974–978.
33. LeCun, Y.; Bengio, Y.; Hinton, G.E. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
34. Zipfel, J.; Verworner, F.; Fischer, M.; Wieland, U.; Kraus, M.; Zschech, P. Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. *Comput. Ind. Eng.* **2023**, *177*, 109045. [CrossRef]
35. Xie, G.; Wang, J.; Liu, J.; Lyu, J.; Liu, Y.; Wang, C.; Zheng, F.; Jin, Y. IM-IAD: Industrial Image Anomaly Detection Benchmark in Manufacturing. *IEEE Trans. Cybern.* **2024**, *54*, 2720–2733. [CrossRef] [PubMed]
36. Liu, W.; Li, R.; Zheng, M.; Karanam, S.; Wu, Z.; Bhanu, B.; Radke, R.J.; Camps, O. Towards Visually Explaining Variational Autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
37. Wang, K.; Yan, C.; Mo, Y.; Wang, Y.; Yuan, X.; Liu, C. Anomaly detection using large-scale multimode industrial data: An integration method of nonstationary kernel and autoencoder. *Eng. Appl. Artif. Intell.* **2024**, *131*, 107839. [CrossRef]
38. Kim, S.; Jo, W.; Shon, T. APAD: Autoencoder-based Payload Anomaly Detection for industrial IoE. *Appl. Soft Comput.* **2020**, *88*, 106017. [CrossRef]
39. Ravanbakhsh, M.; Nabi, M.; Sangineto, E.; Marcenaro, L.; Regazzoni, C.S.; Sebe, N. Abnormal event detection in videos using generative adversarial nets. In Proceedings of the 2017 IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, 17–20 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1577–1581. [CrossRef]
40. Zhang, L.; Dai, Y.; Fan, F.; He, C. Anomaly Detection of GAN Industrial Image Based on Attention Feature Fusion. *Sensors* **2023**, *23*, 355. [CrossRef]
41. Liu, W.; Luo, W.; Lian, D.; Gao, S. Future Frame Prediction for Anomaly Detection—A New Baseline. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; Computer Vision Foundation: New York, NY, USA; IEEE Computer Society: Washington, DC, USA, 2018; pp. 6536–6545. [CrossRef]

42. Georgescu, M.I.; Barbalau, A.; Ionescu, R.T.; Khan, F.S.; Popescu, M.; Shah, M. Anomaly Detection in Video via Self-Supervised and Multi-Task Learning. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 12742–12752.
43. Zhou, X.; Hu, Y.; Liang, W.; Ma, J.; Jin, Q. Variational LSTM Enhanced Anomaly Detection for Industrial Big Data. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3469–3477. [CrossRef]
44. Ullah, W.; Hussain, T.; Ullah, F.U.M.; Lee, M.Y.; Baik, S.W. TransCNN: Hybrid CNN and transformer mechanism for surveillance anomaly detection. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106173. [CrossRef]
45. Lu, S.; Dong, H.; Yu, H. Abnormal Condition Detection Method of Industrial Processes Based on the Cascaded Bagging-PCA and CNN Classification Network. *IEEE Trans. Ind. Inform.* **2023**, *19*, 10956–10966. [CrossRef]
46. Smith, A.D.; Du, S.; Kurien, A. Vision transformers for anomaly detection and localisation in leather surface defect classification based on low-resolution images and a small dataset. *Appl. Sci.* **2023**, *13*, 8716. [CrossRef]
47. Yao, H.; Luo, W.; Yu, W.; Zhang, X.; Qiang, Z.; Luo, D.; Shi, H. Dual-attention transformer and discriminative flow for industrial visual anomaly detection. *IEEE Trans. Autom. Sci. Eng.* **2023**, 1–15. [CrossRef]
48. Jézéquel, L.; Vu, N.; Beaudet, J.; Histace, A. Efficient Anomaly Detection Using Self-Supervised Multi-Cue Tasks. *IEEE Trans. Image Process.* **2023**, *32*, 807–821. [CrossRef]
49. Tang, X.; Zeng, S.; Yu, F.; Yu, W.; Sheng, Z.; Kang, Z. Self-supervised anomaly pattern detection for large scale industrial data. *Neurocomputing* **2023**, *515*, 1–12. [CrossRef]
50. Yan, S.; Shao, H.; Xiao, Y.; Liu, B.; Wan, J. Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. *Robot. Comput. Integr. Manuf.* **2023**, *79*, 102441. [CrossRef]
51. Rosero-Montalvo, P.D.; István, Z.; Tözün, P.; Hernandez, W. Hybrid Anomaly Detection Model on Trusted IoT Devices. *IEEE Internet Things J.* **2023**, *10*, 10959–10969. [CrossRef]
52. Borré, A.; Seman, L.O.; Camponogara, E.; Stefenon, S.F.; Mariani, V.C.; dos Santos Coelho, L. Machine Fault Detection Using a Hybrid CNN-LSTM Attention-Based Model. *Sensors* **2023**, *23*, 4512. [CrossRef] [PubMed]
53. Abir, F.F.; Chowdhury, M.E.H.; Tapotee, M.I.; Mushtak, A.; Khandakar, A.; Mahmud, S.; Hasan, A. PCovNet+: A CNN-VAE anomaly detection framework with LSTM embeddings for smartwatch-based COVID-19 detection. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106130. [CrossRef] [PubMed]
54. Niu, Z.; Yu, K.; Wu, X. LSTM-Based VAE-GAN for Time-Series Anomaly Detection. *Sensors* **2020**, *20*, 3738. [CrossRef] [PubMed]
55. Putzu, L.; Loddo, A.; Ruberto, C.D. Invariant Moments, Textural and Deep Features for Diagnostic MR and CT Image Retrieval. In Proceedings of the Computer Analysis of Images and Patterns: 19th International Conference, CAIP 2021, Virtual Event, 28–30 September 2021; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2021; pp. 287–297. [CrossRef]
56. Mukundan, R.; Ong, S.; Lee, P. Image analysis by Tchebichef moments. *IEEE Trans. Image Process.* **2001**, *10*, 1357–1364. [CrossRef] [PubMed]
57. Teague, M.R. Image analysis via the general theory of moments∗. *J. Opt. Soc. Am.* **1980**, *70*, 920–930. [CrossRef]
58. Teh, C.H.; Chin, R. On image analysis by the methods of moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **1988**, *10*, 496–513. [CrossRef]
59. Oujaoura, M.; Minaoui, B.; Fakir, M. Image Annotation by Moments. *Moments Moment Invariants Theory Appl.* **2014**, *1*, 227–252. [CrossRef]
60. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. I, ISSN: 1063-6919. [CrossRef]
61. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]
62. Putzu, L.; Di Ruberto, C. Rotation Invariant Co-occurrence Matrix Features. In Proceedings of the Image Analysis and Processing—ICIAP 2017, Catania, Italy, 11–15 September 2017; Lecture Notes in Computer Science; Battiato, S., Gallo, G., Schettini, R., Stanco, F., Eds.; Springer: Cham, Switzerland, 2017; pp. 391–401. [CrossRef]
63. He, D.-c.; Wang, L. Texture Unit, Texture Spectrum, And Texture Analysis. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 509–512. [CrossRef]
64. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]
65. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
66. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
67. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
68. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015. IEEE Computer Society: Washington, DC, USA, 2015; pp. 1–9. [CrossRef]

69. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of Machine Learning Research, Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019*; Volume 97, pp. 6105–6114.
70. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; IEEE Computer Society: Washington, DC, USA, 2016; pp. 2818–2826. [CrossRef]
71. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; AAAI Press: Washington, DC, USA, 2017; pp. 4278–4284.
72. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 18–23 June 2018; pp. 8697–8710. [CrossRef]
73. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
74. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
75. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
76. Petrovska, B.; Zdravevski, E.; Lameski, P.; Corizzo, R.; Štajduhar, I.; Lerga, J. Deep Learning for Feature Extraction in Remote Sensing: A Case-Study of Aerial Scene Classification. *Sensors* **2020**, *20*, 3906. [CrossRef] [PubMed]
77. Barbhuiya, A.A.; Karsh, R.K.; Jain, R. CNN based feature extraction and classification for sign language. *Multimed. Tools Appl.* **2021**, *80*, 3051–3069. [CrossRef]
78. Varshni, D.; Thakral, K.; Agarwal, L.; Nijhawan, R.; Mittal, A. Pneumonia Detection Using CNN based Feature Extraction. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 20–22 February 2019; pp. 1–7. [CrossRef]
79. Rippel, O.; Mertens, P.; König, E.; Merhof, D. Gaussian Anomaly Detection by Modeling the Distribution of Normal Data in Pretrained Deep Features. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [CrossRef]
80. Yang, J.; Shi, Y.; Qi, Z. Learning deep feature correspondence for unsupervised anomaly detection and segmentation. *Pattern Recognit.* **2022**, *132*, 108874. [CrossRef]
81. Reiss, T.; Cohen, N.; Bergman, L.; Hoshen, Y. PANDA: Adapting Pretrained Features for Anomaly Detection and Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, Virtual, 19–25 June 2021; Computer Vision Foundation: New York, NY, USA; IEEE: Piscataway, NJ, USA, 2021; pp. 2806–2814. [CrossRef]
82. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), Miami, FL, USA, 20–25 June 2009; IEEE Computer Society: Washington, DC, USA, 2009; pp. 248–255. [CrossRef]
83. Alam, S.; Sonbhadra, S.K.; Agarwal, S.; Nagabhushan, P. One-class support vector classifiers: A survey. *Knowl. Based Syst.* **2020**, *196*, 105754. [CrossRef]
84. Cheng, X.; Zhang, M.; Lin, S.; Zhou, K.; Zhao, S.; Wang, H. Two-Stream Isolation Forest Based on Deep Features for Hyperspectral Anomaly Detection. *IEEE Geosci. Remote. Sens. Lett.* **2023**, *20*, 5504205. [CrossRef]
85. Delussu, R.; Putzu, L.; Fumera, G. Synthetic Data for Video Surveillance Applications of Computer Vision: A Review. *Int. J. Comput. Vis.* **2024**, 1–37. [CrossRef]
86. Foszner, P.; Szczesna, A.; Ciampi, L.; Messina, N.; Cygan, A.; Bizon, B.; Cogiel, M.; Golba, D.; Macioszek, E.; Staniszewski, M. CrowdSim2: An Open Synthetic Benchmark for Object Detectors. In Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Lisbon, Portugal, 19–21 February 2023; Radeva, P., Farinella, G.M., Bouatouch, K., Eds.; Volume 5, pp. 676–683.
87. Murtaza, H.; Ahmed, M.; Khan, N.F.; Murtaza, G.; Zafar, S.; Bano, A. Synthetic data generation: State of the art in health care domain. *Comput. Sci. Rev.* **2023**, *48*, 100546. [CrossRef]
88. Boutros, F.; Struc, V.; Fiérrez, J.; Damer, N. Synthetic data for face recognition: Current state and future prospects. *Image Vis. Comput.* **2023**, *135*, 104688. [CrossRef]
89. Zhang, Z.; Zhao, Z.; Zhang, X.; Sun, C.; Chen, X. Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction. *Comput. Ind.* **2023**, *151*, 103990. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.