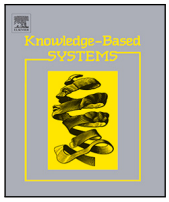
Knowledge-Based Systems 295 (2024) 111833

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Understanding cheese ripeness: An artificial intelligence-based approach for hierarchical classification

Luca Zedda *, Alessandra Perniciano, Andrea Loddo, Cecilia Di Ruberto

Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124, Cagliari, Italy

ARTICLE INFO

Keywords:
Cheese ripeness classification
Hybrid approach
Machine learning
Hierarchical machine learning
Computer vision
Deep learning
Vision transformer
Convolutional neural networks
Handcrafted features

ABSTRACT

Within the contemporary dairy industry, the effective monitoring of cheese ripeness constitutes a critical yet challenging task. This paper proposes the first public dataset encompassing images of cheese wheels that depict various products at distinct stages of ripening and introduces an innovative hybrid approach, integrating machine learning and computer vision techniques to automate the detection of cheese ripeness. By leveraging deep learning and shallow learning techniques, the proposed method endeavors to overcome the limitations associated with conventional assessment methodologies. It aims to provide automation, precision, and consistency in the evaluation of cheese ripeness, delving into a hierarchical classification for the simultaneous classification of distinct cheese types and ripeness levels and presenting a comprehensive solution to enhance the efficiency of the cheese production process. By employing a lightweight hierarchical feature aggregation methodology, this investigation navigates the intricate landscape of preprocessing steps, feature selection, and diverse classifiers. We report a noteworthy achievement, attaining a best F-measure score of 0.991 through the merging of features extracted from EfficientNet and DarkNet-53, opening the field to concretely address the complexity inherent in cheese quality assessment.

1. Introduction

The dairy industry, of which cheese is a prominent product, generates substantial revenue, creates jobs, and contributes to the agricultural and food processing sectors. Dairy products are both a nutritional powerhouse and a catalyst for positive health outcomes. The high content of proteins, calcium, and vital micronutrients promotes bone and muscle health. Additionally, their probiotics enhance digestive well-being and nurture a healthy microbiome.

Beyond its gastronomic significance, cheese has a profound economic impact, especially in regions where cheese production is a vital part of the agricultural sector. The dairy industry, bolstered by cheese production, generates substantial revenue and creates employment opportunities, supporting the agricultural and food processing sectors. Moreover, cheese plays a crucial role in international trade, with various countries engaging in the import and export of diverse cheese varieties to cater to the preferences of a global consumer base.

Determining cheese quality involves evaluating its chemical components, internal structure, and sensory aspects triggered by specific properties and components [1]. The critical step in the cheese-making process is detecting cheese ripeness, which is a specialized task predominantly reliant on the keen observation and sensory evaluation of experts who assess cheese wheels visually, by scent, or by weight checking. This artisanal approach, although conducted by trained professionals, is characterized by its inherent time-intensive nature and the necessity for rigorous personnel training. Apart from being susceptible to the subjective inclinations of individual examiners, introducing an element of personal bias into the assessment process, it is further complicated by biochemical changes like lipolysis and proteolysis during the ripening phase, significantly impacting flavor, aroma, and texture [2,3].

These issues are further emphasized by the colossal scale of cheese production in the industry [4], with leading manufacturers producing over 4 million cheese wheels annually. This volume necessitates an ongoing commitment to staffing and training to maintain the quality and consistency of cheese products.

In general, accurately assessing cheese ripeness can be challenging due to factors such as the season, the origin of the milk, the various processing steps, and storage temperature, which can be unpredictable. Mistakes in determining cheese ripeness may lead to lower-quality products being released into the market, harming a company's reputation and revenue [5].

To overcome these challenges, the dairy industry is increasingly interested in adopting cutting-edge technologies for efficient monitoring. Non-invasive techniques based on physicochemical, chromatographic,

* Corresponding author.
E-mail address: luca.zedda@unica.it (L. Zedda).

https://doi.org/10.1016/j.knosys.2024.111833
Received 12 February 2024; Received in revised form 28 March 2024; Accepted 17 April 2024
Available online 19 April 2024

and electrophoretic analyses have been explored, but they are costly and time-consuming [6].

Some alternative methodologies focus on computer vision (CV), spectral analysis, and ultrasound-based techniques. These methodologies have evolved considerably in recent years, offering rapid, non-destructive, and non-invasive means of monitoring cheese production processes. Particularly within the domain of CV, advancements have led to its widespread application across the agro-food chain, owing to its cost-effectiveness, objectivity, reliability, and speed [1]. CV techniques encompass diverse tasks ranging from qualitative and quantitative analyses to defect identification, and various cheese varieties such as Cheddar, Mozzarella, Parmigiano Reggiano, Grana Padano, Queijo de Nisa, and Pecorino cheeses. However, the quality of CV-based analyses heavily relies on consistent lighting conditions, prompting the exploration of more sophisticated techniques such as X-ray, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT), albeit with associated challenges and costs. In contrast, spectral analysis approaches use light reflection to assess cheese properties, providing valuable insights into cheese structure and composition. Meanwhile, ultrasound-based methods offer non-invasive means of monitoring cheese integrity, particularly in detecting eyes and cracks.

Despite the advancements, the current state of cheese quality analysis remains constrained by data acquisition challenges and privacy concerns. Hence, this study endeavors to contribute to the field through the development of computer vision and machine learning (ML) techniques, focusing on digital image analysis acquired via a simple digital camera, emphasizing its non-intrusive, non-destructive, and cost-effective nature. More precisely, we propose an innovative hybrid approach to automate cheese ripeness detection with a hierarchical classification strategy. By combining deep learning and shallow learning techniques, this approach aims to address the limitations of traditional methods and reduce reliance on human experts.

The contributions of this paper can be summarized as follows:

- **Public dataset release.** We hereby provide public access to the first dataset encompassing images of cheese wheels that depict various products at distinct stages of ripening. It is available at the following GitHub repository.
- **Hybrid approach.** We introduce an innovative artificial intelligence-based solution to automate the challenging task of cheese ripeness detection.
- **Hierarchical classification.** We present a hierarchical classification strategy for the simultaneous differentiation of various cheese types and ripeness levels, addressing the multifaceted nature of cheese production.
- **Feature combination methodology.** We propose a hierarchical feature aggregation methodology that streamlines the feature selection process and offers a lightweight yet effective approach to feature combination.
- **Insights into feature importance.** We provide valuable insights into the significance of preprocessing steps, classifiers, and feature selection techniques through comprehensive experiments, contributing to the understanding of optimal feature combinations.
- **Source code release.** The source code of the framework realized in this work is publicly available at the following GitHub repository.

The work is subdivided into the following sections: Section 2 reviews existing techniques in cheese quality analysis, establishing a foundation for the hybrid approach, while Section 3 gives the details about the dataset, feature extraction, selection and normalization, and machine learning strategies, laying the groundwork for the study. In Section 4, the overall framework proposed is presented, outlining the study's main focuses: a hierarchical classification approach and a unique feature aggregation methodology. In Section 5 the selected experimental environment and hyperparameters are described, along with the presentation of the entire range of experimental results. Finally, in Section 6, we formulate our conclusions about the conducted study and provide insightful possible new ways to improve our work and possibly inspire future works based on our findings.

2. Related work

Within the realm of cheese production, different methodologies, including CV with digital or hyperspectral imaging, near-infrared (NIH) spectroscopy, Fourier-transformed infrared (FTIR) spectroscopy, and other analytical techniques have been developed for monitoring the cheese production process [1]. These methods serve not only for quality assessment but also for determining the geographical origin and detecting potential adulteration in cheese products.

In addition, they offer the desired features of being a rapid, non-destructive, and non-invasive methodology. In contrast to traditional approaches, they can complement human visual inspection in evaluating the attributes and sensory quality of cheese, for example, with color assessments, identification of cheese imperfections such as gas or mechanical holes, the presence of calcium lactate crystals, excessive rind halo formation, and oiling off [7,8] without direct contact with the samples.

Despite the ever-growing need for improvement in the quality and quantity of cheese production, the state of the art for several sub-fields of cheese quality analysis is still limited. The high difficulty in data acquisition and the necessity of maintaining high privacy standards due to industrial patents and techniques can explain this scarcity. However, non-destructive approaches are the key to industrial-scale solutions. Such approaches can be divided into three categories: computer vision, spectral, and ultrasound-based, as presented in Fig. 1.

Computer Vision-based Approaches. In recent years, the use of CV approaches has gained prominence across various stages of the agro-food chain due to its objectivity, reliability, speed, and cost-effectiveness [8–10]. A CV-based pipeline involves the use of a digital camera to capture images, which are then processed for further analysis in a multitude of tasks, from the qualitative [9,11] and quantitative [12,13] analysis to the inspection and identification of defects [10,14] of fruits and vegetables.

In this context, CV techniques have been used across various cheese varieties, including Cheddar [15], Mozzarella [16], Parmigiano Reggiano [17,18], Grana Padano [18,19], Queijo de Nisa [8], and Pecorino [20] cheeses. In addition, they were also employed for ingredient distribution inspection [21].

In any case, the quality of the images highly depends on the constancy of the direction and strength of the light source. To overcome this issue, more complex techniques such as X-ray, MRI, and CT can provide a more robust representation of the image, along with a 3D representation of the inside of the cheese wheel, enabling cheese eye analysis [1,22–24]. Of course, these approaches require specific material that is normally used in a different context, such as the medical context. For this reason, a minimally invasive and, at the same time, inexpensive approach is based on digital images acquired with a digital camera, as done in this work.

Spectral Analysis-based Approaches. Spectral analysis approaches employ the use of light reflected by cheese wheels using probes or various sources of light emitters. Fluorescence spectroscopy is a very effective type of spectral approach that determines the intensity of fluorescent components in the cheese, such as vitamin A, which determines the ripeness and the molecular structure of the cheese [25,26]. NIR provides another key spectral information for various applications. In detail, for cheese analysis, it provides data about the cheese structure, which can be used to determine quality, protein percentage, moisture, and several other parameters [27–29]. Similar to the NIR approach, the FTIR approach employs infrared spectroscopy to determine various aspects of the cheese. This particular technology is usually employed in rapid or real-time scenarios [1,30,31].

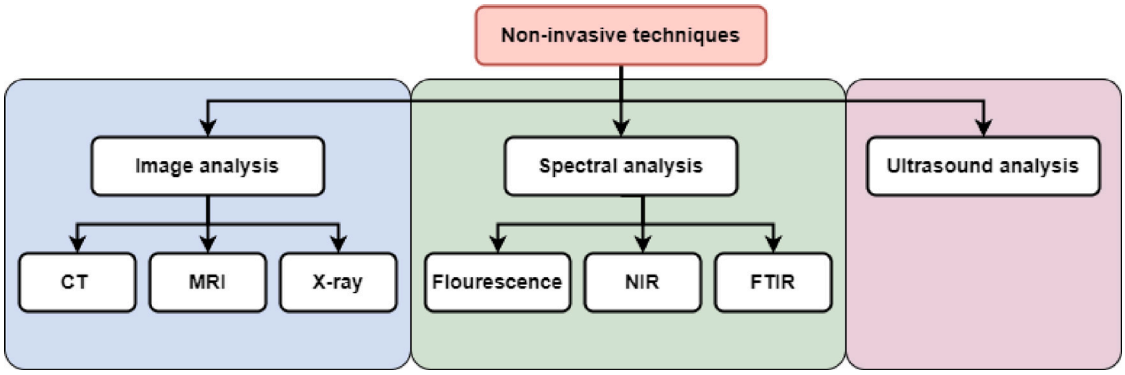


Fig. 1. Hierarchical representation of the different non-destructive techniques for cheese quality analysis.

Ultrasound-based Approaches. Similar to CT scans, X-rays, and MRI imaging, ultrasound methods have emerged as valuable tools for monitoring the development of eyes and cracks within cheese [1]. For instance, by exploiting transducers and oscilloscopes, these tool obtain wave data that describes the integrity and structure of the cheese wheel [32]. Also, Eskelinen et al. demonstrated the capability of ultrasonic techniques in monitoring the progression of eyes and cracks in cheese as it matures, often reconstructing three-dimensional ultrasound images of cheese samples [33]. Meanwhile, Nassar et al. investigated cheese eye formation through ultrasonic methods, highlighting limitations encountered, particularly with mature cheeses that exhibit extensive openings To address this issue, an alternative technique known as the tap test acoustic method was devised [34].

This study revolves around the advancement of computer vision and machine learning methodologies employed in analyzing digital images captured using a basic digital camera. Unlike prevailing methods, this approach is characterized by its non-intrusive and non-destructive nature, prioritizing simplicity in setup and cost-effectiveness.

3. Materials and methods

This section describes the analytical pipeline followed in the study, including the key elements of this framework. Section 3.1 presents the image dataset used. Section 3.2 describes the feature extraction methods, distinguished in deep and handcrafted (HC) features, while Section 3.3 outlines the criteria and techniques for selecting the appropriate features. Then, Section 3.4 discusses the approaches used to standardize and normalize the features. Lastly, Section 3.5 details the specific methodologies and algorithms used to apply machine learning to the processed data.

3.1. Dataset

In this work, we present and provide the first dataset encompassing images of cheese wheels that depict various products at distinct stages of ripening, called the CHEESE-Hierarchical Image Data Base (CHEESE-HIDB).

CHEESE-HIDB was built with the support of the Sardinian agency for the implementation of regional agricultural and rural development programs (LAORE[1]) and BiosAbbey S.r.l.

All the images in this dataset represent cheese wheels acquired at the Italian dairy company Lattebusche-Conad. They are centered on cardboard with a single light source and have a resolution of 6016×4016 pixels.

The images are hierarchically categorized into two levels. The first one represents three different product types, called **Semi-Hard**, **Hard**, and **Extra-Hard**. Some sample images from the three first-level classes are depicted in Fig. 2. Moreover, each product type is divided into two further categories, viz. “Target” and “Not Target”, representing cheese wheels with adequate ripeness and cheese wheels that require additional aging or have exceeded the target, respectively. The number of images for each class is depicted in Table 1. As can be seen, each class has an uneven number of images, skewed towards the “Not Target” class.

Table 1
Count of cheese wheel image samples for each of the represented class in CHEESE-HIDB.

Class	Target	Not target
Semi-Hard	42	84
Hard	42	84
Extra-Hard	42	84

The maturation process for the distinct cheese classes within the dataset exhibits specific temporal patterns. In the case of the Semi-Hard variety, the target ripeness is achieved over a span of 92 days. For the Hard cheese, classified under the Hard class, the target category undergoes maturation for 207 days. Finally, the Extra-Hard class achieves its target ripeness over 460 days.

3.2. Feature extraction

In this section, we present the feature extraction process and the different preprocessing steps employed in our analysis.

3.2.1. Data preprocessing

The naive approach to ripeness and product classification is to use the full RGB image to obtain semantically relevant features and descriptors. However, modern deep learning architectures and HC image descriptor extraction are not well-suited for handling such high-dimensional data. The first step in our preprocessing strategy is to correctly resize the image to the input size of the deep neural networks utilized or to 224×224 if HC descriptors are employed. Another crucial step is to center-crop the image to eliminate most of the background portion.

As evident from Fig. 2, all the images share several common elements, with the most significant one being the presence of a cheese label that denotes the type and date of the wheel. To minimize any potential bias or OCR-related information, we conducted an experiment to select the RGB channel with the lowest cheese label intensity. A clear example is illustrated in Fig. 3.

Fig. 3 reveals that the blue channel provides the strongest image response with the lowest label response. Therefore, for all our experiments, we utilized it for HC image features and a 3-channel stacked representation for deep feature extraction approaches.

[1] https://www.sardegnaagricoltura.it

Fig. 2. Illustrative examples showcasing images from the CHEESE-HIDB. Each column corresponds to samples representing distinct cheese classes, including Semi-Hard, Hard, and Extra-Hard. The rows delineate ripeness classes, depicting instances categorized as Target and Not Target.

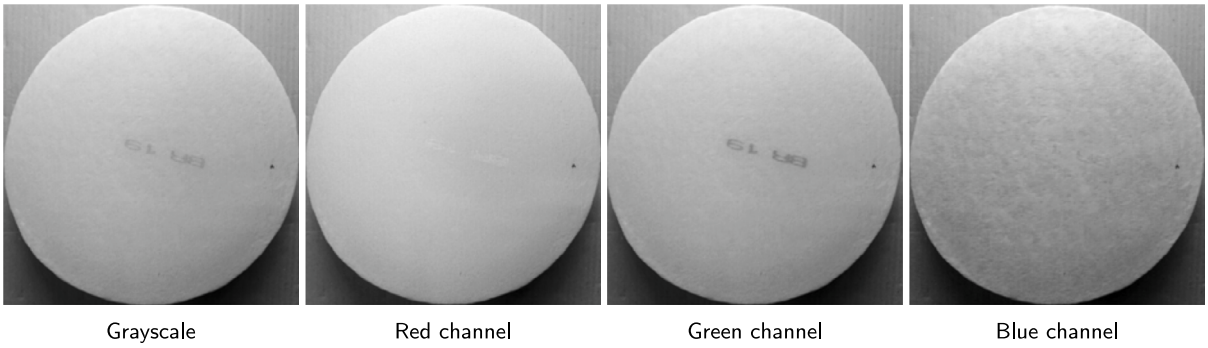


Fig. 3. Illustration of a sample CHEESE-HIDB image categorized as belonging to the Semi-Hard class (Target). From left to right: grayscale conversion of the original RGB image, followed by representations of the red, green, and blue channels, respectively.

3.2.2. Deep learning features

Deep learning feature extraction and classification, in conjunction with shallow learning classifiers, have established themselves as a successful approach for augmenting the predictive capabilities of traditional deep learning models [35]. These deep models often grapple with the challenge of high dimensionality, commonly referred to as the "curse of dimensionality". To mitigate this challenge, feature selection and aggregation processes can be employed [36].

Vision Transformers (ViT) have sparked a revolution in the field of computer vision, delivering cutting-edge performance across various subfields. ViT, originally introduced by Dosovitskiy et al. in 2020 [37], marked the first successful application of the transformer architecture to vision-related tasks. Vision transformers operate by segmenting the input image into fixed-sized patches, embedding them into a hidden dimension, and incorporating positional encoding. Typically, a learnable class patch, also referred to as a class token, is concatenated with the image patches, and the resulting sequence is processed through a series of transformer layers. Notably, the versatility of vision transformers as foundational models [38–43] has enabled the utilization of unlabeled data for efficient image representation and improved pretraining strategies.

While the application of ViT for feature extraction and shallow learning-based classification is relatively unexplored, it has garnered increasing interest in recent research [44]. Depending on the adopted architecture and the utilization of the class token, the feature extraction process can vary. Specifically, if the class token is used, it serves as the source of features. Otherwise, features are extracted by averaging the image patches from the last transformer layer [37]. Detailed information regarding the chosen architectures for feature extraction, input size, and model size for each ViT can be found in Table 2.

Table 2
Details of Vision Transformer architectures employed in this study, including the reference paper, model size, patch size, input shape, and feature extraction method.

Reference	Model size	Patch size	Input shape	Feature extraction
ViT [37]	Base	16	224 × 224	CLS
DINO [39]	Base	16	224 × 224	Average
DINO-v2 [42]	Base	14	224 × 224	Average
EVA [40]	Large	14	336 × 336	Average
CLIP [43]	Base	32	384 × 384	Average
Swin-v2 [40]	Tiny	16	256 × 256	CLS
MAE [41]	Base	16	224 × 224	Average
I-JEPA [38]	Huge	14	224 × 224	Average

Convolutional Neural Networks (CNNs) have proven their worth as effective deep feature extractors in various studies [36,45,46]. CNNs excel at capturing global features from images by guiding the input through multiple convolutional filters and progressively reducing dimensionality across various architectural stages. For our experiments, we selected several pre-trained off-the-shelf architectures based on the Imagenet1k dataset [47]. Detailed information regarding the chosen

Table 3
Specifications of employed CNNs, including the reference paper, number of trainable parameters in millions, input shape, and feature extraction layer.

Reference	Parameters (M)	Input shape	Feature layer
AlexNet [48]	60	224 × 224	Pen. FC
GoogLeNet [49]	5	224 × 224	Loss3
ResNet-18 [50]	11.7	224 × 224	Pool5
ResNet-50 [50]	26	224 × 224	Avg. Pool
ResNet-101 [50]	44.6	224 × 224	Pool5
Inception-v3 [51]	21.8	299 × 299	Last FC
Inception-ResNet-v2 [52]	55	299 × 299	Avg. pool
DarkNet-53 [53]	20.8	224 × 224	Conv53
DenseNet-201 [54]	25.6	224 × 224	Avg. Pool
EfficientNet B0 [55]	5.3	224 × 224	Avg. Pool

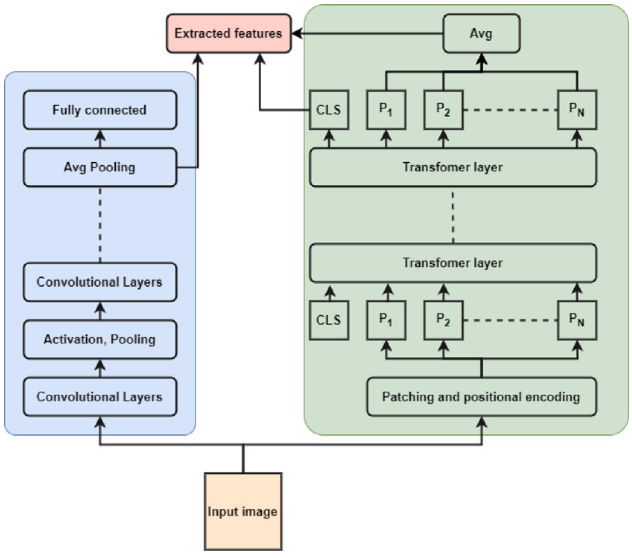


Fig. 4. Schematic representation of the differences in feature extraction methodologies between CNNs and ViTs.

layers for feature extraction, input size, and the number of trainable parameters for each CNN can be found in Table 3. A brief description is provided hereafter.

A schematic representation of the differences between ViT and CNN-based feature extraction approaches is depicted in Fig. 4

3.2.3. Handcrafted features

HC image features encompass a wide range of techniques and methodologies utilized for extracting morphological, pixel-level, and textural information from an image. According to [56], these features can be further categorized into three primary categories: invariant moments, textural features, and color-based features. Let us briefly describe them and indicate the specific descriptors adopted.

Invariant moments - An image moment refers to a weighted average, denoted as the moment, of the pixel intensities within an image, employed for the extraction of specific properties. Moments find application in image analysis and pattern recognition to characterize segmented objects. In the present study, three distinct types of moments, namely, Zernike, Legendre, and Chebyshev, were utilized. A concise overview of these moment types follows.

Chebyshev Moments (CH). Firstly introduced by [57], these are a collection of orthogonal moments derived from Chebyshev polynomials [58]. We utilized both first-order and second-order moments, denoted as CH and CH_due, respectively. Specifically, we calculated both CH and CH_due of order 5.

Second-order Legendre Moments (LGMS). Initially proposed by [59], these moments are derived from Legendre orthogonal polynomials [60] and are employed to represent the shape and spatial characteristics of objects within an image. In our analysis, we utilized Legendre moments of order 5.

Zernike Moments (ZM). Initially introduced by [61], these are another set of orthogonal moments derived from Zernike polynomials, and they are employed to describe the shape and structure of objects in an image. We applied Zernike moments of order 6 with a repetition of 4.

Texture features - They were a focal point of evaluation in this proposal, particularly emphasizing fine textures. The adopted texture features are following described.

Haar Features (Haar). These features consist of adjacent rectangles with alternating positive and negative polarities, taking forms like edge features, line features, four-rectangle features, and center–surround features. The computation of Haar features is often facilitated by using an integral image, which allows for the rapid calculation of pixel value sums within rectangular regions. Notably, Haar features are integral to cascade classifiers, a key component of the Viola–Jones object detection framework, where a series of stages employ subsets of Haar features to efficiently determine the presence of a target object. [62].

Rotation-Invariant Haralick Features (HARri). Thirteen Haralick features [63] were derived from the Gray Level Co-occurrence Matrix (GLCM), and subsequently transformed into rotation-invariant features (refer to [64] for details). To achieve rotation invariance, the computation involved four variations of the GLCM, each with parameters set to $d = 1$ and angular orientations $\theta = [0°, 45°, 90°, 135°]$.

Local Binary Pattern (LBP). This technique characterizes texture and patterns within an image, as described by [65]. In this work, we computed the histogram of the LBP, converted to a rotation invariant form, viz. LBP_ri [66], was extracted and used as the feature vector. The LBP map was generated within a neighborhood defined by a radius $r = 1$ and a number of neighbors $n = 8$.

Color features - These kinds of features aim at extracting color intensity information from the images. In this study, these descriptors were calculated from images that underwent a conversion to grayscale, streamlining the process of analysis and computation.

Grayscale Histogram Features (Hist). The color histogram characterizes the global color distribution within the image. We computed seven statistical descriptors from it, including mean, standard deviation, smoothness, skewness, kurtosis, uniformity, and entropy.

3.3. Feature selection

Feature selection is a crucial step in machine learning and data analysis, as it involves identifying and choosing the most informative attributes from a dataset while discarding irrelevant or redundant ones. The significance of feature selection lies in its ability to enhance model performance, reduce computational complexity, and mitigate the risk of overfitting. Selecting a subset of relevant features, not only improves model interpretability but also speeds up the training process and enhances generalization to unseen data [67–69]. Univariate feature selection using ranking is a technique that assesses individual features' importance based on their statistical properties and assigns each feature a ranking score. By ranking features according to their relevance to the target variable. In our experiments, we employed some common ranking methods, which include:

Chi-squared (chi2). It measures the statistical dependence between a categorical target variable and each categorical or discrete feature in the dataset using the chi-squared statistic.

ANOVA F-statistic (f_classif). The f_classif employs the ANOVA F-statistic to evaluate the variance between different classes and the variance within those classes.

Mutual Information (mutual_info_classif). This strategy quantifies the mutual information between a feature and a categorical target variable. It measures the dependency between features and the target, capturing any relationship, be it linear or nonlinear.

3.4. Feature normalization

Feature normalization is a crucial preprocessing step in ML and data analysis. It involves transforming the numerical features in a dataset to a common scale or distribution. Its primary purpose is to ensure that the features have comparable magnitudes, preventing certain variables from dominating others during model training [70]. In our experiments, the following feature normalization methods were used:

Normalizer. This technique revolves around normalizing the input data along unitary norms, usually the L1 or L2 norms.

Standard Scaler. It standardizes attributes by centering data around zero and scaling to unit variance. It does so by subtracting the mean and dividing by the standard deviation. While effective for normalization, it can overemphasize outliers, leading to narrow inlier distributions.

MinMaxScaler. It is employed to translate and normalize data within a predetermined range, typically spanning from 0 to 1. This process involves subtracting the minimum value from each data point and then dividing by the range. In the absence of outliers, its impact is similar to that of the standard scaler. Nevertheless, in scenarios where outliers are present, the min–max scaler may exhibit limitations in achieving parity in means and variances across distributions.

MaxAbsScaler. This scaler scales each example by the maximum absolute value in its attribute. This makes it sensitive to outliers, resulting in narrower distributions and unequalized means.

RobustScaler. It transforms the data, centering on the median and scaling using the interquartile range. This equalizes variances across attributes robustly, even with outliers, since scaling uses the interquartile magnitude [70].

QuantileTransformer. This scaler non-linearly maps data to uniform distributions, such as normal and uniform, using quantiles. It transforms attributes independently, making it robust to outliers.

PowerTransformer. It employs a power transformation, specifically the Yeo-Johnson transform in our case study, on a feature-wise basis to induce a Gaussian-like distribution in the data. This approach proves advantageous in addressing modeling challenges associated with non-constant variance or instances where a normal distribution is sought after.

3.5. Machine learning strategies

The HC and deep-learning extracted features employed in this study are used as inputs of several ML classifiers, such as Random Forest Classifier (RF), k-nearest Neighbor (k-NN), Support Vector Machine (SVM), Gradient Boosting Classifier (GB), and a Stacked classifier (SC) approach. Let us briefly introduce them.

k-Nearest Neighbor. The k-NN classifier makes categorical determinations based on the classes of the k-training examples that are nearest in distance to a given observation. By considering the proximity of neighboring instances, this method employs a local strategy to classify observations.

Support Vector Machine. This classifier discern categories by mapping examples to specific sides of a decision boundary. Here, a radial basis function kernel is employed to handle non-linear relationships, allowing for a more nuanced representation of complex data patterns. The one-vs-rest approach is applied to address multiclass problems, training individual classifiers to distinguish each class from the rest.

Random Forest. It amalgamates the predictions of numerous decision trees, each trained on random subsets of features and examples. This ensemble technique enhances model robustness by introducing diversity among the trees, contributing to improved resilience against data imbalance and mitigating overfitting. Notably, the specified inclusion of 100 trees further bolsters the predictive power of the random forest.

Gradient Boosting Classifier. Operating through a sequential ensemble-building process, the Gradient Boosting Classifier constructs a series of weak learners, typically decision trees. Each successive tree aims to rectify the errors of its predecessor, thereby iteratively refining predictive accuracy. This method excels in capturing intricate relationships within the data, making it particularly effective in scenarios with complex and non-linear patterns.

Stacked Classifier. The stacked classifier aims to increase classification accuracy by integrating predictions from diverse base classifiers. This is achieved through the training of a meta-classifier on the outputs of the individual classifiers, harnessing the collective intelligence of various algorithms. By synergizing the strengths of multiple approaches, the stacked classifier produces a comprehensive and robust classification outcome.

3.6. Performance evaluation measures

The classification performance has been measured in terms of accuracy, precision, recall, and F-measure. Clear explanations of these measures for binary classification tasks are provided below, followed by their extensions to encompass multiclass scenarios. To assess a binary classifier's performance on a dataset, each instance within the dataset will be categorized as either negative or positive based on the classifier's predictions. The classification result and the true target value will determine whether an instance contributes to one of the following measures:

True Negatives (TN). The count of instances from the negative class that have been accurately predicted as negative.

False Positives (FP). The count of instances from the negative class that have been erroneously predicted as positive.

False Negatives (FN). The count of instances from the positive class that have been erroneously predicted as negative.

True Positives (TP). The count of instances from the positive class that have been correctly predicted as positive.

Using these quantities, the aforementioned measures can be defined as follows:

Accuracy is a measure of how often the classifier correctly predicts both positive and negative instances. It provides a general overview of the classifier's performance, measuring the overall correctness of predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision (also known as Positive Predictive Value) is the ratio of correctly predicted positive instances to the total instances predicted as positive. It measures how accurate the classifier is when it predicts the positive class, focusing on minimizing FPs.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall (also known as Sensitivity or True Positive Rate) is the ratio of correctly predicted positive instances to the total actual positive instances. It quantifies how well the classifier identifies positive instances and aims to minimize false negatives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

F-measure, or F-score, is the harmonic mean of precision and recall. It provides a balance between the two.

$$\text{F-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

When extending these measures to multiclass classification, we consider the concept of macro averaging. In a multiclass scenario, the macro average calculates the measure for each class independently and then takes the average of these individual class measures. This approach provides a balanced evaluation across all classes.

Accuracy Macro Avg. It evaluates the overall correctness of the classifier's predictions by computing the average of the accuracy calculated for each class separately.

$$\text{Accuracy Macro Avg} = \frac{1}{C} \sum_{i=1}^{C} \text{Accuracy}_i \tag{5}$$

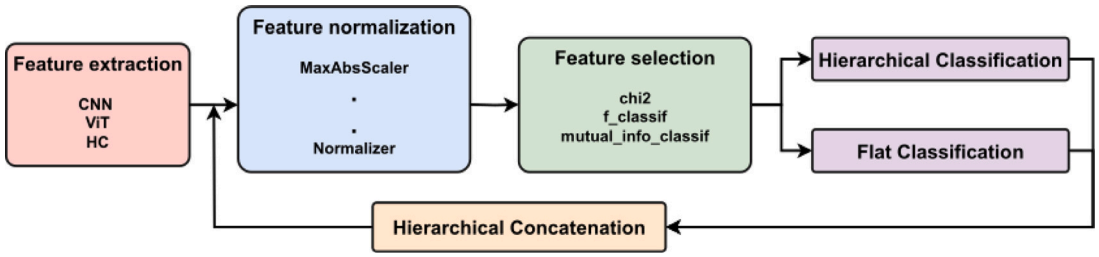


Fig. 5. Schematic representation of the steps involved in our experimental process, representing in one schema the whole possible range of experiments, in the figure as in our experiments the Hierarchical Concatenation is an optional step.

Precision Macro Avg. The precision macro average measures the average precision of the classifier across all classes.

$$\text{Precision Macro Avg} = \frac{1}{C}\sum_{i=1}^{C}\text{Precision}_i \tag{6}$$

Recall Macro Avg. The recall macro average calculates the average recall across all classes.

$$\text{Recall Macro Avg} = \frac{1}{C}\sum_{i=1}^{C}\text{Recall}_i \tag{7}$$

F-Score Macro Avg. The F-Score macro average is the mean of the F-Scores computed for each class.

$$\text{F-Score Macro Avg} = \frac{1}{C}\sum_{i=1}^{C}\text{F-Score}_i \tag{8}$$

Algorithm 1: Hierarchical Feature Concatenation

Data: FeatureSets
Result: MergedFeatureSet with the highest F-Score
for ***each*** *SingularFeatureSet* ***in*** *FeatureSets* **do**
  Classify SingularFeatureSet;
while *number of FeatureSets is greater than 1* **do**
  for ***each*** *Pair of FeatureSets* $< set_n, set_{n+1} >$ **do**
    MergedFeatureSet ← MergeFeatures(set_n, set_{n+1});
    Classify(MergedFeatureSet);
    if *F-Score of MergedFeatureSet is greater than max(F-Score of* set_n*, F-Score of* set_{n+1}*)* **then**
      Keep MergedFeatureSet;
    else
      Keep the FeatureSet with the higher F-Score between set_n and set_{n+1};

4. Proposed approach

The primary objectives of our study revolve around the assessment of hierarchical classification techniques in the context of ripeness classification. Noteworthy precedents in the literature have employed such methodologies to enhance conventional approaches [71,72]. In our investigation, first, a flat classification approach is employed, wherein the simultaneous categorization of the three distinct cheese types into two ripeness levels is treated as a comprehensive 6-class classification problem. Then, a hierarchical approach is adopted, wherein initial training of classifiers is conducted to classify cheese wheels into the three distinct cheese-type categories. Subsequently, binary classifiers are trained for each category to discern the ripeness level within that specific cheese type.

The second aim of our study is to ascertain an effective and semantically robust method for aggregating features from diverse sources, as elucidated in this manuscript. The exhaustive exploration of all feasible feature combinations through a naive, full search approach yields $2^n - 1$ combinations, rendering it impractical and, due to computational constraints, infeasible in our scenario where n equals 32. To address this challenge, we propose a streamlined approach that hierarchically transforms the problem into a tree-like structure of potential combinations. This method necessitates only $n - 1$ steps to converge towards a locally optimal solution.

Table 4
Hyperparameters of classification algorithms used throughout the experimental procedures.

Classifier	Hyperparameter	Value
k-NN	`number of neighbors`	5
	`distance`	Euclidean distance
SVM	`C`	1.0
	`kernel`	rbf
	`tipology`	one-versus-rest
Random Forest	`n. estimators`	100
	`criterion`	gini-score
Gradient Boosting Classifier	`n. estimators`	100
	`learning_rate`	0.1
Stacked Classifier	`final estimator`	logistic regression
	`estimators`	SVM, RF, k-NN

Our methodology first extracts a combination of a classifier, a feature selection algorithm, and a feature scaler. Subsequently, individual sets of features undergo independent preprocessing, involving scaling and filtering, followed by classification. The iterative process involves concatenating features in pairs and repeating the preprocessing step. Ultimately, features are classified, and consideration is given to retaining only those among the single elements and concatenated pairs exhibiting the highest classification measures, notably the F-Score in our analysis. The iterative procedure persists until only a singular set of features remains.

Our final setup for the second step is carried out by repeating the experiments k times, with each trial involving the shuffling of the order of features. A schematic representation of the pseudocode is presented in Algorithm 1. This algorithm distinguishes itself through its efficiency and efficacy in navigating the intricate domain of feature synergies. Employing an iterative refinement of feature combinations through a lightweight and hierarchical methodology, our algorithm not only automates the feature selection process but also furnishes a data-driven mechanism for discerning optimal feature sets. This innovative approach contributes substantively to the wider field of machine learning by presenting a systematic methodology for identifying the most influential features in the context of cheese classification. Furthermore, its applicability extends beyond our specific domain, offering a promising avenue for feature exploration in diverse domains.

A schematic representation of the experimental procedure is delineated in Fig. 5, elucidating each step of the experimentation process. Notably, the incorporation of our innovative hierarchical feature concatenation strategy is introduced as an optional iterative refinement step.

Finally, a visual representation of the Flat and Hierarchical classification approaches is provided in Fig. 6, manifesting in a tree-like structure that partitions the dataset classes hierarchically. This representation accentuates the divergence in the generated classifiers for

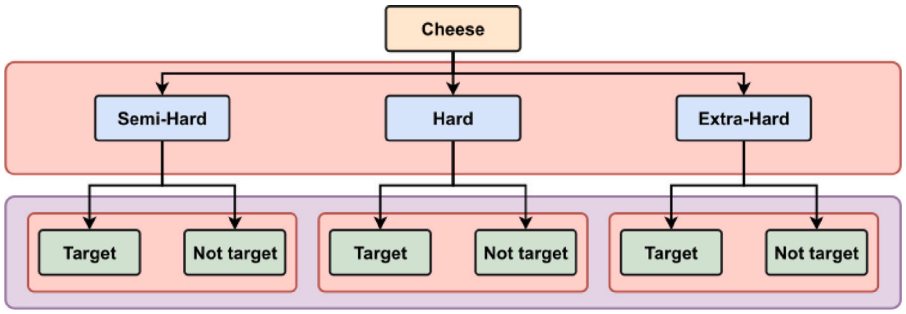


Fig. 6. Schematic representation depicting the classifiers generated as part of our experimental procedure. Classes within the flat classifier are highlighted in purple, while classes at each level of the hierarchical classifier are highlighted in red.

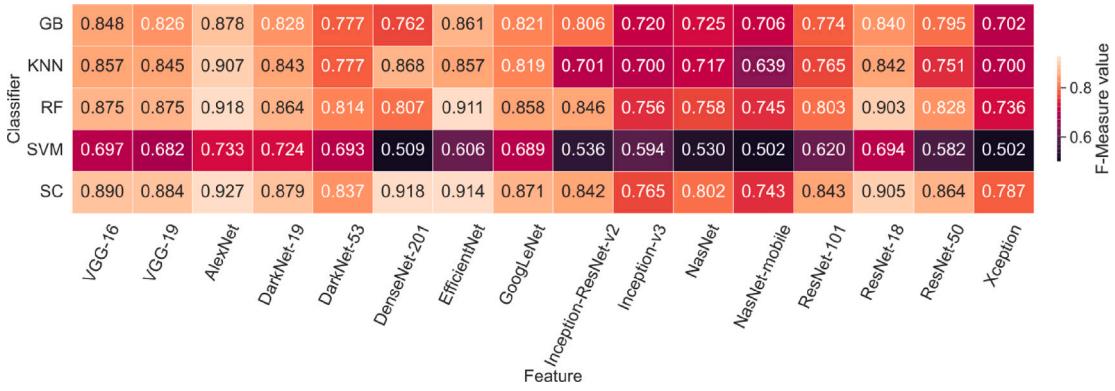


Fig. 7. Comparative analysis of ML classifiers' performance trained with different CNN-extracted features using a hierarchical classification strategy. The heatmap illustrates aggregated F-score values obtained by averaging F-score values across different setups encompassing all combinations of employed feature selectors, selected feature counts, and feature scalers.

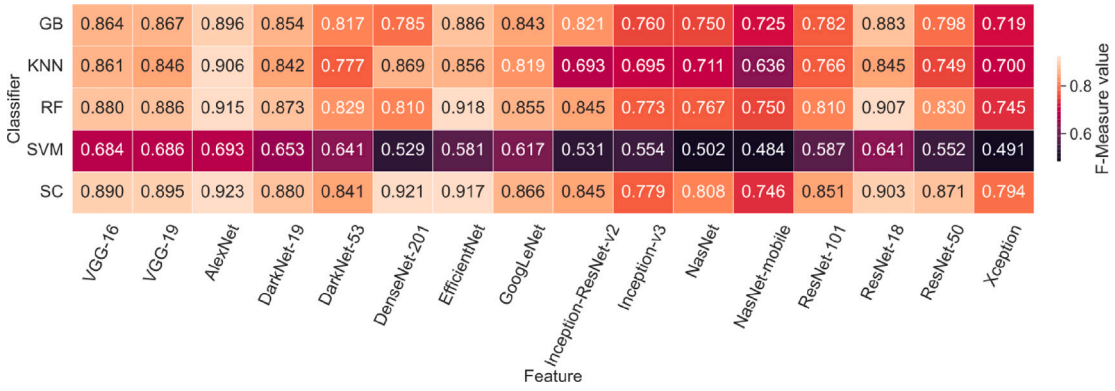


Fig. 8. Comparative analysis of ML classifiers' performance trained with different CNN-extracted features using a flat classification strategy. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature selection methods, feature counts, and feature scaling techniques employed.

each approach, thereby facilitating a nuanced comprehension of their distinctions.

5. Experimental results

We present the outcomes of our experiments employing the proposed methodology on CHEESE-HIDB. In Section 5.1, an in-depth elucidation is provided regarding the varied classification performances of features, encompassing both deep-based and HC features. Section 5.2 undertakes a comparative analysis between the flat and hierarchical approaches. Subsequently, Section 5.3 and Section 5.4 inspect the impact of feature normalization and feature selection, respectively, on the classification performance. Section 5.5 offers a comprehensive exposition of the outcomes derived from the hierarchical feature concatenation process. Finally, Sections 5.6 and 5.7 presents a comparison with methods found in literature on similar contexts and with other feature aggregation methods, respectively.

The experiments were executed on a workstation featuring an Intel(R) Core(TM) i7-12700 @ 2.1 GHz CPU, 16 GB RAM, and an NVIDIA GTX1660 Super GPU with 6 GB of memory. All classifiers utilized in the experiments were imported from the scikit-learn library [73]. A comprehensive list of selected hyperparameters is provided in Table 4.

5.1. Influence of features on classification performance

5.1.1. CNN-extracted features results

Here, we present the results obtained from the classification of features extracted by CNNs. Elucidating trends in F-measure, the heatmaps presented in Figs. 7 and 8 delineate performance variations across

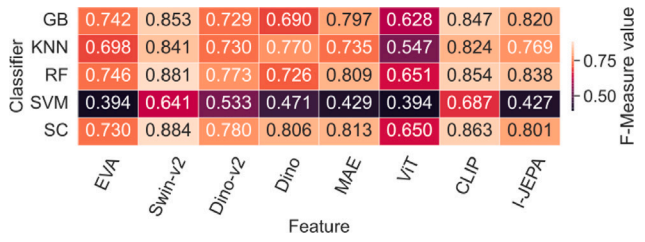


Fig. 9. Comparative analysis of ML classifiers' performance trained with different ViT-extracted features using a hierarchical classification strategy. The heatmap presents a trend of ML classifiers performance with their aggregated F-score values computed by averaging F-score values across different setups encompassing all combinations of employed feature selectors, selected feature counts, and feature scalers.

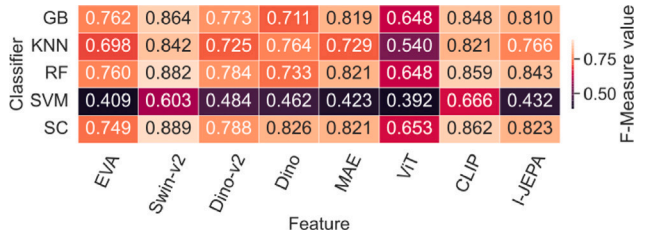


Fig. 10. Comparative analysis of ML classifiers' performance trained with different ViT-extracted features using a flat classification strategy. The heatmap presents a trend of the ML classifiers F-score performance computed by averaging F-score values across different setups encompassing all combinations of employed feature selectors, selected feature counts, and feature scalers.

various classifiers in the two analyzed settings. For conciseness, we offer an overview of average results, accommodating variations in feature scaling and ranking algorithms.

Discernible patterns emerge from the presented tables, indicating that the most resilient features are associated with specific architectures, including AlexNet, EfficientNet, and ResNet-18. Notably, the results appear to exhibit a positive correlation with shallower architectures, exemplified by AlexNet, or those characterized by fewer parameters in comparison to others, such as EfficientNet and ResNet-18, with respective parameter counts of 5.3 million and 11.7 million.

5.1.2. ViT-extracted features results

In this subsection, we present the outcomes derived from the classification of features extracted by the Vision Transformer. A comprehensive depiction of trends in F-measure is provided in Figs. 9 and 10 across diverse classifiers. For brevity, we present solely the average results, accounting for variations in feature scaling and ranking algorithms.

The discerned patterns from both figures indicate superior performance for Swin-v2 and CLIP features compared to other extracted features. This observation can be elucidated by two crucial aspects: Swin-v2, owing to its architectural design, adeptly captures hierarchical structures within images, resulting in a denser and more enriched representation. In contrast, CLIP, built on the ViT architecture, although lacking an inherent hierarchical representation, divides images into larger 32×32 patches—largest among the studied architectures. This results in fewer patches containing more information, contributing to the extraction of valuable features.

Noteworthy among various classifiers are the commendable results exhibited by both Swin-v2 and CLIP features, even with the Support Vector Machine classifier, which otherwise demonstrates suboptimal performance across other features. Significantly, the Stacked approach emerges as a standout among diverse classifiers, consistently yielding optimal results by effectively leveraging the strengths of the selected base classifiers.

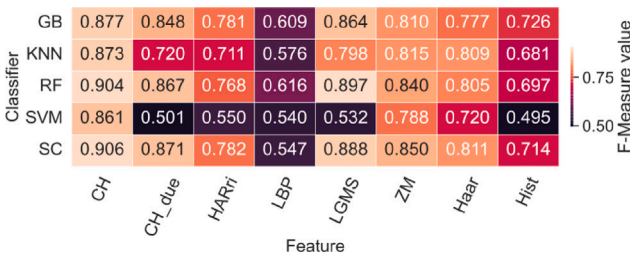


Fig. 11. Comparative analysis of ML classifiers' performance trained with different HC features using a hierarchical classification strategy. The heatmap presents a trend of ML classifiers performance with their aggregated F-score values computed by averaging F-score values across different setups encompassing all combinations of employed feature selectors, selected feature counts, and feature scalers.

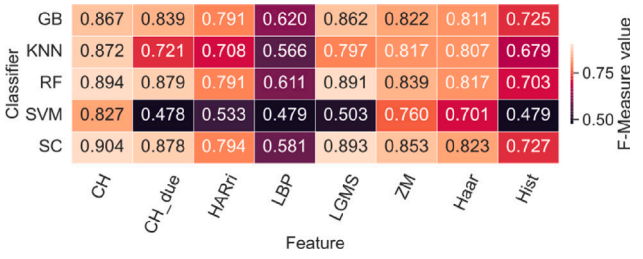


Fig. 12. Comparative analysis of ML classifiers' performance trained with different HC features using a flat classification strategy. The heatmap presents a trend of the ML classifiers F-score performance computed by averaging F-score values across different setups encompassing all combinations of employed feature selectors, selected feature counts, and feature scalers.

5.1.3. Handcrafted features results

In addition to examining features extracted by ViTs and CNNs, we present the outcomes derived from the classification of HC features. Maintaining methodological consistency, the heatmaps depicted in Figs. 11 and 12 adhere to the previously explained structure. Analysis of the presented heatmaps reveals superior performance associated with the utilization of invariant moments and first-order statistics. Conversely, textural features such as LBP exhibit limited discriminative capability in the context of the studied case across the array of selected classifiers.

5.2. Comparison between flat and hierarchical classification results

The previously discussed results have showcased the remarkable achievements of both flat and hierarchical classification approaches. To delve deeper into our analysis, Figs. 13–15 present the differences between the result values obtained from the hierarchical and flat classification approaches. Positive results indicate the dominance of the hierarchical approach, while negative values suggest the superiority of the flat one.

The outcomes exhibit a strong dependency on the selected features, where the type of feature significantly influences the approach either positively or negatively. Moreover, when scrutinizing the impact of different classifiers on the approach, it becomes evident that classifiers appear to be approach-invariant, showing no clear improvement or decrease in performance. Except for the SVM classifier, which exhibits a positive influence from the hierarchical classification approach, clear examples can be drawn from LBP results. When using the SVM classifier, the hierarchical approach achieves, on average, a 6% higher performance compared to its flat counterpart. The best ten results achieved across the whole set of experiments are depicted in Table 5.

5.3. Impact of the feature scalers

In this section, we delve into a detailed exploration of our observations centering on the effectiveness of various selected feature scalers across the extensive set of experiments.

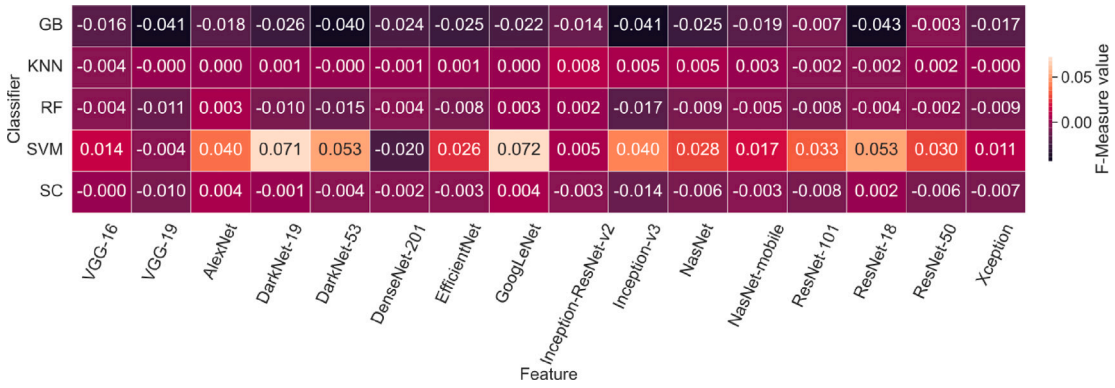


Fig. 13. Comparative heatmap of aggregated ML classifiers F-scores: flat vs. hierarchical classification strategies with CNN-extracted features. The heatmap presents a comparative analysis of the aggregated F-score performance the flat classification reported in Fig. 8, and the hierarchical classification detailed in Fig. 7, providing an overview of their performance trends.

Table 5
Performance measures for hierarchical and flat classification experiments. The results include F-measure, Accuracy, Precision, and Recall scores, along with standard deviations for each experimental setup. Classifiers, feature selectors, feature scalers, and the number of features are specified. For brevity, mutual_info_classif feature selection criteria are abbreviated as m.i.c and QuantileTransformer as QT. Hierarchical approach classifiers are denoted with an H- prefix.

F-measure	Accuracy	Precision	Recall	Classifier	Feature selector	Feature scaler	Feature number	Feature name
0.987 ± 0.038	0.987 ± 0.038	0.984 ± 0.018	0.987 ± 0.038	SVM	m.i.c	QT(Normal)	300	DenseNet-201
0.987 ± 0.038	0.987 ± 0.038	0.984 ± 0.018	0.987 ± 0.038	SVM	m.i.c	QT(Uniform)	150	EfficientNet
0.987 ± 0.038	0.987 ± 0.038	0.984 ± 0.018	0.987 ± 0.038	SC	m.i.c	QT(Normal)	150	EfficientNet
0.987 ± 0.038	0.987 ± 0.038	0.984 ± 0.018	0.987 ± 0.038	SVM	chi2	QT(Normal)	100	EfficientNet
0.987 ± 0.038	0.987 ± 0.038	0.984 ± 0.018	0.987 ± 0.038	H-SC	m.i.c	QT(Normal)	100	EfficientNet
0.987 ± 0.038	0.987 ± 0.038	0.984 ± 0.018	0.987 ± 0.038	SVM	f_classif	QT(Normal)	100	EfficientNet
0.986 ± 0.042	0.987 ± 0.038	0.982 ± 0.018	0.987 ± 0.038	H-RF	f_classif	QT(Uniform)	100	ResNet-50
0.985 ± 0.038	0.985 ± 0.038	0.982 ± 0.018	0.985 ± 0.038	RF	m.i.c	Normalizer	50	ResNet-50
0.985 ± 0.038	0.985 ± 0.038	0.982 ± 0.018	0.985 ± 0.038	H-SVM	f_classif	QT(Normal)	100	EfficientNet
0.985 ± 0.038	0.985 ± 0.038	0.982 ± 0.018	0.985 ± 0.038	H-SVM	chi2	QT(Normal)	100	EfficientNet

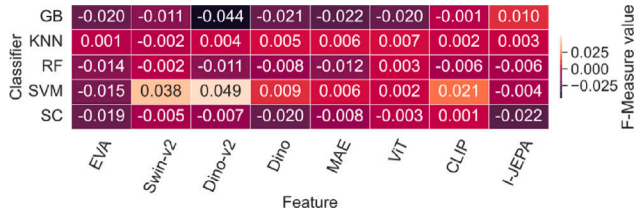


Fig. 14. Comparative heatmap of aggregated ML classifiers F-scores: flat vs. hierarchical classification with ViT-extracted features. The heatmap presents a comparative analysis of the aggregated F-score performance between the flat classification reported in Fig. 10, and the hierarchical classification presented in Fig. 9, providing an overview of their performance trends.

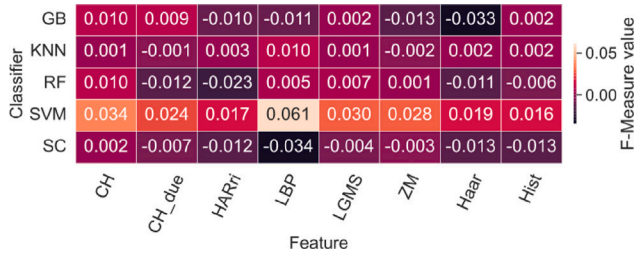


Fig. 15. Comparative heatmap of aggregated ML classifiers F-scores: flat vs. hierarchical classification with HC features. The heatmap presents a comparative analysis of the aggregated F-score performance between the flat classification presented in Fig. 12, and the hierarchical classification reported in Fig. 11, providing an overview of their performance trends.

Throughout the experimentations, two standout feature scalers, namely the PowerTransformer and QuartileTransformer, exhibited remarkable improvements in the obtained results. The impact was particularly pronounced, with instances where the gap between these preferred scalers and alternatives like MaxAbsScaler reached a significant margin exceeding 24% in terms of F-measure.

The PowerTransformer and QuartileTransformer share a common underlying principle, they both contribute to enhancing the distribution characteristics of the initial data. By transforming the features into Gaussian-like or normal distributions, these scalers meticulously address each attribute individually. This tailored treatment of attributes, combined with the distributional shift, may have played a pivotal role in elevating the overall classification performance.

The results of this analysis are reported in Figs. 16–18 showing the performance with the features extracted from CNN, ViT, and HC features, respectively.

5.4. Impact of the feature selectors

Here, we present the observations and the results that focus on the efficacy of the different feature selectors, varying the ranking score across the whole set of experiments. Our results show that f_classif and mutual_info_classif outperform chi2 criteria in several scenarios. This behavior can be explained by an implicit non-linear correlation between features, which is not discernible by feature ranking using chi2 as a criterion.

The impact of different feature selection techniques on model performance is presented in the heatmaps shown in Figs. 19–21. These heatmaps depict the performance results when using features extracted from CNN and ViT, and HC features, respectively.

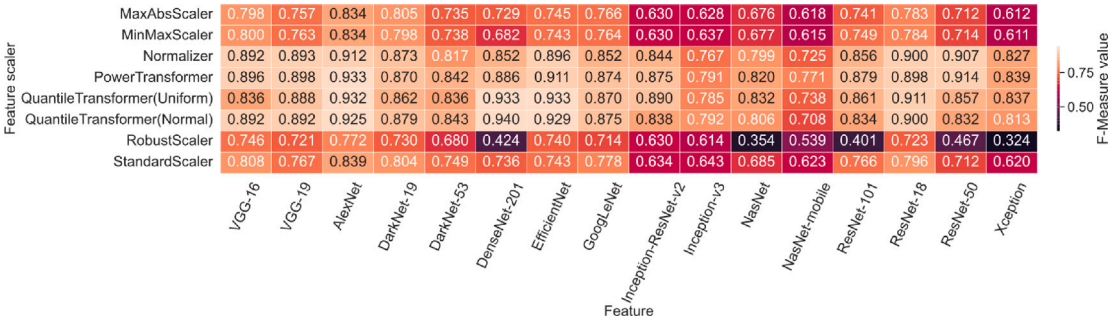


Fig. 16. Comparative analysis of performance obtained with the ML classifiers trained with different CNN-extracted features using a hierarchical classification strategy, from the feature scaler technique perspective. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature selection methods, feature counts, and ML classification methods employed.

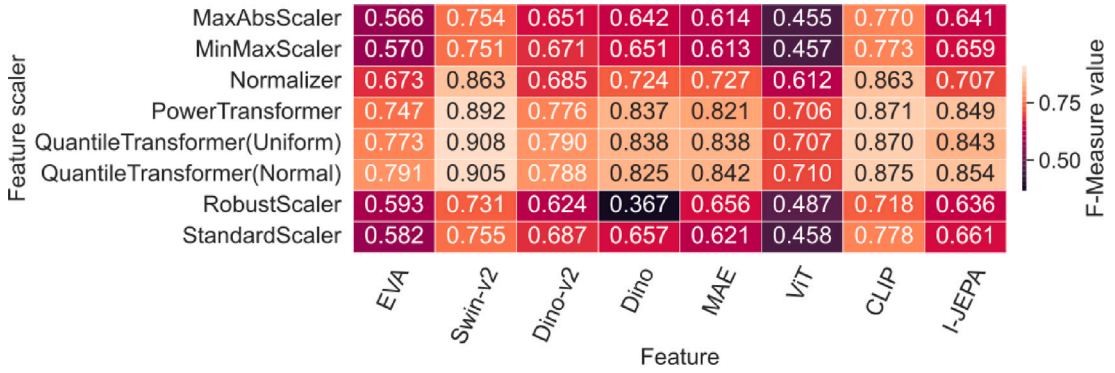


Fig. 17. Comparative analysis of performance obtained with the ML classifiers trained with different ViT-extracted features using a hierarchical classification strategy, from the feature scaler technique perspective. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature selection methods, feature counts, and ML classification methods employed.

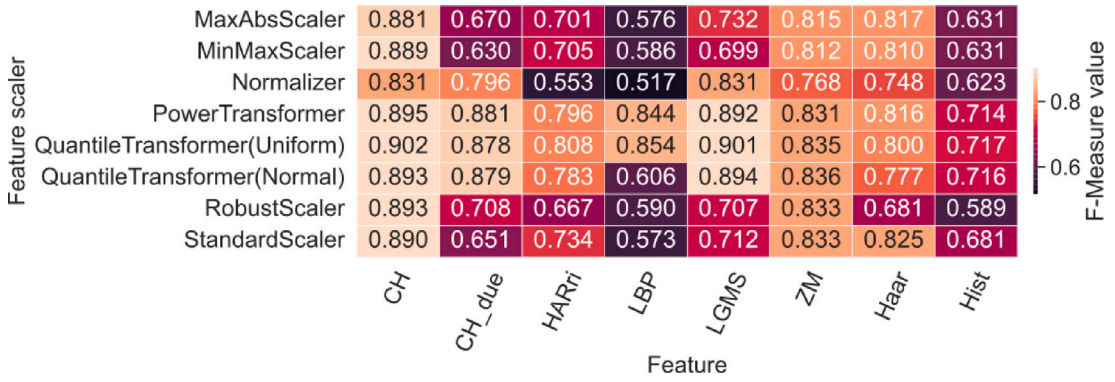


Fig. 18. Comparative analysis of performance obtained with the ML classifiers trained with different HC features using a hierarchical classification strategy, from the feature scaler technique perspective. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature selection methods, feature counts, and ML classification methods employed.

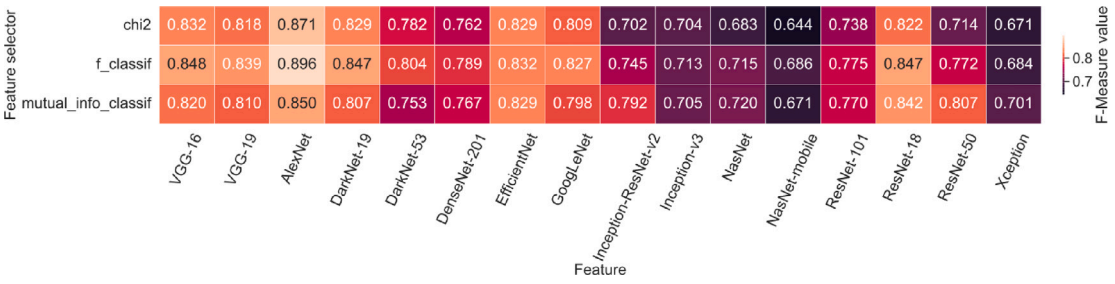


Fig. 19. Comparative analysis of performance obtained with the ML classifiers trained with different CNN-extracted features using a hierarchical classification strategy, from the feature selection technique perspective. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature scaler methods, feature counts, and ML classification methods employed.

Table 6
Results for hierarchical feature concatenation approach. The table displays F-measure scores with standard deviations for each setup, including classifier, feature selector, feature scaler, maximum feature count, and feature names. For clarity, mutual_info_classif is abbreviated as m.i.c and QuantileTransformer as QT.

F-measure	Classifier	Feature selector	Feature scaler	Feature number	Feature list
0.987 ± 0.038	SVM	chi2	QT(Uniform)	300	HARri, EfficientNet
0.987 ± 0.038	RF	m.i.c	MinMaxScaler	300	DenseNet-201
0.987 ± 0.038	SVM	chi2	QT(Normal)	300	EfficientNet, DarkNet-53
0.987 ± 0.038	SC	m.i.c	QT(Normal)	150	EfficientNet
0.987 ± 0.038	SVM	m.i.c	QT(Normal)	150	EfficientNet
0.987 ± 0.038	SC	f_classif	QT(Uniform)	150	ResNet-50, ResNet-18, HARri
0.991 ± 0.040	SC	m.i.c	QT(Normal)	100	EfficientNet, DarkNet-53
0.987 ± 0.038	SC	m.i.c	QT(Normal)	100	EfficientNet
0.987 ± 0.038	SVM	m.i.c	QT(Normal)	100	EfficientNet
0.987 ± 0.038	SC	m.i.c	QT(Normal)	100	EfficientNet

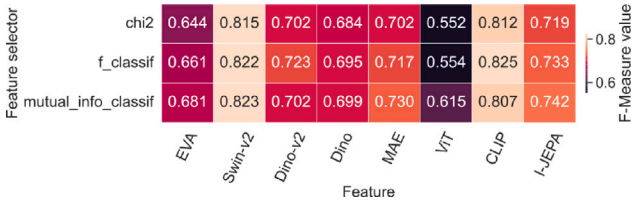


Fig. 20. Comparative analysis of performance obtained with the ML classifiers trained with different ViT-extracted features using a hierarchical classification strategy, from the feature selection technique perspective. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature scaler methods, feature counts, and ML classification methods employed.

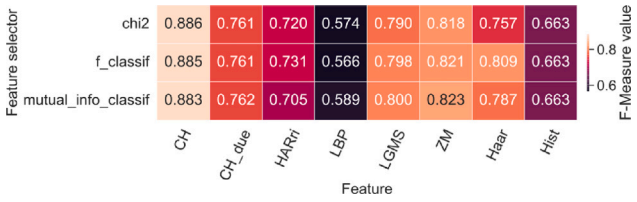


Fig. 21. Comparative analysis of performance obtained with the ML classifiers trained with different HC features using a hierarchical classification strategy, from the feature selection technique perspective. The heatmap presents an aggregated analysis of the average F-score performance obtained by varying the feature scaler methods, feature counts, and ML classification methods employed.

5.5. Hierarchical feature concatenation results

The comprehensive set of experiments detailed in the preceding paragraphs encompassed a total of 38,400 distinct and unique experimental configurations. These experiments systematically investigated the influence of various classifiers, feature selectors, and scalers on the outcomes. Notably, the execution of this extensive experimental evaluation necessitated a computational duration of 30 days. Despite the noteworthy results obtained, accomplishing such an undertaking would pose a hard and time-consuming task.

Our hierarchical feature concatenation methodology yielded superior results, particularly with a modest value for the parameter k, specifically set at 50. The parameter k denotes the number of shuffles employed in the process. To maintain simplicity and coherence in our experiments, we employed the hierarchical classification approach, and the comprehensive set of experiments is presented in Table 6.

The findings elucidate a discernible pattern: QuantileTransformer emerges as the most robust and effective feature scaler for the employed features, while mutual_info_classif proves to be the optimal feature selection criterion. Notably, EfficientNet consistently preserves informative features for classification, featuring in 9 out of the top 10 experiments.

Experiments incorporating a larger number of features, specifically 300 and 150 features, exhibit a proclivity towards favoring feature combinations. Conversely, in instances of reduced selected features, the merging process tends to retain only individual groups of features without merging.

Significantly, the most favorable outcome arises from the amalgamation of features from EfficientNet and DarkNet-53, both falling under the CNN category. This optimal experiment achieves a remarkable F-measure of 0.991 with a standard deviation of 0.040, surpassing F-measure values attained in all other proposed experiments.

5.6. Comparison with literature

To the best of our knowledge, the proposed approach represents the first attempt to evaluate cheese ripening by using a hierarchical classification approach, while a previous method proposed by our team analyzes images acquired through a photographic camera [20], with accuracy up to 0.98 in the classification of four different ripeness stage of a *Pecorino* soft cheese produced by a Sardinia dairy company.

In contrast, alternative methods in the literature primarily focus on chemical and spectroscopic parameters or address various types of cheese [74,75] and diverse ripening durations [7,76]. For these reasons, a direct head-to-head comparison with the methods previously proposed is not feasible.

Nevertheless, some studies have reported quantitative performance measures. Specifically, Del Campo et al. introduced a Principal Component Analysis-based model capable of discriminating among four ripening stages of Emmental cheeses using mid-infrared spectroscopy. Their model achieved a 0.87 cross-validation accuracy based on 14 samples and a 0.57 accuracy on a separate test set of 14 samples. Notably, our work differs from theirs in that our characterization of ripening extends over a longer time frame [76].

Additionally, Soto-Barajas et al. proposed an artificial neural network trained on data related to the fatty acid composition and NIR spectra of sixteen milk mixtures. Their approach achieved 0.50 accuracy in discriminating among different cheese types and 1.00 accuracy in classifying unknown cheese samples based on ripening time [75].

Despite the similarities with these prior proposals, our work stands out as innovative due to its remarkable 0.99 classification accuracy in distinguishing between different ripeness phases with a hierarchical approach. Moreover, its non-invasive nature and end-to-end automation make it valuable for identifying potential issues within cheese storage, such as inappropriate temperature conditions.

5.7. Comparison with other feature aggregation methods

Our approach stands in contrast to the standard practice in various subfields of machine learning, particularly in radiomics, where feature combination is typically performed at a global level, followed by a

subsequent global feature selection process [77,78]. Recent advancements in feature fusion have focused on feature map-level fusion within deep neural networks. This particular approach can be leveraged not only for single-network fusion but also to incorporate feature maps originating from different network backbones [79]. Such methods aim to learn the optimal feature combination through backpropagation. An alternative approach to merging multiple feature map sources can be achieved through a hierarchical or pairwise strategy akin to our feature concatenation technique. Interestingly, this method has demonstrated strong performance not only in our work but also in the deep learning paradigm [80].

The distinguishing aspect of our proposed approach is its ability to capture and preserve the inherent hierarchical structure of the feature representations, which is often overlooked in traditional global feature combination and selection techniques. By employing a hierarchical classification framework, we can leverage the meaningful relationships and dependencies among the various feature subsets, leading to a more comprehensive and effective feature aggregation process. This hierarchical perspective allows for a more nuanced and informed feature fusion, ultimately contributing to the robust performance observed in our experiments.

6. Conclusion

This study proposed the first public dataset encompassing images of cheese wheels depicting various products at distinct stages of ripening, on which an extensive study was conducted. The conclusions derived from the study affirm the efficacy of the proposed methodology within the domain of cheese quality analysis. The combination of the hierarchical classification approach with a distinctive feature aggregation technique has yielded promising outcomes. The investigation has provided insights into the potential utility of ViTs, CNNs, and HC features for feature extraction, showcasing discernible performance variations across diverse classifiers. Notably, the hierarchical feature concatenation strategy emerges as a valuable contribution, presenting a streamlined yet potent approach to feature selection and combination. The experimental findings underscore the pivotal role of preprocessing steps, classifiers, and feature selection techniques in attaining optimal results.

Future research endeavors should prioritize the seamless integration of the proposed methodology into authentic cheese production environments, addressing practical challenges and ensuring adaptability in industrial settings. Further exploration into advanced hierarchical classification strategies, coupled with the incorporation of expert knowledge, holds the potential to fortify the system's robustness and accuracy. The development of user-friendly interfaces and the conduction of robustness analyses against adversarial conditions are essential for enhancing the applicability and reliability of technology-driven solutions within the dairy industry. Interdisciplinary collaboration with experts in fields such as food science, dairy technology, and computer vision can facilitate the development of comprehensive, domain-specific solutions tailored to the unique requisites of cheese quality analysis. Such collaborative efforts stand to propel advancements in non-destructive techniques for food quality analysis, contributing significantly to the enhancement of quality control and assurance standards across the broader food industry.

In light of our findings, we suggest a future direction for research involving the development of a modular framework for cheese classification. This framework could serve as a versatile tool, allowing practitioners to explore and customize the best combinations of features, classifiers, and preprocessing techniques for their specific cheese production contexts. Beyond automating ripeness detection, such a framework holds the potential to be adapted for broader applications in food quality assessment. This avenue of research encourages an iterative approach to refining and optimizing classification processes, thereby fostering technological integration and continuous improvement in the dairy industry.

Acronyms

The following acronyms are used in this manuscript:

CV	Computer Vision
RF	Random Forest
k-NN	k-Nearest Neighbor
SVM	Support Vector Machine
GB	Gradient Boosting
SC	Stacked Classifier
CNNs	Convolutional Neural Networks
ViT	Vision Transformers
TN	True Negatives
FP	False Positives
FN	False Negatives
TP	True Positives
NIR	Near-Infrared Spectroscopy
FTIR	Fourier Transform Infrared Spectroscopy
GLCM	Gray Level Co-occurrence Matrix

Code availability

The source code for both our hierarchical classification training process and hierarchical feature concatenation is provided and available at the following GitHub repository. The employed dataset, CHEESE-HIDB, is publicly available at the following GitHub repository.

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 December 30, 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 June 23, 2022 by the Italian Ministry of University and Research (MUR).

CRediT authorship contribution statement

Luca Zedda: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alessandra Perniciano:** Writing – review & editing, Writing – original draft, Validation, Investigation, Conceptualization. **Andrea Loddo:** Supervision, Investigation, Formal analysis, Conceptualization, Validation, Writing – original draft, Writing – review & editing. **Cecilia Di Ruberto:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Luca Zedda reports financial support was provided by University of Cagliari. Alessandra Perniciano reports financial support was provided by University of Cagliari. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared source code and image dataset.

References

[1] T. Lei, D.W. Sun, Developments of nondestructive techniques for evaluating quality attributes of cheeses: A review, Trends Food Sci. Technol. 88 (2019) 527–542.

[2] A. Forde, G.F. Fitzgerald, Biotechnological approaches to the understanding and improvement of mature cheese flavour, Curr. Opin. Biotechnol. 11 (5) (2000) 484–489.

[3] P.L. McSweeney, Biochemistry of cheese ripening, Int. J. Dairy Technol. 57 (2–3) (2004) 127–144.

[4] P. Franceschi, M. Malacarne, E. Bortolazzo, F. Coloretti, P. Formaggioni, A. Garavaldi, V. Musi, A. Summer, Automatic milking systems in the production of parmigiano reggiano cheese: Effects on the milk quality and on cheese characteristics, Agriculture 12 (1) (2022) http://dx.doi.org/10.3390/agriculture12010104, URL: https://www.mdpi.com/2077-0472/12/1/104.

[5] P. Fox, J. Wallace, S. Morgan, C. Lynch, E. Niland, J. Tobin, Acceleration of cheese ripening, Antonie van Leeuwenhoek 70 (2) (1996) 271–297.

[6] L. Sakkas, C.S. Pappas, G. Moatsou, FT-MIR analysis of water-soluble extracts during the ripening of sheep milk cheese with different phospholipid content, Dairy 2 (4) (2021) 530–541.

[7] A.R. Khattab, H.A. Guirguis, S.M. Tawfik, M.A. Farag, Cheese ripening: A review on modern technologies towards flavor enhancement, process acceleration and improved quality assessment, Trends Food Sci. Technol. 88 (2019) 343–360.

[8] J. Dias, P. Lage, A. Garrido, E. Machado, C. Conceição, S. Gomes, A. Martins, A. Paulino, M.F. Duarte, N. Alvarenga, Evaluation of gas holes in "Queijo de Nisa" PDO cheese using computer vision, J. Food Sci. Technol. 58 (2021) 1072–1080.

[9] I.R. Donis-González, D.E. Guyer, Classification of processing asparagus sections using color images, Comput. Electron. Agric. 127 (2016) 236–241, http://dx.doi.org/10.1016/J.COMPAG.2016.06.018.

[10] F. Wu, Z. Yang, X. Mo, Z. Wu, W. Tang, J. Duan, X. Zou, Detection and counting of banana bunches by integrating deep learning and classic image-processing algorithms, Comput. Electron. Agric. 209 (2023) 107827, http://dx.doi.org/10.1016/J.COMPAG.2023.107827.

[11] M.S. Hossain, M.H. Al-Hammadi, G. Muhammad, Automatic fruit classification using deep learning for industrial applications, IEEE Trans. Ind. Inf. 15 (2) (2019) 1027–1034, http://dx.doi.org/10.1109/TII.2018.2875149.

[12] J. Rong, H. Zhou, F. Zhang, T. Yuan, P. Wang, Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion, Comput. Electron. Agric. 207 (2023) 107741, http://dx.doi.org/10.1016/J.COMPAG.2023.107741.

[13] D.C. Hernández, R. Gutierrez, K. Kung, J. Rodriguez, O. Lao, K. Contreras, K. Jo, J.E. Sánchez-Galán, Recent advances in automatic feature detection and classification of fruits including with a special emphasis on Watermelon (Citrillus lanatus): A review, Neurocomputing 526 (2023) 62–79, http://dx.doi.org/10.1016/J.NEUCOM.2023.01.005.

[14] N. Saranya, K. Srinivasan, S.K.P. Kumar, Banana ripeness stage identification: a deep learning approach, J. Ambient Intell. Humaniz. Comput. 13 (8) (2022) 4033–4039, http://dx.doi.org/10.1007/S12652-021-03267-W.

[15] T. Lei, X.H. Lin, D.W. Sun, Rapid classification of commercial cheddar cheeses from different brands using PLSDA, LDA and SPA–LDA models built by hyperspectral data, J. Food Meas. Charact. 13 (2019) 3119–3129.

[16] P.S. Minz, C.S. Saini, Comparison of computer vision system and colour spectrophotometer for colour measurement of mozzarella cheese, Appl. Food Res. 1 (2) (2021) 100020.

[17] M. Alinovi, G. Mucchetti, F. Tidona, Application of NIR spectroscopy and image analysis for the characterisation of grated parmigiano-reggiano cheese, Int. Dairy J. 92 (2019) 50–58.

[18] R. Iezzi, F. Locci, R. Ghiglietti, C. Belingheri, S. Francolino, G. Mucchetti, Parmigiano reggiano and grana padano cheese curd grains size and distribution by image analysis, LWT 47 (2) (2012) 380–385.

[19] G. Mulas, R. Anedda, D. Longo, T. Roggio, S. Uzzau, An MRI method for monitoring the ripening of Grana Padano cheese, Int. Dairy J. 52 (2016) 19–25.

[20] A. Loddo, C. Di Ruberto, G. Armano, A. Manconi, Automatic monitoring cheese ripeness using computer vision and artificial intelligence, IEEE Access 10 (2022) 122612–122626, http://dx.doi.org/10.1109/ACCESS.2022.3223710, URL: https://ieeexplore.ieee.org/document/9956763/.

[21] T. Jeliński, C.J. Du, D.W. Sun, J. Fornal, Inspection of the distribution and amount of ingredients in pasteurized cheese by computer vision, J. Food Eng. 83 (1) (2007) 3–9, http://dx.doi.org/10.1016/j.jfoodeng.2006.12.020, URL: https://www.sciencedirect.com/science/article/pii/S0260877407000271.

[22] D. Guggisberg, P. Schuetz, H. Winkler, R. Amrein, E. Jakob, M.T. Fröhlich-Wyder, S. Irmler, W. Bisig, I. Jerjen, M. Plamondon, J. Hofmann, A. Flisch, D. Wechsler, Mechanism and control of the eye formation in cheese, Int. Dairy J. 47 (2015) 118–127, http://dx.doi.org/10.1016/j.idairyj.2015.03.001, URL: https://www.sciencedirect.com/science/article/pii/S0958694615000631.

[23] D. Huc, F. Mariette, S. Challois, J. Barreau, G. Moulin, C. Michon, Multi-scale investigation of eyes in semi-hard cheese, Innov. Food Sci. Emerg. Technol. 24 (2014) 106–112, http://dx.doi.org/10.1016/j.ifset.2013.10.002, URL: https://www.sciencedirect.com/science/article/pii/S1466856413001550.

[24] P. Schuetz, D. Guggisberg, I. Jerjen, M.T. Fröhlich-Wyder, J. Hofmann, D. Wechsler, A. Flisch, W. Bisig, U. Sennhauser, H.P. Bachmann, Quantitative comparison of the eye formation in cheese using radiography and computed tomography data, Int. Dairy J. 31 (2) (2013) 150–155, http://dx.doi.org/10.1016/j.idairyj.2012.12.007, URL: https://www.sciencedirect.com/science/article/pii/S0958694613000277.

[25] A. Kulmyrzaev, E. Dufour, Y. Noël, M. Hanafi, R. Karoui, E.M. Qannari, G. Mazerolles, Investigation at the molecular level of soft cheese quality and ripening by infrared and fluorescence spectroscopies and chemometrics—relationships with rheology properties, Int. Dairy J. 15 (6) (2005) 669–678, http://dx.doi.org/10.1016/j.idairyj.2004.08.016, URL: https://www.sciencedirect.com/science/article/pii/S0958694604002869.

[26] Z. Ozbekova, A. Kulmyrzaev, Fluorescence spectroscopy as a non destructive method to predict rheological characteristics of Tilsit cheese, J. Food Eng. 210 (2017) 42–49, http://dx.doi.org/10.1016/j.jfoodeng.2017.04.023, URL: https://www.sciencedirect.com/science/article/pii/S0260877417301784.

[27] R. Karoui, A.M. Mouazen, E. Dufour, L. Pillonel, E. Schaller, D. Picque, J. De Baerdemaeker, J.O. Bosset, A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European Emmental cheese, Eur. Food Res. Technol. 223 (1) (2006) 44–50, http://dx.doi.org/10.1007/s00217-005-0110-2.

[28] E.S. Madalozzo, E. Sauer, N. Nagata, Determination of fat, protein and moisture in ricotta cheese by near infrared spectroscopy and multivariate calibration, J. Food Sci. Technol. 52 (3) (2015) 1649–1655, http://dx.doi.org/10.1007/s13197-013-1147-z.

[29] M.L. Oca, M.C. Ortiz, L.A. Sarabia, A.E. Gredilla, D. Delgado, Prediction of Zamorano cheese quality by near-infrared spectroscopy assessing false non-compliance and false compliance at minimum permitted limits stated by designation of origin regulations, Talanta 99 (2012) 558–565, http://dx.doi.org/10.1016/j.talanta.2012.06.035, URL: https://www.sciencedirect.com/science/article/pii/S0039914012005024.

[30] C. Cevoli, A. Gori, M. Nocetti, L. Cuibus, M.F. Caboni, A. Fabbri, FT-NIR and FT-MIR spectroscopy to discriminate competitors, non compliance and compliance grated parmigiano reggiano cheese, Food Res. Int. 52 (1) (2013) 214–220, http://dx.doi.org/10.1016/j.foodres.2013.03.016, URL: https://www.sciencedirect.com/science/article/pii/S0963996913001798.

[31] M.J. Lerma-García, A. Gori, L. Cerretani, E.F. Simó-Alfonso, M.F. Caboni, Classification of pecorino cheeses produced in Italy according to their ripening time and manufacturing technique using Fourier transform infrared spectroscopy, J. Dairy Sci. 93 (10) (2010) 4490–4496, http://dx.doi.org/10.3168/jds.2010-3199, URL: https://www.sciencedirect.com/science/article/pii/S0022030210004790.

[32] A. Crespo, A. Martín, S. Ruiz-Moyano, M.J. Benito, M. Rufo, J.M. Paniagua, A. Jiménez, Application of ultrasound for quality control of Torta del Casar cheese ripening, J. Dairy Sci. 103 (10) (2020) 8808–8821, http://dx.doi.org/10.3168/jds.2020-18160, URL: https://www.sciencedirect.com/science/article/pii/S0022030220306159.

[33] J. Eskelinen, A. Alavuotunki, E. Hæggström, T. Alatossava, Preliminary study of ultrasonic structural quality control of swiss-type cheese, J. Dairy Sci. 90 (9) (2007) 4071–4077.

[34] G. Nassar, F. Lefbvre, A. Skaf, J. Carlier, B. Nongaillard, Y. Noêl, Ultrasonic and acoustic investigation of cheese matrix at the beginning and the end of ripening period, J. Food Eng. 96 (1) (2010) 1–13.

[35] J. Bodapati, N. Veeranjaneyulu, Feature extraction and classification UsingDeep convolutional neural networks, J. Cyber Secur. Mobil. 8 (2019) 261–276, http://dx.doi.org/10.13052/jcsm2245-1439.825.

[36] B. Petrovska, E. Zdravevski, P. Lameski, R. Corizzo, I. Štajduhar, J. Lerga, Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification, Sensors 20 (14) (2020) 3906, http://dx.doi.org/10.3390/s20143906, URL: https://www.mdpi.com/1424-8220/20/14/3906, Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[37] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021, OpenReview.net, 2021.

[38] M. Assran, Q. Duval, I. Misra, P. Bojanowski, P. Vincent, M. Rabbat, Y. LeCun, N. Ballas, Self-supervised learning from images with a joint-embedding predictive architecture, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Vancouver, BC, Canada, 2023, pp. 15619–15629, http://dx.doi.org/10.1109/CVPR52729.2023.01499, URL: https://ieeexplore.ieee.org/document/10205476/.

[39] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, A. Joulin, Emerging properties in self-supervised vision transformers, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, IEEE, Montreal, QC, Canada, 2021, pp. 9630–9640, http://dx.doi.org/10.1109/ICCV48922.2021.00951, URL: https://ieeexplore.ieee.org/document/9709990/.

[40] Y. Fang, W. Wang, B. Xie, Q. Sun, L. Wu, X. Wang, T. Huang, X. Wang, Y. Cao, EVA: Exploring the limits of masked visual representation learning at scale, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Vancouver, BC, Canada, 2023, pp. 19358–19369, http://dx.doi.org/10.1109/CVPR52729.2023.01855, URL: https://ieeexplore.ieee.org/document/10203681/.

[41] K. He, X. Chen, S. Xie, Y. Li, P. Dollar, R. Girshick, Masked autoencoders are scalable vision learners, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, New Orleans, LA, USA, 2022, pp. 15979–15988, http://dx.doi.org/10.1109/CVPR52688.2022.01553, URL: https://ieeexplore.ieee.org/document/9879206/.

[42] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.Y. Huang, S.W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski, DINOv2: Learning robust visual features without supervision, 2023, http://dx.doi.org/10.48550/arXiv.2304.07193, arXiv:2304.07193 [cs].

[43] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, in: Proceedings of the 38th International Conference on Machine Learning, PMLR, (ISSN: 2640-3498) 2021, pp. 8748–8763, URL: https://proceedings.mlr.press/v139/radford21a.html.

[44] J. Kim, K. Shim, J. Kim, B. Shim, Vision transformer-based feature extraction for generalized zero-shot learning, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, (ISSN: 2379-190X) 2023, pp. 1–5, http://dx.doi.org/10.1109/ICASSP49357.2023.10095217, URL: https://ieeexplore.ieee.org/document/10095217.

[45] A.A. Barbhuiya, R.K. Karsh, R. Jain, CNN based feature extraction and classification for sign language, Multimedia Tools Appl. 80 (2) (2021) 3051–3069, http://dx.doi.org/10.1007/s11042-020-09829-y.

[46] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, A. Mittal, Pneumonia detection using CNN based feature extraction, in: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT, 2019, pp. 1–7, http://dx.doi.org/10.1109/ICECCT.2019.8869364, URL: https://ieeexplore.ieee.org/abstract/document/8869364.

[47] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, F.F. Li, ImageNet: a Large-Scale Hierarchical Image Database, IEEE, 2009, http://dx.doi.org/10.1109/CVPR.2009.5206848, Journal Abbreviation: IEEE Conference on Computer Vision and Pattern Recognition Pages: 255 Publication Title: IEEE Conference on Computer Vision and Pattern Recognition.

[48] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, 60, (6) 2017, pp. 84–90,

[49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, IEEE Computer Society, (ISSN: 1063-6919) 2015, pp. 1–9, http://dx.doi.org/10.1109/CVPR.2015.7298594, URL: https://www.computer.org/csdl/proceedings-article/cvpr/2015/07298594/12OmNyOq4YE.

[50] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Las Vegas, NV, USA, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90, URL: http://ieeexplore.ieee.org/document/7780459/.

[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Las Vegas, NV, USA, 2016, pp. 2818–2826, http://dx.doi.org/10.1109/CVPR.2016.308, URL: http://ieeexplore.ieee.org/document/7780677/.

[52] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI '17, AAAI Press, San Francisco, California, USA, 2017, pp. 4278–4284.

[53] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, 2018, arXiv abs/1804.02767, URL: https://api.semanticscholar.org/CorpusID:4714433.

[54] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Honolulu, HI, 2017, pp. 2261–2269, http://dx.doi.org/10.1109/CVPR.2017.243, URL: https://ieeexplore.ieee.org/document/8099726/.

[55] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: Proceedings of the 36th International Conference on Machine Learning, PMLR, (ISSN: 2640-3498) 2019, pp. 6105–6114, URL: https://proceedings.mlr.press/v97/tan19a.html.

[56] L. Putzu, A. Loddo, C.D. Ruberto, Invariant moments, textural and deep features for diagnostic MR and CT image retrieval, in: Computer Analysis of Images and Patterns: 19th International Conference, CAIP 2021, Virtual Event, September 28–30, 2021, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2021, pp. 287–297, http://dx.doi.org/10.1007/978-3-030-89128-2_28.

[57] R. Mukundan, S. Ong, P. Lee, Image analysis by Tchebichef moments, IEEE Trans. Image Process. 10 (9) (2001) 1357–1364, http://dx.doi.org/10.1109/83.941859, URL: https://ieeexplore.ieee.org/document/941859, Conference Name: IEEE Transactions on Image Processing.

[58] C. Di Ruberto, L. Putzu, G. Rodriguez, Fast and accurate computation of orthogonal moments for texture analysis, Pattern Recognit. 83 (2018) 498–510, http://dx.doi.org/10.1016/j.patcog.2018.06.012, URL: https://www.sciencedirect.com/science/article/pii/S003132031830222X.

[59] M.R. Teague, Image analysis via the general theory of moments*, J. Opt. Soc. Amer. 70 (8) (1980) 920–930, http://dx.doi.org/10.1364/JOSA.70.000920, URL: https://opg.optica.org/abstract.cfm?URI=josa-70-8-920.

[60] C.-H. Teh, R. Chin, On image analysis by the methods of moments, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 496–513, http://dx.doi.org/10.1109/34.3913, URL: https://ieeexplore.ieee.org/document/3913, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[61] M. Oujaoura, B. Minaoui, M. fakir, Image annotation by moments, in: Moments and Moment Invariants - Theory and Applications, 2014, pp. 227–252, http://dx.doi.org/10.15579/gcr.vol1.

[62] P. Viola, M. Jones, 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, Vol. 1, CVPR 2001, (ISSN: 1063-6919) 2001, p. I, http://dx.doi.org/10.1109/CVPR.2001.990517, URL: https://ieeexplore.ieee.org/document/990517.

[63] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. SMC-3 (6) (1973) 610–621, http://dx.doi.org/10.1109/TSMC.1973.4309314, URL: https://ieeexplore.ieee.org/document/4309314, Conference Name: IEEE Transactions on Systems, Man, and Cybernetics.

[64] L. Putzu, C. Di Ruberto, Rotation invariant co-occurrence matrix features, in: S. Battiato, G. Gallo, R. Schettini, F. Stanco (Eds.), Image Analysis and Processing - ICIAP 2017, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2017, pp. 391–401, http://dx.doi.org/10.1007/978-3-319-68560-1_35.

[65] D.-c. He, L. Wang, Texture unit, texture spectrum, and texture analysis, IEEE Trans. Geosci. Remote Sens. 28 (4) (1990) 509–512, http://dx.doi.org/10.1109/TGRS.1990.572934, URL: https://ieeexplore.ieee.org/document/572934, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

[66] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.

[67] B. Venkatesh, J. Anuradha, A review of feature selection and its methods, Cybern. Inf. Technol. 19 (2019) 3, http://dx.doi.org/10.2478/cait-2019-0001.

[68] N. Pudjihartono, T. Fadason, A.W. Kempa-Liehr, J.M. O'Sullivan, A review of feature selection methods for machine learning-based disease risk prediction, Front. Bioinform. 2 (2022) URL: https://www.frontiersin.org/articles/10.3389/fbinf.2022.927312.

[69] Y. Bouchlaghem, Y. Akhiat, S. Amjad, Feature selection: A review and comparative study, E3S Web Conf. 351 (2022) 01046, http://dx.doi.org/10.1051/e3sconf/202235101046, URL: https://www.e3s-conferences.org/10.1051/e3sconf/202235101046.

[70] L.B.V. de Amorim, G.D.C. Cavalcanti, R.M.O. Cruz, The choice of scaling technique matters for classification performance, Appl. Soft Comput. 133 (2023) 109924, http://dx.doi.org/10.1016/j.asoc.2022.109924, URL: https://www.sciencedirect.com/science/article/pii/S1568494622009735.

[71] C. Dünner, T.P. Parnell, D. Sarigiannis, N. Ioannou, A. Anghel, G. Ravi, M. Kandasamy, H. Pozidis, Snap ML: a hierarchical framework for machine learning, in: S. Bengio, H.M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3–8, 2018, Montréal, Canada, 2018, pp. 250–260.

[72] L. Zhang, B. Zhang, Hierarchical machine learning – a learning methodology inspired by human intelligence, in: G.Y. Wang, J.F. Peters, A. Skowron, Y. Yao (Eds.), Rough Sets and Knowledge Technology, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, pp. 28–30, http://dx.doi.org/10.1007/11795131_3.

[73] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[74] P. Schuetz, D. Guggisberg, M.T. Fröhlich-Wyder, D. Wechsler, Software comparison for the analysis of cheese eyes in X-ray computed tomography, Int. Dairy J. 63 (2016) 62–69.

[75] M.C. Soto-Barajas, M.I. González-Martín, J. Salvador-Esteban, J.M. Hernández-Hierro, V. Moreno-Rodilla, A.M. Vivar-Quintana, I. Revilla, I.L. Ortega, R. Morón-Sancho, B. Curto-Diego, Prediction of the type of milk and degree of ripening in cheeses by means of artificial neural networks with data concerning fatty acids and near infrared spectroscopy, Talanta 116 (2013) 50–55.

[76] S.T.M. Del Campo, N. Bonnaire, D. Picque, G. Corrieu, Initial studies into the characterisation of ripening stages of emmental cheeses by mid-infrared spectroscopy, Dairy Sci. Technol. 89 (2) (2009) 155–167.

[77] R. Chang, S. Qi, Y. Yue, X. Zhang, J. Song, W. Qian, Predictive radiomic models for the chemotherapy response in non-small-cell lung cancer based on computerized-tomography images, Front. Oncol. 11 (2021) 646190, http://dx.doi.org/10.3389/fonc.2021.646190, URL: https://www.frontiersin.org/articles/10.3389/fonc.2021.646190/full.

[78] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J.W.L. Aerts, Machine learning methods for quantitative radiomic biomarkers, Sci. Rep. 5 (1) (2015) 13087, http://dx.doi.org/10.1038/srep13087, URL: https://www.nature.com/articles/srep13087.

[79] S. Jiang, W. Min, L. Liu, Z. Luo, Multi-scale multi-view deep feature aggregation for food recognition, IEEE Trans. Image Process. 29 (2020).

[80] N. Rashid, M.A.F. Hossain, M. Ali, M. Islam Sukanya, T. Mahmud, S.A. Fattah, AutoCovNet: Unsupervised feature learning using autoencoder and feature merging for detection of COVID-19 from chest X-ray images, Biocybern. Biomed. Eng. 41 (4) (2021) 1685–1701, http://dx.doi.org/10.1016/j.bbe.2021.09.004, URL: https://linkinghub.elsevier.com/retrieve/pii/S020852162100108X.