This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

CRDet: An Artificial Intelligence-Based Framework for Automated Cheese Ripeness Assessment From Digital Images

Alessandra Perniciano, Luca Zedda, Cecilia Di Ruberto, Barbara Pes, *Member, IEEE*, and Andrea Loddo, *Member, IEEE*

***Abstract*—Assessing cheese quality and ripeness is a crucial challenge in the dairy industry, with significant implications for product quality, consumer satisfaction, and economic impact. Traditional evaluation methods relying on visual inspection and human expertise are susceptible to errors and time constraints. This study proposes an innovative approach leveraging machine learning and computer vision techniques for automated cheese ripeness detection to address these limitations.**

The key contributions of this work include the release of the first comprehensive public dataset of cheese wheel images depicting various products at different ripening stages comprising more than 775 images, CR-IDB, an extensive comparative analysis of the performance of machine learning classifiers trained with features extracted from convolutional neural networks and handcrafted descriptors, along with the evaluation of different feature selection techniques, and finally, a proposal of a novel AI-based framework built upon a Random Forest classifier for cheese ripeness detection, called CRDet.

The novelty of CRDet lies in its enforceability across multiple types and dairy industries, which has not been previously addressed in the literature. Unlike earlier methodologies that focused on specific cheese types or relied on subjective visual inspections, this study introduces a comprehensive, noninvasive, and automated approach that demonstrates superior classification performance in differentiating ripeness phases. Thus, it overcomes the limitations of traditional methods and enhances the reliability of cheese ripening assessments.

With performance in terms of F1 above 90%, the proposed approach reduces reliance on human expertise, ensuring efficient and reliable evaluation methods for the diverse cheese production landscape. The findings provide valuable insights into the potential of feature selection methods for advancing cheese quality analysis, with implications for the broader dairy industry.

***Index Terms*—Cheese monitoring, computer vision, feature selection, machine learning, quality control.**

This work was supported by 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 – Next Generation EU. 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 MUR. Recommended by Associate Editor Shuai Li. *(Corresponding author: Andrea Loddo.)*

Citation: A. Perniciano, L. Zedda, C. Di Ruberto, B. Pes, and A. Loddo, "CRDet: An artificial intelligence-based framework for automated cheese ripeness assessment from digital images," *IEEE/CAA J. Autom. Sinica*, 2025, DOI: 10.1109/JAS.2024.125061

The authors are with the Department of Mathematics and Computer Science, University of Cagliari, Sardegna 09124, Italy (e-mail: alessandra.perniciano@unica.it; luca.zedda@unica.it; cecilia.dir@unica.it; pes@unica.it; andrea.loddo@unica.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2024.125061

I. Introduction

In the food industry, dairy products, especially cheese, hold immense economic value and are pivotal in promoting positive health outcomes. Rich in proteins, calcium, and vital micronutrients, dairy products contribute significantly to bone and muscle health while fostering a healthy digestive system and microbiome through their probiotic content. The cheese industry, in particular, wields substantial economic influence, particularly in regions where cheese production is integral to the agricultural sector. Furthermore, cheese is a versatile and widely consumed dairy product globally, contributing as a standalone consumable and a vital ingredient in various culinary creations.

Assessing cheese quality presents a multifaceted challenge involving evaluating chemical components, internal structure, and sensory attributes influenced by specific properties [1]. The ripening phase, marked by biochemical changes like lipolysis and proteolysis, profoundly affects flavor, aroma, and texture [2], [3]. However, determining cheese ripeness is intricate due to unpredictable factors such as seasonal variations, milk origin, processing steps, and storage temperatures. Errors in ripeness assessment may lead to the market release of subpar products, tarnishing a company's reputation and revenue [4].

Traditional assessment methods, including visual inspection and weight measurement conducted by trained personnel, are susceptible to occasional errors and time constraints. To surmount these challenges, the dairy industry is increasingly exploring advanced technologies for efficient monitoring. Although noninvasive techniques based on physicochemical, chromatographic, and electrophoretic analyses have been studied, they are often costly and time-consuming [5]. Moreover, the subjective nature of artisanal assessments, relying on experts' visual and olfactory evaluations, adds further complexity.

In response to the complexities involved in cheese ripeness evaluation, this paper proposes an innovative artificial intelligence (AI)-based framework leveraging and combining computer vision (CV), machine learning (ML), and feature selection (FS) techniques. This framework aims to automate the cheese ripeness detection task, addressing the limitations of traditional methods and providing precision and consistency in evaluating cheese ripeness. Such advancement can poten-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

tially reduce reliance on human expertise, ensuring efficient and reliable evaluation methods for the vast and diverse world of cheese production.

Despite the advancements, cheese quality analysis remains constrained by data acquisition challenges and privacy concerns. Therefore, with the extensive study conducted, we contribute to the field with the **CRDet** framework to detect the ripeness stage of cheese based on visual cues in the cheese wheel digital image, acquired via a simple digital camera, emphasizing its nonintrusive, nondestructive, and cost-effective nature.

The contributions of this paper can be summarized as follows:

- **Public dataset release.** We 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 URL.
- **Extensive comparison.** We analyzed and compared the results obtained with the features extracted from 14 different Convolutional Neural Networks, seven handcrafted features, and six feature selection approaches to obtain insights into the process of cheese ripeness.
- **Novel framework: CRDet.** Based on the comparison, we propose an innovative AI-based framework to automate the challenging task of cheese ripeness detection by combining computer vision, machine learning, and feature selection techniques named CRDet.

The work is subdivided into the following sections: Section II reviews existing techniques in cheese quality analysis, establishing a foundation for the proposed methodology, while Section III gives the details about the proposed dataset and its preprocessing, feature extraction and selection, and machine learning strategies, laying the groundwork for the study. The experimental results are presented in Section IV, and in Section V are comprehensively discussed. Finally, in Section VI, we formulate our findings about the conducted study and provide insightful possible new ways to improve our work and possibly inspire future works based on our findings.

II. Related Work

In the domain of cheese production, various methodologies have been developed for monitoring the cheese production process, including CV with digital or hyperspectral imaging, near-infrared (NIR) spectroscopy, Fourier-transformed infrared (FTIR) spectroscopy, and other analytical techniques. These methods serve not only for quality assessment but also for determining geographical origin and detecting potential adulteration in cheese products [1]. They offer advantages such as rapidity, non-destructiveness, and non-invasiveness. In contrast to traditional approaches, they can supplement human visual inspection in evaluating cheese attributes and sensory quality, such as color assessments and identification of imperfections like gas or mechanical holes, presence of calcium lactate crystals, excessive rind halo formation, and oiling off [6], [7], without direct contact with the samples.

Despite the increased interest in automating and improving cheese quality control, the state of the art in cheese quality analysis remains limited. The causes are challenges in data acquisition and strict privacy politics inherent in industrial patents and processes. However, nondestructive approaches are pivotal for industrial-scale solutions, categorized into three main groups: computer vision, spectral, and ultrasound-based methods.

Regarding the **CV** approaches, a typical pipeline involves using a digital camera to capture images, which are then processed for further analysis [1]. These techniques are versatile and applicable for assessing the maturity of entire cheese wheels [8] or in different scenarios, such as inspecting ingredient distribution [9]. Also, CV techniques have been applied to various varieties, including Cheddar [10], Mozzarella [11], Parmigiano Reggiano [12], [13], Grana Padano [12], [14], Queijo de Nisa [7], and Pecorino [8] cheeses. In parallel with advances in automation technologies with CV-based systems, there are several robotics calibration systems—highlighted by recent works on calibration technology for industrial robots. In particular, there is potential for integrating sophisticated calibration techniques into automated cheese production systems. For instance, Li *et al.* emphasized using open datasets to enhance robot arm accuracy through machine learning approaches [15], which, in this context, may be useful to set up the camera system. This aligns with findings from [16], which underscores the importance of calibration technology in improving robotic precision across various applications. Moreover, some works introduced innovative training methodologies that could be adapted to enhance robotic manipulator performance in food processing environments [17]−[19].

Nonetheless, these methods' considerable dependence on the light source's direction and strength constancy is a significant issue. More sophisticated and reliable methods, such as magnetic resonance imaging (MRI), computed tomography (CT), and X-rays, can address this problem. These previously mentioned techniques allow for an eye cheese analysis in the context of dairy products, especially when combined with a 3D model of the cheese wheel's interior [20]−[23]. Nevertheless, these approaches require specific equipment typically used in medical contexts. Hence, a minimally invasive and inexpensive approach based on digital images acquired with a basic digital camera remains valuable.

The methods of **Spectral Analysis** use probes or different light emitting sources to measure the light reflected by cheese wheels. One of the most important techniques is Fluorescence spectroscopy, which determines the ripeness and the molecular structures of the cheese, extracting and analyzing the intensity of fluorescent components in the cheese [24], [25]. Near-infrared spectroscopy (NIR) uses additional spectral data to evaluate the quality of the cheese. More specifically, quality, moisture content, and protein percentage can all be established by using NIR data regarding the cheese's structure [26]−[28]. It is essential to mention the Fourier transform infrared spectroscopy (FTIR) method, which uses infrared spectroscopy to identify different cheese characteristics. It is usually utilized in fast or real-time scenarios [20], [29], [30].

With transducers and oscilloscopes, **Ultrasound techniques** generate wave data that defines the integrity and structure of the cheese wheel [31].

This study focuses on developing the cheese ripeness detec-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

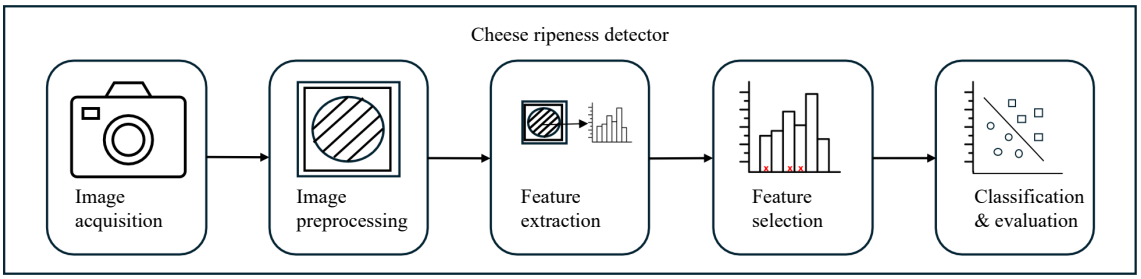


Fig. 1. Schematic representation of the proposed workflow in the context of CRDet.

tor (CRDet), a novel AI-based framework to automate the task of cheese ripeness detection by analyzing digital cheese wheel images taken with a basic digital camera. Unlike other methods, this approach is nonintrusive and nondestructive, prioritizing simplicity in setup and cost-effectiveness.

III. Materials and Methods

This section provides an overview of the materials and methods employed in this study, beginning with Section III-A, which introduces the image dataset analyzed, and Section III-B, which describes the preprocessing steps followed to realize the dataset and make the images consistent and coherent throughout the subsets. Section III-C delineates the feature extraction methods, categorized into deep learning-based (referred to as "deep features" from now on) and handcrafted (HC) features, while Section III-D discusses the FS methodologies employed to identify the most suitable features. Section III-E outlines the specific ML algorithms for classifying the processed data. Lastly, Section III-F presents the evaluation measures used to assess the classification performance. The pipeline of this research is shown in Fig. 1.

A. Dataset

This study introduces the cheese ripeness image data base (CR-IDB), the first dataset featuring images of cheese-wheel varieties obtained from three distinct Italian dairy industries. The Sardinian agency supported the development of CR-IDB for the implementation of regional agricultural and rural development programs (LAORE) and BiosAbbey S.r.l.[1].

CR-IDB comprises three subsets sourced from each dairy: *Podda-Granarolo*, *Gennargentu-Conad*, and *Lattebusche-Conad*. All images were captured with a single light source, centered on cardboard, and have an initial resolution of 6016×4016 pixels, further resized at 1800×1800 pixels after the preprocessing steps (see Section III-B). Image acquisition was performed using a Nikon D750 camera with a CMOS 35.9×24.0 mm sensor at 24 megapixels resolution.

The dataset is publicly available at the following URL [32]. We now provide a detailed description of the three subsets.

The **Podda-Granarolo** subset (hereafter referred to as IDB-1) features a specific soft cheese product that matures within 20–25 days. This subset includes four classes representing different ripeness stages. Notably, the ripening process culminates at 24 days, with "Day 18" and "Day 22" classes indicating under-ripeness and "Day 30" representing over-ripeness. All the classes are represented by 50 images except for the "Day 18", with 45. The total amount of images available in this subset is 195. Table I provides an overview of the different cheese ripeness stages. Examples of images from this subset are illustrated in Fig. 2.

TABLE I
Description of the IDB-1 Subset From CR-IDB

Class name	Target	Days	Num. images
Day 18		18	45
Day 22		22	50
Day 24	✓	24	50
Day 30		30	50

The **Gennargentu-Conad** subset (hereafter referred to as IDB-2) showcases images of Sardinian *Pecorino* cheese wheels. Like the previous subset, even this one is a soft cheese product. However, it reaches its optimal ripeness within 30 days. Consequently, "Day 26" and "Day 28" represent under-ripeness, whereas "Day 32" and "Day 34" indicate over-ripeness. Every single class comprises 40 images, for a total amount of 200. Sample images from this subset are shown in Fig. 3, while the class details are given in Table II.

The **Lattebusche-Conad** subset (hereafter referred to as IDB-3) contains images of three cow's milk cheese varieties: "Semi Hard", "Hard", and "Extra Hard". Each variety is further categorized into "Target" and "Non target" groups based on ripeness. The former includes optimally ripe cheese, while the latter consists of cheeses needing further aging or those past the desired ripeness level. In the context of this subset, every variety contains 84 "Non target" images and 42 "Target" images, for a total of 378. Sample images from this subset can be seen in Fig. 4. Detailed ripeness information for each category is provided in Table III. In addition, please note that in this subset, specific ripeness days were not assigned as in the other subsets due to the operator's observations. Certain forms that were expected to be mature based on the reference period were found to be immature upon inspection. Therefore, for this subset, it was deemed appropriate to categorize forms simply as target (mature) and non-target (immature) based on visual assessment.

B. Image Preprocessing

As previously stated in Section III-A, the images were initially captured at 6016×4016 pixels resolution. However, this resolution included significant unnecessary image areas, as depicted in Fig. 5. Consequently, a center-cropping process was employed to accurately describe and analyze the cheese

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

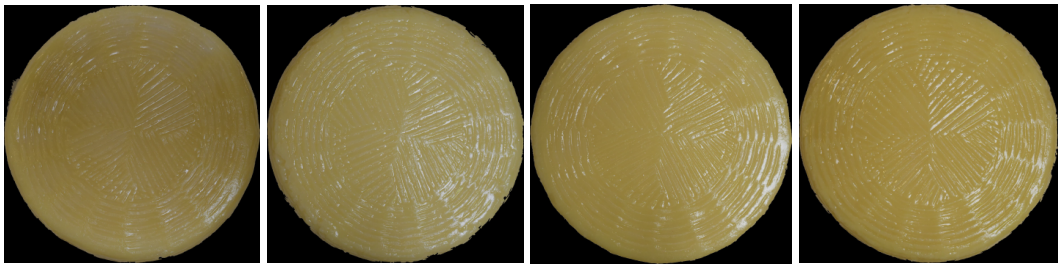


Fig. 2. Representative samples from the IDB-1 subset, showcasing various stages of cheese ripeness: Day 18 and Day 22 show two examples of under-ripeness, Day 24 is an example of the target, and Day 30 shows an example of over-ripeness.

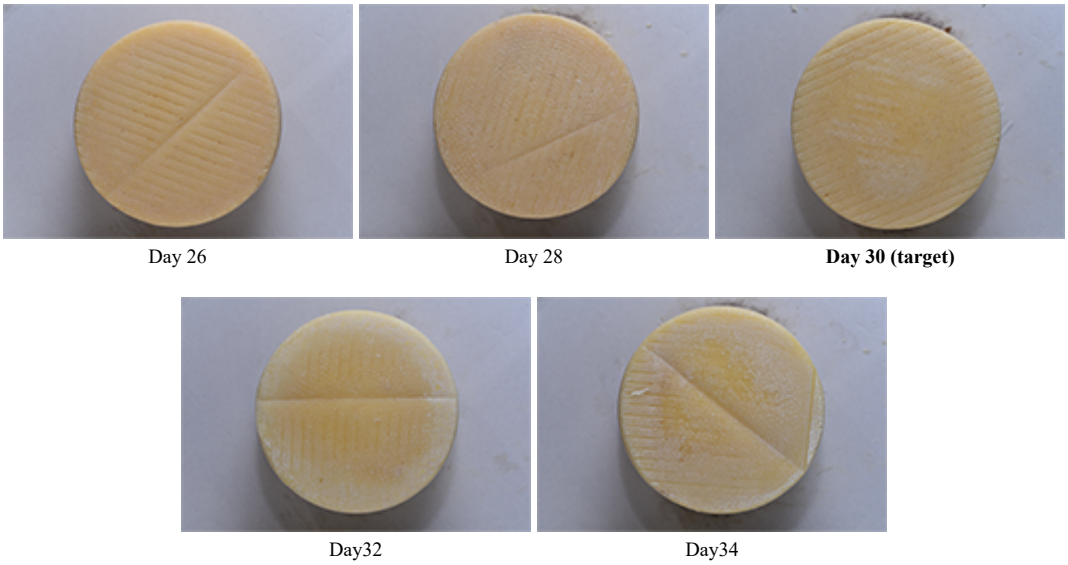


Fig. 3. Representative samples from the IDB-2 subset, showcasing various stages of cheese ripeness: Day 26 and Day 28 show two examples of under-ripeness, Day 30 is an example of the target, while Day 32 and Day 34 show two examples of over-ripeness.

TABLE II
DESCRIPTION OF THE IDB-2 SUBSET FROM CR-IDB

Class name	Target	Days	Num. images
Day 26		26	40
Day 28		28	40
Day 30	✓	30	40
Day 32		32	40
Day 34		34	40

shapes, aiming for uniform and well-contrasted backgrounds in the images. This procedure produced images with a 1800×1800 pixels resolution, as depicted in Fig. 5.

In addition, as noticed from Fig. 4, all the images in the IDB-3 subset exhibit several common characteristics, notably the presence of a discernible cheese label specifying the type and date of the wheel. To mitigate possible biases or information from OCR, the feature extraction procedure has been conducted on the channel with minimal intensity corresponding to the cheese label, viz. the blue one. A clearer exemplification is provided in Fig. 6.

C. Feature Extraction

In this specific context, we extracted various combinations of features to train the ML models. These attributes are categorized into four primary groups, as outlined by Putzu *et al.* [33]: deep, texture, histogram, and invariant moments features. Notably, all features within the deep features group were extracted from convolutional neural network (CNN) architectures, while the latter three groups are HC. The number of features extracted from each approach is detailed in Table IV for HC methods and in Table V for CNN-based features.

1) Handcrafted Features: A brief overview of the three categories of HC features is provided below.

Texture Features. Texture features are typically selected to emphasize the fine characteristics of cheese texture. Specifically, we considered Haar-like features, introduced by Viola *et al.* [34], that 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.

In addition, we extracted thirteen Haralick features from the gray level co-occurrence matrix (GLCM) [35] and converted them into rotation-invariant features named HAR_ri. Specifically, four different types of GLCM were computed, each

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Fig. 4. Representative samples of the IDB-3 subset. The columns showcase instances from three distinct cheese categories (Hard, Semi-Hard, Extra-Hard), while the rows illustrate ripeness categorizations as target and non-target).

TABLE III
DESCRIPTION OF THE IDB-3 SUBSET FROM CR-IDB

Class name	Target	Days (δ)	Num. images
Semi-Hard	✓	$\delta \in [60;180]$	42
Semi-Hard		$\delta \leq 60 \vee \delta \geq 180$	84
Hard	✓	$\delta = 180$	42
Hard		$\delta < 180$	84
Extra-Hard	✓	$\delta \geq 360$	42
Extra-Hard		$\delta < 360$	84

Fig. 5. Representative samples from the Podda subset, showcasing various stages of cheese ripeness: Day 18 and Day 22 show two examples of under-ripeness, Day 24 is an example of the target, and Day 30 shows an example of over-ripeness.

with $d = 1$ and $\theta = [0, 45, 90, 135]$.

Finally, we used the rotation-invariant form [36] of the feature vector derived from the histogram of local binary pattern (LBP). We refer to them as LBP_ri from now on.

Histogram Features. They are provided by the histogram, which characterizes the overall color distribution within the image. Specifically, from this color histogram, we derived seven statistical descriptors: mean, standard deviation, smoothness, skewness, kurtosis, uniformity, and entropy. We refer to these seven features as *Hist*.

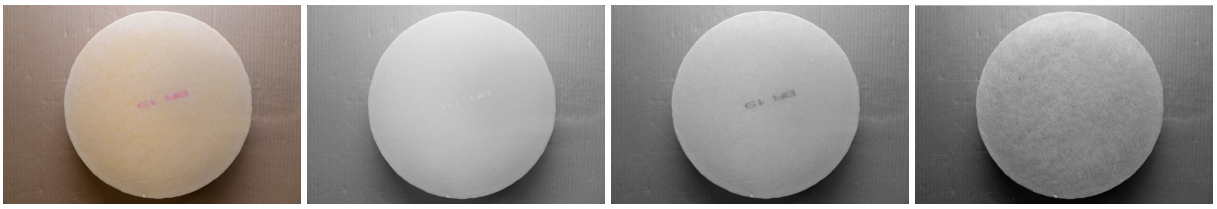
Invariant Moments. The weighted average of pixel intensities in an image, known as the moment, is utilized to extract specific attributes. Moments are commonly used in pattern recognition to describe objects post-segmentation. In this study, we considered three types of moments, following the same implementations provided in a previous work [8]:

- *Chebyschev moments (CH) [37]:* Related to discrete orthogonal moments, their implementation does not involve numerical approximation. Chebyshev polynomials form the basis of these moments, which can extract global image features by varying the moment order [38]. We employed first-order (*CH_1*) and second-order (*CH_2*) Chebyshev moments with vectors of order 5 and 4, respectively;
- *Legendre moments (LM):* Introduced by Teague [39], these are orthogonal moments designed to achieve minimal redundancy in a set of moment functions, thereby highlighting independent characteristics [40]. We used a continuous five-order vector of Legendre moments approximated using Simpson's rule;
- *Zernike moments (ZM):* Based on Zernike polynomials, which form an orthogonal set defined on the unit disk. These moments efficiently represent image characteristics without introducing redundancy, making them valuable in various image analysis applications [41]. This study used a Zernike vector of order 6 with 4 repetitions.

2) Deep Features: CNNs have emerged as robust architectures, proving highly effective in tackling diverse image classification challenges across various domains [53]–[55]. Notably, they have revolutionized the conventional approach to feature extraction by automating the process and using their layers' activations to extract learned features. Various models were trained using the features retrieved by the chosen CNNs.

Precisely, in the context of the end-to-end DL approach, emphasis was placed on utilizing pre-trained models with weights trained on ImageNet. The optimization process involved fine-tuning these pre-trained models, a common practice that entails freezing all layers, excluding the last three fully-connected layers, during training. This approach, incorporating transfer learning, was implemented to address challenges related to limited training data and mitigate the poten-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



Original image Red channel Green channel Blue channel

Fig. 6. Illustration of the three channels composing the original images of IDB-3 subset. The text label on the cheese wheels is clearly visible in the red and green channels, while it is completely missing in the blue one.

TABLE IV
SHORT SUMMARY OF THE HANDCRAFTED FEATURES EXTRACTED FROM IMAGES (WITH THE NUMBER OF FEATURES AND SHORT DESCRIPTION)

Group	# Features	Category	Description
Haar	192	Texture	Haar-like features
HAR_ri	26	Texture	Rotation invariant Haralick texture features
LBP_ri	36	Texture	Local binary patterns (with a radius of 1 and 8 neighbors)
Hist	7	Histogram	Gray-level histogram features
CH_1	21	Invariant Moments	Chebyshev 1st-order moments of order 5
CH_2	15	Invariant Moments	Chebyshev 2nd-order moments of order 4
LM	21	Invariant Moments	Legendre moments of order 5
ZM	21	Invariant Moments	Zernike vector with order 6 and 4 repetitions

TABLE V
EMPLOYED CONVOLUTIONAL NEURAL NETWORKS DETAILS INCLUDING REFERENCE PAPER, NUMBER OF TRAINABLE PARAMETERS IN MILLIONS, INPUT SHAPE, FEATURE EXTRACTION LAYER, AND RELATED FEATURE VECTOR SIZE

Reference	Parameters (M)	Input shape	Feature layer	# Features
AlexNet [42]	60	224×224	Pen. FC	4096
DarkNet-53 [43]	20.8	224×224	Conv53	1000
DenseNet-201 [44]	25.6	224×224	Avg. Pool	1920
EfficientNetB0 [45]	5.3	224×224	Avg. Pool	1280
GoogLeNet [46]	5	224×224	Loss3	1000
Inception-v3 [47]	21.8	299×299	Last FC	1000
Inception-ResNet-v2 [48]	55	299×299	Avg. pool	1536
NasNetL [49]	88.9	331×331	Avg. Pool	4032
ResNet-18 [50]	11.7	224×224	Pool5	512
ResNet-50 [50]	26	224×224	Avg. Pool	1024
ResNet-101 [50]	44.6	224×224	Pool5	1024
VGG16 [51]	138	224×224	Pen. FC	4096
VGG19 [51]	144	224×224	Pen. FC	4096
XceptionNet [52]	22.9	299×299	Avg. Pool	2048

tial for overfitting [46]. In particular, we employed the following CNNs:

● *AlexNet:* Composed by a series of convolutional and max pooling layers with three fully connected layers in the end, it was first introduced by Krizhevsky *et al.* [42]. Since it only has five convolutional layers, it is the shallowest architecture used in the present work;

● *DarkNet:* Is primarily built upon the established principles of inception and batch normalization. This work used two particular versions of DarkNet, which incorporates 18 and 53 convolutional layers; This architectural configuration is the foundational network for the you only look once (YOLO) object detection method [43];

● *DenseNet* proposed by Huang *et al.* [44] to overcome the CNNs' characteristic to have some layers equal to the number of connections. Indeed, the number of connections is $L(L+1)/2$, whereas L is the number of layers. In every layer, the input consists of the output generated by all preceding layers, and this information serves as the input for the subsequent layer. The quantity of filters employed in each convolutional layer fluctuates based on the growth rate parameter denoted as k. In this study, DenseNet-201 with k = 32 has

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

been used;

- *EfficientNet* has the characteristic of scaling the network's width, depth, and resolution in a uniform and efficient way by using compound scaling. It was proposed by Tan *et al.* [45], and eight versions of this network exist (B0–B7). In this study, we adopted the EfficientNet B0;
- *GoogLeNet:* Is based on blocks of inception layers where the filters used can vary from 1×1 to 5×5, thus allowing multi-scale learning [46];
- *Inception-v3:* Built on the inception layers concept, it enhances GoogLeNet by incorporating factorized, smaller, and asymmetric convolutions [47]. The Inception models are renowned for their multi-branch architectures, comprising filters such as 1×1, 3×3, 5×5, and so forth, combined through concatenation within each branch;
- *Inception-ResNet-v2:* Is a fusion of ResNet and Inception architectures [48]. Indeed, the Inception-ResNet block integrates various-sized convolutional filters along with residual connections. It has 4 max-pooling layers and 160 convolutional layers.
- *NasNet:* Is a CNN created through automated neural architecture search, facilitating efficient scaling and adaptability to different devices. Its cell-based design and inclusion of reduction cells contribute to effective feature extraction and spatial downsampling [49];
- *ResNet:* Is associated with a series of deep architectures that leverage residual learning [50]. These architectures incorporate skip-connections or recurrent units connecting blocks of convolutional and pooling layers. Additionally, each block is succeeded by batch normalization [56]. In this study, three variations of ResNet were employed, namely ResNet-18, ResNet-50, and ResNet-101, with the specified numbers indicating the depth of the respective networks;
- *VGG:* Comprises a series of convolutional layers followed by max-pooling, contributing to its deep representation capability [51]. In this study, we employed *VGG16* and *VGG19* where 16 and 19 denote the respective number of layers;
- *XceptionNet:* Developed by Google [52], it is an extension of the *Inception* architecture, utilizing a depth-wise separable convolutional strategy to improve efficiency and decrease parameter count. Its goal is to capture complex feature dependencies by emphasizing cross-channel correlations.

D. Feature Selection

FS involves choosing the most pertinent attributes while eliminating irrelevant or duplicate ones [57]. As FS techniques focus on selecting a subset of features without altering the original ones, there is no loss of information. The primary objective of feature selection is to create a concise subset of features that comprehensively captures the essential aspects of the entire input data [58]. Typically, when feature selection methods are applied within classification tasks, they can be classified into three groups based on how they interact with the classifier [59], [60]:

- *Filters:* They are employed as a preliminary step and rely solely on intrinsic data characteristics to evaluate a feature's relevance without any interaction with the classifier;
- *Wrappers:* They evaluate various feature subsets to select the most optimal one, aiming to enhance the performance of a specific classifier. Although this approach typically outperforms filter methods, it is susceptible to overfitting;
- *Embedded:* These approaches give significance to features based on a classifier's inherent capability to identify the most suitable ones.

In this work, we incorporated various techniques that represent distinct heuristics. Primarily, we focused on predominantly univariate methods (**Pearson's Correlation**, **Gain Ratio**, **Information Gain**, **One Rule**, and **Symmetrical Uncertainty**), which individually evaluate the importance of each feature regardless of others. Additionally, we included one multivariate method, **ReliefF**, which considers the interrelationships among the features. Notably, all these methods are categorized as Filters except for one embedded method (One Rule). In detail:

- *Pearson's Correlation (Correl):* It employs the linear correlation between each feature and the target class to evaluate their significance for prediction purposes [61]. The relevance of a feature is indicated by its stronger correlation with the prediction target Specifically, it is defined as

$$\rho(X,Y)=\frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{1}$$

where X is a generic feature, Y is the class label, σ_{XY} is the covariance of X and Y while σ_X and σ_Y are, respectively, the standard deviations of X and Y;

- *Gain Ratio (GainR):* A variation of *Information Gain*, it is meant to overcome its innate inclination to prioritize features with higher values [62]. Specifically, the Gain Ratio is expressed in the following way:

$$GainR(X)=\frac{InfoG(X)}{SplitInfo(X)}. \tag{2}$$

SplitInfo is a normalization factor that expresses how broadly X splits the data:

$$SplitInfo(X)=-\sum_{i=1}^{r}\frac{|X_i|}{R}\log_2\frac{|X_i|}{R}. \tag{3}$$

Here, $|X_i|$ represents the cases in which X assumes the value X_i, r denotes the number of different values of X, and R represents the total number of training examples;

- *Information Gain (InfoG):* It evaluates how much the value of a particular feature can be used to reduce the entropy of the class (i.e., the degree of uncertainty about its prediction) [62]:

$$InfoG(X)=H(Y)-H(Y|X) \tag{4}$$

where $H(Y)$ and $H(Y|X)$ represent the entropy of the class Y before and after observing the feature X;

- *ReliefF:* Measures the feature relevance by considering its ability to distinguish between instances close to each other in the attribute space [63]. Specifically, it repetitively selects a sample instance R_i from the training set. Then, it identifies its nearest neighbors, specifically one belonging to the same class (*nearest hit H*) and another from the contrasting class (*nearest miss M*). Then, a weight $W(X)$ is computed for each fea-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

ture X:

$$W(X)=\sum_{i=1}^{m}\left[\frac{diff(X,R_i,M)}{m}-\frac{diff(X,R_i,H}{m}\right] \tag{5}$$

where m is the number of instances considered, $diff(X,R_i,M)$ and $diff(X,R_i,H)$ are differences between the values of X in R_i and, respectively, M and H. This method is based on the idea that relevant features should have distinct values for instances of different classes and the same value for instances of the same class;

● *One Rule (OneR):* Is an embedded FS methods [64]. For each training feature, this method constructs a basic classification rule by determining the majority class for each value of that feature. Subsequently, it calculates the accuracy of each rule, and the features are then ranked based on the quality of their respective rules;

● *Symmetrical Uncertainty (SU):* Is another modification of the *Information Gain*. In particular, it expressed the correlation between two attributes X and Y in the following way:

$$U(X,Y)=2\frac{H(X)+H(Y)-H(X,Y)}{H(X)+H(Y)} \tag{6}$$

where $H(X,Y)$ is the joint entropy of X and Y, $H(X)$ and $H(Y)$ are, respectively, the entropy of the feature X and Y.

E. Machine Learning Methods

This study compared the performance of several classification methods, including the K-nearest neighbors, Support Vector Machine with RBF kernel, *random forest (RF)*, Decision Tree, XGBoost, and an Artificial Neural Network with 10 hidden layers. We will only delve into the details of the RF algorithm, which yielded the most significant results. The Random Forest algorithm is a product of bagging decision trees. Each tree within the ensemble is constructed by selecting the best splitting attribute from a subset of random features for every internal node, thereby introducing diversity among the ensemble components.

F. Evaluation Measures

This subsection presents all the evaluation measures employed to evaluate the classification performance. In particular, we present the definition of binary classification and their generalizations for the multiclass problem. According to the binary classification, one class is defined *positive* and one *negative*. In our case, which class is *positive* or *negative* is irrelevant. Depending on the classification result and the target value, an instance will contribute to the increment of one of the following values:

● *True Positives (TP)*. The count of instances predicted correctly as positive;

● *False Positives (FP)*. The count of instances that are predicted as positive but belong to the negative class;

● *False Negatives (FN)*. The count of instances predicted correctly as negative;

● *True Negatives (TN)*. The count of instances that are predicted as negative but belong to the positive class.

The evaluation metrics used in this work are the following:

● *Accuracy (ACC)*. It indicates the percentage of correctly classified records ((7)):

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN}. \tag{7}$$

● *True Positive Rate (TPR)*. It measures how correctly the positive instances are classified ((8)):

$$TPR=\frac{TP}{TP+FN}. \tag{8}$$

● *True Negative Rate (TNR)*. It measures how correctly the negative instances are classified ((9)):

$$TNR=\frac{TN}{TN+FP}. \tag{9}$$

● *Precision*. It measures the percentage of correct prediction among the positive instances ((10)):

$$Precision=\frac{TP}{TP+FP}. \tag{10}$$

● *F-measure (F1)*. It is the harmonic mean between precision and TPR. It also takes into consideration the FP and FN instances ((11)):

$$F\_measure=\frac{2\times precision\times TPR}{precision+TPR}. \tag{11}$$

● *Matthews Correlation Coefficient (MCC)*. It ranges from −1 to 1 and expresses the correlation between the observed and predicted classification. It is typically employed in an imbalanced learning context. The equation is formulated as follows ((12)):

$$MCC=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{12}$$

As mentioned above, we must extend the previous measures to multiclass classification. An important concept is macro averaging (MacroAvg), which calculates the metric for each class independently and then averages them. This method offers a balanced evaluation for all classes. The following equations define the multiclass measures, where the number of classes is indicated with C.

● *MacroAvg Accuracy*. It is defined as follows ((13)):

$$MacroAvgAccuracy=\frac{1}{C}\sum_{i=1}^{C}Accuracy\_i. \tag{13}$$

● *MacroAvg Precision*. The precision for a multiclass problem is defined as follows ((14)):

$$MacroAvgPrecision=\frac{1}{C}\sum_{i=1}^{C}Precision\_i. \tag{14}$$

● *MacroAvg TPR*. It is defined as follows ((15)):

$$MacroAvgTPR=\frac{1}{C}\sum_{i=1}^{C}TPR\_i. \tag{15}$$

● *MacroAvg TNR*. It is defined as follows ((16)):

$$MacroAvgTNR=\frac{1}{C}\sum_{i=1}^{C}TNR\_i. \tag{16}$$

● *MacroAvg F1*. The F1 for multiclass problems is defined as follows ((17)):

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

$$MacroAvgf\_measure = \frac{1}{C}\sum_{i=1}^{C} fmeasure\_i. \quad (17)$$

- *MacroAvg MCC*. It is defined as follows ((12)):

$$MacroAvgMCC = \frac{1}{C}\sum_{i=1}^{C} MCC\_i. \quad (18)$$

IV. Experimental Results

This section presents the experimental results of the proposed methodology on the datasets included in CR-IDB. Section IV-A outlines the experimental setup employed in this study. Then, Sections IV-B and IV-C first present the baseline results, i.e., the results obtained without any FS algorithm, and, then, present the results obtained after applying FS techniques.

A. Experimental Setup

All the experiments followed a 5-fold cross-validation without repetitions as the evaluation protocol. The latter implies that 80% of the dataset is used for the training phase, while the other 20% is used for testing. The assessments related to FS techniques involved four cut-offs (top 100, 50, 25, and 10 features) for the deep features and three cut-offs (top 10, 5, and 2 features) for HC ones. Moreover, to ensure comparable magnitudes, preventing certain variables from dominating others during model training [56], [65], [66], feature normalization was applied with the *Normalizer* technique as provided by the scikit-learn package [67]. This technique normalizes each sample separately along a norm. The $L2$ norm was used in this case.

The employed classifiers were trained with their default parameters, except for the RF classifier implemented with 100 trees and $\log_2(n)+1$ random features.

The workstation used for the study has an AMD Ryzen 5 5600X @ 4.6 GHz CPU, 32 GB RAM @ 3600 MHz, and an NVIDIA GeForce RTX 3070 with 8 GB of memory.

B. Baseline Results

This section provides an overview of the baseline results, i.e., without any FS algorithm, obtained from the three subsets employed in this study. Specifically, two tables are presented for each subset, delineating the comparative performance of an RF classifier trained using either deep or HC features. The evaluation measures encompass ACC, F1, TPR, and TNR, thoroughly assessing the classifier's performance. As previously introduced in Section III-E, in this section, we only present the outcomes achieved by the RF classifier to avoid overwhelming this description. This selection is motivated by the F1 scores obtained by the RF since it outperformed every other classifier in absolute terms, with a 3% margin on the runner-up, XGBoost.

1) On the IDB-1 Subset

a) Deep features: The feature extracted from DenseNet-201 emerges as the top-performing deep feature, achieving the highest scores (all five metrics are tied at 0.91), as shown in Table VI. This suggests that DenseNet-201 features are highly discriminative, leading to precise and balanced classification.

TABLE VI
Performance Comparison of Deep Features on IDB-1 Subset

Feature	ACC	F1	TPR	TNR	MCC
AlexNet	0.82	0.82	0.82	0.82	0.81
DarkNet-19	0.86	0.86	0.86	0.86	0.84
DarkNet-53	**0.90**	**0.91**	**0.90**	**0.91**	**0.90**
DenseNet-201	**0.91**	**0.91**	**0.91**	**0.91**	**0.91**
EfficientNetB0	0.88	0.88	0.88	0.88	0.88
GoogLeNet	0.76	0.75	0.76	0.76	0.73
Inception-ResNet--v2	0.85	0.85	0.85	0.86	0.84
Inception v3	0.82	0.82	0.82	0.81	0.80
NASNet-Large	0.85	0.84	0.84	0.85	0.83
NASNet-Mobile	0.79	0.80	0.79	0.80	0.79
ResNet-18	0.85	0.85	0.85	0.85	0.87
ResNet-50	0.88	0.88	0.88	0.88	0.88
ResNet-101	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
VGG-16	0.85	0.85	0.85	0.85	0.85
VGG-19	0.85	0.85	0.85	0.85	0.85
Xception	0.82	0.82	0.82	0.82	0.82

DarkNet-53, ResNet-101, EfficientNetB0 and ResNet-50 follow closely.

Conversely, GoogLeNet's features appear to be the least effective, with an ACC of 0.76 and a lower F1 (0.75). While GoogLeNet demonstrates a moderate TPR and TNR, its overall classification performance is weaker than other deep features investigated.

b) HC features: Moving to Table VII, CH_2 stands out as the top-performing HC feature, achieving high values across all five metrics, tied with LM.

TABLE VII
Performance Comparison of HC Features on IDB-1 Subset

Feature	ACC	F1	TPR	TNR	MCC
Haar	0.64	0.66	0.64	0.64	0.61
HAR_ri	0.82	0.82	0.82	0.82	0.80
LBP_ri	0.70	0.71	0.70	0.70	0.68
Hist	0.69	0.69	0.69	0.69	0.66
CH_1	**0.87**	**0.87**	**0.87**	**0.87**	**0.86**
CH_2	**0.89**	**0.89**	**0.89**	**0.89**	**0.88**
LM	**0.89**	**0.89**	**0.89**	**0.89**	**0.87**
ZM	0.74	0.74	0.75	0.74	0.74

On the other hand, the Haar feature demonstrates the lowest performance among HC features, with an ACC of 0.64 and an F1 of 0.66. The TPR and TNR are also relatively lower, indicating challenges in accurately classifying positive and negative instances. This implies that Haar features may not be as effective in capturing relevant patterns for the given classification task on this subset.

2) On the IDB-2 Subset

a) Deep features: Analyzing the results shown in Table VIII, it is evident that different CNN architectures yield varying performance levels. Even in this subset, DenseNet-201 fea-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE VIII
PERFORMANCE COMPARISON OF DEEP FEATURES ON IDB-2 SUBSET

Feature	ACC	F1	TPR	TNR	MCC
AlexNet	0.9	0.87	0.85	0.90	0.88
DarkNet-19	0.84	0.80	0.78	0.84	0.79
DarkNet-53	0.84	0.81	0.8	0.85	0.78
DenseNet-201	**0.92**	**0.91**	**0.9**	**0.92**	**0.90**
EfficientNetB0	**0.91**	**0.89**	**0.88**	**0.91**	**0.90**
GoogLeNet	0.81	0.78	0.77	0.80	0.77
Inception-ResNet--v2	0.72	0.65	0.65	0.74	0.60
Inception v3	0.72	0.68	0.67	0.73	0.66
NASNet-Large	0.85	0.84	0.82	0.87	0.81
NASNet-Mobile	0.7	0.64	0.64	0.68	0.60
ResNet-18	0.86	0.83	0.81	0.87	0.79
ResNet-101	**0.93**	**0.92**	**0.9**	**0.94**	**0.91**
ResNet-50	0.9	0.89	0.87	0.92	0.91
VGG-16	0.89	0.88	0.87	0.89	0.87
VGG-19	0.89	0.88	0.87	0.89	0.87
Xception	0.82	0.82	0.8	0.84	0.81

tures stand out with the highest ACC and F1 of 0.95, indicating its effectiveness in correctly classifying instances. The TPR (0.94) and TNR (0.96) are also commendable, suggesting a robust classification performance.

EfficientNetB0 follows closely, with an ACC of 0.92 and an F1 of 0.90. The TPR and TNR are also quite satisfactory (0.90 and 0.91, respectively). These results indicate that EfficientNetB0 is a strong contender in terms of ACC and can minimize false positives and negatives.

ResNet-50 and VGG-16 also demonstrate noteworthy performance, achieving accuracies and F1 of 0.94 and 0.93, respectively. These models show promising results in correctly classifying instances from the dataset.

On the other end of the spectrum, Inception v3 and Inception-ResNet-v2 exhibit lower performance, with ACC scores of 0.68 and 0.76, respectively. The F1 scores for these models are also lower (0.62 and 0.71), suggesting they may require alternative strategies to improve their classification performance.

It is important to note that the choice of CNN architecture significantly impacts the overall performance of the RF classifier. DenseNet-201 and EfficientNetB0 emerge as top performers in this analysis, demonstrating their suitability for feature extraction in this dataset.

b) HC features: Firstly, examining the best results from Table IX, LM stands out with the highest values across all measures. It achieves an ACC of 0.83 and F1 of 0.81. The high TPR (0.81) and TNR (0.82) suggest that LM performs well in correctly classifying positive and negative instances. This feature proves to be a strong candidate for effective feature extraction in this subset.

Conversely, Haar exhibits the lowest performance among the HC features when considering the worst results. With an ACC of 0.62, F1 of 0.63, and TPR of 0.62, Haar lags behind other features in correctly classifying instances. The low TNR of 0.64 suggests challenges in accurately identifying true negatives.

TABLE IX
PERFORMANCE COMPARISON OF HC FEATURES ON IDB-2 SUBSET

Feature	ACC	F1	TPR	TNR	MCC
Haar	0.62	0.63	0.62	0.64	0.61
HAR_ri	0.78	0.75	0.75	0.76	0.71
LBP_ri	0.77	0.73	0.73	0.75	0.73
Hist	0.68	0.66	0.66	0.68	0.65
CH_1	**0.79**	**0.76**	**0.75**	**0.77**	**0.75**
CH_2	**0.80**	**0.77**	**0.77**	**0.78**	**0.76**
LM	**0.83**	**0.81**	**0.81**	**0.82**	**0.82**
ZM	0.69	0.65	0.65	0.66	0.61

3) On the IDB-3 Subset: **Deep Features** Table X shows that DenseNet-201 features achieve outstanding results across all measures, with perfect scores for all five metrics (1.00). This suggests that DenseNet-201 features are highly discriminative, leading to flawless classification of this subset.

TABLE X
PERFORMANCE COMPARISON OF DEEP FEATURES ON IDB-3 SUBSET

Feature	ACC	F1	TPR	TNR	MCC
AlexNet	0.98	0.98	0.98	0.98	0.97
DarkNet-19	0.96	0.96	0.97	0.96	0.96
DarkNet-53	0.94	0.94	0.95	0.93	0.92
DenseNet-201	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
EfficientNetB0	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
GoogLeNet	0.94	0.94	0.95	0.93	0.93
Inception-ResNet--v2	0.97	0.96	0.96	0.96	0.96
Inception v3	0.89	0.88	0.90	0.87	0.86
NASNet-Large	0.96	0.95	0.96	0.95	0.95
NASNet-Mobile	0.91	0.90	0.92	0.89	0.89
ResNet-18	0.98	0.98	0.98	0.98	0.98
ResNet-50	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
ResNet-101	0.98	0.98	0.98	0.98	0.98
VGG-16	0.97	0.97	0.98	0.97	0.97
VGG-19	0.98	0.98	0.98	0.97	0.97
Xception	0.94	0.94	0.95	0.93	0.93

Conversely, Inception v3 exhibits relatively lower performance with an ACC of 0.89 and F1 of 0.88. While still achieving commendable results, the lower TPR (0.90) and TNR (0.87) indicate that Inception v3 features might be less robust for this dataset than other CNN architectures.

HC Features Concerning HC features, Table XI showcases that the CH_5 feature stands out with the highest ACC (0.89), F1 (0.88), TPR (0.87), and TNR (0.90). This results suggests that CH_1 is this dataset's most effective HC feature, providing a solid foundation for accurate classification.

On the other hand, the LBP_ri feature demonstrates the lowest performance among HC features, with an ACC of 0.60 and an F1 of 0.59. The TPR (0.59) and TNR (0.59) further high-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE XI
PERFORMANCE COMPARISON OF HC FEATURES ON IDB-3 SUBSET

Feature	ACC	F1	TPR	TNR	MCC
Haar	0.85	0.84	0.83	0.85	0.79
HAR_ri	0.68	0.68	0.68	0.69	0.61
LBP_ri	0.60	0.59	0.59	0.59	0.58
Hist	0.75	0.75	0.75	0.75	0.71
CH_1	**0.89**	**0.88**	**0.87**	**0.90**	**0.88**
CH_2	0.88	0.87	0.85	0.89	0.83
LM	**0.88**	**0.87**	**0.85**	**0.89**	**0.89**
ZM	**0.88**	**0.87**	**0.87**	**0.88**	**0.88**

light the challenges associated with this feature in capturing relevant information for the classification task on this subset.

C. Results With Feature Selection

This section presents the results obtained following the implementation of six FS techniques (*Correlation*, *Information Gain*, *Gain Ratio*, *ReliefF*, *One Rule*, and *Symmetrical Uncertainty*) on the features extracted from the three subsets of CR-IDB. Specifically, we considered the three best-performing features, deep or HC, identified in the baseline results for applying these FS methods.

1) On the IDB-1 subset: **Deep Features** In light of the baseline results, shown in Table VI, we considered the deep features extracted from DarkNet-53, DenseNet201, and ResNet-101. Table XII shows the performance results.

Considering the results obtained with the DarkNet-53 features, Symmetrical Uncertainty and Correlation reached the highest F1 value with 10 features (0.92) in contrast to the performance of Gain Ratio selector with 10 features (0.71). Symmetrical Uncertainty has become the best choice for these features in three cases where it achieved the best result. Furthermore, the results showed that decreasing the number of features did not affect the performance of the three selectors. Specifically, Information Gain had the same F1 value (0.89) with 100 and 50 features, and with Correlation and One Rule, the classifier reached the same performance with 50 and 25 features. Gain Ratio and Symmetrical Uncertainty yield another interesting finding: a 0.1 drop in the F1 value is achieved by halving the features.

In addition, examining the behavior of ReliefF with 100 and 50 features, there is an improvement in performance with 50 features as opposed to 100. This phenomenon suggests that with 50 features, the selectors deleted some redundant or irrelevant features, improving performance.

We can observe intriguing results using the features extracted from DenseNet-201. In particular, the highest (0.92) and the lowest (0.74) F1 values are obtained, respectively, by Correlation with 100 features and Gain Ratio with 10 features. Furthermore, Correlation proves to be the most compelling feature selection technique for each of the four cut-offs. The same performance attained with 50 and 25 features illustrates that not all the 50 features were necessary for the classification, which is another intriguing aspect of Correlation. Indeed, the F1 value with 50 features differs from the one with 100 features of only 0.1. In addition, a variance of 0.1 exists in the F1 when comparing the values obtained with 50 features versus those with 100 features. The phenomenon above is also evident in the case of the Gain Ratio and One Rule selectors, where there is a constant decrease of 0.1 in the F1 value from the 100 features to the 25 features. Similar to how the feature selection method, ReliefF, interacted with the feature set extracted by DarkNet-53, it demonstrates an enhancement in performance alongside a reduction in the number of features utilized.

About the features extracted from ResNet-101, four selectors, i.e., Correlation, Information Gain, One Rule, and ReliefF, achieved the best F1 value (0.88) with 100 features. The worst performance was obtained by selecting 10 features with the Gain Ratio selector (0.75). In two of the four scenarios, the selectors with the best results are Symmetrical Uncertainty and Information Gain. The only selector whose performance remains constant with 50 and 25 features is One Rule. Another noteworthy finding relates to the F1 value with 50 features, which drops by just 0.1 compared to the value of 100 features using the selectors Information Gain and Gain Ratio. A comparable pattern of behavior can be observed in the ReliefF selection, whose F1 value with 25 features is only 0.1 less than that of 50 features. It is noteworthy that the two selectors had better performances with fewer features. Specifically, One Rule had a higher F1 value with 50 features than with 100 features, and Information Gain achieved a better result with 25 features instead of 100.

After comparing all the deep features for the IDB-1 subset, we can conclude that DenseNet-201 extracted the best features, yielding results ranging from 0.92 to 0.87 with 100 features, 0.91 to 0.86 with 50 features, 0.91 to 0.84 with 25 features, and between 0.87 and 0.74 with 25 features.

HC Features In light of the baseline results, shown in Table VI, we considered the deep features extracted from DarkNet-53, DenseNet201, and ResNet-101. Table XII shows the performance results.

The three best HC features resulting from the baseline results presented in Table VII were CH_2, CH_1, and LM. The results of this experimentation are shown in Table XIII.

The optimal results using CH_2 were obtained when 10 features were selected using ReliefF and Gain Ratio and instead, by choosing just two features, Correlation generated the least important set of features, with an F1 value of 0.59. The best features are generally selected by Gain Ratio, which reached the best performance in two out of three cases.

The results obtained with CH_1 yield a little worse performance than CH_2 feature extraction. ReliefF yielded the greatest F1 of 0.86 with 10 features, while Gain Ratio yielded the worst (0.53) with 2 features. With this set of features, Correlation, One Rule, and ReliefF reached the best F1 value in two out of three cases.

Above all the HC features, LM obtained the lowest performance. The maximum F1 value reached is 0.84 with 10 features selected by Correlation, Gain Ratio and, Symmetrical Uncertainty and the minimum is 0.53 with 2 features selected by Gain Ratio. Moreover, Correlation and Gain Ratio

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE XII

PERFORMANCE COMPARISON OF THE THREE BEST DEEP FEATURES OUTLINED IN TABLE VI ON IDB-1 SUBSET WITH SIX FS TECHNIQUES. HIGHLIGHTED VALUES INDICATE THE BEST RESULTS FOR THE RESPECTIVE CUT-OFF POINTS

		Corr					GainR					InfoG					ReliefF					OneR					SU				
Features	TOP	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC
DarkNet-53	100	0.92	**0.92**	0.92	0.97	0.89	0.87	0.87	0.87	0.96	0.83	0.89	0.89	0.89	0.96	0.85	0.89	0.89	0.89	0.96	0.86	0.91	0.91	0.91	0.97	0.88	0.92	**0.92**	0.92	0.97	0.89
	50	0.90	0.90	0.90	0.97	0.87	0.86	0.86	0.86	0.95	0.82	0.89	0.89	0.89	0.96	0.86	0.90	0.90	0.90	0.97	0.86	0.87	0.86	0.87	0.95	0.82	0.91	**0.91**	0.91	0.97	0.88
	25	0.90	**0.90**	0.90	0.97	0.87	0.81	0.81	0.81	0.94	0.75	0.86	0.86	0.86	0.95	0.82	0.87	0.86	0.87	0.95	0.83	0.86	0.86	0.86	0.95	0.82	0.86	0.86	0.86	0.95	0.81
	10	0.81	0.81	0.81	0.94	0.75	0.71	0.71	0.71	0.90	0.62	0.82	**0.82**	0.82	0.94	0.76	0.82	**0.82**	0.82	0.94	0.76	0.80	0.79	0.79	0.93	0.73	0.83	**0.82**	0.83	0.94	0.77
DenseNet-201	100	0.92	**0.92**	0.92	0.97	0.90	0.91	0.91	0.91	0.97	0.89	0.90	0.90	0.90	0.97	0.88	0.89	0.89	0.89	0.96	0.86	0.87	0.87	0.87	0.96	0.84	0.91	0.91	0.91	0.97	0.89
	50	0.91	**0.91**	0.91	0.97	0.89	0.90	0.90	0.90	0.97	0.87	0.88	0.88	0.88	0.96	0.85	0.92	**0.91**	0.92	0.97	0.89	0.86	0.86	0.86	0.95	0.82	0.89	0.89	0.89	0.97	0.86
	25	0.91	**0.91**	0.91	0.97	0.89	0.89	0.89	0.89	0.96	0.86	0.84	0.84	0.84	0.95	0.79	0.87	0.87	0.87	0.96	0.83	0.86	0.85	0.95	0.86	0.81	0.84	0.84	0.84	0.96	0.79
	10	0.87	**0.87**	0.87	0.96	0.83	0.74	0.74	0.74	0.91	0.66	0.78	0.77	0.78	0.93	0.71	0.84	0.83	0.84	0.94	0.79	0.82	0.82	0.82	0.94	0.76	0.81	0.80	0.81	0.94	0.75
ResNet-101	100	0.88	**0.88**	0.88	0.96	0.84	0.87	0.87	0.87	0.96	0.83	0.88	**0.88**	0.88	0.96	0.84	0.88	**0.88**	0.88	0.96	0.85	0.88	**0.88**	0.88	0.96	0.85	0.87	0.86	0.87	0.96	0.83
	50	0.86	0.86	0.86	0.95	0.82	0.86	0.86	0.86	0.95	0.82	0.87	0.87	0.87	0.96	0.83	0.85	0.85	0.85	0.95	0.80	0.86	0.86	0.86	0.95	0.82	0.88	**0.88**	0.88	0.96	0.84
	25	0.83	0.83	0.83	0.94	0.78	0.80	0.79	0.79	0.93	0.73	0.88	**0.88**	0.88	0.96	0.84	0.84	0.84	0.84	0.95	0.79	0.86	0.86	0.86	0.95	0.81	0.87	0.87	0.87	0.96	0.83
	10	0.77	0.77	0.77	0.92	0.70	0.75	0.75	0.75	0.92	0.67	0.78	0.78	0.78	0.93	0.72	0.78	0.78	0.78	0.93	0.71	0.80	0.78	0.78	0.93	0.71	0.80	**0.79**	0.79	0.93	0.74

TABLE XIII

PERFORMANCE COMPARISON OF THE THREE BEST HC FEATURES OUTLINED IN TABLE VII ON IDB-1 SUBSET WITH SIX FS TECHNIQUES. HIGHLIGHTED VALUES INDICATE THE BEST RESULTS FOR THE RESPECTIVE CUT-OFF POINTS

		Corr					GainR					InfoG					ReliefF					OneR					SU				
Features	TOP	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC
CH_1	10	0.83	0.83	0.83	0.94	0.79	0.83	0.83	0.83	0.94	0.78	0.84	0.84	0.84	0.95	0.79	0.86	**0.86**	0.86	0.95	0.81	0.80	0.80	0.80	0.93	0.75	0.84	0.84	0.84	0.95	0.80
	5	0.75	**0.76**	0.75	0.92	0.69	0.75	0.75	0.75	0.92	0.67	0.73	0.73	0.73	0.91	0.64	0.76	**0.76**	0.76	0.92	0.69	0.76	**0.76**	0.76	0.92	0.69	0.72	0.72	0.72	0.91	0.63
	2	0.62	**0.61**	0.62	0.87	0.49	0.54	0.53	0.54	0.85	0.38	0.56	0.55	0.56	0.86	0.41	0.60	0.59	0.59	0.87	0.49	0.62	**0.61**	0.62	0.87	0.46	0.56	0.55	0.54	0.86	0.41
CH_2	10	0.90	0.90	0.90	0.97	0.87	0.91	**0.91**	0.91	0.97	0.88	0.89	0.89	0.89	0.96	0.86	0.91	**0.91**	0.91	0.97	0.89	0.90	0.90	0.90	0.97	0.87	0.90	0.90	0.90	0.97	0.87
	5	0.84	0.84	0.84	0.95	0.79	0.85	0.84	0.85	0.95	0.80	0.86	**0.86**	0.86	0.95	0.82	0.82	0.82	0.82	0.94	0.77	0.78	0.77	0.78	0.93	0.71	0.84	0.84	0.84	0.95	0.79
	2	0.60	0.59	0.59	0.87	0.47	0.62	**0.62**	0.62	0.87	0.50	0.62	0.61	0.62	0.87	0.49	0.61	0.61	0.61	0.87	0.49	0.62	0.61	0.62	0.87	0.48	0.62	**0.62**	0.62	0.87	0.50
LM	10	0.84	**0.84**	0.84	0.95	0.80	0.84	**0.84**	0.84	0.95	0.79	0.83	0.83	0.83	0.94	0.78	0.83	0.83	0.83	0.94	0.78	0.83	0.82	0.83	0.94	0.76	0.84	**0.84**	0.84	0.95	0.79
	5	0.75	0.75	0.75	0.92	0.67	0.76	**0.76**	0.76	0.92	0.69	0.73	0.73	0.73	0.91	0.64	0.75	0.75	0.75	0.92	0.67	0.74	0.74	0.74	0.91	0.67	0.73	0.73	0.73	0.91	0.64
	2	0.64	**0.63**	0.64	0.88	0.52	0.54	0.53	0.54	0.85	0.39	0.57	0.56	0.57	0.86	0.42	0.59	0.58	0.59	0.86	0.48	0.62	0.60	0.62	0.87	0.45	0.56	0.55	0.56	0.85	0.42

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

achieved the best result in two of the three examples.

CH_2 are the optimal HC features for this dataset. With them, the RF obtained an F1 value ranging from 0.91 to 0.89 when using 10 features, 0.86 to 0.77 when using five features, and 0.62 to 0.59 when using two features.

2) On the IDB-2 Subset: **Deep Features** Table XIV displays the results obtained employing the feature selection with features of the IDB-2 subset extracted by the three best deep features: DenseNet-201, EfficientNetB0 and ResNet-101, as shown in Table VIII.

Selecting 100 features using Information Gain yielded the best results for DenseNet-201 (0.95) and the worst was obtained by selecting 10 features with the ReliefF selector (0.77). Moreover, in two of the four cases, the Information Gain selector is the best for the features extracted by DenseNet-201. Furthermore, it can be observed that halving the number of features has no differences in the performance obtained with Gain Ratio, Correlation, and Symmetrical Uncertainty. For example, the top 25 and 50 of Gain Ratio have produced the same results. The performance with 100, 50, and 25 features is the same when ACC is considered. Furthermore, the F1 value decreases by 0.1 for Gain Ratio and Information Gain when the features are cut in half from 100 to 50. Symmetrical Uncertainty exhibits the same behavior but with 50 and 25 features.

When comparing the features extracted by EfficientNetB0 to the previous two, we can see a slight decrease in performance. Indeed, the best performance result is obtained by the Gain Ratio and Symmetrical Uncertainty selectors with an F1 value of 0.91 selecting 100 features and the worst with the selector Correlation, which achieved 0.73 of F1 with 10 chosen features. In this setting, unlike the others, only Correlation, One Rule, and Information Gain did not produce the best outcomes in nearly two out of every four cases. Additionally, we can see that the F1 value remains constant when comparing the ReliefF results with 100 and 50 features, indicating that not all 100 characteristics are relevant or crucial for the classification task. One Rule has another intriguing behavior when it chooses 50 and 25 features with F1 values that differ by 0.1.

Regarding the features extracted from ResNet-101, One Rule and Symmetrical Uncertainty reached the best results in two out of four cases. The best result is obtained by the ReliefF and Symmetrical Uncertainty selectors with 100 features (0.92) and the worst using 10 features selected by Correlation (0.69). One Rule and Symmetrical Uncertainty present an additional intriguing observation: reducing the features by half, respectively, from 100 to 50 and from 50 to 25, results in a 0.1 decrease in the F1 value.

Considering the methods discussed above, it is evident that the feature extracted from DenseNet-201 gave the RF superior performance. This combination achieved an F1 between 0.95 and 0.91 using 100 features, 0.94 and 0.90 with 50 features, 0.92 and 89 with 25 features, and 0.88 and 0.77 with 10 features.

HC Features Table XV presents the results obtained employing the feature selection with the three best HC features of the IDB-2 subset, as shown in Table IX.

In particular, we acknowledge that the HC features used for the IDB-2 subset are the same as those employed for the Podda one: LM, CH_2 and CH_1

Examining the results shown in Table XV, it is evident that the ReliefF with 10 chosen features (0.73) yielded the highest performance with CH_2. Instead, the lowest F1 value (0.50) was obtained by selecting two features using the One Rule method. Furthermore, we can see that no feature selection technique is superior to the others. Indeed, the best results for the three cut-offs (top 10, 5, and 2) are reached by three different selectors (ReliefF, Information Gain, and Correlation). In addition, the F1 values for Correlation with five and two features selected are the same, suggesting that the two features in the subset of five were more significant than the other three features.

Regarding the CH_1 feature outcomes, we can see that the application of the ReliefF selector with 100 features yielded the best value (0.77), while the Gain Ratio with 2 features results in the worst performance (0.55). Gain Ratio is the best selection for these features, outperforming the others in two out of three cases.

Using LM, the greatest F1 score obtained was 0.75. The selector used for reaching this result is Information Gain with 10 features. Moreover, the selection of two features by Gain Ratio and One Rule (0.51) produced the lowest performance. In this scenario, any particular feature selection technique has no superiority over others.

The best HC feature for this subset is certainly CH_1. Using 10 features, it attained an F1 between 0.73 and 0.77; for five features, the range was between 0.64 and 0.69, while for two features, it was between 0.51 and 0.55.

From a general point of view, we notice that the HC features struggled in this subset, while the deep feature produced the best performance overall.

3) On the IDB-3 Subset: **Deep Features** In the same way as the IDB-2 subset, the CNNs DenseNet-201, EfficientNetB0, and ResNet-50 were used to extract deep features for applying FS techniques. This was done after evaluating the baseline results in Table X. The results obtained after applying FS on the deep features in this subset are shown in Table XVI.

The DenseNet-201 features provided the best performance compared to the dataset's other features, exhibiting the same behavior as previously discovered. When Correlation selects 100 features, the F1 value is highest (0.99), and the lowest value is 0.86 obtained with 10 features selected by Gain Ratio. Correlation and Information Gain performed best in two out of four cases. Furthermore, Correlation and One Rule, when 50 and 25 features are selected, had the same performance. An interesting pattern emerges from the selection of 50 features using Correlation, Gain Ratio, and Symmetrical Uncertainty, where the F1 value differs by 0.1 compared to the value obtained using 100 features. The same behavior can be observed with ReliefF selecting 25 features. Furthermore, Information Gain performed better when the features were

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TABLE XIV

PERFORMANCE COMPARISON OF THE THREE BEST DEEP FEATURES OUTLINED IN TABLE VIII ON IDB-2 SUBSET WITH SIX FS TECHNIQUES. HIGHLIGHTED VALUES INDICATE THE BEST RESULTS FOR THE RESPECTIVE CUT-OFF POINTS

		Corr					GainR					InfoG					ReliefF					OneR					SU				
Features	TOP	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC
DenseNet-201	100	0.94	0.93	0.94	0.98	0.92	0.91	0.91	0.91	0.98	0.89	0.96	**0.95**	0.95	0.99	0.95	0.94	0.94	0.94	0.98	0.93	0.94	0.94	0.94	0.98	0.93	0.94	0.93	0.93	0.98	0.92
	50	0.90	0.90	0.90	0.97	0.87	0.91	0.90	0.91	0.98	0.89	0.94	**0.94**	0.94	0.98	0.93	0.91	0.91	0.91	0.98	0.89	0.92	0.91	0.92	0.98	0.90	0.93	0.93	0.93	0.98	0.92
	25	0.90	0.90	0.90	0.97	0.88	0.91	0.90	0.91	0.98	0.89	0.91	0.91	0.91	0.98	0.89	0.90	0.89	0.90	0.97	0.87	0.90	0.90	0.90	0.97	0.88	0.93	**0.92**	0.93	0.98	0.91
	10	0.80	0.79	0.80	0.95	0.75	0.86	**0.88**	0.89	0.97	0.86	0.89	0.86	0.86	0.96	0.83	0.78	0.77	0.78	0.94	0.72	0.81	0.81	0.81	0.95	0.77	0.87	0.87	0.87	0.97	0.84
EfficientNetB0	100	0.89	0.88	0.89	0.97	0.86	0.92	**0.91**	0.92	0.98	0.90	0.87	0.87	0.87	0.97	0.84	0.89	0.89	0.89	0.98	0.87	0.90	0.90	0.90	0.98	0.88	0.91	**0.91**	0.91	0.98	0.89
	50	0.86	0.85	0.85	0.96	0.82	0.90	**0.89**	0.90	0.97	0.87	0.88	0.88	0.88	0.97	0.85	0.90	**0.89**	0.90	0.97	0.87	0.88	0.87	0.88	0.97	0.85	0.86	0.86	0.86	0.97	0.83
	25	0.79	0.78	0.79	0.95	0.74	0.84	0.83	0.84	0.96	0.80	0.84	0.84	0.83	0.96	0.80	0.88	**0.87**	0.88	0.97	0.85	0.86	0.86	0.86	0.96	0.83	0.82	0.82	0.83	0.95	0.78
	10	0.74	0.73	0.73	0.93	0.67	0.79	0.79	0.79	0.95	0.74	0.80	**0.80**	0.80	0.95	0.75	0.77	0.76	0.77	0.94	0.71	0.81	**0.80**	0.80	0.95	0.76	0.81	**0.80**	0.80	0.95	0.76
ResNet-101	100	0.87	0.87	0.87	0.97	0.84	0.91	0.90	0.91	0.98	0.88	0.89	0.89	0.89	0.97	0.86	0.92	**0.92**	0.98	0.93	0.91	0.92	0.91	0.91	0.98	0.90	0.92	**0.92**	0.92	0.98	0.90
	50	0.84	0.84	0.84	0.96	0.80	0.87	0.87	0.87	0.97	0.84	0.87	0.86	0.87	0.97	0.84	0.87	0.86	0.86	0.97	0.83	0.91	**0.90**	0.91	0.98	0.88	0.85	0.85	0.85	0.96	0.82
	25	0.81	0.80	0.81	0.95	0.77	0.84	0.83	0.83	0.96	0.80	0.84	**0.84**	0.84	0.96	0.80	0.82	0.82	0.82	0.96	0.77	0.85	0.84	0.84	0.96	0.81	0.85	**0.84**	0.85	0.96	0.81
	10	0.70	0.69	0.70	0.92	0.62	0.74	0.73	0.74	0.93	0.67	0.77	0.77	0.77	0.94	0.72	0.80	0.79	0.80	0.95	0.75	0.81	**0.81**	0.81	0.95	0.77	0.72	0.71	0.72	0.93	0.65

TABLE XV

PERFORMANCE COMPARISON OF THE THREE BEST HC FEATURES OUTLINED IN TABLE IX ON IDB-2 SUBSET WITH SIX FS TECHNIQUES. HIGHLIGHTED VALUES INDICATE THE BEST RESULTS FOR THE RESPECTIVE CUT-OFF POINTS

		Corr					GainR					InfoG					ReliefF					OneR					SU				
Features	TOP	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC
CH_1	10	0.75	0.74	0.75	0.94	0.74	0.74	0.74	0.74	0.94	0.74	0.76	0.75	0.76	0.94	0.75	0.77	**0.77**	0.77	0.94	0.77	0.75	0.75	0.75	0.94	0.75	0.74	0.73	0.73	0.93	0.73
	5	0.65	0.64	0.65	0.91	0.64	0.70	**0.69**	0.70	0.93	0.69	0.67	0.65	0.67	0.92	0.65	0.67	0.66	0.67	0.92	0.66	0.70	0.69	0.69	0.92	0.69	0.70	0.68	0.69	0.93	0.68
	2	0.54	0.52	0.53	0.88	0.52	0.55	**0.55**	0.57	0.89	0.55	0.57	0.54	0.55	0.89	0.54	0.54	0.54	0.54	0.89	0.53	0.54	0.53	0.54	0.88	0.54	0.55	0.54	0.55	0.89	0.54
CH_2	10	0.70	0.69	0.70	0.92	0.69	0.71	0.71	0.71	0.93	0.71	0.72	0.71	0.72	0.93	0.71	0.74	**0.73**	0.74	0.93	0.73	0.71	0.71	0.71	0.93	0.71	0.71	0.70	0.70	0.93	0.70
	5	0.64	0.63	0.64	0.91	0.63	0.67	0.66	0.67	0.92	0.66	0.69	**0.68**	0.68	0.92	0.68	0.66	0.65	0.66	0.91	0.65	0.66	0.65	0.65	0.91	0.65	0.66	0.65	0.65	0.91	0.65
	2	0.64	**0.63**	0.64	0.91	0.63	0.52	0.51	0.52	0.88	0.51	0.55	0.54	0.55	0.88	0.54	0.62	0.61	0.61	0.90	0.50	0.51	0.50	0.51	0.88	0.61	0.52	0.51	0.52	0.88	0.51
LM	10	0.75	0.74	0.74	0.94	0.74	0.74	0.74	0.74	0.94	0.74	0.76	**0.75**	0.76	0.94	0.75	0.75	0.74	0.75	0.94	0.74	0.75	0.74	0.74	0.94	0.74	0.74	0.73	0.74	0.94	0.73
	5	0.66	0.64	0.65	0.91	0.64	0.68	0.66	0.68	0.92	0.66	0.66	0.65	0.66	0.92	0.65	0.68	0.67	0.67	0.92	0.67	0.72	**0.71**	0.72	0.93	0.71	0.68	0.66	0.68	0.92	0.66
	2	0.53	0.52	0.53	0.88	0.52	0.53	0.51	0.53	0.88	0.51	0.54	0.53	0.54	0.88	0.53	0.55	**0.54**	0.55	0.89	0.51	0.52	0.51	0.52	0.88	0.54	0.54	0.53	0.54	0.88	0.53

TABLE XVI
PERFORMANCE COMPARISON OF THE THREE BEST DEEP FEATURES OUTLINED IN TABLE X ON IDB-3 SUBSET WITH SIX FS TECHNIQUES. HIGHLIGHTED VALUES INDICATE THE BEST RESULTS FOR THE RESPECTIVE CUT-OFF POINTS

		Corr					Gain Ratio					Information Gain					ReliefF					One Rule					SU				
Features	TOP	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC
DenseNet	100	0.99	**0.99**	0.99	1.00	0.99	0.98	0.98	0.98	1.00	0.98	0.98	0.98	0.98	1.00	0.98	0.98	0.98	0.98	1.00	0.98	0.98	0.98	0.98	1.00	0.98	0.99	**0.99**	0.99	1.00	0.98
	50	0.98	0.98	0.98	1.00	0.98	0.97	0.97	0.97	0.99	0.96	0.99	**0.99**	0.99	1.00	0.98	0.95	0.95	0.95	0.99	0.94	0.97	0.96	0.97	0.99	0.96	0.98	0.98	0.98	0.99	0.97
	25	0.98	**0.98**	0.98	0.99	0.97	0.95	0.94	0.94	0.99	0.93	0.96	0.95	0.96	0.99	0.95	0.94	0.94	0.94	0.98	0.93	0.96	0.96	0.96	0.99	0.95	0.95	0.95	0.95	0.99	0.94
	10	0.89	0.89	0.89	0.97	0.86	0.86	0.86	0.86	0.97	0.83	0.91	**0.91**	0.91	0.98	0.89	0.87	0.87	0.87	0.97	0.85	0.91	**0.91**	0.91	0.98	0.90	0.91	0.90	0.90	0.98	0.89
EfficientNetB0	100	0.98	0.98	0.98	1.00	0.98	0.98	0.98	0.98	0.99	0.97	0.99	**0.99**	0.99	1.00	0.98	0.97	0.97	0.97	0.99	0.96	0.99	**0.99**	0.99	1.00	0.98	0.98	0.98	0.98	1.00	0.97
	50	0.96	0.95	0.95	0.99	0.95	0.95	0.95	0.95	0.99	0.94	0.98	**0.98**	0.98	0.99	0.97	0.95	0.95	0.95	0.99	0.94	0.97	0.97	0.97	0.99	0.97	0.97	0.97	0.97	0.99	0.97
	25	0.93	0.92	0.93	0.98	0.91	0.92	0.91	0.92	0.98	0.90	0.94	0.94	0.94	0.99	0.93	0.94	0.94	0.94	0.99	0.93	0.95	0.95	0.95	0.99	0.94	0.96	**0.96**	0.96	0.99	0.95
	10	0.83	0.83	0.83	0.96	0.80	0.85	0.84	0.85	0.96	0.82	0.85	0.84	0.85	0.96	0.81	0.88	**0.87**	0.88	0.97	0.85	0.85	0.84	0.85	0.96	0.80	0.87	0.86	0.87	0.97	0.84
ResNet-50	100	0.98	0.98	0.98	1.00	0.98	0.97	0.97	0.97	0.99	0.96	0.99	**0.99**	0.99	1.00	0.99	0.98	0.98	0.98	1.00	0.98	0.99	**0.99**	0.99	1.00	0.99	0.99	**0.99**	0.99	1.00	0.99
	50	0.98	0.98	0.98	0.99	0.97	0.96	0.96	0.96	0.99	0.96	0.99	**0.99**	0.99	1.00	0.99	0.98	0.98	0.98	0.99	0.97	0.99	**0.99**	0.99	1.00	0.99	0.99	**0.99**	0.99	1.00	0.98
	25	0.96	0.96	0.96	0.99	0.95	0.82	0.80	0.82	0.96	0.77	0.97	**0.97**	0.97	0.99	0.97	0.96	0.96	0.96	0.99	0.95	0.96	0.96	0.96	0.99	0.96	0.96	0.96	0.96	0.99	0.95
	10	0.85	0.85	0.85	0.96	0.81	0.73	0.71	0.73	0.93	0.66	0.90	0.90	0.90	0.98	0.88	0.88	0.88	0.88	0.97	0.86	0.92	**0.92**	0.92	0.98	0.90	0.89	0.89	0.89	0.97	0.87

TABLE XVII
PERFORMANCE COMPARISON OF THE THREE BEST HC FEATURES OUTLINED IN TABLE XI ON IDB-3 SUBSET WITH SIX FS TECHNIQUES. HIGHLIGHTED VALUES INDICATE THE BEST RESULTS FOR THE RESPECTIVE CUT-OFF POINTS

		Corr					GainR					InfoG					ReliefF					OneR					SU				
Features	TOP	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC	ACC	F1	TPR	TNR	MCC
CH_1	10	0.92	0.91	0.92	0.98	0.90	0.91	0.91	0.91	0.98	0.89	0.92	0.91	0.92	0.98	0.90	0.92	**0.92**	0.92	0.98	0.91	0.92	0.91	0.92	0.98	0.89	0.92	**0.92**	0.92	0.98	0.90
	5	0.89	0.89	0.89	0.98	0.87	0.87	0.87	0.87	0.97	0.85	0.90	**0.90**	0.90	0.98	0.88	0.88	0.88	0.88	0.97	0.85	0.87	0.87	0.87	0.97	0.85	0.88	0.88	0.88	0.97	0.85
	2	0.83	**0.83**	0.83	0.96	0.80	0.66	0.66	0.66	0.92	0.58	0.74	0.74	0.74	0.94	0.69	0.83	**0.83**	0.83	0.96	0.80	0.73	0.73	0.73	0.94	0.67	0.77	0.76	0.76	0.95	0.71
LM	10	0.92	0.91	0.92	0.98	0.89	0.91	0.91	0.91	0.98	0.89	0.92	0.91	0.92	0.98	0.90	0.92	**0.92**	0.92	0.98	0.90	0.91	0.91	0.91	0.98	0.89	0.91	0.91	0.91	0.98	0.89
	5	0.88	0.88	0.88	0.97	0.85	0.88	0.88	0.88	0.97	0.85	0.88	0.87	0.88	0.97	0.85	0.88	0.88	0.88	0.97	0.85	0.89	**0.89**	0.89	0.98	0.87	0.88	0.88	0.88	0.97	0.86
	2	0.83	**0.83**	0.83	0.96	0.80	0.62	0.61	0.62	0.92	0.54	0.70	0.70	0.70	0.93	0.63	0.83	**0.83**	0.83	0.96	0.80	0.73	0.72	0.73	0.94	0.67	0.77	0.77	0.77	0.95	0.72
ZM	10	0.92	**0.92**	0.92	0.98	0.91	0.91	0.91	0.91	0.98	0.89	0.91	0.91	0.91	0.98	0.90	0.91	0.91	0.91	0.98	0.89	0.92	0.91	0.91	0.98	0.90	0.91	0.91	0.91	0.98	0.89
	5	0.90	0.90	0.90	0.98	0.88	0.90	0.90	0.90	0.98	0.88	0.91	0.90	0.90	0.98	0.89	0.91	**0.91**	0.91	0.98	0.89	0.89	0.89	0.89	0.98	0.87	0.90	0.90	0.90	0.98	0.88
	2	0.72	0.72	0.72	0.94	0.66	0.77	0.76	0.77	0.95	0.72	0.84	**0.83**	0.84	0.96	0.80	0.78	0.77	0.78	0.95	0.73	0.83	**0.83**	0.83	0.96	0.80	0.84	**0.83**	0.84	0.96	0.80

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

reduced, probably to eliminate redundant or unnecessary features.

The experiments performed with the features extracted by EfficientNet led to good performance. In particular, the best was reached by selecting 100 features with Information Gain and One Rule selectors (0.99) and the worst with 10 features selected by Correlation (0.83). The best selector for these features is Information Gain, which had the best result in two out of four cases.

Moreover, Information Gain and Symmetrical Uncertainty exhibited comparable performances with 100 and 50 selected features, differing only by a marginal 0.1. The same pattern can be observed with ReliefF and Symmetrical Uncertainty with 50 and 25 features.

Using the features extracted by ResNet-50, the F1 values obtained are lower when compared to DenseNet-201 results but higher if examined against EfficientNetB0 ones. Indeed, the highest F1 value is 0.99, resulting in selecting 100 features with Information Gain, One Rule, and Symmetrical Uncertainty, and the lowest is 0.71 obtained with 10 features by Gain Ratio. Information Gain and One Rule stand out as the most effective selectors for these features, achieving the top results in three out of four cases. The results of selecting 100 and 50 features are identical with all selectors, except for Gain Ratio, whose F1 value changes by 0.1.

As previously stated, DenseNet-201 extracts the best features for this dataset, with an F1 value ranging from 0.99 and 0.98 with 100 features, 0.99 and 0.95 with 50 features, 0.98 and 0.94 with 25 features and 0.91 and 0.86 with 10 features.

HC Features The HC methods employed for the IDB-3 subset include ZM, LM, and CH_1.

CH_1 demonstrated worthy performance, achieving the highest F1 value of 0.92 when utilized by ReliefF and Symmetrical Uncertainty, which selected 10 features. Conversely, the lowest F1 value was observed when Gain Ratio chose 2 features (0.62). In two out of three cases, the most effective selector is ReliefF. The F1 value varies by only 0.1 when comparing Information Gain outcomes with 100 and 50 features, indicating that not all 10 obtained features are necessary for the classification.

Even the performance with LM is outstanding. Indeed, the best result, obtained with the selection of ReliefF, reached an F1 value of 0.92 while the worst (0.61) is obtained with the Gain Ratio selector. As with CH_1, ReliefF was the top selector in two out of three cases.

When compared to the other HC features, ZM produced the best results. Correlation reached the best (0.92) and the worst (0.72) F1 value, selecting respectively 10 and 2 features. Similar to the previous scenarios examined, ReliefF is the best selector in two of the three cases. Furthermore, the same performance is obtained when ReliefF selected 10 and 5 features. Moreover, the performance obtained by Gain Ratio, Information Gain, and Symmetrical Uncertainty with 5 features differs from the one with 10 features of only 0.1.

ZM features consistently yielded the most favorable performances for this dataset, showcasing F1 values spanning from 0.92 to 0.91 with 10 features, 0.91 to 0.89 with 5 features, and 0.83 to 0.72 with 2 features.

V. Discussion

This section provides a comprehensive analysis of classification performance, covering the baseline results (Section V-A) and the results obtained after applying FS techniques (Section V-B). In addition, in Section V-C, we present a comparison with previous research to emphasize the limitations of traditional visual inspection and expert experience methods, highlighting the need to utilize artificial intelligence methods for inspection. Finally, Section V-D briefly discusses the main limitations of the proposed approach.

A. On Baseline Results

When comparing Tables VI and VII, it is noteworthy that the RF Classifier with deep features generally performs better than the one with HC features on the IDB-1 subset. DenseNet-201, among CNN features, and CH_2, among HC features, have emerged as the best-performing representations, highlighting that, according to the considered scenario, selecting the right features can maximize the classification results, even if applied to HC features.

Comparing Tables VIII and IX, it is evident that the RF trained with deep features outperforms the one with HC features on the IDB-2 subset. DenseNet-201, among the deep features, demonstrates superior performance, emphasizing the significance of deep learning representations for effective feature extraction.

While HC features, such as CH_1, show moderate success, they generally fall behind CNN features regarding classification accuracy and robustness. The results reinforce that deep learning representations, particularly from models like DenseNet-201, are well-suited for capturing complex patterns in this subset.

Upon comparing Tables X and XI, it is clear that the RF trained with deep features obtained exceptional performance on the IDB-3 subset, mainly when using DenseNet-201 ones. On the other hand, the RF trained with HC features yielded more diverse results, with CH_1 outperforming the other HC features and LBP_ri showing lower effectiveness.

Finally, the MCC scores achieved demonstrate the exceptional performance of the proposed approach. These consistently high MCC values underscore the model's robustness and ability to maintain classification accuracy even in the presence of unbalanced datasets, further validating the effectiveness of our methodology.

B. On FS Results

Fig. 7 showcases the efficacy of various FS methods across the different subsets, taking into account varying cut-off points (100, 50, 25, 10 for the deep features, and 10, 5, and 2 for the HC). For the sake of comparison, the baseline results (all features) are also reported. The best-performing features, encompassing both HC and deep, were chosen for each subset based on the F1 value obtained (see Section IV-C).

Nearly all FS techniques exhibited constant performance for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

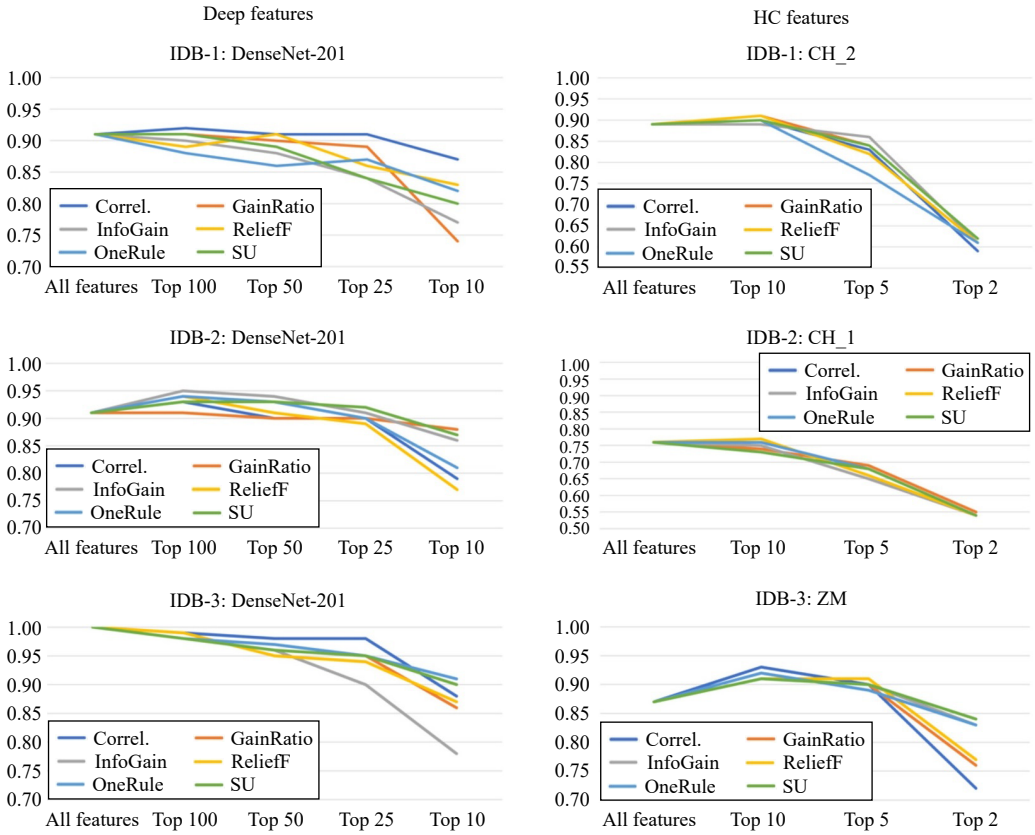


Fig. 7. Comparison of the results obtained by the different FS methods combined with RF and considering different cut-offs. The figure reports the F1 trends for each dataset with the different cut-offs for the best HC and deep features. The best HC and deep features were selected based on the F1 score obtained. Each graph also compares the baseline results obtained without applying any feature selection, therefore considering all the features and not a subset of them. The label "all features" is the baseline for each graph.

the first three cut-offs regarding the deep features trend. When limiting the selection to just 10 features, the performance deteriorates, indicating that this amount is insufficient to accurately describe the dataset. Furthermore, no FS approach outperforms the others overall; instead, they all have comparable tendencies.

Examining the HC features trend, the performance diminished with each successive cut-off. This implies that the existing features are adequate for the given task and that an additional FS operation is unnecessary.

Comparing the FS and baseline tables reveals that the outcomes are relatively similar despite substantially reducing the number of features. The reduction is notable, decreasing from around 1000 features to 100 with CNNs and halving for HC features. This indicates that not all extracted features contribute significantly to classification, and performing dimensionality reduction simplifies the model without compromising performance.

Thanks to the results obtained, we propose CRDet, a novel AI-based framework built upon an RF classifier trained with features extracted from a pretrained DenseNet-201 and further reduced with a Corr FS technique with a cut-off of 100. The RF classifier uses the same hyperparameters as in the experiments, consisting of 100 trees and $\log_2(n)+1$ random features. The proposed framework addresses the cheese ripeness task from multiple points of view, represented by the three different image subsets contained in CR-IDB, but also offers a robust and practical solution that relies on off-the-shelf methods without the burden of training complex networks.

Finally, using the feature selection strategies, the MCC scores confirmed the excellent performance of the proposed approach, further validating the effectiveness of our methodology even in presence of a reduced number of features.

C. Comparison With Literature

To the best of our knowledge, the approach proposed in this study represents the first systematic evaluation of cheese ripening of multiple types and from different dairy industries. In contrast, in previous methodologies developed in this context, our research team employed photographic imaging techniques, achieving an accuracy of up to 0.98 in classifying four distinct ripeness stages of *Pecorino* soft cheese produced by a dairy company in Sardinia [8] and 0.99 in the classification of six ripeness stage of cow's milk [68].

While valuable, traditional visual inspection and expert evaluation methods are inherently limited by subjectivity and variability in human judgment. Studies have shown that expert assessments can lead to inconsistent classifications, particularly in cases where ripeness stages exhibit subtle differences. This inconsistency underscores the necessity for more reliable and objective methods, such as those offered by artificial

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

intelligence.

In the existing literature, alternative methodologies predominantly focus on chemical and spectroscopic parameters or explore various cheese types and ripening durations [69], [70], [71], [6]. Consequently, a direct head-to-head comparison with previously established methods is not feasible.

Nevertheless, some studies have reported quantitative performance measures that highlight the limitations of traditional approaches. For example, Del Campo *et al.* introduced a Principal Component Analysis-based model capable of discriminating among four ripening stages of *Emmental* cheese using mid-infrared spectroscopy. Their model achieved a cross-validation accuracy of 0.87 based on a sample size of 14, with a separate test set yielding an accuracy of only 0.57 [71]. This significant drop in accuracy illustrates the challenges faced by conventional methods, particularly when applied to diverse samples over extended periods.

Additionally, Soto-Barajas *et al.* proposed an artificial neural network trained on data related to the fatty acid composition and near-infrared (NIR) spectra of sixteen milk mixtures. Their approach achieved an accuracy of 0.50 in discriminating among different cheese types while achieving a perfect accuracy of 1.00 in classifying unknown cheese samples based on ripening time [70]. However, the limited accuracy in distinguishing cheese types further emphasizes the need for more robust methodologies.

In contrast to these prior studies, our work is distinguished by its accurate classification performance in differentiating between various ripeness phases in three dairy industry case studies, represented in a novel proposed dataset. Furthermore, our method's noninvasive nature and end-to-end automation enhance its applicability in real-world settings, making it particularly valuable for identifying potential issues in cheese storage, such as inappropriate temperature conditions. This advancement not only addresses the limitations of traditional inspection methods but also underscores the critical role of artificial intelligence in enhancing the accuracy and reliability of cheese ripening assessments.

D. Limitations

The use of CV systems like CRDet has shown potential in monitoring cheese production, but they do have limitations. One major issue is their reliance on consistent lighting conditions, as variations in light direction and intensity can result in inaccurate assessments of cheese quality and characteristics. This aspect can be addressed with ad hoc robotic systems to improve environmental and acquisition conditions [15], [16]. Additionally, CV techniques may struggle with complex surface textures or irregularities, which can obscure important visual information needed for precise evaluations. Moreover, the consequent challenges associated with data acquisition can limit their effectiveness in real-world applications. Therefore, while CV approaches are valuable tools, it may be necessary to complement them with other analytical techniques to ensure comprehensive quality control in cheese production.

VI. Conclusion

This study proposed the first public dataset encompassing images of different cheese wheels depicting various products at distinct stages of ripening, CR-IDB, on which extensive research was conducted.

We delved into numerous experiments to discern the ripeness of various cheese types using three innovative image sets. The outcomes produced by RF, leveraging features extracted from 14 CNN architectures and seven HC features, combined with six FS techniques, underscore the proposed methodology's pivotal role within the cheese quality analysis domain. Furthermore, the FS procedure pinpointed the most relevant and distinctive features crucial for accurately classifying cheese wheels.

The results presented in this work highlight the importance of selecting appropriate feature representations for cheese ripeness classification. Deep features generally performed better than HC ones. Still, the specific architecture or HC feature selected played a crucial role in determining the classifier's performance on all three subsets of CR-IDB.

Our investigation demonstrated that selecting optimal features tailored for both the specific task and the classifier used is critical. We found that streamlining the feature set did not compromise classifier performance, and we achieved remarkable results with a less complex model, using only one-tenth of the considered initial features.

The promising results obtained led us to propose a new AI-based framework called CRDet. This framework uses an RF classifier that is trained with features extracted from a pre-trained DenseNet-201 and further reduced with a Corr FS technique. The excellent results obtained with this strategies in terms of F1-score (above 90%) and MCC (in line with F1) underscore the method's impact of the cheese ripeness classification task on multiple and diverse data subsets, proving its effectiveness even in presence of a reduced number of feature. The aim of CRDet is to address the cheese ripeness task from multiple perspectives, represented by the three different image subsets contained in CR-IDB. In addition, the proposed framework offers a practical and robust solution that does not require the training or fine-tuning of complex networks. It relies on off-the-shelf methods, making it an accessible and efficient solution.

However, CV systems, such as CRDet, have some limitations in monitoring cheese production. While these systems can be effective, they face challenges like dependence on consistent lighting conditions and difficulties in assessing complex surface textures, which can hinder accurate evaluations. To enhance their effectiveness, it may be necessary to integrate CV techniques with other analytical methods for comprehensive quality control.

Finally, as for future improvements, a more thorough investigation will include further classifiers to make a holistic comparison. Additionally, integrating hybrid or embedded feature selection methods would enhance the analysis of cheese ripeness classification. Finally, exploring and optimizing feature sets could potentially improve classification results and consider training end-to-end deep learning methods, like CNN or Vision Transformer architectures, with adequate strategies to overcome the data limitation issue and prevent overfitting.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

REFERENCES

[1] T. Lei and D.-W. Sun, "Developments of nondestructive techniques for evaluating quality attributes of cheeses: A review," *Trends Food Sci. Technol.*, vol. 88, pp. 527–542, Jun. 2019.

[2] P. L. H. McSweeney, "Biochemistry of cheese ripening," *Int. J. Dairy Technol.*, vol. 57, no. 2-3, pp. 127–144, May 2004.

[3] A. Forde and G. F. Fitzgerald, "Biotechnological approaches to the understanding and improvement of mature cheese flavour," *Curr. Opin. Biotechnol.*, vol. 11, no. 5, pp. 484–489, Oct. 2000.

[4] P. F. Fox, J. M. Wallace, S. Morgan, C. M. Lynch, E. J. Niland, and J. Tobin, "Acceleration of cheese ripening," *Antonie van Leeuwenhoek*, vol. 70, no. 2-4, pp. 271–297, Oct. 1996.

[5] L. Sakkas, C. S. Pappas, and G. Moatsou, "FT-MIR analysis of water-soluble extracts during the ripening of sheep milk cheese with different phospholipid content," *Dairy*, vol. 2, no. 4, pp. 530–541, Sep. 2021.

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

[7] J. Dias, P. Lage, A. Garrido, E. Machado, C. Conceição, S. Gomes, A. Martins, A. Paulino, M. F. Duarte, and N. Alvarenga, "Evaluation of gas holes in "Queijo de Nisa" PDO cheese using computer vision," *J. Food Sci. Technol.*, vol. 58, no. 3, pp. 1072–1080, Mar. 2021.

[8] A. Loddo, C. Di Ruberto, G. Armano, and A. Manconi, "Automatic monitoring cheese ripeness using computer vision and artificial intelligence," *IEEE Access*, vol. 10, pp. 122612–122626, Jan. 2022.

[9] T. Jeliński, C.-J. Du, D.-W. Sun, and J. Fornal, "Inspection of the distribution and amount of ingredients in pasteurized cheese by computer vision," *J. Food Eng.*, vol. 83, no. 1, pp. 3–9, Nov. 2007.

[10] T. Lei, X.-H. Lin, and 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.*, vol. 13, no. 4, pp. 3119–3129, Dec. 2019.

[11] P. S. Minz and C. S. Saini, "Comparison of computer vision system and colour spectrophotometer for colour measurement of mozzarella cheese," *Appl. Food Res.*, vol. 1, no. 2, p. 100020, Dec. 2021.

[12] R. Iezzi, F. Locci, R. Ghiglietti, C. Belingheri, S. Francolino, and G. Mucchetti, "Parmigiano Reggiano and Grana Padano cheese curd grains size and distribution by image analysis," *LWT*, vol. 47, no. 2, pp. 380–385, Jul. 2012.

[13] M. Alinovi, G. Mucchetti, and F. Tidona, "Application of NIR spectroscopy and image analysis for the characterisation of grated Parmigiano-Reggiano cheese," *Int. Dairy J.*, vol. 92, pp. 50–58, May 2019.

[14] G. Mulas, R. Anedda, D. L. Longo, T. Roggio, and S. Uzzau, "An MRI method for monitoring the ripening of Grana Padano cheese," *Int. Dairy J.*, vol. 52, pp. 19–25, Jan. 2016.

[15] Z. Li, S. Li, A. Francis, and X. Luo, "A novel calibration system for robot arm via an open dataset and a learning perspective," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 69, no. 12, pp. 5169–5173, Dec. 2022.

[16] Z. Li, S. Li, and X. Luo, "An overview of calibration technology of industrial robots," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 1, pp. 23–36, Jan. 2021.

[17] Z. Li, S. Li, O. O. Bamasag, A. Alhothali, and X. Luo, "Diversified regularization enhanced training for effective manipulator calibration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8778–8790, Nov. 2023.

[18] G.-N. Zhu, Y. Zeng, Y. S. Teoh, E. Toh, C. Y. Wong, and I.-M. Chen, "A bin-picking benchmark for systematic evaluation of robotic-assisted food handling for line production," *IEEE/ASME Trans. Mech.*, vol. 28, no. 3, pp. 1778–1788, Jun. 2023.

[19] Y. Wang, M. Kang, Y. Liu, J. Li, K. Xue, X. Wang, J. Du, Y. Tian, Q. Ni, and F.-Y. Wang, "Can digital intelligence and cyber-physical-social systems achieve global food security and sustainability?," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 11, pp. 2070–2080, Nov. 2023.

[20] T. Lei and D.-W. Sun, "Developments of nondestructive techniques for evaluating quality attributes of cheeses: A review," *Trends Food Sci. Technol.*, vol. 88, pp. 527–542, Jun. 2019.

[21] 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, and D. Wechsler, "Mechanism and control of the eye formation in cheese," *Int. Dairy J.*, vol. 47, pp. 118–127, Aug. 2015.

[22] D. Huc, F. Mariette, S. Challois, J. Barreau, G. Moulin, and C. Michon, "Multi-scale investigation of eyes in semi-hard cheese," *Innov. Food Sci. Emerg. Technol.*, vol. 24, pp. 106–112, Aug. 2014.

[23] P. Schuetz, D. Guggisberg, I. Jerjen, M. T. Fröhlich-Wyder, J. Hofmann, D. Wechsler, A. Flisch, W. Bisig, U. Sennhauser, and H. P. Bachmann, "Quantitative comparison of the eye formation in cheese using radiography and computed tomography data," *Int. Dairy J.*, vol. 31, no. 2, pp. 150–155, Aug. 2013.

[24] A. Kulmyrzaev, É. Dufour, Y. Noël, M. Hanafi, R. Karoui, E. M. Qannari, and 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.*, vol. 15, no. 6-9, pp. 669–678, Jun.-Sep. 2005.

[25] Z. Ozbekova and A. Kulmyrzaev, "Fluorescence spectroscopy as a non destructive method to predict rheological characteristics of tilsit cheese," *J. Food Eng.*, vol. 210, pp. 42–49, Oct. 2017.

[26] M. L. Oca, M. C. Ortiz, L. A. Sarabia, A. E. Gredilla, and 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*, vol. 99, pp. 558–565, Sep. 2012.

[27] R. Karoui, A. M. Mouazen, É. Dufour, L. Pillonel, E. Schaller, D. Picque, J. De Baerdemaeker, and 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.*, vol. 223, no. 1, pp. 44–50, May 2006.

[28] E. S. Madalozzo, E. Sauer, and N. Nagata, "Determination of fat, protein and moisture in ricotta cheese by near infrared spectroscopy and multivariate calibration," *J. Food Sci. Technol.*, vol. 52, no. 3, pp. 1649–1655, Mar. 2015.

[29] M. J. Lerma-García, A. Gori, L. Cerretani, E. F. Simó-Alfonso, and 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.*, vol. 93, no. 10, pp. 4490–4496, Oct. 2010.

[30] C. Cevoli, A. Gori, M. Nocetti, L. Cuibus, M. F. Caboni, and A. Fabbri, "FT-NIR and FT-MIR spectroscopy to discriminate competitors, non compliance and compliance grated Parmigiano Reggiano cheese," *Food Res. Int.*, vol. 52, no. 1, pp. 214–220, Jun. 2013.

[31] A. Crespo, A. Martín, S. Ruiz-Moyano, M. J. Benito, M. Rufo, J. M. Paniagua, and A. Jiménez, "Application of ultrasound for quality control of torta del casar cheese ripening," *J. Dairy Sci.*, vol. 103, no. 10, pp. 8808–8821, Oct. 2020.

[32] A. Loddo, CR-IDB: The cheese ripeness image data base, 2024.

[33] L. Putzu, A. Loddo, and C. Di Ruberto, "Invariant moments, textural and deep features for diagnostic MR and CT image retrieval," in *Proc. 19th Int. Conf. Computer Analysis of Images and Patterns*, 2021, pp. 287–297.

[34] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. 2001 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Kauai, HI, USA, 2001, pp. I.

[35] L. Putzu and C. Di Ruberto, "Rotation invariant co-occurrence matrix features," in *Proc. 19th Int. Conf. Image Analysis and Processing*, Catania, Italy, 2017, pp. 391–401.

[36] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

[37] R. Mukundan, S. H. Ong, and P. A. Lee, "Image analysis by tchebichef moments," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1357–1364, Sep. 2001.

[38] C. Di Ruberto, L. Putzu, and G. Rodriguez, "Fast and accurate computation of orthogonal moments for texture analysis," *Pattern Recognit.*, vol. 83, pp. 498–510, Nov. 2018.

[39] M. R. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Am.*, vol. 70, no. 8, pp. 920–930, Aug. 1980.

[40] C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496–513, Jul. 1988.

[41] M. Oujaoura, B. Minaoui, and M. Fakir, "Image annotation by moments," in *Moments and Moment Invariants-Theory and Applications*, G. A. Papakostas, Ed. Xanthi, Greece: Science Gate Publishing, 2014, pp. 227–252.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Information Processing Systems*, Lake Tahoe, Nevada, 2012, pp. 1097–1105.

[43] J. Redmon, S. K. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 779–788.

[44] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, 2017, pp. 2261–2269.

[45] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Machine Learning*, Long Beach, CA, USA, 2019, pp. 6105–6114.

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 1–9.

[47] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 2818–2826.

[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2017, pp. 4278–4284.

[49] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, 2018, pp. 8697–8710.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770–778.

[51] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learning Representations*, San Diego, CA, USA, 2015.

[52] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 1800–1807.

[53] E. Raei, A. Akbari Asanjan, M. R. Nikoo, M. Sadegh, S. Pourshahabi, and J. F. Adamowski, "A deep learning image segmentation model for agricultural irrigation system classification," *Comput. Electron. Agric.*, vol. 198, p. 106977, Jul. 2022.

[54] A. Loddo, M. Loddo, and C. Di Ruberto, "A novel deep learning based approach for seed image classification and retrieval," *Comput. Electron. Agric.*, vol. 187, p. 106269, Aug. 2021.

[55] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[56] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Machine Learning*, Lille, France, 2015, pp. 448–456.

[57] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[58] R. R. Zebari, A. Mohsin Abdulazeez, D. Q. Zeebaree, D. A. Zebari, and J. N. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *J. Appl. Sci. Technol. Trends*, vol. 1, no. 1, pp. 56–70, May 2020.

[59] A. M. A. and P. A. Thomas, "Comparative review of feature selection and classification modeling," in *Proc. 2019 Int. Conf. Advances in Computing, Communication and Control*, Mumbai, India, 2019, pp. 1–9.

[60] L. M. Cannas, N. Dessi, and B. Pes, "Assessing similarity of feature selection techniques in high-dimensional domains," *Pattern Recogn. Lett.*, vol. 34, no. 12, pp. 1446–1453, Sep. 2013.

[61] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*. 2nd ed. London, UK: Pearson, 2019.

[62] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. 4th ed. Amsterdam, The Netherlands: Morgan Kaufmann, 2017.

[63] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Inform.*, vol. 85, pp. 189–203, Sep. 2018.

[64] N. Dessì and B. Pes, "Similarity of feature selection methods: An empirical study across data intensive classification tasks," *Expert Syst. Appl.*, vol. 42, no. 10, pp. 4632–4642, Jun. 2015.

[65] J.-M. Jo, "Effectiveness of normalization pre-processing of big data to the machine learning performance," *J. Korea Inst. Electron. Commun. Sci.*, vol. 14, no. 3, pp. 547–552, Jun. 2019.

[66] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Appl. Soft Comput.*, vol. 133, p. 109924, Jan. 2023.

[67] 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, and E. Duchesnay, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[68] L. Zedda, A. Perniciano, A. Loddo, and C. Di Ruberto, "Understanding cheese ripeness: An artificial intelligence-based approach for hierarchical classification," *Knowl. Based Syst.*, vol. 295, p. 111833, Jul. 2024.

[69] P. Schuetz, D. Guggisberg, M.-T. Fröhlich-Wyder, and D. Wechsler, "Software comparison for the analysis of cheese eyes in x-ray computed tomography," *Int. Dairy J.*, vol. 63, pp. 62–69, Dec. 2016.

[70] 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, and 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*, vol. 116, pp. 50–55, Nov. 2013.

[71] S. T. M. Del Campo, N. Bonnaire, D. Picque, and G. Corrieu, "Initial studies into the characterisation of ripening stages of emmental cheeses by mid-infrared spectroscopy," *Dairy Sci. Technol.*, vol. 89, no. 2, pp. 155–167, Mar. 2009.

Alessandra Perniciano received the B.Sc. and the M.Sc. degrees from the University of Cagliari in 2020 and 2023, respectively. Currently, she is pursuing a Ph.D. candidate at the Department of Mathematics and Computer Science at the University of Cagliari (Italy), focusing on artificial intelligence and machine learning techniques for analyzing and processing hybrid biomedical data. Her research interests include data mining and machine learning, with a particular emphasis on class balancing, feature

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

extraction, feature selection, and bioinformatics. She has authored multiple scientific publications and presented her work at international conferences.

Luca Zedda is a Ph.D. candidate in computer science at the University of Cagliari, specializing in artificial intelligence and its applications in biomedical and clinical research. His work spans foundational topics in deep learning, computer vision, and machine learning, with a strong focus on medical imaging, radiomics, and high-dimensional biomedical data analysis. He is particularly interested in advancing self-supervised learning techniques, transformer architectures, and foundation models to address challenges in healthcare and life sciences. Luca's research aims to develop state-of-the-art AI systems that enhance disease detection, diagnosis, and data-driven healthcare solutions. He has contributed to various scientific publications and presented at international conferences, reflecting his commitment to pushing the boundaries of AI-driven innovations in scientific and clinical contexts.

Cecilia Di Ruberto is an Associate Professor of computer science at the Department of Mathematics and Computer Science, University of Cagliari, Italy. She received the M.Sc. in computer science from the University of Salerno, Italy, in 1990 and the Ph.D. degree in computer science from the University of Naples, Italy, in 1995. Currently, her research interests include computer vision, image retrieval, medical image analysis, pattern recognition, and machine learning. She has been working in microscopic image analysis, particularly in blood smear image analysis for cell counting, malaria parasites detection and classification, and leukemia detection. Author of over 100 scientific papers in peer-reviewed journals and international conference proceedings.

Barbara Pes (Member, IEEE) is an Associate Professor at the Department of Mathematics and Computer Science of the University of Cagliari (Italy), where she taught/teaches foundations of computer science, database and data mining courses. She has participated in several research projects on web-based information systems, service-oriented architectures, data integration, high-dimensional data analysis, and bioinformatics. She is the author/co-author of over 80 papers published in international conferences, books, and journals. Currently, her main research interests are in data mining and machine learning, classification of high-dimensional data, and advanced feature selection methods.

Andrea Loddo (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from the University of Cagliari, in 2012, 2014, and 2019, respectively. His Ph.D. thesis addressed blood cell image analysis and classification issues to create new tools for automatic diagnosis to support medical analysis. He is currently an Assistant Professor at the Department of Mathematics and Computer Science, University of Cagliari. He authorizes over 50 scientific manuscripts in peer-reviewed journals and international conference proceedings. His research interests include image analysis and processing, computer vision, pattern recognition, and machine and deep learning, with a focus on medical tasks.