



A novel deep ordinal classification approach for aesthetic quality control classification

Riccardo Rosati¹ · Luca Romeo^{1,2} · Víctor Manuel Vargas³ · Pedro Antonio Gutiérrez³ · César Hervás-Martínez³ · Emanuele Frontoni¹

Received: 19 January 2022 / Accepted: 31 January 2022 / Published online: 2 March 2022
© The Author(s) 2022, corrected publication 2022

Abstract

Nowadays, decision support systems (DSSs) are widely used in several application domains, from industrial to healthcare and medicine fields. Concerning the industrial scenario, we propose a DSS oriented to the aesthetic quality control (AQC) task, which has quickly established itself as one of the most crucial challenges of Industry 4.0. Taking into account the increasing amount of data in this domain, the application of machine learning (ML) and deep learning (DL) techniques offers great opportunities to automatize the overall AQC process. State-of-the-art is mainly oriented to approach this problem with a nominal DL classification method which does not exploit the ordinal structure of the AQC task, thus not penalizing the error among distant AQC classes (which is a relevant aspect for the real use case). The paper introduces a DL ordinal methodology for the AQC classification. Differently from other deep ordinal methods, we combined the standard categorical cross-entropy with the cumulative link model and we imposed the ordinal constraint via the thresholds and slope parameters. Experimental results were performed for solving an AQC task on a novel image dataset originated from a specific company's demand (i.e., aesthetic assessment of wooden stocks). We demonstrated how the proposed methodology is able to reduce misclassification errors (up to 0.937 quadratic weight kappa loss) among distant classes while overcoming other state-of-the-art deep ordinal models and reducing the bias factor related to the item geometry. The proposed DL approach was integrated as the main core of a DSS supported by Internet of Things (IoT) architecture that can support the human operator by reducing up to 90% the time needed for the qualitative analysis carried out manually in this specific domain.

Keywords Ordinal classification · Deep ordinal models · Quality control · Decision support system

1 Introduction

Learning to classify patterns from labeled examples and predicting discrete classes in a new unseen set is the main goal of supervised machine learning (ML). Starting from the notion of statistical learning, different losses can be potentially tailored for solving specific learning tasks in different challenging pattern recognition domains, from industrial [1, 2] to diagnostic applications [3, 4]. In the context of Industry 4.0, the increasing availability of data, the advancements in computing power and breakthroughs in algorithm development have led ML and deep learning (DL) methodologies to develop appealing solutions in different industrial areas such as predictive maintenance

[1, 5], decision support system (DSS) [6] and quality control (QC) [2, 7, 8]. Indeed, the QC task has quickly established itself as one of the most crucial and challenging industrial 4.0 scenarios [9], where the main objective is to detect production issues and classify the quality of the final product. The quality monitoring of the instrumentation/products/materials may potentially enable manufacturers to support the technicians during the process while reducing resource costs and improving productivity [10].

The application of ML and DL techniques offers great opportunities to automatize the overall QC process [11]. Indeed, these methodologies for QC have been employed in several industrial areas, but state-of-the-art is mainly oriented to present ad hoc rather than vanilla ML solutions capable of dealing with challenges of this domain, namely the intrinsic variability of the annotation procedure and the

Extended author information available on the last page of the article

difficulty to generalize across different sets. The aesthetic quality control (AQC) task is a non-metric QC task where the aesthetic aspect of the material is not measurable and is based on expert observation. In this domain, the classes of the target variable often exhibit a natural ordering. However, the natural ordinal structure of the problem is not usually exploited and modeled in the learning procedure. For these reasons, the state-of-the-art solutions include standard classification and regression models that do not completely solve the ordinal structure of the AQC task. This gap in the scientific literature lies the foundations to introduce a DL-based DSS driven by ordinal constraints for solving an AQC task. The proposed approach allows penalizing the misclassification errors that are further from the correct AQC classes. This outcome is also in line with the industrial demands in order to provide a DL-based DSS for AQC that is as aligned as possible with the human operator (human agency and oversight [12]).

1.1 Aesthetic quality control task

Quality control (QC) is a growing area in Industry 4.0 and a fundamental step for detecting production issues and for classifying the compliance of the finished product. Recently, the increasing amount of data in this scenario offers a great opportunity for ML and DL techniques to be the main core of a DSS that is able to automatize the overall QC process, saving time and resources and maximizing the performances, while easily generalizing in different contexts. As evidence of this, these methodologies for QC task have been employed in different domains. In the fabric and textile industry, DL approaches have been applied to perform leather and stitching classification, avoiding the operator visual inspection phase to identify stitching defects on material surfaces [13]. In the printing industry, a deep neural network soft sensor has been proposed, which compares the scanned surface to the used engraving file and performs an automatic quality control process by learning features through exposure to training data [14]. In the automotive industry, a DL-based approach has been adopted for automatic fault detection and isolation [15] and for the quality control of complex multistage manufacturing processes, where the product dimensional variability is a crucial factor and undetected defects can easily be propagated downstream [16]. All of these solutions focus on quantitative and deterministic analyses: the dimensional control, the inspection of the roughness of materials, the patterned fabric defect detection and the test of production parameters are all measurable evaluation procedures. In our previous work [17], we dealt with an unexplored and challenging QC application, which is the aesthetic evaluation of material, introducing the topic of aesthetic quality control (AQC) task. In this case, the DL

algorithm should model all those qualitative analyses that are strictly human dependent, subjective and not directly measurable: this aspect clearly increases the complexity of the classification task, and it becomes more and more apparent as the number of classes to be considered is higher. As demonstrated, approaching this problem with a nominal DL classification method (which does not exploit class order) causes a substantial drop in accuracy performance and an increase in misclassification errors even between widely distant classes, which represents the main fault from the industrial production perspective. Considering the ordinal nature of the problem, these issues can be addressed by overcoming the limitation of the nominal approach, in which the classes are not arranged in an appropriate ordered scale, by exploiting the gradual rank of the dataset classes with specific methodologies for ordinal classification.

1.2 Ordinal classification

Recently ordinal classification (also called ordinal regression) methods have been proven useful in different research areas, including medical research [18–20], computer vision [21–23], finance application [24], and environmental management [25]. An extensive review of ordinal classification approaches was provided in [26]. However, the introduction of these methodologies for solving an AQC task is not still explored in the ML literature. It is worth noting that ordinal classification approaches differentiate from the multipartite ranking problems where learning to rank strategy is applied to automatically construct a ranking model from training data [27, 28]. The multipartite ranking problem represents the state-of-the-art in many information retrieval applications [29]. Although ordinal classification can be potentially scaled for solving a multipartite ranking problem, they are pointwise approaches for classifying data, where a naturalistic order is encoded in the label.

Ordinal classification problems can be easily simplified into other standard problems using the round prediction of a regression model or a cost-sensitive penalty. These are considered standard approaches for solving the ordinal classification task, with the main limitation that they assume a distance between class labels which can influence the performance of the classifier. A specific method based on a cost-sensitive ordinal hyperplanes ranking algorithm has been used for human age estimation using face images [22]. The authors designed the cost of an individual binary classifier so that the misranking cost can be bounded by the total misclassification costs.

Another class of ordinal-based approaches is the ordinal decomposition strategy. Within this category, the multiple model approaches use several binary classification

branches to compute a series of cumulative probabilities. Although this approach introduces a large number of hyperparameters to be tuned, there are some works [20] that try to reduce the effect of this problem, by redesigning the output layer of the conventional deep neural network. Moreover, in the ordinal decomposition approaches, the relationships among different binary classifiers are often neglected. To try to alleviate this issue, it was proposed to learn an ordinal distribution of the problem and to optimize those binary classifiers simultaneously [30]. Similarly, a multiple ordinal regression algorithm to estimate the preferences of humans was proposed [31]. They maximized the sum of the margins between every consecutive class with respect to one or more rankings (e.g., perceived length and weight). An ordinal decomposition approach combined with a fully 3D convolutional neural network (CNN) network was used for assessing the level of neurological damage in Parkinson's disease (PD) patients and exploring the potential classification performance improvement in using ordinal label information [18]. A standard sigmoid function is provided in the output node, rather than using a softmax function for the output nodes. They trained a single convolutional model for solving simultaneously individual binary classification tasks, which were treated as multiple fully connected blocks.

The most natural strategy to handle the ordinal structure extends the standard regression task by assuming that a latent variable underlies the ordinal classes. In this general approach, called the threshold model, both the latent variable and the thresholds, which act, respectively, as a mapping function and ordinal constraints, need to be learned from the data. A threshold-based loss function is designed to model the ordinal values among multiple output variables [32]. The authors applied the kernel trick to provide a nonlinear extension of the model. Another work presented a structural distance metric for video-based face recognition [23]. Here the ordinal problem is designed as a non-convex integer program problem that firstly learns stable ordinal filters by projecting video data into a large-marginal ordinal space and then self-corrected the projected data in a structure low-rank strategy. A large margin ordinal regression formulation was also provided as a feature selection strategy for detecting minimum and maximum feature relevance bounds by inducing sparsity in the model [33]. The authors in [34] proposed the introduction of the l_p -norm for deriving the ordinal threshold with class centers with the aim to alleviate the influence of outliers (i.e., non-i.i.d. noises). Their approach provided an optimization algorithm and corresponding convergence analysis for computing the l_p -centroid. In [35], two neural network threshold ensemble models were proposed for ordinal regression problems. They generated a different formulation of the learned threshold by generating different

projections for the parameter updating. Another approach consists in imposing the ordinal constraints on the weights that connect the hidden layer with the output layer [36]. The formulation allows determining the optimum ones analytically according to the closed-form solution of the inequality constrained least-squares problem estimated from the Karush–Kuhn–Tucker conditions. In [37] is proposed a deep convolutional neural network model for ordinal regression by considering a family of probabilistic ordinal link functions in the output layer. These ordinal link functions fall within cumulative link models (CLMs). They split the ordinal space into the different classes of the problem by using a set of ordered thresholds. The thresholds are learned during the training process by minimizing a loss function that takes into account the distance between the categories, based on the weighted Kappa index.

Other ordinal approaches include ensemble decision tree and random forest models [19, 38] based on a weighted entropy function for selecting the predictors in the tree that reflect the magnitude of potential classification errors. A different approach based on conditional ordinal random field model was proposed for context-sensitive modeling of the facial action unit intensity by answering the context question in terms of temporal correlation between the ordinal outputs [21].

1.3 Limitation of state-of-the-art

Similar to the regression model, the main problem of standard ordinal classification approaches based on regression is the lack of a direct relationship between the prediction error of the regression model and the misclassification error. A different problem arises for the cost-sensitive penalty approach where there is the need to have a priori knowledge of the task in order to properly define the cost matrix. Accordingly, the ordinal binary decomposition approaches are highly influenced by how the overall problem is decomposed and how the results of all decompositions are aggregated into a single final classification. Some recent work in literature tried to overcome these problems by learning a single model for solving simultaneously individual binary classification tasks. However, these methodologies only model a static relationship among the ordinal classes that originate on how the problem is decomposed in binary subtasks. The threshold-based models proposed in literature often require multiple hyperparameters for setting the ordinal probability thresholds. Indeed, most of the state-of-the-art threshold-based approaches require highly demanding optimization procedures, which do not always guarantee optimal convergence and robustness against outliers.

The most related work to our proposal is the paper [37] that introduced the CLMs and quadratic weight kappa for

solving an ordinal problem. The main differences with our work lie in the (i) loss function we adopted, (ii) the different hyperparameters (i.e., slope) we learned in the learning process, (iii) a different unexplored task we aim to solve (AQC task) and (iv) the multiple objectives we aim to achieve, i.e., both an increase in generalization performance and also mitigation of unwanted bias related to the geometry. Indeed, we solved the ordinal problem by modeling the cumulative distribution of the AQC classes through the hyperparameters we learn in the CLM. Moreover, in our work, we exploited the standard cross-entropy loss for solving the ordinal AQC problem. As we shall see in results section, our deep ordinal model performs favorably over the CLMs for deep ordinal classification in [37].

1.4 Main contributions

To summarize, the main contributions of this paper are:

- the introduction of a deep learning methodology for ordinal classification specifically tailored for solving a topical and unexplored challenge on Industry 4.0, i.e., the aesthetic quality control classification. We introduced a novel dataset for the evaluation of wooden stocks. The task at the basis of the overall project originated from a specific company's demand;
- the introduction of a deep learning methodology for ordinal classification based on cumulative link model and categorical cross-entropy. We demonstrated how there is a sort of redundancy between the maximization of an ordinal loss and the modeling of cumulative distribution. The proposed approach overcomes this limitation by combining categorical cross-entropy with the cumulative link model and imposing the ordinal constraint via the thresholds and slope parameters. The introduction of the slope is effective in order to model the transient between adjacent cumulative link functions.
- the demonstration on how the proposed methodology is able from one side to reduce misclassification errors among distant classes (which is a relevant aspect for the real use case) and from the other side to reduce the bias factor related to the geometry. This fact has been demonstrated through an insightful explanation of the proposed DL behavior on the most discriminative shotgun parts. The ordinal constraints allow the network to learn the characteristics that properly describe the quality of shotgun (i.e., wood grains), rather than other confounds/bias characteristics (e.g., geometry).

The rest of this paper is organized as follows: Section 2 introduces the novel dataset for solving the quality control task, i.e., the evaluation of wooden stocks; in Sect. 3 the

proposed deep ordinal method is described; in Sect. 4 the evaluation procedure with respect to the state-of-the-art models is reported; in Sect. 5 the results are presented; in Sect. 6 the integration of the proposed approach in a decision support system is reported; and finally, in Sect. 7 the conclusions, limitations and future work of the proposed approach are discussed.

2 Materials

The QC phase is a fundamental step in the production of a rifle as the finished product must guarantee high performances both at mechanical and aesthetic levels. As regards the wooden stocks, these items must comply with the quality requirements related to aesthetics and surface manufacturing. In particular, the task consists of assigning a certain grade to each stock according to the wood grain: this implies a natural order between classes, because the more rich and fancy the veining pattern, the higher the quality class for the item. Each different type of rifle model manufactured by the company is equipped with a stock belonging to a specific grade class, and this coupling is at the total discretion of the company according to its market decisions.

The collected dataset is composed of both left- and right-side images belonging to different shotguns, for a total of 2120 RGB images with a size of 1000×500 pixels. The stocks have been classified into 4 main grades (1, 2, 3, 4) and their relative minor grades (2-/+, 3-/+, 4-/+), resulting in 10 different classes. Figure 1 shows an example of each of the 10 quality classes and the dataset distribution. Then, the original images were cropped to 470×270 , in order to focus better on the wooden region and remove the background. The images were acquired using a dedicated acquisition bench, composed of an industrial lamp and a high-definition RGB camera installed at the top of a photographic box, and annotation software.

3 Method

In this section, we present the proposed deep ordinal model, which consists of convolutional modules for extracting feature maps and an ordinal classification module (see Fig. 2). The main aspect of the ordinal module is the integration of the cumulative link model (CLM) in the output layer, parameterized by slope and thresholds, for encoding the ordinal nature of the label.

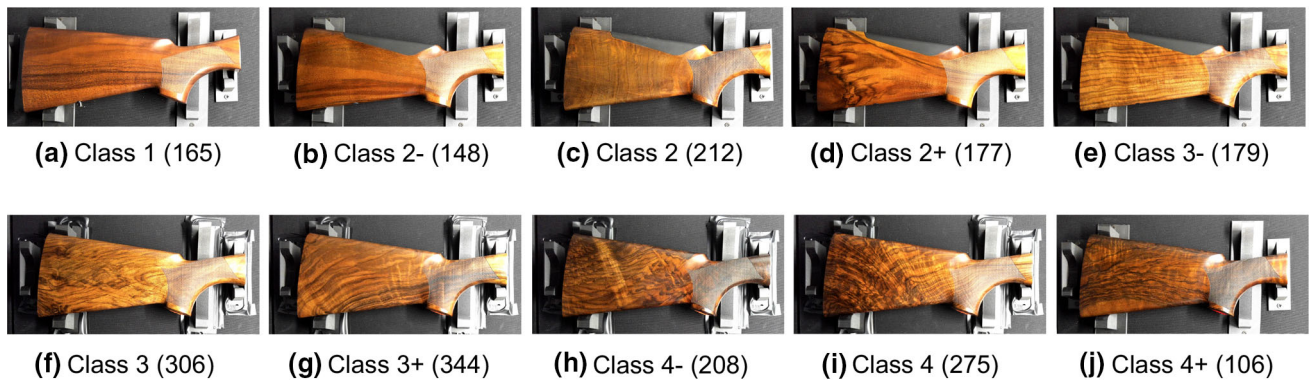


Fig. 1 Example of different stocks for each aesthetic quality class belonging to the collected dataset. The relative number of images/stocks is indicated in the parentheses

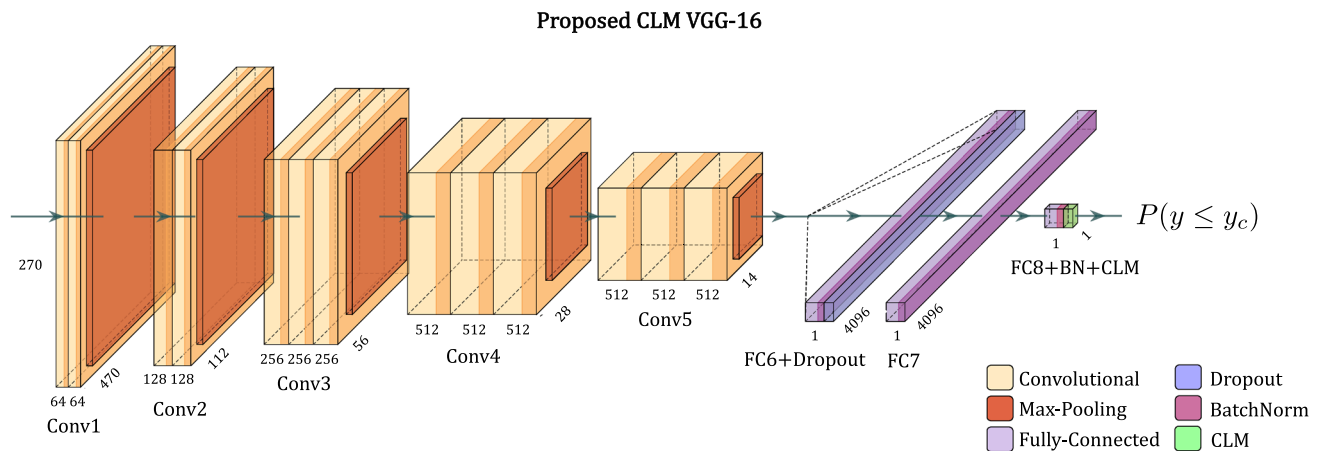


Fig. 2 The proposed CLM VGG-16 architecture, which consists of convolutional layers for features extraction and an ordinal head based on CLM

3.1 Feature extractor

We adopted as feature extractor the convolutional part of VGG-16 CNN [39] with 13 convolutional layers. Each of the 5 convolutional blocks has filters with a 3×3 pixels receptive field and is followed by a ReLU activation function. For CNN-parameter dimensionality reduction, max-pooling layers are used after 2 convolutional layers for the first 2 convolutional blocks and after 3 convolutional layers for the other blocks. The activation of the last convolutional block is used as the embedded features ($F \in \mathcal{R}^d$) learned from the feature extractor and is then fed to the ordinal classification head, which computes the output decision of the model. We let $\mathbf{x} \in \mathcal{X} \subseteq \mathcal{R}^d$ and $y \in \mathcal{Y} = \{y_1, y_2, \dots, y_C\}$ the input space and output space of C different ordinal classes, respectively.

3.2 Ordinal classification module

The output of the convolutional part of the CNN is fed to a sequence of 2 fully connected (FC) layers, followed by the

output layer. Dropout regularization layer was inserted between the first and the second FC layer with a rate of 0.3. The dropout rate was chosen in the validation stage (see Table 1). A batch normalization layer was added in order to stabilize the learning process and reduce the number of training epochs. The last FC layer has only one neuron as it provides the model projection in a 1-dimensional space: its value is used to classify the sample into the corresponding class according to the threshold model. Inspiring from [37],

Table 1 Network hyperparameters and cumulative link model parameters explored in the validation set

CLM parameters	Values
Slope	10, 50, 100
<i>Network hyperparameters</i>	
Dropout rate	0.1, 0.3, 0.5
Batch size	8, 16, 32
Learning rate	10^{-4} , 10^{-3} , 10^{-2}

the threshold-based approach we adopt in the output layer of the CNN is the CLM. In the CLM formulation [40], the class order is enforced by the following latent constraint:

$$f^{-1}(P(y \leq y_c | \mathbf{x})) = t_c - f(\mathbf{x}) \tag{1}$$

where $c = 1, \dots, C - 1$, $f^{-1} : [0, 1] \rightarrow \infty$ is a monotonic function (inverse link function) and t_c is the threshold defined for class y_c . Hence, the class y_c is predicted if and only if $f(\mathbf{x}) \in [t_{c-1}, t_c]$.

We integrated in the output layer of the architecture different forms of CLM exploring many link functions, all following the form $link[P(y \leq y_c | \mathbf{x})] = t_c - f(\mathbf{x})$. They are defined as follows:

- *logit* link function defined as follows:

$$P(y \leq y_c | \mathbf{x}) = \frac{1}{1 + e^{-s(t_c - f(\mathbf{x}))}} \tag{2}$$

- *probit* link function defined as follows:

$$P(y \leq y_c | \mathbf{x}) = \int_{-\infty}^{t_c - f(\mathbf{x})} \frac{s}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2 x^2} dx \tag{3}$$

- *clog-log* link function defined as follows:

$$P(y \leq y_c | \mathbf{x}) = 1 - e^{-e^{s(t_c - f(\mathbf{x}))}} \tag{4}$$

where s controls the slope of the CLM. Notice how the introduction of the slope represents one of the main contributions of the proposed work to control the transient between each monotonic link function with the purpose to be adapted according to the specific ordinal problem. It is worth noting that the function f is learned from the training data.

3.3 Setting the slope and thresholds parameters

The CLMs are highly influenced by the right choice of the thresholds and slope. The thresholds represent the cutting point between adjacent ordinal classes, while the slope controls the transient of $P(y \leq y_c | \mathbf{x})$. For instance, a small slope value may lead to a high transient in the CLM that does not enable the ordinal structure modeling (see Fig. 3).

For that reason, we have explored different formulations for the optimization of the thresholds ($\mathbf{t} = (t_1, t_2, \dots, t_{C-2}, t_{C-1})$) and the slope (s):

- (A): learning the slope s and the thresholds \mathbf{t} from data.
- (B): preliminary fixing the values of the slope s and the thresholds \mathbf{t} ;
- (C): preliminary fixing the values of the slope s and learning the thresholds \mathbf{t} from data;

It is assumed that $t_0 = -\infty$ and $t_C = +\infty$, defining C consecutive intervals that divide the real line defined by $f(x)$.

In formulation (A), both the slope s and the thresholds are learned during the training process. In particular, the threshold is learned from the following equation:

$$t_q = t_1 + \sum_{c=2}^{C-1} \gamma_c^2, \tag{5}$$

where t_1 is learned to obtain the first threshold, γ is learned to obtain the other thresholds and C is the number of classes. This formulation for the thresholds ensures that the constraints $t_1 \leq t_2 \leq \dots \leq t_{C-1}$ are fulfilled, which is needed for obtaining increasing $P(y \leq y_c | \mathbf{x})$ with c .

In formulation (B), rather than learning the parameter s in the training stage, we have tuned this parameter in the validation stage. Moreover, the imbalanced setting of the ordinal classes is taken into account by fixing the thresholds instead of learning them during the training stage. In particular, we set the thresholds according to the prior probability of each class as follows:

$$t_1 = \frac{\sum_{i=1}^N 1_{y=y_1}}{N}, \tag{6}$$

$$\gamma_c = \sqrt{P(y = y_c)} = \sqrt{\frac{\sum_{i=1}^N 1_{y=y_c}}{N}}, \tag{7}$$

where t_1 is the value of the first threshold related to the prior probability of the first class, γ_c is the vector of the prior probabilities $P(y = y_c)$ associated to each class $c = \{2, \dots, C - 1\}$ and N is the total number of training points.

In the hybrid formulation (C), only the thresholds are learnable parameters, while the slope is tuned in the validation stage.

3.4 Loss function

The loss function was defined in terms of categorical cross-entropy (CCE) as follows:

$$L(\hat{y}, y) = - \sum_{i=1}^C y_i \log(\hat{y}_i), \tag{8}$$

where C is the number of classes indicating the output size, \hat{y}_i is the i th scalar value in the model output and y_i is the corresponding target value.

4 Experimental procedure

4.1 Experimental comparisons

It is worth noting here that the goal of our paper is to predict the aesthetic quality classes of the rifle models. Despite the difference in the task definition, we have decided to perform the experimental comparisons with

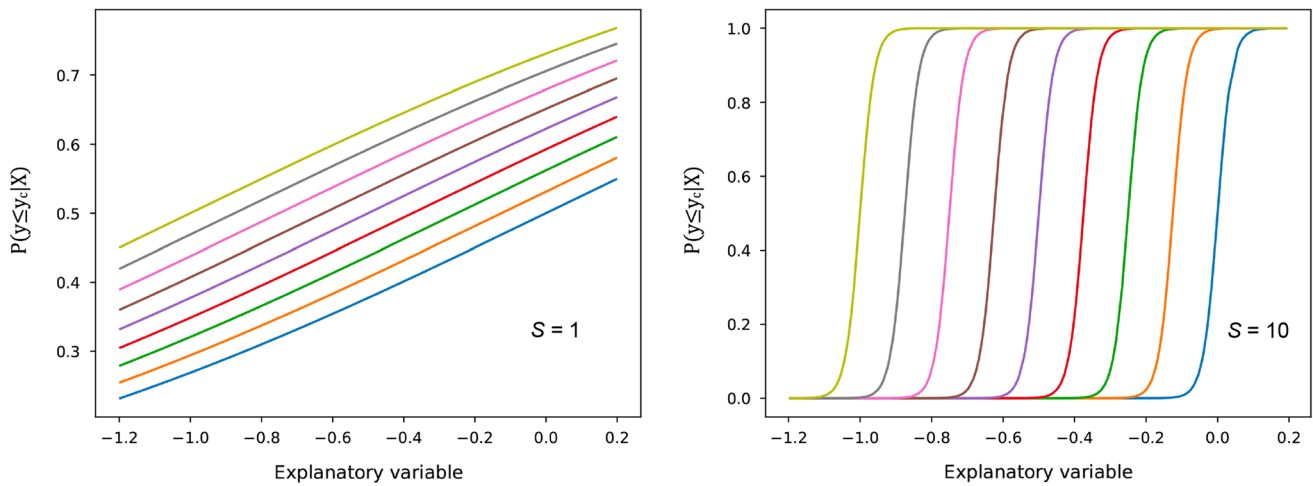


Fig. 3 The effect of the slope parameter s regularization in logit link function for defining the $C - 1$ thresholds

respect to baseline nominal VGG-16 [39] and other state-of-the-art ordinal DL methodologies, including ordinal binary decomposition VGG-16 [18] and CLM VGG-16 with weight kappa loss [37]. Figure 4 shows the architectures of the state-of-the-art methodologies employed for comparisons.

4.1.1 Nominal VGG-16

In the nominal classification, the VGG-16 model presents the classic architecture, where the convolutional part is followed by 3 FC layers and the last one has dimension C as the number of class labels. The output of this last FC layer is fed to a softmax activation function which maps the output of the CNN model into a set of probabilities belonging to each class. The loss function is the CCE loss, as defined in 3.4.

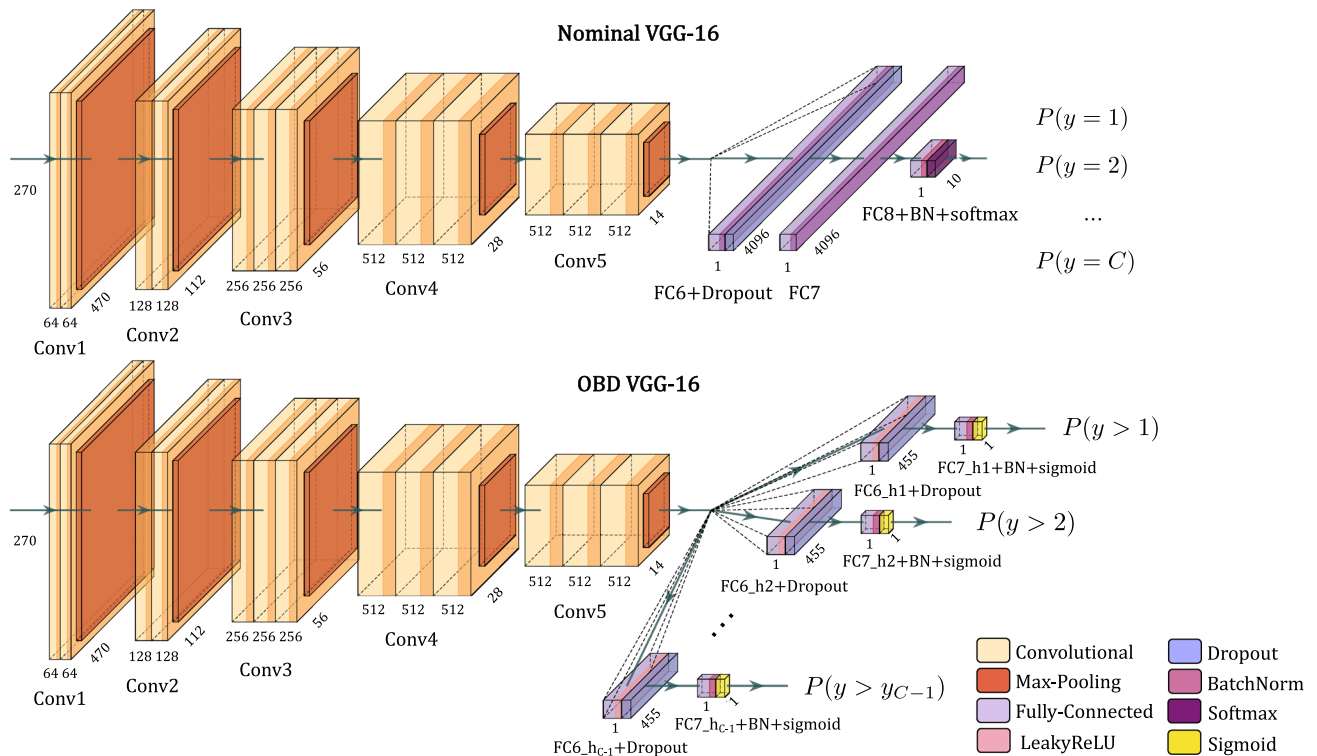


Fig. 4 The other state-of-the-art architectures: baseline nominal VGG-16 and ordinal binary decomposition VGG-16

4.1.2 Ordinal binary decomposition VGG-16

The ordinal binary decomposition (OBD) is an ordinal approach that consists of decomposing the ordinal problem into a set of $C - 1$ binary problems, where each problem c must determine if $y > y_c$ conditioned to $1 \leq c < C$. Following the implementation in [18], the convolutional part of the VGG-16 is the input of multiple FC blocks, all of the same dimension. Each block consists of a FC layer, followed by a Leaky ReLU activation function and dropout layer. A final output layer computes the final classification given by the model solving an individual binary classification subproblem. The output of each of the $C - 1$ FC blocks has a sigmoid activation function representing the probability $o_k = P(\mathbf{y} > y_k | \mathbf{x}) \in (0, 1)$. The adopted loss functions include the mean squared error (MSE) defined as follows:

$$L(\hat{y}, y) = \frac{1}{C - 1} \sum_{k=1}^{C-1} (y_k - \hat{y}_k)^2, \tag{9}$$

and the mean absolute error (MAE):

$$L(\hat{y}, y) = \frac{1}{C - 1} \sum_{k=1}^{C-1} ||y_k - \hat{y}_k||, \tag{10}$$

where C is the number of classes indicating the output size, \hat{y}_k is the predicted probability of the model output to be greater than y_k and y_k is the corresponding target value that is equal to 1 when $y_i = C_q$ and 0 otherwise.

4.1.3 Cumulative link model VGG-16 with weighted kappa loss

Differently from our approach, in the work made by [37] the CLM structure in the output layer is combined with the continuous version of the quadratic weighted kappa (QWK) loss function [41]. We employed the QWK according to [41] as follows:

$$QWK = 1 - \frac{\sum_{i,j} \omega_{i,j} O_{i,j}}{\sum_{i,j} \omega_{i,j} E_{i,j}}, \tag{11}$$

where N is the number of training data, N_i is the number of samples for each i -th class, ω is the penalization matrix, O is the confusion matrix, $E_{ij} = \frac{O_{i \cdot} O_{\cdot j}}{N}$, $O_{i \cdot}$ is the sum of the i -th row and $O_{\cdot j}$ is the sum of the j -th column. In our experimental comparisons, linear weights ($\omega_{i,j} = \frac{(i-j)}{(C-1)}$, $\omega_{i,j} \in [0, 1]$) and quadratic weights ($\omega_{i,j} = \frac{(i-j)^2}{(C-1)^2}$, $\omega_{i,j} \in [0, 1]$) are considered.

4.2 Experimental design

A transfer learning approach was used to fine-tune the networks on ImageNet pre-trained weights, in order to reduce computational time while improving the generalization performance [42]. For this reason, all the convolutional layers were frozen. As a preprocessing step, the mean value was removed for each image. The dataset was split by a stratified hold-out procedure, i.e., using 60% of images as training, 20% as validation and 20% as a test. Images belonging to the same shotgun (front and back) were maintained in the same set, ensuring that the model may be able to generalize across different unseen shotgun stocks. In order to cope with the small dimension of the dataset and the slight unbalance of the classes, a balanced data augmentation strategy was performed on the fly on all the training set samples, applying a horizontal flip to original images. In this process, we ensure that, during training, the number of samples per class follows a uniform distribution, performing an oversampling of the minority classes. We adopted Adam as optimizer and we explored the best batch size, the initial learning rate and dropout rate as network hyperparameters (see Table 1). These network hyperparameters together with the slope parameter for formulations (B) and (C) were tuned in a separate validation set using a grid-search approach. The number of training epochs was set to 50 while adopting the early stopping strategy with the patience of 10 epochs monitoring validation loss. All the experiments were performed using TensorFlow 2.0 and Keras 2.3.1 frameworks on Intel Core i7-4790 CPU 3.60 GHz with 16 GB of RAM and NVIDIA GeForce GTX 970. All the code used in the experiments and the employed dataset is available in a public repository.¹

4.3 Evaluation metrics and error criteria

Both nominal and ordinal metrics were considered to provide quantitative performance results of the proposed model. Taking into account the ordinal nature of this problem, the ordinal metrics are potentially more relevant for evaluating our ordinal classification task.

4.3.1 Nominal metrics

The correct classification rate (CCR) or accuracy is the most standard metric for evaluating classification models, indicating the percentage of correctly classified samples. In our application context, CCR presents two main problems:

¹ https://github.com/rosati1392/AQC_Ordinal.git.

- all the prediction mistakes are equally penalized, without considering how much is the deviation from the ground truth (according to the ordinal scale);
- when in presence of class imbalance, it can become an unreliable measure of model performance, as it can be trivially increased by assigning all patterns to the majority class.

The CCR is defined as follows:

$$CCR = \frac{1}{N} \sum_{i=1}^N 1\{\hat{y}_i = y_i\}, \tag{12}$$

where N denotes the number of test samples, y_i is the class label for sample x_i and \hat{y}_i is the predicted label for sample x_i .

Other accuracy-based metrics are Top-2 CCR and Top-3 CCR, which are the accuracy where true class matches with any one of the two or three, respectively, most probable classes predicted by the model.

Another nominal metric considered is minimum sensitivity (MS), which expresses the lowest percentage of samples correctly predicted to belong to a certain class:

$$MS = \min \left\{ S_c = \frac{O_{cc}}{O_{c\bullet}}, c = 1, \dots, C \right\}, \tag{13}$$

where O is the confusion matrix, C is the number of classes and S_c is the sensitivity computed for the class c .

4.3.2 Ordinal metrics

The quadratic weighted kappa (QWK) is a relevant metric for ordinal problems as it gives a higher weight to the errors that are further from the correct class. We have reported the continuous formulation of QWK (for the results, the values reported are generally those from discrete QWK, while the continuous version is used only for

the state-of-the-art experimental comparison in the training process [37]) according to Eq. (11).

Other ordinal metrics are 1-off accuracy, which indicates that the predicted label is off at most by 1 adjacent class from the ground truth one, and MAE, which is the average absolute deviation of the prediction from the ground truth, defined as:

$$MAE = \frac{1}{N} \sum_{i,j=1}^C ||i - j|| O_{ij}, \tag{14}$$

where N is the number of test samples, C is the number of classes, and O is the confusion matrix.

5 Results

In Sect. 5.1, we report the predictive performance of the proposed method for each formulation. Afterward, in Sect. 5.2 we describe the experimental comparisons with baseline nominal and state-of-the-art DL ordinal methodologies. Finally, in Sect. 5.3 we show results on how the proposed approach is relevant in order to mitigate the bias.

5.1 Predictive performance for each formulation

The predictive performance of the proposed approach was provided by tuning the network hyperparameters. It is worth noting that for formulation A the CLM parameters were learned in the training set, while for formulations B and C the slope was tuned in a separate validation set and kept fixed during the training stage. Table 2 shows the predictive performance of the proposed approach (in terms of QWK and MS) for each formulation and for each final activation. The adoption of these two metrics is related to the aim of our classification task: we want to maximize the model performance in the ordinal problem while being

Table 2 Predictive performance on the test set of the proposed approach for each formulation and for each final CLM activation

Formulations	Final activation	Slope	Thresholds	BS	LR	QWK	MS
Experiment A	Logit	Trainable	Trainable	16	10 ⁻²	0.926	0.128
	Probit	Trainable	Trainable	16	10 ⁻²	0.928	0.205
	Clog-log	Trainable	Trainable	16	10 ⁻²	0.892	0.179
Experiment B	Logit	Fixed	Fixed	16	10 ⁻²	0.846	0
	Probit	Fixed	Fixed	16	10 ⁻²	–	–
	Clog-log	Fixed	Fixed	16	10 ⁻²	–	–
Experiment C	Logit	Fixed	Trainable	16	10 ⁻²	0.937	0.231
	Probit	Fixed	Trainable	16	10 ⁻²	0.922	0.282
	Clog-log	Fixed	Trainable	16	10 ⁻²	0.927	0.282

In bold, we reported the experiment that leads to the best results both in terms of quadratic weight kappa (QWK) and minimum sensitivity (MS)

consistent in prediction among all the dataset classes despite the imbalanced setting.

With respect to these formulations, experiment C achieved the best results both in terms of QWK (with logit as CLM link function) and MS (with probit and clog-log). Notice how in this formulation we fixed and tuned the optimal slope value in the validation set. This procedure provides better generalization performance than fully learn slope and thresholds in the training procedure. Another relevant aspect is that fixing the thresholds to a preset value does not allow the model to converge for probit and clog-log activations. This highlights that the slope parameter has no effect if the flexibility provided by the threshold model structure, where the threshold of each class is independently adjusted during training, is not guaranteed.

5.2 Experimental comparisons with state-of-the-art DL ordinal models

Figure 5 shows the test confusion matrices of the proposed approach and the baseline approach (nominal approach). The confusion matrix of the proposed method is more focused on the diagonal, thus penalizing the error among distant AQC classes.

Table 3 shows the experimental results of our approach with respect to the baseline approach (nominal approach) and other state-of-the-art deep ordinal methods (OBD VGG-16 [18] and CLM VGG-16 [37]). Experiment C was chosen as the best formulation of our proposed approach (see Sect. 5.1). Our proposed deep ordinal model outperforms the nominal approach in terms of QWK, MS and 1-OFF of about 8%, 68.9%, and 17.4%, respectively. It is worth noting that QWK, MS and 1-OFF represent the most important metrics in order to reduce misclassification errors among distant classes. This requisite fully

corresponds to the original company’s demand, i.e., the reduction of errors among distant AQC classes. Moreover, the proposed approach overcomes in terms of QWK, MS and 1-OFF the OBD [18] of about 1,8%, 282% and 4,3%, respectively, and CLM [37] of about 0.03%, 37.6% and 1%, respectively.

The comparison with the other ordinal methodologies and the highest values of QWK, MS and 1-OFF highlighted how the proposed method is more effective to model the ordinal structure of the AQC classes by penalizing the distance of incorrect prediction from the ground truth class.

5.3 Model interpretability and Bias mitigation

From the point of view of a domain expert such as a human operator who is responsible for AQC task, explanation and interpretability are key points that may increase the usefulness and the trustworthiness of the overall decision support system (DSS). An explanation, specifically tailored to the end-users (i.e., human operator) on how the DL model achieved the prediction, is relevant in order (i) to uncover valuable information that otherwise would have remained hidden within the complexity of the model and (ii) to empower users with powerful new insights. Starting from this concept, our objective was to encourage the prediction of the proposed model to be as aligned as possible with the human annotation. By designing an ordinal DL methodology, our claim was to penalize large errors (misclassification errors among distant classes) that do not usually happen in human annotation. After being demonstrated this outcome, we would go further by describing that from one side the model is potentially able to provide new insights on finer-grained wood patterns that can be sometimes unseen by a human operator, and from the other side the model is completely aligned on what the human

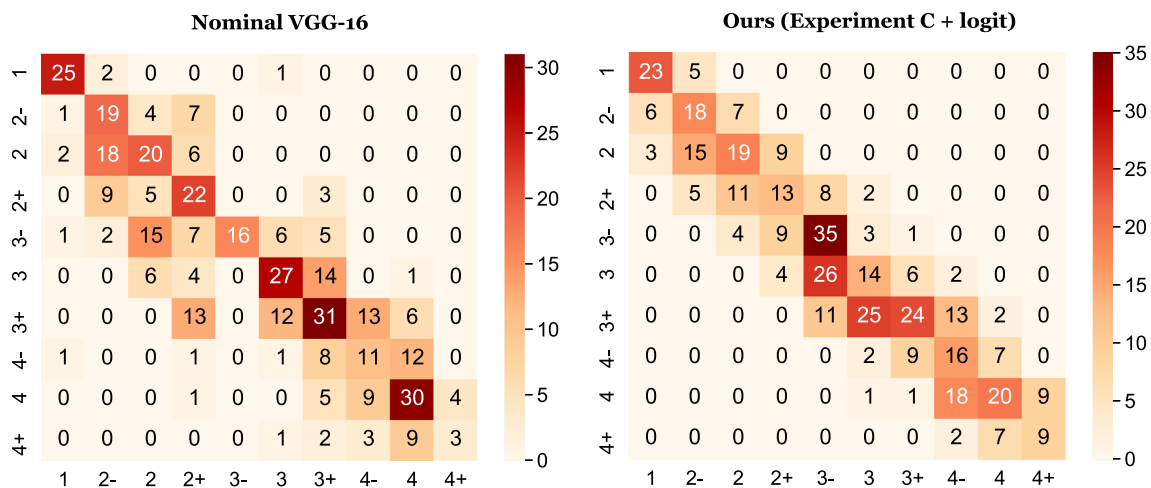


Fig. 5 Confusion matrices for nominal and proposed approach

Table 3 Experimental results comparison on the test set in terms of both ordinal and nominal metrics of the proposed approach with respect to baseline nominal VGG-16 model (NOM), ordinal binary

decomposition (OBD) implementation for CNN [18] and state-of-the-art cumulative link models (CLM) for deep ordinal classification [37]

Method	Final activation	Loss	BS	LR	QWK	MS	MAE	CCR	TOP_2	TOP_3	1-OFF
NOM	Softmax	CCE	16	10 ⁻²	0.867	0.167	0.124	0.481	0.731	0.863	0.788
OBD	Sigmoid	MSE	16	10 ⁻²	0.920	0	0.158	0.528	0.783	0.906	0.880
OBD	Sigmoid	MAE	16	10 ⁻²	0.923	0	0.142	0.554 ↑	0.802 ↑	0.896	0.887
CLM	Logit(th)	QWK	16	10 ⁻²	0.924	0	0.128	0.424	0.709	0.851	0.875
CLM	Probit(th)	QWK	16	10 ⁻²	0.929	0	0.124	0.429	0.726	0.868	0.908
CLM	Clog-log(th)	QWK	16	10 ⁻²	0.911	0	0.132	0.394	0.670	0.863	0.835
CLM	Logit(th)	LWK	16	10 ⁻²	0.918	0	0.122	0.443	0.715	0.844	0.877
CLM	Probit(th)	LWK	16	10 ⁻²	0.917	0	0.127	0.392	0.698	0.851	0.858
CLM	Clog-log(th)	LWK	16	10 ⁻²	0.909	0	0.135	0.382	0.687	0.849	0.844
CLM	Logit(th,slope)	QWK	16	10 ⁻²	0.925	0.205	0.113	0.460	0.776	0.911	0.894
CLM	Probit(th,slope)	QWK	16	10 ⁻²	0.934	0.180	0.112 ↓	0.467	0.776	0.910	0.915
CLM	Clog-log(th,slope)	QWK	16	10 ⁻²	0.928	0.051	0.112 ↓	0.462	0.743	0.858	0.899
Ours	Logit(th)	CCE	16	10 ⁻²	0.937 ↑	0.231	0.137	0.455	0.774	0.927 ↑	0.925 ↑
Ours	Probit(th)	CCE	16	10 ⁻²	0.922	0.282 ↑	0.140	0.434	0.729	0.901	0.881
Ours	Clog-log(th)	CCE	16	10 ⁻²	0.927	0.282 ↑	0.137	0.451	0.776	0.927 ↑	0.889

The best value for each metric is highlighted in bold. For the final activation, the learnable parameters are specified in parentheses, where “th” stands for thresholds

operator is checking, thus focusing on the aesthetic quality classification of rifles based on the analysis of wood grains and avoiding any unwanted bias related to the geometry and shotgun series [17].

This fact is confirmed by exploring the saliency map of the proposed ordinal DL approach according to the approach proposed by [43]. Accordingly, the extracted saliency maps are constrained to focus on wood grains rather than the geometrical edges (see Fig. 6). Thus, this strategy allows alleviating the bias by separating the two

tasks and providing the prediction of quality classes for each shotgun macro-series model.

The overall methodology allows the network to learn the characteristics that properly describe the quality of shotgun (i.e., wood grains), rather than other confounds/bias characteristics (e.g., geometry).

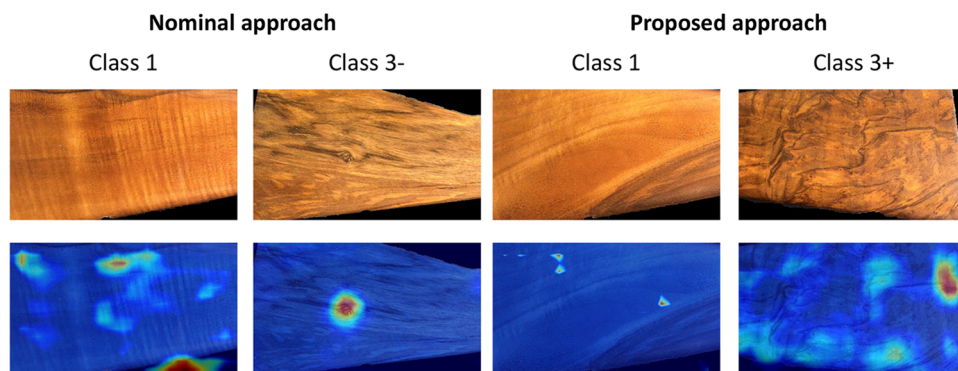


Fig. 6 Saliency maps obtained from test images correctly predicted by the nominal and proposed approach. In class 1, it can be seen how the nominal approach is more focused outside the stock, whereas for the proposed approach the map does not show any hot point because

veins are not relevant in this class. For the higher classes, notice how the proposed model better focuses on the wood features, following the attention on the pattern of grains

6 Aesthetic quality control decision support system

Taking into account the high variability among different wooden stocks, the aesthetic quality classification of rifles stocks based on the analysis of wood grains represents a challenging and relevant step during the overall production chain. The function of AQC is to build a method that can make objective the result of the visual inspections carried out, which are still purely dependent on the evaluation of human operators (inter-operator and intra-operator variability). Thus, the aim is to create a DSS for the automation of this aesthetic assessment of wooden stock images using DL techniques, with the purpose of making this control more reliable, fast and standardized. The integration of the proposed ordinal DL methodology as the main core of a DSS for solving AQC tasks is described in Fig. 7.

A container logic was adopted for packaging the DL model and all its dependencies so the inference phase runs reliably from one computing environment to another. A docker image is essentially a snapshot of a container. Microsoft Azure framework was adopted for providing a cloud-based environment using virtualized containers. This environment can ensure hardware and software isolation, flexibility and inter-dependencies between data collection, model building and prediction phases. Indeed, the proposed approach is integrated into a AQC serverless platform where the predicted quality class is obtained by ingestion event. The technician may trigger a cloud function that could invoke the DL model to provide the inference. This setting may ensure the high scalability of the system while allowing the continuous fine-tuning of the DL model based on the new images of rifles available. All the prediction results were stored in the Azure blob storage and displayed to the human operator in a GUI interface.

The DSS platform is comprised of the acquisition box, rifle placement and GUI interface (see Fig. 8). The DSS platform was also used to collect the employed image dataset described in Sect. 2. The integration of the proposed DL model in the DSS platform allows us to reduce up to 90% the time needed for the qualitative analysis

carried out manually in this specific field (inference computation time of our proposed methodology is 4 s for each image on Intel Core i7-4790 CPU 3.60 GHz with 16 GB of RAM).

The great flexibility and invariance to environmental conditions of these techniques will also allow a high level of replicability of the project, even between companies with production lines characterized by different processes, minimizing the impact on the phases before and after the AQC. We are currently testing the generalization power of the proposed DL approach in a different company's production chain for supporting the AQC process of the technician in a different environment and operating condition.

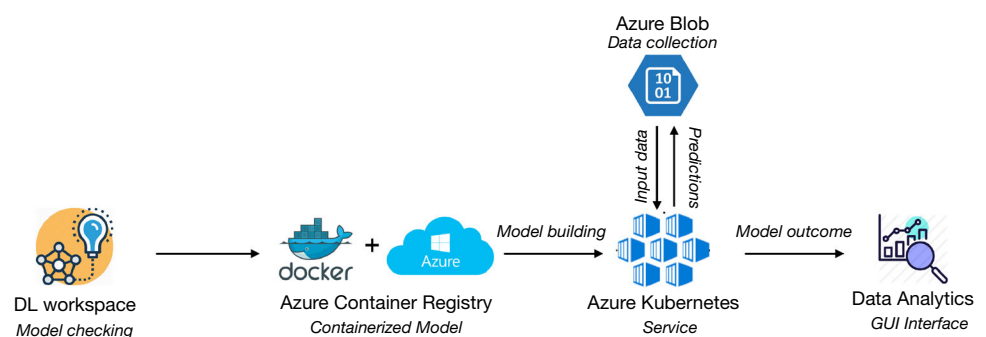
7 Discussion and conclusions

Our work aims to propose an ordinal deep learning (DL) approach, specifically tailored for providing the aesthetic quality classification of shotguns based on the analysis of wood grains. Being trained on examples annotated by experts rather than composed of strict descriptive rules, this model is able, with the necessary training, to generalize across different unseen shotgun stocks. The proposed DL methodology is integrated as the main core of a decision support system (DSS) for solving a challenging aesthetic quality control (AQC) task in Industry 4.0 scenario.

However, the main advantage of the proposed approach is not limited to the automatization of the overall AQC procedure and the minimization of the annotator's variability. The proposed DL-based DSS driven by ordinal constraints was properly conceived to model the natural ordinal structure of the AQC task while penalizing the misclassification errors that are far from the correct AQC classes. The introduction of the slope parameter allows to model the transient between CLM functions for each learnable ordinal threshold.

The higher performance obtained by the proposed method for quality class prediction with respect to a baseline nominal DL approach and state-of-the-art ordinal

Fig. 7 DSS cloud interface



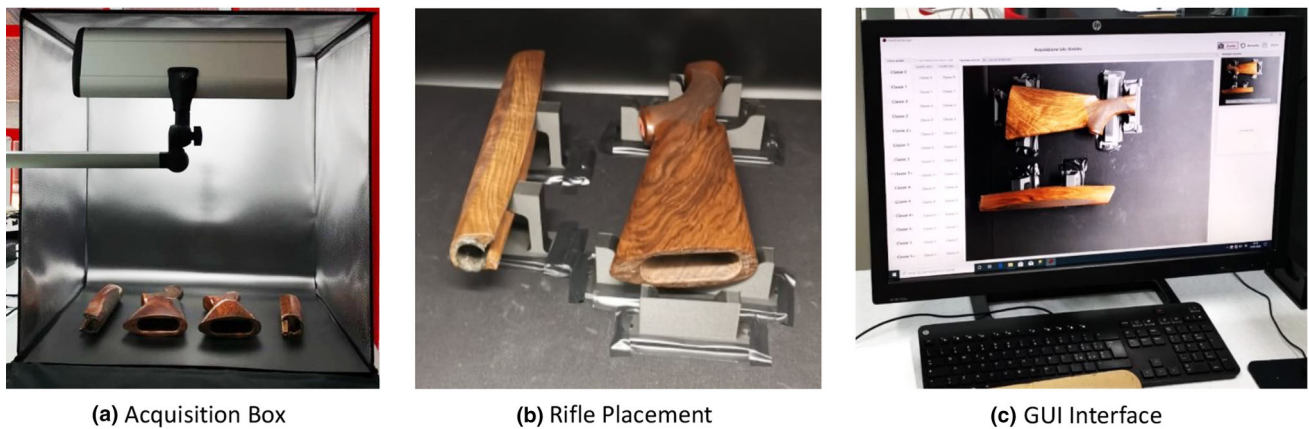


Fig. 8 Decision support system (DSS) platform. The DSS platform is comprised of the acquisition box, rifle placement and GUI interface. The DSS platform was also used to collect the employed image dataset

DL approaches suggests how the proposed approach represents a valuable solution for automatizing the overall AQC procedure. In fact, the experimental findings demonstrated how a standard CCE together with CLM is sufficient to model ordinal structure of the label, without requiring the minimization of an ordinal loss (e.g., QWK). This is also in line with recent findings in the ordinal classification literature [44]. Moreover, the ordinal constraints allow the network to learn the characteristics that properly describe the quality of shotgun (i.e., wood grains), rather than other confounds/bias characteristics (e.g., geometry). The experimental results demonstrated how the impact of the proposed approach both in terms of predictive performance and interpretability is also compliant with the ethic guidelines by the European Commission (Human agency and oversight, [12]) in order to provide a DSS based on DL that is as aligned as possible with the human operator for supporting the AQC task.

As a result, the potential impact of the proposed approach could be measured according to (i) introducing an ordinal DL technique in a challenging and unexplored Industrial 4.0 scenario, namely AQC task; (ii) ensuring the robustness of the approach with respect to possible bias factors, i.e., the network learns the characteristics that really describe the quality of shotgun; (iii) integrating the DL in a DSS that can support the human operator by reducing up to 90% the time needed for the qualitative analysis carried out manually in this specific domain.

Future work may be addressed to combine domain adaptation techniques with the proposed approach in order to generalize across different conditions while providing the integration in the proposed DSS platform for easier serialization. Another interesting future direction includes the possibility to model inter-operator variability by providing multiple annotations by different operators for the same image. This direction includes the possibility to

design a multi-task deep ordinal approach to simultaneously monitor correlation and variability among raters.

Funding Open access funding provided by CRUI-CARE. This work was supported within the research agreement between Università Politecnica delle Marche and Benelli Armi Spa for the “4USER Project” (User and Product Development: from Virtual Experience to Model Regeneration) funded on the POR MARCHE FESR 2014-2020-ASSE 1-OS 1-ACTION 1.1-INT. 1.1.1. Promotion of industrial research and experimental development in the areas of smart specialization-LINEA 2-Bando 2019, approved with DDPF 293 of 22/11/2019. This work has been also partially subsidised by “Agencia Española de Investigación (España)” (grant reference: PID2020-115454GB-C22 / AEI / 10.13039 / 501100011033), by “Consejería de Salud y Familia (Junta de Andalucía)” (Grant reference: PS-2020-780) and by “Consejería de Transformación Económica, Industria, Conocimiento y Universidades (Junta de Andalucía)” y Programa Operativo FEDER 2014-2020” (Grant references: UCO-1261651 and PY20_00074). Víctor Manuel Vargas’s research has been subsidised by the FPU Predoctoral Program of the Spanish Ministry of Science, Innovation and Universities (MCIU), Grant reference FPU18/00358.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


References

- Sakib N, Wuest T (2018) Challenges and opportunities of condition-based predictive maintenance: a review. *Proc CIRP* 78:267–272. <https://doi.org/10.1016/j.procir.2018.08.318>
- Cadavid JPU, Lamouri S, Grabot B, Pellerin R, Fortin A (2020) Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. *J Intell Manuf* <https://doi.org/10.1007/s10845-019-01531-7>
- Aydın S (2020) Deep learning classification of neuro-emotional phase domain complexity levels induced by affective video film clips. *IEEE J Biomed Health Inform* 24(6):1695–1702. <https://doi.org/10.1109/JBHI.2019.2959843>
- Aydın S, Güdücü Ç, Kutluk F, Öniz A, Özgören M (2019) The impact of musical experience on neural sound encoding performance. *Neurosci Lett* 694:124–128. <https://doi.org/10.1016/j.neulet.2018.11.034>
- Calabrese M, Cimmino M, Fiume F, Manfrin M, Romeo L, Ceccacci S, Paolanti M, Toscano G, Ciandrini G, Carrotta A, et al (2020) SOPHIA: an event-based IoT and machine learning architecture for predictive maintenance in Industry 4.0. *Information* 11(4):202. <https://doi.org/10.3390/info11040202>
- Romeo L, Loncarski J, Paolanti M, Bocchini G, Mancini A, Frontoni E (2020) Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. *Expert Syst Appl* 140:112869. <https://doi.org/10.1016/j.eswa.2019.112869>
- Escobar CA, Morales-Menendez R (2018) Machine learning techniques for quality control in high conformance manufacturing environment. *Adv Mech Eng* 10(2):1–16. <https://doi.org/10.1177/1687814018755519>
- Wang T, Chen Y, Qiao M, Snoussi H (2018) A fast and robust convolutional neural network-based defect detection model in product quality control. *Int J Adv Manuf Technol* 94(9–12):3465–3471. <https://doi.org/10.1007/s00170-017-0882-0>
- Dutta G, Kumar R, Sindhwani R, Singh RK (2021) Digitalization priorities of quality control processes for SMEs: a conceptual study in perspective of Industry 4.0 adoption. *J Intell Manuf*. <https://doi.org/10.1007/s10845-021-01783-2>
- Escobar CA, McGovern ME, Morales-Menendez R (2021) Quality 4.0: a review of big data challenges in manufacturing. *J Intell Manuf*. <https://doi.org/10.1007/s10845-021-01765-4>
- Paraschos PD, Koulinas GK, Koulouriotis DE (2020) Reinforcement learning for combined production-maintenance and quality control of a manufacturing system with deterioration failures. *J Manuf Syst* 56:470–483. <https://doi.org/10.1016/j.jmsy.2020.07.004>
- EU (2019) Ethics guidelines for trustworthy AI. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>. Accessed 31 March 2021
- Pazzaglia G, Martini M, Rosati R, Romeo L, Frontoni E (2021) A deep learning-based approach for automatic leather classification in industry 4.0. In: *Pattern recognition. ICPR international workshops and challenges: virtual event, January 10–15, 2021, Proceedings, Part IV*. Springer, pp 662–674. https://doi.org/10.1007/978-3-030-68799-1_48
- Villalba-Diez J, Schmidt D, Gevers R, Ordieres-Meré J, Buchwitz M, Wellbrock W (2019) Deep learning for industrial computer vision quality control in the printing industry 4.0. *Sensors* 19(18):3987. <https://doi.org/10.3390/s19183987>
- Iqbal R, Maniak T, Doctor F, Karyotis C (2019) Fault detection and isolation in industrial processes using deep learning approaches. *IEEE Trans Industr Inf* 15(5):3077–3084. <https://doi.org/10.1109/TII.2019.2902274>
- Peres RS, Barata J, Leitao P, Garcia G (2019) Multistage quality control using machine learning in the automotive industry. *IEEE Access* 7:79908–79916. <https://doi.org/10.1109/ACCESS.2019.2923405>
- Rosati R, Romeo L, Cecchini G, Tonetto F, Perugini L, Ruggeri L, Viti P, Frontoni E (2021) Bias from the wild industry 4.0: are we really classifying the quality or shotgun series? In: *Pattern recognition. ICPR international workshops and challenges*, pp 637–649. Springer, Cham. https://doi.org/10.1007/978-3-030-68799-1_46
- Barbero-Gómez J, Gutiérrez P-A, Vargas V-M, Vallejo-Casas J-A, Hervás-Martínez C (2021) An ordinal CNN approach for the assessment of neurological damage in Parkinson's disease patients. *Expert Syst Appl* 182:115271. <https://doi.org/10.1016/j.eswa.2021.115271>
- Singer G, Ratnovsky A, Naftali S (2021) Classification of severity of trachea stenosis from EEG signals using ordinal decision-tree based algorithms and ensemble-based ordinal and non-ordinal algorithms. *Expert Syst Appl* 173:114707. <https://doi.org/10.1016/j.eswa.2021.114707>
- Liu X, Fan F, Kong L, Diao X, Xie W, Lu J, You J (2020) Unimodal regularized neuron stick-breaking for ordinal classification. *Neurocomputing* 388:34–44. <https://doi.org/10.1016/j.neucom.2020.01.025>
- Rudovic O, Pavlovic V, Pantic M (2015) Context-sensitive dynamic ordinal regression for intensity estimation of facial action units. *IEEE Trans Pattern Anal Mach Intell* 37(5):944–958. <https://doi.org/10.1109/TPAMI.2014.2356192>
- Chang K-Y, Chen C-S (2015) A learning framework for age rank estimation based on face images with scattering transform. *IEEE Trans Image Process* 24(3):785–798. <https://doi.org/10.1109/TIP.2014.2387379>
- He R, Tan T, Davis L, Sun Z (2018) Learning structured ordinal measures for video based face recognition. *Pattern Recogn* 75:4–14. <https://doi.org/10.1016/j.patcog.2017.02.00>
- Hirk R, Hornik K, Vana L (2019) Multivariate ordinal regression models: an analysis of corporate credit ratings. *Stat Methods Appl* 28(3):507–539. <https://doi.org/10.1007/s10260-018-00437-7>
- Balugani E, Lolli F, Pini M, Ferrari AM, Neri P, Gamberini R, Rimini B (2021) Dimensionality reduced robust ordinal regression applied to life cycle assessment. *Expert Syst Appl* 178:115021. <https://doi.org/10.1016/j.eswa.2021.115021>
- Gutiérrez PA, Pérez-Ortiz M, Sánchez-Monedero J, Fernández-Navarro C, Hervás-Martínez C (2016) Ordinal regression methods: survey and experimental study. *IEEE Trans Knowl Data Eng* 28(1):127–146. <https://doi.org/10.1109/TKDE.2015.2457911>
- Uematsu K, Lee Y (2015) Statistical optimality in multipartite ranking and ordinal regression. *IEEE Trans Pattern Anal Mach Intell* 37(5):1080–1094. <https://doi.org/10.1109/TPAMI.2014.2360397>
- Raykar VC, Duraiswami R, Krishnapuram B (2008) A fast algorithm for learning a ranking function from large-scale data sets. *IEEE Trans Pattern Anal Mach Intell* 30(7):1158–1170. <https://doi.org/10.1109/TPAMI.2007.70776>
- Xu D, Zhu F, Liu Q, Zhao P (2021) Arail: Learning to rank from incomplete demonstrations. *Inf Sci* 565:422–437. <https://doi.org/10.1016/j.ins.2021.02.001>
- Zhu H, Shan H, Zhang Y, Che L, Xu X, Zhang J, Shi J, Wang F-Y (2021) Convolutional ordinal regression forest for image ordinal estimation. *IEEE Trans Neural Netw Learn Syst*. <https://doi.org/10.1109/TNNLS.2021.3055816>
- Hamsici OC, Martinez AM (2016) Multiple ordinal regression by maximizing the sum of margins. *IEEE Trans Neural Netw Learn Syst* 27(10):2072–2083. <https://doi.org/10.1109/TNNLS.2015.2477321>

32. Ma Z, Chen S (2019) A convex formulation for multiple ordinal output classification. *Pattern Recogn* 86:73–84. <https://doi.org/10.1016/j.patcog.2018.09.005>
33. Pfannschmidt L, Jakob J, Hinder F, Biehl M, Tino P, Hammer B (2020) Feature relevance determination for ordinal regression in the context of feature redundancies and privileged information. *Neurocomputing* 416:266–279. <https://doi.org/10.1016/j.neucom.2019.12.133>
34. Tian Q, Zhang W, Wang L, Chen S, Yin H (2018) Robust ordinal regression induced by L_p-centroid. *Neurocomputing* 313:184–195. <https://doi.org/10.1016/j.neucom.2018.06.041>
35. Fernández-Navarro F, Gutiérrez PA, Hervás-Martínez C, Yao X (2013) Negative correlation ensemble learning for ordinal regression. *IEEE Trans Neural Netw Learn Syst* 24(11):1836–1849. <https://doi.org/10.1109/TNNLS.2013.2268279>
36. Fernández-Navarro F, Riccardi A, Carloni S (2014) Ordinal neural networks without iterative tuning. *IEEE Trans Neural Netw Learn Syst* 25(11):2075–2085. <https://doi.org/10.1109/TNNLS.2014.2304976>
37. Vargas VM, Gutiérrez PA, Hervás-Martínez C (2020) Cumulative link models for deep ordinal classification. *Neurocomputing* 401:48–58. <https://doi.org/10.1016/j.neucom.2020.03.034>
38. Singer G, Anuar R, Ben-Gal I (2020) A weighted information-gain measure for ordinal classification trees. *Expert Syst Appl* 152:113375. <https://doi.org/10.1016/j.eswa.2020.113375>
39. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
40. Herbrich R, Graepel T, Obermayer K (2000) Large margin rank boundaries for ordinal regression. *Adv Large Margin Classif* 88(2):115–132
41. de la Torre J, Puig D, Valls A (2018) Weighted kappa loss function for multi-class classification of ordinal data in deep learning. *Pattern Recogn Lett* 105(C):144–154. <https://doi.org/10.1016/j.patrec.2017.05.018>
42. Pan SJ, Yang Q (2009) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22(10):1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
43. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*, pp 618–626
44. Kook L, Herzog L, Hothorn T, Dürr O, Sick B (2022) Deep and interpretable regression models for ordinal outcomes. *Pattern Recogn* 122:108263. <https://doi.org/10.1016/j.patcog.2021.108263>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Riccardo Rosati¹  · Luca Romeo^{1,2} · Víctor Manuel Vargas³ · Pedro Antonio Gutiérrez³ · César Hervás-Martínez³ · Emanuele Frontoni¹

✉ Riccardo Rosati
r.rosati@pm.univpm.it

Luca Romeo
l.romeo@univpm.it

Víctor Manuel Vargas
vvargas@uco.es

Pedro Antonio Gutiérrez
pagutierrez@uco.es

César Hervás-Martínez
chervas@uco.es

Emanuele Frontoni
e.frontoni@univpm.it

¹ Department of Information Engineering, Marche Polytechnic University, Via Breccia Bianche 12, Ancona 60131, Italy

² Computational Statistics and Machine Learning, Istituto Italiano di Tecnologia, Via Enrico Melen 83, Genova 16152, Italy

³ Department of Computer Science and Numerical Analysis, University of Córdoba, Campus Universitario de Rabanales, Córdoba 14014, Spain