Applied Soft Computing 138 (2023) 110191

Contents lists available at ScienceDirect

Applied Soft Computing

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

Exponential loss regularisation for encouraging ordinal constraint to shotgun stocks quality assessment

Víctor Manuel Vargas [a,*], Pedro Antonio Gutiérrez [a], Riccardo Rosati [b], Luca Romeo [c], Emanuele Frontoni [d], César Hervás-Martínez [a]

[a] Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain
[b] Department of Information Engineering, Marche Polytechnic University, Ancona, Italy
[c] Department of Economics and Law, University of Macerata, Macerata, Italy
[d] Department of Political Sciences, Communication and International Relations, University of Macerata, Macerata, Italy

ARTICLE INFO

Article history:
Received 17 September 2021
Received in revised form 29 November 2022
Accepted 3 March 2023
Available online 9 March 2023

Keywords:
Ordinal classification
Convolutional neural networks
Loss function
Cumulative link models
Aesthetic quality control

ABSTRACT

Ordinal problems are those where the label to be predicted from the input data is selected from a group of categories which are naturally ordered. The underlying order is determined by the implicit characteristics of the real problem. They share some characteristics with nominal or standard classification problems but also with regression ones. In the real world, there are many problems of this type in different knowledge areas, such as medical diagnosis, risk prediction or quality control. The latter has gained an increasing interest in the Industry 4.0 scenario. Some weapons manufacturer follow an aesthetic quality control process to determine the quality of the wood used to produce the stock of the weapons they manufacture. This process is an ordinal classification problem that can be automatised using machine learning techniques. Deep learning methods have been widely used for multiples types of tasks including image aesthetic quality control, where convolutional neural networks are the most common alternative, given that they are focused on solving problems where the input data are images. In this work, we propose a new exponential regularised loss function that is usedto improve the classification performance for ordinal problems when using deep neural networks. The proposed methodology is applied to a real-world aesthetic quality control problem. The results and statistical analysis prove that the proposed methodology outperforms other state-of-the-art methods, obtaining very robust results.

1. Introduction

Deep learning models [1] have been widely used in the last years since they were proposed for the first time. These kind of models can be applied to multiple types of problems, including: standard classification [2], regression [3], ordinal classification [4], segmentation [5], image aesthetic assessment [6], image retrieval [7], medical diagnosis [8–10], Internet of Things [11], among others. They are mainly focused on solving problems where a large amount of data is available, such as problems where data are images, text or time series. However, thanks to data augmentation techniques, it is possible to train complex models with a reduced amount of data.

The most common type of deep learning model is the convolutional neural network (CNN) [12], which is an artificial neural network that is mainly focused on working with input data that comes in the shape of an image. The main purpose of this model is to extract relevant features from the images and use them to classify each pattern in the correct category or obtain an accurate prediction on regression problems. This process is carried out by stacking multiple convolutional blocks, which extract higher level features from low level characteristics (groups of pixels) and reduce the dimensionality of the initial data. The main element of these blocks is the convolutional layer, which applies convolution operations with multiple filters to the input received from the input images or the previous block output. Pooling operations reduce the number of features, by summarising them with maximum, minimum or average operations. In the same way it is done in shallow neural network, each convolutional block also contains an activation function that introduces non-linear transformations in the model, increasing its expressive capability. Finally, in the last years, the batch normalisation [13] operation has been included in most of the CNN models as it accelerates the training process by reducing the covariance shift. The features obtained from these convolutional blocks are used to obtain a

* Correspondence to: Campus Universitario de Rabanales, "Albert Einstein" building, 3rd floor, 14014 Córdoba, Spain
E-mail address: vvargas@uco.es (V.M. Vargas).

https://doi.org/10.1016/j.asoc.2023.110191

prediction by using standard dense (fully-connected) layers and a output layer, which is often a softmax function.

Ordinal classification [14,15] problems are those problems where the label that we aim to predict from the input data is chosen between a series of categories that follow a natural order. This order is determined by the characteristics of the real problem that is being solved. The main difference between ordinal and nominal classification is the aforementioned order between categories. Ordinal classification, also named ordinal regression, shares some aspects with regression, as, in both types of problems, there is an underlying order. However, in regression problems, the predicted variable is continuous while in ordinal classification it is discrete and limited to a set of labels [16].

In the Industry 4.0, there are interesting applications related to image classification [17] within ordinal problems [18]. Benelli Armi Spa is a widely known weapons manufacturer. When building a shotgun, they want to measure the quality of the wood that is used to create the stock of the weapon. Traditionally, a specialised quality technician has done this task, checking each of the stocks one by one. However, it is possible to automatise and reduce the intrinsic inter-operator and intra-operator variability of this process by taking pictures of both sides of the stock, which allows a subsequent classification by a CNN model. Also, this problem is an ordinal classification one as the different categories that should be taken into account are the different commercial quality grades which are usually expressed in an ordinal scale.

In this work, we propose a new regularisation function [1] that transforms the standard one-hot (0/1) encoding of the labels into a soften alternative, enhancing the classification quality and reflecting possible noise or errors in the labelling process. The main novelty of the proposed regularisation approach is that it provides enhanced flexibility and robustness by introducing a parameter that leads to better classification performance in ordinal problems compared to previously proposed alternatives. Also, it is worth noting that the proposed methodology is applied to a novel real-world problem that is related to the manufacturing industry and was generated by the demand of a specific company. Our proposal is compared with a baseline approach and other state-of-the-art methods, and the average results of three different metrics are statistically analysed.

The rest of the paper is organised as follows: in Section 2, previous proposals related to this work are analysed; in Section 3 the proposed method is described; in Section 4, the dataset related to the stocks aesthetic quality control problem is described and the experimental design along with the model are explained; in Section 5, the results of the experiments are presented; and, finally, in Section 6, the conclusions of this work are presented.

2. Related works

2.1. Ordinal classification

Any given classification problem can be defined as the problem of predicting the real class y from an input data $\mathbf{x}$, where $\mathbf{x}$ is a K-dimensional vector $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^K$, and y is chosen from a set of labels $\mathcal{Y} = \{\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{Q-1}\}$, being Q the total number of categories defined for the problem. Ordinal classification or ordinal regression problems can be defined as a special case of standard classification problems, where an order constraint is included between the categories. In this way, for this kind of problems, the labels satisfy the expression $\mathcal{C}_0 \prec \mathcal{C}_1 \prec \mathcal{C}_2 \prec \ldots \prec \mathcal{C}_{Q-1}$. The precedence operator ($\prec$) indicates that the categories follow a natural order but, in contrast to a regression problem, they are discrete labels instead of continuous. Moreover, the distance between each category does not necessarily need to be the same. In these terms, we can define the position of each class in the ordinal scale as an integer like $\mathcal{O}(\mathcal{C}_q) = q$.

Thus, in any ordinal problem, the order information described can be accounted to obtain better classification performance, reducing the errors in distant classes while trying to maximise the number of patterns correctly classified. Also, examples that are misclassified in adjacent classes should produce a lower error when evaluating the classification performance of any ordinal model.

2.2. Cumulative link models

One common approach to address ordinal classification problems taking into account the ordinal information is to use threshold-based models. Cumulative Link Models (CLM) [19] are one type of thresholds models which try to predict the probability for each of the categories accounting for the order information implicit to the problem using a set of thresholds that separate different categories and a projection obtained from the input data. Concretely, these models create a 1-dimensional linear projection from the input data, which can be denoted as $f(\mathbf{x}) \in \mathbb{R}$. This 1-D space is divided in Q segments by using a set of thresholds which can be conveniently defined to suit the classes distribution of the current ordinal problem. However, they are often learned from the training data instead of setting them manually. Thus, the threshold vector can be defined as $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_Q\}$. To divide the output space properly, these thresholds should be in ascending order, satisfying the expression $\beta_0 < \beta_1 < \cdots < \beta_Q$. The first threshold is always $-\infty$ and the last one $+\infty$. To take into account the constraint when using learnable thresholds, they are commonly redefined using an unconstrained scalar value, that represents the first threshold, and a vector of $Q-2$ elements that represent the square root of the increment that must be applied to any threshold to obtain the next one. More specifically, this reformulation can be expressed as:

$$\beta_q = \begin{cases} -\infty, & q = 0, \\ a + \sum_{i=1}^{q-1} b_i^2, & 0 < q < Q, \\ +\infty, & q = Q \end{cases} \tag{1}$$

where a is the scalar representing the first threshold and $\mathbf{b} = \{b_1, b_2, \ldots, b_{Q-2}\}$ is the aforementioned vector. The square that is applied to the b_i term guarantees that the thresholds will always be in ascending order.

Using the 1-D mapping and the thresholds that have just been defined, the CLM predicts the class $\mathcal{C}_q$ if and only if $f(\mathbf{x}) \in [\beta_{q-1}, \beta_q]$. In this way, the probability obtained from the CLM for any given class and input data can be calculated as follows:

$$P(y \prec C_q|\mathbf{x}) = g(\beta_q - f(\mathbf{x})), \tag{2}$$

where $g(x)$ is a monotonic function that is known as the link function. In previous works [14], different alternatives were explored to be used as link functions:

- $\text{logit}(p) = \log \dfrac{p}{1-p}$,
- $\text{probit}(p) = \Phi^{-1}(p)$,
- $\text{cloglog}(p) = \log(-\log(1-p))$,

where p is the probability of the ith category, $\Phi = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right)$ is the normal distribution cumulative distribution function and $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ is the Gauss error function.

The described CLM usually achieves good performance when classifying ordinal categories. However, there are some limitations: this kind of models are affected by the optimal choice of the

[1] https://github.com/victormvy/ur-exponential-loss

selected parameters, i.e. the performance of this model is highly influenced by the learned or fixed thresholds. For this reason, we propose to combine the CLM with a soft labelling approach based on a unimodal regularisation which is described in Section 2.3.

2.3. Soft labels and unimodal regularisation

Label smoothing [20] is a regularisation technique that is applied to the representation of the labels. When used to train a model, it encourages the classifier to be less confident, giving some probability to the other classes instead of focusing only on the true category. This enhances the robustness of the model in the presence of noisy labels. Label smoothing can be very useful for ordinal problems, where misclassifying a pattern in an adjacent class is more probable than predicting a distant category. The way the label smoothing is performed depends on the characteristics of the problem and is a way to introduce the ordinal information of the problem into the model. This extra ordinal information usually accelerates the convergence of the model and reduces the number of training examples needed to optimise the model.

The authors of [21] proposed to sample ordinal smooth labels from a Poisson distribution and a binomial one. Then, they also used an exponential function to obtain soft labels. The probability for class $\mathcal{C}_q$ when the target class is $\mathcal{C}_j$ using the Poisson distribution is given by:

$$P_j(q) = \frac{\lambda_q^j e^{-\lambda_q}}{j!}, \tag{3}$$

where $\lambda_q \in \mathbb{R}^+$ is the parameter of the distribution associated with class $\mathcal{C}_q$. Its mean and variance is determined by the value of its parameter λ_q. Thus, for some classes, it is not possible to obtain a distribution that is centred in the middle of the class interval while keeping the variance low. For this reason, this kind of distributions are not very appropriate and usually have severe performance pitfalls. Therefore, the authors introduced the binomial distribution, which they stated that is more flexible and provides better results. This probability for class $\mathcal{C}_j$ when the target class is $\mathcal{C}_q$ using this distribution is given by:

$$P_j(q) = \binom{Q}{q} p_q^j (1-p_q)^{Q-q}, \tag{4}$$

where p_q is the parameter associated with the distribution of class $\mathcal{C}_q$. In this case, the binomial distribution has two parameters: Q, or the number of classes, and the probability of the event (belong to a specific class). Note that, even though the mean ($E[x] = Q \cdot p$) and the variance ($V[x] = Q \cdot p(1-p)$) are determined by different expressions, it is not easy to achieve distributions correctly centred in the middle of the class interval and with a small variance. Finally, they proposed to use a exponential function as a third alternative. Using this function, the probability for class $\mathcal{C}_j$ when the true class is $\mathcal{C}_q$ is obtained as follows:

$$f_j(q) = e^{-|j-q|}, \tag{5}$$

where $j = \mathcal{O}(\mathcal{C}_j)$ and $q = \mathcal{O}(\mathcal{C}_q)$. Given that said function is not a probability function, a softmax normalisation is applied to obtain the required probabilities. The main drawback of this function is that, in some cases, the probability mass is not sufficiently concentrated in the class interval due to the lack of flexibility. For that reason, in Section 3, an improved version of this loss regularisation is proposed.

2.4. L_p norm

L_p norms have been used in optimisation algorithms in several fields as a generalisation of L_2 and L_1 norms, including binary classification [22], feature selection [23] or generative adversarial networks [24], among others. Depending on the value of p, the objective varies. In [25], the authors proved that L_2 norm methods tend to expand or bleed out over natural boundaries. Therefore, using a L_p norm where $1 < p < 2$ should provide a more suitable alternative when it is optimised properly. The use of this type of generalised norms has drawn a huge attention in multiple applications, such as 3D medical image super-resolution [26].

More specifically, multiple works have discussed the potential advantages of the alternatives to the L_2 or L_1 norm. In [27], the authors presented an L_p norm alternative to Least Squares Support Vector Machine (LSSVM) [28]. The authors of [29] proposed a new method that achieves better robustness by replacing the L_2 norm in conventional linear discriminant analysis by L_p norm in within-class distances and by L_s norm in between-class distances. However, for several of these tasks, the L_p norms have raised a huge attention. Bregman divergences is one of the standard tools for analysing online machine learning algorithms [30], allowing a generalisation of the least mean squared algorithm. In this sense, the loss bounds for these L_p norm algorithms involve others than the standard L_2 or L_1 norms [22].

In this work, we propose to apply the L_p norm to the exponential regularisation that was described in Section 2.3 to obtain soft labels with a more flexible distribution for an ordinal classification problem.

3. Proposed methodology: L_p norm exponential regularised cross-entropy loss

The exponential regularised soft labelling presented in Section 2.3 (Eq. (5)) applies a L_1 norm. In this work, we propose to sample on a more flexible exponential function based on the introduction of the L_p norm, which means that there is an extra tunable parameter that can be adjusted by the learning algorithm. In this way, a more flexible L_p normalised exponential function can be defined. The probability for $\mathcal{C}_j$ when the target class is $\mathcal{C}_q$ is defined as:

$$f_j(p, q) = e^{-|j-q|^p}, \quad 1 \leq p \leq 2, \tag{6}$$

where $j = \mathcal{O}(\mathcal{C}_j)$, $q = \mathcal{O}(\mathcal{C}_q)$, and p parameter can be tweaked manually or cross-validated. Hence, the p parameter controls how much a pattern is penalised when it is classified in class $\mathcal{C}_j$ and its real class is $\mathcal{C}_q$. Lower values of p mean that less relevance is given to that error. In this context, cost functions other than the Squared Euclidean norm (L_2) or the Manhattan Distance norm (L_1) might provide better results due to its enhanced flexibility. The range of possible values for the aforementioned parameter should be restricted to the interval of real numbers to respect the formal definition of a geometrical norm, i.e. $p \in [1, 2]$.

This regularisation of the labels is applied as an alternative to the standard categorical cross-entropy loss function:

$$\mathscr{L}(\mathbf{x}, q) = \sum_{j=1}^{Q} h(j, q)[-\log P(y = C_j|\mathbf{x})], \tag{7}$$

where $h(j, q) = \delta_{j,q}$, $j = \mathcal{O}(\mathcal{C}_j)$ and $q = \mathcal{O}(\mathcal{C}_q)$ are the predicted and ground truth classes, respectively, and $\delta_{j,q}$ is the Dirac delta, which equals to 1 for $j = q$, and 0 otherwise. $h(j, q)$ can be replaced with a soft version $h'(j, q)$, which can be obtained by applying the aforementioned exponential function. In this way, the standard definition of the loss function can be replaced by:

$$\mathscr{L}(\mathbf{x}, q) = \sum_{j=1}^{Q} h'(j, q)[-\log P(y = C_j|\mathbf{x})], \tag{8}$$

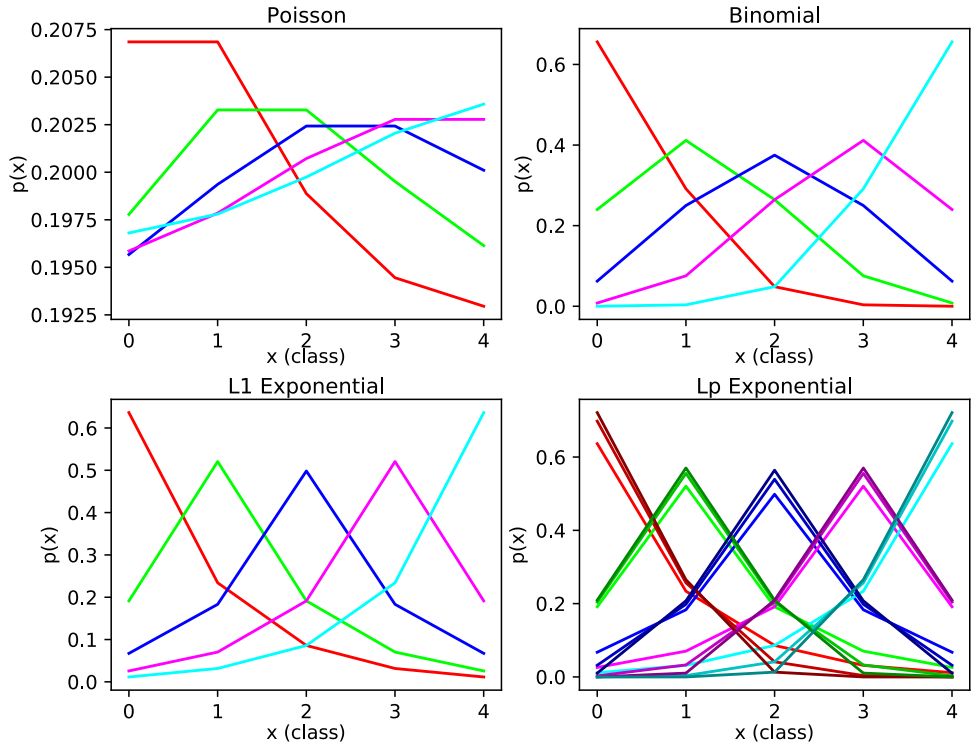

Fig. 1. Different types of distributions for a problem with 5 classes. The *x*-axis represents the evaluated class, the *y* axis represents the value of the smooth label given for the class and the colours are associated to the true class (red: 0, green: 1, blue: 2, pink: 3, and cyan: 4). Thus, each line represents the probability distribution for one real label. In the L_p Exponential plot, different intensities of the colour show different values of the p parameter (1.0, 1.5 and 2.0, where a lighter colour means higher value).

where $h'(j,q) = (1-\eta)\delta_{j,q} + \eta f_j(p,q)$ and η is a parameter that ranges from 0 to 1 controlling the smoothness of the labels. When $\eta = 0$, no smooth factor is applied. On the other side, when $\eta = 1$, the labels are completely smooth and the standard labels are not used.

Fig. 1 shows classes distributions for the proposed exponential function along with the Poisson, binomial and standard exponential distributions that were described before. The colour represents the true class of the pattern, while the *x*-axis represents the class being examined and the *y*-axis represents the soft label applied. In the case of the L_p exponential, the distributions obtained with the lower and upper bounds, and an intermediate value of the parameter are considered. Therefore, for each class, the distributions for $p \in \{1.0, 1.5, 2.0\}$ are shown. Any other distribution that can be obtained by tweaking this parameter will be in-between the lower and upper bounds distributions. $p = 1.0$ is represented with the darkest colours while $p = 2.0$ has the lightest colours in the plot. As mentioned before, the L_1 exponential is equivalent to using the L_p exponential with $p = 1$, as the latter is a more general and flexible version of the exponential function.

4. Experiments

In this Section, the details relative to the experimental design are described. First, the dataset available for the problem that is being solved is characterised. Then, the model used to solve this problem is described. Finally, the test and validation scheme, the optimisation procedure, the loss and output functions, and the hyperparameters used are shown.

4.1. Data

The wooden stocks of shotguns are commercially classified in categories ranging from grade 1, which indicates almost veinless wood, up to grade 5, which refers to a very twisted and variegated grain pattern. Nowadays, the quality control process of these stocks is performed by human eye. However, Benelli Armi Spa has created a dataset composed of images of these wooden stocks that have been labelled by a highly specialised quality control technician. The detention and conservation of this dataset are regulated by an agreement between Benelli Armi Spa and Università Politecnica delle Marche.

Table 1
Benelli dataset classes distribution.

Label	1	2^-	2	2^+	3^-	3	3^+	4^-	4	4^+
Images	165	148	212	177	179	306	344	208	275	106

The dataset contains both left and right side images belonging to different shotguns, which comprises a total of 2120 1000×500 colour images. Each of these pictures was assigned one of ten ordinal categories, which are ordered according to the existing commercial grades. In this way, 4 main grades have been defined (1, 2, 3, 4) and their relative minor grades (2^-, 2^+, 3^-, 3^+, 4^-, 4^+). The number of patterns belonging to each class is represented in Table 1.

The original images have been cropped to 470×270 and the background has been removed and replaced by a plain black colour. In these terms, the percentage of wooden stock in the image have been maximised. Fig. 2 shows two images that belong to this dataset and have already been preprocessed.

Given that the dataset was directly provided to Università Politecnica delle Marche by Benelli Armi Spa, there are no dataset splits defined, which allows us to perform the partitioning from scratch. The test and validation schemes will be discussed in Section 4.3.

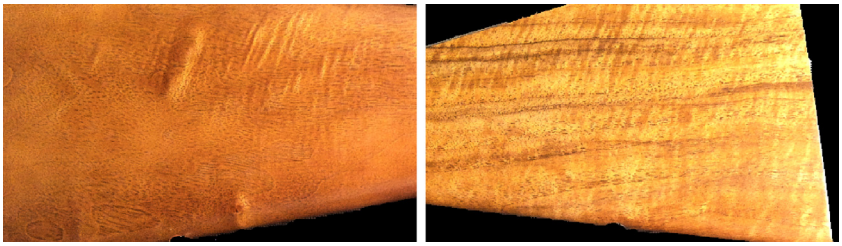

Fig. 2. Cropped dataset images from class 1 (left) and 2^+ (right).

4.2. Model

The method proposed in this paper will be applied to a convolutional neural network (CNN) model. Specifically, the well-known VGG-16 architecture is used. For the convolutional part of the model, the pre-trained ImageNet weights (included in TensorFlow Keras applications module) are used and the parameters of these layers are set to be non-learnable. Therefore, only the top part of the model is adjusted. The same architecture with the same pre-trained weights was used in [18] for solving this problem. This method accelerates the convergence and reduces the computational time significantly. In the fully-connected part of the model, a 50% dropout is applied before two dense layers with 4096 units and ReLU activation. The function used in the output layer is the softmax, for the baseline experiments, and the CLM with different link types for the rest of the experiments.

The model comprises a total of 266M of parameters where 251M are trainable, and the remaining are fixed parameters that belong to the convolutional layers that use the pre-trained ImageNet weights.

4.3. Experimental design

The performed experiments follow a 10-fold cross-validation scheme that is repeated 3 times (with 3 different seeds) to achieve 30 executions. Each of these partitions defines different and non-overlapping train and test splits with 90% of data for training and 10% for test. Also, for each of the training partitions, a hold-out is performed to divide this whole set into train and validation. In this way, the validation set can be used to lead the early-stopping strategy that prevents the overfitting by stopping the training process when the validation loss has not improved for several epochs. Moreover, validation metrics are used to adjust the hyperparameters.

Training data is fed into the model during the training process using a generator that performs data augmentation based on random horizontal flips and also creates balanced batches regarding the different classes of the problem. For the problem considered, it is important to use balanced batches as the dataset is fairly imbalanced.

The Adam [31] algorithm is used to optimise the model, and the learning rate is fixed to 0.01 for the whole learning process. During the training process, the training data is processed in batches of size 16 and the learning stage is run for a maximum of 50 epochs. For comparison purposes, different sets of experiments are performed:

1. Baseline. Softmax in the output layer and standard categorical cross-entropy (CCE) as loss function.
2. CLM with logit, probit and complementary log–log links in the output layer, and Poisson regularised CCE.
3. CLM with logit, probit and complementary log–log links in the output layer, and binomial regularised CCE.
4. CLM with logit, probit and complementary log–log links in the output layer, and exponential regularised CCE.
5. Proposed method. CLM with logit, probit and complementary log–log links, and L_p exponential regularised CCE.

As shown in Section 3, the proposed loss function has an hyperparameter (p) that must be adjusted. In this work, mentioned hyperparameter is cross-validated monitoring the validation Quadratic Weighted Kappa (QWK) [32] metric for each fold and seed. Therefore, different values of the p parameter can be obtained for different folds and seeds, as we noticed that the optimal value of p depends on the data considered. Thus, the experimental procedure to adjust this hyperparameter is described in Algorithm 1.

```
foreach seed do
    Split whole dataset in 10 folds that will be used for
     training and test.
    foreach p do
        foreach fold do
            Split all the data not included in the current fold
             into 80% for training and 20% for validation.
            Train for the number of epochs determined by
             early stopping and evaluate on the validation
             set.
        end
    end
end
foreach seed do
    foreach fold do
        Find p value that achieved the best validation QWK
         for this fold and seed.
        Evaluate on the test set (current fold) with the best
         validation p value.
    end
end
```

Algorithm 1: p parameter cross-validation procedure.

The η parameter has been fixed to $\eta = 1.0$, which achieves fully soft labels, instead of combining the standard labels with the soft labels obtained through the unimodal distributions.

5. Results

In this Section, the results of the experiments described in Section 4 are shown. We considered three metrics that are appropriate for ordinal problems and imbalanced datasets: Quadratic Weighted Kappa (QWK) [32], Minimum Sensitivity (MS) [33] and Minimum Absolute Error (MAE) [33].

The QWK can be calculated according to the following expression:

$$\text{QWK} = 1 - \frac{\sum_{i,j}^{N} \omega_{i,j} O_{i,j}}{\sum_{i,j}^{N} \omega_{i,j} E_{i,j}}, \tag{9}$$

Table 2
Mean results for 30 executions of each of the alternatives on the test set.

Loss	Output	N	Mean p	QWK	MS	MAE
CCE	Softmax	30	–	0.88712	0.14839	0.76572
CCE-Poisson	CLM Logit	30	–	0.78042	0.00000	1.73867
CCE-Poisson	CLM Probit	30	–	0.77503	0.00000	1.78581
CCE-Poisson	CLM CLogLog	30	–	0.77921	0.00000	1.65733
CCE-Binomial	CLM Logit	30	–	0.92068	0.19775	**0.70338**
CCE-Binomial	CLM Probit	30	–	0.91929	*0.20380*	0.71846
CCE-Binomial	CLM CLogLog	30	–	0.89189	0.06812	0.86134
CCE-Exp-L_p	CLM Logit	30	1.61	**0.92391**	**0.21883**	*0.70563*
CCE-Exp-L_p	CLM Probit	30	1.70	**0.92391**	0.19875	0.71013
CCE-Exp-L_p	CLM CLogLog	30	1.66	0.91368	0.15572	0.77624
CCE-Exp-L_1	CLM Logit	30	–	*0.92237*	0.18290	0.72664
CCE-Exp-L_1	CLM Probit	30	–	0.91596	0.10117	0.78666
CCE-Exp-L_1	CLM CLogLog	30	–	0.89992	0.04978	0.85809

where N is the number of samples rated, ω is the penalisation matrix (in this case, quadratic weights are considered, $\omega_{i,j} = \frac{(i-j)^2}{(Q-1)^2}$, $\omega_{i,j} \in [0, 1]$), Q is the number of classes, O is the confusion matrix, $E_{ij} = \frac{O_{i\bullet}O_{\bullet j}}{N}$, $O_{i\bullet}$ is the sum of the i-th row and $O_{\bullet j}$ is the sum of the j-th column of the matrix.

The MS metric is defined as

$$\text{MS} = \min\left\{S_q = \frac{O_{qq}}{O_{q\bullet}}, \quad q = 1, \ldots, Q\right\}, \tag{10}$$

where S_q is the sensitivity of the class $\mathcal{C}_q$, O is the confusion matrix and Q is the number of classes.

The ordinal MAE can be calculated as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i,j=1}^{Q}|i-j|O_{ij}, \tag{11}$$

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

Table 2 contains the mean results for 30 executions of each of the experiments. For the experiments where the L_p-exponential regularised categorical cross-entropy is used, the value of the p parameter was adjusted through cross-validation for each fold and seed, and the mean value is displayed under the Mean p column. The best value for each metric is highlighted with bold font face and the second best with italic font.

As can be observed from the results in said table, the L_p-exponential regularised categorical cross-entropy with the logit link obtained the best result for QWK and MS and the second-best for MAE. Also, the same loss function with the probit link achieved the same result for QWK.

The confusion matrices of the best proposed alternative (CCE-Exp-Lp + CLM Logit) and the baseline method (CCE + Softmax) are shown in Figs. 3, 4 and 5, for seeds 0, 1 and 2, respectively. Each figure represents the confusion matrices of each of the 3 seeds considered. Each matrix is obtained by accumulating the confusion matrices of all the 10 folds.

From these matrices, it can be observed that the baseline method misclassifies some patterns in distant classes, which implies an important cost for this real problem, while the proposed ordinal method has almost all the errors in the adjacent classes.

5.1. Statistical analysis

In this Section, a statistical analysis has been performed to check whether the proposed alternative provides significantly better results than the baseline and previous proposed methods. To do that, each of the metrics has been analysed separately.

First, using the QWK metric, a Kolmogorov–Smirnov [34] test has been performed to check if the 30 QWK test values are normally distributed. The test confirmed that the values of this metric follow a normal distribution (p-value < 0.05). After that, a Friedman rank test [35] has been performed to obtain the rank related to each method. The results of this test are shown in Table 3. Note that the highest rank value represents the best method.

Table 3
Friedman test results for the QWK metric. The best method according to this metric is highlighted in bold.

Method	Rank
CCE + Softmax (Baseline)	5.43
CCE-Exp-L_p + CLM CLogLog	8.20
CCE-Exp-L_p + CLM Logit	10.87
CCE-Exp-L_p + CLM Probit	**10.90**
CCE-Exp-L_1 + CLM CLogLog	5.77
CCE-Exp-L_1 + CLM Logit	10.17
CCE-Exp-L_1 + CLM Probit	8.73
CCE-Poisson + CLM Probit	2.00
CCE-Poisson + CLM Logit	1.97
CCE-Poisson + CLM CLogLog	2.07
CCE-Binomial + CLM Probit	9.70
CCE-Binomial + CLM Logit	10.13
CCE-Binomial + CLM CLogLog	5.07

The results of the Friedman test show that the proposed L_p Exponential regularised CCE with the probit link achieved the best rank concerning the QWK metric. Also, the same loss with the logit link obtained the second-best rank, which is very close to the first one.

Given that the CCE-Exp-L_p + CLM Logit and the CCE-Exp-L_p + CLM Logit have similar ranks, and the logit link function has better overall results considering all the methods, the L_p Exponential regularised CCE with logit link has been compared with the other methods using a paired sample t-test. The results of this test are shown in Table 4. The Paired differences columns show the mean and standard deviation of the differences between both methods indicated in the first column. The t column shows the value of the statistical, the df column indicates the degrees of freedom and, finally those p-values lower than $\alpha = 0.05$ indicate that there are significant differences between both methods.

As can be observed in Table 4, the logit and probit links perform similarly with the L_p exponential regularisation (p-value $= 1.0$). Also, it shows no significant differences with the standard exponential with logit link (p-value $= 0.129$), and the binomial regularisation with probit (p-value $= 0.129$) or logit link (p-value $= 0.314$). However, it performs significantly different than the rest of the methods.

Also, in Table 5, the results of the paired t-test comparing the baseline with the other approaches is shown. These results show that there are significant differences between the baseline and all the other methods (except the binomial regularisation with

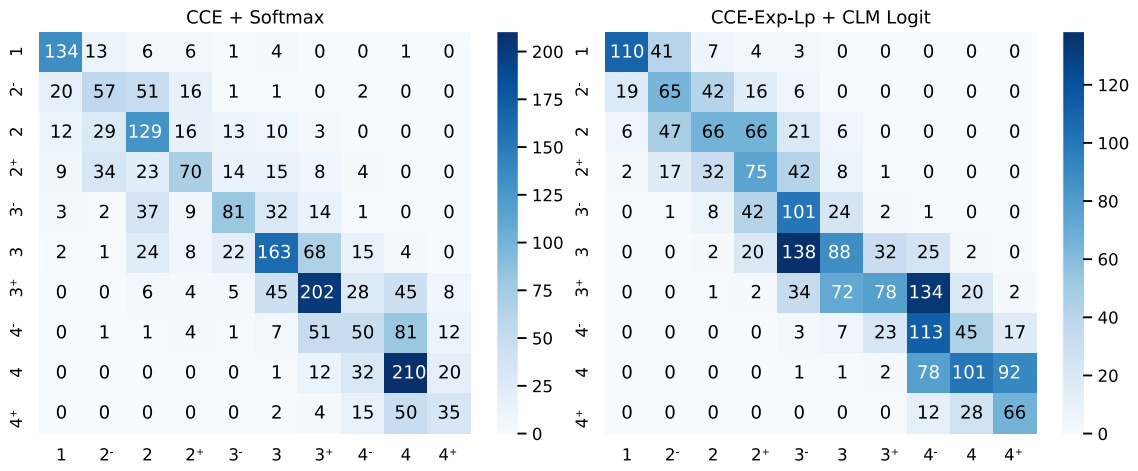

Fig. 3. Confusion matrices obtained for the seed 0.

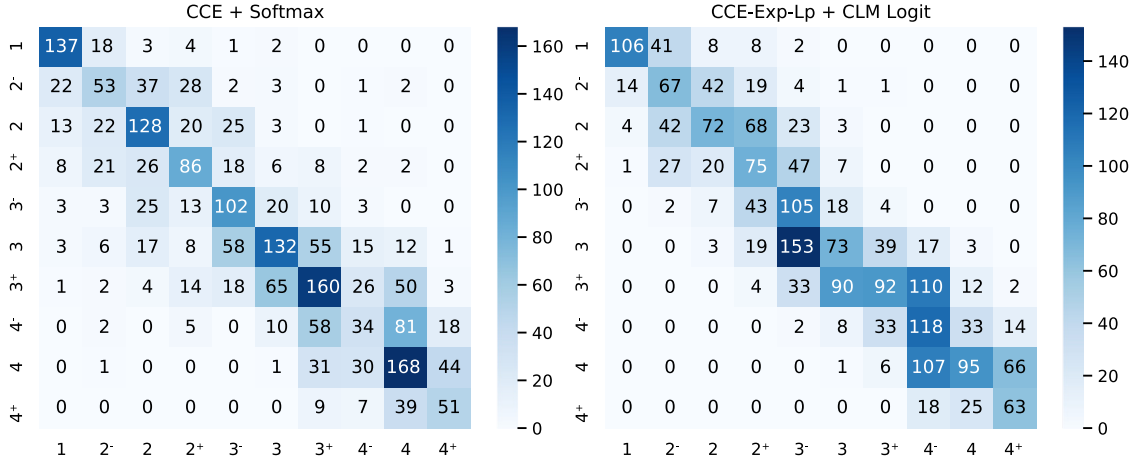

Fig. 4. Confusion matrices obtained for the seed 1.

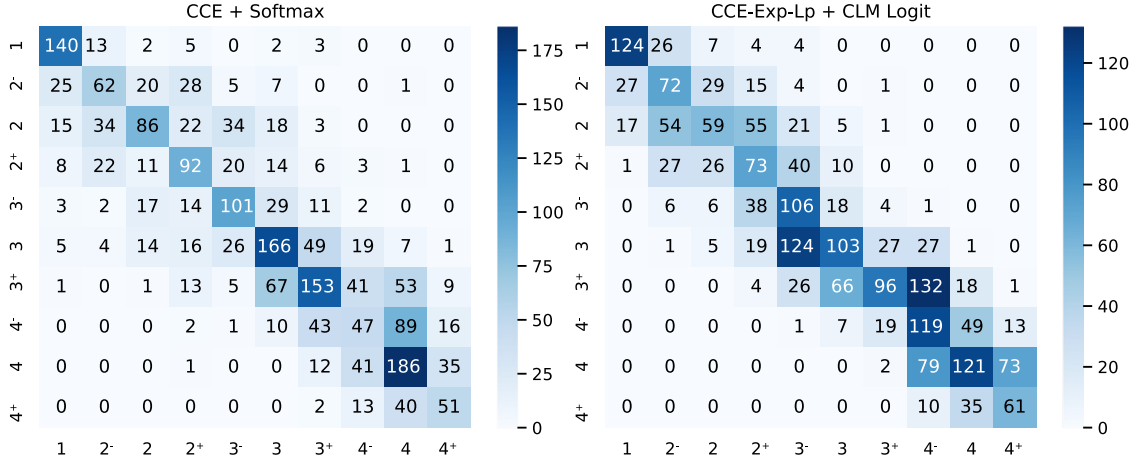

Fig. 5. Confusion matrices obtained for the seed 2.

the complementary log–log link). The method proposed in this work obtained better results than the baseline with all the link functions.

The same analysis has been performed accounting for the MS metric results on the test set. The Kolmogorov–Smirnov test reported that the values are distributed following a normal distribution (p-value < 0.05). Therefore, a Friedman rank test has been performed and the results are shown in Table 6. The highest rank in this table represents the best method.

In this case, the L_p exponential regularised CCE loss combined with the CLM with logit link obtained the best rank (10.87). The same loss function with the probit link obtained also high results. After this test, a paired sample t-test has been performed to compare the best alternative according to the ranking with the other methods. The results of this test are shown in Table 7.

Table 4
Paired sample t-test to compare L_p Exponential regularised CCE + CLM Logit with other methods regarding QWK. The *p*-values of those methods which are significantly different is highlighted in bold font.

Methods	Paired differences		*t*	df	*p*-value
	Mean	Std. Dev.			
Exp-L_p + Logit - Exp-L_p + CLogLog	0.01023	0.00917	6.110	29	<**0.001**
Exp-L_p + Logit - Exp-L_p + Probit	−0.00001	0.00676	0.000	29	1.000
Exp-L_p + Logit - CCE + Softmax	0.03679	0.02132	9.454	29	<**0.001**
Exp-L_p + Logit - Exp-L_1 + CLogLog	0.02400	0.01069	12.393	29	<**0.001**
Exp-L_p + Logit - Exp-L_1 + Logit	0.00154	0.00620	1.562	29	0.129
Exp-L_p + Logit - Exp-L_1 + Probit	0.00795	0.00995	4.377	29	<**0.001**
Exp-L_p + Logit - Poisson + Probit	0.14888	0.05688	14.336	29	<**0.001**
Exp-L_p + Logit - Poisson + Logit	0.14350	0.05042	15.587	29	<**0.001**
Exp-L_p + Logit - Poisson + CLogLog	0.14471	0.06075	13.047	29	<**0.001**
Exp-L_p + Logit - Binomial + Probit	0.00484	0.16969	1.562	29	0.129
Exp-L_p + Logit - Binomial + Logit	0.00303	0.01619	1.024	29	0.314
Exp-L_p + Logit - Binomial + CLogLog	0.03202	0.01793	9.782	29	<**0.001**

Table 5
Paired sample t-test to compare L_p CCE + Softmax (baseline) with other methods regarding QWK. The *p*-values of those methods which are significantly different is highlighted in bold font.

Methods	Paired differences		*t*	df	*p*-value
	Mean	Std. Dev.			
CCE + Softmax - Exp-L_p + CLogLog	−0.02656	0.02171	−6.702	29	<**0.001**
CCE + Softmax - Exp-L_p + Logit	−0.03679	0.02131	−9.454	29	<**0.001**
CCE + Softmax - Exp-L_p + Probit	−0.03679	0.02132	−9.453	29	<**0.001**
CCE + Softmax - Exp-L_1 + CLogLog	−0.01279	0.02208	−3.173	29	**0.004**
CCE + Softmax - Exp-L_1 + Logit	−0.03525	0.02157	−8.951	29	<**0.001**
CCE + Softmax - Exp-L_1 + Probit	−0.02884	0.02331	−6.776	29	<**0.001**
CCE + Softmax - Poisson + Probit	0.11209	0.05466	11.232	29	<**0.001**
CCE + Softmax - Poisson + Logit	0.10670	0.05949	9.824	29	<**0.001**
CCE + Softmax - Poisson + CLogLog	0.10792	0.05823	10.150	29	<**0.001**
CCE + Softmax - Binomial + Probit	−0.03195	0.03031	−5.774	29	<**0.001**
CCE + Softmax - Binomial + Logit	−0.03376	0.02759	−6.703	29	<**0.001**
CCE + Softmax - Binomial + CLogLog	−0.00477	0.03090	−0.845	29	0.405

Table 6
Friedman test results for the MS metric. The best method according to this metric is highlighted in bold.

Method	Rank
CCE + Softmax (Baseline)	8.02
CCE-Exp-L_p + CLM CLogLog	8.43
CCE-Exp-L_p + CLM Logit	**10.87**
CCE-Exp-L_p + CLM Probit	10.10
CCE-Exp-L_1 + CLM CLogLog	4.78
CCE-Exp-L_1 + CLM Logit	9.40
CCE-Exp-L_1 + CLM Probit	6.15
CCE-Poisson + CLM Probit	2.50
CCE-Poisson + CLM Logit	2.50
CCE-Poisson + CLM CLogLog	2.50
CCE-Binomial + CLM Probit	10.30
CCE-Binomial + CLM Logit	10.08
CCE-Binomial + CLM CLogLog	5.37

The results related to the Poisson regularisation have been omitted in this table since all the MS results for this method obtained a value of 0. As can be observed in Table 7, the L_p exponential regularised loss with logit link resulted significantly better than most of the other alternatives. Only the binomial regularised loss with probit and logit links obtained similar results. Another paired t-test was performed to compare the baseline with the rest of the methods. The results of this test shown that the proposed method obtained better results than the baseline and is significantly better when using the logit (p-value = 0.002) or probit (p-value = 0.036) links.

Finally, the results concerning the MAE metric have been analysed in the same way that in the previous analyses. First, a Kolmogorov–Smirnov test has been used to confirm that the results are normally distributed (p-value < 0.05). Then, a Friedman rank test has been performed to obtain a ranking of the methods regarding the MAE metric. The results are shown in Table 8. Note that, in this case, the lowest rank shows the method that obtained the best performance.

The test reported that the best method is the one that uses the Binomial regularisation combined with the CLM with logit link. However, the method that uses the L_p exponential regularisation with the logit link obtained very close results. In these terms, to compare this method with the rest of alternatives, a paired sample t-test was performed. The results of the aforementioned test are shown in Table 9.

The analysed method shows significant differences with almost all the other alternatives. However, the L_p exponential regularisation with the probit link and the binomial regularisation with probit or logit link are not statistically different. In the same way we did for the other metrics, another paired t-test was performed to compare the baseline with the other approaches. The test reported significant differences between the baseline and the proposed method when using the logit (p-value = 0.002) or probit (p-value = 0.001) links.

To sum up, the Exp-L_p + Logit obtained the best overall results for most of the metrics. It is the best method in terms of QWK and MS metrics, showing significant differences for the latter. Also, it is better than the standard exponential regularisation (L1) in all the three metrics and provides significant improvements for MS and MAE. Finally, the Binomial regularisation with logit link achieved slightly better results than the Exp-L_p with the same link concerning the MAE metric, but there are no significant differences between these methods. Nevertheless, it is worth mentioning that the L_p exponential improved not only the baseline results but also the results of the standard exponential function.

Table 7
Paired sample t-test to compare L_p Exponential regularised CCE + CLM Probit with other methods regarding MS. The *p*-values of those methods which are significantly different is highlighted in bold font.

Methods	Paired differences		*t*	df	*p*-value
	Mean	Std. Dev.			
Exp-L_p + Logit - Exp-L_p + CLogLog	0.06311	0.07868	4.393	29	<**0.001**
Exp-L_p + Logit - Exp-L_p + Probit	0.02009	0.05444	2.021	29	**0.049**
Exp-L_p + Logit - CCE + Softmax	0.07044	0.11548	3.341	29	**0.002**
Exp-L_p + Logit - Exp-L_1 + CLogLog	0.16905	0.07357	12.585	29	<**0.001**
Exp-L_p + Logit - Exp-L_1 + Logit	0.03593	0.05634	3.494	29	**0.002**
Exp-L_p + Logit - Exp-L_1 + Probit	0.11766	0.07202	8.949	29	<**0.001**
Exp-L_p + Logit - Binomial + Probit	0.02193	0.10174	1.181	29	0.247
Exp-L_p + Logit - Binomial + Logit	0.01418	0.10152	0.765	29	0.450
Exp-L_p + Logit - Binomial + CLogLog	0.15071	0.08263	9.990	29	<**0.001**

Table 8
Friedman test results for the MAE metric. The best method according to this metric is highlighted in bold.

Method	Rank
CCE + Softmax (Baseline)	5.75
CCE-Exp-L_p + CLM CLogLog	6.48
CCE-Exp-L_p + CLM Logit	3.23
CCE-Exp-L_p + CLM Probit	3.48
CCE-Exp-L_1 + CLM CLogLog	9.13
CCE-Exp-L_1 + CLM Logit	4.13
CCE-Exp-L_1 + CLM Probit	7.20
CCE-Poisson + CLM Probit	12.17
CCE-Poisson + CLM Logit	12.17
CCE-Poisson + CLM CLogLog	11.67
CCE-Binomial + CLM Probit	3.87
CCE-Binomial + CLM Logit	**2.90**
CCE-Binomial + CLM CLogLog	8.82

6. Conclusions

The main contribution of this work was to propose a novel more flexible exponential regularisation method based on introducing a L_p norm into a previously proposed exponential regularised loss. This loss regularisation is appropriate for ordinal problems where the misclassification errors should be in adjacent classes instead of distant classes, encouraging labels distribution to be soft and unimodal, being centred in the middle of the real class interval. Moreover, the softmax that is commonly used in the output of the model was replaced with the cumulative link model with different types of links, which is also more adequate for ordinal problems.

The described method was applied to solve a novel real-world problem that was generated by the specific demand of an industrial manufacturing company and consists in classifying shotgun stocks accounting for the quality of the wood that they are made of. Different categories reference different levels of quality, which are naturally ordered. The method proposed was compared with a baseline approach, which uses the standard categorical cross-entropy loss and the softmax at the output of the model. Also, experiments using previously proposed regularisation methods were run to do further comparisons. Three ordinal metrics were used to compare all these methods: QWK, MS and MAE.

The results demonstrated that the proposed alternative achieved the best result for QWK and MS, and the second-best result for MAE. Also, the statistical tests demonstrated the robustness and the effectiveness of the proposed approach and the gain with respect to previous alternatives, which is significant and not caused by a random component or noise in the data. Moreover, the results obtained for the application problem that we aimed to solve have been very successful, achieving a high QWK value, which implies small classification errors, where most of them occur in the adjacent classes. In this way, this method can be applied in Industry 4.0 as a Decision Support System (DSS) to help classify weapon stocks according to their quality. Also, it can be used on other industrial problems where the input data are images and the categories are naturally ordered. The setup of the system is fairly simple, as it only requires the acquisition box, which takes the pictures of the stocks, and the deep learning-based DSS. It helps the human operator in the task of classifying each shotgun stock, significantly reducing the inference time. Moreover, the current dataset can be complemented with every new stock that is classified, increasing the robustness of the model and providing excellent scalability properties.

In future works, the regularised loss function and the ordinal output model described in this manuscript can be applied to more complex CNN models, which could lead to enhanced performance. Also, the same model used in this work can be applied to different ordinal problems where the input data are images and the categories follow an order. In these terms, new DSS for other real problems can be developed.

CRediT authorship contribution statement

Víctor Manuel Vargas: Methodology, Software, Writing – original draft, Investigation, Visualization. **Pedro Antonio Gutiérrez:** Conceptualization, Methodology, Validation, Writing – review & editing. **Riccardo Rosati:** Methodology, Validation, Data curation, Writing – review & editing. **Luca Romeo:** Methodology, Data curation, Resources, Writing – review & editing. **Emanuele Frontoni:** Formal analysis, Supervision, Writing – review & editing, Project administration, Funding acquisition. **César Hervás-Martínez:** Formal analysis, Supervision, Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work has been supported by "Agencia Española de Investigación (España)" (grant reference: PID2020-115454GB-C22 /AEI/10.13039/501100011033). This work has been also 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

Table 9
Paired sample t-test to compare L_p Exponential regularised CCE + CLM Probit with other methods regarding MAE. The *p*-values of those methods which are significantly different is highlighted in bold font.

Methods	Paired differences		*t*	df	*p*-value
	Mean	Std. Dev.			
Exp-L_p + Logit - Exp-L_p + CLogLog	−0.07061	0.06233	−6.205	29	<**0.001**
Exp-L_p + Logit - Exp-L_p + Probit	−0.00450	0.04988	−0.494	29	0.625
Exp-L_p + Logit - CCE + Softmax	−0.06009	0.09471	−3.475	29	**0.002**
Exp-L_p + Logit - Exp-L_1 + CLogLog	−0.15246	0.06992	−11.943	29	<**0.001**
Exp-L_p + Logit - Exp-L_1 + Logit	−0.02101	0.03605	−3.192	29	**0.003**
Exp-L_p + Logit - Exp-L_1 + Probit	−0.08103	0.06855	−6.475	29	<**0.001**
Exp-L_p + Logit - Poisson + Probit	−1.08018	0.32809	−18.033	29	<**0.001**
Exp-L_p + Logit - Poisson + Logit	−1.03304	0.28901	−19.578	29	<**0.001**
Exp-L_p + Logit - Poisson + CLogLog	−0.95170	0.34386	−15.159	29	<**0.001**
Exp-L_p + Logit - Binomial + Probit	−0.01324	0.09093	−0.798	29	0.432
Exp-L_p + Logit - Binomial + Logit	0.00266	0.07882	0.185	29	0.854
Exp-L_p + Logit - Binomial + CLogLog	−0.15571	0.07550	−11.296	29	<**0.001**

experimental development in the areas of smart specialisation - LINEA 2 -Bando 2019, approved with DDPF 293 of 22/11/2019. Víctor Manuel Vargas's research has been subsidised by the FPU Predoctoral Program of the Spanish Ministry of Science, Innovation and Universities (MCIU), grant references FPU18/00358 and EST22/00163. Funding for open access charge: Universidad de Córdoba / CBUA.

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, http://dx.doi.org/10.1038/nature14539.
[2] A. Gupta, S. Gupta, R. Katarya, et al., InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray, Appl. Soft Comput. 99 (2021) 106859, http://dx.doi.org/10.1016/j.asoc.2020.106859.
[3] Y. Dong, X. Ma, T. Fu, Electrical load forecasting: A deep learning approach based on K-nearest neighbors, Appl. Soft Comput. 99 (2021) 106900, http://dx.doi.org/10.1016/j.asoc.2020.106900.
[4] M. Pérez-Ortiz, M. Cruz-Ramírez, M.D. Ayllón-Terán, N. Heaton, R. Ciria, C. Hervás-Martínez, An organ allocation system for liver transplantation based on ordinal regression, Appl. Soft Comput. 14 (2014) 88–98, http://dx.doi.org/10.1016/j.asoc.2013.07.017.
[5] L. Li, X. Zhao, W. Lu, S. Tan, Deep learning for variational multimodality tumor segmentation in PET/CT, Neurocomputing 392 (2019) 1–19, http://dx.doi.org/10.1016/j.neucom.2018.10.099.
[6] H. Zeng, Z. Cao, L. Zhang, A.C. Bovik, A unified probabilistic formulation of image aesthetic assessment, IEEE Trans. Image Process. 29 (2019) 1548–1561, http://dx.doi.org/10.1109/TIP.2019.2941778.
[7] A. Qayyum, S.M. Anwar, M. Awais, M. Majid, Medical image retrieval using deep convolutional neural network, Neurocomputing 266 (2017) 8–20, http://dx.doi.org/10.1016/j.neucom.2017.05.025.
[8] S. Zhou, B. Tan, Electrocardiogram soft computing using hybrid deep learning CNN-ELM, Appl. Soft Comput. 86 (2020) 105778, http://dx.doi.org/10.1016/j.asoc.2019.105778.
[9] D. Ezzat, A.E. Hassanien, H.A. Ella, An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization, Appl. Soft Comput. (2020) 106742, http://dx.doi.org/10.1016/j.asoc.2020.106742.
[10] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, Y. Dong, The ensemble deep learning model for novel COVID-19 on CT images, Appl. Soft Comput. 98 (2021) 106885, http://dx.doi.org/10.1016/j.asoc.2020.106885.
[11] Y.-S. Su, C.-F. Ni, W.-C. Li, I.-H. Lee, C.-P. Lin, Applying deep learning algorithms to enhance simulations of large-scale groundwater flow in IoTs, Appl. Soft Comput. 92 (2020) 106298, http://dx.doi.org/10.1016/j.asoc.2020.106298.
[12] G. Marques, D. Agarwal, I. de la Torre Díez, Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network, Appl. Soft Comput. 96 (2020) 1–11, http://dx.doi.org/10.1016/j.asoc.2020.106691.
[13] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on Machine Learning, Vol. 37, 2015, pp. 448–456.
[14] V.M. Vargas, P.A. Gutiérrez, C. Hervás, Deep ordinal classification based on the proportional odds model, in: Proceedings of the International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, 2019, pp. 441–451, http://dx.doi.org/10.1007/978-3-030-19651-6_43.
[15] J.L. Suárez, S. García, F. Herrera, Ordinal regression with explainable distance metric learning based on ordered sequences, Machine Learning 110 (10) (2021) 2729–2762, http://dx.doi.org/10.1007/s10994-021-06010-w.
[16] P. Bellmann, F. Schwenker, Ordinal classification: Working definition and detection of ordinal structures, IEEE Access 8 (2020) 164380–164391.
[17] J. Villalba-Diez, D. Schmidt, R. Gevers, J. Ordieres-Meré, M. Buchwitz, W. Wellbrock, Deep learning for industrial computer vision quality control in the printing industry 4.0, Sensors 19 (18) (2019) 3987, http://dx.doi.org/10.3390/s19183987.
[18] R. Rosati, L. Romeo, G. Cecchini, F. Tonetto, L. Perugini, L. Ruggeri, P. Viti, E. Frontoni, Bias from the wild industry 4.0: Are we really classifying the quality or shotgun series? in: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV, Springer, 2021, pp. 637–649, http://dx.doi.org/10.1007/978-3-030-68799-1_46.
[19] A. Agresti, Analysis of Ordinal Categorical Data, Vol. 656, J. Wiley & Sons, 2010.
[20] C.-B. Zhang, P.-T. Jiang, Q. Hou, Y. Wei, Q. Han, Z. Li, M.-M. Cheng, Delving deep into label smoothing, IEEE Trans. Image Process. 30 (2021) 5984–5996, http://dx.doi.org/10.1109/TIP.2021.3089942.
[21] X. Liu, F. Fan, L. Kong, Z. Diao, W. Xie, J. Lu, J. You, Unimodal regularized neuron stick-breaking for ordinal classification, Neurocomputing 388 (7) (2020) 34–44, http://dx.doi.org/10.1016/j.neucom.2020.01.025.
[22] C. Gentile, The robustness of the p-norm algorithms, Mach. Learn. 53 (3) (2003) 265–299, http://dx.doi.org/10.1023/A:1026319107706.
[23] Y.-F. Ye, Y.-H. Shao, C.-N. Li, Wavelet Lp-norm support vector regression with feature selection, J. Adv. Comput. Intell. Intell. Inform. 19 (3) (2015) 407–416, http://dx.doi.org/10.20965/jaciii.2015.p0407.
[24] C. Zhou, J. Zhang, J. Liu, Lp-WGAN: Using lp-norm normalization to stabilize Wasserstein generative adversarial networks, Knowl.-Based Syst. 161 (2018) 415–424, http://dx.doi.org/10.1016/j.knosys.2018.08.004.
[25] M. Liu, D.F. Gleich, Strongly local p-norm-cut algorithms for semi-supervised learning and local graph clustering, 2020, arXiv preprint arXiv:2006.08569.
[26] K. Thurnhofer-Hemsi, E. López-Rubio, N. Roe-Vellve, M.A. Molina-Cabello, Multiobjective optimization of deep neural networks with combinations of Lp-norm cost functions for 3D medical image super-resolution, Integr. Comput.-Aided Eng. (Preprint) (2020) 1–19, http://dx.doi.org/10.3233/ICA-200620.
[27] T. Ke, L. Zhang, X. Ge, H. Lv, M. Li, Construct a robust least squares support vector machine based on L_p-norm and L_∞-norm, Eng. Appl. Artif. Intell. 99 (2021) 104134, http://dx.doi.org/10.1016/j.engappai.2020.104134.
[28] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300, http://dx.doi.org/10.1023/A:1018628609742.
[29] Q. Ye, L. Fu, Z. Zhang, H. Zhao, M. Naiem, Lp-and Ls-norm distance based robust linear discriminant analysis, Neural Netw. 105 (2018) 393–404, http://dx.doi.org/10.1016/j.neunet.2018.05.020.
[30] J. Kivinen, M.K. Warmuth, B. Hassibi, The p-norm generalization of the LMS algorithm for adaptive filtering, IEEE Trans. Signal Process. 54 (5) (2006) 1782–1793, http://dx.doi.org/10.1109/TSP.2006.872551.
[31] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the International Conference on Learning Representations, 2015, pp. 1–15.
[32] J. de la Torre, D. Puig, A. Valls, Weighted kappa loss function for multi-class classification of ordinal data in deep learning, Pattern Recognit. Lett. 105 (2018) 144–154, http://dx.doi.org/10.1016/j.patrec.2017.05.018.
[33] M. Cruz-Ramírez, C. Hervás-Martínez, J. Sánchez-Monedero, P.A. Gutiérrez, Metrics to guide a multi-objective evolutionary algorithm for ordinal classification, Neurocomputing 135 (2014) 21–31, http://dx.doi.org/10.1016/j.neucom.2013.05.058.
[34] F.J. Massey Jr., The Kolmogorov-Smirnov test for goodness of fit, J. Amer. Statist. Assoc. 46 (253) (1951) 68–78, http://dx.doi.org/10.1080/01621459.1951.10500769.
[35] G.A. Mack, J.H. Skillings, A Friedman-type rank test for main effects in a two-factor ANOVA, J. Amer. Statist. Assoc. 75 (372) (1980) 947–951, http://dx.doi.org/10.1080/01621459.1980.10477577.