

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/284096440>

# On the Use of Nominal and Ordinal Classifiers for the Discrimination of States of Development in Fish Oocytes

Article in *Neural Processing Letters* · October 2016

DOI: 10.1007/s11063-015-9476-8

CITATIONS

6

READS

205

6 authors, including:



**María Pérez-Ortiz**

University College London

113 PUBLICATIONS 1,650 CITATIONS

[SEE PROFILE](#)



**E. Cernadas**

University of Santiago de Compostela

66 PUBLICATIONS 3,680 CITATIONS

[SEE PROFILE](#)



**Rosario Domínguez-Petit**

Instituto Español de Oceanografía

97 PUBLICATIONS 979 CITATIONS

[SEE PROFILE](#)



**Pedro Antonio Gutiérrez**

University of Cordoba (Spain)

220 PUBLICATIONS 3,545 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



ALGORITMOS DE CLASIFICACION ORDINAL Y PREDICCIÓN EN ENERGÍAS RENOVABLES (ORDINAL CLASSIFICATION AND PREDICTION ALGORITHMS IN RENEWABLE ENERGY, ORCA-RE) TIN2014-54583-C2-1-R Financial Entity: Ministerio de Economía y Competitividad.MINECO [View project](#)



INVIPESCA-FORT [View project](#)

# On the Use of Nominal and Ordinal Classifiers for the Discrimination of States of Development in Fish Oocytes

M. Pérez-Ortiz<sup>1</sup> · M. Fernández-Delgado<sup>3</sup> · E. Cernadas<sup>3</sup> · R. Domínguez-Petit<sup>4</sup> · P. A. Gutiérrez<sup>2</sup> · C. Hervás-Martínez<sup>2</sup>

© Springer Science+Business Media New York 2015

**Abstract** The analysis of microscopic images of fish gonad cells (*oocytes*) is a useful tool to estimate parameters of fish reproductive ecology and to analyze fish population dynamics. The study of oocyte dynamics is needed to understand ovary development and reproductive cycle of fish. Oocytes go through different developmental states in a continuum temporal sequence providing an interesting example of ordinal classification, which is not exploited by the current oocyte analysis software. This promising paradigm of machine learning known as ordinal classification or ordinal regression focus on classification problems where there exist a natural order between the classes, thus requiring specific methods and evaluation metrics. In this paper we compare 11 ordinal and 15 nominal state-of-the-art classifiers using oocytes of three fish species (*Merluccius merluccius*, *Trisopterus luscus* and *Reinhardtius hippoglossoides*). The best results are achieved by SVMOD, an ordinal decomposition method of the labelling space based on the Support Vector Machine, varying strongly with the number of states for each specie (about 95 and 80 % of accuracy with three and six states respectively). The classifiers designed specially for ordinal classification are able to capture the underlying nature of the state ordering much better than common nominal classifiers. This is demonstrated by several metrics specially designed to measure misclassification errors associated to states far in the ranking scale.

**Keywords** Fish oocytes · Ordinal classification · Texture analysis · *Reinhardtius hippoglossoides* · Decomposition methods

---

✉ M. Fernández-Delgado  
manuel.fernandez.delgado@usc.es

<sup>1</sup> Department of Mathematics and Engineering, University of Loyola Andalucía, Third Building, 14004 Córdoba, Spain

<sup>2</sup> Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain

<sup>3</sup> CITIUS: Centro de Investigación en Tecnoloxías da Información da USC, University of Santiago de Compostela, Campus Vida, Santiago de Compostela, Spain

<sup>4</sup> Department of Fisheries Ecology, Instituto de Investigaciones Marinas, Agencia Estatal Consejo Superior de Investigaciones Científicas, Vigo, Spain

## 1 Introduction

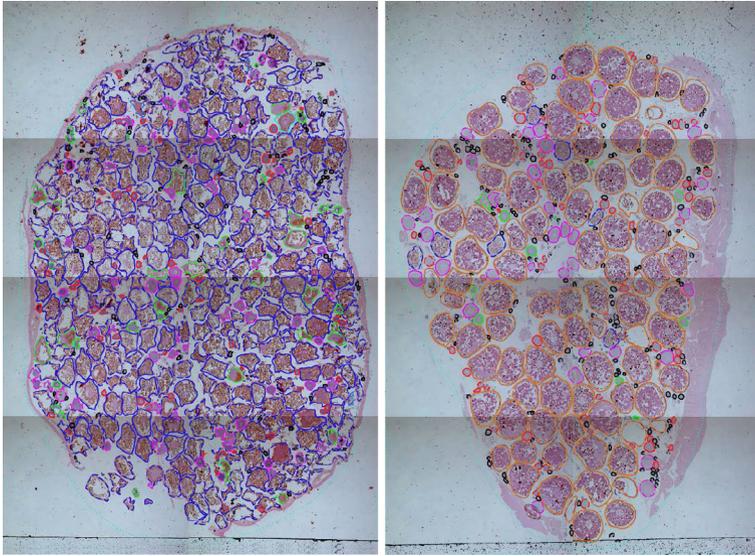
The assessment of oocyte development dynamic and fecundity is a fundamental topic in the study of reproductive biology and population dynamics [19]. To estimate fecundity with accuracy, only mature oocytes must be considered, which requires a reliable classification of oocytes according to its state of development. The best method to classify oocytes is histology, although experienced personnel is required. The main developmental states of oocytes are: *Primary Growth* (PG), *Cortical Alveoli* (CA), *Vitellogenic* (VIT), *Hydrated* (HYD) and *Atretic* (AT). The PG state corresponds to immature oocytes; CA, VIT and HYD to mature ones; and AT corresponds to those mature oocytes that will be resorpted (i.e., non-ovulated). Depending on the objective of the study, these main states could be divided in sub-states. Specifically, the specie *Reinhardtius hippoglossoides*, also known as Greenland halibut, presents some irregularities in the maturation processes [20,27] that could suggest that individual spawning does not necessarily occur on an annual basis as for most exploited fish. This specie presents a unique reproductive development pattern, with ovaries simultaneously containing oocytes developing for the current and subsequent reproductive seasons [22,30]. Four sub-levels of development within the VIT state have been identified (VIT1, VIT2, VIT3 and VIT4) in this specie (see Fig. 1). When maturation begins, a group of oocytes evolves from PG to CA and progresses until reach VIT2; then some oocytes (called the leading cohort) continue the progression (VIT3-VIT4-HYD), while the rest of mature oocytes (secondary cohort) remains in VIT2 (likely until the next spawning season) or become AT. To analyze oocyte cohort dynamic and estimate egg production it is necessary to classify correctly the VIT sub-states.

In a previous work [15] we developed Govocitos,<sup>1</sup> an automatic image analysis software which uses color texture classification to discriminate oocytes in the four main states of development (CA, VIT, HYD and AT), although states VIT and AT could not be reliably distinguished. Govocitos achieves acceptable accuracies with oocytes of two gadiform species, *Merluccius merluccius* and *Trisopterus luscus*, but it does not consider the VIT sub-levels nor the temporal evolution of states, which limits its usefulness for understanding the oocyte development. In this paper we solve this deficiency considering the whole time line of developmental states, using ordinal classifiers to fully capture the temporal evolution of states and extending the study to species with more states, such as *Reinhardtius hippoglossoides*. Several issues need to be taken into account in order to exploit the presence of this order structure. First of all, the learner (classifier, in this case) could benefit from this implicit ordering in order to construct more robust and fairer decision regions for the data, since the classification errors to be minimized vary from the ones considered in the nominal classification paradigm. Secondly, with the final aim of evaluating the performance of those classifiers, different measures or metrics could be developed and used. Section 2 introduces the ordinal classification setting and describes the most commonly used ordinal classifiers. Section 3 presents the experiments and discusses the results. Section 4 compiles some conclusions of the work.

## 2 Ordinal Classification Methods

The classification of patterns into naturally ordered labels is usually referred to as ordinal regression or classification. This paradigm is receiving much attention from the pattern

<sup>1</sup> <http://citius.usc.es/w/govocitos>.



**Fig. 1** Examples of histological images of fish specie *Reinhardtius hippoglossoides*. The cell outlines were manually annotated by experts using the Govocitos software tool. The color identifies the state of development of the oocyte: black (PG), red (CA), pink (VIT1), cyan (VIT2), blue (VIT3), orange and green (VIT4)

recognition and machine learning communities, given its applicability to real world problems (economy, medicine, psychology and others). Ordinal classification is usually said to lie between both classification and regression: as opposed to multinomial (or nominal) classification, there exists some ordering between the categories in the labelling space  $\mathcal{Y}$  and both standard classifiers and the common zero-one loss function do not capture and reflect this ordering; in contrast to regression,  $\mathcal{Y}$  is a finite set and a non-metric space (i.e. distances between categories are unknown). More formally, the aim of ordinal classifiers is to learn a prediction rule  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , where  $\mathcal{X} \subset \mathbb{R}^d$  corresponds to the input space and  $\mathcal{Y}$  to the labelling space. Therefore,  $f$  will assign an input pattern  $\mathbf{x}_i \in \mathcal{X}$  to one of the  $K$  discrete classes  $C_k, k \in \{1, \dots, K\}, C_k \in \mathcal{Y}$ , where there exist a given ordering between the labels (i.e.,  $C_1 < C_2 < \dots < C_K, <$  denoting this order information).

In the current paper, the classes correspond to the states of development of oocytes, which are naturally ordered by its growing along the time. This natural order requires to penalize differently the misclassification errors: it is less wrong e.g. to assign a oocyte in state 1 to state 2 than to state 5, because the oocyte developments are more similar between states 2 and 1 than between states 5 and 1. Concerning ordinal problems, a common (although not correct) approach is to use nominal classifiers (obviating the ordinal information) or regressors (assuming that the distances between different categories are known and equal). Contrarily to these approaches, ordinal classifiers have been shown to achieve better performance (in terms of the class ordering) for multiple ordinal classification problems [16]. In the current paper, we test this hypothesis comparing the most outstanding ordinal classifiers (described briefly in the following subsections) and nominal classifiers in the classification of developmental states of fish oocytes.

## 2.1 Threshold Methods

Thresholds models assume that an underlying, unobservable real-valued outcome (the latent variable) exists for ordered crisp classes. These methodologies estimate: (1) a function  $f(\mathbf{x})$  to predict the nature of the latent variable, i.e., a projection that maintains the classes ordered according to their rank; and (2) a vector of thresholds  $\mathbf{b} = (b_1, b_2, \dots, b_{K-1}) \in \mathbb{R}^{K-1}$  (where  $K$  is the number of classes) to represent the intervals in the range of  $f(\mathbf{x})$ , where  $b_1 \leq b_2 \leq \dots \leq b_{K-1}$ . In our problem, a threshold model would try to uncover the latent variable related to the actual level of development of an oocyte, and the thresholds would divide this latent variable into the states of development considered. The first method in this category is the proportional odds model (POM) [26], a reformulation of Logistic Regression for ordinal classification, which links the cumulative probabilities to a linear predictor  $f$  and imposes a stochastic ordering of the input space. The ordinal version of discriminant learning, called Kernel discriminant learning ordinal regression (KDLOR) [33], constraints the classes to be ordered according to their ranking in the projection to optimize. Finally, the support vector ordinal regression with implicit constraints (SVORIM) is a reformulation of the Support Vector paradigm [7] which seeks for  $K - 1$  parallel separating hyperplanes to divide the data.

## 2.2 Decomposition Methods

These techniques rely on the idea of decomposing the original ordinal problem into sets of simpler binary classification tasks [11, 34], which can be solved either by a single model or by a set of models. The subproblems are defined by a very natural methodology, considering whether a pattern  $\mathbf{x}$  belongs to a class greater than a fixed  $k$  and combining the binary predictions in a unique ordinal class [24]. This idea has demonstrated very powerful for ordinal classification, in the same way as one-vs-one and one-vs-all approaches for nominal multi-class classification. The first decomposition method [11] computes  $K - 1$  binary classification models and relabels the dataset considering whether a pattern belongs to a class greater than a fixed  $k$  (which ranges from 1 to  $K - 1$ ). The posteriori output probabilities of each model are then fused to provide a unique ordinal prediction. Originally, the C4.5 decision tree classifier was used as the base binary methodology, but it has been recently demonstrated [34] that the SVM paradigm also leads to good performance for this purpose. Furthermore, it has been shown that the use of different weights per pattern (derived from the distances to the class  $k$ ) helps to improve the performance. The combination of decomposed labels, weights per pattern and SVM base methodology will be referred in the experimental section as SVM with ordinal decompositions (SVMOD). A reformulation of the extreme learning machine, called extreme learning machine for ordinal regression (ELMOR) [9], uses the one-of- $K$  coding matrix for the outputs (commonly used with artificial neural networks) and considers whether a pattern belongs to a class greater than a fixed  $k$ . Finally, the ensemble learning for ordinal regression with product combiner and SVM (EPSVM) combines binary and ternary classification tasks, trying to distinguish each class from the previous and subsequent ones and making use of a probability fusion function [29].

## 2.3 Reduction Methods

These methods can also be seen as decomposition techniques, although with slight differences. The REDuction SVM (REDSVM) [24] transforms the training data  $(\mathbf{x}_i, y_i)$  to extended data  $(\mathbf{x}_i^k, y_i^k)$ ,  $1 \leq k \leq K - 1$ , in such a way that  $\mathbf{x}_i^k = (\mathbf{x}_i, k)$ ,  $y_i^k = 2\llbracket k < y_i \rrbracket - 1$ , being

$\llbracket \cdot \rrbracket$  a Boolean test which is 1 if the inner condition is true, and 0 otherwise, and using specific misclassification weights:  $w_{y_i,k} = |c_{y_i,k} - c_{y_i,k+1}|$ , where  $\mathbf{C}$  is a cost matrix, with  $c_{y_i,k-1} \geq c_{y_i,k}$  if  $k \leq y_i$  and  $c_{y_i,k} \leq c_{y_i,k+1}$  if  $k \geq y_i$ . Then, a binary classifier  $f$  is used with the extended data generating probabilistic values which are used to give an output prediction. The data replication method in [4], that will be referred as ordinal neural network (ONN), represents a similar framework, except that it is based on a multi-layer perceptron (MLP) neural network instead of SVM, being also less flexible because it assumes the absolute cost for the  $\mathbf{C}$  matrix.

## 2.4 Ensemble Based Techniques

Opposed to the previous methods, some efforts have been made to derive a Boosting algorithm for ordinal regression by using thresholded ensemble models, with robustness for approximating complex labelling spaces [25]. This model is composed of confidence functions, and their weighted linear combination is used as the projection for the data. We tested two different approaches, ordinal regression boosting (ORBoost) and ordinal regression boosting using perceptrons (ORBoostP), which use MLP neural networks and single perceptrons as base learners respectively.

## 3 Experimental Work

Subsamples of fixed ovaries were embedded in paraffin, sectioned at  $3.5 \mu\text{m}$  and stained with *Haematoxylin-Eosin* standard protocol. We used *Leica*<sup>2</sup> hardware and software: a *DRE* research microscope to digitalize the histological sections, connected to a *DFC320* digital camera with *IM50* software, and the *Application Suite v.4.1* software to create mosaic images. The exposure time and color balance were set automatically. The spatial resolution at which the images were captured was  $1.095 \mu\text{m}$  per pixel for species *Merluccius merluccius* (MC) and *Trisopterus luscus* (TL), and  $3.943 \mu\text{m}$  per pixel for *Reinhardtius hippoglossoides* (RH). The outline of cells was manually drawn and classified by expert technicians using the Govocitos software. Color texture analysis relates the chromatic and textural information of images, providing good results in the classification of three states of development (CA, HYD and VIT/AT) for species MC and TL using nominal classifiers [15]. Govocitos uses a 25-length color-texture feature vector with 10 grey level texture features and 15 chromatic features. Grey level texture descriptors model the spatial relationship of a pixel and its neighbors, providing information of the image structure such as smoothness and regularity, among others. Specifically, we used the Local Binary Patterns [28], taking the uniform patterns with radius  $R = 1$  and 8 neighbors. The chromatic features provide information about the distribution of the levels on each RGB channel, including the mean, variance, third and fourth statistical moments and entropy. The input patterns are preprocessed to have zero-mean and standard deviation one before being fed to the classifier (the mean and deviation values are calculated using only the training set).

### 3.1 Validation Methodology

The data include patterns from three species. The first one is *Merluccius merluccius*, with 1022 patterns with 3 states of development (classes): CA (25.3 % of the total patterns),

<sup>2</sup> <http://www.leica.com>.

HYD (6.0 %) and VIT/AT (68.7 %). We also have 912 patterns of specie *Trisopterus luscus* with the same 3 states: CA (57.6 %), HYD (1.5 %) and VIT/AT (40.9 %). We use the same experimental methodology for both species: the data are divided in equal-sized training and validation sets (with 511 and 456 training and validation patterns for *Merluccius* and *Trisopterus* respectively), and for each classifier we selected the values of the tunable parameters with the lowest mean absolute error (MAE, see Sect. 3.3) on the validation set. Using these best values, the test stage uses 5-fold cross validation (the available data are divided in 5 folds, averaging the performance over the 5 trials: in each trial, 4 folds are used for training and the remaining one for test). The third species is *Reinhardtius hippoglossoides*, with 16 images (one image per individual, see Fig. 1) with 7915 cells and 6 states: PG (37.6 %), CA (18.8 %), VIT1 (20.8 %), VIT2 (11.6 %), VIT3 (8.0 %) and VIT4 (3.2 %). We performed first a leave-one-image-out (LOIO) validation, with 16 trials, one for each image. In each trial, a training and a validation set (with 600 and 300 patterns respectively) are created selecting 100 training and 50 validation patterns, respectively, of each state from 15 images, excluding one image which is used for test (this image is different in each trial). The values of the tunable parameters are selected to minimize the MAE on the validation set. Using this best value, the classifier is trained using the training and validation sets (900 patterns), and tested using the excluded image (whose number of patterns is different for each trial). The performance is averaged over the 16 trials. Additionally, we developed a second experiment with mixed images (MIX) randomly selecting 10 trios of training (600 patterns), validation (300 patterns) and test sets (600 patterns), where 100, 50 and 50 patterns respectively belong to each of the six states. The patterns are extracted from all the 16 images. The training and validation sets are used for parameter tuning as in LOIO, and the test performance is averaged over the 10 test sets. In both methodologies LOIO and MIX, the training, validation and test data are the same for all the classifiers. The whole data set is publically available.<sup>3</sup>

### 3.2 Tested Methods

Eleven ordinal approaches, described in the Sect. 2, are tested: the linear method POM; different methods based on SVM: SVORIM, SVMOD, REDSVM and EPSVM; one method based on Discriminant Analysis: KDLOR; two methods based on Artificial Neural Networks concepts: ELMOR and ONN; and two ensemble models: ORBoost and ORBoostP. Additionally, we also compare to the well-known technique Support Vector Regression (SVR) in order to analyze whether a pure regression perspective could be suitable. All the SVM-based ordinal or nominal methods use LibSVM [6], tuning the regularization parameter  $C$  and the inverse  $\gamma$  of the kernel spread with values in  $\{2^i\}_{-5}^{14}$  and  $\{2^i\}_{-16}^8$  respectively. We use the sigmoidal activation function for ELMOR, ONN, ORBoost and ORBoostP, tuning the number of nodes in the hidden layer with values  $\{5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ . For ONN and ORBoost, the number of hidden neurons was adjusted in the range  $\{5, 10, 15, 20, 30, 40\}$ . As proposed by [25], the number of ensemble members in ORBoost was 25 and 2000 for ORBoostP. The range of  $\epsilon$  for  $\epsilon$ -SVR was  $\{10^i\}_0^3$ . These ordinal methods are compared to the following thirteen nominal classifiers, selected among the classifiers which exhibited the best behavior in the comprehensive study developed in [10]:

1. ABR Adaboost.M1 ensemble of classification trees [12] implemented in the R language<sup>4</sup> (boosting function in the `adabag` package [1]).

<sup>3</sup> [https://wiki.citius.usc.es/datasets/fish\\_ovary](https://wiki.citius.usc.es/datasets/fish_ovary).

<sup>4</sup> <http://www.r-project.org>.

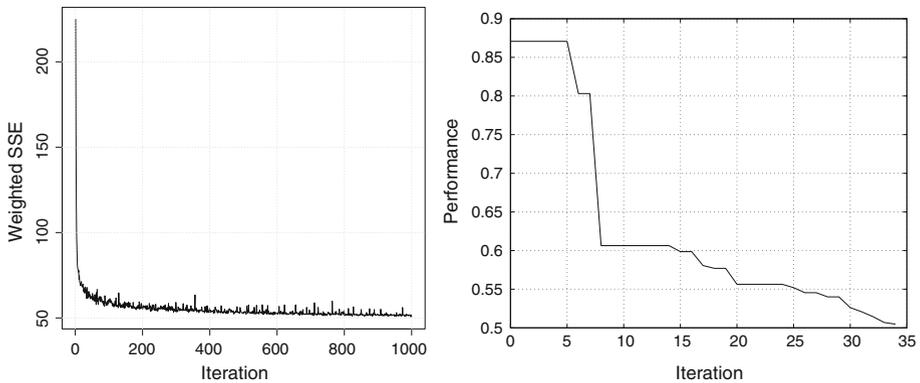
2. *ABW* AdaBoostM1 ensemble of decision stump classifiers (one-node decision trees) implemented in Weka [17]. The percentage of weight mass for base training is tuned with values 25, 50, 75 and 100 %.
3. *AvNN* ensemble of five MLP neural networks, from the *Caret*<sup>5</sup> package in R, trained with different random weight initializations, tuning the number of hidden neurons and the learning decay with values {1, 3, 5} and {0, 0.1,  $10^{-4}$ } respectively, values recommended in the *caret* documentation.
4. *BAG* Bagging ensemble of decision trees [2] (*bagging* function, *ipred* package in R).
5. *ELM* Extreme learning machine, using publically available Matlab code.<sup>6</sup> Following the recommendations found in the literature [18,35] and in the software documentation, we selected the best activation function among sine, sign, hardlimit, triangular, radial basis and sigmoid functions, and the best number of hidden neurons using 33 values between 3 and 5000, and we scaled the inputs between  $-1$  and  $+1$ .
6. *GELM* Gaussian kernel extreme learning machine using Matlab code from the same previous location. The parameters  $C$  and  $\gamma$  are tuned similarly to LibSVM (see above).
7. *GSVM* Support vector machine (SVM) with Gaussian kernel, implemented using LibSVM as the remaining SVM-based methods (see above).
8. *LBR* LogitBoost ensemble of decision stumps [13] (*LogitBoost* function, *caTools* package in R) with 200 Boosting iterations.
9. *LBW* LogitBoost ensemble of decision stumps implemented in Weka [17], with the 100% of weight mass to train, five runs for internal cross-validation, shrinkage parameter  $H = 1$  and 10 iterations.
10. *LDA* The classical Linear Discriminant Analysis [31] (*lda* function, *MASS* package in R).
11. *MLP-T* Multi-Layer perceptron neural network, tuning the number of hidden neurons with values {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}, implemented in R (*m1p* function, *RSNNS* package) and accessed via the *Caret* package) with the default parameters (standard Backpropagation, 100 training epochs, topological order for weight update and learning rate between 0.1 and 1).
12. *MLP-R* the same MLP network from the *RNNS* package, but used directly from R without *Caret*: 1000 training epochs are used, tuning the number of hidden neurons as MLP, and selecting the training algorithm between 8 methods: Standard Backpropagation (BP), Batch BP, Chunk BP, Momentum BP, BP with weight decay, Resilient BP, Quickprop and Scaled Conjugate Gradient. The Fig. 2 (left panel) reports an example of Sum of Squared Errors versus training epoch for the MLP-R and species *Trisopterus luscus*.
13. *MLP-PSO* Multi-Layer perceptron trained using particle swarm optimization (PSO), using the Matlab PSO Research Toolbox<sup>7</sup> [14] and the Neural Network add-in for PSORT.<sup>8</sup> The number of particles is tuned in the set {5, 10, 50, 20, 25, 30} and the number of hidden neurons with the same values as the MLP. The Fig. 2 (right panel) reports an example of performance vs. the training epoch for the MLP-PSO and species *Trisopterus luscus*.

<sup>5</sup> <http://caret.r-forge.r-project.org>.

<sup>6</sup> <http://www.extreme-learning-machines.org>.

<sup>7</sup> [http://www.georgeevers.org/psor\\_research\\_toolbox.htm](http://www.georgeevers.org/psor_research_toolbox.htm).

<sup>8</sup> <http://www.mathworks.com/matlabcentral/fileexchange/29565-neural-network-add-in-for-psort>.



**Fig. 2** Evolution of the sum of squared errors (SSE) during training for the MLP-R (*left panel*) and evolution of the performance for the MLP-PSO for specie *Trisopterus luscus*

14. *MLRM* Multinomial logistic regression model [23], using the `Logistic` class in Weka, with unlimited iterations and log-likelihood ridge  $10^{-8}$ .
15. *RF* Random forest [3] ensemble of 500 trees (`randomForest` function in the R homonym package), tuning the parameter `mtry` with values {2, 4, 7, 9, 12, 14, 17, 19, 22, 25}.

Any parameter value or configuration setting not specified in the previous list has been set to its default value, specified by the classifier documentation.

### 3.3 Evaluation Metrics

The first measure used to evaluate the previous classifiers is the well-known *Classifier Accuracy* (*Acc*, in %), the percentage of agreements between the desired and real classes without considering the class ordering. The *Cohen Kappa* ( $\kappa$ , in %), is based on *Acc* but discarding the probability of success by chance [5]. We also used other metrics specially designed for ordinal classification [8, 16]. The *mean absolute error* (MAE) is defined by  $MAE = \frac{1}{N} \sum_{i=1}^N |r(y_i^*) - r(y_i)|$ , being  $y_i^*$  and  $y_i$  the predicted and the true class respectively for pattern  $i$ , and  $r(y)$  the rank of  $y$  (its position in the ordinal scale), being  $N$  the number of patterns. The MAE value ranges from 0 to  $K - 1$  (maximum deviation in the number of ranks between two labels). The last two metrics measure the correlation between predicted targets and true targets: the *Kendall tau rank correlation coefficient* ( $\tau$ ) measures the association between predicted and true class [21] as  $\tau = \left( \sum_{ij} c_{ij}^* c_{ij} \right) \left( \sum_{ij} c_{ij}^{*2} \sum_{ij} c_{ij}^2 \right)^{-1/2}$ , where  $i, j \in \{1, \dots, N\}$ ,  $c_{ij} = +1$  if  $y_i > y_j$  (in the ordinal scale), being  $c_{ij} = 0$  when  $y_i = y_j$ , and  $c_{ij} = -1$  when  $y_i < y_j$  (the same for  $c_{ij}^*$ , using  $y_i^*$  instead of  $y_i$ ). The  $\tau$  values range from  $-1$  (maximum disagreement between prediction and true label), to 0 (no correlation between them) and to 1 (maximum agreement). Finally, the *Spearman rank correlation coefficient* ( $\rho$ ) is the Pearson correlation coefficient between the ranked predicted and true class [32], taking values in  $[-1, 1]$  with the same significance as  $\tau$ .

### 3.4 Results and Discussion

Table 1 reports the selected values for each classifier and experiment. For those classifiers with hidden neurons (ELMOR, ONN, avNNet, ELM, MLP-T, MLP-R and MLP-PSO), this

**Table 1** Selected values for those classifiers with tunable parameters, for each specie and experiment

Classifier	Parameter	Merluccius	Trisopterus	Reinhardtius	
				LOIO	MIX
KDLOR	$\gamma$	$2^{-7}$	$2^{-7}$	$2^3$	$2^3$
SVORIM	$C$	$2^6$	$2^7$	$2^8$	$2^9$
SVMOP	$\gamma$	$2^{-4}$	$2^{-7}$	$2^{-7}$	$2^{-7}$
	$C$	$2^8$	$2^{10}$	$2^{11}$	$2^{11}$
ELMOR	$\gamma$	$2^{-4}$	$2^{-7}$	$2^{-7}$	$2^{-7}$
	#hid	80	100	70	70
EPSVM	$C$	$2^6$	$2^3$	$2^6$	$2^7$
REDSVM	$\gamma$	$2^{-5}$	$2^{-4}$	$2^{-7}$	$2^{-7}$
	$C$	$2^8$	$2^{10}$	$2^8$	$2^{11}$
ONN	$\gamma$	$2^{-4}$	$2^{-8}$	$2^{-7}$	$2^{-6}$
	#hid	90	10	10	5
SVR	$C$	$2^3$	$2^8$	$2^{10}$	$2^6$
ELM	$\gamma$	$2^{-10}$	$2^{-10}$	$2^{-7}$	$2^{-10}$
	$\xi$	10	100	1000	100
	activ.funct.	triang.	sigmoid	sign	sine
GELM	#hid	170	110	110	116
	$C$	$2^{12}$	$2^{11}$	$2^{11}$	$2^{11}$
GSVM	$\gamma$	8	8	7.8	8.4
	$C$	$2^{13}$	$2^9$	$2^{12}$	$2^{11}$
MLP-T	$\gamma$	$2^{-7}$	$2^{-8}$	$2^{-6.9}$	$2^{-7.2}$
	#hid	15	11	15.4	14
MLP-R	train func.	Resil. BP	Std. BP	Resil. BP (6)	Chunk. BP (4)
MLP-PSO	#hid	7	11	8.1	12.4
	#part	5	5	6.2	7.0
RF	#hid	5	3	4.2	4.0
	mtry	2	8	6.0	6.0

For species Reinhardtius and LOIO and MIX experiments, the average values of the selected values are reported (the most selected train function and its number of trials for classifier MLP-R). The parameter  $P$  of ABW did not influenced the results. The AvNNNet used decay = 0.1 and 5 hidden neurons in all the experiments

value is relatively stable among classifiers in the same family: 70–100 for ELMOR and 110 for ELM; 5 for AvNNNet because it is an ensemble of MLPs; more variable for ONN, between 5 and 90; 11–15 for MLP-T, 7–12 for MLP-R and 3–5 for MLP-PSO. The classifiers with Gaussian kernels also exhibit similar values of spread  $\gamma$ : about  $2^{-7}$  for SVORIM, SVMOP, EPSVM, REDSVM, SVR and GSVM; about 8 for GELM; and very variable in KDLOR ( $2^{-7}$ -8). The parameter  $C$  also varies between  $2^6$  and  $2^{13}$  for SVORIM, SVMOP, EPSVM, REDSVM, SVR, GELM and GSVM. Finally, the mtry values of RF are also relatively stable in the range 2–8. Tables 2 and 3 report the results in terms of classification accuracy (Acc, %),  $\kappa$ , MAE, Kendall  $\tau$  and Spearman  $\rho$  rank correlation coefficients for species MC and TL (Table 2) and for specie RH with LOIO and MIX methodologies (Table 3). The classifiers

**Table 2** Classification results: accuracy and Cohen  $\kappa$  (both in %), MAE, Kendall  $\tau$  and Spearman  $\rho$  for species *Merluccius merluccius* and *Trisopterus luscus* with 3 states (CA, HYD, VIT/AT)

Classifier	<i>Merluccius merluccius</i>					<i>Trisopterus luscus</i>				
	Acc.	$\kappa$	MAE	$\tau$	$\rho$	Acc.	$\kappa$	MAE	$\tau$	$\rho$
POM	87.8	73.1	0.164	0.800	0.834	92.0	83.7	0.140	0.856	0.865
KDLOR	85.4	71.1	0.160	0.832	0.870	84.8	72.9	0.166	0.854	0.894
SVORIM	89.5	78.6	0.132	0.839	0.874	94.7	89.1	0.095	0.902	0.908
SVMOD	<b>94.1</b>	<b>86.9</b>	<b>0.110</b>	<b>0.856</b>	0.858	94.9	89.7	0.097	0.900	0.901
ELMOR	93.2	85.4	0.116	0.848	0.859	91.5	83.0	0.152	0.842	0.850
EPSVM	89.0	74.0	0.161	0.777	0.828	92.3	84.2	0.143	0.855	0.858
REDSVM	89.1	77.0	0.140	0.834	0.861	93.8	87.4	0.112	0.886	0.890
ONN	86.8	73.6	0.162	0.810	0.853	91.5	83.8	0.112	0.889	0.911
ORBoost	90.1	79.0	0.124	<i>0.854</i>	<b>0.882</b>	93.1	86.5	0.096	0.905	0.922
ORBoostP	90.2	78.9	0.125	0.851	<i>0.878</i>	93.3	86.7	0.098	0.902	<b>0.917</b>
SVR	84.8	69.7	0.186	0.792	0.818	84.2	71.4	0.190	0.832	0.862
Mean	89.1	77.0	<b>0.144</b>	<b>0.827</b>	<b>0.856</b>	<b>91.5</b>	<b>83.5</b>	<b>0.127</b>	<b>0.875</b>	<b>0.889</b>
ABR	93.1	85.1	0.130	0.832	0.835	94.5	88.8	0.109	0.887	0.888
ABW	81.4	59.1	0.309	0.671	0.688	79.8	56.6	0.388	0.610	0.615
AvNN	93.6	86.1	<i>0.115</i>	0.851	0.854	<b>95.6</b>	<b>91.1</b>	<i>0.088</i>	<i>0.909</i>	0.910
BAG	91.8	81.8	0.148	0.806	0.811	91.0	81.5	0.179	0.815	0.815
ELM	93.5	85.8	0.119	0.845	0.849	94.6	89.0	0.105	0.892	0.893
GELM	<i>93.8</i>	86.4	0.119	0.845	0.847	95.2	90.1	0.093	0.904	0.906
GSVM	92.8	84.2	0.134	0.823	0.826	95.5	<i>90.9</i>	<b>0.087</b>	<b>0.911</b>	<i>0.912</i>
LBR	91.4	82.3	0.164	0.790	0.795	92.6	87.1	0.148	0.830	0.835
LBW	91.1	80.7	0.162	0.789	0.794	90.3	80.3	0.188	0.805	0.807
LDA	92.5	83.8	0.128	0.829	0.835	93.1	85.9	0.134	0.861	0.863
MLP-T	<i>93.8</i>	<i>86.5</i>	0.129	0.834	0.837	<b>95.6</b>	<b>91.1</b>	0.088	<i>0.909</i>	0.910
MLP-R	91.6	81.6	0.156	0.800	0.804	94.3	88.4	0.113	0.883	0.883
MLP-PSO	72.5	48.2	0.308	0.688	0.722	68.2	47.3	0.395	0.633	0.668
MLRM	93.3	85.4	0.127	0.835	0.837	93.4	86.8	0.127	0.869	0.871
RF	93.1	84.9	0.128	0.831	0.835	93.5	86.7	0.133	0.862	0.862
Mean	<b>90.6</b>	<b>80.1</b>	0.158	0.805	0.811	91.1	82.8	0.158	0.839	0.843

Ordinal (resp. nominal) classifiers are in the upper (resp. lower) half of the table. The best and second best results are in bold and italics respectively

are divided into ordinal (upper part) and nominal (lower part). The highest Acc,  $\kappa$ ,  $\tau$  and  $\rho$ , and the lowest MAE, are highlighted for each specie and experiment (the second best value is italicized). The mean values for the whole set of ordinal and nominal methods are also included. From the application point of view, almost all of the results are very promising both in Acc and MAE: for species MC and TL we achieve Acc = 94.1 %, MAE = 0.110 and Acc = 95.6 %, MAE = 0.087 respectively (this MAE value means that each state is misclassified with neighbor states less than 10 %). For specie RH the best results are slightly worse in terms of Acc (67.8 and 80.4 % for LOIO and MIX experiments),  $\kappa$  and MAE, but very similar for the  $\tau$  and  $\rho$  correlation coefficients, where indeed the results are outstanding. This

**Table 3** Classification results for the specie *Reinhardtius hippoglossoides* with 6 states (PG, CA, VIT1, VIT2, VIT3, VIT4) using the Leave-one-image-out (LOIO) and mixed images (MIX) methodologies

Classifier	Leave-one-image-out					Mixed images				
	Acc.	$\kappa$	MAE	$\tau$	$\rho$	Acc.	$\kappa$	MAE	$\tau$	$\rho$
POM	63.9	49.5	0.394	0.770	0.851	70.5	64.6	0.310	0.879	0.940
KDLOR	66.7	52.9	0.360	0.789	0.856	78.9	74.7	0.222	0.912	0.957
SVORIM	67.2	53.5	0.354	0.788	0.858	79.0	74.8	0.222	0.911	0.957
SVMOD	<b>67.8</b>	<b>54.6</b>	<b>0.352</b>	0.796	<b>0.861</b>	80.1	76.2	0.215	0.912	0.957
ELMOR	62.7	49.0	0.418	0.762	0.831	77.5	73.0	0.247	0.898	0.948
EPSVM	66.9	53.5	0.358	<b>0.798</b>	0.860	75.6	70.7	0.259	0.897	0.950
REDSVM	67.2	53.4	0.362	0.789	0.856	78.7	74.4	0.225	0.910	0.957
ONN	63.7	49.4	0.393	0.768	0.851	74.3	69.1	0.276	0.890	0.945
ORBoost	64.5	50.2	0.376	0.772	0.849	77.6	73.1	0.232	0.909	0.957
ORBoostP	64.4	50.2	0.378	0.772	0.848	77.5	73.0	0.233	0.909	0.956
SVR	66.1	52.3	0.363	0.785	0.855	79.1	74.9	0.218	<b>0.914</b>	<b>0.959</b>
Mean	<b>65.6</b>	<b>51.7</b>	<b>0.373</b>	<b>0.781</b>	<b>0.852</b>	<b>77.2</b>	<b>72.6</b>	<b>0.242</b>	<b>0.904</b>	<b>0.953</b>
ABR	51.8	34.4	0.597	0.748	0.799	77.8	73.3	0.255	0.891	0.941
ABW	30.8	10.5	0.895	0.572	0.619	32.8	19.4	0.771	0.710	0.805
AvNN	53.0	37.8	0.626	0.772	0.822	79.3	75.2	0.229	0.905	0.952
BAG	50.6	35.0	0.675	0.714	0.771	70.7	64.8	0.382	0.824	0.889
ELM	64.4	50.2	0.424	0.750	0.808	78.3	73.9	0.249	0.894	0.942
GELM	65.6	51.7	0.400	0.766	0.821	80.1	76.1	0.221	0.908	0.953
GSVM	67.4	53.7	0.363	0.790	0.647	<b>80.4</b>	<b>76.5</b>	<b>0.212</b>	0.913	0.785
LBR	45.7	33.1	0.801	0.600	0.663	71.6	70.9	0.433	0.801	0.865
LBW	56.3	40.6	0.592	0.679	0.737	72.3	66.8	0.338	0.851	0.912
LDA	53.5	39.0	0.578	0.785	0.839	77.5	72.9	0.248	0.897	0.948
MLP-T	52.3	36.0	0.599	0.768	0.824	79.0	74.9	0.234	0.903	0.950
MLP-R	51.4	35.1	0.613	0.769	0.827	76.8	72.2	0.265	0.888	0.940
MLP-PSO	32.0	16.6	0.907	0.516	0.582	36.7	24.1	0.757	0.700	0.816
MLRM	66.1	53.0	0.372	0.794	0.848	78.3	73.9	0.237	0.903	0.951
RF	52.6	36.8	0.639	0.748	0.802	78.2	73.8	0.254	0.891	0.940
Mean	52.9	37.6	0.605	0.718	0.761	71.3	65.9	0.349	0.859	0.906

result could indicate that the states are properly ordered from a purely ranking perspective, but the predictions might be displaced one or two values (recall that the correlation coefficients  $\tau$  and  $\rho$  consider for example whether a pattern belonging to class  $C_2$  is ranked higher than a pattern belonging to class  $C_1$  but it does not consider whether they actually are included in this two states in the prediction, i.e. the pattern belonging to  $C_2$  could be predicted to belong to  $C_3$  and the one belonging to  $C_1$  to  $C_2$ ). The average MIX accuracy and  $\kappa$  are 11 and 21 points, respectively, above their LOIO counterparts (see the *Mean* row), because in the MIX experiments the oocytes of each image may be selected for the training or test set. Therefore, implicit information about the sample acquisition and processing is included in the training and test sets, which justifies better results compared to LOIO experiments (more realistic

from the application point of view), where any information about an individual fish in the testing set is not included in the training set.

From the classifiers perspective, Tables 2 and 3 identify some outstanding methods: SVMOD (ordinal, the best for species MC and RH-LOIO and very near to the best in RH-MIX) and GSVM (nominal, very near to the best for specie TL and the best for RH in MIX experiments). Furthermore, the measures MAE,  $\tau$  and  $\rho$  show the real difference between ordinal and nominal more clearly than Acc and  $\kappa$ . A general conclusion from this table is that, although any ordinal or nominal classifier has chance of obtaining the best result, in general, ordinal classifiers perform better in mean in terms of measures (MAE,  $\tau$ ,  $\rho$ ) that consider the ordinal nature of the dataset, sometimes at the expense of lower Acc values. The *Mean* rows show that for specie MC (3 states), the average Acc and  $\kappa$  are better for nominal classifiers (which do not consider the state order), although for specie TL the ordinal classifiers are better, while MAE,  $\tau$  and  $\rho$  are better for ordinal classifiers and both species. However, in species RH (6 states) the five measures are better for the ordinal classifiers. This suggests that the superiority of ordinal with respect to nominal classifiers increases with the number of states, being this superiority not so clear (at least, for non-ordinal quality measures) with fewer states. In fact, with 3 states is not probable for a nominal classifier to assign a pattern to a non-neighbor state, because 4 of 6 possible errors (corresponding to non-diagonal elements in the 3-order square confusion matrix) respect the state ordering. However, in specie RH (6 states) the probability of non-ordinal errors is biggest (20 of 30 possible errors involve non-neighbor states), enhancing the difference between ordinal and nominal classifiers.

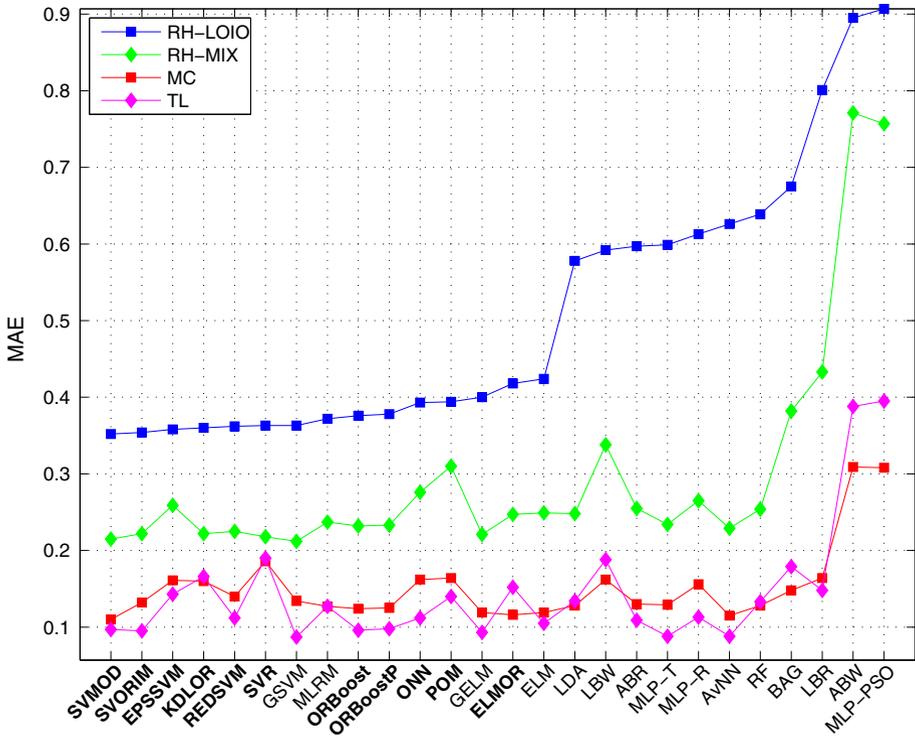
Table 4 reports the ranking results obtained for each method and metric (averaged over all experiments) ordered by increasing MAE: the SVMOD performs properly for all the metrics, obtaining promising mean ranking values, as well as SVORIM, which also obtains competitive results in terms of MAE,  $\tau$  and  $\rho$ , while the nominal method GSVM obtains good ranking results for Acc,  $\kappa$  and MAE. Since GSVM uses the one-vs-one paradigm, which is not designed to specifically minimise the ordinal errors, it presents a good general performance, but in terms of ordinal metrics it is generally worse. The last column of Table 4 shows the  $p$  value of the  $T$  test comparing the mean MAE value obtained by SVMOD to each one of the other classifiers: SVMOD is significantly better than 7 classifiers (REDSVM, MLRM, EPSVM, ONN, POM, ABW and MLP-PSO, whose  $p$ -value is in bold), with high  $p$ -values (lower difference) for AvNN, MLP-T, MLP-R, GELM, SVORIM, ABR, RF and LDA. The MLP-T works slightly better than the MLP-R, showing that the more sophisticated setup of MLP-R (tuning of training algorithm and more epochs than MLP-T, accessed via Caret) does not achieve better performance. The MLP-PSO achieves the worst results, being not competitive for any specie, number of classes and experiment type.

Figure 3 plots the mean MAE values of the classifiers for each experiment ordered by increasing MAE in RH-LOIO. Considering the LOIO plot (blue), the best accuracies are achieved by the ordinal classifiers (names in bold), which occupy most of the places in the left half of the horizontal axis. The nominal classifiers are in the right part of the axis, except GSVM and MLRM (7th and 8th positions respectively). There are clearly two groups: classifiers SVMOD to ELM (MAE about 0.35–0.40), and the remaining ones (LDA and following, MAE about 0.6 and higher). In the MIX experiments (green) many ordinal and nominal classifiers are below 0.25, and just a few are sub-optimal (EPSVM, ONN, POM, LBW, BAG and LBR). In specie MC (red) SVMOD is the best, followed by GELM, ELMOR, ELM and AvNNNet, while SVR, ONN, POM, LBW, BAG and LBR achieve bad results. In specie TL (magenta) the bests are GSVM, avNNNet, MLP, GELM, SVORIM and SVMOD. The MLP-T is slightly better than MLP-R in all the species and experiments, being MLP-PSO the worst one except for RH-MIX.

**Table 4** Ranking results averaged over all the experiments (ordinal classifiers are in bold), ordered by increasing MAE ranking, and  $p$ -values of the  $T$  test comparing the best classifier (SVMOD) and the remaining ones (significant differences for  $p < 0.05$  are in bold face)

Position	Classifier	Acc.	$\kappa$	MAE	$\tau$	$\rho$	$p$ -value
1	<b>SVMOD</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	4.5	–
2	GSVM	2	2	2	4.5	23	0.519
3	<b>SVORIM</b>	6	6	3	2	3	0.261
4	<b>ORBoost</b>	9	11	4	3	<b>1</b>	0.083
5	GELM	3	3	5	7	9	0.287
6	<b>ORBoostP</b>	8	10	6	6	2	0.064
7	<b>REDSVM</b>	7	7	7	8	6	<b>0.041</b>
8	MLRM	4	4	8	10	11	<b>0.004</b>
9	ELM	5	5	9	12	14	0.133
10	<b>KDLOR</b>	18	20	10	11	4.5	0.120
11	<b>EPSVM</b>	11	13	11	19	12	<b>0.038</b>
12	<b>ELMOR</b>	10	8	12	16	16	0.058
13	<b>ONN</b>	17	18	13	15	7	<b>0.024</b>
14	<b>SVR</b>	20	23	14	20	13	0.137
15	<b>POM</b>	19	21	15	21	15	<b>0.018</b>
16	MLP-T	13	12	16	9	10	0.331
17	AvNN	12	9	17	4.5	8	0.372
18	LDA	16	15	18	13	17	0.210
19	ABR	15	16	19	14	18	0.249
20	MLP-R	21	17	20	17	19	0.197
21	RF	14	14	21	18	20	0.235
22	LBW	22	22	22	23	22	0.052
23	BAG	23	24	23	22	21	0.094
24	LBR	24	19	24	24	24	0.132
25	ABW	25	25	25	25	26	<b>0.022</b>
26	MLP-PSO	26	26	26	26	25	<b>0.021</b>

Table 5 reports the average confusion matrix, sensitivities (Se) and positive predictivities (PP) achieved by SVMOD and GSVM on specie RH with LOIO experiments (matrices for species MC and TL are not reported due to their low numbers of states). In both matrices the diagonal values are the highest in each row and column, and only the PP of state CA is below 50 %, due to the overlap between states PG and CA (the largest non-diagonal values correspond to these two neighbor states). The only high non-diagonal values are adjacent to the diagonal, corresponding to patterns assigned to a state neighbor to the right one. Comparing SVMOD and GSVM, the latter achieves higher values outside the diagonal, excepting the (PG,CA) and (PG,VIT1) values, learning worse the ordinal information (remember from Table 3 that SVMOD wins GSVM with specie RH and LOIO experiments not only in Acc. and  $\kappa$  but also in MAE,  $\kappa$  and  $\rho$ ). Besides, GSVM achieves lower Se and PP for all the states excepting PG, because it assigns more patterns CA to PG than SVMOD. In fact, the sensitivity of SVMOD is above 60 % for all the states, while GSVM is below 55 % for states CA and VIT3. Regarding PP, the SVMOD wins GSVM in all the states except CA, with high difference in states VIT1-VIT3.



**Fig. 3** MAE values achieved by each classifier (ordinal classifiers are in *bold*) for species RH (LOIO and MIX experiments), MC and TL, ordered by increasing the MAE for RH-LOIO

**Table 5** Confusion matrices and sensitivities/positive predictivities for each state (in %) achieved by SVMOD (upper) and GSVM (lower) for specie RH and LOIO experiments

SVMOD	PG	CA	VIT1	VIT2	VIT3	VIT4	Se (%)	PP (%)
PG	<b>22.29</b>	10.15	0.94	0.37	0.03	0.00	66.0	80.6
CA	5.00	<b>11.32</b>	2.29	0.09	0.01	0.00	60.5	45.3
VIT1	0.37	3.47	<b>14.23</b>	2.12	0.05	0.10	70.0	74.0
VIT2	0.00	0.04	1.68	<b>7.16</b>	0.92	0.15	71.9	68.3
VIT3	0.00	0.01	0.08	0.60	<b>6.00</b>	2.14	68.0	71.0
VIT4	0.00	0.00	0.02	0.14	1.44	<b>6.78</b>	80.9	74.0
GSVM	PG	CA	VIT1	VIT2	VIT3	VIT4	Se (%)	PP (%)
PG	<b>24.51</b>	7.66	1.17	0.29	0.14	0.00	72.5	80.3
CA	5.41	<b>10.36</b>	2.79	0.15	0.01	0.00	55.4	48.4
VIT1	0.58	3.35	<b>14.29</b>	2.02	0.03	0.08	70.2	71.6
VIT2	0.01	0.04	1.63	<b>7.00</b>	1.05	0.22	70.3	64.3
VIT3	0.00	0.00	0.05	1.23	<b>4.68</b>	2.88	53.0	62.5
VIT4	0.00	0.00	0.03	0.19	1.57	<b>6.59</b>	78.6	67.5

## 4 Conclusions

This paper uses 11 ordinal and 15 nominal approaches to classify states of development of fish oocytes from histological microscopy images. Twenty-five features are extracted from every oocyte, including 10 grey level texture (Local Binary Patterns) and 15 statistical color features. Three fish species are considered: *Merluccius merluccius* and *Trisopterus luscus*, which present 3 states of biological interest, and *Reinhardtius hippoglossoides*, with 6 states with and without leaving one image out. The experiments demonstrate that ordinal classifiers exhibit improved robustness and performance compared to nominal methods for all the species considered: SVMOD achieves accuracies about 94 and 95 % for species MC and TL and 67–80 % for specie RH with and without leave one image out respectively. Several standard nominal techniques can also obtain promising results for some cases (GSVM for specie RH, without leave one image out, and AvNNet for specie TL). However, SVMOD has the best Friedman rank for all the five measures considered (Accuracy, Cohen  $\kappa$ , Mean Averaged Error, Kendall  $\tau$  and Spearman  $\rho$ ), and SVORIM is the second for the last three measures, which consider the ordinal nature of the classification problem (although GSVM is the second for Acc,  $\kappa$  and MAE). The difference between ordinal and nominal techniques has been shown to be higher when the number of states increases, being clearly reflected by ordinal quality measures (Kendall  $\tau$  and Spearman  $\rho$ ). The confusion matrix of SVMOD shows that the ordinal classifiers locate their errors in states near to the true ones, with sensitivities and positive predictions above 60 % for almost all the states. On the whole, it can be said that ordinal regression techniques should be preferred to regression and multinomial classification methods when dealing with datasets that present an ordinal nature. This also motivates the improvement of the current techniques in the ordinal classification literature, which, given the novelty of the topic, are still in constant development.

**Acknowledgments** We acknowledge support from the “Junta de Andalucía” under project P11-TIC-7508, from the Spanish Ministry of Science and Innovation (MICINN) under projects TIN2011-22935, TIN2012-32262 and TIN2014-54583-C2-1-R, and from FEDER funds

## References

1. Alfaro E, Gámez M, García N (2007) Multiclass corporate failure prediction by adaboost.m1. *Int Adv Econ Res* 13:301–312
2. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
3. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
4. Cardoso JS, da Costa JFP (2007) Learning to classify ordinal data: the data replication method. *J Mach Learn Res* 8:1393–1429
5. Carletta J (1996) Assessing agreement on classification tasks: the kappa statistic. *Comput Linguist* 22(2):249–254
6. Chang C, Lin C (2008) LibSVM: a library for Support Vector Machines. URL: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
7. Chu W, Keerthi SS (2007) Support vector ordinal regression. *Neural Comput* 19:792–815
8. Cruz-Ramírez M, Hervás-Martínez C, Sánchez-Monedero J, Gutiérrez P (2014) Metrics to guide a multi-objective evolutionary algorithm for ordinal classification. *Neurocomputing* 135:21–31
9. Deng WY, Zheng QH, Lian S, Chen L, Wang X (2010) Ordinal extreme learning machine. *Neurocomputing* 74(1–3):447–456
10. Fernández-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 15:3133–3181
11. Frank E, Hall M (2001) A simple approach to ordinal classification. In: *Proceedings of 12th European Conference on Machine Learning*, pp 145–156

12. Freund Y, Schapire R (1996) Experiments with a new Boosting algorithm. In: International Conference on Machine Learning. Morgan Kaufmann, pp 148–156
13. Friedman J, Hastie T, Tibshirani R (1998) Additive logistic regression: a statistical view of boosting. *Ann Stat* 28:2000
14. Evers G (2009) An automatic regrouping mechanism to deal with stagnation in particle swarm optimization. Ph.D. thesis, The University of Texas-PanAmerican, Edinburg, TX
15. González-Rufino E, Carrión P, Cernadas E, Fernández-Delgado M, Domínguez-Petit R (2013) Exhaustive comparison of colour texture features and classification methods to discriminate cells categories in histological images of fish ovary. *Pattern Recogn* 46:2391–2407
16. Gutiérrez P, Pérez-Ortiz M, Fernandez-Navarro F, Sánchez-Monedero J, Hervás-Martínez C (2012) An experimental study of different ordinal regression methods and measures. In: 7th International Conference on Hybrid Artificial Intelligence Systems (HAIS), Lecture Notes in Computer Science, vol. 7209, pp 296–307
17. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The Weka data mining software: an update. *SIGKDD Explor News* 11(1):10–18
18. Huang G, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B Cybern* 42:513–529
19. Hunter JR, Macewicz BJ, Lo N, Kimbrell CA (1992) Fecundity, spawning and maturity of female Dover Sole, *Microstomus pacificus*, with an evaluation of assumptions and precision. *Fish Bull* 90:101–128
20. Junquera S, Román E, Morgan J, Sainza M, Ramilo G (2003) Time scale of ovarian maturation in Greenland halibut (*Reinhardtius hippoglossoides*, Walbaum). *ICES J Mar Sci* 60:767–773
21. Kendall M (1938) A new measure of rank correlation. *Biometrika* 30:81–89
22. Kennedy J, Gundersen A, Hoines A, Kjesbu O (2011) Greenland halibut (*Reinhardtius hippoglossoides*) spawn annually but successive cohorts of oocytes develop over 2 years, complicating correct assessment of maturity. *Can J Fish Aquat Sci* 68:201–209
23. Le Cessie S, Van Houwelingen J (1992) Ridge estimators in logistic regression. *Appl Stat* 41(1):191–201
24. Li L, Lin HT (2007) Ordinal regression by extended binary classification. *Adv Neural Inf Process Syst* 19:865–872
25. Lin HT, Li L (2006) Large-margin thresholded ensembles for ordinal regression: theory and practice. In: Balcázar J, Long P, Stephan F (eds) Algorithmic learning theory, vol 4264, Lecture Notes in Computer Science. Springer, Berlin, pp 319–333
26. McCullagh P, Nelder JA (1989) Generalized linear models. Monographs on statistics and applied probability, 2nd edn. Chapman & Hall/CRC, London
27. Morgan MJ, Bowering WR (1997) Temporal and geographic variation in maturity at length and age of Greenland halibut (*Reinhardtius hippoglossoides*) from the Canadian North-West Atlantic with implications for fisheries management. *ICES J Mar Sci* 54:875–885
28. Ojala T, Piatikäinen M, Mäenpää T (2002) Multiresolution grey-scale and rotation invariant texture classification with local binary pattern. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
29. Pérez-Ortiz M, Gutiérrez PA, Hervás-Martínez C (2014) Projection based ensemble learning for ordinal regression. *IEEE Trans Cybern* 44(5):681–694
30. Rideout RM, Maddock DM, Burton MPM (1999) Oogenesis and the spawning pattern in Greenland halibut from the north-west Atlantic. *J Fish Biol* 54:196–207
31. Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge
32. Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15: 72–101
33. Sun BY, Li J, Wu DD, Zhang XM, Li WB (2010) Kernel discriminant learning for ordinal regression. *IEEE Trans Knowl Data Eng* 22:906–910
34. Waegeman W, Boullart L (2009) An ensemble of weighted support vector machines for ordinal regression. *Int J Comput Syst Sci Eng* 3(1):1–7
35. Wang N (2015) Generalized single-hidden layer feedforward networks for regression problems. *IEEE Trans Neural Netw* 26:1161–1176