# Boosting Rare Benthic Macroinvertebrates Taxa Identification With One-Class Classification

Fahad Sohrab
*Faculty of Information Technology and Communication Sciences*
*Tampere University*
Tampere, Finland
fahad.sohrab@tuni.fi

Jenni Raitoharju
*Programme for Environmental Information*
*Finnish Environment Institute*
Jyväskylä, Finland
jenni.raitoharju@environment.fi

*Abstract*—**Insect monitoring is crucial for understanding the consequences of rapid ecological changes, but taxa identification currently requires tedious manual expert work and cannot be scaled-up efficiently. Deep convolutional neural networks (CNNs) provide a promising way to increase the biomonitoring volumes significantly. However, taxa abundances are typically very imbalanced, and the amounts of training images for the rarest classes are too low for deep CNNs. As a result, the samples from the rare classes are often completely missed, while detecting them has biological importance. On the other hand, one-class classifiers are traditionally trained with much fewer samples to model a single class of interest. In this paper, we examine their capability to complement deep CNN based taxa identification by indicating samples potentially belonging to the rare classes of interest for human inspection. Our experiments confirm that the proposed approach may indeed support moving towards partial automation of the taxa identification task.**

*Index Terms*—**biomonitoring, taxa identification, machine learning, one-class classification, support vector data description**

## I. INTRODUCTION

To understand the consequences of climate change and other anthropogenic changes in different aquatic ecosystems, it is crucial to widely monitor different animal groups. Also international environmental legislation, such as the EU Water Framework Directive (WFD) [1], acknowledges the task of monitoring aquatic ecosystems. Since changes in the abundances of benthic macroinvertebrate taxa can provide an early warning sign of environmental problems in aquatic ecosystems, they are widely used as indicating factors in WFD-compliant ecological status assessment and environmental decision making [2], [3]. Simultaneously, they have also been identified as one of the most challenging groups to be monitored [4]. The task currently requires tedious manual expert work making it expensive, time-consuming, and error-prone. The recent advances in machine learning, especially deep convolutional neural networks (CNNs), provide a promising way to scale-up monitoring and provide faster information for environmental decision making. In the future, the samples of benthic macroinvertebrates may be imaged with an automated imaging device and then identified using a deep learning model trained with a sample dataset.

The overall accuracy obtained by automatic identification of benthic macroinvertebrates is approaching human expert-level [5] and, already in the near future, it may be possible to use machines to handle the majority of the samples, while human experts manually identify only the difficult and interesting cases, such as specimens potentially belonging to rare species. A significant challenge that needs to be addressed is induced by the very imbalanced taxa abundances. For some rare species, the number of training images is simply too low for a deep CNN and, as a result, the identification often fails. This problem is largely overlooked in the recent works [5], [6] that consider only the overall identification accuracy. The low number of misclassified specimens from rare species hardly affects the overall accuracy, while they are important for monitoring biodiversity. In this paper, we propose a mechanism that can indicate a reasonably-sized subset of specimens as potential samples of rare species for human expert inspection. To this end, we propose complementing the deep CNN based identification with one-class classifiers. One-class classifiers are traditionally trained with much fewer samples than deep networks, and our experimental results support the assumption that they can help detect samples from the rare species.

The rest of this paper is organized as follows. In Section II, we discuss the related work on machine learning in biomonitoring and one-class classification. The proposed system is described in detail in Section III. The details of experimental setup along with the results are provided in detail in Section IV. Finally, the conclusion and future work are presented in Section V.

## II. RELATED WORK

### A. Machine learning in biomonitoring

Machine learning is rapidly gaining recognition as a promising tool for many biomonitoring applications, such as identifying fish species [7], forest surveillance [8], or monitoring Arctic flowering seasons [9]. In this paper, we concentrate on benthic macroinvertebrate identification. Nevertheless, primary challenges are similar for most biomonitoring applications, and the solutions may be easily applied to other applications. For example, the identification task is very fine-grained. Fine-grained classification aims at distinguishing subordinate categories of a common superior category [10]. Domain experts
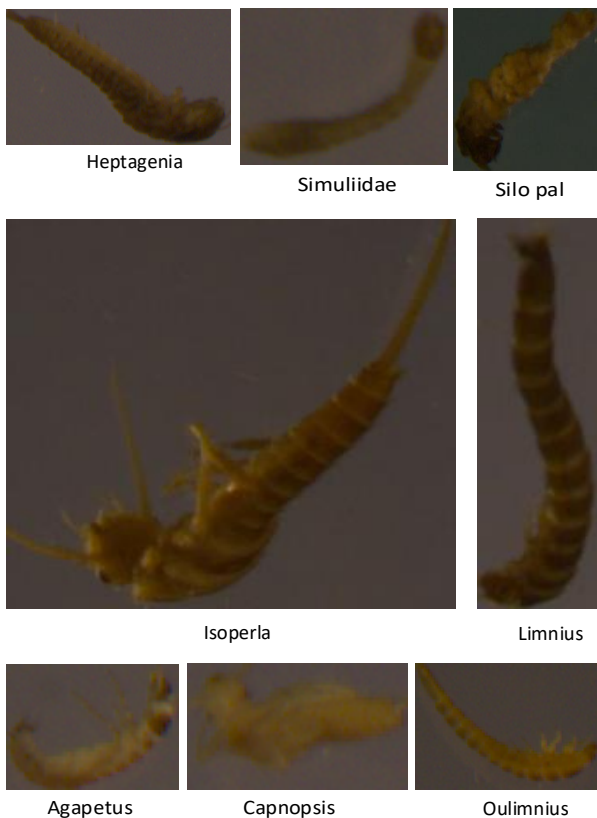
Fig. 1. Example images of different benthic macroinvertebrate taxa from FIN-Benthic 2 dataset [5]

usually define those subordinate categories with complicated rules, which typically focus on subtle differences in particular regions. For a non-expert, it may be hard to see any difference between similar species as illustrated by example images in Fig. 1. At the same time, the intra-class variance may be large due to different development stages [11]. Taxa distributions in the nature, and thus also the available reference datasets, are very imbalanced [5]. Furthermore, some taxonomists continue to object the shift toward automated methods due to different doubts and fears [12]. The last problem may be eased by providing better mechanisms for dividing the identification task between machines and human experts so that the machine first handles only the most routine-like cases [6].

The machine learning techniques for image-based taxa identification have developed from using handcrafted features with shallow networks [13], [14] towards using deep neural networks, which operate on images as inputs [5], [6]. A significant challenge with deep neural networks is the need of huge amounts of training data. This has lead to efforts to create imaging devices capable of providing high-quality images with minimal manual effort [11], [15]. Nevertheless, the existing datasets, such as FIN-Benthic2 [5] used in this paper, typically have very imbalanced classes. The smallest taxa simply do not provide enough information for training deep neural networks. However, such rare taxa and changes in their abundances may be biologically and environmentally

interesting. The performance of the deep neural networks for the very rare species may be enhanced, e.g., by data-augmentation [16] or special loss functions [17], but also these approaches tend to overfit to the few training samples and do not generalize well for unseen samples. In this paper, we suggest combining one-class classifiers with the trained deep neural network to provide an additional mechanism for detecting samples potentially belonging to the rare classes for human inspection.

### B. One-class classification

The main idea in one-class classification is to create a representative model of a class of interest, typically called target or positive class, using data from this class only. During inference, the model is used to predict whether unseen samples belong to the target class or are outliers. Traditional one-class classifiers can produce successful class models with only tens and hundreds of samples [18]–[20]. At the same time, computational complexity of the training phase makes training on very large or high-dimensional datasets infeasible. However, testing the trained one-class classifiers is fast, and very large datasets can be evaluated with a low computational cost. One-class classifiers have been applied, for example, on highly imbalanced phage-bacteria datasets for fast identification of potential phage candidates for a given bacteria [21]. In [19], one-class classification was applied to the facial image analysis problem. River target detection in remote sensing images was studied in [22], where the proposed one-class classification based system reduced the time complexity in target detection. A document classification system based on principal component analysis (PCA) and one-class classification was proposed in [23], where PCA helped achieve dimensionality reduction for one-class classification.

We denote the target data as $X = [x_1, ..., x_n]$, where $n$ is the number of target items and $x_i$ are $D$-dimensional vectors. One-Class Support Vector Machine (OC-SVM) [24] separates all the data points from the origin and maximizes the distance from this hyperplane to the origin:

$$\min_{w, \xi_i, \rho} \quad \frac{1}{2} \|w\|^2 + \frac{1}{Cn} \sum_{i=1}^{n} \xi_i - \rho$$
$$\text{s.t.} \quad w * x_i \geq \xi_i - \rho, \ \forall i \in \{1, \ldots, n\}$$
$$\xi_i \geq 0, \ \forall i \in \{1, \ldots, n\}, \quad (1)$$

where $w$ is a weight vector, slack variables $\xi_i$ allow some data points to lie within the margin, and hyper-parameter $C$ sets an upper bound on the fraction of training samples allowed within the margin and a lower bound on the number of training samples used as Support Vector.

Another classical one-class classification method is Support Vector Data Description (SVDD) [25]. An SVDD model is trained by forming the smallest hypersphere, which includes
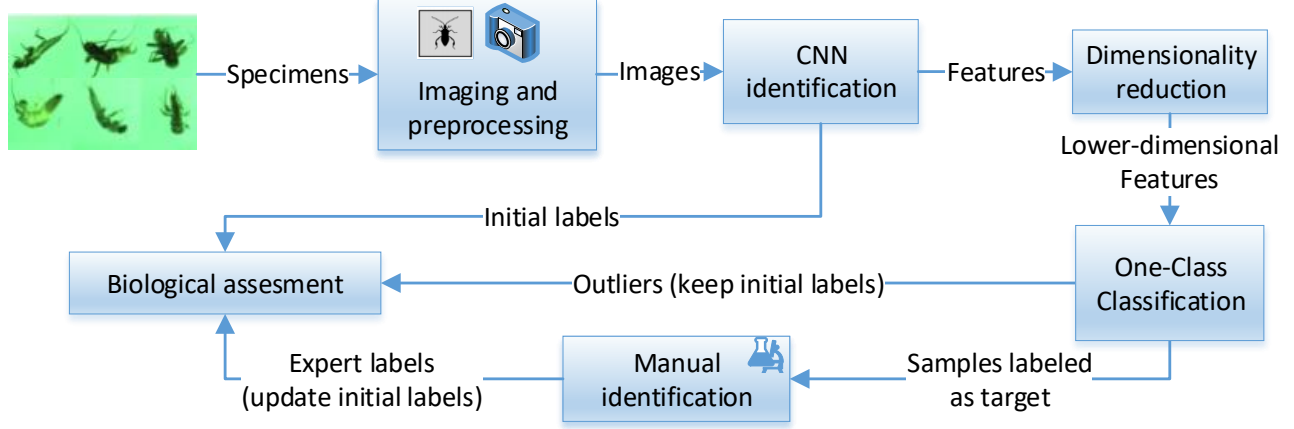
Fig. 2. The proposed taxa identification pipeline

all the target data. SVDD minimized the following function:

$$\min \quad F(R, \mathbf{a}) = R^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad \|\mathbf{x}_i - \mathbf{a}\|_2^2 \leq R^2 + \xi_i, \ \forall i \in \{1, \ldots, n\},$$

$$\xi_i \geq 0, \ \forall i \in \{1, \ldots, n\}, \tag{2}$$

where $R$ is the radius, $\mathbf{a}$ is the center of hypersphere, $\xi_i$ are slack variables allowing some training samples to be left outside the hypersphere, and hyper-parameter $C$ controls the number of allowed outliers. Both OC-SVM and SVDD can be solved in one step using Lagrange multipliers.

A recent extension of SVDD, Subspace Support Vector Data Description (S-SVDD) [26] maps the data to an optimised $d$-dimensional subspace suitable for one-class classification as $\mathbf{Qx}_i$. S-SVDD is solved iteratively, alternating the steps of solving SVDD in the current subspace and improving the subspace projection Q. The second step computes the gradient of Lagrangian of Eq. (2), $\Delta \mathcal{L}$, and updates Q as

$$\mathbf{Q} = \mathbf{Q} - \eta(\Delta \mathcal{L} + \beta \Delta \Psi), \tag{3}$$

where $\Psi$ is an additional regularization term enforcing more variance, $\beta$ is a weight for controlling the importance of $\Psi$, and $\eta$ is a learning rate. $\Psi$ is calculated as

$$\Psi = \mathrm{Tr}(\mathbf{QX}\lambda\lambda^\intercal \mathbf{X}^\intercal \mathbf{Q}^\intercal), \tag{4}$$

where $\mathrm{Tr}(\cdot)$ is a trace operator and different values for $\lambda$ result in different versions of S-SVDD. In this paper, we use unregularized version (i.e. $\lambda_i = 0$) denoted as S-SVDD and two regularized versions S-SVDDr1 with $\lambda_i = 1$ and SVDDr2, where $\lambda$ is used to select only the support vectors.

### III. PROPOSED SYSTEM

Our work aims at helping the move from fully manual taxa identification of benthic macroinvertebrates to a semi-automated approach, where a trained machine learning model

can handle most of the specimens, while the human experts can concentrate on difficult and potentially most interesting cases. As our starting point, we assume the typical scenario where we have a trained deep neural network model that gives a satisfactory overall accuracy, while it fails to correctly identify specimens from rare species, which have biological/environmental importance. We propose using one-class classifiers as an additional mechanism that can be used together with a deep CNN to pinpoint specimens that potentially belong to the rare species for human expert inspection. The proposed approach does not require changing the existing CNN in any way and it can be used to complement any kind of CNN. Here, we only assume that the features extracted by the deep CNN provide sufficient information to discriminate the classes of interest even though the CNN itself fails to identify the samples belonging to these classes correctly.

The proposed overall framework is shown in Fig. 2. In the first phase, collected macroinvertebrate samples are imaged, and the images are preprocessed as needed (e.g., normalization, resizing). The images are fed to a trained deep CNN for initial identification. Features extracted from the second last layer of the network are mapped to a lower dimensionality to make the one-class models smaller and more focused on the key information. The mapping may be done using the traditional dimensionality reduction techniques, such as PCA or Linear Discriminant Analysis (LDA), or learned specifically for the one-class classification task, such as in S-SVDD. The lower-dimensional features are then fed to one or more one-class classifiers, each trained to model a single rare class of interest. Finally, the specimens which are classified to the target class are manually identified by a human expert, while otherwise the initial CNN identification is used in the subsequent biological assessment of the results. Note that the experts use the actual specimens with microscopic analysis, while the machine learning components rely on images and

features extracted from the images.

As one-class classifiers use only target class data for training, they may not be able to accurately distinguish unseen target samples from outliers, which have a high similarity with the target class. However, this may be even a benefit in our application. Trying to separate target samples from very similar outliers is naturally error-prone. Therefore, it is better to direct also these unclear cases for expert identification instead of trying to build a model as accurate as possible. In general, our goal is to detect as many samples from the target class as possible with the minimum amount of samples that require manual identification. However, it is not straightforward how to evaluate the performance of different one-class classifiers on the given task. Depending on biomonitoring goals and importance of the target class, it may vary how much human effort is acceptable to maximize the number of detected target class specimens.

## IV. EXPERIMENTAL SETUP AND RESULTS

### A. Dataset

We used FIN-Benthic2 dataset [5] in our experiments. The dataset is publicly available[1] and consists of 460004 images of 9631 benthic macroinvertebrate specimens belonging to 39 different taxa. The number of images per taxon varies from 490 to 44240 making the dataset very imbalanced. The images are of varying sizes and in PNG-format. FIN-Benthic2 provides 10 different data splits for training, validation, and testing. Each split has been formed so that the images of a single specimen (max 50) are in the same set (train/validation/test). In this paper, we consider only image-based identification, and we leave it for future work to investigate how to exploit the fact that we actually have several images corresponding to the same specimen. We used Split 1 as our data splitting.

For one-class classification, we selected three different taxa, *Capnopsis schilleri*, *Nemoura cinerea*, and *Leuctra nigra*, as our target classes. Each of these taxa is rare, and VGG16 has poor performance on them. The target classes were selected as a proof-of-concept, not based on their environmental importance. The image numbers for the selected classes are shown in Table I.

TABLE I
IMAGE NUMBERS IN SPLIT 1 OF FIN-BENTHIC2 DATASET

|  | Train | Validation | Test |
|---|---|---|---|
| *Capnopsis schilleri* | 600 | 100 | 350 |
| *Nemoura cinerea* | 650 | 100 | 50 |
| *Leuctra nigra* | 1100 | 50 | 200 |
| Whole dataset | 321407 | 45912 | 92685 |

### B. Classifiers and their parameters

As our base-model, we fine-tuned a VGG16 network [27] pre-trained on ImageNet using FIN-Benthic2 dataset. To make

---

[1]FIN-Benthic2 dataset is available at http://urn.fi/urn:nbn:fi:att:dc6440ad-43bd-4349-8fb9-0e0d1971a7e8

VGG16 suitable for our task, we added two dense layers on top of the VGG16 convolutional output. The first added layer is composed of 4069 neurons with ReLU activation. The second added layer is the output layer composed of 39 neurons using soft-max activation. We also added two dropout layers on top of the mentioned dense layers to avoid overfitting. The dropout rate was set to 40 percent. We fine-tuned the whole network for 50 epochs using Stochastic Gradient Descent with a learning rate of 0.007 and selected the final network based on the validation set accuracy. As the original images are of varying size, we first scaled them to 64x64. The overall accuracy of the network on the test set was 0.872. This is similar to earlier published results [5], while we did not concentrate on optimizing this step in this work.

We extracted the output of the second last VGG16 layer (i.e., 4096-D) for further analysis and first applied PCA for dimensionality reduction. We used only the target class training samples to obtain the PCA mapping and then applied this mapping for all the remaining data. We kept the first 100 principal components as our final feature vectors used for training and testing the one-class classifiers. Finally, we trained different one-class classifiers (separate models for each target species) using feature vectors of the training images of the target species. The hyper-parameters were optimized using the validation set. In the end, we tested the models with the full test set, where all the images not belonging to the target class were considered as outliers.

Note that both the CNN architecture (VGG16) and the dimensionality reduction techniques (PCA) are well-known and widely used techniques and they were selected for the experiments just as a proof-of-concept. The proposed approach can be used with any CNN model and more advanced approaches for learning the optimal dimensionality reduction will be a topic for future work.

The one-class classifiers considered were OC-SVM, SVDD, S-SVDD, S-SVDDr1, and S-SVDDr2 (See Section II-B). We used both linear and non-linear (kernel) versions. For the kernel version, we used the RBF kernel, i.e.,

$$\mathbf{K}_{ij} = \exp\left( \frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2} \right), \qquad (5)$$

where $\sigma$ is an additional hyper-parameter. The hyper-parameters $C$, $d$, $\eta$, $\beta$ and $\sigma$ were selected from the following values:

- $C \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$,
- $d \in \{1, 2, 3, 4, 5, 10, 20, 50, 100\}$,
- $\eta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$,
- $\beta \in \{0.01, 0.1, 1, 10, 100\}$,
- $\sigma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$.

For comparison purposes, we used the PCA output as the input also for S-SVDD even though it could be also used to learn the projection from the original features. For both linear and non-linear case, we also report the outcome of an ensemble of all the considered one-class classifiers (Ensemble-OCC). In Ensemble-OCC, a test instance is classified to the

TABLE II
ONE-CLASS CLASSIFIER RESULTS FOR DIFFERENT TARGET SPECIES

| | *Capnopsis schilleri* | | | | *Nemoura cinerea* | | | | *Leuctra nigra* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TPR | GM | TP | TP+FP | TPR | GM | TP | TP+FP | TPR | GM | TP | TP+FP |
| **CNN classification** | | | | | | | | | | | | |
| VGG16 | 0.046 | 0.214 | 16 | 101 | 0.020 | 0.141 | 1 | 39 | 0.170 | 0.412 | 34 | 174 |
| **Linear one-class classification** | | | | | | | | | | | | |
| OC-SVM | 0.906 | 0.613 | 317 | 54367 | 0.660 | 0.357 | 33 | 74739 | 0.625 | 0.437 | 125 | 64304 |
| SVDD | 0.346 | 0.586 | 121 | 701 | 0.280 | 0.525 | 14 | 1422 | 0.730 | 0.832 | 146 | 4860 |
| S-SVDD | 0.557 | 0.740 | 195 | 1893 | 0.480 | 0.676 | 24 | 4385 | 0.805 | 0.838 | 161 | 11910 |
| S-SVDDr1 | 0.609 | 0.773 | 213 | 1977 | 0.340 | 0.567 | 17 | 5209 | 0.805 | 0.837 | 161 | 12103 |
| S-SVDDr2 | 0.706 | **0.825** | 247 | 3573 | 0.560 | **0.702** | 28 | 11178 | 0.855 | **0.876** | 171 | 9625 |
| Ensemble-OCC | 0.620 | 0.779 | 217 | 2077 | 0.380 | 0.601 | 19 | 4487 | 0.820 | 0.848 | 164 | 11497 |
| **Non-linear one-class classification** | | | | | | | | | | | | |
| OC-SVM | 0.034 | 0.185 | 12 | 87 | 0.000 | 0.000 | 0 | 51 | 0.220 | 0.469 | 44 | 102 |
| SVDD | 0.331 | 0.574 | 116 | 658 | 0.300 | 0.543 | 15 | 1441 | 0.730 | 0.832 | 146 | 4904 |
| S-SVDD | 0.503 | 0.705 | 176 | 1169 | 0.440 | 0.649 | 22 | 3890 | 0.815 | 0.853 | 163 | 10085 |
| S-SVDDr1 | 0.540 | 0.730 | 189 | 1404 | 0.400 | 0.622 | 20 | 3138 | 0.780 | 0.854 | 156 | 6221 |
| S-SVDDr2 | 1.000 | 0.000 | 350 | 92685 | 0.220 | 0.465 | 11 | 1762 | 0.995 | 0.003 | 199 | 92683 |
| Ensemble-OCC | 0.503 | 0.705 | 176 | 1160 | 0.300 | 0.542 | 15 | 1883 | 0.775 | 0.851 | 155 | 6245 |

target class if most of the one-class classifiers predict the instance as positive.

## C. Performance metrics

In one-class classification, sensitivity and specificity are popular performance metrics used to evaluate the performance of the trained models. Sensitivity, also known as recall or True Positive Rate (TPR), is the fraction of correctly classified target class samples correctly:

$$TPR = \frac{TP}{P}, \qquad (6)$$

where TP is the total number of correctly predicted positive samples, and P is the total number of positive samples in the dataset. Specificity, also called the True Negative Rate (TNR), is the proportion of outlier samples that are correctly identified:

$$TNR = \frac{TN}{N}, \qquad (7)$$

where TN is an acronym for True Negatives, i.e., the total number of correctly predicted outliers, and N represents the total number of outliers in the dataset. Geometric Mean (GM) takes into account both TPR and TNR, and it is defined as the square root of the product of TPR and TNR:

$$GM = \sqrt{TPR \times TNR}. \qquad (8)$$

GM reflects both the ability of the model to detect target class samples and its ability to keep the overall amount of samples to be manually identified low. Therefore, we use it as our primary performance measure and use it for optimizing the hyper-parameters also. Furthermore, we report the total number of correctly identified target samples, i.e., True Positives (TP), and the total number of samples needing manual identification, i.e., True Positives and False Positives (TP+FP).

## D. Experimental results

We give the experimental results in Table II. We see that one-class classifiers, using the same features as VGG16, can indeed detect samples from rare species much better than the deep network with a reasonable overhead (TP+FP). Here, it should be remembered that up to 50 images can represent the same specimen and, therefore, the actual number of specimens needing manual inspection may be significantly smaller than the reported number of images.

Comparing the different one-class classifiers, we see that the results for the linear versions are more robust, while the kernel versions in some cases produce models that classify (almost) all the samples as target class or outliers. The best one-class classifier for all the considered taxa in terms of GM is the linear S-SVDDr2 model. We also notice that ensembling the one-class classifiers did not further improve the results.

## V. CONCLUSION AND FUTURE WORK

We proposed a taxa identification framework, where specimens potentially representing rare species are directed for human expert inspection. We showed that one-class classifiers can complement a deep neural network with high overall classification accuracy in a way that allows dividing the tasks between machine and human experts. This supports moving from fully manual to semi-automated taxa identification in biomonitoring. The best one-class classification model in terms of Geometric Mean was regularized linear Subspace Support Vector Data Description.

In this paper, we considered images separately, while we actually have multiple images of a single specimen. In our future work, we will consider how to exploit this information. For example, we may require a certain fraction of images to be

classified as target class to assign the specimen for human inspection, or we may use multi-modal one-class classifiers, e.g., [28], by considering each image as a separate modality. We will experiment with more advanced dimensionality reduction techniques and consider how to use classification confidences of both the CNN and one-class classifiers to further reduce the number of samples requiring human inspection. We will also experiment with different classifier types, such as class-specific classifiers, in our general identification framework.

## ACKNOWLEDGMENT

## REFERENCES

[1] EU Water Framework Directive (WDF), Directive 2000/60/EC, *Journal of the European Communities*, vol. L327/1, 2000.

[2] D. Buchner, A. Beermann, A. Laini, P. Rolauffs, S. Vitecek, D. Hering, and F. Leese, "Analysis of 13,312 benthic invertebrate samples from german streams reveals minor deviations in ecological status class between abundance and presence/absence data," *PloS one*, vol. 14, no. 12, 2019.

[3] Y. Sun, Y. Takemon, and Y. Yamashiki, "Freshwater spring indicator taxa of benthic invertebrates," *Ecohydrology & Hydrobiology*, 2019.

[4] S. Poikane, R. Johnson, L. Sandin, A. Schartau *et al.*, "Benthic macroinvertebrates in lake ecological assessment: a review of methods, intercalibration and practical recommendations," *Science of the total environment*, vol. 543, pp. 123–134, 2016.

[5] J. Ärje, J. Raitoharju, A. Iosifidis, V. Tirronen, K. Meissner, M. Gabbouj, S. Kiranyaz, and S. Kärkkäinen, "Human experts vs. machines in taxa recognition," *Signal Processing: Image Communication*, p. 115917, 2020.

[6] J. Raitoharju and K. Meissner, "On confidences and their use in (semi-) automatic multi-image taxa identification," in *IEEE Symposium Series on Computational Intelligence*, 2019.

[7] A. A. dos Santos and W. N. Gonçalves, "Improving pantanal fish species recognition through taxonomic ranks in convolutional neural networks," *Ecological Informatics*, vol. 53, 2019.

[8] N. Wyniawskyj, M. Napiorkowska, D. Petit, P. Podder, and P. Marti, "Forest monitoring in guatemala using satellite imagery and deep learning," in *IEEE International Geoscience and Remote Sensing Symposium*, 2019.

[9] J. Ärje, D. Milioris, D. Tran, A. Iosifidis, J. Raitoharju, M. Gabbouj, J. Jepsen, and T. H. ye, "Automatic flower detection and classification system using a light-weight convolutional neural network," in *EUSIPCO workshop on "Signal Processing, Computer Vision and Deep Learning for Autonomous Systems"*, 2019.

[10] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, "Learning to navigate for fine-grained classification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 420–435.

[11] J. Raitoharju, E. Riabchenko, I. Ahmad, A. Iosifidis, M. Gabbouj, S. Kiranyaz, V. Tirronen, J. Ärje, S. Kärkkäinen, and K. Meissner, "Benchmark database for fine-grained image classification of benthic macroinvertebrates," *Image and Vision Computing*, vol. 78, 2018.

[12] M. Kelly, S. Schneider, and L. King, "Customs, habits, and traditions: the role of nonscientific factors in the development of ecological assessment methods," *Wiley Interdisciplinary Reviews: Water*, vol. 2, no. 3, 2015.

[13] D. Lytle, G. Martinez-Munoz, W. Zhang, N. Larios, L. Shapiro, R. Paasch, A. Moldenke, E. Mortensen, S. Todorovic, and T. Dietterich, "Automated processing and identification of benthic invertebrate samples," *Journal of the North American Benthological Society*, vol. 29, no. 3, 2010.

[14] S. Kiranyaz, T. Ince, J. Pulkkinen, M. Gabbouj, J. Ärje, S. Kärkkäinen, V. Tirronen, M. Juhola, T. Turpeinen, and K. Meissner, "Classification and retrieval on macroinvertabrate image databases," *Computers in Biology and Medicine*, vol. 41, no. 7, 2011.

[15] J. Ärje, C. Melvad, M. R. Jeppesen, S. A. Madsen, J. Raitoharju, M. S. Rasmussen, A. Iosifidis, V. Tirronen, M. Gabbouj, K. Meissner *et al.*, "Automatic image-based identification and biomass estimation of invertebrates," *Methods in Ecology and Evolution*, 2020.

[16] J. Raitoharju, E. Riabchenko, K. Meissner, I. Ahmad, A. Iosifidis, M. Gabbouj, and S. Kiranyaz, "Data enrichment in fine-grained classification of aquatic macroinvertebrates," in *ICPR Workshop on Computer Vision for Analysis of Underwater Imagery*, 2016.

[17] C. Huang, Y. Li, C. Change Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[18] S. S. Khan and M. G. Madden, "One-class classification: taxonomy of study and review of techniques," *The Knowledge Engineering Review*, vol. 29, no. 3, 2014.

[19] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "One class classification applied in facial image analysis," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.

[20] F. Sohrab, J. Raitoharju, A. Iosifidis, and M. Gabbouj, "Ellipsoidal subspace support vector data description," *IEEE Access*, vol. 8, pp. 122 013–122 025, 2020.

[21] J. F. López, J. A. L. Sotelo, D. Leite, and C. Peña-Reyes, "Applying one-class learning algorithms to predict phage-bacteria interactions," in *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. IEEE, 2019.

[22] S. Bo and Y. Jing, "One-class classification based river detection in remote sensing image," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2017.

[23] B. S. Kumar and V. Ravi, "Text document classification with pca and one-class svm," in *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*. Springer, 2017.

[24] B. Schölkopf, R. Williamson, A. Smola, and J. Shawe-Taylor, "Sv estimation of a distribution's support," 1999.

[25] D. Tax and R. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, 2004.

[26] F. Sohrab, J. Raitoharju, M. Gabbouj, and A. Iosifidis, "Subspace support vector data description," in *International Conference on Pattern Recognition*, 2018.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[28] F. Sohrab, J. Raitoharju, A. Iosifidis, and M. Gabbouj, "Multimodal subspace support vector data description," *Pattern Recognition*, p. 107648, 2020.