***Title**:* "Using Deep Convolutional Networks for the Automatic Recognition of MacroInvertebrate in Rivers and Affluents in Panama"

***Authors**: Carlos Quintero[1], Fernando Merchán[1], Aydeé Cornejo[2], Javier E. Sánchez-Galán[1,3]*

Research Problem: The determination of water quality in rivers is a very important task for ensuring the well-being of any population [1]. Moreover, It is very laborious task, involving a combination of chemical and microbiological analysis techniques. A low cost alternative to achieve this same goal is the use of *biomonitoring* or recognition of macroinvertebrate organisms which presence or absence can be used to determine the water quality [2].

Biomonitoring is often used in rural communities that use of natural streams for consumption and daily life. In those cases community members are often the ones obtaining the quality reports. Although it is an efficient method it has the disadvantage that it requires biological training both for the capture, and identification of macroinvertebrate present in water and watersheds [3].

Motivation: This work focuses on providing a tool to automate the recognition of freshwater macroinvertebrate images based on deep convolutional networks using the Python programming language, and the Tensorflow [4] and Keras libraries.

Technical Contribution: The contributions of the works are presented as follows: First, the development of a database of macroinvertebrate images is presented. Then the details of the implementation of the automatic recognition algorithm are described.

*Database of Training Images***:** A database of images (subfolders) was created for each of the 14 families of macroinvertebrates in the study, examples of characteristic images of each of the families are shown in **Figure 1**.

Figure 1. Test images of Macroinvertebrate families including shells classes.

Due there were few test images (n<100) for some of the families studied, the *Data augmentation (DA)* [6] technique was used to obtain new images from the initial images. To achieve this, different methods of rotation, cutting, translation and change in the intensity of colors are applied to each image, thus obtaining new images for training the model in the network.

[1] Universidad Tecnológica de Panamá
[2] Instituto Conmemorativo Gorgas de Estudios de la Salud (ICGES). Panamá
[3] Corresponding author. javier.sanchezgalan@utp.ac.pa

Model Training: The Inception-v3 algorithm [5] was used as the reference model, the *Transfer Learning module (TL)* was added, which implies that only the last layers of the network were trained and the classification of the family resulted in the output layer of macroinvertebrate. Finally, all the training models were applied the *Fine Tuning (FT)* method, balancing the values of the network in the last layers to improve the classification. Network Architecture is shown in **Figure 2**.

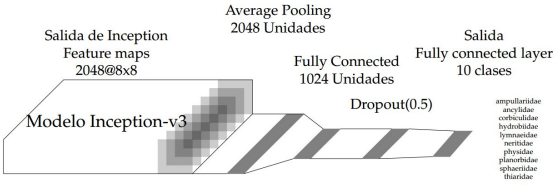

Figure 2. Network Architecture. Inception-v3 is used to predict macroinvertebrate classes.

Results: Various experiments were sought after, however we present the results of Experiment #4, the automation of the recognition of 14 families of macroinvertebrates (10 families of shells and 4 of the families "ephemeroptera" and "odonata"), in Figure 2.

	Entrenamiento					Validación				
Modelo	top-1	top-3	top-5	κ de Cohen	error	top-1	top-3	top-5	κ de Cohen	error
TL	0.9901	1.0000	1.0000	0.9892	0.0401	0.7007	0.8849	0.9309	0.6734	1.1625
TL-DA	0.6641	0.9062	0.9766	0.6299	0.9673	0.6809	0.9046	0.9605	0.6486	0.9643
FT-DA	0.8594	0.9844	1.0000	0.8436	0.5727	0.7270	0.9013	0.9737	0.6993	0.8500

	Prueba				
Modelo	top-1	top-3	top-5	κ de Cohen	error
TL	0.6863	0.8529	0.9706	0.6542	1.2423
TL-DA	0.7059	0.8922	0.9608	0.6760	0.9835
FT-DA	0.8235	0.9804	0.9902	0.8044	0.6856

Results of Experiment #4. The FT-DA method obtains the lowest classification errors in the validation and test datasets, but the performance of the TL method obtains a minimum error in the training phase.

Conclusions: It is possible to develop a macro invertebrate image recognition tool using a small volume database. In overall, the best results in the recognition tool were obtained by using fine adjustment in the final layers of the convolutional model and a data augmentation module. To obtain a good model, it is necessary to balance the volume of images between the different classes as much as possible or to weight the training in the classes with the lowest volume, this can be seen in the high value obtained in Cohen's κ coefficient.

Acknowledgements: This project is carried out under the framework of the Senacyt-funded project entitled "*Participatory Biomonitoring of Water Quality with Rural Aqueduct Management Boards (JAAR): A Tool for the Sustainability of Water Resources in Panama*". FM, AC and JSG are members of the National Investigators System (SNI).

Preliminary results were presented as: *"Uso de Redes Neuronales Convolucionales para el Reconocimiento Automático de Imágenes de Macroinvertebrados para el Biomonitoreo Participativo" at 6th Engineering, Science and Technology Conference - Panama 2017 (ESTEC 2017).* https://knepublishing.com/index.php/KnE-Engineering/article/view/1462/3528

References:

[1] Karr, J. R. (1999), Defining and measuring river health. Freshwater Biology, 41: 221-234. doi:10.1046/j.1365-2427.1999.00427.x

[2] Buss, D.F., Carlisle, D.M., Chon, TS. et al. Environ Monit Assess (2015) 187: 4132.

[3] Perla Alonso-EguíaLis, José Manuel Mora, Bruce Campbell, Monika Springer. Diversidad, conservación y uso de los macroinvertebrados dulceacuícolas de México, Centroamérica, Colombia, Cuba y Puerto Rico. Publisher: Instituto Mexicano de Tecnología del Agua, Jiutepec, Morelos, México.Editor: Alonso-Eguía Lis et al. (Eds.).ISBN: 978-607-9368-21-0. 2014.

[4] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[5] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818-2826. 2016.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.