

# Overview of SnakeCLEF 2022: Automated Snake Species Identification on a Global Scale

Lukáš Pícek<sup>1</sup>, Marek Hruží<sup>1</sup>, Andrew M. Durso<sup>2</sup> and Isabelle Bolon<sup>4</sup>

<sup>1</sup>*Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Czechia*

<sup>2</sup>*Department of Biological Sciences, Florida Gulf Coast University, Florida, USA*

<sup>3</sup>*Institute of Global Health, Department of Community Health and Medicine, University of Geneva, Switzerland*

## Abstract

The main goal of the third year of the SnakeCLEF challenge was to provide an evaluation platform that helps track the performance of AI-driven methods for snake species recognition systems on a global scale and allows direct comparison with human experts. We ran two challenges separately for humans – experts and novices – and AI methods in order to lay the groundwork for future comparison between human and machine-based snake species identification. We have provided 187,129 snake observations with 318,532 photographs – 270,251 for training and 48,281 for testing – of 1,572 snake species collected in 208 countries. The human performance evaluation was conducted on a tailored subset with 150 images derived from the full test set. We report (i) a description of the provided data, (ii) evaluation methodology and principles, (iii) an overview of the methods submitted by the participating teams, and (iv) a discussion of the obtained results.

## Keywords

LifeCLEF, SnakeCLEF, fine grained visual categorization, global health, epidemiology, snake bite, snake, reptile, benchmark, biodiversity, species identification, machine learning, computer vision, classification

## 1. Introduction

A robust image-based identification system for snake species is an important goal for biodiversity, conservation, and global health. With over half a million victims of death and disability from venomous snakebite annually, such a system could significantly improve eco-epidemiological data and treatment outcomes (e.g. selection of specific antivenoms) [1, 2]. Importantly, most herpetological expertise and most snake images are concentrated in developed countries where snake diversity is relatively low and snakebite is not a major public health concern. In contrast, remote parts of developing countries tend to lack expertise and images, even in areas where snake diversity is high and snakebites are common [3, 4]. Thus, snake species identification assistance has a bigger potential to save lives in areas with the least information.

A primary difficulty of snake species identification lies in the high intra-class and low inter-class variance in appearance, which may depend on geographic location, color morph, sex, or age. At the same time, many species are visually similar to other species, i.e., mimicry (Figure 1).

---


*CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy*

✉ picekl@kky.zcu.cz (L. Pícek)

🆔 0000-0002-6041-9722 (L. Pícek); 0000-0002-2287-0985 (M. Hruží); 0000-0002-3008-7763 (A. M. Durso); 0000-0001-5940-2731 (I. Bolon)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Furthermore, our knowledge of which snake species occur in which countries is incomplete, and it is common that most or all images of a given snake species might originate from a small handful of countries or even a single country. Furthermore, many snake species resemble species found on other continents, with which they are entirely allopatric. Incorporating metadata on the geographic origin of an unidentified snake almost always narrows down the possible correct identifications considerably because only about 125 of the approximately 3,900 snake species co-occur in any given location [5]. It is known that more widespread species with more images are over-predicted relative to rare species with few images [6], and this can be a particularly vexing problem when trying to predict the identity of species that are widespread across areas of the world with few images.

The main goal of the SnakeCLEF 2022 competition was to provide a reliable evaluation ground for automatic snake species recognition. Like other LifeCLEF competitions, the SnakeCLEF 2022 competition was hosted on Kaggle<sup>1</sup> primarily to attract machine learning experts to participate and present their ideas.



**Figure 1:** Harmless mimic species *Cemophora coccinea* ssp. *coccinea* (top row) and poisonous lookalike species. *Micrurus pyrrhocryptus*, *Micrurus ibiboboca*, and *Micrurus nigrocinctus* (left to right, bot. row). ©roadmom–iNaturalist, ©Anthony Damiani–iNaturalist, ©Adam Cushen–iNaturalist, ©Alexander Guñazu–iNaturalist, ©Tarik Câmara–iNaturalist, and ©Cristhian Banegas–iNaturalist.

<sup>1</sup><https://www.kaggle.com/competitions/fungiclef2022>



**Figure 2:** Two snake observations from SnakeCLEF2022 dataset – three images for each individual. ©André Giraldi – iNaturalist, ©Harshad Sharma – iNaturalist.

**Table 1**

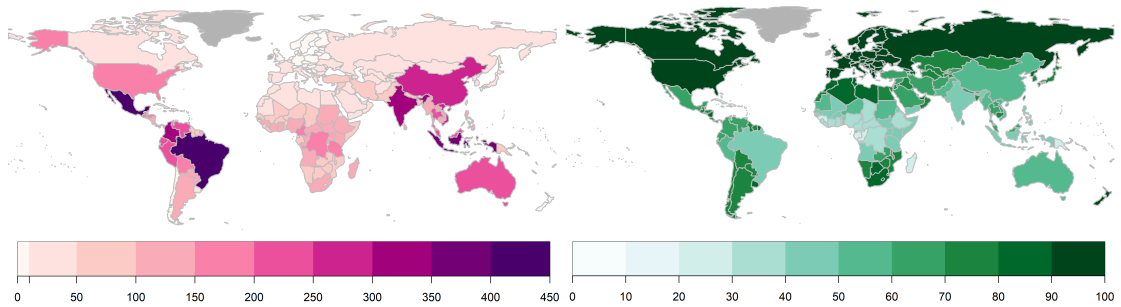
Details of the SnakeCLEF 2022 datasets and their comparison with previous editions.

Dataset	Species	Images	Observation	Countries	min / max samples
SnakeCLEF 2020	783	259,214	×	145	19 / 14,433
SnakeCLEF 2021	772	386,006	×	188	10 / 22,163
SnakeCLEF 2022	1,572	318,532	187,129	208	5 / 6,472
SnakeCLEF 2022–Training	1,572	270,251	158,698	207	3 / 5,518
SnakeCLEF 2022–Test	1,572	48,281	28,431	183	2 / 954

## 1.1. Dataset

For this year, the dataset used in previous editions [7, 8] has been extended with new and rare species. The number of species was doubled and the number of images from remote geographic areas with no or just a few samples was increased considerably, i.e., the uneven species distributions across all the countries was straightened. The SnakeCLEF 2022 dataset is based on 187,129 snake observations – including some instances of multiple images of the same individual (refer to Figure 2) – with 318,532 photographs belonging to 1,572 snake species and observed in 208 countries. The dataset has a heavy long-tailed class distribution, where the most frequent species (*Natrix natrix*) is represented by 6,472 images and the least frequent species just by 5 samples. The difference in the number of images between the species with the most and fewest was reduced by an order of magnitude relative to SnakeCLEF2021. All the data were gathered from the online biodiversity platform iNaturalist<sup>2</sup>. Additional dataset parameters and comparison with previous editions are listed in Table 1.

<sup>2</sup><https://www.inaturalist.com/>



**Figure 3:** **Left:** Worldwide snake species distribution, i.e., The number of species found in each country. Large countries in the tropics (Brazil, Mexico, Colombia, India, and Indonesia) have more than 300 species. **Right:** Percentage of snake species per country included in the SnakeCLEF 2022 dataset. The countries with adequate species coverage are those from Europe, Oceania, and North America, i.e., the countries with the smallest diversity.

For testing, two sets were created: (i) the full test set for a machine evaluation, with 48,280 images from 28,431 observations, and (ii) the subset from the full test set with 150 observations, tailored for the human performance evaluation. Unlike in other LifeCLEF competitions, where the final testing set remained undisclosed, we provided the test data without labels to the participants. To prevent over fitting to the leaderboard, the evaluation method was composed of two stages; the first being the public leaderboard where the user scores were calculated on an unknown 20% of the test set, and the second a private leaderboard where participants were scored on the remaining part of the test set. In addition to image data, we provide:

- human verified species labels that allow up-scaling to higher taxonomic ranks,
- the country-species mapping file describing country-species presence to allow better regularization towards all geographical locations, based on The Reptile Database [9], and
- information about endemic species — species that occur only in one geographical region, e.g., Australia or Madagascar.

Geographical information, i.e., state/province and country labels, was included for approximately 95% of the training and test images. Additionally, we provide a mapping matrix ( $MM_{cs}$ ) describing country-species presence to allow better worldwide regularization.

$$MM_{cs} = \begin{cases} 1 & \text{if species } S \in \text{country } C, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Unlike last year’s dataset, where the vast majority (77%) of all images came from the United States and Canada, the SnakeCLEF 2022 dataset includes just a fraction of the data (28.3%) from the United States and Canada. The rest of the data is distributed across remaining regions, e.g., Europe, Asia, Africa, Australia and Oceania. This was achieved by limiting the number of observations per species in a given country to 400. The estimated worldwide snake distribution and their coverage is visualized in Figure 3.

## 1.2. Timeline

The SnakeCLEF 2022 competition was announced together with the data in late February 2022 through the LifeCLEF, Kaggle, and FGVC challenge pages. Anyone with research ambitions was allowed to register and participate in the competition. The test data were provided jointly with the training data allowing continuous evaluation. Each team could submit up to 2 submissions a day. The competition deadline was May 16, setting the competition for roughly three months. Participants had to submit CSV files containing the Top1 prediction for each snake observation. Once the submission phase was closed (mid-May), the participants were allowed to submit post-competition submissions to evaluate their ablation studies.

## 1.3. Evaluation Protocol

The main goal of this challenge was to build a system that is capable of recognizing 1,572 snake species based on the given snake observation – unseen set of images – and relevant geographical location. As a main metric, we use the macro F1 score ( $F_1^m$ ). The  $F_1^m$  is defined as the mean of class-wise F1 scores:

$$F_1^m = \frac{1}{N} \sum_{s=0}^N F_{1s}, \quad F_{1s} = 2 \times \frac{P_s \times R_s}{P_s + R_s}, \quad (2)$$

where  $s$  is species index,  $N$  equals to the number of classes in a training set. The F1 score for each class represents harmonic mean of the class precision  $P_s$  and recall  $R_s$ . This type of evaluation is suitable for data with long-tail distribution, since the quantity of samples from individual classes does not effect the outcome. For the additional evaluation of the performance of the teams we also compute the micro classification accuracy, which is a ratio between the number of correctly classified samples and all samples.

## 1.4. Working Notes

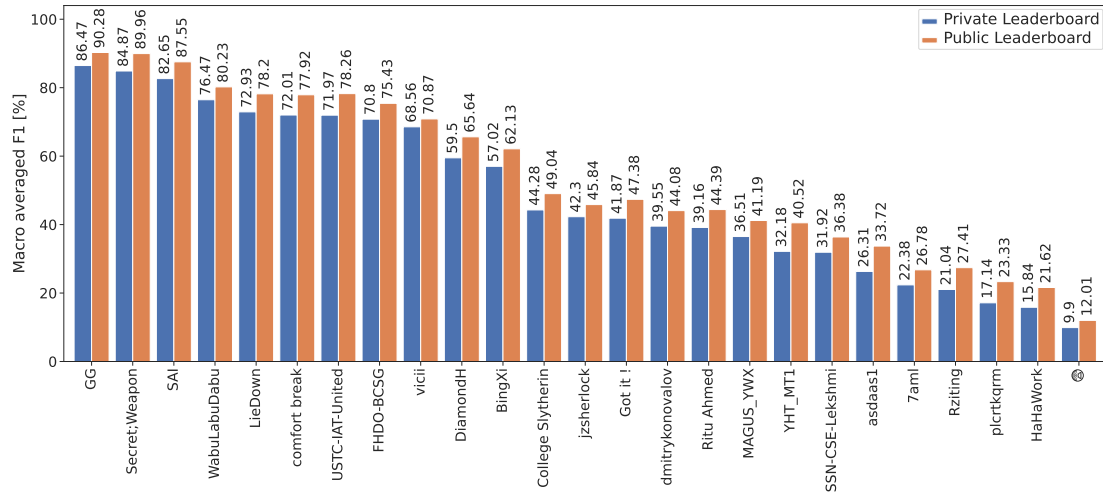
All participants were asked to provide code and a *Working Note* paper – a technical report with information needed to reproduce the results of all submissions. All submitted *Working Notes* were reviewed by 2–3 reviewers with a decent publication history in the field of Computer Vision and Machine Learning, ensuring a sufficient reproducibility and quality. The review process was single-blind and offered up to two rebuttals. The acceptance rate was 66.66%.

## 2. Results

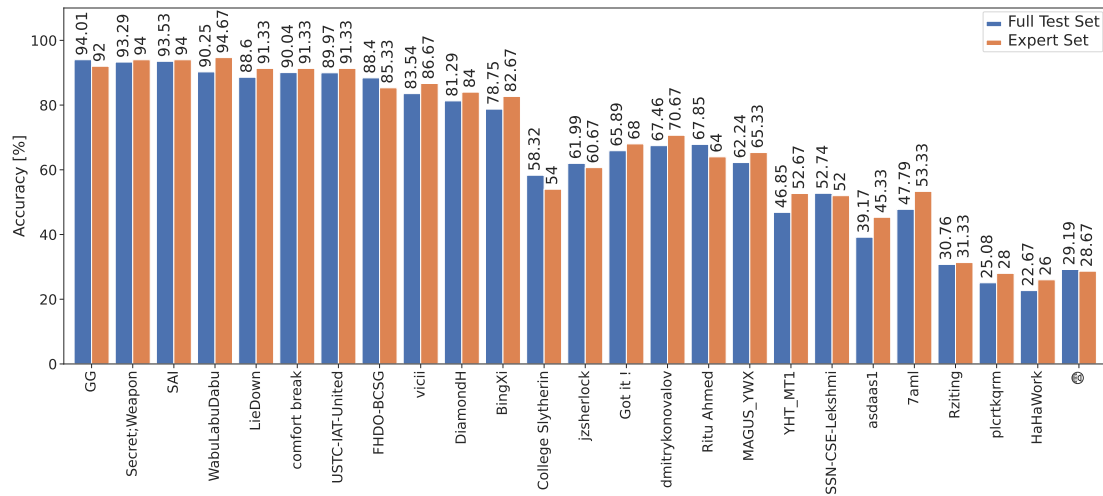
The best performing team achieved  $F_1^m$  of 86.47% on the private part of the test set and 94.01% classification accuracy on the full test set. Both scores are the best ones in the given category. We observe a steady decrease in the  $F_1^m$  score for the first three teams achieving 84.87% and 82.65%  $F_1^m$  score. Then there are several significant drops in performance. Similar behavior is present in the classification accuracy with the second team achieving 93.29% and the third one achieving 93.53%. It can be seen that the accuracy is not correlated with the  $F_1^m$  score which

suggests that some teams did not cope well with the long tail distribution. Further performance evaluation for Top-25 teams is provided in Figure 4.

On the expert set, the best performing team – 4<sup>th</sup> in terms of  $F_1^m$  score – achieved impressive accuracy of 94.67%. Eight teams achieved satisfactory accuracy on the expert set, around 90%. Further performance evaluation for the Top-25 teams is provided in Figure 5.



**Figure 4:** Official SnakeCLEF 2022 competition results, sorted by performance on the private set.



**Figure 5:** Identification accuracy on the full test set and the test set for human performance evaluation.

## 2.1. Participants and Methods

A total of 31 teams participated in the SnakeCLEF 2022 challenge and submitted 676 submissions. Everyone who submitted a solution better than baseline submission, i.e., random predictions, was considered a participant. The number of participants quadrupled since last year, primarily because Kaggle was used as an evaluation platform. More details can be found in the individual working notes of participants [10, 11, 12, 13, 14, 15] who passed the review process, ensuring a sufficient level of reproducibility and quality. The main outcomes we can derive from the achieved results are as follows:

**GG** [10]: The winner of the challenge. They have introduced a novel architecture CoLKA-Net based on VAN [16] (Visual Attention Network) and CoAtNet [17]. It is a combination of large kernel attention and vision transformer. In an ablation study, the model outperforms other tested models – ViT [18], Swin [19], VOLO [20], and ConvNeXt [21] by around 2 points of  $F_1^m$  score (78.00% to 80.10%). In addition, the team used techniques such as Label Aware Smoothing [22], Pseudo labelling for tail classes, FixRes mitigation [23], and augmentations. When TrivialAugment [24] was deployed during the middle stage of experimentation the team observed a rise in  $F_1^m$  of around 0.5%. Progressively, Random Erasing [25], CutMix [26] and Mixup [27] were added which helped with regularization. The final submission score was achieved by an ensemble of six models  $2 \times$  ConvNeXt, VOLO, CoLKA-Net, Swin, and ViT. The novel CoLKA-Net is an interesting contribution with potential outside of the scope of this competition.

**Secret;Weapon** [11]: The runner-up focused solely on ViT models, ensembling three different variants; two Large models trained on different resolutions (384; 432) and one Huge model (392). The observed trend from the ablation study is that larger models with higher input resolution perform better. The ViT models were pretrained as Masked autoencoders [28] on ImageNet-1K [29]. The team designed a new Effective Logit Adjustment Loss combining Logit adjustment loss [30] and Class-balanced Loss [31]. They observe a better overall performance when compared to Cross Entropy and further analyze that the improvement comes from the tail classes. The metadata was integrated by applying an estimated species a priori distribution in individual regions to the ViT estimates. Interestingly, when all metadata (code, endemic, country) were added the  $F_1^m$  score on validation data rose but dropped rapidly on the test data. The best combination on the test data was when the country code was omitted. This hints at a distribution gap between the training and test metadata.

**SAI** [12]: The team used MetaFormer [32]. The metadata passes through an Embedding, individual tags (code, endemic, country) are concatenated and the final MLP is used to produce the Meta token. The Metaformer seems suitable for integrating categorical tokens. The team shows in an ablation study a significant raise in  $F_1^m$  score from 66.64% to 74.18%. The Logit adjustment loss function is used to optimize the model. It works better than Seesaw [33] loss but is outperformed by a post-hoc logit adjustment. Filtering the predictions based on possible locations of species raises the  $F_1^m$  score from 64.32% to 69.09%. SimCLR [34] with InfoLoss [35] is used to train the model. The team shows significant improvement when compared to only

supervised models pre-trained on different datasets (eg. from ImageNet-1k pre-train  $F_1^m$  of 63.76% to 68.83% when SimCLR is used). In accordance with the findings of previous teams, a bigger resolution of input images leads to better scores. The final score was achieved by the multi-scale, multi-crop, and multi-resolution model ensemble.

**USTC-IAT-United** [13]: The team combines models of Swin Transformer, EfficientNet, and BEiT [36]. They experiment with several loss functions (CE Loss, Seesaw Loss, Focal Loss [37]) and show that Focal Loss performs the best, but only marginally (less than 1%  $F_1^m$  score improvement when compared to CE). From the individual models, the EfficientNet-L2 is the best one. Given the highest number of parameters (480M), it is expected. From this point of view, it is interesting that Swin with 197M parameters achieves the worst score ( $F_1^m$  of 61.70%) and is outperformed by EfficientNet-B7 ( $F_1^m$  of 64.99%) with only 66M parameters. The authors experimented with a fine-grained classification technique – PIM [38] – to find informative and discriminative features. They show the improvement on the Swin model ( $F_1^m$  from 62.38% to 63.80%) but then do not use it because of the relatively big computational burden when training the model. Finally, the models are ensembled by averaging their softmax outputs. The authors observe a steady increase in  $F_1^m$  score when adding the models to the ensemble one by one. But an unexpected drop is observed when the last model – Swin – is added. The drop is present only in the test data and not the validation data. This opens the question of how, when, and what models to add to the ensemble.

**FHDO-BCSG** [14]: The team adopted the two-stage principle of the best solution from Snake-CLEF2021 [39]. In the first stage, the snake is detected and in the second stage, the detected region is classified by a CNN. However, this year it was not the best choice, substantially lacking in performance when compared to the winning team (70.7%  $F_1^m$  versus 85.4%  $F_1^m$ ). Even so, the detection helps and the YOLOv5 [40] detection network improved the private  $F_1^m$  by 13% on average. The team experimented with two CNN models – EfficientNet and ConvNeXt. They provide an in-depth study of the behavior of the models when combined with different techniques of optimizing, adding metadata, types of augmentations, and so on. The best competition score was achieved by ensembling seven models – EfficientNet-B4 with and without object detection, EfficientNet-v2-m with no object detection, and EfficientNet-v2-m with and without object detection. However this score of 70.79%  $F_1^m$  is only a bit ahead of a single EfficientNet-v2-m model with score 70.23%  $F_1^m$ . This begs the question of how important it is to use ensemble in a real-life application when the improvement is limited. Lastly, the team experimented with different types of metadata representation. In all cases, the metadata were used as a priori distributions of snake species in different locations that were multiplied with the resulting softmax. The worst option was to use the estimated a priori distribution, followed by the binarized version (with a threshold of 0) and finally the post-competition best result (73.90%  $F_1^m$ ) when the binarized distribution was multiplied by the a priori country distribution of the training dataset.

**anonymous\_rice** [15]: The team experimented with several CNN models – ResNet, ResNext, and EfficientNet to produce deep features of the images and concatenate them with the categorical representation of metadata. In the final solution, ResNet-101 and EfficientNet-B0 were



used. The concatenated features are inputted into the XGBoost Ensemble Classifier [41] to produce the final classification. The nice property of the XGBoost algorithm is that the relative importance of the ensembled features is computed. Unfortunately, the team achieved low score of 3.6%  $F_1^m$  but after the competition when the backbones were trained further, the score raised significantly to 51.39%  $F_1^m$ . This shows some potential of the XGBoost algorithm that may be interesting to study in the future.

### 3. Conclusions and Perspectives

This paper presents an overview and results evaluation of the third edition of the SnakeCLEF challenge organized in conjunction with the Conference and Labs of the Evaluation Forum (CLEF<sup>3</sup>) and LifeCLEF<sup>4</sup> research platform [42], and FGVC9 Workshop<sup>5</sup> – The Ninth Workshop on Fine-Grained Visual Categorization organized within the CVPR conference. The main outcomes we can derive from the this year’s evaluation are as follows.

**Transformer-based architectures outperformed CNNs.** This year various deep neural network architectures – Convolutional Neural Networks and Transformers – were evaluated; ConvNext [21], EfficientNet [43], Vision Transformer [18], Swin Transformer [19], and MetaFormer [32]. Unlike last year, where the CNN architectures overwhelmed the performance, Vision Transformer architectures were a vital asset for most methods submitted this year. The second best method with  $F_1^m$  score of 84.56% was based on an ensemble of exclusively ViT models and performed slightly worse (−0.9%) than the best performing system that used a combination of Transformer and CNN models. An ensemble of MetaFormer models achieved the third-best score of 82.65%. It seems that Transformers and CNNs benefit from each other in an ensemble, whereas a standalone Transformer ensemble performs better than a pure CNN ensemble which achieved an  $F_1^m$  score of "only" 70.80%

**Loss Function matters.** Several loss functions were evaluated: Label Aware Smoothing [22], (modified) Categorical Cross-Entropy, Seesaw [33], and Focal Loss [37]. Overall, any Loss function if used is better than standard CrossEntropy. The winning team used Label Aware Smoothing. The runner-up used an Effective Logit Adjustment Loss and showed an improvement of around 2% of  $F_1^m$  score when compared to Cross Entropy, reducing the error rate by 15%. The the third team used Logit adjustment to outperform the Seesaw loss from an  $F_1^m$  score of 76.49% to 78.57%. The team USTC-IAT-United compared CE Loss, Seesaw Loss, and Focal Loss with EfficientNet model. Focal Loss performed the best, but only with marginal improvement over CE Loss and Seesaw Loss.

**Self-supervision has potential.** Adding unlabeled data to the train set is a welcome option when not many observations of a species are available. The third team used the SimCLR [34] method with InfoNCE [35] loss function to increase the  $F_1^m$  score from 63.76% to 68.83% when

---

<sup>3</sup> <http://www.clef-initiative.eu/>

<sup>4</sup> <http://www.lifeclef.org/>

<sup>5</sup> <https://sites.google.com/view/fgvc9/home>

compared to an ImageNet-1k pretrained models. Overall, performance on tail classes was higher this year.

**Geographical metadata improves classification performance.** Most teams report accuracy improvement when adding the metadata into the learning process. The second team achieved an improvement of 10.89% in terms of the  $F_1^m$  score using a simple location filtering approach. The third team described an absolute improvement of 7.54% when adding the metadata into the MetaFormer. Using the a priori country distribution from the training data outperformed other approaches tried by the fifth team, but in real life this will frustrate users in countries that lack many snake images to use as training data. Representation of many locations in both training and testing data remain important and challenging.

**Ensemble helps, but at what cost?** Most teams used ensembling to increase the accuracy of classification. The standard approach was to compute an average of the individual models' decisions. Some teams used a late fusion of deep features by concatenation as an ensemble technique. Even though the improvement in accuracy is observable (around 1 percentage point of  $F_1^m$  across the board), it would be interesting to measure the added computational complexity vs the added accuracy. In the case of snakebite, the system's inference time plays a crucial role.

## Acknowledgments

LP was supported by the UWB grant, project No. SGS-2022-017. LP and MH were supported by the Technology Agency of the Czech Republic, project No. SS05010008. AMD was supported by the Florida Gulf Coast University Office of Scholarly Innovation and Student Research. The work described herein has been supported by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. LM2018101 LINDAT/CLARIAH-CZ.

## References

- [1] I. Bolon, A. M. Durso, S. Botero Mesa, N. Ray, G. Alcoba, F. Chappuis, R. Ruiz de Castañeda, Identifying the snake: First scoping review on practices of communities and healthcare providers confronted with snakebite across the world, *PLoS ONE* 15 (2020) e0229989.
- [2] R. Ruiz de Castañeda, A. M. Durso, N. Ray, J. L. Fernández, D. J. Williams, G. Alcoba, F. Chappuis, M. Salathé, I. Bolon, Snakebite and snake identification: empowering neglected communities and health-care providers with ai, *The Lancet Digital Health* 1 (2019) e202–e203.
- [3] A. M. Durso, R. Ruiz de Castañeda, C. Montalcini, M. R. Mondardini, J. L. Fernandez-Marques, F. Grey, M. M. Müller, P. Uetz, B. M. Marshall, R. J. Gray, et al., Citizen science and online data: Opportunities and challenges for snake ecology and action against snakebite, *Toxicon: X* (2021) 100071.

- [4] A. M. Durso, I. Bolon, A. Kleinhesselink, M. Mondardini, J. Fernandez-Marquez, F. Gutsche-Jones, C. Gwilliams, M. Tanner, C. E. Smith, W. Wüster, et al., Crowdsourcing snake identification with online communities of professional herpetologists and avocational snake enthusiasts, *Royal Society open science* 8 (2021) 201273.
- [5] U. Roll, A. Feldman, M. Novosolov, A. Allison, A. M. Bauer, R. Bernard, M. Böhm, F. Castro-Herrera, L. Chirio, B. Collen, et al., The global distribution of tetrapods reveals a need for targeted reptile conservation, *Nature Ecology & Evolution* 1 (2017) 1677–1682.
- [6] A. M. Durso, G. K. Moorthy, S. P. Mohanty, I. Bolon, M. Salathé, R. Ruiz de Castañeda, Supervised learning computer vision benchmark for snake species identification from photographs: Implications for herpetology and global health, *Frontiers in Artificial Intelligence* 4 (2021) 17.
- [7] L. Picek, R. Ruiz de Castañeda, A. M. Durso, S. P. Mohanty, Overview of the snakeclef 2020: Automatic snake species identification challenge, in: *CLEF task overview 2020*, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece., 2020.
- [8] L. Picek, A. M. Durso, R. Ruiz De Castañeda, I. Bolon, Overview of SnakeCLEF 2021: Automatic snake species identification with country-level focus, in: *Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum*, 2021.
- [9] P. Uetz, P. Freed, J. Hošek, et al., The reptile database, 2020. URL: [https://reptile-database.reptarium.cz/advanced\\_search](https://reptile-database.reptarium.cz/advanced_search).
- [10] Y. Shen, X. Sun, Z. Zhu, When large kernel meets vision transformer: A solution for snakeclef & fungiclef, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [11] L. Yang, X. Li, R. Song, K. Zhu, G. Li, Solution for snakeclef 2022 by tackling long-tailed categorization, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [12] C. Zou, F. Xu, M. Wang, W. Li, Y. Cheng, Solutions for fine-grained and long-tailed snake species recognition in snakeclef 2022, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [13] J. Yu, H. Chang, Z. Cai, G. Xie, L. Zhang, K. Lu, S. Du, Z. Wei, Z. Liu, F. Gao, F. Shuang, An efficient model integration-based snake classification algorithm, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [14] L. Bloch, J.-F. Böckmann, B. Bracke, C. M. Friedrich, Combination of object detection, geospatial data, and feature concatenation for snake species identification, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [15] M. Palaniappan, K. Desingu, H. Bharathi, E. A. Chodisetty, A. Bhaskar, Deep learning and gradient boosting ensembles for classification of snake species, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [16] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, S.-M. Hu, Visual attention network, *arXiv preprint arXiv:2202.09741* (2022).
- [17] Z. Dai, H. Liu, Q. V. Le, M. Tan, Coatnet: Marrying convolution and attention for all data sizes, *Advances in Neural Information Processing Systems* 34 (2021) 3965–3977.
- [18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint arXiv:2010.11929* (2020).

- [19] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
- [20] L. Yuan, Q. Hou, Z. Jiang, J. Feng, S. Yan, Volo: Vision outlooker for visual recognition, arXiv preprint arXiv:2106.13112 (2021).
- [21] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.
- [22] Z. Zhong, J. Cui, S. Liu, J. Jia, Improving calibration for long-tailed recognition, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 16489–16498.
- [23] H. Touvron, A. Vedaldi, M. Douze, H. Jégou, Fixing the train-test resolution discrepancy, Advances in neural information processing systems 32 (2019).
- [24] S. G. Müller, F. Hutter, Trivialaugment: Tuning-free yet state-of-the-art data augmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 774–782.
- [25] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, in: Proceedings of the AAAI conference on artificial intelligence, volume 34, 2020, pp. 13001–13008.
- [26] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6023–6032.
- [27] H. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz, mixup: Beyond empirical risk minimization, arXiv preprint arXiv:1710.09412 (2017).
- [28] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, R. Girshick, Masked autoencoders are scalable vision learners, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16000–16009.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.
- [30] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, S. Kumar, Long-tail learning via logit adjustment, arXiv preprint arXiv:2007.07314 (2020).
- [31] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, S. Belongie, Class-balanced loss based on effective number of samples, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 9268–9277.
- [32] Q. Diao, Y. Jiang, B. Wen, J. Sun, Z. Yuan, Metaformer: A unified meta framework for fine-grained recognition, arXiv preprint arXiv:2203.02751 (2022).
- [33] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, D. Lin, Seesaw loss for long-tailed instance segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9695–9704.
- [34] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International conference on machine learning, PMLR, 2020, pp. 1597–1607.
- [35] A. Van den Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive

- coding, arXiv e-prints (2018) arXiv-1807.
- [36] H. Bao, L. Dong, F. Wei, Beit: Bert pre-training of image transformers, arXiv preprint arXiv:2106.08254 (2021).
  - [37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.
  - [38] P.-Y. Chou, C.-H. Lin, W.-C. Kao, A novel plug-in module for fine-grained visual classification, arXiv preprint arXiv:2202.03822 (2022).
  - [39] R. Borsodi, D. Papp, Incorporation of object detection models and location data into snake species classification, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.
  - [40] G. Jocher, A. Stoken, A. Chaurasia, J. Borovec, NanoCode012, TaoXie, Y. Kwon, K. Michael, L. Changyu, J. Fang, A. V, Laughing, tkianai, yxNONG, P. Skalski, A. Hogan, J. Nadar, imyhxy, L. Mammana, AlexWang1900, C. Fati, D. Montes, J. Hajek, L. Diaconu, M. T. Minh, Marc, albinxavi, fatih, oleg, wanghaoyang0106, ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support, 2021. URL: <https://doi.org/10.5281/zenodo.5563715>. doi:10.5281/zenodo.5563715.
  - [41] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, New York, NY, USA, 2016, pp. 785–794. URL: <http://doi.acm.org/10.1145/2939672.2939785>. doi:10.1145/2939672.2939785.
  - [42] A. Joly, H. Goëau, S. Kahl, L. Picek, T. Lorieul, E. Cole, B. Deneu, M. Servajean, A. Durso, H. Glotin, R. Planqué, W.-P. Vellinga, A. Navine, H. Klinck, T. Denton, I. Eggel, P. Bonnet, M. Šulc, M. Hruz, Overview of lifeclef 2022: an evaluation of machine-learning based species identification and species distribution prediction, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2022.
  - [43] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.