

Overview of SnakeCLEF 2023: Snake Identification in Medically Important Scenarios

Lukáš Pícek¹, Rail Chamidullin¹, Marek Hruz¹ and Andrew M. Durso²

¹*Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Czech Republic*

²*Department of Biological Sciences, Florida Gulf Coast University, Florida, USA*

Abstract

Developing an effective automatic system for snake species identification has significant importance for biodiversity, conservation, and global health. Snakebites result in over half a million deaths and disabilities worldwide each year, highlighting the urgent need for a system to enhance eco-epidemiological data and improve treatment outcomes, especially in remote regions that lack the necessary expertise and data but have high snake diversity and a high incidence of snakebites. The SnakeCLEF challenge provide an evaluation ground that helps track the performance of AI-driven methods for snake species recognition systems on a global scale. The fourth edition of the SnakeCLEF challenge focuses on (i) evaluation of gradual improvements in automatic snake species identification, (ii) testing worldwide generalization on two specific scenarios, i.e., India and Central America, and (iii) evaluation with uneven costs for different errors, such as mistaking a venomous snake for a harmless one. This paper showcases the vital role of a robust automatic identification system for snakes, particularly in regions with limited resources, and highlights the potential impact on biodiversity conservation and global health outcomes. We report (i) a comprehensive description of the provided data, (ii) an evaluation methodology, (iii) an overview of the submitted methods, and (iv) perspectives derived from the achieved 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

With the high number of annual deaths and disabilities exceeding half a million caused by venomous snakebites, a robust image-based system for snake species identification might play a crucial role not just in global health but also has the potential to significantly enhance eco-epidemiological data and treatment outcomes, including the selection and distribution of antivenoms [1, 2]. Most expert knowledge (herpetological) and snake data are accumulated in developed countries (i.e., North America, Australia, and the EU) where snake diversity is relatively low or snakebite is not a major public health concern. In contrast, remote regions of developing countries often lack both expertise and data, even in areas with high snake diversity and many snakebites [3, 4]. Therefore, assistance with snake species identification has great potential for saving lives, particularly in those regions where it is needed most.

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece

*Corresponding author.

✉ picekl@kky.zcu.cz (L. Pícek)

🆔 0000-0002-6041-9722 (L. Pícek); 0000-0003-1728-8939 (R. Chamidullin); 0000-0002-7851-9879 (M. Hruz);

0000-0002-7851-9879 (A. M. Durso)

© 2023 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)

The main challenge in snake species identification arises from the significant intra-class and low inter-class variance in appearance, which can be influenced by factors such as geographic location, color morph, sex, or age. Additionally, many species have evolved to look visually similar to other species to protect themselves (i.e., mimicry). Interestingly, many snake species resemble species found on different continents, with which they have no overlapping ranges (e.g., Figure 1). Therefore, exploiting information about an unknown snake’s geographic origin could narrow the identification process, as only approximately 125 out of the roughly 4,000 snake species co-occur in any given location [5]. It is known that species with a wider distribution and a greater number of images are more frequently predicted than rare species with limited images [6]. This can pose significant challenges when attempting to predict the identity of species that are widespread across regions with sparse image data.

The primary objective of the SnakeCLEF 2023 challenge is to provide an evaluation ground that helps track the performance of AI-driven methods for snake species recognition systems on a global scale. Taking part in the LifeCLEF [7, 8], and FGVC-CVPR workshops, the SnakeCLEF 2023 competition was hosted on HuggingFace to attract participation from machine learning experts and facilitate the presentation of their ideas. The fourth edition of the SnakeCLEF challenge focuses on (i) evaluation of gradual improvements in automatic snake species identification, (ii) testing worldwide generalization on two specific scenarios, i.e., India and Central America, and (iii) evaluation with uneven costs for different errors, such as mistaking a venomous snake for a harmless one. This paper showcases the vital role of a robust automatic identification system for snakes, particularly in regions with limited resources, and highlights the potential impact on biodiversity conservation and global health outcomes. We report (i) a comprehensive description of the provided data, (ii) an evaluation methodology, (iii) an overview of the submitted methods, and (iv) perspectives derived from the achieved results.



Figure 1: Medically-important venomous snakes. Top row: *Daboia russelii* (Russell’s Viper – Asia), *Bitis arietans* (Puff Adder – Africa), and *Crotalus adamanteus* (Eastern Diamond-backed Rattlesnake – North America). **Bottom row:** *Bothrops atrox* (Common Lancehead – South America), *Acanthophis antarcticus* (Common Death Adder – Australia), and *Vipera ammodytes* (Nose-horned Viper – Europe). Photos by @chinmay_c_m, @kevin_ulett, @tom_slade, @eduardo_navia, @laylab11, and @vojtechvita.

2. Objective

From previous SnakeCLEF editions [9, 10, 11] we have learned that "machines" can accurately recognize snake species, achieving $F_1^C \approx 90\%$ and Top1 Accuracy $\approx 90\%$ even in scenarios with long-tailed distributions and 1,600 species. Thus, testing real Medically Important Scenarios and specific countries/regions and integrating the medical importance of species is the next step that will allow more comprehensive performance evaluation. Therefore, the main objective for all participants of this competition is to create a machine learning model that can accurately predict snake species for given observation data, i.e., images and location, and: (i) fits limits for memory footprint (max size of 1GB), (ii) minimizes the danger to human life, i.e., the venomous \longleftrightarrow harmless confusion, and (iii) generalizes to all countries and geographic regions.

2.1. Test Scenarios

This year we focused on two geographic scenarios, both important for snakebite globally. First, we emphasize a large, biodiverse country with historical emphasis on just a few species of medically-important venomous snakes (MIVS). Second, we focus on a region composed of four adjacent countries that differ in socio-economic status but overlap in snake fauna.

India: Estimates suggest that 1.1–1.7 million snakebites take place in India every year [12], leading to 40,000–60,000 deaths from snakebite annually [13], more than any other country. More than 75 MIVS occur in India [14], although the "Big Four"—spectacled cobra (*Naja naja*), common krait (*Bungarus caeruleus*), Russell's viper (*Daboia russelii*) and saw-scaled viper (*Echis carinatus*)—have historically been considered responsible for the majority of dangerous snakebites and have been the focus of almost all specific antivenom research, development, and production [15]. In this challenge, we include 43 MIVS (Big Four + 39 others) and 49% (125/253) of non-venomous snakes found in India, including several close lookalikes; e.g., venomous kraits (genus *Bungarus*) and harmless wolfsnakes (genus *Lycodon*) (Figure 2).



Figure 2: Top row: *Lycodon laoensis* (Laotian Wolfsnake), a non-venomous species. Photos by @seiha, @bangtran, and @Rueangrit Promdam. Bottom row: *Bungarus fasciatus* (Banded Krait), a dangerous, venomous snake. Photos by @alexericsonlee, @hmheinz, and @Lawrence Hylton.

Central America (Panama, Costa Rica, Nicaragua, and Honduras): Snakebite is a significant public health problem in all four countries, mainly affecting people living in rural areas [16]; Panama reports the highest incidence of snakebite cases in the Americas [17]. Snakebite is a reportable disease in Panama, Costa Rica, and Nicaragua [17], but there are few epidemiological data from Honduras [18]. Detailed comparisons highlight significant advantages enjoyed by Costa Rica [19], the regional leader in epidemiological data and healthcare quality as well as a global leader in antivenom research, development, and production [20].

Collectively, 43 MIVS occur in these 4 countries, of which we cover 38 as well as 67% (147/218) of non-venomous snake species. Central America is notable for its relatively high diversity of venomous coralsnakes (genus *Micrurus*) and their harmless mimics [21] (Figure 3), which are among the most frequently mis-classified of all snake species by both humans [22] and computer vision [23]. Just as all antivenom used in Panama, Nicaragua, and Honduras is manufactured in Costa Rica. [24], we anticipate that geographical bias can be reduced and regional sufficiency achieved through sharing of training data and improved algorithm performance.



Figure 3: **Top row:** *Pliocercus elapoides* (Variegated False Coralsnake), a non-venomous species. Photos by © *pnlosmarmoles*, © *sonoran*, and © *davidruizdominguez*. **Bottom row:** *Micrurus nigrocinctus* (Central American Coralsnake), a dangerous, venomous snake. Photos by © *viperinus*, © *arthurchp*, and © *jo22nfrog*.

3. Dataset

The dataset was constructed from three sources. Training, validation, and *public testing* (Public Test), subsets are based on observations submitted to the citizen science platforms – iNaturalist and HerpMapper. The competition test set (Private Test) is composed of private images from individuals and natural history museums who have not put those images online in any form and therefore is undisclosed, and participants do not have access to this data.

In total, all the subsets combine roughly 110,000 real snake observations with community-verified species labels. The number of species was extended to 1,784 snake species (up from 1,572 in 2022 [11] and 772 in 2021 [10]). Besides the increase in the number of species, the number of observations from remote geographic areas with no or just a few samples was increased

considerably, and the uneven species distributions across all the countries were straightened.

Despite these improvements, there are still fewer observations and poorer species coverage from remote parts of developing countries that tend to lack herpetological expertise and have high snake diversity and snakebite incidence (i.e., southern Asia, Sub-Saharan Africa, and parts of Central/South America) [3]. Unfortunately, snakebites are common in many of these same areas [25]. In addition to image data, we have provided information about medical importance, i.e., whether or not a species is a medically-important venomous snake (MIVS), for each species, as well as a country-species relevance matrix. We list the dataset statistics in Table 1 and direct comparison with previous editions of the SnakeCLEF datasets in Table 2.

Table 1

SnakeCLEF 2023 dataset statistics for each split and subset. Taxonomic and geographic coverage.

| Subset | Species | Countries | Images | Observations |
|--------------------------|---------|-----------|---------|--------------|
| Training | 1,784 | 212 | 168,144 | 95,588 |
| ↳ <i>iNaturalist</i> | 1,784 | 210 | 154,301 | 85,843 |
| ↳ <i>HerpMapper</i> | 889 | 119 | 13,843 | 9,745 |
| Validation | 1,599 | 177 | 14,117 | 7,816 |
| Public Test | 1,784 | 191 | 28,274 | 15,632 |
| Private Test | 182 | 8 | 8,080 | 3,765 |
| ↳ <i>India</i> | 76 | 1 | 2,892 | 2,395 |
| ↳ <i>Central America</i> | 107 | 4 | 5,188 | 1,370 |

Table 2

SnakeCLEF datasets statistics comparison. All editions. Taxonomic and geographic coverage.

| 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 2023 | 1,784 | 218,615 | 122,801 | 212 | 3 / 2,079 |

4. Evaluation Protocol

To motivate research in recognition scenarios with uneven costs for different errors, such as mistaking a venomous snake for a harmless one, this year’s challenge goes beyond the 0-1 loss common in classification. We make some assumptions to reduce the complexity of the evaluation. We pretend that there exists a universal antivenom that is applicable to all venomous snake bites¹. Furthermore, we pretended that our hypothetical antivenom is not lethal or seriously harmful when applied to a healthy human². Hence, we penalize the misclassification

¹In reality, there are hundreds of snake antivenoms [26], each with its own range of efficacy across species, and the correct one is not always available in real snakebite cases [27].

²In reality, treatment with antivenom when unneeded is a waste of a limited resource at best and can cause severe anaphylaxis or death at worst; e.g. [28]; although patient harm in the reverse case is more common [29].

of a venomous species with a harmless one more than the other way around. Although this solution is not perfect, it is a first step into a more complex evaluation of snake identification and snakebite treatment pipelines. We specify two metrics (T_1, T_2) reflecting these different scenarios.

4.1. Evaluation Metrics

To motivate research in recognition scenarios with uneven costs for different errors, such as mistaking a venomous snake for a harmless one, this year's challenge goes beyond the 0-1 loss common in classification. In addition to previously used metrics, e.g., accuracy and macro averaged F_1 , we introduce novel metrics that consider venomous \longleftrightarrow harmless confusion, and different error cost.

Track1 metric: The first track metric is the sum of (L) over all test observations:

$$\mathbf{L} = \sum_i L(y_i, \hat{y}_i) \quad (1)$$

Let us consider a function p such that $p(s) = 1$ if species s is venomous; otherwise, $p(s) = 0$. For a correct species y and predictor species \hat{y} , the loss $L(y, \hat{y})$ is given as follows:

$$L(y, \hat{y}) = \begin{cases} 0 & \text{if } y = \hat{y} \\ 1 & \text{if } y \neq \hat{y} \text{ and } p(y) = 0 \text{ and } p(\hat{y}) = 0 \\ 2 & \text{if } y \neq \hat{y} \text{ and } p(y) = 0 \text{ and } p(\hat{y}) = 1 \\ 2 & \text{if } y \neq \hat{y} \text{ and } p(y) = 1 \text{ and } p(\hat{y}) = 1 \\ 5 & \text{if } y \neq \hat{y} \text{ and } p(y) = 1 \text{ and } p(\hat{y}) = 0 \end{cases} \quad (2)$$

Note: The costs were selected to illustrate a higher cost when a venomous snake is mistaken for a harmless one. We do not claim the selected costs reflect the risks in a practical scenario: practical costs would have to be determined by assessing what exactly follows after the recognition process. One can imagine several aspects, e.g., the species of snake and potency of its venom, the cost and side-effects of the antivenom, patient attributes, and so on.

Track2 metric: The second track metric includes the overall classification rate (macro averaged F_1) and the venomous species confusion error. The metric is a weighted average between the macro F_1 -score and the weighted accuracies of different types of confusion.

$$M = \frac{w_1 F_1 + w_2 C_{h \rightarrow h} + w_3 C_{h \rightarrow v} + w_4 C_{v \rightarrow v} + w_5 C_{v \rightarrow h}}{w_1 + w_2 + w_3 + w_4 + w_5} \quad (3)$$

where $w_1 = 1, w_2 = 1, w_3 = 2, w_4 = 2$, and $w_5 = 5$ are the weights of individual confusions. $C_{v \rightarrow h}$ is the percentage of wrongly classified venomous species as a harmless species, $C_{h \rightarrow v}$ is the percentage of wrongly classified harmless species as a venomous species, $C_{v \rightarrow v}$ is the percentage of wrongly classified venomous species as another venomous species, $C_{h \rightarrow h}$ is the percentage of wrongly classified harmless species as another harmless species, and the F_1 is the macro averaged F_1 -score.

This metric has a lower bound of 0% and upper bound of 100%. The lower bound is achieved if you misclassify every species and furthermore you misclassify every harmless species as a venomous one and vice-versa. On the other hand, if F_1 is 100% (every species is classified correctly), every P_i must be equal to zero by definition and 100% will be achieved.

4.2. Timeline

The SnakeCLEF 2023 competition was announced together with the dataset release on the 14th of February 2023 through the LifeCLEF, HuggingFace, and FGVC challenge pages and anyone was allowed to register and participate in the competition. The competition deadline was the 24th of May, arranging the competition for around 3.5 months. Differently from last year, the test data was kept secret. However, we allowed one submission a day using the HuggingFace evaluation platform and public test set. Two weeks prior to the deadline, five submissions a day were allowed. The final evaluation was possible only via the "code+model submission" submission. Therefore, one week prior to the deadline, we provided a dockerizer sample "code+model submission" and we asked all the participants to provide their "models". Once the submission phase was closed, all participants could submit post-competition submissions to evaluate their ablation studies.

4.3. Working Notes

All participants were encouraged to provide code and a technical report (Working Note) with the information needed to reproduce the results of all submissions. All submitted Working Note papers were reviewed by 2–3 reviewers with a decent publication history in Computer Vision and Machine Learning, ensuring sufficient reproducibility and quality. The review process was single-blind and provided up to two rebuttals.

5. Challenge Results

The official challenge results, based on Track 1 and Track 2 Metrics, are displayed in Figure 4. Apart from that, we show the public leaderboard scores in Figure 5. The best-performing team on the private leaderboard – word2vector – achieved 91.43% and 908 in terms of Metric 1 and Metric 2, respectively, on the private part of the test set. Interestingly, this team did not achieve the best F_1^m score trailing behind by 1.36%. The top 3 teams achieved a competitive performance, while the remaining two heavily underperformed. The differences in ranking between public and private leaderboards revealed a small generalization capability to neglected regions of BBracke team’s solution.

Central America: In the Central America scenario where we cover snake species observations from Costa Rica, Nicaragua, Honduras, and Panama, the top 3 teams achieved similar scores in all measured metrics—around 93% in Track 1 and 250 in Track 2.

India: The Indian scenario seems to be a deal breaker. Even though first- and second-runner-up teams performed slightly better in the Central America scenario in terms of all metrics than the winning team, the error in Metric 2 in the India scenario doubled for those teams.

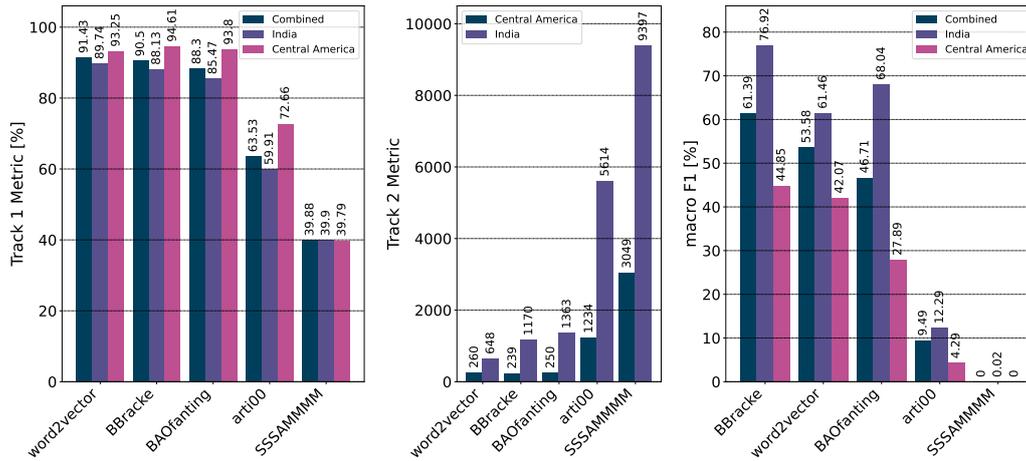


Figure 4: Official SnakeCLEF 2023 competition results – 5 teams. Private Leaderboard.

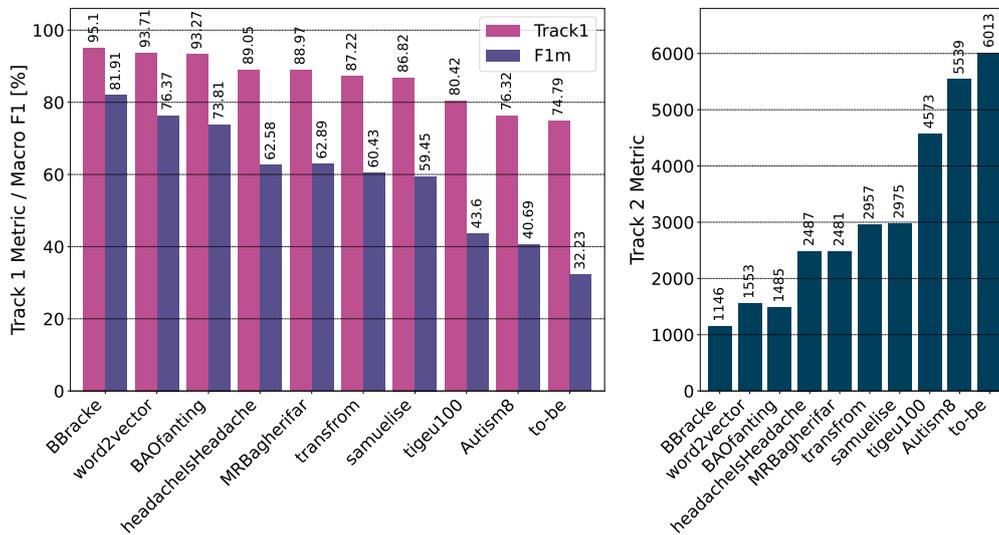


Figure 5: SnakeCLEF 2023 competition results – Top10 teams on public test set. Public Leaderboard.

6. Participants and Methods

This year a total of 16 teams participated in the SnakeCLEF challenge. However, just five teams submitted their models for private evaluation together with the working notes. An overview of the methods used just by teams that passed the review process is synthesized below. More details can be found in their individual working notes [30, 31, 32, 33].

word2vector [30]: The team uses textual input to help with the classification process. They input the textual representations of the geographical locations (e.g. country codes) into the

CLIP [34] model. The output embedding vectors are used to compute a PCA model and their dimension is reduced. Although the experiments are not provided, the reduction seems like a logical step, because the variety of textual inputs is very limited. The output of this procedure is used to train a small neural network to predict the geographical prior given the observed image and its prediction from the visual model. In this regard, the authors experiment with several architectures of neural networks including ResNet [35], BEiT [36], EVA [37], Swin Transformer [38], VOLO [39], and ConvNeXt [40]. The last one was used for further experimentation, although the baseline experiments did not show any clear winner regarding the backbone. The only takeaway is that ResNet50 underperformed when compared to the other models. The authors also use features from the penultimate convolution stage of the ConvNext and after applying MaxPooling they concatenate them with the final features. The authors used Seesaw Loss [41] to handle the long-tailed distribution of training data and Real-World Weighted Cross-Entropy [42] in the last three epochs to accommodate for the weighted penalization in the challenge metric. They report that CutMix [43] augmentation outperforms others such as RandomMix [44]. As a last step, the authors want to make sure that they do not classify a venomous species as a harmless one. That is why when the model exhibits uncertainty in its predictions for a particular observation, they classify it as a venomous snake species if any venomous species appears in the top-5 predictions. This approach resulted in the best Track1 metric of 91.31%. However, they achieved the second best F_1 score of 53.58%.

BBracke [31]: The team used ConvNeXt as the visual backbone. The model pre-trained on iNaturalist21 had better performance than the one pre-trained on ImageNet21k. The model is trained by optimizing a loss combining Focal Loss [45] and ArcFace Loss [46]. The authors incorporate the metadata in the form of a learned embedding of the alphabetically sorted region codes and endemcity flag. They are handled as separate inputs and combined in a small neural network. The embedding is learned end to end together with visual species classification network. The authors make use of several images in a single observation by applying Multi-Instance Learning methods. The authors achieve the second-best Track1 result of 90.19% and the best F_1 score of 61.39%. This shows that optimizing the F_1 score might not lead to the best results when analyzing in a more realistic scenario when different errors have different costs.

BAOfanting [32]: The authors use Metaformer [47], which naturally combines the visual and textual information. The textual information represents the meta-data in the form of country code, endemcity, and binomial name, which are one-hot encoded and passed through a non-linear embedding layer. The model is trained using the ArcFace Loss function. Each image in the test set is augmented five times and the predictions are averaged. The final decision about an observation is obtained by three differently trained Metaformers with or without the SimCLR [48] method. The method achieved 88.30% in the Track1 metric and F_1 score of 46.71%

arti00 [33]: The author uses a method for detecting region of interest in the image based on the activations of neurons in a classification neural network. Three architectures were tested – EfficientNet_B0, Dino v2 [49], and DEiT [50]. The DEiT model trained with Equalized Focal Loss [51] achieved the best results which resulted in fifth place in the competition with Track1 metric of 38.13% and F_1 score of 9.49%.

7. Conclusions and Perspectives

This paper presents an overview and results evaluation of the fourth edition of the SnakeCLEF challenge organized in conjunction with the CLEF LifeCLEF lab, and CVPR-FGVC10 – The Tenth Workshop on Fine-Grained Visual Categorization organized within the CVPR conference. The main outcomes we can derive from this year’s evaluation are as follows.

NLP model encoded metadata might be the next big thing. As in previous years, most of the teams used the provided metadata and showed that by doing so the competition metric improves. CLIP [34] – a strong multi-modal descriptor, was used for the first time in this competition to encode the metadata. This trend may lead to the utilization of bigger NLP models.

Transformers for the win. But do not rule out CNNs yet. On the vision part, convolutional models (ResNet [35], EfficientNet [52], ConvNext [40]) and Transformer models (MetaFormer [47], Swin [38], VOLO [39]) were used to extract the visual features. When teams compared the architectures side-by-side, most of the time the Transformer architecture performed better. However, the winning team used ConvNextv2. Due to the lack of a fair and exhaustive ablation study, it is not clear how a Transformer model would fare.

Task-tailored losses and self-supervision are the keys to learning. Traditionally, Seesaw loss [41] and SimCLR [48] were used to cope with the long-tailed data. Some teams introduced a weighted version of the loss functions tackling the different penalization for different errors. Multi-Instance Learning [53] was applied to make use of more images per observation.

Medically important scenarios might be on to something. The final results on the private dataset show an interesting behavior of the models. The best team (named *word2vector*) achieved a macro F_1 score of 53.58% with the competition score of 91.31%. The runner-up (BBracke) actually achieved a much better F_1 score of 61.39% but had a lower competition score of 90.19%. We hypothesize that this was possible due to the post-processing step of the team *word2vector*. They used a conservative strategy to prevent the most strictly penalized misclassification of non-venomous species as venomous ones: whenever their top-5 results contained a venomous species, the observation was classified as such.

Three months-long competition period is not enough. Given the complexity of this task and considering additional requirements like the limit for memory footprint (max size of 1GB) and code submission, many participants were discouraged from participation and code submission. Out of 58 registered teams, 15 submitted reasonable predictions, and just five submitted *code+model* for final evaluation.

Acknowledgments

RCH, LP, and MH were supported by the Technology Agency of the Czech Republic, project No. S05010008. RCH was supported by the UWB grant, project No. SGS-2022-017.

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* 9 (2021) 100071.
- [4] J. Longbottom, F. M. Shearer, M. Devine, G. Alcoba, F. Chappuis, D. J. Weiss, S. E. Ray, N. Ray, D. A. Warrell, R. R. de Castañeda, et al., Vulnerability to snakebite envenoming: a global mapping of hotspots, *The Lancet* 392 (2018) 673–684.
- [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] A. Joly, C. Botella, L. Picek, S. Kahl, H. Goëau, B. Deneu, D. Marcos, J. Estopinan, R. Chamidullin, M. Šulc, M. Hruz, M. Servajean, B. Kellenberger, E. Cole, H. Glotin, et al., Overview of lifeclef 2023: evaluation of ai models for the identification and prediction of birds, plants, snakes and fungi, in: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 14th International Conference of the CLEF Association, CLEF 2023, Thessaloniki, Greece, September 18–23, 2023, Proceedings, Springer, 2023*.
- [8] A. Joly, H. Goëau, S. Kahl, L. Picek, C. Botella, D. Marcos, M. Šulc, M. Hruz, T. Lorieul, S. S. Moussi, et al., Lifeclef 2023 teaser: Species identification and prediction challenges, in: *European Conference on Information Retrieval, Springer, 2023*, pp. 568–576.
- [9] L. Picek, R. Ruiz De Castañeda, A. M. Durso, P. M. Sharada, 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*.
- [10] L. Picek, A. M. Durso, I. Bolon, R. Ruiz de Castañeda, Overview of snakeclef 2021: Automatic snakespecies identification with country-level focus, in: *Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021*, pp. 1463–1476.
- [11] L. Picek, A. M. Durso, M. Hruz, I. Bolon, Overview of SnakeCLEF 2022: Automated snake species identification on a global scale, in: *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, 2022*, pp. 1957–1969.
- [12] W. Suraweera, D. Warrell, R. Whitaker, G. Menon, R. Rodrigues, S. H. Fu, R. Begum, P. Sati, K. Piyasena, M. Bhatia, et al., Trends in snakebite deaths in india from 2000 to 2019 in a nationally representative mortality study, *Elife* 9 (2020) e54076.

- [13] B. Mohapatra, D. A. Warrell, W. Suraweera, P. Bhatia, N. Dhingra, R. M. Jotkar, P. S. Rodriguez, K. Mishra, R. Whitaker, P. Jha, et al., Snakebite mortality in india: a nationally representative mortality survey, *PLoS Negl Trop Dis* 5 (2011) e1018.
- [14] R. Whitaker, A. Captain, *Snakes of india. the field guide*. 2nd edi, 2008.
- [15] R. S. Laxme, S. Khochare, H. F. de Souza, B. Ahuja, V. Suranse, G. Martin, R. Whitaker, K. Sunagar, Beyond the 'big four': Venom profiling of the medically important yet neglected indian snakes reveals disturbing antivenom deficiencies, *PLoS neglected tropical diseases* 13 (2019) e0007899.
- [16] J. M. Gutiérrez, Clinical toxicology of snakebite in central america, in: *Handbook of: Clinical Toxicology of Animal Venoms and Poisons*, CRC Press, 2017, pp. 645–665.
- [17] J. M. Gutiérrez, H. W. Fan, C. L. Silvera, Y. Angulo, Stability, distribution and use of antivenoms for snakebite envenomation in latin america: report of a workshop, *Toxicon* 53 (2009) 625–630.
- [18] J. Alger, E. E. Boza-Oviedo, R. E. Mejía, F. Navas, P. Simons-Morales, R. T. Velázquez, J. M. Gutiérrez, A multi-sectorial approach for addressing the problem of snakebite envenoming in honduras, *Toxicon: official journal of the International Society on Toxinology* 159 (2019) 61–62.
- [19] J. M. Gutiérrez, K. Maduwage, G. Iiyasu, A. Habib, Snakebite envenoming in different national contexts: Costa rica, sri lanka, and nigeria, *Toxicon: X* 9 (2021) 100066.
- [20] J. M. Gutiérrez, C. F. Teixeira, H. W. Fan, Instituto butantan and instituto clodomiro picado: A long-standing partnership in science, technology, and public health, *Toxicon* 202 (2021) 75–81.
- [21] A. R. Davis Rabosky, C. L. Cox, D. L. Rabosky, P. O. Title, I. A. Holmes, A. Feldman, J. A. McGuire, Coral snakes predict the evolution of mimicry across new world snakes, *Nature communications* 7 (2016) 11484.
- [22] A. M. Durso, I. Bolon, A. R. Kleinhesselink, M. R. Mondardini, J. L. 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.
- [23] I. Bolon, L. Picek, A. M. Durso, G. Alcoba, F. Chappuis, R. Ruiz de Castañeda, An artificial intelligence model to identify snakes from across the world: Opportunities and challenges for global health and herpetology, *PLoS neglected tropical diseases* 16 (2022) e0010647.
- [24] J. M. Gutiérrez, Global availability of antivenoms: The relevance of public manufacturing laboratories, *Toxins* 11 (2018) 5.
- [25] A. Kasturiratne, A. R. Wickremasinghe, N. de Silva, N. K. Gunawardena, A. Pathmeswaran, R. Premaratna, L. Savioli, D. G. Lalloo, H. J. de Silva, The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths, *PLoS medicine* 5 (2008) e218.
- [26] A. Silva, G. K. Isbister, Current research into snake antivenoms, their mechanisms of action and applications, *Biochemical society transactions* 48 (2020) 537–546.
- [27] J. Potet, D. Beran, N. Ray, G. Alcoba, A. G. Habib, G. Iiyasu, B. Waldmann, R. Ralph, M. A. Faiz, W. M. Monteiro, et al., Access to antivenoms in the developing world: a multidisciplinary analysis, *Toxicon: X* 12 (2021) 100086.
- [28] P. V. P. Silveira, S. d. A. Nishioka, Non-venomous snake bite and snake bite without

- envenoming in a brazilian teaching hospital: analysis of 91 cases, *Revista do Instituto de Medicina Tropical de São Paulo* 34 (1992) 499–503.
- [29] 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.
- [30] F. Hu, P. Wang, Y. Li, C. Duan, Z. Zhu, F. Wang, F. Zhang, Y. Li, X.-S. a. Wei, Metaformer model with arcface loss and contrastive learning for snakeclef2023 fine-grained classification, in: *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*, 2023.
- [31] B. Bracke, L. Bloch, M. Bagherifar, C. M. Friedrich, Combination of object detection, geospatial data, and feature concatenation for snake species identification, in: *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*, 2023.
- [32] Z. Shi, H. Chen, C. Liu, J. Qiu, Metaformer model with arcface loss and contrastive learning for snakeclef2023 fine-grained classification, in: *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*, 2023.
- [33] A. Balana, Metric-weighted ensemble focal loss for snake species identification, in: *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*, 2023.
- [34] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: *International conference on machine learning*, PMLR, 2021, pp. 8748–8763.
- [35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [36] H. Bao, L. Dong, F. Wei, Beit: Bert pre-training of image transformers, *arXiv preprint arXiv:2106.08254* (2021).
- [37] Y. Fang, W. Wang, B. Xie, Q. Sun, L. Wu, X. Wang, T. Huang, X. Wang, Y. Cao, Eva: Exploring the limits of masked visual representation learning at scale, 2022. *arXiv:2211.07636*.
- [38] 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.
- [39] L. Yuan, Q. Hou, Z. Jiang, J. Feng, S. Yan, Volo: Vision outlooker for visual recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- [40] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, S. Xie, Convnext v2: Co-designing and scaling convnets with masked autoencoders, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16133–16142.
- [41] 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.
- [42] Y. Ho, S. Wookey, The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling, *IEEE access* 8 (2019) 4806–4813.
- [43] 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.
- [44] X. Liu, F. Shen, J. Zhao, C. Nie, Randommix: A mixed sample data augmentation method with multiple mixed modes, *arXiv preprint arXiv:2205.08728* (2022).

- [45] 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.
- [46] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4690–4699.
- [47] 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).
- [48] 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.
- [49] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al., Dinov2: Learning robust visual features without supervision, arXiv preprint arXiv:2304.07193 (2023).
- [50] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention, in: International conference on machine learning, PMLR, 2021, pp. 10347–10357.
- [51] B. Li, Y. Yao, J. Tan, G. Zhang, F. Yu, J. Lu, Y. Luo, Equalized focal loss for dense long-tailed object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 6990–6999.
- [52] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR, 2019, pp. 6105–6114.
- [53] M. Ilse, J. Tomczak, M. Welling, Attention-based deep multiple instance learning, in: International conference on machine learning, PMLR, 2018, pp. 2127–2136.