Overview of SnakeCLEF 2024: Revisiting Snake Species Identification in Medically Important Scenarios

Lukas Picek[1,2], Marek Hrúz[1] and Andrew M. Durso[3]

[1]*Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Czech Republic*
[2]*Inria, LIRMM, University of Montpellier, CNRS, Montpellier, France*
[3]*Department of Biological Sciences, Florida Gulf Coast University, Florida, USA*

Abstract

The SnakeCLEF challenge serves as a major benchmark for evaluating the performance of AI-driven methods in snake species recognition on a global scale. The 5th edition of the SnakeCLEF challenge builds on last year's training data and extends the test set with new data from private collections originating from southern Africa. Similar to last year, SnakeCLEF 2024 focuses on (i) evaluating incremental improvements in automatic snake species identification, (ii) testing global generalization in three specific scenarios: India, Central America, and southern Africa, and (iii) assessing the impact of uneven error costs, such as mistaking a venomous snake for a harmless one. In this paper, we highlight the crucial importance of a robust automatic snake identification system, especially in resource-limited environments and in neglected regions, and its potential benefits for biodiversity conservation and global health. We present (i) a detailed description of the provided data, (ii) the evaluation methodology, (iii) an overview of the submitted methods, and (iv) insights gained from the 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 annual deaths and disabilities exceeding half a million due to venomous snakebites, a robust image-based system for snake species identification could significantly enhance eco-epidemiological data and treatment outcomes [1, 2]. Most herpetological expertise is concentrated in developed countries like those in North America and the EU, where snake diversity is relatively low and snakebites are not a major public health concern. In contrast, remote regions of developing countries with high snake diversity and many snakebites often need more expertise and data. Therefore, providing assistance with snake species identification has excellent potential for saving lives, especially in these high-need regions [3, 4]. Enhancing the ability to correctly identify snake species can directly impact the effectiveness of medical responses to snakebites, ensuring the correct antivenom is provided. Additionally, AI-powered tools can enhance education, helping populations in remote and neglected regions recognize dangerous species and take preventative measures.

The main challenge in identifying snake species arises from high intra-class variation and low inter-class similarities in appearance, influenced by factors such as geographic location, color morph, sex, or age. Besides, many species have evolved to have visual similarities to other species as a defense mechanism (e.g., mimicry). Interestingly, some snake species resemble species found on different continents without overlapping ranges. Therefore, leveraging information about an unknown snake's geographic origin could help with the identification process, given that only about 125 out of approximately 4,000 snake species co-occur in any specific location [5]. Species with broader distributions and more images tend to be more accurately predicted than rare species with limited photographic records [6]. This poses significant challenges when identifying species that are widespread across regions with sparse image data.

In order to assess the performance of AI-driven methods for snake species recognition on a global scale, we are organizing an annual competition – SnakeCLEF. The SnakeCLEF 2024 took part in the LifeCLEF [7], and FGVC-CVPR workshops and was hosted on the HuggingFace platform to attract participation from machine learning experts and boost idea exchange. This edition of SnakeCLEF focuses on (i) evaluating incremental improvements in automatic snake species identification, (ii) testing generalization across three specific scenarios (e.g., India, Central America, and southern Africa), and (iii) assessing error costs, particularly mistaking venomous snakes for harmless ones. In order to underscore the importance of a robust automatic snake identification system, especially in resource-limited regions, and highlight its potential impact on biodiversity conservation and global health outcomes, we provide (i) a comprehensive description of the dataset, (ii) details of the evaluation methodology, (iii) an overview of submitted methods, and (iv) insights derived from the results.

2. Objective

In previous SnakeCLEF editions [8, 9, 10, 11], we have learned that AI-based methods achieve high performance (F_1^C and Accuracy $> 90\%$) even in scenarios with long-tailed distributions and thousands of species but tend to underperform in medically important neglected regions. In light of that, we kept the training data and objective the same as last year but extended the test set. This allows us to easily compare this year's performance with last year's and allows for a more comprehensive generalization evaluation in neglected regions. The main objective was to create a classification model that returns a ranked list of predicted species for each real snake observation and minimize the danger to human life and the waste of antivenom if a bite from those observations were treated as coming from the top-ranked prediction. The classification model had to fit memory footprint limits and a prediction time limit (60 minutes) within a given server instance (Nvidia T4 small 4vCPU, 15GB RAM, 16GB vRAM).

2.1. Test Scenarios

This year, we focused on three geographic scenarios significant for global snakebites: (i) a biodiverse country historically focused on a few species of medically important venomous snakes (MIVS), (ii) a small tropical region comprising four adjacent countries with varying socio-economic statuses but shared snake fauna, and (iii) a large temperate region socio-economically similar to (ii) with more realistic test data.

Central America: In this region, snakebite is a significant public health issue, particularly affecting rural populations [12]. Panama reports the highest incidence of snakebite cases in the Americas [13]. Costa Rica is a global leader in antivenom research with robust epidemiological data and good healthcare quality [14, 15]. Together, these four countries host 43 MIVS, of which we cover 38, as well as 68% of non-venomous species. Central America has many species of venomous coralsnakes (genus *Micrurus*) and their harmless mimics [16], which are frequently misclassified by both humans and AI systems [17, 18]. However, most bites are caused by *Bothrops asper* (Terciopelo) [19], a pit viper with several harmless lookalikes (see Figure 1).

Figure 1: Left: *Leptodeira rhombifera* (Rhombic Cat-eyed Snake), a non-venomous species. **Right:** *Bothrops asper* (Terciopelo), a dangerous, venomous snake. Photos by ©leolara28, ©jcsullivan, ©add03e, and ©johnwmyles.

Southern Africa: Across South Africa, Namibia, Botswana and Zimbabwe, there are a total of 160 snake species (32 MIVS and 128 non-venomous species) [20, 21], of which we cover 121 [75.6%] (28 MIVS [87.5%] and 93 non-venomous [72.7%]). About 43% of species are endemic to one of the countries, and 31% of species are found in all the countries, with the remaining 26% being found in either two or three. This region is over 8× larger than Central America and is temperate and arid rather than tropical. Snakebite is a public health issue in rural areas, but snakebite epidemiology in Africa generally lags behind other regions [22, 23]. South Africa, the most developed country, has lower per-capita incidence and mortality. There is a high potential for generalization to adjacent countries, as political boundaries do not align with ecoregions.

Figure 2: Top: *Philothamnus semivariegatus* (Western Natal Green Snake), a non-venomous species. Photos by *©carasylvia, ©happyasacupcake, and ©gbmalawi*. **Bottom:** *Dispholidus typus* (Boomslang), a dangerous, venomous snake. Photos by *©garydewinnaar, ©gusbenson, and ©alexsoethoudt*.

India: India is the global leader in snakebites [24] with 40,000–60,000 annual deaths [25]. Although >75 MIVS are found in India [26], the "Big Four"—spectacled cobra (*Naja naja*), common krait (*Bungarus caeruleus*), Russell's viper (*Daboia russelii*), and saw-scaled viper (*Echis carinatus*) — are the most well-known and widespread and receive the most research attention and focus [27]. In this challenge, we include 44 MIVS (Big Four + 40 others) and 51% (128/253) of non-venomous snakes found in India, including many visually similar species (see Figure 3).

Figure 3: Top: *Eryx conicus* (Rough-scaled Sand Boa), a non-venomous species. Photos by *© murugeshgeo, © vojtechvita, and © knakrani*. **Bottom:** *Daboia russelii* (Russell's Viper), a dangerous, venomous snake. Photos by *©ravinaidu, ©murugeshgeo, and ©jayantdeshpande*.

3. Dataset

Same as in the previous edition [11], the dataset was assembled from three sources. The training, and validation subsets were derived from observations submitted to the citizen science platforms iNaturalist and HerpMapper. The competition test set (*Private Test*) consists of private images from individuals and natural history museums, which have not been published online, and thus, participants do not have access to this data.

In total, the collected data include approximately 110,000 real snake observations of 1,784 species with community-verified species labels. The number of selected species covers roughly half of all existing snake species, which makes it the dataset with the biggest species and geographical coverage. Compared to previous editions (i.e., 2020–2022), there has been a significant increase in observations from remote geographic areas that previously had few or no samples, improving the distribution of species data across various countries. The dataset comparison with previous SnakeCLEF editions is available in Table 2. Despite coverage improvements, there are still fewer observations and poorer species coverage in some developing countries.

Apart from the image data, the dataset includes information about the medical importance of each species, indicating whether a species is venomous or not. A country-species relevance matrix is also provided. Detailed statistics of the dataset are presented in Table 1, with a direct comparison to previous editions of the SnakeCLEF datasets shown in Table 2.

Table 1
SnakeCLEF 2024 dataset Statistics for each subset, including taxonomic and geographic coverage.

Subset	# of Species	# of Countries	# of Images	# of Observations
Training	1,784	212	168,144	95,588
↪ *iNaturalist*	*1,784*	*210*	*154,301*	*85,843*
↪ *HerpMapper*	*889*	*119*	*13,843*	*9,745*
Validation	1,599	177	14,117	7,816
Private Test	199	12	8,865	4,226
↪ *India*	*76*	*1*	*2,892*	*2,395*
↪ *Central America*	*107*	*4*	*5,188*	*1,370*
↪ *Southern Africa*	*80*	*4*	*786*	*462*

Table 2
SnakeCLEF datasets statistics comparison. All editions. Taxonomic and geographic coverage.

Dataset	Species	Images	Observation	Countries	min / max samples
SnakeCLEF 2020 [8]	783	259,214	×	145	19 / 14,433
SnakeCLEF 2021 [9]	772	386,006	×	188	10 / 22,163
SnakeCLEF 2022 [10]	1,572	318,532	187,129	208	5 / 6,472
SnakeCLEF 2023–2024	1,784	218,615	122,801	212	3 / 2,079

4. Evaluation Protocol

Similarly, as in the previous edition [11], we ask participants to provide a code that predicts snake species on unseen snake observations. Participants had to submit their solutions as HuggingFace models and submit them through the competition space. Participants could run any model or architecture but must fit limits for memory footprint and a prediction time limit (60 minutes) within a given HuggingFace server instance (Nvidia T4 small 4vCPU, 15GB RAM, 16GB VRAM). For an easier start, we provided a sample submission based on ONNX and PyTorch.

4.1. Evaluation Metrics

To allow a direct comparison to last year's performance, we used the same metric, which motivates research in recognition scenarios with uneven costs for different errors, such as mistaking a venomous snake for a harmless one. Last year, we introduced two metrics that consider *venomous* $\longleftrightarrow$ *harmless* confusion with different error costs. Those are defined as follows:

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

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

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:

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

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 accuracy of different types of confusion.

$$M = \frac{w_1 F_1 + w_2 C_{h \to h} + w_3 C_{h \to v} + w_4 C_{v \to v} + w_5 C_{v \to h}}{w_1 + w_2 + w_3 + w_4 + w_5}, \tag{3}$$

where $w_1 = 1, w_2 = 1, w_3 = 2, w_4 = 2$, and $w_5 = 5$ are the weights of individual confusions. $C_{v \to h}$ is the percentage of wrongly classified venomous species as a harmless species. $C_{h \to v}$ is the percentage of wrongly classified harmless species as a venomous species, $C_{v \to v}$ is the percentage of wrongly classified venomous species as another venomous species, $C_{h \to h}$ 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 ranges from 0% to 100%. A score of 0% is achieved if every species is misclassified, especially if all harmless species are mistaken for venomous ones and vice versa. Conversely, a score of 100% (where every species is correctly classified) means every P_i must be zero by definition.

4.2. Timeline

The SnakeCLEF 2024 competition was announced together with the dataset release on February 29, 2023, through the LifeCLEF, HuggingFace, and FGVC challenge pages. Anyone was allowed to register and participate in the competition. The competition deadline was May 24, making the competition duration approximately 3 months. The test data was kept secret, as was the case last year. Participants were allowed five submissions per day. Two weeks before the deadline, this limit was increased to 10 submissions per day. After the submission phase closed, all participants could submit post-competition submissions to evaluate their ablation studies.

4.3. Working Notes

All participants were encouraged to submit both their code and a technical report (Working Note) detailing the information required to reproduce their results. Each Working Note was reviewed by 2–3 reviewers with strong publication records in computer vision and machine learning, ensuring high standards of reproducibility and quality. The review process was single-blind and allowed for up to two rounds of rebuttals.

4.4. Baseline

To allow an easier start for all participants and straightforward model evaluation, we provide a relatively strong baseline based on the ONNX Swin-v2-tiny model [28] wrapped into a HuggingFace repository that can be directly submitted for evaluation on the private test set. Besides the PyTorch-based submission, we also provide an example for submitting the PyTorch model that was primarily provided for another LifeCLEF [7, 29] competition – FungiCLEF 2024 [30]. The Swin-v2-tiny model was trained for 50 epochs with SeeSaw loss [31] using the FGVC framework. Other training hyperparameters are as follows: batch size of 128, initial learning rate of 0.01, SGD optimizer, plateau learning rate schedule, and Randaugment [32].

5. Challenge Results

In Figure 4, we report private leaderboard performance achieved by individual teams in terms of Track 1 (T_1) and Track 2 (T_2) metrics. The best-performing team in both tracks *upupup* [33], achieved 83.6% in Metric 1 and 1,840 in Metric 2 on the private part of the test set. Interestingly, only 5 teams outperformed the provided baseline. Unlike last year, when the top two participants achieved a competitive and comparable performance, this year's difference between the participants was relatively large, 3.6% and 640 points in terms of Track1 and Track2 metrics between the top1 and top2 teams. More precisely, in the case of the Track2 metric, the best-performing team reduced the relative error of the runner-up by more than 25%.
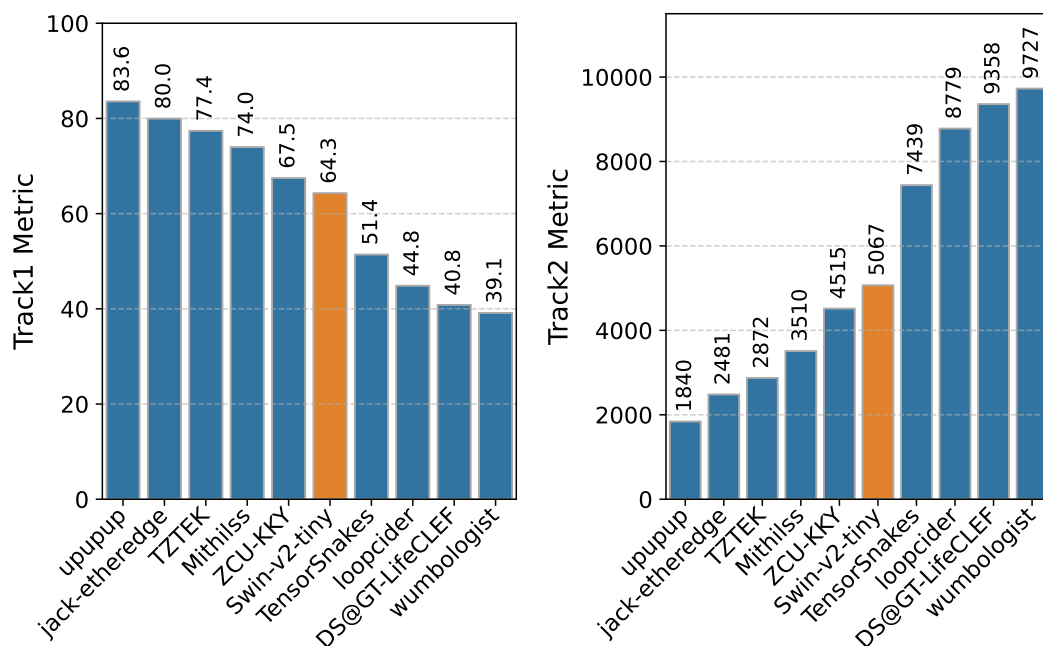

Figure 4: The official SnakeCLEF 2024 competition results; private leaderboard. We compare all 9 teams in terms of Track1 and Track2 metrics. Only 5 teams out of 9 provided a submission that outperformed the baseline. The orange color depicts baseline performance.

Comparing the performance of SnakeCLEF 2023 and 2024 participants (see Table 3) revealed a relatively large drop in performance from last year. This noticeable decline in the participants' performance is most likely related to the more stringent hardware limitations imposed in SnakeCLEF 2024. The hardware constraints introduced this year have posed significant challenges for the teams, as the computational resources allowed are considerably less than those available in the previous year's competition. None of the teams from SnakeCLEF 2023 would have been able to meet these new, stricter hardware requirements, highlighting the substantial impact these limitations have had on the ability to develop

and deploy high-performance solutions. Consequently, the stricter hardware constraints have likely necessitated adjustments in strategy and implementation, resulting in a performance drop as teams adapt to the new conditions. This change underscores the importance of hardware resources in ML competitions and the direct correlation between computational power and algorithmic efficiency and effectiveness.

Table 3
SnakeCLEF2023 and 2024 performance comparison. We compare the top 3 participants from SnakeCLEF2023 with the top 5 teams in SnakeCLEF2024.

Team name	SnakeCLEF	India		Central America	
		Track1	Track2	Track1	Track2
word2vector [34]	2023	89.74	648	93.25	260
BBracke [35]	2023	88.13	1,170	94.61	239
BAOfanting [36]	2023	85.47	1,363	93.80	250
upupup	2024	81.16	1,844	91.01	325
jack-etheredge	2024	76.70	2,657	89.11	401
TZTEK	2024	75.22	2,922	85.97	541
Mithilss	2024	71.26	3,719	85.69	549
ZCU-KKY	2024	64.83	4,591	78.51	933

Central America: In the Central America scenario, with snake species observations from Costa Rica, Nicaragua, Honduras, and Panama, the competition was "intense". The top four teams achieved comparable scores across both tracks' metrics, demonstrating a high level of performance and consistency. Specifically, in Track 1, these teams attained impressive accuracy rates ranging from 85% to 90%. Similarly, in Track 2, which measures performance based on a different set of criteria, the scores ranged from 358 to 504. These results reflect the teams' adeptness in handling the varied and complex data from the Central American region. The high Track 1 and low Track 2 scores highlight the robustness of the provided approaches and the effectiveness of their methods in adapting to the specific environment, such as Central America.

India: As with last year, the Indian scenario proved to be a significant challenge and was a deal breaker for many participants. Despite the first-runner-up team's strong performance in the Central America and Africa scenarios, where they matched the winning team in Track1 and Track2 metrics, their error in Metric 2 for the India scenario increased by around 50%. This highlights the unique and complex difficulties of the Indian scenario, possibly due to a diverse range of snake species, varied environmental conditions, and more complex data sets, although none of these seem fundamentally different from Central America (in fact, mimicry of Indian snakes is less exact than in the Neotropics). The consistent struggle with this scenario emphasizes the need for robust models that can generalize well across different geographic contexts and suggests that further research and targeted improvements are necessary to address these challenges effectively.

Southern Africa: The African scenario stands out due to its notably low Track 1 scores across the board, with a baseline of 56% and a top score of 75%, but relatively low Track 2 scores (i.e., 688 for baseline and 358 for the best team). These scores are particularly significant because the test data are the most realistic, having been drawn from completely independent sources and collected by hand from relevant scenarios, especially individuals in Southern Africa seeking snake identifications. This method of data collection ensures a high level of authenticity and relevance, making the results particularly indicative of real-world performance. Interestingly, the relatively low number of mimic species in Southern Africa might explain why the Track 2 metric wasn't as high as in India, where mimicry is more common. This scenario's unique challenges highlight the need for models that can accurately interpret and analyze data from diverse and independently sourced datasets, emphasizing the importance of adaptability and precision in developing effective snake identification tools.
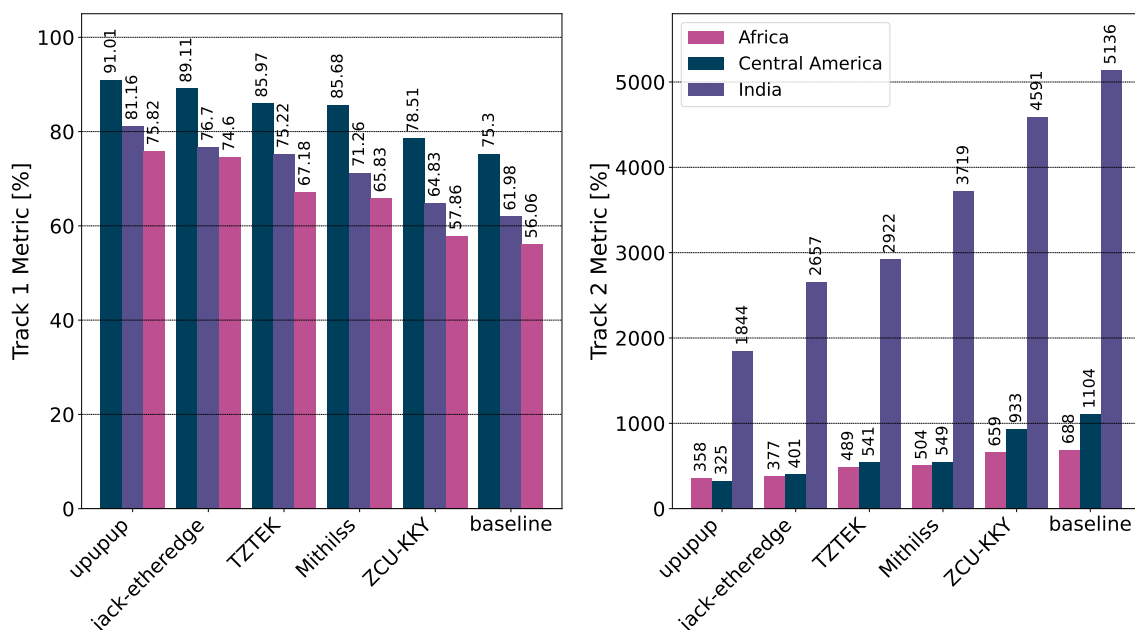
Figure 5: SnakeCLEF2024 performance comparison in different geographically-based scenarios.

6. Participants and Methods

This year, a total of 14 teams participated in the SnakeCLEF. However, just nine teams submitted solutions different from the baseline, and four submitted working notes. Details of the best methods and systems used are synthesized below, with further elaboration available in the individual working notes submitted by the participants [33, 37, 38, 39].

upupup (Top1) [33]: The main idea used by the team is that the venomous and the harmless snakes should be recognized by different models. This is in the spirit of the competition when different misclassifications are penalized. The rationale behind this approach is that common visual features are detected for both types of snakes, and the later stages of the network focus on fine-grained visual features of the venomous and harmless species, but then each branch uses different weights for the fourth stage of the ConvNeXt model. The first branch is used for the classification of all the species and is also responsible for the computation of the gating parameter. This parameter is estimated from the globally averaged features produced by the first three stages of the network, which are combined with the features produced by the first branch of the model. The authors use global maximum pooling for both feature maps individually and combine them afterward. Then, the sigmoid activation function is used to compute the scalar gating parameter. The second branch focuses on venomous snakes, while the third one focuses on harmless species. The gating parameter is used to decide which of these branches will be used for prediction and gradient-based weight optimization. This approach is used only in the training phase and is omitted for the inference. However, the authors show that this training setup helps the model to perform well overall. A combination of Seesaw loss [31] and CE loss is used for optimization. The team reached the best overall results in the competition with 81.16% and 91.01% in Metric 1 in India and Central America, respectively, and penalties of 1,844 pts and 325 pts in Metric 2.

jack-etheredge (Top2) [37]: The team uses a CAFormer [40] model in the final solution. They introduce a new venom loss, which considers the different penalties for misclassification. A cost matrix between the predicted class and the misclassification penalty is used to re-weight the softmax values of the prediction. This loss is similar to the weighted cross entropy loss but uses the costs directly instead of utilizing a weighted log loss. The addition of the venom loss significantly improves the performance

of the tested models across all metrics, even the F1 score. The team uses an ensemble of models trained on different data splits. Contrary to open set problems, the LogitNorm [41] did not improve the recognition rate. The team reached the second overall result in the competition with 76.7% and 89.11% in Metric 1 in India and Central America, respectively, and penalty of 2,657 pts and 401 pts in Metric 2.

ZCU-KKY (Top5) [39]: The team uses a Swin-v2 Tiny [28] model for the recognition. This is so that the model can be used on mobile devices for fast and practical inference. This is aligned with the mission of this competition when a fast response of the system is essential. The team combines two heads - one is for the species classification, and the other one is for venomous/harmless classification. This approach is similar to the "upupup" team and shows that handling the venomous and harmless species as two distinct groups produces better results. A combination of the Seesaw loss for species recognition and the binary cross entropy loss for venomousness detection is used for optimization. The one-hot target for the species identification CE loss is transformed by using a temperature parameter that is based on the different penalties set for the misclassification of venomous and harmless species. The authors show a relative improvement of around 9% when compared to the CE loss with one-hot targets. Even though the results are not as good as the results of other teams, they show an improvement over the baseline model by introducing the head responsible for venomousness recognition. The team reached the result of 64.84% and 78.51% in Metric 1 in India and Central America, respectively, and penalty of 4,591 pts and 933 pts in Metric 2.

DS@GT-LifeCLEF (Top14) [38]: Although the results of this team are not as good as the baseline approach, the presented methods show the potential of self-supervised models. The authors use the Dinov2 model (ViT-B/14 distilled) pre-trained on the LVD-142M dataset to extract the features from the image. Both the CLS token is used and a DCT of the patch features. The approach to represent the patches is such that the 16×16 patch representations with 768 dimensions are reshaped into a matrix of size 256×768, and a 2D DCT is computed on this signal. The coefficients of the top left 8×8 cosine representation (i.e., the smallest frequencies) are flattened into a 64-dimensional final representation. The authors use this representation together with the CLS token representation to visualize their discriminative capabilities. The DCT representation does not seem to form any useful clusters, unlike the CLS representation. However, the pipeline used to compute the DCT coefficients might not take full advantage of the patch representation. The reshaping of the patch tokens might be the culprit and, if handled in a more effective way, might provide some additional insides into the process of self-supervised learning. The authors train a linear layer on top of the representations to classify them into predefined classes. The results are sub-par, far behind the baseline experiment. The idea itself is interesting. However, the results do not show the full potential of the solution.

7. Conclusions and Perspectives

This paper presents an overview and results evaluation of the 5th edition of the SnakeCLEF challenge, organized in conjunction with the CLEF, LifeCLEF lab and the CVPR-FGVC11 – The 11th Workshop on Fine-Grained Visual Categorization, organized in conjunction with the CVPR conference. The SnakeCLEF challenge focuses on the challenging task of snake species identification from images, promoting advancements in fine-grained visual categorization and the development of robust machine learning models not just for global health. In summary, the fifth edition of the SnakeCLEF challenge has not only advanced state-of-the-art snake species recognition but also promoted a collaborative environment for researchers to exchange ideas and methodologies. The main outcomes from this year's evaluation are as follows:

Respect the penalties: Every team showed that customizing the loss according to given penalties leads to better results. The systems are more careful when dealing with possibly venomous snakes. Even though this is an expected behavior given the evaluation metric, it is proof that carefully designed penal-

ties lead to the desired properties of the systems. Other desired properties may be reached by carefully designing the evaluation metric. For example, it may incorporate the economic, social, or medical needs of different geographical areas. To conclude, the introduction of a custom loss that accounts for the different penalties for misclassification always helps. It seems to be the leading factor in improving the results.

Handle venomous and harmless species differently: A branching or multi-head approach to the classification of venomous vs. harmless species is another important factor in achieving better results. Although these mechanisms predominantly aim at optimizing the competition metric, they also improve the F1 scores. This is interesting because it shows that there are recognizable visual queues for venomousness, and it is best to model them explicitly. The winning team uses the detection of venomousness as a proxy task, which leads the model to specialize on different species in its branches. However, the main branch that provides the decision for the other branches then becomes a better species recognizer.

CNNs are back in the game? Maybe with no metadata.: The architecture of the model (CNN vs. Transformer) is not a major cause of the success. The winning team used a CNN, while the second team used a Transformer. Choosing the architecture according to other factors, such as run time or memory limitations, might be possible. Also, in the last year's competition, it has been shown that metadata helps a lot. In such cases, a transformer-based architecture is able to utilize them more efficiently. One may estimate that the future belongs to Transformers, however, for purely vision-based tasks one may not rule out CNNs.

Acknowledgments

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, L. Picek, S. Kahl, H. Goëau, V. Espitalier, C. Botella, B. Deneu, D. Marcos, J. Estopinan, C. Leblanc, T. Larcher, M. Šulc, M. Hrúz, M. Servajean, et al., Overview of lifeclef 2024: Challenges

on species distribution prediction and identification, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2024.

[8] L. Picek, R. Ruiz De Castañeda, A. M. Durso, P. M. Sharada, Overview of the snakeclef 2020: Automatic snake species identification challenge, in: Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, 2020.

[9] 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.

[10] L. Picek, A. M. Durso, M. Hrúz, 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.

[11] L. Picek, M. Šulc, R. Chamidullin, A. M. Durso, Overview of snakeclef 2023: Snake identification in medically important scenarios, in: Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, 2023.

[12] 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.

[13] 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.

[14] J. M. Gutiérrez, K. Maduwage, G. Iliyasu, A. Habib, Snakebite envenoming in different national contexts: Costa rica, sri lanka, and nigeria, Toxicon: X 9 (2021) 100066.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] R. Otero-Patino, Epidemiological, clinical and therapeutic aspects of bothrops asper bites, Toxicon 54 (2009) 998–1011.

[20] D. G. Broadley, F. V. F. M., Fitzsimons' Snakes of Southern Africa, Delta Books, Johannesburg, 1983.

[21] P. Uetz, P. Freed, J. Hošek, et al., The reptile database, 2024. URL: https://reptile-database.reptarium.cz/advanced_search.

[22] J.-P. Chippaux, Estimate of the burden of snakebites in sub-saharan africa: a meta-analytic approach, Toxicon 57 (2011) 586–599.

[23] S. Halilu, G. Iliyasu, M. Hamza, J.-P. Chippaux, A. Kuznik, A. G. Habib, Snakebite burden in sub-saharan africa: estimates from 41 countries, Toxicon 159 (2019) 1–4.

[24] 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.

[25] 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.

[26] R. Whitaker, A. Captain, Snakes of india. the field guide. 2nd edi, 2008.

[27] 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.

[28] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, et al., Swin

transformer v2: Scaling up capacity and resolution, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 12009–12019.

[29] A. Joly, L. Picek, S. Kahl, H. Goëau, V. Espitalier, C. Botella, B. Deneu, D. Marcos, J. Estopinan, C. Leblanc, et al., Lifeclef 2024 teaser: Challenges on species distribution prediction and identification, in: European Conference on Information Retrieval, Springer, 2024, pp. 19–27.

[30] L. Picek, M. Sulc, J. Matas, Overview of FungiCLEF 2024: Revisiting fungi species recognition beyond 0-1 cost, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[31] 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.

[32] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, Randaugment: Practical automated data augmentation with a reduced search space, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 702–703.

[33] P. Wang, Y. Li, B.-F. Tan, Y.-C. Zhou, Y. Li, X.-S. Wei, Multibranch co-training to mine venomous feature representation: A solution to snakeclef2024, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[34] F. Hu, P. Wang, Y. Li, C. Duan, Z. Zhu, F. Wang, F. Zhang, Y. Li, X.-S. a. Wei, Metaformer model with arcfaceloss and contrastive learning for snakeclef2023 fine-gained classification, in: Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, 2023.

[35] 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.

[36] Z. Shi, H. Chen, C. Liu, J. Qiu, Metaformer model with arcfaceloss and contrastive learning for snakeclef2023 fine-gained classification, in: Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, 2023.

[37] J. Etheredge, Generalizable training techniques for fine-grained long-tailed image recognition: Transferring methods optimized for fungiclef 2024 to snakeclef 2024, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[38] A. Miyaguchi, M. Gustineli, A. Fischer, R. Lundqvist, Transfer learning with self-supervised vision transformers for snake identification, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[39] M. Sieber, T. Železný, Do not lose to losses for snakeclef2024, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[40] W. Yu, C. Si, P. Zhou, M. Luo, Y. Zhou, J. Feng, S. Yan, X. Wang, Metaformer baselines for vision, IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).

[41] H. Wei, R. Xie, H. Cheng, L. Feng, B. An, Y. Li, Mitigating neural network overconfidence with logit normalization, in: International conference on machine learning, PMLR, 2022, pp. 23631–23644.