The Visual Computer (2026) 42:301
https://doi.org/10.1007/s00371-026-04482-2

RESEARCH

# Reptile species classification using Swin Transformers for biodiversity conservation in the Canary Islands

Ruymán Hernández-López[1] · Francisco A. Delgado-Rajó[2] · Sergio Celada-Bernal[1] · Alejandro Piñan-Roescher[1] · Carlos M. Travieso-González[1]

Received: 31 July 2025 / Accepted: 30 March 2026


**Abstract**

The Canary Islands, internationally recognized as a biodiversity hot spot, possess unique ecological characteristics including endemic reptile species that face substantial threats from invasive alien species. Particularly, concerning is the California kingsnake (*Lampropeltis californiae*), which exhibits remarkable adaptability and inflicts severe ecological damage on endangered endemic fauna. This proposal aims to address biodiversity conservation challenges in the Canary Islands ecosystem by advancing reptile species classification through the implementation of Swin Transformer architectures. The integration of these advanced neural network architectures with conservation biology provides an automated, precise tool for species identification across different taxonomic levels that can enhance monitoring and control strategies for biodiversity preservation worldwide. This technical approach addresses urgent conservation requirements in the Canary Islands while simultaneously establishing a methodological framework applicable to other ecological contexts facing comparable biodiversity threats. The methodology employed pre-trained Swin Transformer models, initially developing fine-grained reptile species discrimination capabilities through hyperparameter optimization on a limited-scale database (160 images, 40 per species) using nested cross-validation to ensure statistical rigor and independence. The optimized configuration was subsequently transferred to a substantially larger database (4,400 images) for binary classification distinguishing snake from non-snake reptiles, where training on only 400 samples achieved 99.38% accuracy on 4,000 independent test samples, directly addressing the invasive alien species monitoring challenge. The *Swin-B* Transformer model demonstrated robust performance, achieving up to 100% accuracy in multiclass species classification and 99.38% accuracy in the challenging, taxonomically broader binary snake versus non-snake classification. These results establish the transferability and scalability of the hyperparameter optimization methodology across datasets of substantially different magnitudes.

**Keywords** Swin Transformer · Reptile classification · Biodiversity conservation

## 1 Introduction

Biodiversity encompasses the variability of living organisms, ecosystems and their ecological complexes; this includes diversity within species, between species and of ecosystems [1], providing essential ecosystem services such as sustainable food production, freshwater supply, pollination, and

✉ Ruymán Hernández-López
ruyman.hernandez@ulpgc.es

Francisco A. Delgado-Rajó
paco.rajo@ulpgc.es

Sergio Celada-Bernal
sergio.celada@ulpgc.es

Alejandro Piñan-Roescher
alejandro.pinan@ulpgc.es

Carlos M. Travieso-González
carlos.travieso@ulpgc.es

[1] Signals and Communications Department (DSC), Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), Campus Universitario de Tafira, C/ Practicante Ignacio Rodríguez, 35017 Las Palmas de Gran Canaria, Canarias, Spain

[2] Telematic Engineering Department (DIT), Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), Campus Universitario de Tafira, C/ Practicante Ignacio Rodríguez, 35017 Las Palmas de Gran Canaria, Canarias, Spain

Published online: 18 May 2026

protection against natural hazards [2]. Natural ecosystems form intricate networks of species interactions maintained in a delicate dynamic equilibrium. Any alteration of ecosystem biota through invasion or extinction of species can significantly compromise the capacity of the ecosystem to perform vital functions [3].

Currently, biodiversity loss and ecosystem degradation represent one of the most critical threats to humanity in this decade [4]. This ecological crisis is primarily driven by five fundamental factors: land- and sea-use changes, overexploitation of natural resources, climate change, pollution and the proliferation of *Invasive Alien Species* (IAS). These factors have collectively contributed to a 60% decline in global wildlife populations over the last four decades. In the European context, approximately one-quarter of wild species face extinction risk, with Spain having the highest number of species on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species.

The Canary Islands are widely recognized as a biodiversity hot spot, featuring distinctive ecological traits due to their high habitat diversity, adaptive radiation, and the fact that have never been connected with any mainland [5, 6]. The reptilian fauna of the archipelago comprise a group of 15 living species with well-known insular distributions, 14 of which are endemic and only have limited capability to disperse across marine barriers [7]. However, the island ecosystems are particularly vulnerable to biological invasions due to their distinctive evolutionary conditions, including lack of predator adaptations and low genetic diversity.

Several IAS identified as concerning for the outermost region of the Canary Islands in the Spanish IAS catalogue include the Yemen chameleon, *Chamaeleo calyptratus* (Duméril & Duméril, 1851), the ball python, *Python regius* (Shaw, 1802), and the Cuban green anole, *Anolis porcatus* (Gray, 1840). However, the California kingsnake, *Lampropeltis californiae* (Blainville, 1835), represents a particularly noteworthy case in Gran Canaria. This species has demonstrated remarkable adaptability, thriving due to favorable climatic conditions, abundant food resources, and the absence of natural predators, resulting in rapid territorial expansion. The ecological impact of this IAS has been severe, particularly on endemic reptile populations. Notable affected species include the endangered Gran Canaria skink, *Chalcides sexlineatus* (Steindachner, 1891), and the critically endangered Gran Canaria giant lizard, *Gallotia stehlini* (Schenkel, 1901). The subsequent population decline of these endemic species has initiated trophic cascades with consequential effects on local agricultural productivity [8].

Despite regional authorities implementing various control strategies, including active search operations by specialized personnel, citizen participation in early warning systems, and strategic trap deployment, these measures have proven insufficient. Additional protocols have been established for cargo inspection at ports and airports, incorporating specialized canine detection units to prevent both the dispersal of this species within Gran Canaria and its potential spread to other islands of the archipelago. However, the complex interplay of technical, political, economic, and social aspects has hindered effective control of this ecological threat, highlighting the need for more innovative approaches to preserve the unique biodiversity of the archipelago.

On the other hand, in recent decades, the rapid advancement of *artificial intelligence* (AI) has revolutionized numerous fields, offering novel opportunities to address significant global challenges. Among these, biodiversity conservation stands out as a critical area where AI technologies can exert a substantial impact.

### 1.1 Related work

Recent advances in computer vision have enabled significant progress in automatic species recognition, with approaches employing convolutional neural networks (CNNs), *Vision Transformers* (ViTs), and their respective variants demonstrating effectiveness in this domain. While the primary focus of the following review remains on reptile identification systems, relevant studies involving other fauna have been incorporated due to the limited research applying cutting-edge technologies specifically to reptile recognition.

Species identification presents significant challenges due to morphological similarities across taxonomic groups [9]. This complexity has driven the evolution of automatic species identification in camera-trap imagery through distinct methodological approaches, where the two fundamental data processing tasks are animal presence detection and species classification [10]. These data processing principles extend to a wide range of applications, such as automated animal behavior recognition [11].

In the development of visual recognition systems for wildlife, Chen et al. [12] introduced what they considered the first fully automatic computer vision-based species recognition system using real camera-trap images. Their deep CNN-based approach surpassed the performance of the traditional bag of visual words model, proposed as the baseline species recognition algorithm, when evaluated on a public dataset comprising 20 North American species.

Over the past decade, advancements in neural network architectures and the availability of more extensive datasets have driven significant improvements in the accuracy and efficiency of these systems. Among CNN architectures, *ResNet* models, which incorporate residual connections to address the vanishing gradient problem [13], have demonstrated remarkable performance in wildlife recognition studies. For instance, Tabak et al. developed a highly accurate species classification system using *ResNet-18*, achieving 97.6% accuracy on species identification with over 3.7 mil-

lion images (27 classes) [14]. A more recent study using *Inception-ResNet V2* reached 94.82% accuracy, classifying 6 species across 1000 images per species [15].

Another survey highlighted the ecological significance of monitoring endangered species, particularly reptiles and amphibians which face threats including the introduction of invasive species, habitat modification, and climate change. The study classified three herpetofaunal groups (snakes, lizards, and toads) using pre-trained models *VGG16* and *ResNet-50*, achieving accuracies of 87% and 86%, respectively, outperforming the self-trained CNN model at 72%, thus demonstrating the superiority of the transfer learning approach [16].

Continuing with this approach, advances in wildlife recognition have led to more specialized research in reptile identification. A survey [17], which achieved 93% accuracy in classifying 14 Indonesian reptile species, aimed to determine the right CNN model for obtaining the best performance. Yet, while architectures can be customized through variations in layer count, composition, ordering, and activation functions, selecting the optimal architecture depends heavily on the specific dataset and classification task requirements [18].

Going a step further, works in species-specific reptile identification have yielded notable results. For example, Patel et al. achieved 75% accuracy across 9 racer snakes from the Galápagos Islands combining R-CNN (region-based CNN) for object detection and CNN for image classification [19]. Similarly, Rajabizadeh and Rezghi conducted a comparative study for image classification, where the *MobileNetV2* model outperformed both *VGG16* model and traditional methods, reaching 93.16% accuracy in classifying 6 snake species of Lar National Park (Iran) [20].

In addition, research has also explored CNNs for distinguishing between venomous and non-venomous species classification, where automated visual identification systems can assist healthcare professionals in selecting appropriate antivenoms and support pharmaceutical studies for venom-derived drug development [21–23].

As deduced from the aforementioned surveys, CNNs have made substantial contributions to computer vision, but while CNNs have revolutionized computer vision, they have inherent limitations. The core of CNNs is rooted in the use of the convolution operator. Their operations are fundamentally local, providing only translational equivariance, and their fixed weights during inference restrict their ability to capture relationships between distant pixels [24].

To address these limitations, advances in neural architectures have led to the development of attention-based *Transformers*, which excel at processing long-range dependencies and global features while showing competitive performance in many tasks compared to traditional CNNs.

ViTs have further advanced the field, demonstrating error patterns that better align with human visual perception and revealing distinct operational characteristics from CNNs. However, while powerful for computer vision tasks due to their global image processing capabilities, lack the ability to capture local image details that CNNs excel at. This limitation has led to the development of hybrid architectures that combine both CNNs and Transformers, resulting in improved performance across various vision applications.

On the other hand, a key development in ViT architectures, which has proven highly effective in various computer vision tasks, is the Swin Transformer [25]. This hierarchical model uses shifted windows to efficiently scale to high-resolution images while maintaining computational efficiency. The Swin Transformer achieves this by limiting self-attention computation to non-overlapping local windows while also allowing for cross-window connection through the shifted window partitioning approach, which can effectively capture local and global contextual information [26].

A study [27] focused on wildlife monitoring demonstrated its effectiveness in handling long-tailed data distributions, where it naturally achieved high accuracy (88.76% for the *Wildlife Conservation Society* (WCS) [28] dataset and 94.97% for Snapshot Serengeti [29]) and performed better for most of the datasets even without specific imbalance-handling techniques. In contrast, another publication [30] on mammal and bird recognition, as well as taxonomy classification, demonstrated that the *ConvNeXt* architecture outperforms the Swin Transformer architecture, even when both models have the same number of parameters.

Recent studies have demonstrated the versatility of Tran sformer-based architectures across various ecological applications, from species monitoring to biodiversity conservation. Particularly, in the realm of species identification, ViTs have exhibited remarkable performance. Further extending this approach, Bolon et al. demonstrated the potential of ViT-based models for global snake identification [31].

Similarly, several innovative architectural adaptations of ViTs have emerged to address domain-specific challenges. These adaptations include, for example, the implementation of an ensemble model combining ViT and EfficientNet-B4 for snake species distinction [32].

Particularly noteworthy for this exploration focus are the achievements of Swin Transformer in species identification. A comprehensive comparison of various architectures, including CNN, ViT, Swin Transformer, and hybrid approaches for sika deer individual recognition, demonstrated the potential of the architecture [33]. Additionally, the integration of the *Fine-Grained Visual Classification Plug-in Module* (FGVC-PIM) with Swin Transformer showcased enhanced discrimination of key image regions while maintaining robust hierarchical feature extraction [34].

While previous works have extensively applied various deep learning architectures to species identification across multiple taxa, the specific implementation of Swin Transformer models for reptile classification focused on Canary Islands species represents a distinct contribution to the literature.

### 1.2 Proposed approach

This work advances reptile species classification by enhancing the identification of both endemic and invasive alien species, building upon previous work documented in *Reptile Identification for Endemic and Invasive Alien Species Using Transfer Learning Approaches* [35]. While the previous study employed *convolutional neural networks* (CNNs)-based architectures, this work implements *Swin Transformer* architectures to address biodiversity conservation challenges within the Canary Islands ecosystem.

The study implements multi-class classification—in the machine learning sense of distinguishing among more than two categories—using different architectures for the four distinct reptile classes from the *Multi-Species Database*, followed by comprehensive hyperparameter optimization to achieve optimal performance. The Swin Transformer architecture achieves high *Accuracy* in species classification through its capacity to capture intricate visual distinctions and perform hierarchical feature extraction.

The feature extraction and classification capabilities developed in discriminating individual reptile species at this fine-grained level are then strategically leveraged for a broader ecological application. Specifically, the best-performing architecture and configuration are transferred to the *Snake/Non-Snake Database* for binary classification—distinguishing between snake and non-snake reptile species from a total of six species. This approach directly addresses the increasing snake proliferation in Gran Canaria. The overall methodology of the workflow is summarized in Fig. 1.

The methodology demonstrates effectiveness in detecting morphological variations between species through detailed analysis of both local and global image features. This technical approach serves immediate conservation requirements in the Canary Islands while offering potential applications in other ecological contexts that face similar biodiversity challenges, thus presenting an effective approach for automated species identification particularly for rapid screening and monitoring support in conservation management contexts.

The innovative application of Swin Transformer architectures to reptile species classification represents a significant advancement in conservation biology, specifically addressing a gap in automated monitoring of reptile species relevant to the Canary Islands biodiversity. This *deep learning* (DL) approach introduces novel capabilities for taxonomic differentiation between snakes and non-snake reptiles, a distinction of critical importance for conservation management in these historically snake-free insular environments now facing invasive predator challenges. The methodology leverages hierarchical feature extraction capabilities inherent to Swin Transformers to capture complex morphological characteristics across diverse reptilian species, establishing a foundation for enhanced ecological monitoring systems that can effectively support biodiversity preservation efforts in vulnerable island ecosystems.

This paper is structured as follows: Initially, the materials and methods section describes in detail the resources utilized. Subsequently, the experimental methodology outlines the specific procedures and protocols implemented in this study. Following this, the results and discussion section presents the findings along with a critical analysis, interpreting their significance within the broader context of the field. Finally, the conclusion synthesizes the main findings and indicates their implications.

## 2 Materials and methods

This section outlines both the dataset structure and the techniques employed in this work. The diverse specimen collection forms the foundation for the survey, while deep learning approaches—including specialized neural network architectures and optimization strategies—were implemented for taxonomic identification. The computational techniques address the specific challenges of biological specimen recognition through targeted training methods and performance evaluation protocols.

### 2.1 Dataset

This study utilizes custom image collections of reptile species organized into two databases, the Multi-Species Database and the Snake/Non-Snake Database, for sequential experimental phases. Both databases are classified as private due to usage restrictions on portions of their constituent imagery. The Multi-Species Database is employed in the first experimental phase to perform species-level classification, where each class represents a distinct species, while the Snake/Non-Snake Database is utilized in the second experimental phase to conduct binary classification distinguishing between the two taxonomic groups represented in this database.

The dataset sizes reflect the available photographic documentation of these conservation-priority species, representing realistic constraints common in conservation biology applications, particularly for endemic species with restricted geographic distributions.

The complete data corpus comprises 4560 samples focusing on six species of ecological significance within the

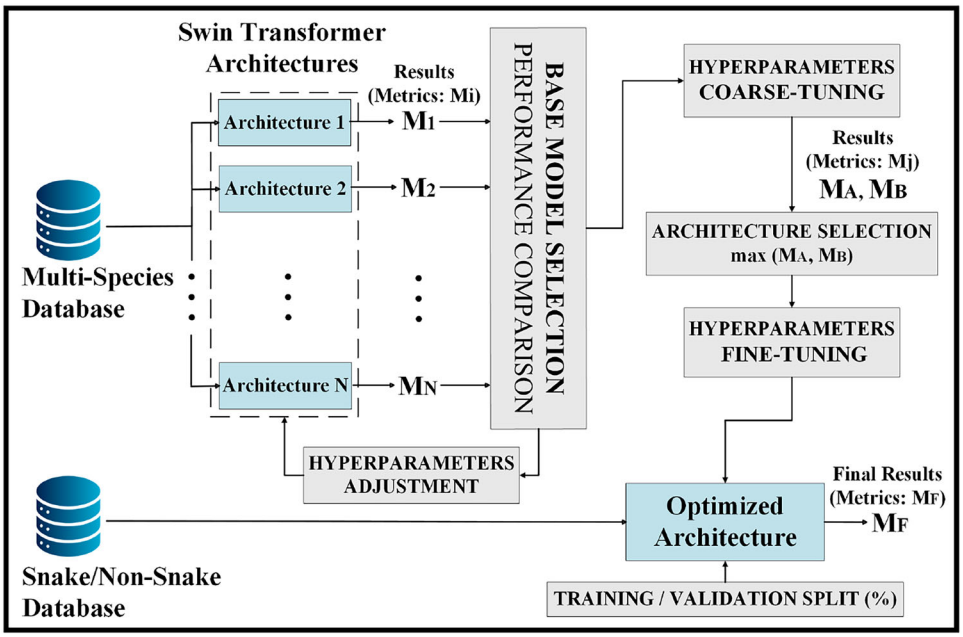



**Fig. 1** Conceptual schematic overview of the workflow

**Table 1** Summary of the databases used in the study

| Data collection overview | | | |
|---|---|---|---|
| Database | Class | Samples | Subtotal |
| *Multi-Species* | *G. stehlini* | 40 | 160 |
| | *C. sexlineatus* | 40 | |
| | *C. calyptratus* | 40 | |
| | *P. regius* | 40 | |
| *Snake/Non-Snake* | Snake | 2200 | 4400 |
| | Non-snake | 2200 | |
| Overall Total | | | 4560 |

Canary Islands ecosystem, including two endemic species and four invasive alien species.

Species selection criteria prioritized both conservation-priority native species and potentially disruptive invasive elements relevant to the biodiversity management challenges of the archipelago. All images in both databases underwent meticulous tagging and categorization according to their respective species classification, with the integrity and accuracy of these labels being essential for the performance and reliability of the classifier. Furthermore, the diversity and representative nature of these classes are fundamental to ensuring the capacity of the models to effectively generalize to new, unseen data.

From this corpus, appropriate datasets were derived to form the foundation for model development, evaluation, and classification system implementation. Table 1 provides a comprehensive overview of the samples that constitute the complete database.

An overview of the ecological relevance of these reptile species, along with representative specimens and their corresponding collection metadata, is provided in Appendix A.

#### 2.1.1 Composition and characteristics of multi-species database

The Multi-Species Database comprises images of four reptile species: the two endemic species, the Gran Canaria giant lizard (*G. stehlini*) and the Gran Canaria skink (*C. sexlineatus*), and two IAS of concern for the outermost region of the Canary Islands, as specified in the Spanish catalogue of invasive alien species, the Yemen chameleon (*C. calyptratus*) and the ball python (*P. regius*).

The database consists of a balanced dataset containing 160 images, with 40 samples per species, organized into two distinct subsets: the *Good Insight Dataset* and the *Wild Dataset*. The former encompasses high-quality images that provide clear views of specimens from multiple angles, facilitating robust feature extraction during the training process. The latter contains challenging images acquired under suboptimal conditions, including inadequate lighting, partial occlusion, or multiple specimens within a single frame. This dual-dataset structure ensures that the system can both learn distinctive species characteristics effectively and maintain performance under real-world conditions.

This heterogeneous collection of images facilitates the assessment of model discrimination capabilities in reptile classification, with particular emphasis on species exhibiting significant morphological differences. The Multi-Species Database is identical to that employed in the previous study

[35], where additional detailed information and example images can be found.

#### 2.1.2 Composition and characteristics of Snake/Non-Snake Database

In contrast, the Snake/Non-Snake Database contains images of all six reptile species under study. The California kingsnake (*L. californiae*) imagery was provided by the technical and biological team of the Government of the Canary Islands for protected species and invasive alien species in the Canary Islands. This dataset comprises photographs of specimens documented within the Canarian archipelago, whereas images for the remaining species were sourced from the specialized biodiversity platform *iNaturalist* [36].

In the curation process of this database, samples were excluded based on several criteria: specimens photographed alongside other species, images with insufficient resolution or severe focus issues, photographs depicting only shed skin or skeletal remains, images showing only tracks rather than specimens themselves, photographs of deceased specimens or those exhibiting severe trauma, etc.

Special emphasis was placed on documenting the pattern diversity of *L. californiae* compared to other species in this work, given its status as the most prolifically spreading invasive reptile species in the Canarian ecosystems.

From the valid *L. californiae* samples, specimens were selected to maximize morphological pattern diversity. The dataset was carefully curated to enhance the ability of the classifier to identify all known phenotypic variations present in *L. californiae* populations from Gran Canaria. The analysis included both typically pigmented and amelanistic specimens, each type displaying ring-shaped, linear, and composite (combined linear and ring-shaped) patterns. While the wild-type linear pattern represents the most frequent phenotype in this dataset, a substantial proportion of amelanistic specimens was included due to the *founder effect* from the initial introduction of this IAS. That is to say, although amelanistic morphs are generally uncommon in native populations, their frequency is notably higher in the Canary Islands.

The sample distribution between *L. californiae* and the remaining species obtained from iNaturalist resulted in an imbalanced dataset. A subset of samples was selected from this collection to create a dataset that met the requirements for the second experimental phase. This phase focused on developing a binary classifier to distinguish between snake and non-snake species samples. For this purpose, a balanced dataset (Snake/Non-Snake Database) was created containing 2200 samples per class, for a total of 4400 samples, where the snake class consists of *L. californiae* and *P. regius* specimens, while the non-snake class includes *G. stehlini*, *C. sexlineatus*, *C. calyptratus*, and *A. porcatus*.

### 2.2 Pre-trained Swin Transformer

The original Swin Transformer architecture [25] addresses fundamental challenges in vision applications through a hierarchical design and the shifted window partitioning approach. This hierarchical structure enables the model to capture multi-scale features effectively, making it particularly suitable for dense prediction tasks.

Building upon this foundation, the Swin Transformer v2 [37] introduces significant advancements over its predecessor, including enhanced scalability, training stability, and the capability to process higher-resolution inputs. These advancements encompass a *scaled cosine attention* mechanism and a *post-normalization* approach, which collectively enhance training stability and model scalability across different computational scales.

Both architectures encompass multiple model variants that share a common framework but differ in scale and complexity. The primary distinctions among variants lie in the number of transformer blocks and the overall parameter count, which directly influence computational requirements and model capacity. For instance, tiny, small, and base configurations represent different scales of the architecture, each designed to accommodate different computational constraints and performance requirements.

The Swin Transformer architecture implements window-based multi-head self-attention (W-MSA) and shifted wind ow-based multi-head self-attention (SW-MSA) modules within successive transformer blocks. Each block incorporates residual connections, layer normalization, and multi-layer perceptron components with GELU activation.

The self-attention mechanism within each window is computed according to the following equation:

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \qquad (1)$$

where $B$ represents the learnable relative position bias matrix that encodes spatial relationships between patches.

Furthermore, Swin Transformer incorporates patch merging modules that reduce spatial resolution by concatenating feature vectors from neighboring patches and applying a linear projection.

The Swin Transformer v2 architecture modifies the original formulation through implementation of post-normalization (applying Layer Normalization subsequent to each module), substitution of dot product attention with a scaled cosine attention mechanism, and introduction of a log-spaced continuous relative position bias approach.

Unlike the global attention used in ViT [38], Swin Transformer computes self-attention within local windows. For a window containing $M \times M$ patches, the computational com-

plexity of self-attention is reduced from:

$$\Omega(\text{MSA}) = 4hwC^2 + 2(hw)^2C \tag{2}$$

to the window-based multi-head self-attention (W-MSA):

$$\Omega(\text{W-MSA}) = 4hwC^2 + 2M^2hwC \tag{3}$$

where $hw$ represents the total number of patches across the entire image, and $C$ is the dimension of the embedding vectors per patch.

This window-based attention mechanism substantially reduces computational complexity from quadratic to linear scaling with image size, enabling efficient processing of high-resolution inputs. The shifted window approach alternates between successive layers, allowing cross-window information flow and enabling the model to capture both local fine-grained details and global contextual information effectively.

This approach leverages pre-trained models trained on large-scale datasets, enabling the effective application of previously learned features and significantly reducing training time and computational requirements while improving model accuracy. This technique is particularly advantageous when working with limited training data or computational resources, as it leverages robust feature extraction capabilities developed during pre-training.

The availability of pre-trained Swin Transformer variants facilitates the application of transfer learning principles to specific computer vision tasks.

### 2.3 Multilayer perceptron

The multilayer perceptron represents a fundamental computational model that mimics biological neural networks, operating on the fundamental principle of a feedforward flow of information through a hierarchically organized structure. This architecture comprises multiple layers, where information propagates unidirectionally from input to output through intermediate layers. The layers positioned between the input and output layers are termed hidden layers.

The input units serve a purely passive function, merely distributing signals to the first hidden layer without computational processing. All hidden units share the same activation function but maintain distinct learnable parameters, while output units represent simplified versions of hidden units. Each hidden unit transforms signals from the preceding layer into a single output signal for the next layer. Every hidden unit employs an activation function that is typically nonlinear and uniform across all hidden units. The output results from applying this activation function to the weighted sum of input signals plus an individual bias term. Output units typically employ either linear transformations or specific activation functions depending on the task requirements.

Mathematically, an MLP with L layers computes outputs through successive affine transformations followed by element-wise nonlinear activations. Each hidden layer $i$ applies a transformation of the form $h(i) = \sigma(W(i)h(i-1) + b(i))$, where $W(i)$ and $b(i)$ represent learnable weight matrices and bias vectors, and $\sigma$ denotes the activation function. The output layer typically applies a final linear transformation. This hierarchical composition enables MLPs to approximate complex nonlinear mappings through the universal approximation theorem [39].

## 3 Experimental methodology

This section presents the methodological framework employed to assess Swin Transformer architectures in reptile species classification. The evaluation approach details the network architecture, implements rigorous nested cross-validation with systematic data partitioning, establishes comprehensive performance metrics, and describes a two-phase experimental design to evaluate the proposed system, providing a robust foundation for analyzing model performance in identifying reptile species, including both endemic and invasive alien species in the Canary Islands ecosystem.

### 3.1 The network architecture

The network architecture of the proposed system consists of a pre-processing stage followed by a configurable modular architecture. This architecture incorporates a selectable Swin Transformer model and an optional multilayer perceptron. The system allows for two distinct operational modes: one that utilizes the MLP in cascade with the Swin Transformer and another that bypasses the MLP, using only the output of the Swin Transformer. This flexible configuration enables comparative analysis of performance under different architectural arrangements. The schematic diagram of the comprehensive modular architecture is illustrated in Fig. 2.

The preprocessing stage involves image standardization operations to ensure compatibility with pre-trained model requirements. This includes spatial resizing to match expected input dimensions and pixel value normalization to align with the training data distributions from large-scale datasets used during pre-training.

All images were processed using the RGB color model and stored in JPEG format to ensure consistent data representation. The preprocessing stage implemented standardized resizing to dimensions of $224 \times 224$ pixels, selected for optimal balance between detail preservation and computational efficiency. Additionally, pixel values were normalized using mean values of [0.485, 0.456, 0.406] and standard devia-

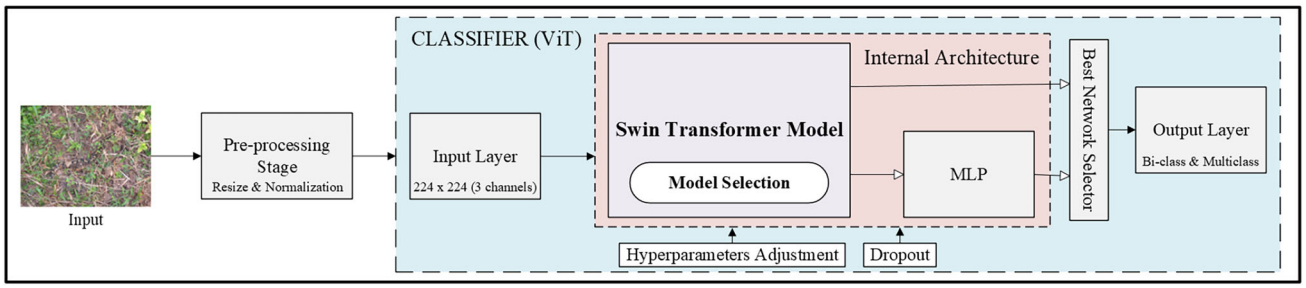



**Fig. 2** Representative diagram of the architecture

tion values of [0.229, 0.224, 0.225], corresponding to the RGB channel statistics of the ImageNet dataset. This preprocessing approach establishes dataset uniformity, mitigates brightness and contrast variations, and optimizes computational resource utilization for model training.

Subsequently, preprocessed samples are fed into the input layer, which adapts to the structure of the input data before entering the internal architecture of the classifier. The main component of the system is the Swin Transformer model, which serves as the backbone of the classifier architecture. The Swin Transformer variants encompass different architectural scales available in both the original and v2 versions, each offering different computational complexities and model capacities to accommodate various system requirements.

The implementation leveraged all Swin Transformer model variants available in *TorchVision v0.17*, encompassing the tiny, small, and base configurations from both the original and v2 architectures. All models were utilized in their pre-trained state, having been trained on *ImageNet-1K v1*, a large-scale dataset comprising 1000 object classes, with 1,281,167 training images, 50,000 validation images, and 100,000 test images. The parameter counts for each variant are as follows: *Swin-T* with 28,288,354 parameters; *Swin-S* with 49,606,258 parameters; *Swin-B* with 87,768,224 parameters; *SwinV2-T* with 28,351,570 parameters; *SwinV2-S* with 49,737,442 parameters; and *SwinV2-B* with 87,930,848 parameters.

Following the complete Swin Transformer backbone network, the processed feature representation follows one of two pathways: It can either be directed immediately to the output classification layer, or alternatively, it can be routed through an additional cascaded multilayer perceptron module before final classification. This optional MLP pathway exists externally to the Swin Transformer architecture itself and serves as a supplementary processing stage.

The MLP configuration consists of a fully connected five-layer structure with rectified linear unit (ReLU) activation functions applied to each neuron. The first layer comprises 1000 neurons, and the subsequent four layers implement a progressively decreasing neuron architecture to facilitate feature compression and adaptation to the target classification task. The neuronal distribution across layers is as follows: layer 1 with 1000 neurons; layer 2 with 512 neurons; layer 3 with 256 neurons; layer 4 with 128 neurons; and layer 5 with 64 neurons. The final output layer was configured with a number of neurons equal to the number of classes in the target dataset.

*Dropout* regularization was applied throughout the internal layers of the architecture. During training, the Dropout mechanism stochastically deactivates neurons with a probability determined by the rate parameter, thereby reducing model dependence on specific neuronal pathways and mitigating overfitting risks. To maintain activation magnitude consistency, the outputs of remaining active neurons are scaled by a factor of $1/(1-\text{rate})$, ensuring that the expected sum of outputs remains invariant across training and inference phases.

The methodology incorporated several critical hyperparameters: the *Learning Rate* (LR), the number of *Epochs*, and the *Patience* parameter for the Early Stopping mechanism when implementing the *Reduce Learning Rate On Plateau* scheduling strategy. Additionally, the super-convergence technique *One Cycle Learning Rate* [40] was employed as an alternative scheduling approach. The One Cycle LR scheduler systematically modulates learning rates throughout the training process, initially increasing to a maximum value before gradually decreasing, thereby facilitating faster convergence and achieving superior generalization performance.

The classification algorithms were implemented using *Python v3.9.19* and *PyTorch v2.2.1*, an open-source machine learning framework developed by *Facebook AI Research* (FAIR). PyTorch is specifically designed for deep learning and numerical computation, enabling the construction of dynamic neural network models through a *define-by-run* approach. This framework provides comprehensive tools and libraries for building, training, and deploying machine learning models, making it particularly suitable for the classification tasks undertaken in this study.

### 3.2 Nested cross-validation

In the first experimental phase, *nested cross-validation* was employed to provide a rigorous methodological approach for evaluating classification model performance while maintaining statistical independence during data partitioning. This technique effectively mitigates overfitting risks and delivers an unbiased assessment of predictive capabilities across diverse data subsets.

The methodology employs a comprehensive performance evaluation strategy that addresses inherent variability in model performance, particularly when confronting complex classification tasks with limited dataset dimensions, as is the case with the Multi-Species Database. By implementing multiple data partitioning and model assessment iterations, the approach generates a statistically robust representation of model generalization potential.

Nested cross-validation provides superior protection agai nst overfitting compared to other validation methods, including standard $k$-fold cross-validation, by maintaining strict separation between hyperparameter optimization and performance evaluation. This methodology is particularly valuable for limited-scale datasets, as it maximizes data utilization while ensuring that model selection and performance assessment are conducted on completely independent data partitions, preventing information leakage that could artificially bias performance metrics.

The nested cross-validation procedure implements a complete cross-validation framework within both outer and inner loops. In the outer loop partitioning, the dataset is segmented into multiple test set components, with the remaining data reserved for subsequent analysis. In the inner loop subdivision, within each outer fold, the data undergo further division into training and validation sets.

Both outer and inner loops were configured with $k = 5$ iterations, and the stratified sampling strategy was implemented to preserve original dataset class distribution, with or without data shuffling. Regardless of whether dataset distribution is employed with or without data shuffling, the number of samples per class remains consistent across each partition, with the key difference being that without data shuffling, samples are extracted with ordered indices in each partition, thereby preserving the original sequence of the data.

The nested cross-validation framework, comprising 5 independent test sets each evaluated across 5 distinct validation configurations (25 total iterations), provides robust statistical evidence of model generalization capabilities rather than data-specific overfitting.

The experimental protocol unfolded as follows:

1. Dataset segmentation into 5 distinct test groups, each containing 32 samples, collectively representing the entire dataset.
2. For each test group, the remaining samples were further subdivided into 5 unique validation and training sets.
3. Execution of 5 classification tasks per test group, utilizing distinct models trained on varied validation and training set combinations.
4. Comprehensive evaluation resulting in 25 unique classifications, each performed by a model trained on specific data subsets.

Figure 3 illustrates a representative example of the data partitioning structure employed in nested cross-validation. As observed in Fig. 3a, the dataset is divided such that in the first iteration (outer loop, iteration $k = 1$), the test set consists of the initial samples of each class (dataset distribution without data shuffling). In subsequent iterations (outer loop, iterations $k = 2$, $k = 3$, $k = 4$, and $k = 5$), the test set comprises subsequent partitions per class of the dataset.

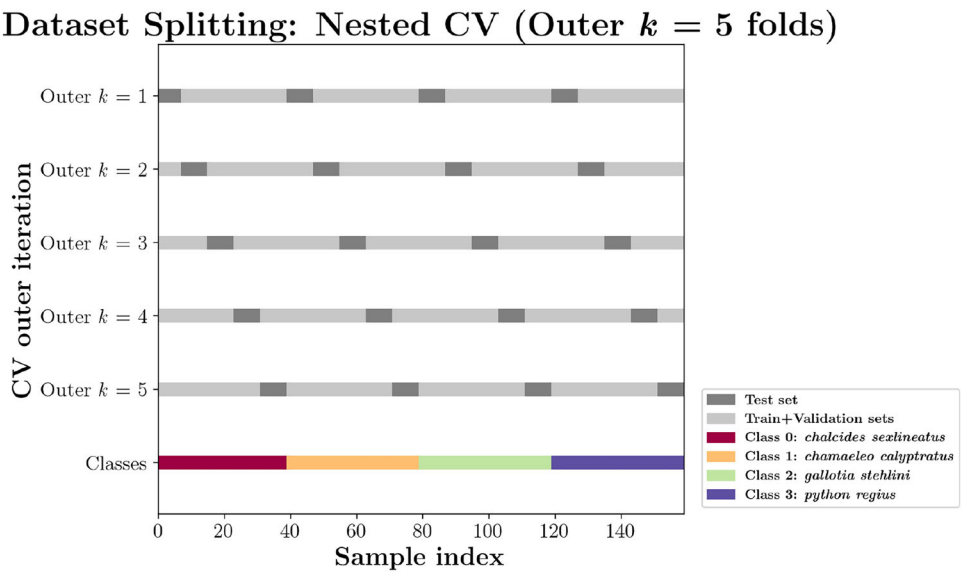
On the other hand, Fig. 3b illustrates the distribution of both validation and training sets for the inner loop (outer loop, iteration $k = 1$). The validation set follows the same sampling strategy within the inner loop as the test set did in the outer loop. Specifically, after extracting the test set samples, the remaining samples are partitioned so that: in the first inner iteration, the validation set consists of the initial samples of each class, and in subsequent inner iterations, the validation set comprises subsequent partitions per class. Each inner iteration features a different validation set, ensuring that distinct samples are used across iterations. The distribution of validation and training samples follows an analogous pattern for the remaining outer loop iterations (outer loop, iterations $k = 2$, $k = 3$, $k = 4$, and $k = 5$), with these sets being formed after extracting the respective test subsets for each iteration.

This nested cross-validation approach ensures: a comprehensive dataset evaluation with every sample being tested exactly once, and multiple model training and validation across diverse data portions. The methodology presents a statistically robust framework for assessing machine learning model performance, minimizing potential bias and enhancing generalizability estimation. The described cross-validation scheme was implemented using scikit-learn v1.4.2.

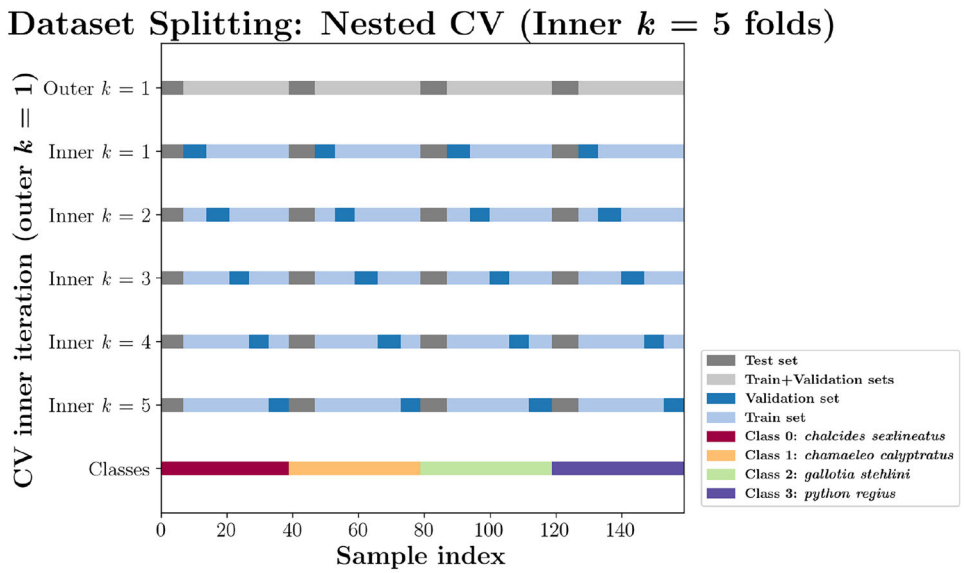
### 3.3 Performance metrics

In the domain of machine learning and image classification, the systematic evaluation criteria of classification models necessitate the implementation of robust metrics and comparative frameworks. These analytical tools facilitate a rigorous qualitative assessment of proposed models and enable a systematic comparison of their respective outcomes.

Specifically, this study employs performance metrics to evaluate the efficacy of algorithms in classifying and identi-




(a) Outer loop dataset splitting.




(b) Inner loop dataset splitting (outer loop $k = 1$).

**Fig. 3** Representative sample distribution across Nested Cross-Validation

fying various species in images, where each image is labeled as belonging to a certain class so that there are $N$ reference classes, which are mutually exclusive.

The primary performance indicators utilized include *accuracy*, *precision*, *recall*, and *F1 score*, with the formulation of these metrics fundamentally derived from the confusion matrix. The results are categorized as either positives ($p$) or negatives ($n$), creating four potential classification outcomes: true positive ($TP$), which correctly identifies the presence of a condition or attribute; true negative ($TN$), which correctly identifies the absence of a condition or attribute; false positive ($FP$), which incorrectly indicates the presence of a condition or attribute when it is absent; and false negative ($FN$), which incorrectly indicates the absence of a condition or attribute when it is present.

Accuracy is defined as the fraction of correct predictions made by the classifier out of the total number of predictions, and it is calculated according to Eq. (4):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

On the other hand, for each class $C_i$, where $i \in \{1, 2, \ldots, N\}$ and $N$ is the total number of classes, precision is computed as described in Eq. (5), while recall is determined as outlined in Eq. (6). The F1 score, which provides a harmonic balance between these two metrics, is subsequently

calculated according to Eq. (7).

$$Precision_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + FP_{C_i}} \tag{5}$$

$$Recall_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + FN_{C_i}} \tag{6}$$

$$F1\ Score_{C_i} = 2 \cdot \left( \frac{Precision_{C_i} \cdot Recall_{C_i}}{Precision_{C_i} + Recall_{C_i}} \right) \tag{7}$$

These metrics are calculated using standard contingency table elements, where $TP_{C_i}$, $FP_{C_i}$, and $FN_{C_i}$ represent the true positives, false positives, and false negatives specific to class $C_i$, respectively.

In multi-class classification, as is the case with the first experimental phase, where each observation receives a single label from four possible species from the Multi-Species Database, performance evaluation requires metrics that account for all classes. While accuracy simply measures correctly classified elements against total elements, the F1 score necessitates multi-class calculations of precision and recall integrated into the harmonic mean.

Various implementation approaches exist; however, given the use of balanced datasets in this study, where each class holds equal importance, it was essential to identify any potential bias in class classification relative to others. Therefore, the *macro-average (Macro-AVG)* approach was employed to obtain the results presented in the first experimental phase. This approach weighs all classes equally regardless of sample size, making it particularly relevant for balanced datasets where equal treatment of all classes is desired.

Macro-AVG calculates precision, recall, and F1 score separately for each class, then averages these values. These metrics can be obtained as shown in Eq. (8), Eq. (9), and Eq. (10), respectively.

$$Macro - AVG\ Precision = \frac{\sum_{i=1}^{N} Precision_{C_i}}{N} \tag{8}$$

$$Macro - AVG\ Recall = \frac{\sum_{i=1}^{N} Recall_{C_i}}{N} \tag{9}$$

$$Macro - AVG\ F1\ Score = \frac{\sum_{i=1}^{N} F1\ Score_{C_i}}{N} \tag{10}$$

The resulting metric evaluates the algorithm by giving equal weight to each class: High macro-average F1 score values indicate that the algorithm performs effectively across all classes, whereas low macro- average F1 score values signify that certain classes are poorly predicted by the model. All performance metrics were computed using scikit-learn v1.4.2.

**Table 2** Baseline hyperparameter setup used across all experiments

| Baseline hyperparameter configuration | |
|---|---|
| Hyperparameter | Setting |
| Loss function | *Cross-Entropy Loss* |
| Optimization algorithm | *Adam* |
| Batch size | 8 |
| Dropout rate | 0.2 |

### 3.4 Proposed experiments

The experimental design validates the proposed classification approach through two distinct phases, each addressing different scales and taxonomic resolutions while demonstrating the transferability of hyperparameter optimization across datasets of substantially different magnitudes.

The first experimental phase encompasses multiclass classification using the limited-scale Multi-Species Database (160 images across four species), incorporating comprehensive architectural comparison and systematic hyperparameter optimization through nested cross-validation. This phase identifies optimal model configurations while ensuring statistical independence and minimizing overfitting risks through rigorous data partitioning.

The second experimental phase strategically leverages the optimized configuration identified in phase 1 to conduct large-scale binary classification using the Snake/Non-Snake Database (4,400 images). This phase demonstrates the scalability of the methodology by applying the refined model to a substantially larger dataset with taxonomically broader classification objectives—distinguishing snakes from non-snakes rather than individual species. Critically, this phase trains on only 400 images while testing on 4,000 independent samples, providing strong evidence of genuine generalization and the practical applicability of the optimized configuration to conservation-priority monitoring tasks. The taxonomically coarser binary classification addresses the ecological reality that distinguishing invasive snakes from non-snake reptiles represents the primary conservation management requirement in the Canary Islands context, where the archipelago was historically snake-free.

To establish a baseline for systematic hyperparameter optimization, a set of fixed values was established. This configuration (Table 2) served as the constant substrate upon which subsequent iterative refinement and performance improvement were built.

During the initial assessments in the first experimental phase, the performance of the models proved significantly suboptimal. Subsequently, the hyperparameter tuning process commenced with an initial learning rate of $1 \cdot 10^{-3}$, implementing the LR reduction using the *Reduce Learning Rate on Plateau* scheduler, a *Patience* of 40 epochs, and

a maximum training duration of 100 epochs. This configuration yielded improved performance across the evaluated architectures.

Based on these improved results, a comprehensive architectural comparison was conducted to identify the optimal model for subsequent experimentation. Given the superior performance and stability of the cascaded *Swin-B* Transformer and MLP architecture, the *Swin-B* model was selected as the base model for subsequent experimentation. To investigate potential performance improvements, a systematic hyperparameter optimization process was implemented.

This optimization process commenced with coarse-tuning hyperparameter adjustments. The *Epochs* hyperparameter was increased from 100 to 200, and a comprehensive sweep of the *Patience* hyperparameter was conducted across the range $Patience \in \{10, 20, 30, \ldots, 100\}$. This methodical expansion of the hyperparameter search space was designed to examine whether the already strong performance of the *Swin-B* model could be further enhanced through extended training periods and varied convergence thresholds. The performance comparison included accuracy values, reported as means and standard deviations, using both sequential and shuffle data sampling strategies.

The increased *Epochs* value allowed the model to potentially discover more complex patterns, while the systematic variation of *Patience* values provided insights into optimal early stopping conditions. Following this patience parameter sweep, the configuration that demonstrated superior performance utilized the cascaded Swin Transformer and MLP architecture with data shuffling enabled and a patience value of 50 epochs.

To validate the stability and reproducibility of this optimal hyperparameter configuration, the same experiment was repeated four additional times, varying only the sampling methodology. Specifically, two repetitions employed data shuffling for sampling, while two additional repetitions utilized sequential sampling. This repetition protocol was designed to assess both the consistency of the hyperparameter configuration and the comparative impact of different sampling strategies on model performance.

Given the inconsistency observed between sampling strategies and the marginal differences between the two approaches, it was determined that subsequent experimentation would proceed without employing data shuffling, thereby simplifying the experimental protocol without compromising performance potential.

Subsequently, coarse-grained hyperparameter adjustments were conducted utilizing the *One Cycle Learning Rate* scheduler. This learning rate scheduler implements the one cycle learning rate policy, which differs from the previously used scheduler in that it updates the learning rate after every batch rather than at the end of each epoch. In addition, given that this scheduler is not chainable, the number of epochs remains fixed, thus eliminating the need for the *Patience* hyperparameter. In other words, unlike the *Reduce LR On Plateau* scheduler, which adaptively adjusts the learning rate or halts training based on performance plateaus, the *One Cycle LR* scheduler requires training with a fixed number of epochs.

In this stage of the experimentation, with the aim of improving the previously obtained results, a more comprehensive coarse-grained hyperparameter adjustment was conducted. Unlike the previous stage, this time an exhaustive sweep was implemented that encompassed both diverse learning rate values and different numbers of *Epochs*, thus allowing a more complete exploration of the hyperparameter space to optimize the performance of the models. This systematic approach enabled a thorough investigation into the effects of these hyperparameters on model convergence and overall performance.

The hyperparameter optimization involved an experiment for each combination of epoch count and learning rate. Epoch values were systematically varied from 10 to 100 in increments of 10, while learning rate values followed a logarithmic scale from $10^{-8}$ to $10^{-3}$, including both full ($10^n$) and half ($0.5 \cdot 10^n$) orders of magnitude.

Following this broad optimization phase, given the superior performance demonstrated by the standalone *Swin-B* Transformer architecture compared to the cascaded configuration, this architecture was selected for fine-tuning to achieve precise hyperparameter calibration. The concentration of optimal results within the lower learning rate range motivated the fine-grained hyperparameter adjustment in this specific high-performance zone. The new sweep was conducted using systematic incremental steps across this interval.

Following the comprehensive hyperparameter optimization conducted on the limited-scale dataset that identified the optimal configuration for the *Swin-B* Transformer architecture, the second experimental phase was systematically implemented to leverage this optimized model for large-scale dataset evaluation.

This comprehensive analysis focused on adapting the *Swin-B* Transformer architecture to a binary classification schema for image classification within the Snake/Non-Snake Database. The research methodology centered on conducting classification experiments with strategic variations in training-validation data partitioning, aimed at comprehensively assessing the performance of the model across diverse methodological approaches.

The experimental design incorporated the following critical methodological considerations:

1. Architecture Selection: The *Swin-B* Transformer architecture was selected, predicated on its demonstrated superior performance in comparative evaluations.

2. Learning Rate Selection: The learning rate was definitively established at $0.5 \cdot 10^{-5}$, predicated on its demonstrated superior performance in preceding experimental iterations.
3. Epochs Determination: The hyperparameter *Epochs* was fixed at 30, representing the minimum number of training epochs required to achieve perfect classification performance.

The primary objective of this experimental phase was to demonstrate the transferability and scalability of the hyperparameter optimization achieved on the limited-scale dataset to a substantially larger dataset, thereby validating the robustness and generalizability of the optimized neural network architecture through a rigorous, multifaceted evaluation approach.

By systematically modulating the training–validation data allocation strategy, the research sought to generate definitive performance metrics that would provide comprehensive insights into the classificational capabilities of the model.

From the total dataset of 4400 images (comprising 2200 snake and 2200 non-snake images), a strategic partition was implemented: 4000 images (2000 per class) were allocated to the test subset, while the remaining 400 images were designated for training purposes. This partitioning methodology maintained balanced class distribution across both subsets, ensuring statistical representativeness while facilitating rigorous model evaluation.

The experimental protocol adheres to a standard hold-out validation methodology, delineating clear training, validation, and testing phases. The 400 images for the training stage underwent internal partitioning across five experiments with progressive validation set proportions: 10%, 20%, 30%, 40%, and 50%. A stratified sampling approach was employed to maintain the original class distribution across training and validation subsets. The subsequent test stage leverages the entire 4000-image test dataset to comprehensively evaluate the generalization of the classifier performance.

## 4 Results and discussion

This section presents the outcomes obtained from the classification experiments conducted with the implemented models. An in-depth discussion of various comparative aspects follows based on these results.

The initial step involved evaluating different model architectures to identify the optimal foundation for the classification tasks. Table 3 provides a comprehensive comparison of accuracy metrics across different architectural configurations. This configuration enabled mean accuracy values exceeding 95% across the evaluated architectures and the data demonstrate that the cascaded *Swin-B* Transformer and MLP architecture achieves superior performance, exhibiting both the highest mean accuracy (98.75%) and the lowest standard deviation ($\pm$1.53%), Table 3

The reduced variability exhibited by the cascaded *Swin-B* Transformer and MLP configuration is particularly noteworthy, as it indicates more consistent performance across different experimental folds. This enhanced stability represents a significant advantage over alternative architectures and underscores the robustness of this particular configuration.

Table 4 presents a granular analysis of the performance metrics achieved by the cascaded *Swin-B* Transformer and MLP architecture throughout all experimental iterations. This detailed breakdown reveals that while optimal performance (100% accuracy) was not achieved in 10 out of 25 total iterations, the performance degradation was minimal. In these non-optimal instances, only a single sample out of 32 available test samples was misclassified, resulting in a 96.88% accuracy rate.

Furthermore, the convergence characteristics demonstrate the effectiveness of the implemented *Early Stopping* mechanism. The number of training epochs across iterations ranged from 41 to 47, closely aligning with the configured *Patience* value of 40. This indicates that the training process appropriately terminated when no further improvements were observed beyond the designated patience threshold.

Following the model selection process, systematic hyperparameter optimization was conducted to further enhance performance. Table 5 presents the performance comparison of standalone *Swin-B* Transformer and cascaded *Swin-B* Transformer with MLP architectures.

The patience parameter sweep yielded substantial performance improvements, as evidenced by the data presented in Table 5. The configuration that yielded the best performance achieved an accuracy of **99.88**% ($\pm$**0.61**), in Table 5 surpassing the results previously shown in Table 4. This superior performance was attained using the cascaded Swin Transformer and MLP architecture with data shuffling enabled and a patience value of 50 epochs.

Table 6 presents a comprehensive analysis of the performance metrics achieved by the cascaded *Swin-B* Transformer and MLP architecture throughout all experimental iterations. Analysis of the confusion matrix (Fig. 4) reveals only a single misclassification among the 32 test samples, specifically in outer loop iteration 2 and inner loop iteration 1, where a *G. stehlini* specimen was incorrectly identified as *C. sexlineatus*.

The stability verification through repeated experiments with this configuration yielded results shown in Table 7. Although these results are slightly lower than the initial experiment, and the standard deviations show some variability, the hyperparameter configuration demonstrated imperfect stability across repetitions.

**Table 3** Comparison of architectural performance for base model selection—multi-species database

| Comparative analysis of performance - accuracy (%): Mean (Std. Dev.) | | | | |
|---|---|---|---|---|
| Base model | Swin transformer | | Cascaded swin transf.+MLP | |
| *Swin-T* | 98.12 (±2.65) | | 96.00 (±3.48) | |
| *Swin-S* | 97.38 (±2.28) | | 97.62 (±2.38) | |
| *Swin-B* | 97.38 (±3.27) | | **98.75 (±1.53)** | |
| *SwinV2-T* | 95.88 (±3.39) | | 96.88 (±3.06) | |
| *SwinV2-S* | 96.75 (±3.59) | | 97.25 (±3.10) | |
| *SwinV2-B* | 97.62 (±2.54) | | 97.88 (±2.46) | |
| Hyperparameter Framework | | | | |
| LR Scheduler: *Reduce LR on Plateau* | LR: $1 \cdot 10^{-3}$ | *Epochs*: 100 | *Patience*: 40 | Sampling: *Shuffle* |

Bold values indicate the best result among those compared in each table

**Table 4** Detailed performance metrics of the experiment with cascaded *Swin-B* transformer and MLP architecture ($LR = 1 \cdot 10^{-3}$; $Epochs = 100$; and $Patience = 40$)—multi-species database

| Cascaded *Swin-B* Transformer+MLP - Performance Metrics | | | | | | |
|---|---|---|---|---|---|---|
| Loop Iter. ($k$) | | Elapse Epochs | Macro-AVG (%) | | | Accuracy (%) |
| Outer | Inner | | Precision | Recall | F1 Score | |
| 1 | 1, 2, 4, 5 | 41, 46, 47 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 3 | 41 | 97.22 | 96.88 | 96.86 | 96.88 |
| 2 | 1, 4, 5 | 41, 42 | 97.22 | 96.88 | 96.86 | 96.88 |
| | 2, 3 | | 100.00 | 100.00 | 100.00 | 100.00 |
| 3 | 1, 2, 3, 5 | 41, 43 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 4 | 42 | 97.22 | 96.88 | 96.86 | 96.88 |
| 4 | 1–5 | 41 | 97.22 | 96.88 | 96.86 | 96.88 |
| 5 | 1–5 | 41, 42 | 100.00 | 100.00 | 100.00 | 100.00 |
| Hyperparameter Framework | | | | | | |
| LR Scheduler: *Reduce LR on Plateau* | | | LR: $1 \cdot 10^{-3}$ | *Epochs*: 100 | *Patience*: 40 | Sampling: *Shuffle* |

**Table 5** Performance comparison of standalone *Swin-B* Transformer and cascaded *Swin-B* Transformer with MLP architectures with sequential and shuffle data sampling strategies—multi-species database

| Patience sweep for Swin-B archs. - accuracy (%): Mean (Std. Dev.) | | | | |
|---|---|---|---|---|
| *Patience* | Swin Transformer | | Cascaded Swin Transformer+MLP | |
| | Sequential | Shuffle | Sequential | Shuffle |
| 10 | 97.38 (±3.39) | 95.88 (±3.15) | 98.88 (±1.74) | 98.50 (±2.00) |
| 20 | 98.38 (±2.67) | 97.00 (±2.57) | 98.38 (±2.36) | 99.00 (±1.70) |
| 30 | 96.88 (±3.06) | 96.88 (±2.65) | 98.88 (±2.14) | 98.12 (±1.53) |
| 40 | 98.62 (±1.99) | 96.12 (±2.97) | 98.62 (±2.66) | 96.88 (±1.98) |
| 50 | 96.38 (±3.61) | 98.12 (±2.50) | 98.75 (±1.98) | **99.88 (±0.61)** |
| 60 | 96.25 (±3.54) | 97.62 (±2.03) | 98.50 (±2.67) | 98.75 (±1.53) |
| 70 | 97.38 (±3.27) | 96.62 (±2.64) | 98.38 (±1.79) | 97.50 (±3.06) |
| 80 | 96.88 (±3.31) | 96.62 (±2.64) | 99.00 (±1.92) | 98.38 (±2.36) |
| 90 | 97.62 (±3.10) | 97.12 (±2.15) | 98.00 (±2.92) | 98.12 (±2.34) |
| 100 | 98.12 (±3.19) | 97.38 (±3.72) | 97.50 (±2.65) | 98.75 (±1.77) |
| Hyperparameter Framework | | | | |
| LR Scheduler: *Reduce LR on Plateau* | | | LR: $1 \cdot 10^{-3}$ | *Epochs*: 200 |

Bold values indicate the best result among those compared in each table

**Table 6** Detailed performance metrics of the experiment with cascaded *Swin-B* Transformer and MLP architecture ($LR = 1 \cdot 10^{-3}$; $Epochs = 200$; and $Patience = 50$)—multi-species database

| Cascaded Swin-B Transformer+MLP - performance metrics | | | | | | |
|---|---|---|---|---|---|---|
| Loop Iter. ($k$) | | Elapse Epochs | Macro-AVG (%) | | | Accuracy (%) |
| Outer | Inner | | Precision | Recall | F1 Score | |
| 1 | 1–5 | 51, 52, 59 | 100.00 | 100.00 | 100.00 | 100.00 |
| 2 | 1 | 52 | 97.22 | 96.88 | 96.86 | 96.88 |
| | 2–5 | 51, 52 | 100.00 | 100.00 | 100.00 | 100.00 |
| 3 | 1–5 | 51, 52 | 100.00 | 100.00 | 100.00 | 100.00 |
| 4 | 1–5 | 51, 55 | 100.00 | 100.00 | 100.00 | 100.00 |
| 5 | 1–5 | 51, 59 | 100.00 | 100.00 | 100.00 | 100.00 |
| Hyperparameter Framework | | | | | | |
| LR Scheduler: *Reduce LR on Plateau* | | | LR: $1 \cdot 10^{-3}$ | *Epochs*: 200 | *Patience*: 50 | Sampling: *Shuffle* |

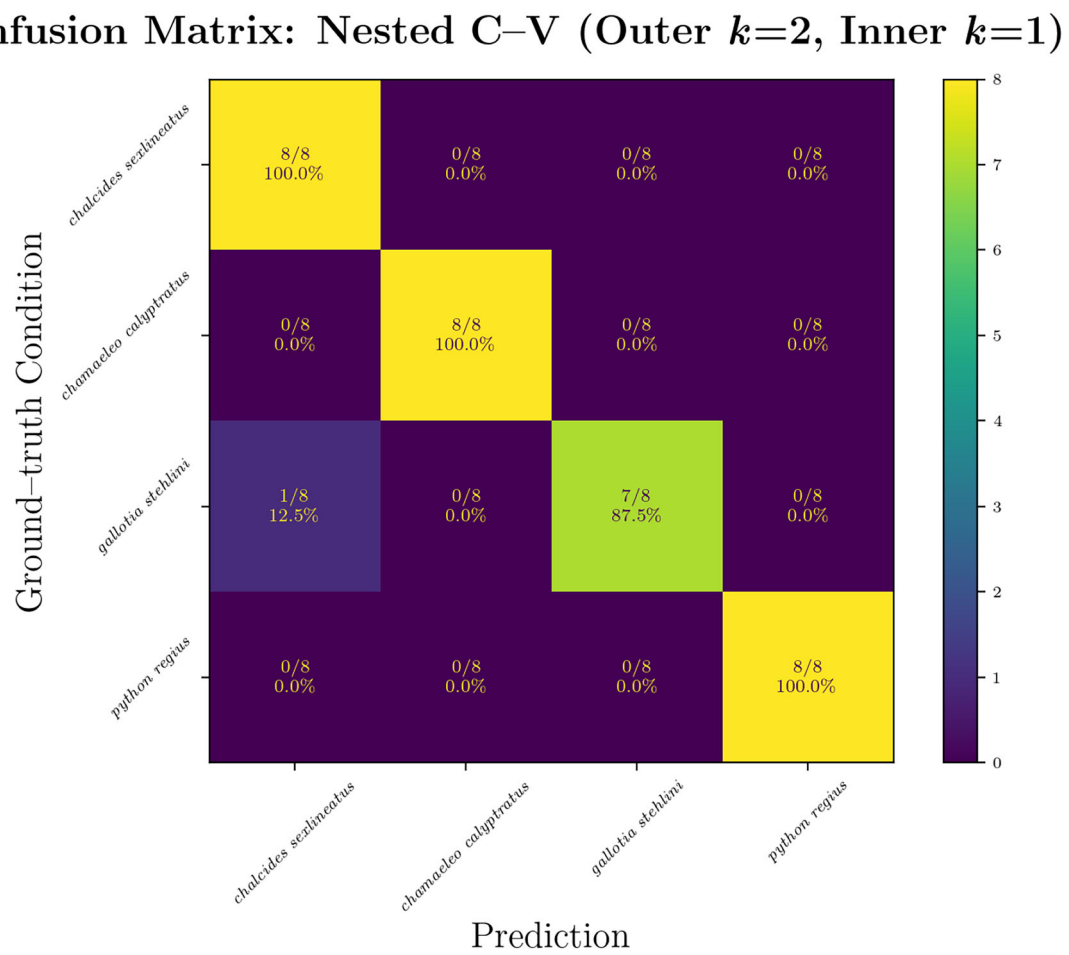



**Fig. 4** Misclassification confusion matrix with the cascaded *Swin-B* Transformer and MLP architecture ($LR = 1 \cdot 10^{-3}$; $Epochs = 200$; $Patience = 50$)

**Table 7** Experimental repetition results with *Swin-B* Transformer and MLP—multi-species database

| Repetition | Sampling | Accuracy (%) | |
|---|---|---|---|
| | | Mean | Std. Dev. |
| 1 | Shuffle | 99.50 | ±1.15 |
| 2 | | 97.75 | ±1.66 |
| 3 | Sequential | 98.38 | ±2.95 |
| 4 | | 99.62 | ±1.02 |
| Hyperparameter Framework | | | |
| LR Sch.: *Red. LR on Plateau* | | LR: $1 \cdot 10^{-3}$ | |
| *Epochs*: 200 | | *Patience*: 50 | |

Regarding the sampling strategies evaluated (sequential versus shuffled), the results from these repetitions and from Table 5 do not demonstrate a consistent pattern indicating that one strategy consistently outperforms the other. Certain architectural configurations with specific hyperparameter settings delivered superior results with data shuffling enabled, while others performed better without data shuffling.

The comprehensive hyperparameter sweep using the *One Cycle LR* scheduler yielded results presented in Fig. 5. Specifically, Fig. 5a displays the experimental outcomes for the *Swin-B* Transformer architecture, while Fig. 5b illustrates the mean accuracy results for experiments conducted with the cascaded *Swin-B* Transformer and MLP architecture.

This comparative evaluation facilitated the identification of the architectural configuration that yielded superior performance metrics. The results presented in Fig. 5 reveal a significant performance disparity between the architectures. The standalone *Swin-B* Transformer architecture demonstrates superior performance, with mean accuracy values ranging between 97.88% and 100.0%, whereas the cascaded *Swin-B* Transformer and MLP architecture exhibited a more variable range, spanning from 25.38% to 96.88%.

Optimal performance for the *Swin-B* architecture was achieved with specific learning rates: $1 \cdot 10^{-7}$, $0.5 \cdot 10^{-6}$, $1 \cdot 10^{-6}$, $0.5 \cdot 10^{-5}$, and $0.5 \cdot 10^{-4}$. Notably, the LR value of $0.5 \cdot 10^{-5}$ demonstrated optimal results across multiple *Epochs* values, specifically at points 30, 60, 80, and 90.

Based on these findings from the coarse-grained sweep, the fine-tuning process was focused on the high-performance region. As evidenced in Fig. 5a, the concentration of high mean accuracy values within a learning rate range of $1 \cdot 10^{-7}$ to $1 \cdot 10^{-5}$ provided the foundation for the subsequent fine-grained hyperparameter adjustment.

The results of this refined learning rate sweep are presented in Fig. 6, offering a more detailed perspective on the *Swin-B* Transformer architecture performance.

In this fine-tuning phase, the range of mean accuracy values shows an improvement compared to the results obtained in the coarse adjustment phase. The lowest mean accuracy value recorded is 98.38%, which was achieved in the experiment with a learning rate of $0.4 \cdot 10^{-5}$ and 60 epochs. However, this set of experiments also produced a result with a higher standard deviation than in the coarse adjustment phase, specifically ±3.75, which was observed in the experiment conducted with a learning rate of $0.6 \cdot 10^{-5}$ and 70 epochs.

As observed by comparing coarse-tuning (Fig. 5) and fine-tuning (Fig. 6), the learning rate of $0.5 \cdot 10^{-5}$ consistently emerges as the optimal hyperparameter, achieving the highest number of perfect classifications across both tuning stages. The comprehensive accuracy analysis in Fig. 7 explores the intricate relationship between adjacent learning rates and training epochs on model performance.

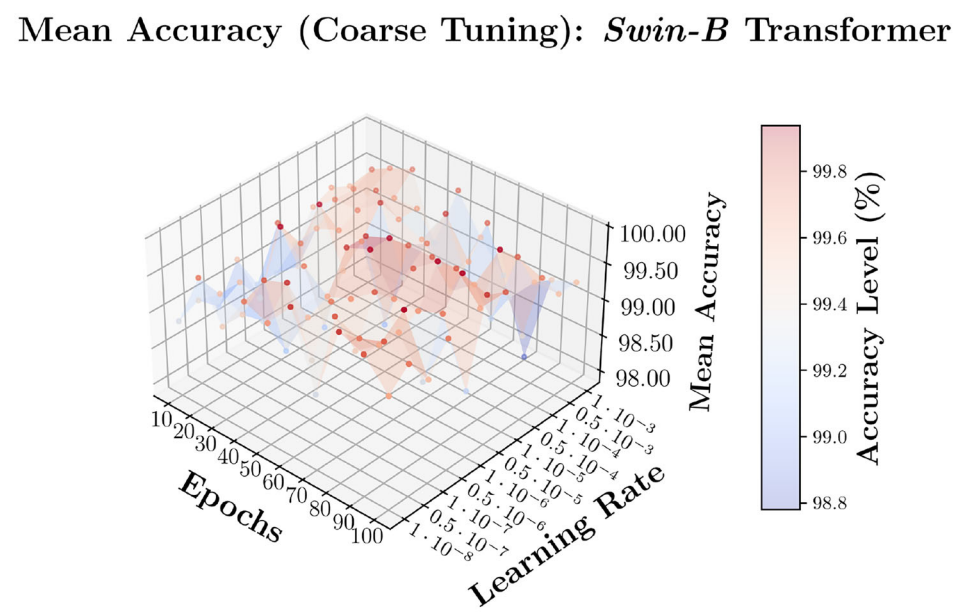
Illustrated in Fig. 7a, this learning rate demonstrates superior performance compared to adjacent rates. Multiple folds reached 100% accuracy, underscoring the critical importance of precise learning rate selection in model optimization. The box and whisker diagram in Fig. 7b provides deeper insights into the accuracy distribution. The data reveal remarkable model stability, with an majority of accuracy values clustering at 100%. Even the lower-performing iterations maintained high accuracy, with most values above 96.88%. The consistent performance across different epochs suggests robust model learning, particularly for the $0.5 \cdot 10^{-5}$ learning rate, which demonstrated generalization capabilities.

Comparative analysis with adjacent learning rates ($0.4 \cdot 10^{-5}$ and $0.6 \cdot 10^{-5}$) further emphasizes the effectiveness of the selected learning rate. While these rates achieved accuracies around 99%, they failed to match the consistent high-performance characteristics of the $0.5 \cdot 10^{-5}$ learning rate.

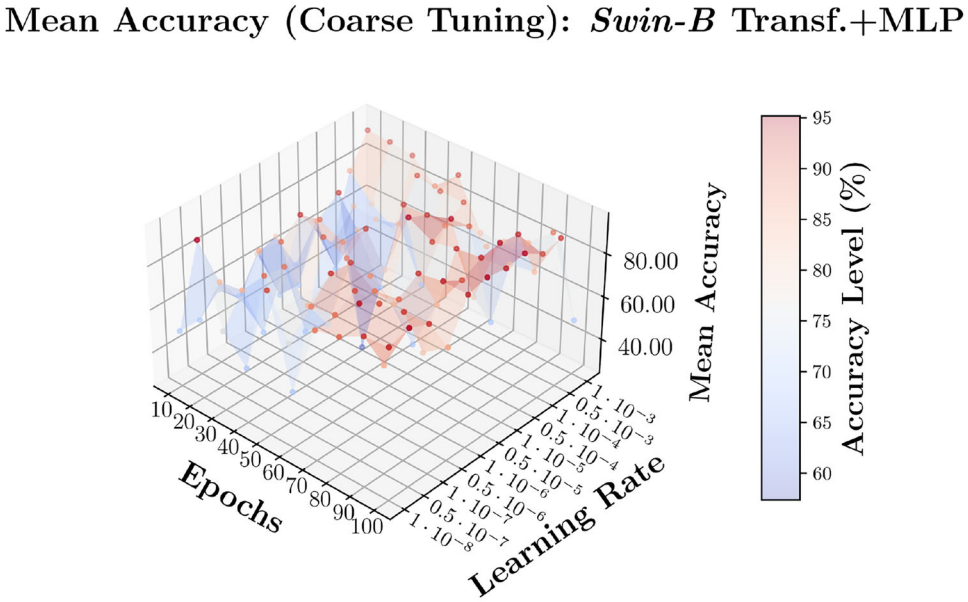
Having established the optimal hyperparameter configuration through the first experimental phase, the second phase evaluated the transferability of these optimized settings to binary classification on the complete Snake/Non-Snake Database.

Table 8 presents a comprehensive evaluation of the performance metrics of the classifier, elucidating the accuracy variations across different validation set proportions. The data presented demonstrate high performance of the classifier designed to distinguish between snake and non-snake images across all validation set ratios.

The model maintains remarkably consistent performance metrics regardless of the validation subset size, with accuracy consistently above 99% across all tested proportions. Precision and recall metrics exhibit balanced performance for both classes, indicating the ability of the model to correctly identify both snake and non-snake images with minimal error. The highest accuracy of 99.38% was achieved using a 20% validation set, though the difference in performance across all tested validation proportions is marginal (ranging from




(a) Mean accuracy heatmap for *Swin-B* Transformer architecture.




(b) Mean accuracy heatmap for cascaded *Swin-B* Transformer and MLP architecture.

**Fig. 5** Comparison of mean accuracy across learning rate and epoch sweep (LR: $10^{-8}$ to $10^{-3}$; *Epochs*: 10 to 100) for the two architectural variants (Coarse-Tuning)

99.08% to 99.38%). This suggests that the performance of the model is robust to changes in validation set size.

The misclassification rates remain consistently low across all configurations, with error rates ranging from 0.3% to 1.4% for both classes. The balanced performance between precision and recall for both classes indicates that the model does not exhibit significant bias toward either class. These quantitative results suggest that the binary snake vs non-snake classification model maintains consistent performance metrics across various validation set configurations.

Furthermore, to highlight the significance of this work within the current state of the art, it is essential to contextualize the results obtained in this study in relation to other investigations that have employed CNN or ViT for reptile species identification.

It is important to note that the following comparative analysis (Table 9) is provided for informational purposes rather than as a rigorous methodological comparison. Given that this study specifically focuses on reptile species relevant to the Canary Islands, the datasets and target species differ significantly from those utilized in previous research. This

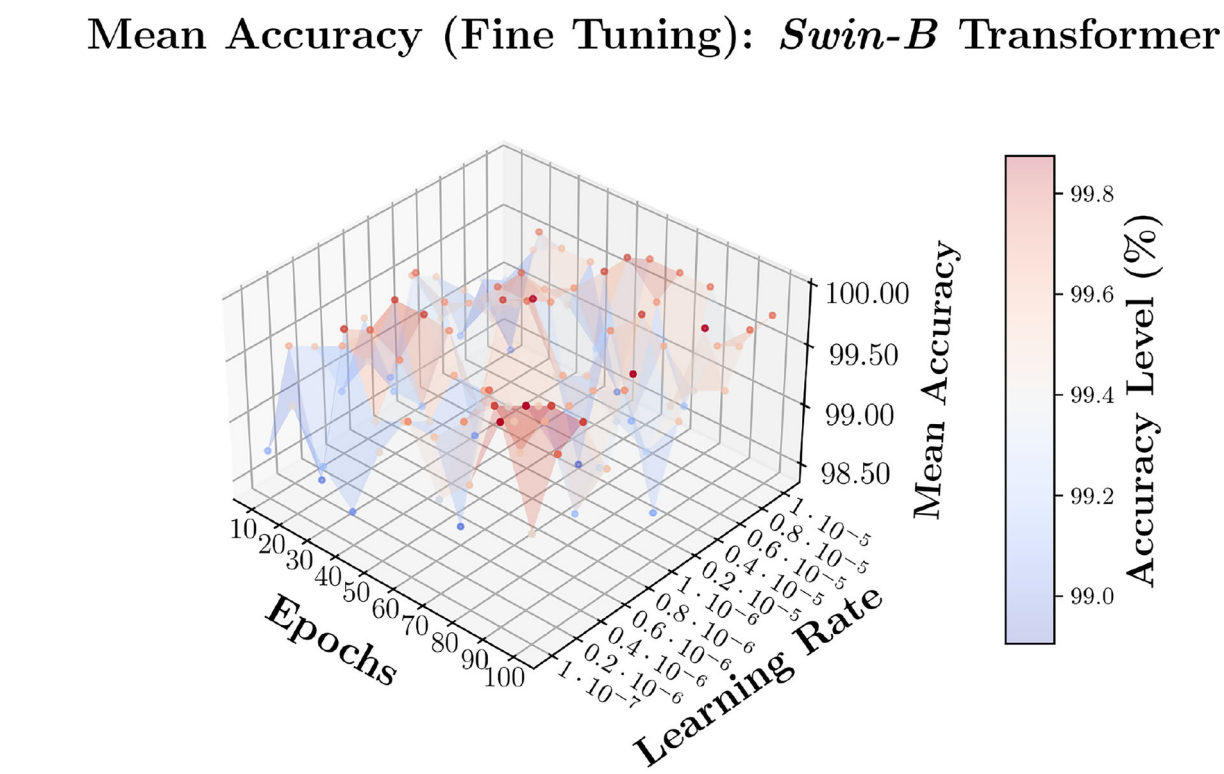
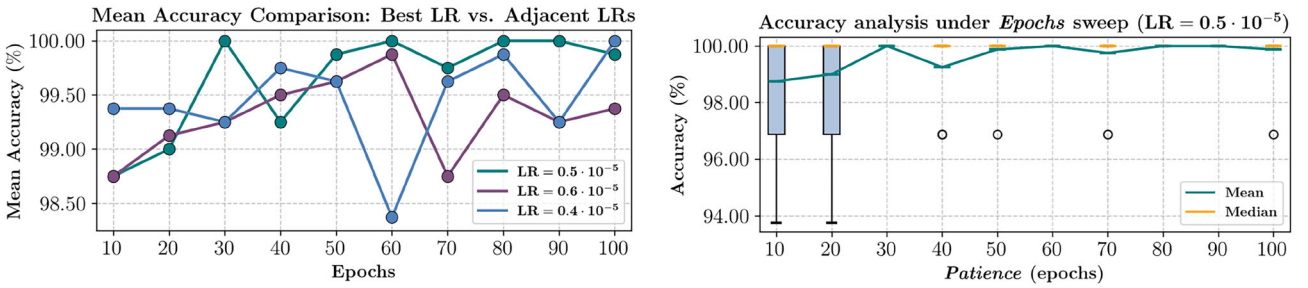



**Fig. 6** Mean accuracy heatmap for *Swin-B* Transformer architecture (Fine-Tuning)




(a) Comparative mean accuracy near $LR = 0.5 \cdot 10^{-5}$.

(b) Accuracy distribution and statistical metrics for the best LR.

**Fig. 7** Comprehensive accuracy analysis: Impact of learning rates and training epochs on model performance

fundamental distinction in data sources and taxonomic focus necessarily affects direct performance comparisons. Below are several studies that developed models for species classification that serve as reference points for understanding the broader research landscape in this domain.

Considering the aforementioned differences in datasets and taxonomic focus that preclude direct methodological comparisons, the comparative results presented in Table 9 demonstrate that our approach exceeds the performance reported in comparable studies focused on reptile identification.

The performance of the *Swin-B* Transformer model is significant considering the specialized nature of the dataset focusing exclusively on Canarian reptile species. While the number of classes in this experimental phase of the study is lower than in several previous studies, such as Bloch & Friedrich [32] and Bolon et al. [31], both of which classified 772 snake species on the SnakeCLEF dataset, the taxonomically broader binary classification presents unique challenges that distinguish it from species-specific categorization tasks.

Notable comparisons include Ahmed et al. [23], who reported 97.09% accuracy with VGG16 on 11,285 samples across 45 snake species, and Bolon et al. [31], who achieved 96% accuracy using a ViT-based model.

These results demonstrate that transformer-based architectures, specifically the *Swin-B* model, exhibit high efficacy in identifying reptile species in geographically constrained ecosystems, indicating potential applications for conserva-

**Table 8** Performance metrics for snake vs non-snake classification model evaluated across varying validation set proportions (10% to 50%)—Snake/Non-Snake Database

| Snake vs non-snake classification model performance across different validation set ratios | | | | | | |
|---|---|---|---|---|---|---|
| Train./Val. | Class | Precision (%) | Recall (%) | F1 Score (%) | Identifications | Misclassifications |
| 90%/10% | Non-snake | 99.70 | 98.90 | 99.30 | 1978 (98.9%) | 22 (1.1%) |
| | Snake | 98.91 | 99.70 | 99.30 | 1994 (99.7%) | 6 (0.3%) |
| | Macro-AVG | 99.30 | 99.30 | 99.30 | Accuracy (%) | 99.30 |
| 80%/20% | Non-snake | 99.70 | 99.05 | 99.37 | 1981 (99.0%) | 19 (0.9%) |
| | Snake | 99.06 | 99.70 | 99.38 | 1994 (99.7%) | 6 (0.3%) |
| | Macro-AVG | 99.38 | 99.38 | 99.37 | Accuracy (%) | 99.38 |
| 70%/30% | Non-snake | 98.96 | 99.50 | 99.23 | 1990 (99.5%) | 10 (0.5%) |
| | Snake | 99.50 | 98.95 | 99.22 | 1979 (99.0%) | 21 (1.1%) |
| | Macro-AVG | 99.23 | 99.23 | 99.22 | Accuracy (%) | 99.22 |
| 60%/40% | Non-snake | 99.50 | 98.65 | 99.07 | 1973 (98.7%) | 27 (1.4%) |
| | Snake | 98.66 | 99.50 | 99.08 | 1990 (99.5%) | 10 (0.5%) |
| | Macro-AVG | 99.08 | 99.08 | 99.07 | Accuracy (%) | 99.08 |
| 50%/50% | Non-snake | 99.15 | 99.50 | 99.33 | 1990 (99.5%) | 10 (0.5%) |
| | Snake | 99.50 | 99.15 | 99.32 | 1983 (99.2%) | 17 (0.8%) |
| | Macro-AVG | 99.33 | 99.33 | 99.32 | Accuracy (%) | 99.33 |
| Hyperparameter Framework | | | | | | |
| LR Scheduler: *One Cycle Learning Rate* | | | | LR: $0.5 \cdot 10^{-5}$ | *Epochs*: 30 | |

**Table 9** Comparison of best results from previous studies and the current work

| Research | Architecture | Accuracy (%) | Samples | Classes | Species |
|---|---|---|---|---|---|
| [19] | ResNet, VGG16 (CNN) | 75.00 | 247 | 9 | Snakes from Galápagos Islands (Ecuador) |
| [17] | CNN | 93.00 | 43,896[1] | 14 | Reptiles from Adiluhur Park (Indonesia) |
| [22] | CNN | 90.50 | 1766 | 2 | Snakes (non-venomous/venomous) |
| [20] | MobileNetV2 (CNN) | 93.16 | 594 | 6 | Snakes of Lar National Park |
| [32] | EfficientNet-B4, ViT-L | 91.17[2] | *Unspecified*[3] | 772 | Snakes |
| [21] | Inception-V3 | 90.00 | 240 | 2 | Snakes |
| [31] | ViT-based model | 96.00 | *Unspecified*[3] | 772 | Snakes |
| [16] | VGG16 (CNN) | 87.00 | 2700 | 3 | Snakes, lizards and toads |
| [23] | VGG16 (CNN) | 97.09 | 11,285 | 45 | Snakes |
| This work | *Swin-B* Transformer | 99.38 | 4400 | 2 | Endemic and IAS Canarian reptiles (snakes/non-snakes) |

[1] The dataset comprised 43,896 samples, including augmented images

[2] While this work identifies an ensemble of EfficientNet-B4 and ViT-L as the superior model (macro-AVG F1 score=82.88%), EfficientNet-B4 arch. delivers the highest accuracy value

[3] The study utilized the SnakeCLEF 2021 dataset containing 386,006 training images and 23,673 test images, though the exact final sample count after filtering is not specified

tion monitoring and management of endemic and invasive alien species in the Canary Islands.

The high classification performance achieved in this study reflects the morphologically distinctive nature of the target taxa within the Canary Islands context. The binary snake vs. non-snake classification exploits fundamental morphological differences between serpents and lizards that are readily apparent in photographic images. The six species in our databases represent distinct families with characteristic visual features, rather than cryptic species complexes requiring detailed morphological examination. This system is designed as a conservation monitoring support tool for rapid screening in a geographically constrained context, not as a universal snake identification system for regions with high species diversity, extensive mimicry complexes, or numerous cryptic species.

## 5 Conclusion

The innovative approach proposed in this study has proven successful in the application of Swin Transformer architectures for reptile species classification in the Canary Islands ecosystem, achieving exceptional accuracy rates for species recognition, including both endemic and invasive alien species identification. The implementation of these advanced deep learning models represents a significant contribution to biodiversity conservation efforts in this ecologically sensitive region, where invasive alien species such as *Lampropeltis californiae* pose substantial threats to endemic fauna.

The experimental results validate the superiority of the *Swin-B* Transformer model for reptile classification tasks, consistently achieving near-perfect accuracy across multiple testing iterations with various hyperparameters configurations. The model exhibits remarkable classification capabilities for both multiclass and binary classification scenarios, with particular effectiveness in distinguishing between snake and non-snake reptile species, addressing the critical conservation challenge of snake proliferation in Gran Canaria.

The observed error margins ($\leq$1.4%) across experiments with progressive validation set proportions represent performance bounds that should be considered when evaluating operational deployment for field-based snake identification applications, where environmental variability, image quality, and specimen positioning may differ from training conditions. Further evaluation under diverse field conditions would strengthen confidence in operational performance parameters.

This study addresses significant gaps in current biodiversity monitoring approaches, offering an automated, highly accurate tool for species identification that can enhance existing control strategies implemented by regional authorities.

Particularly noteworthy is the demonstrated transferability and scalability of the hyperparameter optimization methodology across datasets of substantially different magnitudes—from the limited-scale Multi-Species Database to the large-scale Snake/Non-Snake Database—establishing a robust methodological framework that maintains performance consistency regardless of dataset size. The demonstrated ability of the methodology to progress from individual species classification to broader taxonomic groupings provides a scalable approach for diverse conservation challenges. This represents a critical advancement for practical deployment in conservation contexts where data availability varies considerably across different taxonomic groups and geographical regions.

The integration of advanced transformer-based architectures with conservation biology establishes a methodological framework applicable to other ecological contexts facing similar biodiversity challenges, thereby strengthening the technological foundation for global biodiversity conservation efforts.

## Appendix A Ecological relevance of study species

This section outlines the ecological relevance of the reptile species included in this study, comprising both endangered endemic species and invasive alien species in the Canary Islands. These species form the foundation of the Multi-Species and Snake/Non-Snake databases used in this work.

### Endemic Species

- Gran Canaria giant lizard, *Gallotia stehlini* (Schenkel, 1901): According to the IUCN Red List assessment from June 27, 2024 (Version 2024-1) [41], this species has been classified as *Critically Endangered* (CR A3ce) based on projected population declines of 80% over three generations (spanning 21–24 years). The decline will likely persist until the California kingsnake colonizes most of the island, forcing remaining lizard populations into remote cliffs and uplands.
- Gran Canaria skink, *Chalcides sexlineatus* (Steindachner, 1891): In the same IUCN evaluation (Version 2024-1) [42], this species has been classified as *Endangered* (EN A2be; B1ab (iii, iv, v) + 2ab (iii, iv, v)) because the expansion of the California kingsnake has resulted in rapid declines, inferred to be at a rate greater than 50% in the 10 years since 2014.

### Invasive Alien Species

- Yemen chameleon, *Chamaeleo calyptratus* (Duméril & Duméril, 1851): This species represents a significant threat to biodiversity conservation in the Canary Islands. These reptiles function as reservoirs for multiple pathogenic bacteria, with 70% of the examined specimens harboring at least one pathogen. The prevalence of bacteria such as *Yersinia enterocolitica*, *Salmonella* species, and various mycobacteria indicates a potential risk of pathogen transmission to the native fauna.
  This bacterial reservoir capacity, combined with direct predation on endemic invertebrates, positions this invasive species as a substantial threat to the conservation of the endemic biodiversity of the archipelago [43].
- Cuban green anole, *Anolis porcatus* (Gray, 1840): This species represents a significant ecological threat through multiple mechanisms. Its broad diet of insects, arthropods, and small vertebrates leads to intense predation and resource competition with native species. The species has demonstrated remarkable invasive potential through

**Table 10** Collection metadata and source information for the specimens illustrated in Fig. 8

| Specimen collection data | | | | | | |
|---|---|---|---|---|---|---|
| iNaturalist observations | | | | | | |
| Fig.[1] | Species | Observer | Location | Obs. Date | License | URL[2] |
| Fig. 8a | *G. stehlini* | Justin Philbois | Gran Canaria, Spain | 10/08/2023 | CC0 | https://www.inaturalist.org/observations/177735479 |
| Fig. 8b | *C. sexlineatus* | Martiño Cabana | Gran Canaria, Spain | 25/02/2023 | CC-BY | https://www.inaturalist.org/observations/149640879 |
| Fig. 8c | *A. porcatus* | Francesco Cecere | Bartolomé Masó, Cuba | 29/07/2024 | CC0 | https://www.inaturalist.org/observations/236044088 |
| Fig. 8d | *C. calyptratus* | Jacky Judas | Asir, Saudi Arabia | 15/10/2021 | CC-BY | https://www.inaturalist.org/observations/98347793 |
| Fig. 8f | *P. regius* | Lucy Keith-Diagne | Akamkpa, Nigeria | 16/05/2018 | CC-BY | https://www.inaturalist.org/observations/67137520 |
| *L. californiae* observation | | | | | | |
| Fig. | Observer | | | Location | Obs. Date | Remark |
| Fig. 8e | Gestión y Planeamiento Territorial y Medioambiental, S.A.[3] | | | Gran Canaria, Spain | 23/03/2021 | The image has not been previously published |

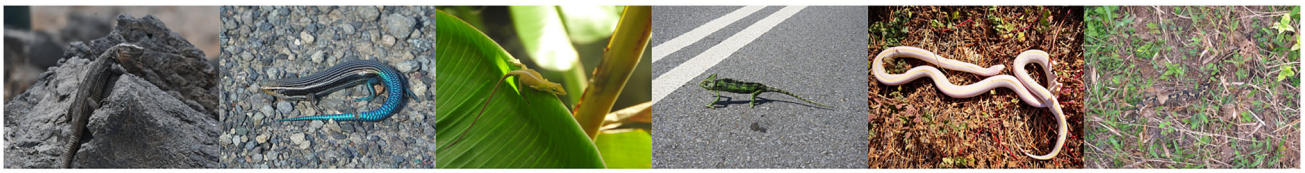
[1]Images resized for presentation purposes
[2]URLs last accessed: February 20, 2025
[3]Gestión y Planeamiento Territorial y Medioambiental (Gesplan) is a public company dedicated to territorial planning and environmental management

rapid reproduction and environmental adaptation, with more than 5500 specimens captured in Tenerife since 2021.

Furthermore, it serves as a potential reservoir of zoonotic bacteria (including *Escherichia coli*, *Campylobacter*, *Staphylococcus*, *Vibrio cholerae*, and *Salmonella*) posing transmission risks to native fauna [44].

- Ball python, *Python regius* (Shaw, 1802 ): This species has been introduced to both the Canary and Balearic archipelagos. In the Canary Islands, documented observations across diverse habitats in the wild include Gran Canaria, Tenerife, La Palma, and Fuerteventura. Although the ball python has not caused as much ecological damage as the California kingsnake in the Canary Islands, it still poses significant ecological risks, particularly to insular ecosystems. As one of the most traded reptiles globally, its high import volumes, reproductive capacity, adaptability, and the absence of large native predators facilitate its potential establishment in the wild. The species threatens native fauna through predation, competition, and the possible transmission of parasites and pathogens, endangering endemic species such as the Gran Canaria giant lizard. The discovery of released individuals in the wild, coupled with its high fecundity, generalist habits, and adaptability, highlights the substantial risk of invasion and ecological impact [45].
- California kingsnake, *Lampropeltis californiae* (Blainville, 1835): This species has devastated the endemic reptile populations of Gran Canaria causing a massive population reductions. The impact is particularly severe in areas invaded earlier and threatens to disrupt entire ecosystem dynamics by potentially interrupting crucial ecological functions such as plant pollination, seed dispersal, and invertebrate population control.

  Urgent management strategies must be developed to prevent the spread of the invasive California kingsnake and mitigate its devastating ecological impact on island ecosystems, recognizing the profound global conservation challenges posed by such invasive predators [46].



(a) *G. stehlini.* (b) *C. sexlineatus.* (c) *A. porcatus.* (d) *C. calyptratus.* (e) *L. californiae.* (f) *P. regius.*

**Fig. 8** Representative samples from the database used in this research

Representative specimens from the database are shown in Fig. 8, and the corresponding collection metadata for each illustrated specimen are detailed in Table 10.

**Acknowledgements** This work has been carried out with the advice of members of the technical and biological team of the Government of the Canary Islands for protected species and invasive alien species in the Canary Islands, who further contributed the California kingsnake (*L. californiae*) imagery employed in this study.

**Author Contributions** All authors contributed to the conception and design of the study. Material preparation, data collection, and analysis were carried out collaboratively. The first draft of the manuscript was collectively written, and all authors reviewed and approved the final version.

**Funding** This work was funded by the Research Training Personnel Program (Personal Investigador en Formación - PIF) of the University of Las Palmas de Gran Canaria, Call 2023-2 (3rd phase), under reference FPI2024010053. The program is sponsored by Banco Santander and funded by the Cabildo de Gran Canaria, the Ministerio de Ciencia, Innovación y Universidades, and the Agencia Canaria de Investigación, Innovación y Sociedad de la Información del Gobierno de Canarias.

**Data Availability** The data presented in this study are available on request from the corresponding author. The two private databases are not publicly available due to privacy restrictions and the proprietary nature of the data.

**Code Availability** The code developed for this study is available on request from the corresponding author.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Ethical approval** Not applicable.

**Consent for publication** Not applicable.

**Materials availability** Not applicable.



## References

1. Biological Diversity (CBD), C.: Convention on biological diversity: article 2. Use of terms. United Nations, Rio de Janeiro, Brazil . https://www.cbd.int/convention/text/ (1992). Accessed: 16 January 2025
2. Costanza, R., d'Arge, R., Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., Belt, M.: The value of the world's ecosystem services and natural capital. Nature **387**(6630), 253–260 (1997). https://doi.org/10.1038/387253a0
3. Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., Wardle, D.A.: Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecol. Monogr. **75**(1), 3–35 (2005). https://doi.org/10.1890/04-0922
4. European Commission, D.-G.f.E.: Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions eu biodiversity strategy for 2030 bringing nature back into our lives. Technical report, The European Commission, Brussels, Belgium . https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0380 (2020). Accessed: 16 January 2025
5. Freitas, R., Romeiras, M., Silva, L., Cordeiro, R., Madeira, P., González, J.A., Wirtz, P., Falcón, J.M., Brito, A., Floeter, S.R., Afonso, P., Porteiro, F., Viera-Rodríguez, M.A., Neto, A.I., Haroun, R., Farminhão, J.N.M., Rebelo, A.C., Baptista, L., Melo, C.S., Martínez, A., Núñez, J., Berning, B., Johnson, M.E., Ávila, S.P.: Restructuring of the 'macaronesia' biogeographic unit: a marine multi-taxon biogeographical approach. Sci. Rep. (2019). https://doi.org/10.1038/s41598-019-51786-6
6. Ferraro, G., Failler, P.: Biodiversity, multi-level governance, and policy implementation in Europe: a comparative analysis at the subnational level. J. Publ. Policy **44**(3), 546–572 (2024). https://doi.org/10.1017/s0143814x24000072
7. Guerrero, J.C., Vargas, J.M., Real, R.: A hypothetico-deductive analysis of the environmental factors involved in the current reptile distribution pattern in the canary islands. J. Biogeogr. **32**(8), 1343–1351 (2005). https://doi.org/10.1111/j.1365-2699.2005.01253.x
8. ARTE GEIE: Re: Schlangenplage Auf Gran Canaria: Die Kettennatter Breitet Sich Aus. ARTE. Länge: 33 Minuten. Ausgestrahlt auf ARTE.tv.https://www.arte.tv/de/videos/111751-010-A/re-schlangenplage-auf-gran-canaria/ (2025). Accessed: 17 January 2025
9. Manasa, K., Paschyanti, D.V., Vanama, G., Vikas, S.S., Kommineni, M., Roshini, A.: Wildlife surveillance using deep learning with YOLOv3 model. In: 2021 6th International Conference on Communication and Electronics Systems (ICCES), pp. 1798–1804. IEEE, Coimbatore, India (2021). https://doi.org/10.1109/ICCES51350.2021.9489121
10. Leorna, S., Brinkman, T.: Human vs. machine: detecting wildlife in camera trap images. Eco. Inform. **72**, 101876 (2022). https://doi.org/10.1016/j.ecoinf.2022.101876
11. Li, G., Shi, G., Zhu, C.: Dynamic serpentine convolution with attention mechanism enhancement for beef cattle behavior recognition. Animals **14**(3), 466 (2024). https://doi.org/10.3390/ani14030466
12. Chen, G., Han, T.X., He, Z., Kays, R., Forrester, T.: Deep convolutional neural network based species recognition for wild animal monitoring. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 858–862. IEEE, Paris, France (2014). https://doi.org/10.1109/icip.2014.7025172
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv (2015). https://doi.org/10.48550/ARXIV.1512.03385
14. Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di Salvo, P.A., Lewis, J.S., White, M.D., Teton, B., Beasley, J.C., Schlichting, P.E., Boughton, R.K., Wight, B., Newkirk, E.S., Ivan, J.S., Odell, E.A., Brook, R.K., Lukacs, P.M., Moeller, A.K., Mandeville, E.G., Clune, J., Miller, R.S.: Machine learning to classify animal species

in camera trap images: applications in ecology. Methods Ecol. Evol. **10**(4), 585–590 (2018). https://doi.org/10.1111/2041-210x.13120
15. Thangaraj, R., Rajendar, S., M, S., K, R.S., Sasikumar, S., L, C.: Automated recognition of wild animal species in camera trap images using deep learning models. In: 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), vol. 3, pp. 1–5. IEEE, Bhimtal, India (2023). https://doi.org/10.1109/icaect57570.2023.10117922
16. Binta Islam, S., Valles, D., Hibbitts, T.J., Ryberg, W.A., Walkup, D.K., Forstner, M.R.J.: Animal species recognition with deep convolutional neural networks from ecological camera trap images. Animals **13**(9), 1526 (2023). https://doi.org/10.3390/ani13091526
17. Annesa, O.D., Kartiko, C., Prasetiadi, A.: Identification of reptile species using convolutional neural networks (CNN). Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) **4**(5), 899–906 (2020). https://doi.org/10.29207/resti.v4i5.2282
18. Christin, S., Hervet, É., Lecomte, N.: Applications for deep learning in ecology. Methods Ecol. Evol. **10**(10), 1632–1644 (2019). https://doi.org/10.1111/2041-210x.13256
19. Patel, A., Cheung, L., Khatod, N., Matijosaitiene, I., Arteaga, A., Gilkey, J.W.: Revealing the unknown: real-time recognition of galápagos snake species using deep learning. Animals **10**(5), 806 (2020). https://doi.org/10.3390/ani10050806
20. Rajabizadeh, M., Rezghi, M.: A comparative study on image-based snake identification using machine learning. Sci. Rep. (2021). https://doi.org/10.1038/s41598-021-96031-1
21. Othman, Z., Abu Mansor, N.N., Azmi, N.F., Mat Zain, N.H., Fariza Abu Samah, K.A., Ismai, I., Ahmad, K.A.: Snake species identification using digital image processing. In: 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), vol. 4, pp. 1–6. IEEE, Kedah, Malaysia (2021). https://doi.org/10.1109/icraie52900.2021.9703898
22. Progga, N.I., Rezoana, N., Hossain, M.S., Islam, R.U., Andersson, K.: A cnn based model for venomous and non-venomous snake classification. In: Applied Intelligence and Informatics, pp. 216–231. Springer, Cham, Switzerland (2021). https://doi.org/10.1007/978-3-030-82269-9_17
23. Ahmed, K., Gad, M.A., Aboutabl, A.E.: Snake species classification using deep learning techniques. Multimed. Tools Appl. **83**(12), 35117–35158 (2023). https://doi.org/10.1007/s11042-023-16773-0
24. Rehman, M.U., Nizami, I.F., Ullah, F., Hussain, I.: IQA vision transformed: a survey of transformer architectures in perceptual image quality assessment. IEEE Access **12**, 183369–183393 (2024). https://doi.org/10.1109/access.2024.3506273
25. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 9992–10002 (2021). https://doi.org/10.1109/ICCV48922.2021.00986
26. Zhang, M., Gao, F., Yang, W., Zhang, H.: Real-time target detection system for animals based on self-attention improvement and feature extraction optimization. Appl. Sci. **13**(6), 3987 (2023). https://doi.org/10.3390/app13063987
27. Cunha, F., Santos, E.M., Colonna, J.G.: Bag of tricks for long-tail visual recognition of animal species in camera-trap images. Eco. Inform. **76**, 102060 (2023). https://doi.org/10.1016/j.ecoinf.2023.102060
28. Society, W.C.: WCS Camera Traps. Lila.science. Dataset containing 1.4 million images of 675 species from 12 countries (2022). https://lila.science/datasets/wcscameratraps
29. Swanson, A.B., Kosmala, M., Lintott, C.J., Simpson, R.J., Smith, A., Packer, C.: Data from: snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Dryad (2016). https://doi.org/10.5061/DRYAD.5PT92
30. Schneider, D., Lindner, K., Vogelbacher, M., Bellafkir, H., Farwig, N., Freisleben, B.: Recognition of European mammals and birds in camera trap images using deep neural networks. IET Comput. Vision (2024). https://doi.org/10.1049/cvi2.12294
31. Bolon, I., Picek, L., Durso, A.M., Alcoba, G., Chappuis, F., Castañeda, R.: An artificial intelligence model to identify snakes from across the world: Opportunities and challenges for global health and herpetology. PLoS Negl. Trop. Dis. **16**(8), 0010647 (2022). https://doi.org/10.1371/journal.pntd.0010647
32. Bloch, L., Friedrich, C.M.: Efficientnets and vision transformers for snake species identification using image and location information. In: Faggioli, G., Ferro, N., Joly, A., Maistro, M., Piroi, F. (eds.) Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021. CEUR Workshop Proceedings, vol. 2936, pp. 1477–1498. CEUR-WS.org, Bucharest, Romania (2021). https://ceur-ws.org/Vol-2936/paper-126.pdf
33. Gong, H., Luo, T., Ni, L., Li, J., Guo, J., Liu, T., Feng, R., Mu, Y., Hu, T., Sun, Y., Guo, Y., Li, S.: Research on facial recognition of sika deer based on vision transformer. Eco. Inform. **78**, 102334 (2023). https://doi.org/10.1016/j.ecoinf.2023.102334
34. Veiga, R.J.M., Rodrigues, J.M.F.: Fine-grained fish classification from small to large datasets with vision transformers. IEEE Access **12**, 113642–113660 (2024). https://doi.org/10.1109/access.2024.3443654
35. Hernández-López, R., Travieso-González, C.M.: Reptile identification for endemic and invasive alien species using transfer learning approaches. Sensors **24**(5), 1372 (2024). https://doi.org/10.3390/s24051372
36. iNaturalist: iNaturalist. California academy of sciences and national geographic society. https://www.inaturalist.org (2025). Accessed: 11 February 2025
37. Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., Guo, B.: Swin transformer v2: scaling up capacity and resolution. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 11999–12009 (2022). https://doi.org/10.1109/CVPR52688.2022.01170
38. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 words: transformers for image recognition at scale. arXiv (2020). https://doi.org/10.48550/ARXIV.2010.11929
39. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge, MA (2016)
40. Smith, L.N., Topin, N.: Super-convergence: very fast training of neural networks using large learning rates. arXiv (2017). https://doi.org/10.48550/ARXIV.1708.07120
41. Bowles, P.: Gallotia stehlini: Bowles, P.: The IUCN red list of threatened species 2024: e.T61506A137850850. IUCN (2024). https://doi.org/10.2305/iucn.uk.2024-1.rlts.t61506a137850850.en
42. Bowles, P.: Chalcides sexlineatus: Bowles, P.: The IUCN red list of threatened species 2024: e.T61487A137848983. IUCN (2024). https://doi.org/10.2305/iucn.uk.2024-1.rlts.t61487a137848983.en
43. Pino-Vera, R., Abreu-Acosta, N., Foronda, P.: Study of zoonotic pathogens in alien population of veiled chameleons (Chamaeleo Calyptratus) in the Canary islands (Spain). Animals **13**(14), 2288 (2023). https://doi.org/10.3390/ani13142288
44. Afonso, O., Fariña, B.: Presencia de Anolis porcatus en Tenerife (Islas Canarias). Boletín de la Asociación Herpetológica Española **34** (2023)

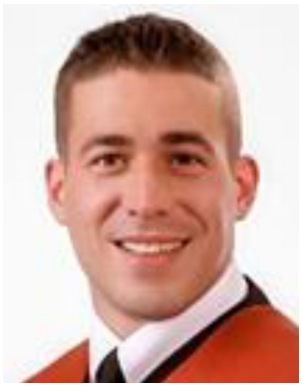
45. Ministerio para la Transición Ecológica y el Reto Demográfico: Catálogo Español de Especies Exóticas Invasoras: Python regius (Shaw, 1802). https://www.miteco.gob.es/content/dam/miteco/es/biodiversidad/temas/conservacion-de-especies/mtjpythonregius_tcm30-523168.pdf. Consultado el 7 de febrero de 2025 (2020)
46. Piquet, J.C., López-Darias, M.: Invasive snake causes massive reduction of all endemic herpetofauna on gran Canaria. Proc. Royal Soc. B: Biol. Sci. (2021). https://doi.org/10.1098/rspb.2021.1939



**Ruymán Hernández-López** received the M.Eng. degree in Telecommunication Engineering from the Universidad de Las Palmas de Gran Canaria (ULPGC), Spain, in 2017. He is currently a Ph.D. candidate at ULPGC, where he also works as a Research Training Fellow. He is a member of the IDeTIC research group (Digital Signal Processing Division). He has participated in European and regional research projects and has authored 4 publications in international peer-reviewed journals. He has served as a session chair and organizing committee member at international conferences. His research interests include deep learning, digital signal processing, biodiversity conservation, and species identification using artificial intelligence techniques.

**Francisco A. Delgado-Rajó** is a Ph.D. holder and Professor at the University of Las Palmas de Gran Canaria (ULPGC), where he leads the Networks and Telematic Services Division at IDeTIC. With a career spanning over 20 years, his research focuses on IoT architectures, Edge Computing, and Artificial Intelligence for environmental monitoring and cybersecurity. He has participated in 27 competitive research projects and published extensively in indexed journals. Currently, as a Principal Investigator, he develops high-impact solutions for distributed intelligent systems, recognized by prestigious awards such as the Alcatel Prize for his doctoral thesis.

**Sergio Celada-Bernal** received his B.Sc. degree in Telecommunication Technologies Engineering and his M.Sc. degree in Telecommunication Engineering from the University of Las Palmas de Gran Canaria (ULPGC). He is currently a Ph.D. candidate in the Doctoral Program in Business, Internet and Communications Technologies at ULPGC, where he also works as a Research Training Fellow. His research focuses on digital twins in healthcare and the application of artificial intelligence for pattern detection and disease analysis. He has authored 7 publications, supervised undergraduate and master's theses, and contributed to teaching in Biomedical Engineering.



**Alejandro Piñan-Roescher** received a B.Sc. in Telecommunication Technologies (Electronic Systems) in 2017 and an M.Sc. in Telecommunication Engineering in 2019, both from ULPGC. He is currently pursuing a Ph.D. in Business, Internet, and Communication Technologies, where he works as a research fellow. His research focuses on time series forecasting using advanced machine learning and deep neural networks. He develops models to analyze complex patterns and deliver accurate predictions in dynamic environments. In this context, his work focuses on estimating solar radiation using multimodal data. He has authored five publications, supervised undergraduate and master's theses, and contributed to university teaching.

**Carlos M. Travieso-González** received his M.Sc. degree from UPC in 1997 and his Ph.D. from ULPGC in 2002. He is currently a Full Professor and Head of the Signals and Communications Department at ULPGC. His research focuses on biometrics, biomedical signal processing, and machine learning. He has authored over 510 publications, 7 patents, supervised 13 Ph.D. theses, and participated in up to 52 research projects. Additionally, he has participated as an expert evaluator for the European Union and for research agencies of Germany, Spain, the UK, France, Slovenia, Poland, Croatia, Slovakia, and other countries. He has chaired numerous international conferences, including the IEEE-IWOBI, APPIS, and CAIP series.