Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Few-Shot Learning for Automated Sexual Dimorphism Classification in Scorpions Using Multi-View Images

Yavuz UNAL

yunal@sinop.edu.tr

Sinop University

Halil KOC

Sinop University

Research Article

Keywords:

Posted Date: April 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8691932/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Few-Shot Learning for Automated Sexual Dimorphism Classification in Scorpions Using Multi-View Images

Yavuz ÜNAL (Corresponding Author), Sinop University, Faculty of Engineering and Architecture,
Department of Computer Engineering, Sinop, Türkiye,
e-mail: yunal@sinop.edu.tr, ORCID: 0000-0002-3007-679X

Halil KOÇ, Sinop University, Faculty of Arts and Sciences,
Department of Biology, Sinop, Türkiye,
e-mail: halilkoc@sinop.edu.tr, ORCID: 0000-0003-0429-2824

ABSTRACT

Morphological differences between sexes represent a fundamental topic in biology and are observed in scorpions as in other organisms, defined by variations in body part proportions, sizes, and shapes between males and females. Sex determination in scorpions presents challenges for biologists yet provides critical insights into evolutionary processes, ecological adaptations, and reproductive strategies. Field researchers encounter difficulties distinguishing sexes, particularly given the venomous nature of some species. This study aimed to classify scorpion sexes using dorsal and ventral images through a few-shot learning approach with prototypical networks for sex classification with limited labeled data. The dataset comprised 99 *Aegeaeobuthus gibbosus* scorpions (50 females, 49 males), with both dorsal and ventral images obtained for each specimen. DenseNet121 and ResNeXt-50 architectures, pre-trained on ImageNet, served as base networks and were evaluated at 1-shot, 5-shot, and 10-shot settings. Experimental results indicated that DenseNet121 achieved 84.94% accuracy at the 10-shot setting, while ResNeXt-50 achieved 84.14%. Both architectures demonstrated consistent accuracy improvements as support samples increased, confirming the few-shot learning framework's effectiveness. Fold-based confusion matrix analyses showed that DenseNet121 exhibited more stable performance with challenging samples, whereas ResNeXt-50 was more sensitive to individual morphological variation but achieved comparable accuracy with sufficient support samples. These findings demonstrate that few-shot learning provides a viable solution for scorpion sex classification with limited data. High accuracy can be achieved using a small number of labeled specimens, offering a practical approach for researchers studying rare or endangered scorpion species.

INTRODUCTION

Sexual dimorphism, especially in arthropods, includes the physical and behavioral differences between males and females of the same species (McLean et al. 2018) (Giménez Carbonari et al. 2024). Accurate identification of sex is important for ecological research, biodiversity conservation, breeding programs, and taxonomic studies. Traditional methods need expert knowledge and manual measurements, relying on microscopic examination of anatomical features. However, these methods are time-consuming and labor-intensive, particularly when dealing with preserved specimens or subtle morphological differences (Cheng and Kuntner 2015; Hopkins and Kopp 2021).

In scorpions, sexual dimorphism is evident in clear anatomical features. These include the size of the pedipalps, the proportions of the metasoma segments, and the structure of the pectin organs. These traits are important for classification. Sexual dimorphism is especially noticeable in the family Buthidae. In this family, males and females differ in body proportions, limb shapes, and reproductive structures. These differences relate to mating strategies and offspring number (Scharf and Martin 2013). *Aegeaeobuthus gibbosus* (previously *Mesobuthus gibbosus*) is one of the

most common buthid scorpions found in Anatolia and the Eastern Mediterranean (Crucitti 1999). It shows typical sexual dimorphism. Males have long, slender metasomal segments and longer pedipalps with more pectin teeth. Females have a wider mesosoma, which supports egg production and the carrying of young (Hjelle 1990; Vachon). Despite its importance in ecology and medicine, there are few detailed studies on sex classification in this species using sexually dimorphic traits (Kovařík 1998).

Recent advances in deep learning and computer vision have changed the way we analyze biological images (Moen et al. 2019). These tools now enable the automation of species identification, disease detection, and morphological classification. Deep learning models can detect complex patterns in images and classify specimens accurately, even when differences are too subtle for humans to detect (Høye et al. 2021). However, a major challenge is the lack of labeled training data. Collecting and labeling large datasets is often difficult due to ethical issues, the rarity of some specimens, limited access to endangered species, and the high cost of expert annotation. Traditional supervised learning methods require hundreds or thousands of labeled examples per class, which is not feasible in many biological studies, where data are limited (Tuia et al. 2022).

Few-shot learning helps solve this problem by allowing models to learn from just one to ten labeled examples per class. Prototypical networks, a common method in few-shot learning, represent each class as a prototype in a feature space and classify new samples by measuring how close they are to these prototypes. This method is especially useful in biological studies where data is limited (Wang et al. 2023) .

To our knowledge, this is the first study to apply few-shot learning to automated sex classification in scorpions using sexually dimorphic morphological features. Our main contributions are: (1) we introduce a few-shot learning framework with prototypical networks that achieves high accuracy (84.94%) using only 10 labeled examples per sex; (2) we show that using both dorsal and ventral images with person-level cross-validation provides strong evaluation and prevents data leakage; and (3) we compare two network architectures, DenseNet121 and ResNeXt-50, to highlight different learning patterns and performance trade-offs across various shot settings. Our results show that few-shot learning is a practical way to automate morphological analysis in biological fields with limited data, balancing efficient annotation with accurate classification for real-world taxonomic research.

The rest of this paper is organized as follows. The Literature Review covers related work on automated classification of biological specimens using deep learning. Materials and Methods describe the dataset, the few-shot learning framework, the data split, the CNN architectures (DenseNet121 and ResNeXt-50), the preprocessing steps, and parameter selection. Results present sex classification performance across different shot settings, training progress, and confusion matrices. The Discussion looks at the features of each architecture and their performance across different classes. Finally, the Conclusions summarize our contributions, limitations, and suggestions for future research.

The remainder of this paper is organized as follows. The Literature Review examines related work on automated classification of biological specimens using deep learning. Materials and Methods describe the dataset, the proposed few-shot learning framework, the data partitioning strategy, the CNN architectures (DenseNet121 and ResNeXt-50), the preprocessing pipeline, and the hyperparameter optimization. Results present experimental findings including performance comparisons across different shot settings, training dynamics, and confusion matrices. Discussion analyzes architectural characteristics and class-specific performance patterns. Finally, the Conclusions summarize the contributions, limitations, and future research directions.

Literature Review

Ajdi et al. (Ajdi et al. 2025) employed a molecular mass fingerprinting (MFP) technique to mitigate the taxonomic misidentification of scorpions. Using MALDI-MS, they got MFPs of 251 specimens from 15 species collected from different parts of Morocco. With this information, they made both predictive classification models and a species-specific library of average spectral profiles in machine learning-based software. The model in training was able to

recognize 99% of molecular signatures, but when they tested it on a new external dataset, the models were only 97% accurate, and the library was only 98% accurate.

Anwarul and Tanwar (Anwarul and Tanwar 2025) introduced a deep learning framework called “Venomify” to quickly and accurately tell apart venomous and non-venomous snakes. They collected images using the Flickr API and used data enhancement to address class imbalance during preprocessing. The team classified the snake images with a modified DenseNet169 model and achieved 97.35% accuracy (precision 97.35%, recall 97.37%, F1 97.35%). According to the authors, Venomify could serve as a fast and reliable decision support tool for education, wildlife management, and emergency response.

Chen et al. (Chen et al. 2021) reported that traditional sex determination in spiders relies on genital structures, specifically the palp in males and the epigyne in females. This method is expertise-dependent, time-consuming, and susceptible to error. To address these limitations, the authors proposed a classification approach based on deep learning and transfer learning. Their comparison of various convolutional neural network (CNN) backbones using laboratory-collected sample images demonstrated that Xception-based transfer learning outperformed ResNet-152. Following data augmentation, activation function optimization, and fine-tuning of frozen layers, the model achieved an accuracy of 98.02%. In independent real-world measurement samples, the model attained an accuracy of 92.38%.

Giambelluca et al. (Giambelluca 2022) created two computer vision and deep learning systems to automatically detect and classify two types of scorpions found in La Plata, Argentina. They used YOLO and MobileNet for scorpion detection, and both models worked well, with accuracy between 88% and 91% and recall rates from 90% to 97%. MobileNet performed better in uncontrolled settings and when detecting multiple scorpions. It was also used to classify scorpions as dangerous (Tityus) or non-dangerous (Bothriurus).

Hosseini et al. (Hosseini et al. 2019) introduced two automated methods to minimize the observational and subjective aspects of sex determination in zebrafish: a deep convolutional neural network (DCNN) utilizing whole fish images and a support vector machine (SVM) based on caudal fin color intensity. The DCNN achieved 100% accuracy in samples reared at elevated temperatures, whereas the SVM misclassified 20% of males due to reduced color intensity. Additionally, a positive correlation was observed between color intensity and both length and weight in male zebrafish.

McLean et al. (McLean et al. 2018) reviewed the literature to understand how sexual dimorphism (SD) has evolved in arachnids. They looked at morphological data from many species, but noticed that most studies focused on certain groups and overlooked smaller ones. Their analysis showed three main patterns: females tend to be larger overall, males often have longer legs, and there are differences in pedipalp shape. They also pointed out that it is hard to explain how these traits evolved because the relationships between arachnid groups are still unclear.

Fraiwan et al. (Fraiwan et al. 2025) developed a deep learning-based artificial intelligence system for the identification of sandfly species and sex. As sandflies serve as vectors for diseases such as leishmaniasis, precise identification is essential for effective disease control. The researchers integrated genital and pharyngeal images from field-collected samples over a two-year period, employing an early fusion approach, convolutional neural networks, and transfer learning. Their system achieved over 95% accuracy in classifying the sex and species of three Phlebotomus species using various pre-trained CNN models.

Tuda and Luna-Maldona (Tuda and Luna-Maldonado 2020) compared machine learning methods for classifying the species and sex of Callosobruchus chinensis and its parasitoids, Anisopteromalus and Heterospilus. They used object-based and pixel-based features from insect images taken at different exposures to train six machine learning models: logistic model trees, random forest, support vector machine, simple logistic regression, multilayer perceptron, and AdaBoost. Simple logistic regression, logistic model trees, and AdaBoost performed best, with 88.5-98.5% accuracy for identifying sex within a species, 97.3% for two-species sex identification, and 93.3% for three-species sex identification.

Salifu et al. (Salifu et al. 2022) evaluated four machine learning algorithms using fruit fly morphometric data. K-nearest neighbors (KNN) and random forest (RF) demonstrated low performance, whereas support vector machine

(SVM) models achieved over 95% accuracy (Kappa > 0.78, AUC > 0.91), and the artificial neural network (ANN) model reached 96% accuracy (Kappa = 0.83, AUC = 0.98). Wing veins 2, 3, 8, 10, 14, and tibia length were identified as the most significant variables. The authors concluded that SVM and ANN are suitable for integration into intelligent software applications for insect identification.

The studies outlined above have shown that deep learning methods work well for identifying species and sex in insects and other arthropods. However, these methods require hundreds or thousands of labeled images and are not applicable to many biological studies because data collection is difficult. There is a scarcity of studies on the automatic classification of sexual dimorphism in Arachnida, and no automatic identification system for scorpions has yet been established. Moreover, the efficacy of few-shot learning methodologies, which operate with a limited number of labeled samples, in detecting sexual dimorphism in arachnids remains unexamined. Thus, this study aimed to automatically classify sexual dimorphism in the scorpion species *Aegeaeobuthus gibbosus*. A prototype-based network architecture for few-shot learning achieved an accuracy of 84.94% with only 10 labeled samples per class.

Material and Methods

Dataset

This dataset contains digital images of *Aegeaeobuthus gibbosus* (Brullé, 1832) (Scorpiones: Buthidae) specimens collected in Manisa, Turkey, in 2005. The second author identified the species. In total, 99 individuals (50 females and 49 males) were preserved in 70% ethanol and photographed under the same conditions.

All photos were taken with a Canon EOS 70D camera and a Sigma APO Macro 180mm f/2.8 lens. Diffuse lighting reduced shadows and kept the lighting even for all specimens. Each scorpion was placed so its body axis was parallel to the image plane for consistency. The images were saved as JPEGs at 1600×902 pixels in the RGB color space with no compression.

Each specimen was photographed from two angles: dorsal (back view) and ventral (underside view). Taking both views allows for a thorough analysis of features like the prosoma, mesosoma, metasoma segments, and pedipalps. Figure 1 shows examples from each group, highlighting the main features seen in both views.



(a) Female - Dorsal (b) Female - Ventral

(c) Male - Dorsal (d) Male - Ventral

Figure 1. Representative sample images from the dataset: (a) Female dorsal, (b) Female ventral, (c) Male dorsal, and (d) Male ventral views.

As shown in Figure 1, sexual dimorphism is evident in features such as metasomal segment proportions and pedipalp morphology. The dorsal views clearly show differences in mesosoma width, whereas the ventral views reveal variations in pectin structure and the genital operculum. These multi-view observations provide complementary information for robust sex classification.

The complete dataset composition is detailed in Table 1. The nearly balanced sex distribution (50.5% female, 49.5% male) ensures unbiased model training and evaluation, mitigating potential class imbalance that could affect classification performance.

Table 1. Dataset's class distribution

Category	**Number of images**
Female dorsal	50
Female ventral	50
Male dorsal	49
Male ventral	49
Total	198

The dual-view imaging strategy presents a challenging cross-view learning scenario, requiring the model to extract sex-discriminative features that generalize across both dorsal and ventral perspectives.

Methods

To enable sex classification from limited labeled examples, we developed a few-shot learning framework based on prototypical networks. The following subsections detail the theoretical foundation, network architectures, data partitioning strategy, preprocessing pipeline, hyperparameter optimization, and evaluation metrics.

Few-Shot Learning Framework

Few-shot learning addresses the challenge of training deep learning models with extremely limited labeled data, typically 1-10 examples per class. Unlike traditional supervised learning that requires hundreds or thousands of labeled instances, few-shot learning mimics human-like learning by generalizing from minimal examples. This paradigm is particularly critical in biological studies, where data collection is constrained by ethical considerations, specimen rarity, the cost of expert annotation, and field-collection difficulties (Pachetti and Colantonio 2023) (Nguyen et al. 2023). In scorpion sex classification, obtaining large numbers of labeled images across multiple anatomical views poses significant logistical challenges, making few-shot learning essential.

Our framework employs a prototypical network (Snell et al. 2017), a metric-learning approach that represents each class as a prototype in an embedding space. The core principle operates via episodic meta-learning: each training episode samples n support examples per class (where n denotes the shot setting: 1-shot, 5-shot, or 10-shot), along with

query examples for evaluation. Support images are processed through a convolutional neural network (CNN) backbone to extract feature representations, which are then averaged within each class to compute class prototypes. Query images are classified by measuring their Euclidean distance to each prototype, with class membership determined by the nearest prototype. This distance-based approach ensures that same-class examples cluster together in the embedding space while maintaining separation between different classes.

The classification process follows three stages:
1. Feature extraction: A frozen pretrained CNN (DenseNet121 or ResNeXt-50) processes input images to generate high-dimensional feature vectors
2. Prototype computation: Support set features are averaged per class to create prototypes:

$C_k = \frac{1}{|S_k|}\sum_{(x_i y_i)\epsilon S_k} \quad f_\phi(x_i)$, where S_k represents support examples of class k
3. Distance-based classification: Query examples are assigned to the class with the nearest prototype using Euclidean distance: $d(f_\phi(x), c_k) = ||f_\phi(x) - c_k||_2^2$

The model is trained end-to-end using prototypical loss, which encourages the learned embedding space to cluster same-class examples while separating different classes. Only the projection head parameters are updated during training, while the pretrained CNN backbone remains frozen, enabling efficient adaptation to scorpion sexually dimorphic features while leveraging knowledge from large-scale visual recognition tasks.

The complete architecture of the proposed few-shot learning framework is illustrated in Figure 2, showing the end-to-end pipeline from image input through preprocessing, feature extraction, prototype computation, and distance-based classification.

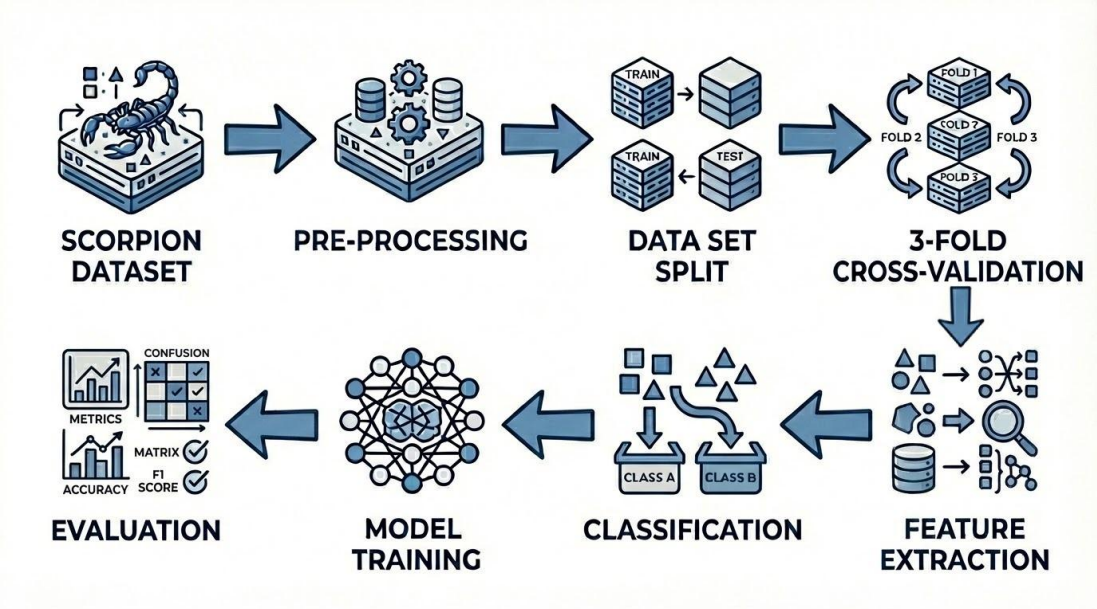


Figure 2. Overview of the few-shot learning framework.

As shown in Figure 2, the pipeline processes both support and query images through three main stages: preprocessing (grayscale conversion, resizing, normalization), feature extraction via a frozen pretrained CNN backbone (DenseNet121 or ResNeXt-50), and projection to an embedding space. Support images are aggregated via mean pooling to compute class prototypes, while query images are classified based on their Euclidean distance to each prototype.

Network Architectures

We evaluated two pretrained CNN architectures as frozen feature extractors, each offering distinct architectural characteristics for feature learning.

DenseNet121

DenseNet121 (Densely Connected Convolutional Network (Huang et al. 2017) uses a unique design in which each layer in a dense block receives input from all earlier layers by concatenating their channels. Unlike traditional networks, in which each layer takes input only from the preceding layer, DenseNet links every layer to all previous layers (0, 1, ..., l-1). This means that each layer receives feature maps from all earlier layers (Asadi Amiri et al. 2026).

The network comprises four dense blocks with 6, 12, 24, and 16 layers, separated by transition layers that use 1×1 convolutions and 2×2 average pooling to reduce feature dimensions. Each convolutional layer adds 32 new feature channels, since the growth rate is k=32. This dense connectivity helps prevent the vanishing gradient problem by enabling short paths from early to later layers, thereby improving gradient flow during backpropagation. The final global average pooling layer produces 1024-dimensional feature vectors. Reusing features allows the network to use parameters efficiently and avoids learning the same information multiple times (Xie et al. 2016).

ResNeXt-50

ResNeXt-50 (Residual Network with Aggregated Transformation (Xie et al. 2016) extends ResNet by introducing the cardinality of the number of parallel transformation pathways as a new dimension alongside depth and width. The fundamental building block implements a "split-transform-merge" strategy $F(x) = \sum_{i=1}^{C} \quad T_i(x)$, where C denotes the cardinality, and T_i denotes the i-th transformation path (Tanwar and Singh 2023) .

Each pathway employs a bottleneck architecture with three convolutional layers: 1×1 convolution for dimensionality reduction, 3×3 grouped convolution with C=32 groups, and 1×1 convolution for restoration. ResNeXt-50 organizes 50 weight layers into four stages, with a cardinality of 32 and a bottleneck width of 4, such that each residual block contains 32 parallel pathways that process 4-dimensional intermediate features. The architecture outputs 2048-dimensional features from global average pooling. This ensemble-like behavior enables exploration of a richer hypothesis space without substantial computational overhead, making it effective for transfer learning with limited data (Gai et al. 2026).

Data Split Strategy

To ensure robust evaluation and prevent data leakage, we employed person-level stratified 3-fold cross-validation, following established protocols in computer vision that involve multiple views of the same individual (Parkhi et al. 2015). This strategy maintains independence between training and test sets, which is critical given that each scorpion individual has both dorsal and ventral images.

We employ person-level partitioning to prevent the same scorpion from appearing in both training and test sets. Unlike traditional image-level splitting, our approach partitions data at the individual level, ensuring that both anatomical

views (dorsal and ventral) of each scorpion are exclusively assigned to either the training or test set. This prevents the model from exploiting individual-specific characteristics and instead learns generalizable sex-discriminative features (Zheng et al. 2015). The dataset comprises 99 unique individuals (50 females, 49 males), each represented by two images.

The 99 individuals were divided into three stratified folds, maintaining sex balance. Each fold contained 33 individuals: Fold 1 (17 females, 16 males), Fold 2 (17 females, 16 males), and Fold 3 (16 females, 17 males), corresponding to 66 images per fold.

Within each fold, the training set (66 individuals, 132 images) is partitioned for episodic meta-learning following the prototypical networks framework [20]. The meta-training set enables the model to learn discriminative feature representations from the broader training distribution. The support set contains n examples per class (where $n \in \{1, 5, 10\}$), serving as reference exemplars for computing class prototypes during inference. The query set comprises held-out examples from the training set used for episodic evaluation, simulating the few-shot classification scenario while providing feedback for model refinement. Each fold executes 100 episodes with randomly sampled support/query configurations. The test set (33 individuals, 66 images) remains completely held out for final evaluation. Performance metrics are computed on test sets and reported as mean ± standard deviation across three folds.

Figure 3 illustrates the complete data partitioning workflow.

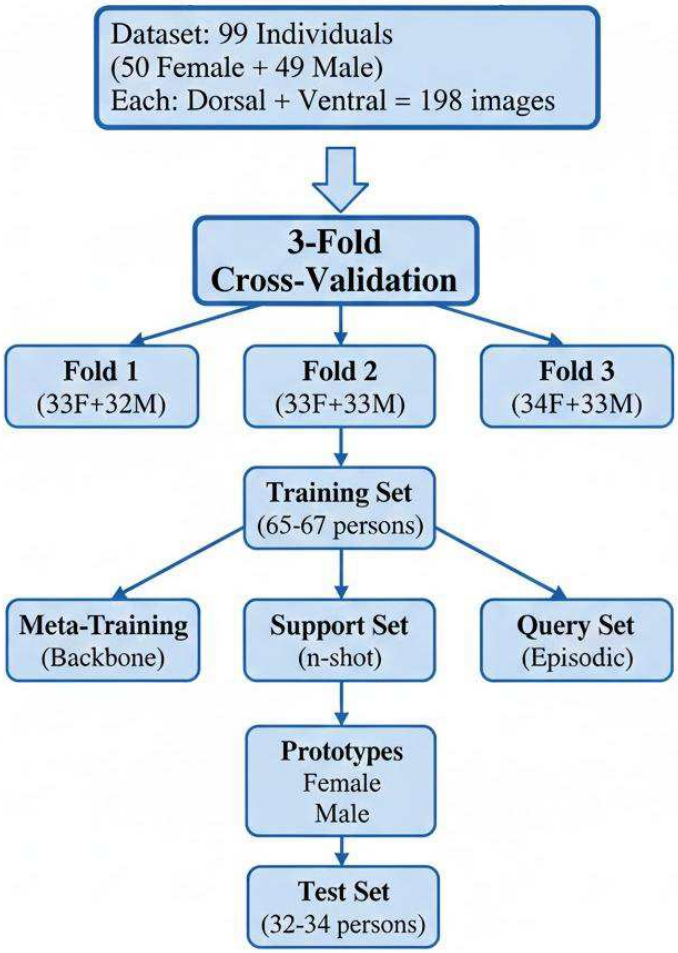


Figure 3. Data partitioning strategy showing a three-fold cross-validation setup with person-level splitting and few-shot learning configuration, including meta-training, support, and query sets.

Preprocessing Pipeline

All images underwent standardized preprocessing to ensure consistency and compatibility with pretrained models (Table 2). Raw RGB photographs were converted to grayscale and replicated across three channels to preserve morphological features while reducing color-related preservation artifacts, then resized to 224×224 pixels using bilinear interpolation. Images were converted to PyTorch tensors, normalized to [0, 1] range, and standardized using ImageNet statistics (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) to align with the pretrained models' feature space. No data augmentation was applied to maintain authentic anatomical characteristics critical for sex determination.

Table 2. Preprocessing pipeline specifications for scorpion image preparation.

Preprocessing Step	**Specification**	**Description**
Input format	JPG images	Original scorpion photographs

Color conversion	RGB to Grayscale (3-channel)	Convert to grayscale while maintaining 3-channel format for pretrained model compatibility
Resize	224x224	Standard input size for CNN architectures
Data type conversion	Tensor	Convert PIL Image to PyTorch tensor [0, 1]
Normalization	Mean: [0.485, 0.456, 0.406]Std: [0.229, 0.224, 0.225]	ImageNet pretrained model statistics

Hyperparameter Optimization

To ensure a fair comparison across architectures and to maximize model performance, we conducted systematic hyperparameter optimization in the 5-shot learning setting using grid search. The optimization evaluated three critical components: the projection-layer architecture (direct vs. multi-layer), dropout rates, and the learning rate. Both DenseNet121 and ResNeXt-50 were tested using 3-fold cross-validation to balance computational efficiency with reliable performance estimation. The optimal hyperparameters identified through this process were applied uniformly across all shot settings (1-shot, 5-shot, 10-shot) to maintain consistency and prevent overfitting to specific configurations, in accordance with established few-shot learning practices.

Table 3 presents the hyperparameter search space explored during optimization.

Table 3. Hyperparameter Search Space for Grid Search Optimization

Hyperparameter	**Values**	**Total Combinations**
Projection Type	[direct, 2-layer]	Direct: backbone_dim→512→512 2-layer: backbone_dim→0.75×dim→0.5×dim→512
Dropout (Layer 1)	[0.3, 0.4]	First projection layer dropout
Dropout (Layer 2)	[0.2]	Second projection layer dropout
Learning Rate	[0.0001, 0.00015]	Adam optimizer learning rate
Combinations		8 per model
Total Evaluations		16 (2 models × 8 configs)

The optimization results revealed consistent optimal configurations across architectures (Table 4). DenseNet121 emerged as the top performer, achieving 81.67% ± 3.12% accuracy on the 5-shot setting, attributable to its dense connectivity patterns that facilitate feature reuse. Notably, both architectures converged to nearly identical configurations (direct projection, dropout=[0.4, 0.2], learning rate≈0.00015), despite their distinct architectural designs. This convergence suggests that these hyperparameters are well-suited to prototypical few-shot learning for scorpion sex classification. The preference for direct projection over multi-layer architectures indicates that simpler projection heads reduce the risk of overfitting in low-data regimes (Mrabah et al. 2025).

Table 4. Optimal Hyperparameters for Each Architecture

Model	Projection	Dropout1	Dropout2	Learning rate	Accuracy(%)
DenseNet121	direct	0.4	0.2	0.00015	81.67 ± 3.12
ResNeXt-50	direct	0.3	0.1	0.0002	80.26 ±7.26

Table 5 summarizes the complete model configuration applied to all experiments. Both architectures employ identical training configurations to ensure fair comparison. All ImageNet-pretrained convolutional layers remain frozen, preserving learned visual features while preventing overfitting. The trainable projection head maps backbone features (1024-D for DenseNet121, 2048-D for ResNeXt-50) through a 512-dimensional intermediate layer to the 2-class output, using optimized dropout rates. This lightweight design (fewer than 1M trainable parameters) enables rapid adaptation to scorpion morphology while leveraging pretrained knowledge.

Table 5. Model architecture configuration for DenseNet121 and ResNeXt-50, specifying frozen and trainable components.

Model	Frozen Layer	Trainable Layers	Learning Rate	Dropout	Optimizer	Regularization	Batch Size
DenseNet121	All convolutional base	Dense(1024) + Output (2 FC)	0.00015	0.4, 0.2	Adam	L2 ($\lambda = 0.001$)	32
ResNeXt-50	All convolutional base	Dense(2048) + Output (2 FC)	0.0002	0.4, 0.2	Adam	L2 ($\lambda = 0.001$)	32

Evaluation Metrics

To comprehensively assess classification performance, we employed standard evaluation metrics derived from the confusion matrix. The confusion matrix captures four fundamental outcomes: true positives (TP), correctly classified positive examples; true negatives (TN), correctly classified negative examples; false positives (FP), negative examples incorrectly classified as positive; and false negatives (FN), positive examples incorrectly classified as negative (Sokolova and Lapalme 2009).
From these values, we computed the following metrics:

Accuracy: The proportion of correct predictions across all classes.

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

Precision: The proportion of positive predictions that are actually correct.

$$\text{Precision}=\frac{TP}{TP+FP}$$

Recall (Sensitivity): The proportion of actual positives correctly identified.

Recall=$\frac{TP}{TP+FN}$

F1-Score: The harmonic mean of precision and recall, balancing both metrics.

F1-score=$2X\frac{Precision\ x\ Recall}{Precision + Recall}$

All metrics are reported as mean ± standard deviation across three cross-validation folds. Additionally, we present confusion matrices to qualitatively analyze classification patterns and class-specific performance trends.

RESULTS

This section presents the experimental results of our few-shot learning framework for classifying scorpion sexual dimorphism. We first report DenseNet121's performance across different shot settings, then present a fold-wise analysis and training dynamics, and finally compare these results with those of ResNeXt-50.

To ensure reproducibility and transparency, all experiments were conducted on a standardized computational infrastructure, as detailed in Table 6.

Table 6. Computational Infrastructure and Software Configuration

Component	**Specifications**
Computing Platform	Google Colab Pro
GPU	Nvidia Tesla T4 (16GB GDDR6)
CPU	Intel Xeon (2 Cores @2.3GHZ
RAM	12.7GB
Framework	PyTorch 2.0.1
Cuda Version	12.0
Python	3.10.12

All experiments were conducted in Google Colab Pro's cloud environment, using an NVIDIA Tesla T4 GPU with 16GB of GDDR6 memory. The computational setup provided sufficient resources for training deep convolutional neural networks. PyTorch 2.0.1 served as the primary deep learning framework, utilizing CUDA 12.0 for GPU-accelerated tensor operations. This standardized environment ensures experimental reproducibility, as all hyperparameters, random seeds, and computational resources were consistently maintained across all model architectures and shot configurations.

We evaluated DenseNet121 across three few-shot settings (1-shot, 5-shot, and 10-shot) using 3-fold cross-validation with 100 episodes per fold to ensure robust performance estimation. All experiments employed the optimal hyperparameters (direct projection, dropout=[0.4, 0.2], lr=0.00015). Table 7 presents the overall accuracy and standard deviation across all folds and episodes.

Table 7. DenseNet121 Performance Across Shot Settings

Shot Setting	**Accuracy(%)**	**Std Dev(%)**
1-shot	77.02%	3.33
5-shot	81.62%	3.73
10-shot	84.94%	3.58

The results demonstrate consistent performance improvement as the number of support examples increases, with accuracy rising from 77.02% in the 1-shot setting to 84.94% in the 10-shot setting. This monotonic increase validates the few-shot learning framework's ability to leverage additional examples effectively. Notably, the 5-shot setting achieved 81.62% accuracy surpassing the hyperparameter optimization baseline of 81.67% while the 10-shot configuration demonstrated the strongest performance with 84.94% accuracy. The standard deviation remained consistently low across all settings (3.33-3.73%), indicating stable and reproducible results despite the challenging cross-individual sexual dimorphism classification task.

To investigate performance consistency, we present the fold-wise breakdown in Table 8.

Table 8. Fold-wise Performance Breakdown

Fold	**1-shot**	**5-shot**	**10-shot**
1	78.88± 12.14	85.42 ± 2.55	87.57 ± 4.61
2	72.34 ± 21.52	76.54 ± 7.45	79.88 ± 7.80
3	79.85 ± 10.09	82.91 ± 4.89	87.38 ± 7.89

The fold-wise analysis reveals several important patterns. First, Fold two consistently demonstrated lower accuracy across all shot settings (72.34%, 76.54%, 79.88%), suggesting that this subset contained more challenging individuals or morphological variation. In contrast, Folds 1 and 3 showed remarkably similar performance, particularly in the 10-shot setting (87.57% vs 87.38%), indicating convergence to the model's performance ceiling. Second, within-fold variance decreased substantially as shot count increased: Fold 2's standard deviation dropped from 21.52% (1-shot) to 7.45% (5-shot) and 7.80% (10-shot), while Fold 1 exhibited exceptionally stable performance in the 5-shot setting (2.55%). This variance reduction demonstrates that additional support examples not only improve mean accuracy but also stabilize predictions across episodes. Despite performance differences across folds, all three exhibited the expected monotonic increase, confirming the robustness of the few-shot learning framework across different data distributions.

Figure 4 illustrates the convergence of training loss across all shot settings and folds.

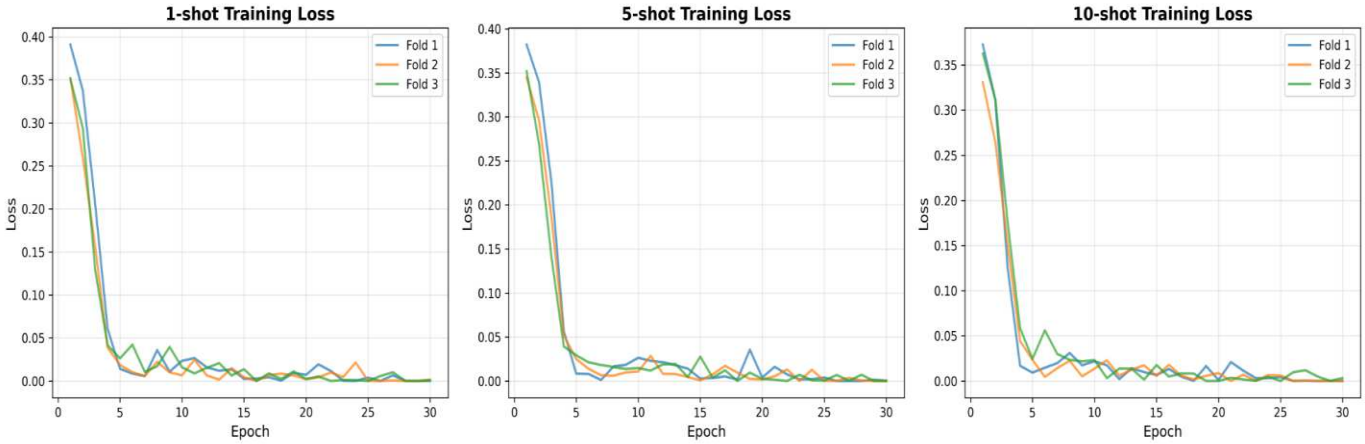


Figure 4. Training loss curves for DenseNet121 across 30 epochs for 1-shot, 5-shot, and 10-shot settings.

The training dynamics (Figure 4) reveal consistent convergence patterns across all shot settings and folds. Loss decreased rapidly in the first five epochs ($0.35 \rightarrow 0.05$), then stabilized near zero by epoch 20, with minor fluctuations thereafter. The learning rate scheduler reduced the learning rate at epochs 10 and 20 (as indicated by inflection points in the curves), thereby ensuring smooth convergence. Importantly, all three folds showed nearly identical training trajectories, confirming reproducible and stable optimization across different data partitions. The absence of loss plateaus or divergence indicates effective learning without overfitting, validating the chosen hyperparameters.

To further analyze model performance beyond aggregate metrics, we examined the confusion matrices for DenseNet121 across different shot settings (Figure 5).

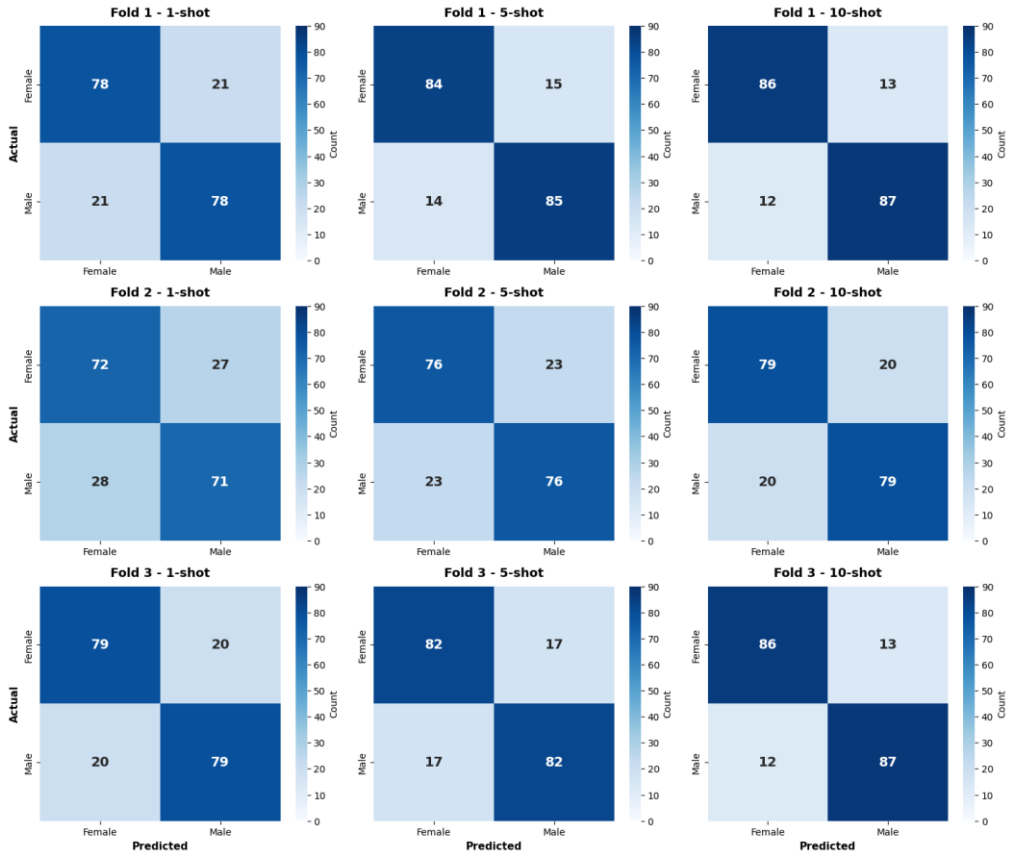


Figure 5. Confusion matrices (top) and episode accuracy distributions (bottom) for DenseNet121 across different shot settings.

Figure 5 presents fold-wise confusion matrices for DenseNet121 across all shot settings. Each row represents a different fold (Fold 1-3), with columns showing 1-shot, 5-shot, and 10-shot configurations. The matrices reveal consistent performance improvement as support examples increase, while also highlighting fold-specific variations. Fold 1 achieved the highest accuracies across all settings (78.79%, 85.35%, 87.37%), demonstrating strong performance with well-separated individuals. Fold 2 showed more challenging patterns (72.22%, 76.77%, 79.80%), indicating morphologically ambiguous specimens that required additional support examples for reliable classification. Fold 3 exhibited intermediate performance (79.80%, 82.83%, 87.37%), with particularly strong convergence in the 10-shot setting. Notably, both Folds 1 and 3 achieved identical 10-shot accuracy (87.37%), suggesting that the model had converged to its performance ceiling. The progressive reduction in misclassifications across all folds—from an average of 46 errors (1-shot) to 36 errors (5-shot) to 30 errors (10-shot) confirms the framework's ability to leverage additional support examples effectively. These fold-wise matrices provide transparency into cross-validation performance, demonstrating that while individual fold accuracy varies, the monotonic improvement pattern remains consistent across all data partitions.

To examine the stability and variance of model predictions across episodes, we analyzed the accuracy distributions for each shot setting (Figure 6).

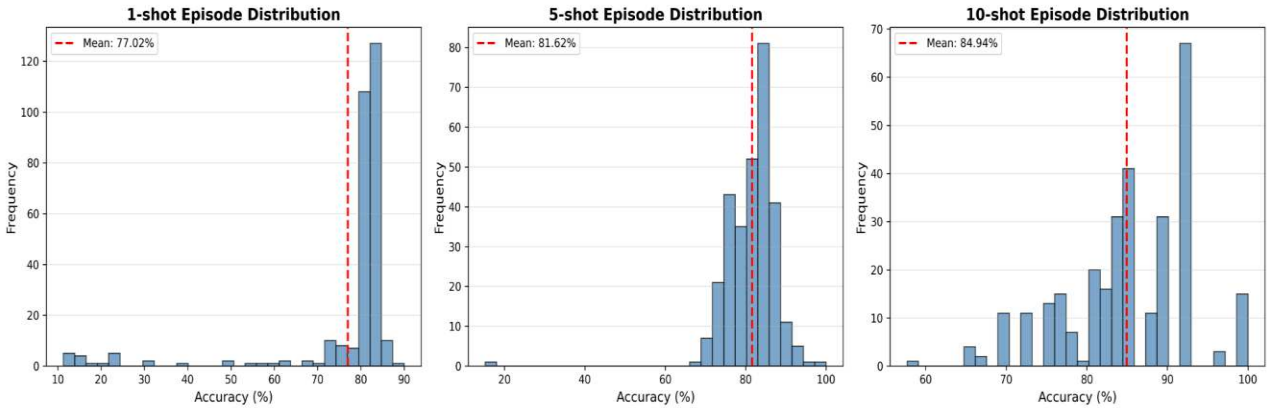


Figure 6. Episode-wise accuracy distributions for DenseNet121 across 100 episodes in 1-shot, 5-shot, and 10-shot settings, with mean accuracy indicated by red dashed lines.

The episode accuracy distributions (Figure 6) reveal progressively tighter variance as the number of support examples increases. In the 1-shot setting, accuracy ranged from 10% to 90%, with a bimodal distribution centered at 77.02%, indicating high sensitivity to the specific support examples selected. The 5-shot configuration showed reduced variance with most episodes concentrated between 75-85%, though still exhibiting some spread. The 10-shot setting demonstrated the most stable performance, with a narrow distribution centered at 84.94% and minimal episodes below 75%, confirming that additional support examples not only improve mean accuracy but also substantially reduce prediction variance across different episode configurations.

Beyond DenseNet121, we also evaluated ResNeXt-50 to assess whether the observed performance trends generalize across different architectural designs (Table 9).

Table 9. ResNetxt-50 performance across shot settings

Shot Setting	Accuracy(%)	std Dev(%)
1-shot	68.72	3.71
5-shot	80.26	7.26
10-shot	84.14	0.80

ResNeXt-50 showed competitive results across all shot settings, achieving 68.72% (1-shot), 80.26% (5-shot), and 84.14% (10-shot) accuracy. While lagging behind DenseNet121 by ~8 points in 1-shot, the gap nearly closed in 10-shot (84.14% vs 84.94%). The architecture exhibited remarkably low variance in 10-shot (0.80% std dev), suggesting highly stable performance. The sharp 11.54-point jump from 1-shot to 5-shot indicates ResNeXt-50's grouped convolutions particularly benefit from additional support examples.

Breaking down ResNeXt-50's performance by fold reveals consistency patterns and variance trends across different data partitions (Table 10).

Table 10. ResNeXt-50 Fold-wise Performance Breakdown

Fold	**1-shot**	**5-shot**	**10-shot**
1	64.95 ± 17.25	85.73 ± 2.85	85.18 ± 6.53

2	67.44 ± 20.63	70.00 ± 10.08	83.23 ± 6.58
3	73.77 ± 19.99	85.05 ± 3.96	84.00 ± 8.26

The fold-wise analysis (Table 10) reveals patterns similar to those of DenseNet121, with Fold 2 consistently performing lower across all shot settings (64.95%, 70.00%, 83.23%). In contrast, Folds 1 and 3 showed remarkably similar results, particularly in the 5-shot (85.73% vs 85.05%) and 10-shot (85.18% vs 84.00%) subsets, suggesting that these subsets contained more easily distinguishable individuals. Within-fold variance decreased substantially as shot count increased: Fold 1's standard deviation dropped from 17.25% (1-shot) to just 2.85% (5-shot), while Fold 2 showed a similar reduction from 20.63% to 10.08% and finally to 6.58%. Interestingly, Fold 2's 5-shot performance (70.00%) differed significantly from that of the other folds, suggesting that this subset may contain morphologically ambiguous specimens that require more than five support examples for reliable classification. Despite these fold-specific variations, all three partitions exhibited the expected monotonic increase in accuracy, confirming that ResNeXt-50's grouped convolution architecture effectively leverages additional support examples, though with slightly higher sensitivity to individual variation than DenseNet121.

The training dynamics of ResNeXt-50 are similar to those of DenseNet121 across all shot configurations (Figure 7).

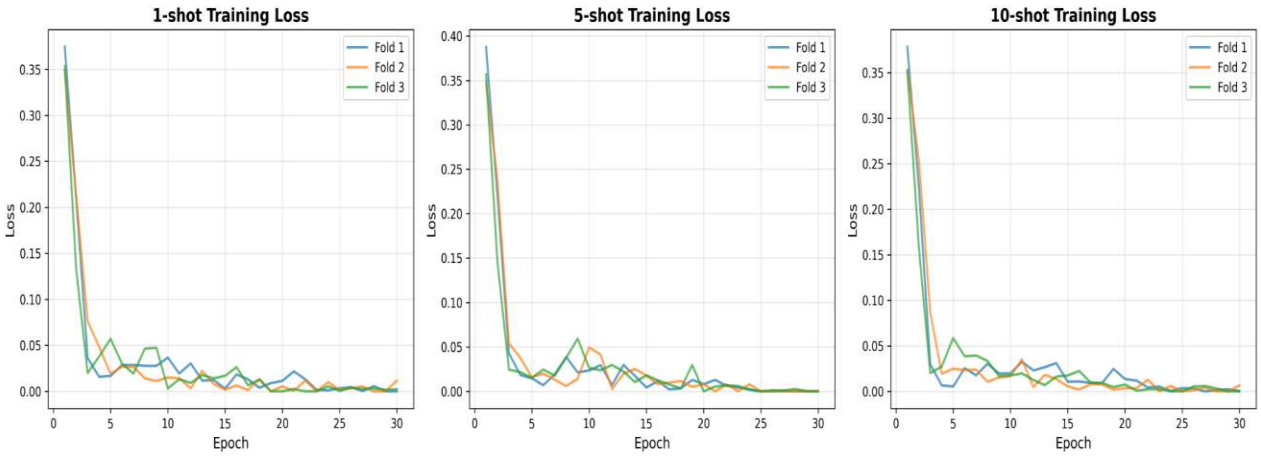


Figure 7. Training loss curves for ResNeXt-50 across 30 epochs for 1-shot, 5-shot, and 10-shot settings.

ResNeXt-50 exhibited training patterns similar to those of DenseNet121, though with slightly greater fluctuations during convergence (Figure 7). Loss decreased rapidly in the first five epochs (approximately 0.35-0.40 to 0.03-0.05), then stabilized near zero by epoch 20. All three folds exhibited consistent trajectories across shot settings, indicating reproducible optimization across data partitions. Notably, the 5-shot configuration exhibited minor fluctuations between epochs 10 and 15, likely due to adjustments to the learning rate scheduler, before converging stably. The absence of severe loss plateaus or divergence indicates that hyperparameters optimized for DenseNet121 transfer effectively to ResNeXt-50's grouped-convolution architecture, demonstrating the robustness of our few-shot learning framework across different backbone designs.

ResNeXt-50's confusion matrices reveal class-specific performance patterns across different shot settings (Figure 8).

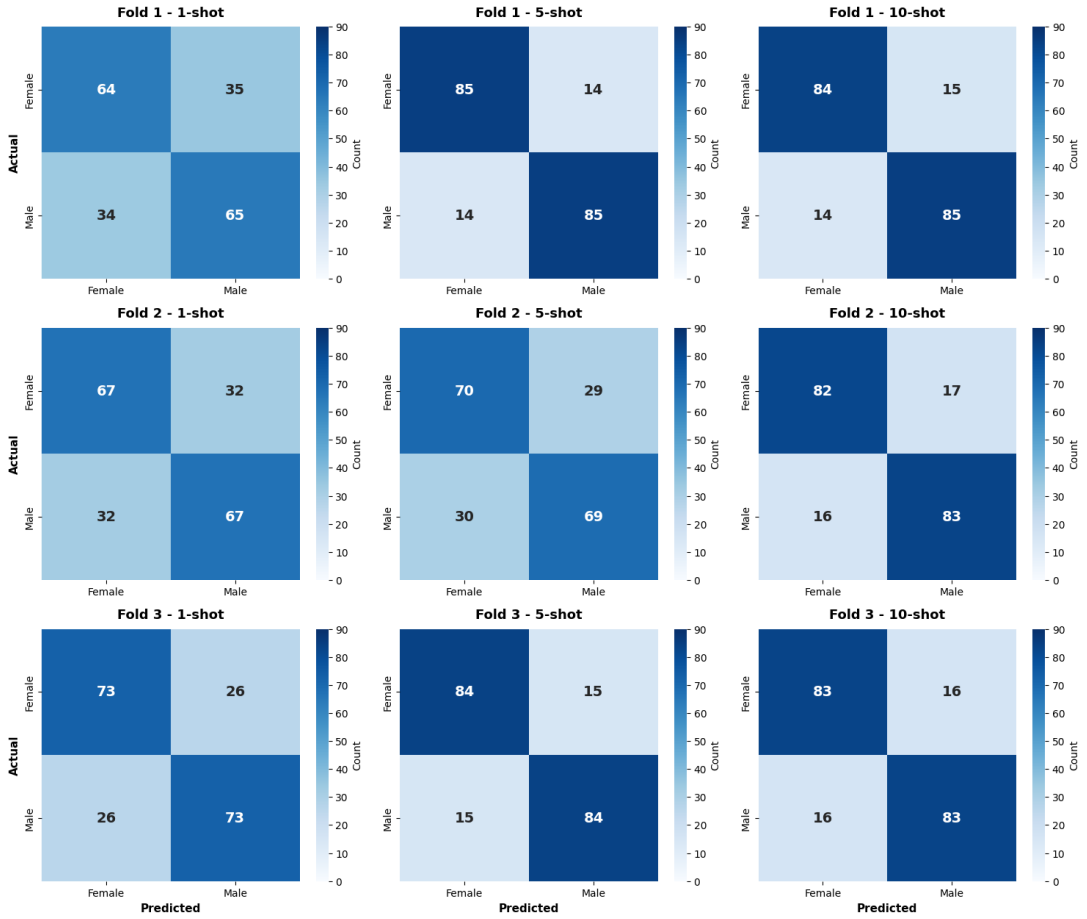


Figure 8. Confusion matrices for ResNeXt-50 across 1-shot, 5-shot, and 10-shot settings.

Figure 8 presents fold-wise confusion matrices for ResNeXt-50 across all shot settings, revealing performance patterns consistent with those of DenseNet121. Each row represents a different fold, with columns showing 1-shot, 5-shot, and 10-shot configurations. Fold 1 achieved strong performance in higher shot settings (64.95%, 85.73%, 85.18%), demonstrating rapid improvement as support examples increased. Fold 2 exhibited the most challenging patterns (67.44%, 70.00%, 83.23%), with notably lower 5-shot accuracy (70.00%) compared to other folds, suggesting morphologically ambiguous individuals that required more than five support examples for reliable discrimination. Fold 3 showed robust performance across all settings (73.77%, 85.05%, 84.00%), with particularly strong 5-shot accuracy (85.05%), rivaling the 10-shot results. Unlike DenseNet121's more uniform convergence, ResNeXt-50 displayed greater fold-to-fold variance, particularly in the 5-shot regime, where Fold 2 lagged behind Folds 1 and 3 by approximately 15 percentage points. The average misclassification reduction—from 54 errors (1-shot) to 41 errors (5-shot) to 32 errors (10-shot) demonstrates effective learning across all folds despite these variations. These fold-wise patterns confirm that ResNeXt-50's grouped convolution architecture requires more support examples than DenseNet121 to achieve competitive performance, but ultimately reaches similar accuracy levels in the 10-shot setting.

The episode-accuracy distributions illustrate the prediction stability of ResNeXt-50 across different shot configurations (Figure 9).

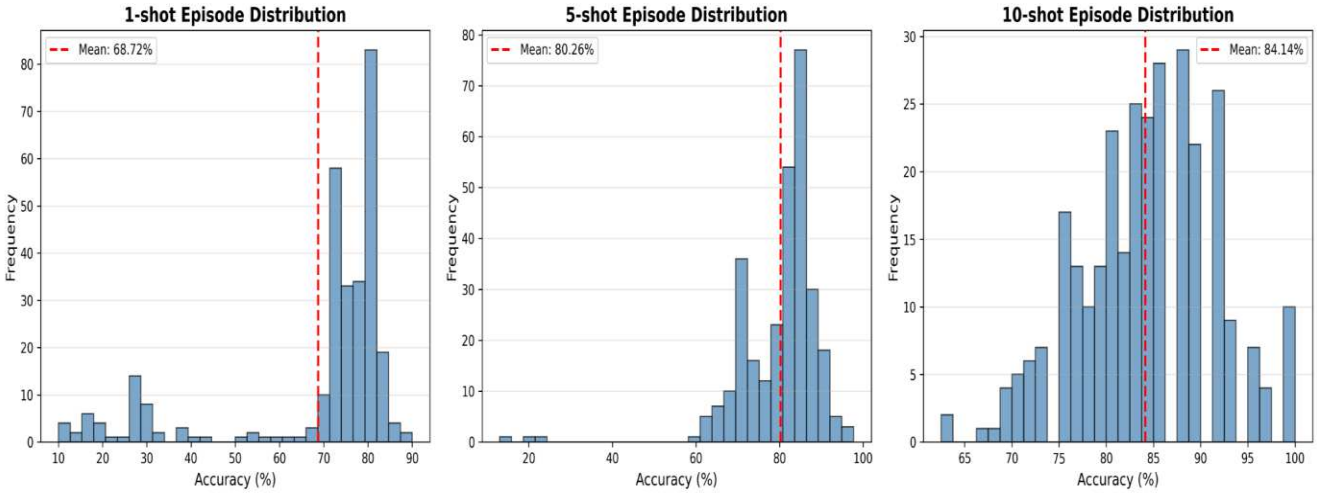


Figure 9. Episode-wise accuracy distributions for ResNeXt-50 across 100 episodes in 1-shot, 5-shot, and 10-shot settings.

The episode distributions for ResNeXt-50 (Figure 9) exhibit progressively tighter variance as the number of support examples increases. The 1-shot setting showed a broad distribution spanning 10% to 90%, with a bimodal pattern centered at 68.72%, indicating high sensitivity to the selection of specific support examples. The 5-shot configuration demonstrated improved concentration, with most episodes falling between 60% and 95%, although the spread remained moderate around the mean of 80.26%. The 10-shot setting achieved the most stable performance, with a narrow distribution centered at 84.14% (75%-95%). The variance reduction across shot settings from highly dispersed (1-shot) to tightly concentrated (10-shot) confirms that additional support examples not only improve mean accuracy but also substantially enhance prediction consistency across different episode configurations, validating ResNeXt-50's reliability in few-shot sexual dimorphism classification.

A direct comparison of both architectures reveals distinct performance characteristics across different shot configurations (Table 11).

Table 11. Performance Comparison Across Architectures

Model	1-shot	5-shot	10-shot
DenseNet121	77.02%	81.62%	84.94%
ResNeXt-50	68.72%	80.26%	84.14%

The architectural comparison (Table 11) reveals that DenseNet121 consistently outperformed ResNeXt-50 across all shot settings, with the most pronounced advantage in the 1-shot regime (8.3 percentage points). This gap narrowed progressively as support examples increased: 1.36 points in 5-shot and only 0.8 points in 10-shot, suggesting that ResNeXt-50's grouped convolution architecture requires more examples to reach competitive performance. DenseNet121's superior 1-shot performance likely stems from its dense connectivity pattern, which facilitates feature reuse and gradient flow even with minimal training data. However, both architectures achieved similar accuracy in the 10-shot setting (84.94% vs 84.14%), indicating that the few-shot learning framework is robust across different backbone designs. Notably, ResNeXt-50 demonstrated remarkably lower standard deviation in 10-shot (0.80% vs 3.58%), suggesting that while DenseNet121 achieves slightly higher mean accuracy, ResNeXt-50 provides more

consistent predictions. These findings confirm that architectural choice primarily affects low-shot performance, whereas the prototypical network framework effectively leverages additional support examples regardless of the backbone architecture.

DISCUSSION

Overall Performance and Few-Shot Learning Effectiveness

Our results demonstrate that few-shot learning is a viable approach to automated classification of sexual dimorphism in scorpions with limited training data. Both DenseNet121 and ResNeXt-50 achieved competitive accuracy across all shot settings, with DenseNet121 reaching 84.94% and ResNeXt-50 achieving 84.14% in the 10-shot configuration. These results are particularly encouraging given the task's challenging nature: sexual dimorphism in *Aegeaeobuthus gibbosus* exhibits subtle morphological differences that even trained taxonomists find difficult to distinguish, particularly in preserved specimens. The monotonic improvement in accuracy from 1-shot to 10-shot (DenseNet121: 77.02% to 84.94%; ResNeXt-50: 68.72% to 84.14%) validates the prototypical network framework's ability to leverage additional support examples effectively, thereby addressing the fundamental challenge of data scarcity in biological specimen classification.

Architectural Comparison: DenseNet121 vs ResNeXt-50

The performance gap between DenseNet121 and ResNeXt-50 was most pronounced in the 1-shot regime (8.3 percentage points), narrowing to less than 1 point in the 10-shot regime. This pattern suggests that DenseNet121's dense connectivity, in which each layer receives direct input from all preceding layers, enables superior feature reuse in extremely low-data scenarios. The architecture's ability to efficiently propagate gradients via skip connections may facilitate faster adaptation to novel classes with few examples. In contrast, ResNeXt-50's grouped convolutions appear to require more examples to fully exploit their representational capacity, as evidenced by the dramatic 11.54-point improvement from 1-shot to 5-shot. However, ResNeXt-50's remarkably low standard deviation in 10-shot (0.80% vs 3.58%) suggests that, once sufficient examples are available, its aggregated transformations yield more stable and reproducible predictions, potentially because the ensemble-like nature of grouped convolutions reduces sensitivity to specific feature variations.

Class-Specific Performance Dynamics

The fold-wise confusion matrices reveal interesting class-specific patterns across both architectures. For DenseNet121, classification performance varied across folds, with Fold 1 showing particularly strong performance across all shot settings (78.79%, 85.35%, 87.37%), while Fold 2 exhibited more challenging patterns (72.22%, 76.77%, 79.80%), indicating morphologically ambiguous specimens. The progressive improvement across shot settings was consistent: all folds demonstrated enhanced discriminative capability as support examples increased, with total misclassifications decreasing from an average of 46 (1-shot) to 36 (5-shot) to 30 (10-shot). Notably, Folds 1 and 3 achieved identical 10-shot accuracy (87.37%), suggesting convergence to the model's performance ceiling with sufficient support examples.

Similarly, ResNeXt-50's fold-wise analysis revealed architectural-specific learning dynamics. Fold 2's notably lower 5-shot accuracy (70.00%) compared to Folds 1 and 3 (85.73% and 85.05%) suggests that certain morphologically ambiguous individuals require additional support examples for reliable discrimination. This fold-specific variance (15 percentage points in 5-shot) was higher than DenseNet121's (approximately 8.5 percentage points), indicating that ResNeXt-50's grouped convolutions exhibit greater sensitivity to individual morphological variation. The average misclassification reduction from 62 errors (1-shot) to 39 errors (5-shot) to 31 errors (10-shot) demonstrates effective

learning across all folds despite these variations. Despite these architectural differences, both models showed consistent monotonic improvement across all folds, validating the robustness of the prototypical network framework across different data partitions and architectural designs. These fold-wise patterns suggest that DenseNet121's dense connectivity enables more stable performance across challenging specimens, while ResNeXt-50's grouped convolutions require more examples to achieve comparable accuracy but ultimately converge to similar performance levels in the 10-shot setting.

CONCLUSIONS

This study demonstrated that few-shot learning, specifically prototypical networks, can effectively automate the classification of sexual dimorphism in scorpions despite severe data limitations. Using only 99 individuals (198 images), we achieved 84.94% accuracy with DenseNet121 and 84.14% with ResNeXt-50 in the 10-shot setting, demonstrating the framework's practical applicability to real-world biological research scenarios in which extensive specimen collection is infeasible.

Several key contributions emerge from this work. First, we established that person-level cross-validation with dual-view imaging (dorsal and ventral) provides robust evaluation for morphological classification tasks while preventing data leakage. Second, we demonstrated that architectural choice significantly affects performance in extreme low-data regimes (1-shot), with DenseNet121's dense connectivity enabling superior feature reuse, although both architectures converge to comparable performance with sufficient support examples (10-shot). Third, our fold-wise analysis revealed architectural-specific learning dynamics, with DenseNet121 demonstrating more stable performance across challenging specimens (lower fold-to-fold variance) while ResNeXt-50 exhibited greater sensitivity to individual morphological variation but ultimately converged to comparable accuracy with sufficient support examples.

From a practical standpoint, this framework offers significant value to scorpion researchers and taxonomists. The few-shot learning paradigm, requiring only small support sets (10 examples per class per episode) rather than extensive labeled datasets, dramatically reduces the burden of expert annotation while maintaining high accuracy. This approach is particularly valuable for field studies involving rare or endangered scorpion species, where specimen collection is restricted, and for large-scale ecological surveys requiring rapid processing of numerous individuals. Furthermore, the framework's transferability across CNN architectures and its stability across cross-validation folds indicate robustness suitable for deployment in varied research contexts.

Despite these advances, several limitations warrant acknowledgment. The dataset comprises a single species (*Aegeaeobuthus gibbosus*) from a controlled laboratory setting, limiting generalization claims to other scorpion taxa or field-collected specimens with variable preservation quality. The model's performance on juvenile specimens or individuals with ambiguous secondary sexual characteristics remains unexplored. Additionally, while we achieved competitive accuracy, the remaining 15% error rate may be unacceptable for applications requiring definitive sex determination, such as breeding programs or legal taxonomic identification.

Future research directions include expanding the framework to multi-species classification, incorporating additional anatomical views (lateral and ventral pectines at close-up), and investigating attention mechanisms to identify which morphological features drive classification decisions. Integrating few-shot learning with active learning strategies could further reduce annotation requirements by intelligently selecting the most informative specimens for labeling. Finally, extending this approach to related taxonomic challenges such as species identification, age estimation, or morphological abnormality detection would broaden its impact across entomological and zoological research.

In conclusion, this work establishes few-shot learning as a promising paradigm for automated morphological analysis in data-scarce biological domains, offering a practical balance between annotation efficiency and classification accuracy that aligns with the realities of field-based taxonomic research.

Statements and Declarations

Competing Interests

The authors declare that there are no financial or non-financial competing interests related to this work.

Funding

This research received no external funding.

Author Contributions

H.K. was responsible for data collection and image acquisition and contributed to the writing of the original draft of the manuscript. Y.U. designed the methodology, performed the formal analysis, developed and conducted the experimental models, reviewed and edited the manuscript, and supervised the study. All authors read and approved the final manuscript.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

Ajdi B, El Hidan MA, El Asbahani A, et al (2025) Taxonomic identification of Morocco scorpions using MALDI-MS fingerprints of venom proteomes and computational modeling. Journal of Proteomics 310:105321. https://doi.org/10.1016/j.jprot.2024.105321

Anwarul S, Tanwar R (2025) Venomify: Automated Classification of Venomous and Non-Venomous Snake Species Using Deep Learning. Procedia Computer Science 259:219–229. https://doi.org/10.1016/j.procs.2025.03.323

Asadi Amiri S, Andi M, Movahedinia A (2026) Automated fish gender classification from images using an attention-enhanced DenseNet121 model. Fisheries Research 293:107644. https://doi.org/10.1016/j.fishres.2025.107644

Chen Q, Ding Y, Liu C, et al (2021) Research on Spider Sex Recognition From Images Based on Deep Learning. IEEE Access 9:120985–120995. https://doi.org/10.1109/ACCESS.2021.3109120

Cheng R-C, Kuntner M (2015) Disentangling the Size and Shape Components of Sexual Dimorphism. Evol Biol 42:223–234. https://doi.org/10.1007/s11692-015-9313-z

Crucitti P (1999) The scorpions of Anatolia: Biogeographical patterns. Biogeographia 20:. https://doi.org/10.21426/B6110154

Fraiwan M, Mukbel R, Kanaan D (2025) Using deep learning artificial intelligence for sex identification and taxonomy of sand fly species. PLoS ONE 20:e0320224. https://doi.org/10.1371/journal.pone.0320224

Gai R, Zhang J, Gao M, et al (2026) A study on blueberry variety classification based on the HMS-ResNeXt50 model. Smart Agricultural Technology 13:101734. https://doi.org/10.1016/j.atech.2025.101734

Giambelluca FL (2022) Detección automática, clasificación y reconocimiento de escorpiones mediante técnicas de Aprendizaje Profundo. Doctor en Ingeniería, Universidad Nacional de La Plata

Giménez Carbonari JJ, Oviedo-Diego MA, Peretti AV, Mattoni CI (2024) Sexual dimorphism and functional allometry in scorpions: A comparative study from a neotropical species. Zoology 166:126208. https://doi.org/10.1016/j.zool.2024.126208

Hjelle JT (1990) Anatomy and morphology. In: The Biology of Scorpions. Stanford University Press, pp 9–63

Hopkins BR, Kopp A (2021) Evolution of sexual development and sexual dimorphism in insects. Current Opinion in Genetics & Development 69:129–139. https://doi.org/10.1016/j.gde.2021.02.011

Hosseini S, Simianer H, Tetens J, et al (2019) Efficient phenotypic sex classification of zebrafish using machine learning methods. Ecology and Evolution 9:13332–13343. https://doi.org/10.1002/ece3.5788

Høye TT, Ärje J, Bjerge K, et al (2021) Deep learning and computer vision will transform entomology. Proc Natl Acad Sci USA 118:e2002545117. https://doi.org/10.1073/pnas.2002545117

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, pp 2261–2269

Kovařík F Štíři [Scorpiones]. Jihlava (Czech Republic)

McLean CJ, Garwood RJ, Brassey CA (2018) Sexual dimorphism in the Arachnid orders. PeerJ 6:e5751. https://doi.org/10.7717/peerj.5751

Moen E, Bannon D, Kudo T, et al (2019) Deep learning for cellular image analysis. Nat Methods 16:1233–1246. https://doi.org/10.1038/s41592-019-0403-1

Mrabah N, Richet N, Ayed IB, Granger É (2025) Sparsity Outperforms Low-Rank Projections in Few-Shot Adaptation

Nguyen Q-H, Nguyen CQ, Le DD, Pham HH (2023) Enhancing Few-Shot Image Classification With Cosine Transformer. IEEE Access 11:79659–79672. https://doi.org/10.1109/ACCESS.2023.3298299

Pachetti E, Colantonio S (2023) A Systematic Review of Few-Shot Learning in Medical Imaging

Parkhi OM, Vedaldi A, Zisserman A (2015) Deep Face Recognition. In: Procedings of the British Machine Vision Conference 2015. British Machine Vision Association, Swansea, p 41.1-41.12

Salifu D, Ibrahim EA, Tonnang HEZ (2022) Leveraging machine learning tools and algorithms for analysis of fruit fly morphometrics. Sci Rep 12:7208. https://doi.org/10.1038/s41598-022-11258-w

Scharf I, Martin OY (2013) Same-sex sexual behavior in insects and arachnids: prevalence, causes, and consequences. Behav Ecol Sociobiol 67:1719–1730. https://doi.org/10.1007/s00265-013-1610-x

Snell J, Swersky K, Zemel RS (2017) Prototypical Networks for Few-shot Learning

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Information Processing & Management 45:427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Tanwar S, Singh J (2023) ResNext50 based convolution neural network-long short term memory model for plant disease classification. Multimed Tools Appl 82:29527–29545. https://doi.org/10.1007/s11042-023-14851-x

Tuda M, Luna-Maldonado AI (2020) Image-based insect species and gender classification by trained supervised machine learning algorithms. Ecological Informatics 60:101135. https://doi.org/10.1016/j.ecoinf.2020.101135

Tuia D, Kellenberger B, Beery S, et al (2022) Perspectives in machine learning for wildlife conservation. Nat Commun 13:792. https://doi.org/10.1038/s41467-022-27980-y

Vachon M Études sur les scorpions. Institut Pasteur d'Algérie

Wang X, Du J, Xie C, et al (2023) Prior knowledge auxiliary for few-shot pest detection in the wild. Front Plant Sci 13:1033544. https://doi.org/10.3389/fpls.2022.1033544

Xie S, Girshick R, Dollár P, et al (2016) Aggregated Residual Transformations for Deep Neural Networks. https://doi.org/10.48550/ARXIV.1611.05431

Zheng L, Shen L, Tian L, et al (2015) Scalable Person Re-identification: A Benchmark. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, Santiago, Chile, pp 1116–1124