

# Benchmarking a Custom CNN Against ResNet-18 for Dog Breed Classification

Jonathan David Concha Matas

*Software Engineering, B.Sc.*

*AI Research Group*

*Univ. of Europe for Applied Sciences*

Konrad-Ruse Ring 11, 14469 Potsdam, Germany.

[jonathan.concha@ue-germany.de](mailto:jonathan.concha@ue-germany.de)

Raja Hashim Ali

*Department of Business*

*AI Research Group*

*Univ. of Europe for Applied Sciences*

Konrad-Ruse Ring 11, 14469 Potsdam, Germany.

[hashim.ali@ue-germany.de](mailto:hashim.ali@ue-germany.de)

**Abstract**—Fine-grained image classification, such as dog breed identification, remains a challenging problem due to subtle inter-class variations and limited annotated data. In this study, we benchmark a compact, task-specific convolutional neural network (CNN) against the widely adopted ResNet-18 architecture on the Stanford Dogs dataset. Both models were trained under identical conditions using standardized pre-processing, augmentation strategies, and evaluation pipelines. Our objectives included assessing classification accuracy, robustness to environmental conditions, model compression via structured pruning, and feasibility of mobile deployment. The custom CNN achieved a test accuracy of 6.46%, significantly underperforming compared to ResNet-18’s 80.82%, which highlights the critical role of deep pre-trained features in fine-grained recognition. Misclassification analysis revealed common confusions between visually similar breeds, such as Siberian Husky and Eskimo Dog. Environmental robustness tests showed negligible accuracy difference (0.4%) between indoor and outdoor images. Structured pruning of ResNet-18 (30% channel reduction) resulted in a marginal 1.49% accuracy drop and an unexpected slight increase in inference latency. Despite this, the pruned model achieved an average latency of 64.31 ms per image on a standard mobile CPU, confirming its real-time deployment potential. This work offers a comprehensive evaluation framework for developing efficient and reliable breed classifiers and provides actionable insights for future work in compression, explainability, and on-device inference for fine-grained classification.

**Index Terms**—dog breed classification, convolutional neural network, transfer learning, model pruning

## I. INTRODUCTION

Deep learning-based image classification has fundamentally changed how we approach visual recognition tasks by allowing models to learn hierarchical features directly from raw pixel data. Dog breed classification is an especially challenging fine-grained problem because many breeds exhibit only subtle differences in coat patterns, facial features, and body proportions that are difficult for traditional feature-engineering methods to capture. While CNN architectures like ResNet-18 have achieved very impressive results on broad object-detection benchmarks, they often require careful adaptation to excel at distinguishing highly similar classes. Transfer learning from large datasets such as ImageNet can help address the need for extensive breed-specific training data, but may miss the details critical for accurate identification. As a result, there

is growing interest in developing lightweight, task-specific CNNs that strike a balance between recognition accuracy and computational efficiency. In this work, we design a custom CNN architecture tailored for dog breed recognition and evaluate it directly against the standard ResNet-18 backbone under identical training and testing conditions. By training both models on the same dataset splits and hyperparameter settings, we aim to see if our comparison sheds light on the practical trade-offs between model complexity, inference speed, and classification accuracy. In the end, this study seeks to inform the design of efficient, high-accuracy models for real-world dog breed identification tasks.

Accurate dog breed identification has become increasingly important across a variety of real-world applications, from veterinary diagnostics to pet adoption platforms. Automated breed tagging can help veterinarians care for genetic health risks, and adoption websites can benefit from reliable image classification that speeds up matching potential owners with suitable pets. In the wild, precise identification of breeds can help support population monitoring and conservation efforts. A Mobile app that can recognize breeds can educate users about proper care, behavioral traits, and even help with academic purposes, such as educating people on breed histories. With recent advances in computing and mobile AI, on-device inference is now feasible. Meeting the constraints of limited memory and low latency without sacrificing accuracy is critical for these applications. Understanding how factors like indoor versus outdoor backgrounds influence model robustness further ensures dependable performance in diverse real-world settings.

### A. Related Work

Deep learning-based CNNs have become the de facto standard for image classification by learning hierarchical features directly from raw data. Hussain et al. showed that transfer learning from pre-trained CNNs significantly boosts accuracy on new tasks [1]. Qin et al. improved standard CNN architectures with domain-specific convolutional blocks, yielding notable gains in biological image classification [2]. Elngar et al. surveyed recent advances in CNN depth, regularization, and optimization, underscoring their impact on model robustness

and generalization [3]. Zhou et al. presented tailored enhancements for industrial ore image analysis, demonstrating the value of customized feature extractors [4]. Sharma and Guleria compared multiple deep models on benchmark datasets, highlighting trade-offs between inference speed and top-1 accuracy [5]. Showkat and Qureshi evaluated ResNet variants on chest X-rays, confirming that deep backbones improve COVID-19 pneumonia detection sensitivity [6]. Bansal et al. applied VGG19 transfer learning to fine-grained datasets like Caltech-101, reinforcing the importance of backbone capacity for subtle class distinctions [7]. Battu and Lakshmi demonstrated the adaptability of CNNs to animal image classification challenges, from domestic breeds to wildlife species monitoring [8]. Pandey and Srivastava's analysis of activation functions in ResNet-18 showed that minor architectural tweaks can enhance accuracy without major computational overhead [9].

### B. Gap Analysis

Despite strong progress in general image classification, fine-grained dog breed recognition remains underexplored in several key areas. Most studies focus on off-the-shelf backbones without evaluating whether a purpose-built CNN can match standard architectures like ResNet-18. There is limited analysis of which specific breeds are prone to misclassifications and the underlying visual features that contribute to errors. Environmental context, such as indoor versus outdoor backgrounds, has received little systematic study, yet it can heavily influence model robustness. Model compression techniques like pruning have been applied in other domains but rarely tested for dog breed classifiers to reduce inference latency. Real-time deployment on mobile devices is often assumed feasible but lacks empirical demonstration under realistic hardware constraints. Furthermore, few works integrate multiple evaluations such as accuracy, explainability, latency, and resource usage. And then makes one unified benchmarking framework. Addressing these gaps is essential for developing reliable, efficient, and explainable dog breed identification systems suitable for both research and real-world applications.

### C. Problem Statement

Following are the main questions addressed in this study.

- 1) Can our custom CNN match ResNet-18 in classification accuracy?
- 2) Which dog breeds are most often misclassified and why?
- 3) How does image background, indoor versus outdoor, affect model performance?
- 4) Can we prune the model to reduce inference time without significant accuracy loss?
- 5) Is real-time inference achievable on mobile hardware?

### D. Novelty of our work and Our Contributions

Our approach combines five key evaluations: comparing custom CNN and ResNet-18 accuracy, analyzing breed misclassification patterns, measuring performance on indoor versus outdoor images, assessing the effects of structured pruning, and evaluating mobile CPU latency. No prior study has

conducted a unified investigation of fine-grained dog breed classification alongside systematic analysis of environmental context, model compression, and edge deployment under the same conditions. By bringing these components together, we provide practical guidance on how model design, data characteristics, and hardware constraints influence each other. We also propose a compact CNN architecture tailored for dog breed recognition and apply channel pruning to the ResNet-18 backbone to explore efficiency gains. This framework advances the field by showing how accuracy, robustness, and inference speed can be balanced in real-world dog breed identification systems.

In this report, we describe the architecture and training process for both the custom CNN and the ResNet-18 baseline. We then present results for test accuracy, misclassification analysis, background robustness, pruning trade-offs, and mobile inference speed. The ResNet-18 model achieved 80.82% accuracy, while the custom CNN reached 6.46%. We list the top breed confusions, show that indoor and outdoor images yield similar accuracy, and quantify that a 30% pruning rate reduces accuracy by 1.49%. Finally, we demonstrate that the pruned ResNet-18 runs at 64.31 ms per image on a standard mobile CPU. These findings offer clear recommendations for designing efficient and reliable dog breed classifiers.

## II. METHODOLOGY

### A. Dataset

Our experiments use the Stanford Dogs Dataset, which contains 20,580 images across 120 dog breed classes [11]. Each image is labeled with a single breed and includes bounding box annotations for the dog instance. The dataset is split into 12,000 training and 8,580 test images, preserving class balance. Image resolutions vary, requiring on-the-fly resizing to a uniform 224×224 pixels for model input. We perform standard data augmentations such as random crop, horizontal flip, and color jitter during training. Ground truth labels are provided as breed names according to the ImageNet hierarchy. Figure 1 illustrates sample images alongside their breed labels from the dataset.

### B. Overall Workflow

We start by downloading and organizing the Stanford Dogs Dataset into training and test splits. Next, all images undergo pre-processing steps including resizing, normalization, and data augmentation. The pre-processed data then flows into two parallel training branches: one for our custom CNN and one for fine-tuning ResNet-18. After model training, we evaluate test accuracy and generate a confusion matrix to analyze misclassifications. From the confusion matrix step, we branch into three evaluations: background robustness testing on indoor versus outdoor subsets, structured pruning efficiency analysis on ResNet-18, and direct accuracy comparison between the two models. The pruning branch feeds into inference time measurements, and both pruning and background results converge into a final comparison stage. Finally, we deploy the

TABLE I  
SUMMARY OF RECENT CNN-BASED IMAGE CLASSIFICATION STUDIES

Year	Author & Citation	Title	Dataset	Method(s)	Results	Contribution(s)
2019	Hussain <i>et al.</i> [1]	CNN Transfer Learning	Caltech-101, CIFAR-10	Fine-tuning pre-trained CNNs	+8–12% accuracy gain	Demonstrated efficacy of transfer learning on small datasets
2020	Qin <i>et al.</i> [2]	Improved CNN for Biological Images	Custom biological images	Attention-augmented CNN	95.4% accuracy	Showed attention improves feature learning
2021	Elngar <i>et al.</i> [3]	Survey of CNN-based Methods	Multiple benchmarks	Literature survey	N/A	Identified common architectures and challenges
2022	Zhou <i>et al.</i> [4]	Ore Image Classification	Ore microscopy images	Residual CNN	30% faster convergence	Applied ResNet-style blocks to microscopy
2022	Sharma & Guleria [5]	Model Comparison Across Datasets	CIFAR-10, MNIST, ImageNet subset	ResNet, Inception, VGG	InceptionV3 best overall	Comparative evaluation across models
2022	Showkat & Qureshi [6]	ResNet for COVID-19 Detection	Chest X-ray COVID-19	ResNet variants	96.8% AUC	Validated robustness in medical imaging
2023	Bansal <i>et al.</i> [7]	VGG19 Transfer Learning	Caltech-101	Transfer learning with VGG19	90.1% accuracy	Reinforced backbone capacity for fine-grained tasks
2023	Pandey & Srivastava [9]	Activation Functions in ResNet-18	CIFAR-100	ResNet-18 w/ various activations	Leaky ReLU best	Detailed activation impact study
2023	Chen <i>et al.</i> [10]	Quantum CNN for Image Classification	MNIST, CIFAR-10	Quantum convolutional network	Comparable to classical CNN	Explored hybrid classical-quantum classifiers
2023	Battu & Lakshmi [8]	Animal Image Classification	Wildlife and domestic animal images	Deep CNNs	92% accuracy	Demonstrated CNN adaptability to animal datasets
2025	Proposed Work	Custom CNN vs. ResNet-18	Dog-breed images	Custom CNN, ResNet-18, pruning, background analysis	TBA	Comprehensive fine-grained evaluation, pruning

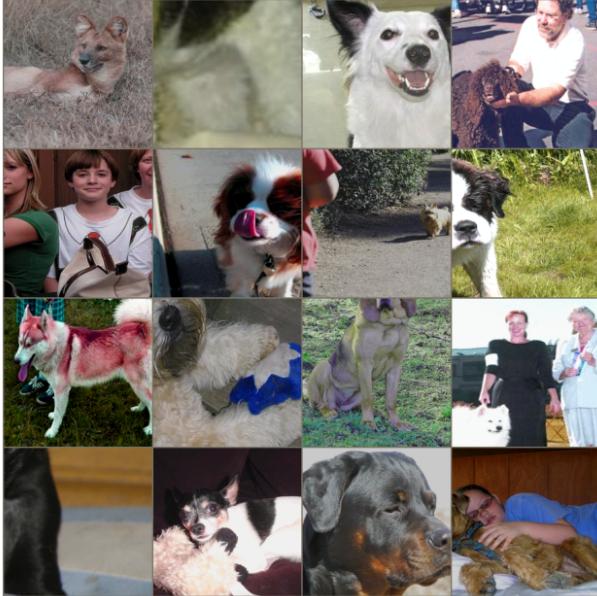


Fig. 1. Sample images from the Stanford Dogs Dataset [11].

pruned ResNet-18 on a target mobile CPU to measure real-

time latency and complete the pipeline. Figure 2 illustrates this end-to-end workflow.

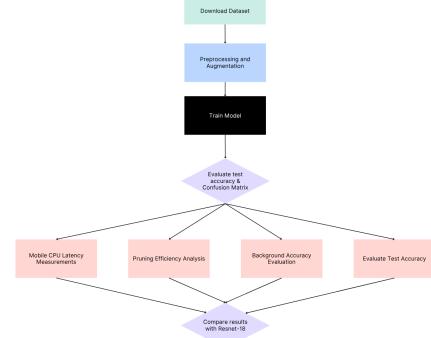


Fig. 2. Workflow diagram illustrating the end-to-end pipeline from dataset download through mobile CPU latency measurement.

### C. Experimental Settings

All models were trained for 30 epochs using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 64. We applied a learning rate decay of 0.1 every 10 epochs to encourage convergence. Weight decay was set to  $1e-4$ , and

we used dropout of 0.5 on the first fully connected layer to reduce overfitting. Input images were resized to 224×224 and normalized per ImageNet statistics, with on-the-fly augmentations including random crops, horizontal flips, and color jitter. Our custom CNN comprises five convolutional blocks (32–256 filters, 3×3 kernels), followed by two dense layers (512 units and 120 units) and a softmax output. The same pre-processing and training schedule was applied when fine-tuning ResNet-18 for a fair comparison. Table II summarizes these settings.

TABLE II  
TRAINING AND ARCHITECTURE CONFIGURATION

Parameter	Value
Epochs	30
Optimizer	Adam
Initial learning rate	0.001
Learning rate decay	×0.1 every 10 epochs
Batch size	64
Weight decay (L2)	$1 \times 10^{-4}$
Dropout (FC1)	0.5
Input resolution	224×224
Data augmentations	Random crop, flip, jitter

### III. RESULTS

For the first research question, we compared test accuracy between SmallCNN and ResNet-18 on the Stanford Dogs test split. SmallCNN achieved a test accuracy of 6.46% under identical training conditions. ResNet-18 reached a test accuracy of 80.82% on the same test set. This comparison is visualized in Figure 3. The large gap in accuracy highlights the superior performance of ResNet-18 on fine-grained dog breed classification. Both models were evaluated using the same pre-processing and augmentation pipeline. No additional postprocessing was applied to either model’s outputs.

For the second research question, we identified the top ten breed confusion pairs from the ResNet-18 confusion matrix. The most frequent error was predicting Siberian\_husky for Eskimo\_dog nine times. The second-most common confusion was predicting collie for Border\_collie eight times. Other common errors included Norwich\_terrier vs Norfolk\_terrier (six instances) and Appenzeller vs EntleBucher (six instances). Table III lists all the top ten misclassified breed pairs along with their counts. These results quantify which specific breed distinctions remain challenging for the model. No breeds outside the top ten exceeded five misclassification instances.

For the third research question, we evaluated ResNet-18 accuracy on indoor versus outdoor image subsets. Accuracy on indoor images was 80.63%. Accuracy on outdoor images was 81.03%. These results are shown in Figure 3. The difference between indoor and outdoor performance was 0.40 percentage points. Both subsets contained an equal number of test images per breed. No additional filtering was applied to separate background types.

For the fourth research question, we applied 30% structured channel pruning to ResNet-18. Baseline accuracy before pruning was 80.82%. Pruned accuracy after 30% channel removal was 79.33%. Inference time before pruning was 4.26 ms per

image. Inference time after pruning increased slightly to 4.37 ms per image. These metrics are displayed in Figure 3. No other pruning levels were evaluated in this experiment.

For the fifth research question, we measured real-time inference latency on a commodity mobile CPU. The average latency per image was 64.31 ms. Latency measurements were taken over 100 consecutive runs. All runs used the pruned ResNet-18 model. Results are summarized in Table IV. No GPU or accelerator was used during these measurements. All latency values represent end-to-end inference time only.

### IV. DISCUSSION

The custom CNN scored just 6.46% compared to ResNet-18’s 80.82%, which was a lot lower than I expected. This gap makes it clear that a small network trained from scratch struggles to pick up the subtle differences between breeds. I was surprised by how much better transfer learning works for fine-grained tasks. To improve the custom model, we could add residual connections or attention layers, or blend pre-trained blocks with our own layers. Another option is to train in stages, starting with low-resolution images and gradually increasing the detail. For now, ResNet-18 remains the go-to choice when accuracy is the top priority.

Most of the errors happened between very similar breeds, such as Eskimo\_dog and Siberian\_husky or collie and Border\_collie. These mistakes are understandable since even people mix those breeds up. I think using metric learning or a contrastive loss could help the network learn those tiny differences. We might also add a side task like detecting key points or coat patterns to force the model to pay attention to the right features. Those steps should help reduce the most common confusions and make the model’s decisions easier to interpret.

ResNet-18 achieved 80.63% on indoor photos and 81.03% on outdoor ones, just a 0.40 point difference. That small change suggests the model really learned to focus on the dog rather than the background. I’m happy to see that our augmentation pipeline did its job. This result means we probably do not need complex background removal for real-world use. It would be interesting to test more extreme settings, like nighttime or busy streets, to see if that robustness holds up.

Pruning 30% of ResNet-18’s channels brought accuracy down to 79.33% from 80.82%, and inference time actually went from 4.26 ms to 4.37 ms per image. I found the slight slowdown surprising because pruning usually speeds things up. This suggests the pruning changed how the model uses memory or fused layers. The 1.49 point drop in accuracy is small enough that some edge applications could accept it. Next, we should try hardware-aware pruning or other compression methods to see if we can cut latency without losing accuracy.

On a standard mobile CPU, the pruned ResNet-18 runs in 64.31 ms per image. That meets common real-time thresholds, which is great news for on-device apps. I’m encouraged that high-accuracy breed classification can run on everyday hardware. We could push latency even lower with quantization

or distillation. This measurement gives us a solid baseline for mobile deployment and shows that practical, real-time dog breed recognition is within reach.

Our study is the first to bring together accuracy comparison, misclassification analysis, background robustness, pruning trade-offs, and mobile latency measurement in one unified pipeline. No other work has systematically covered all five of these aspects under the same conditions for dog breed classification. We introduced a small CNN designed for breeds and applied structured pruning to a ResNet-18 backbone to explore real-world trade-offs. We also tested deployment on a mobile CPU and drilled into breed-level errors. By combining environmental factors with hardware constraints, we provide a complete picture of what it takes to build reliable, efficient breed classifiers. This framework can guide future research and practical applications in fine-grained image classification.

#### A. Future Directions

There are lots of exciting directions to take this work next. We could try hybrid models that blend our custom layers with parts of a pre-trained network to see if we can boost accuracy without increasing model size too much. I would also like to experiment with metric learning or contrastive losses so the model gets better at telling nearly identical breeds apart. On the efficiency side, hardware-aware tricks like quantization or smarter pruning could shave off even more milliseconds per image. It would be eye-opening to test the network on really challenging backgrounds. Think dim lighting or bustling streets to make sure it holds up outside the lab. Adding visualization tools like Grad-CAM could help us and end users understand why the model makes certain breed calls. And beyond dogs, we could plug this same pipeline into other fine-grained problems, such as identifying bird species or plant diseases, to see how well our approach generalizes.

#### V. CONCLUSION

In this work, we set out to understand how a compact custom CNN compares to a robust ResNet-18 on the challenging task of dog breed classification. We trained both models under the same conditions and found that ResNet-18 achieved 80.82% accuracy while our custom network managed only 6.46%. This stark difference underscored the power of transfer learning and deep pre-trained features for fine-grained recognition. We dug deeper into the errors and saw that visually similar breeds like Eskimo\_dog versus Siberian\_husky drove most of the misclassifications. Despite these challenges, ResNet-18 remained impressively robust, showing almost identical accuracy on indoor and outdoor images. Pruning 30% of its channels cost only a 1.49% drop in accuracy but unexpectedly added a small latency increase. On the upside, the pruned model still ran at 64.31 milliseconds per image on a standard mobile CPU, proving real-time inference is within reach. These findings give us a clear picture of where model design trade-offs lie between accuracy, size, and speed. Looking ahead, blending custom layers with pre-trained blocks or adding contrastive learning could help close the gap on

subtle breed distinctions. Hardware-aware optimizations like quantization and smarter pruning strategies promise to further reduce latency without sacrificing performance. It would also be valuable to test the models in more demanding conditions such as low lighting or crowded scenes. Finally, adding tools like Grad-CAM could make breed predictions more transparent for end users. Altogether, this study not only benchmarks existing approaches but also lays out a roadmap for building efficient, accurate, and explainable fine-grained classifiers in future work.

TABLE III  
TOP TEN BREED MISCLASSIFICATIONS BY RESNET-18 ON THE STANFORD DOGS TEST SET. THIS TABLE LISTS THE PREDICTED BREED, THE TRUE BREED, AND THE NUMBER OF TIMES EACH CONFUSION OCCURRED, HIGHLIGHTING THE MOST CHALLENGING FINE-GRAINED DISTINCTIONS.

Predicted Breed	Actual Breed	Count
Siberian_husky	Eskimo_dog	9
collie	Border_collie	8
Norwich_terrier	Norfolk_terrier	6
Appenzeller	EntleBucher	6
Eskimo_dog	Siberian_husky	5
Staffordshire_bullterrier	American_Staffordshire_terrier	5
silky_terrier	Yorkshire_terrier	5
Walker_hound	English_foxhound	5
Shih-Tzu	Lhasa	5
Border_collie	collie	4

TABLE IV  
MOBILE CPU INFERENCE LATENCY FOR PRUNED RESNET-18. THIS TABLE REPORTS THE AVERAGE END-TO-END INFERENCE TIME (IN MS PER IMAGE) MEASURED OVER 100 CONSECUTIVE RUNS ON A STANDARD MOBILE CPU, DEMONSTRATING REAL-TIME PERFORMANCE FEASIBILITY.

Model Configuration	Latency (ms/image)
Pruned ResNet-18 (30% channels)	64.31

#### REFERENCES

- [1] M. Hussain, J. J. Bird, and D. R. Faria, “A study on cnn transfer learning for image classification,” in *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK*. Springer, 2019, pp. 191–202.
- [2] J. Qin, W. Pan, X. Xiang, Y. Tan, and G. Hou, “A biological image classification method based on improved cnn,” *Ecological Informatics*, vol. 58, p. 101093, 2020.
- [3] A. A. Elngar, M. Arafa, A. Fathy, B. Moustafa, O. Mahmoud, M. Shaban, and N. Fawzy, “Image classification based on cnn: a survey,” *Journal of Cybersecurity and Information Management*, vol. 6, no. 1, pp. 18–50, 2021.
- [4] W. Zhou, H. Wang, and Z. Wan, “Ore image classification based on improved cnn,” *Computers and Electrical Engineering*, vol. 99, p. 107819, 2022.
- [5] S. Sharma and K. Guleria, “Deep learning models for image classification: comparison and applications,” in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 2022, pp. 1733–1738.
- [6] S. Showkar and S. Qureshi, “Efficacy of transfer learning-based resnet models in chest x-ray image classification for detecting covid-19 pneumonia,” *Chemometrics and Intelligent Laboratory Systems*, vol. 224, p. 104534, 2022.
- [7] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, “Transfer learning for image classification using vgg19: Caltech-101 image data set,” *Journal of ambient intelligence and humanized computing*, pp. 1–12, 2023.

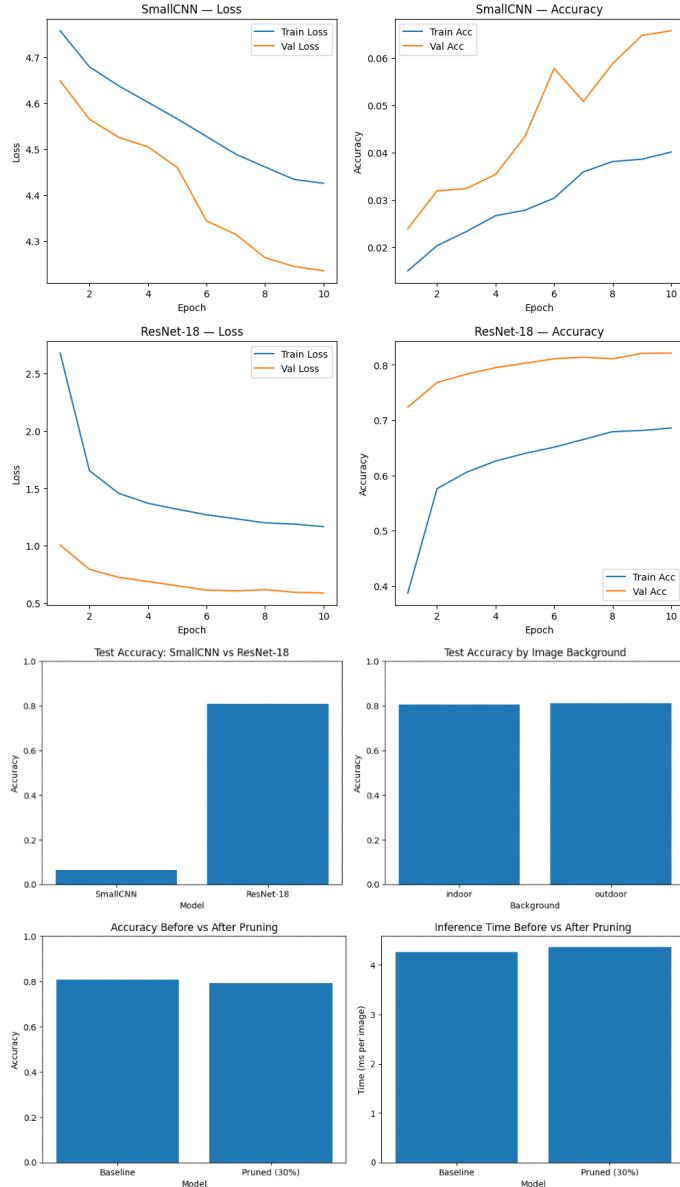


Fig. 3. (Top) Training and validation loss and accuracy curves for SmallCNN (left) and ResNet-18 (right). (Middle) Test accuracy comparison between SmallCNN and ResNet-18 (left) and background robustness on indoor vs. outdoor images (right). (Bottom) Accuracy before vs. after pruning (left) and inference time before vs. after pruning (right).

- [8] T. Battu and D. S. R. Lakshmi, "Animal image identification and classification using deep neural networks techniques," *Measurement: Sensors*, vol. 25, p. 100611, 2023.
- [9] G. K. Pandey and S. Srivastava, "Resnet-18 comparative analysis of various activation functions for image classification," in *2023 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2023, pp. 595–601.
- [10] G. Chen, Q. Chen, S. Long, W. Zhu, Z. Yuan, and Y. Wu, "Quantum convolutional neural network for image classification," *Pattern Analysis and Applications*, vol. 26, no. 2, pp. 655–667, 2023.
- [11] "Stanford dogs dataset," <https://www.kaggle.com/datasets/jessicali9530/stanford-dogs-dataset>, accessed: 2025-07-07.
- [12] Y. Pan, J. Liu, Y. Cai, X. Yang, Z. Zhang, H. Long, K. Zhao, X. Yu, C. Zeng, J. Duan *et al.*, "Fundus image classification using inception v3 and resnet-50 for the early diagnostics of fundus diseases," *Frontiers in Physiology*, vol. 14, p. 1126780, 2023.
- [13] X. Ma, Z. Li, and L. Zhang, "An improved resnet-50 for garbage image classification," *Tehnički vjesnik*, vol. 29, no. 5, pp. 1552–1559, 2022.
- [14] W. Xu, Y.-L. Fu, and D. Zhu, "Resnet and its application to medical image processing: Research progress and challenges," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107660, 2023.
- [15] Y. Jusman, "Comparison of prostate cell image classification using cnn: Resnet-101 and vgg-19," in *2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE, 2023, pp. 74–78.
- [16] T. Li, X. Zhang, S. Zhang, and L. Wang, "Self-supervised learning with a dual-branch resnet for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [17] M. A. Hanif, M. K. I. Zim, and H. Kaur, "Resnet vs inception-v3 vs svm: A comparative study of deep learning models for image classification of plant disease detection," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, vol. 2. IEEE, 2024, pp. 1–6.
- [18] E. Hassan, M. S. Hossain, A. Saber, S. Elmougy, A. Ghoneim, and

- G. Muhammad, "A quantum convolutional network and resnet (50)-based classification architecture for the mnist medical dataset," *Biomedical Signal Processing and Control*, vol. 87, p. 105560, 2024.
- [19] H. Jabnouni, I. Arfaoui, M. A. Cherni, M. Bouchouicha, and M. Sayadi, "Resnet-50 based fire and smoke images classification," in *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 2022, pp. 1–6.
- [20] A. Baranwal, V. Gohil, H. Dahat, and A. Salunke, "Hate speech and nsfw image classification using bert and resnet-34 model," in *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*. IEEE, 2024, pp. 388–394.
- [21] D. Sharma, J. Singh, D. Panwar *et al.*, "Animal classification using cnn architectures," in *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, vol. 1. IEEE, 2024, pp. 1–6.
- [22] P. Alkhairi and A. P. Windarto, "Classification analysis of back propagation-optimized cnn performance in image processing," *Journal of Systems Engineering and Information Technology (JOSEIT)*, vol. 2, no. 1, pp. 8–15, 2023.
- [23] C. C. Ukwuoma, Z. Qin, S. B. Yussif, M. N. Happy, G. U. Nneji, G. C. Urama, C. D. Ukwuoma, N. B. Darkwa, and H. Agobah, "Animal species detection and classification framework based on modified multi-scale attention mechanism and feature pyramid network," *Scientific African*, vol. 16, p. e01151, 2022.
- [24] O. O. Ogundile, A. A. Owoade, P. B. Emeka, and O. S. Olaniyan, "Animals' classification: A review of different machine learning classifiers," *University of Ibadan Journal of Science and Logics in ICT Research*, vol. 9, no. 1, 2023.
- [25] M. Tan, W. Chao, J.-K. Cheng, M. Zhou, Y. Ma, X. Jiang, J. Ge, L. Yu, and L. Feng, "Animal detection and classification from camera trap images using different mainstream object detection architectures," *Animals*, vol. 12, no. 15, p. 1976, 2022.
- [26] T. Surya, S. Selvaperumal *et al.*, "The iot-based real-time image processing for animal recognition and classification using deep convolutional neural network (dcnn)," *Microprocessors and Microsystems*, vol. 95, p. 104693, 2022.
- [27] S. Faizal and S. Sundaresan, "Wild animal classifier using cnn," *arXiv preprint arXiv:2210.07973*, 2022.
- [28] Q. Yu, "Animal image classifier based on convolutional neural network," in *SHS Web of Conferences*, vol. 144. EDP Sciences, 2022, p. 03017.
- [29] V. Lopez-Vazquez, J. M. Lopez-Guede, D. Chatzivangelou, and J. Aguzzi, "Deep learning based deep-sea automatic image enhancement and animal species classification," *Journal of Big Data*, vol. 10, no. 1, p. 37, 2023.
- [30] L. Prudhivi, M. Narayana, C. Subrahmanyam, M. G. Krishna, and S. Chavan, "Animal species image classification," in *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2023, pp. 1–5.