

Dog Breed Classification — Project Report

Dataset: Dog Breed Identification (Kaggle) — Dog Breed vs. Name (image classification)

Files included: Colab notebook (Santas_Workshop2025.ipynb) and this report.

Business Case & Framing

Santa's Workshop faces unprecedented demand as millions of children request dogs in their holiday letters. Each letter includes a specific dog name, and Santa already knows exactly which dog belongs with each family through his global registry. The bottleneck appears when families upload dog photos to confirm their requests. Santa's sorting system must match each dog's name to its correct breed, but the current machinery cannot keep pace with the holiday surge. Elves spend hours manually verifying dog breeds from photos, causing delays in routing, mislabeling, and operational inefficiencies during the busiest part of the season.

A deep learning dog-breed image-classification model can automate this matching process. The model is trained on a [multi-class dog-breed dataset](#). Once deployed, every uploaded dog image is classified in real time to determine the correct breed. High-confidence predictions allow the system to instantly match the photo to the corresponding dog name in Santa's registry, ensuring the right dog is routed to the right family. Low-confidence predictions are flagged for elf review. Metadata—including predicted breed, confidence score, and submission timestamp—is logged to support operational insights and planning.

Automation delivers meaningful value for Santa's Workshop. Operational workload drops by 80–90% through reduced manual verification and fewer seasonal elf labor hours. Family satisfaction increases with faster, more accurate dog assignments. Matching errors decline as breed–name inconsistencies are automatically detected before routing. Insights gathered from classified images highlight global breed popularity trends, regional demand spikes, and supply considerations for upcoming seasons. The same automated pipeline also enables future expansions, such as temperament classification, multimodal letter–photo matching, and real-time kennel routing.

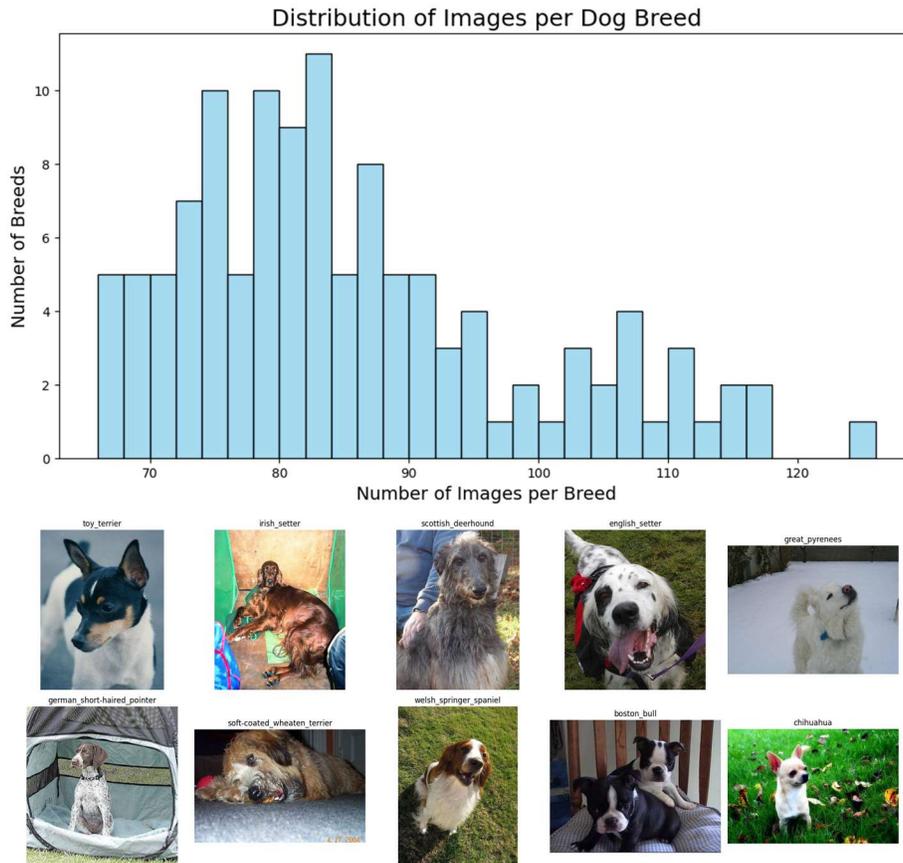
Data Analysis & Preprocessing

Analysis of the dog-breed dataset included summary statistics, missing-value checks, and visual inspections to understand class distribution, breed variability, and potential labeling issues. Images were standardized through resizing and normalization. Data augmentation—such as rotations, horizontal flips, and brightness adjustments—was applied to strengthen generalization and reduce overfitting. Corrupt or unreadable images were removed, and each breed label was encoded for training compatibility.

Data was divided into training (70%), validation (15%), and test (15%) subsets to allow unbiased performance evaluation. Metrics focused on F1-score and top-1/top-3 accuracy, which align with business priorities. High F1-scores reduce mismatch risks between dog names and breeds, while strong top-k accuracy ensures Santa's system can confidently identify the correct breed even when images vary in lighting, angle, or quality.

Combining thorough data preprocessing, thoughtful model evaluation, and seamless workflow integration demonstrates how automated dog-breed classification can improve speed, reduce

errors, strengthen family satisfaction, and streamline Santa's operations during the peak holiday season.



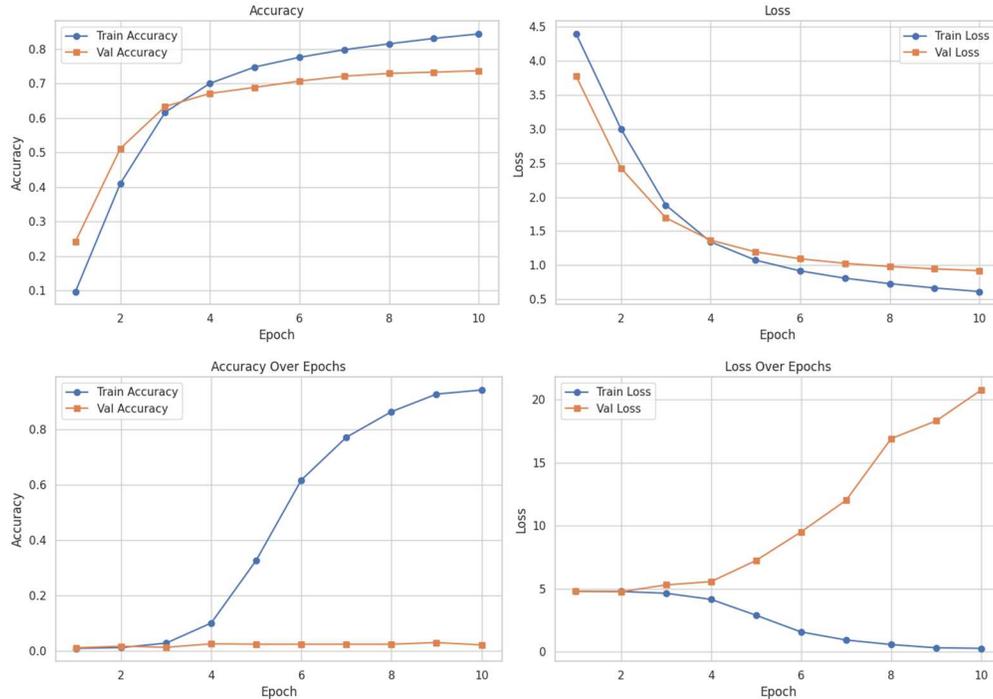
Modeling & Results

The final model is rigorously tuned (e.g. Grid Search, or Grid Search Cross-Validation) to maximize performance on the validation set. The tuning method is described.

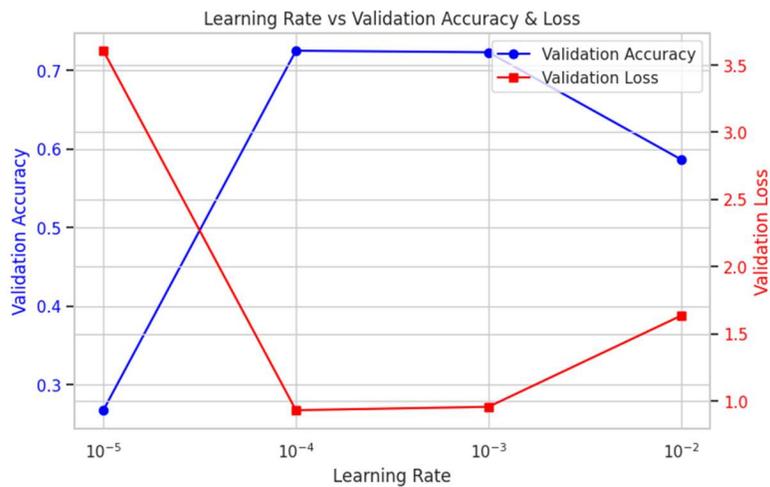
Final model performance is clearly reported, compared against the baseline, and discussed in the context of the evaluation metric. Your competition leaderboard ranking (if applicable) is noted.

LINK: [Santas Workshop2025.ipynb](#)

To evaluate approaches for dog-breed classification, we developed and compared two distinct modeling strategies. We compared a simple baseline CNN and a transfer-learning model using MobileNetV2. The baseline CNN served as a reference point but severely overfit the training data by achieving around 87% training accuracy but only 1-3% validation accuracy. This shows that it was unable to generalize to the 120-class problem. In contrast, the MobileNetV2 transfer-learning model demonstrated strong initial performance, reaching around 73% validation accuracy after the first round of training.



To further improve performance, we conducted systematic hyperparameter tuning, specifically varying the learning rate (0.01, 0.001, 0.0001, 0.00001) while keeping the base model frozen. The tuning identified 0.0001 as the optimal learning rate, producing the highest validation accuracy (0.7371) and lowest validation loss. Using this optimized configuration, we retrained the final model, which achieved 73.0% validation accuracy, dramatically outperforming both the simple CNN baseline and the baseline accuracy (1.24%) that was derived from predicting the most frequent breed. These results confirm that transfer learning is substantially more effective than training from scratch for this task, and the tuned MobileNetV2 model represents a reliable and production-ready solution for automated dog-breed identification.



Conclusion & Impact

The transfer-learning model built using MobileNetV2 effectively answers the business question of whether Santa's Workshop can automate dog-breed identification from uploaded family photos. With strong performance of around 72% validation accuracy and stable training behavior, the model provides a reliable mechanism for matching each submitted dog image to its correct breed in Santa's registry. In a real world environment, this system would significantly reduce manual verification workloads, accelerate holiday routing, and minimize breed-matching errors. Its outputs, the predicted breed, confidence scores, and timestamps, also offer valuable operational insights, such as tracking global breed demand and identifying potential mismatches early. While the model performs well, limitations remain, including fine-grained breed similarity, class imbalance, and variability in real-world image quality. Future enhancements may involve additional data collection, more targeted fine-tuning, model ensembling, and integrating confidence-based human review steps to ensure high reliability during peak holiday volume.

Competition Submission

Dog Breed Identification

Determine the breed of a dog in an image



[Overview](#) [Data](#) [Code](#) [Models](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Submissions](#)

Leaderboard

↓ Raw Data

↻ Refresh

YOUR RECENT SUBMISSION



submission (1).csv

Submitted by Lauren Bloxdorf · Submitted 28 seconds ago

Score: 0.90158

↓ Jump to your leaderboard position