

Identifying Dog Breeds

DS4002 Case Study- Caroline Hagood



In a world where technology increasingly intersects with everyday life, machine learning models for identifying dog breeds have the potential to revolutionize pet care and ownership. Veterinary practices can use such models to quickly assess breed-specific medical needs. Meanwhile, pet owners can use apps that identify dog breeds from photos to learn more about their pets, ensuring proper care, training, and nutrition. These tools could also help dog enthusiasts recognize breeds they encounter in parks, shelters, or during rescues.

Each day, thousands of dogs are adopted from shelters across the country, and many new pet owners rely on breed identification to understand their dog's characteristics and needs. Some shelters are beginning to use new technology to identify the exact breed type of dogs. Accurate breed identification helps families choose dogs that best suit their lifestyles, promotes responsible ownership, and provides valuable insights into training, behavior, and health.

Your task as a data scientist is to create a machine learning breed identification system based on images of different breeds of dogs. You are presented with a dataset of dozens of different breed folders containing images of dogs. By analyzing the dataset and applying cutting-edge convolutional neural network (CNN) modeling techniques, your goal is to build a robust image classification model that predicts the breed of a dog from a given image. Along the way, you must consider how data preprocessing, augmentation, and model training choices impact performance.

Imagine you're designing a system for a national shelter database, and accuracy is necessary. You must deliver the model with proof that it achieves high prediction accuracy of at least 80% and generalizes well to unknown data. You will also explore possible limitations of the model and recommend improvements or strategies to scale the solution for broader adoption. By the end of this project, you will deliver a detailed case study documenting your approach, findings, and final model performance.