Automated Dog Emotion Recognition for Veterinary Applications Using Vision Transformers

Presented by: Adeel Rehan

MS Artificial Intelligence • Fall 2025

PROBLEM STATEMENT

The Challenge of Understanding Animal Emotions

Accurately assessing canine emotional states is crucial for effective veterinary care and animal welfare. Yet current approaches face significant limitations:

- Emotional states judged subjectively by humans
- Interpretations vary between veterinarians and caretakers
- Subtle emotional cues easily missed or misread
- No standardized automated tools available

This gap creates inconsistencies in care and missed opportunities for early intervention.

RESEARCH GOALS

Project Objectives

Automated Recognition

Develop a system to identify dog emotions from images without human interpretation

Vision Transformers

Apply state-of-the-art deep learning architecture for image-based emotion classification

Veterinary Compliance

Ensure alignment with real-world veterinary domain requirements and workflows

Performance Benchmark

Achieve validation accuracy ≥ 75% to demonstrate clinical viability

Practical Impact

Demonstrate AI's potential to enhance animal welfare and veterinary decision-making

Domain & Dataset

Veterinary AI Domain

Dataset Source: Dog Emotion Dataset (Kaggle)

The dataset comprises RGB facial images of dogs labeled across four distinct emotional states, providing a foundation for supervised learning.

Data Preprocessing Pipeline

01

Image Resizing

Standardized to 224×224 pixels for model input consistency

02

Normalization

Pixel values scaled to improve training stability

03

Data Augmentation

Strong augmentation techniques to enhance model generalization

04

Stratified Split

85% training / 15% validation to maintain class balance

Angry

Happy

Relaxed

Sad

ARCHITECTURE

Methodology: Vision Transformer Model

The Vision Transformer (ViT) represents a paradigm shift in computer vision, applying transformer architecture—originally designed for natural language—to image classification tasks.

Patch Division

Images divided into fixed-size patches

Token Embedding

Patches converted into learnable embeddings

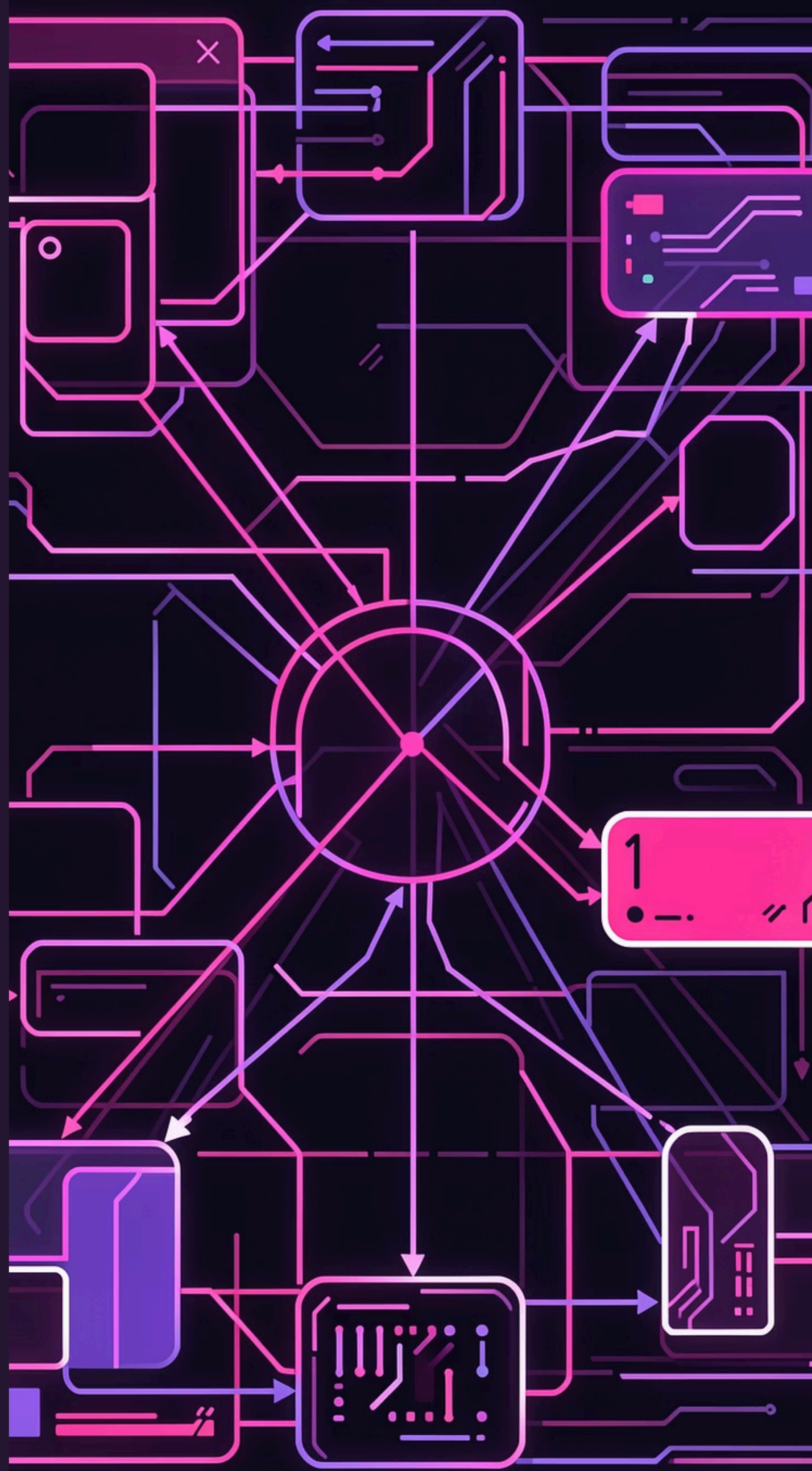
Transformer Layers

Multi-head self-attention processes tokens

Classification

Head predicts final emotion category

Key Constraint: No pretrained models used—single-model approach trained from scratch to demonstrate foundational learning capabilities.



Training Configuration & Setup

Framework

PyTorch
Industry-standard deep learning library for research flexibility

Loss Function

Cross-Entropy Loss
Standard for multi-class classification problems

Optimizer

Adamax
Adaptive learning rate optimization for stable convergence

Hardware

GPU (CUDA-enabled)
Accelerated training on parallel computing architecture

Training Strategy

A carefully designed approach ensures robust model performance and prevents overfitting:

- **Stratified Split:** Maintains class distribution across training and validation sets
- **Regularization:** Techniques applied to enhance generalization
- **Early Stopping:** Monitors validation performance to prevent overtraining

Evaluation Metrics

A comprehensive evaluation framework ensures rigorous assessment of model performance across multiple dimensions:

Accuracy

Overall correctness of predictions across all emotion classes

Precision (weighted)

Proportion of correct positive predictions, adjusted for class imbalance

Recall (weighted)

Ability to identify all relevant instances of each emotion

F1-Score (weighted)

Harmonic mean of precision and recall for balanced assessment

ROC-AUC

Model's ability to distinguish between emotion classes across thresholds

✓ RESULTS

Model Performance & Validation

75.67%	0.7621	0.7567	0.7587
Validation Accuracy	**Precision**	**Recall**	**F1-Score**
Exceeds 75% requirement	Weighted average	Weighted average	Weighted average

✔️ **Course requirement satisfied:** The model achieved validation accuracy ≥ 75%, demonstrating its viability for veterinary applications and establishing a strong baseline for automated dog emotion recognition.

Detailed confusion matrix and ROC curve analyses provide deeper insights into class-specific performance and model discrimination capabilities across all four emotion categories.

Impact, Challenges & Demonstration

Sustainable Development Goals

Technical Challenges

- **Limited Dataset Size:** Small sample constrains model learning capacity
- **Visual Ambiguity:** Similar patterns between emotional states create classification difficulty

Live Demonstration

The demo showcases end-to-end system functionality:

1. Jupyter notebook execution with full training pipeline
2. Comprehensive model evaluation metrics
3. Visual analysis via confusion matrix
4. ROC curve interpretation for multi-class performance

SDG 3: Good Health and Well-Being

Supports veterinary behavioral assessment and early detection of animal distress

SDG 15: Life on Land

Promotes animal welfare through AI-driven insights and compassionate care

Conclusion & Future Directions

Key Findings

This research successfully demonstrates that Vision Transformers can be effectively applied to dog emotion recognition, achieving clinically relevant accuracy levels without pretrained models.

The system validates the feasibility of automated emotion detection for veterinary applications, opening new pathways for AI-assisted animal welfare assessment.

Future Research Opportunities

1 **Dataset Expansion**

Larger, more diverse datasets with multiple breeds and environmental contexts

2 **Real-Time Systems**

Deploy lightweight models for live emotion monitoring in clinical settings

3 **Clinical Integration**

Embed into veterinary decision-support systems for enhanced diagnostic workflows