An Explainable Pit Bull-Type Dog Breed Identification System Using Vision-Language Models and Official Breed Standards

Yu-Kuan Chiang
Department of Computer Science and Information Engineering
National Central University
Taoyuan City, Taiwan
chiangyukuan@g.ncu.edu.tw

Ching-Han Chen
Department of Computer Science and Information Engineering
National Central University
Taoyuan City, Taiwan
pierre@g.ncu.edu.tw

Wei-Jhe Huang
Department of Computer Science and Information Engineering
National Central University
Taoyuan City, Taiwan
jackhuang.miat@gmail.com

***Abstract*— Following Taiwan's designation of pit bull-type dogs as regulated breeds in 2022, front-line enforcement officers must distinguish regulated breeds from visually similar non-regulated ones. To address this challenge, we propose an explainable identification system that utilizes a Vision-Language Model (Gemma3 27B-IT) to extract eight morphological features from multi-view dog images and classifies breeds via weighted Euclidean distance against prototype vectors derived from official breed standards published by Taiwan's Ministry of Agriculture. Evaluation on 784 images demonstrates a binary regulatory classification precision of 93.08%, outperforming a direct VLM classification baseline (90.35%), and the proposed system additionally generates human-readable reports citing specific breed standard criteria, validating its practical applicability for enforcement support.**

***Keywords*— Breed identification, Vision-Language Model, explainable AI, morphological features, weighted distance classification.**

I. Introduction

In 2022, Taiwan's Council of Agriculture designated pit bull-type dogs as regulated breeds, requiring owners to register and comply with specific management measures [1]. However, enforcement faces a fundamental challenge: four closely related breeds—the American Pit Bull Terrier (APBT), American Staffordshire Terrier (AmStaff), Staffordshire Bull Terrier (SBT), and American Bully (Bully) share highly similar phenotypic characteristics, as shown in Fig. 1, yet only APBT and AmStaff are regulated. Prior research has shown that even trained shelter staff identify pit bull-type dogs inconsistently [2], highlighting the inherent unreliability of subjective visual inspection, and the need for an objective, evidence-based classification tool.

While deep learning approaches using Convolutional Neural Networks (CNNs) have achieved significant success in dog breed classification [3], they function as black-box models that can only output a class label without explaining the reasoning, which is problematic in regulatory contexts where decisions must be legally justifiable and transparent.

To address the limitations, we propose an explainable breed identification system integrating a Vision-Language Model (VLM) for structured morphological feature extraction from multi-view photographs. Classification is achieved by comparing these morphological attributes against idealized breed standards via a weighted distance algorithm. Our approach encodes expert knowledge synthesized from the Taiwan Ministry of Agriculture (MOA) guidelines [4] into quantitative feature vectors, which enable the automatic generation of human-readable reports.



Fig.1. Four pit bull-type breeds targeted in this study. Top-left: American Pit Bull Terrier (APBT), top-right: American Staffordshire Terrier (AmStaff), bottom-left: Staffordshire Bull Terrier (SBT), bottom-right: American Bully. Images sourced from [4].

II. Proposed Method

The overall system architecture illustrated as Fig.2, which consists of three main modules.

A. VLM-Based Feature Extraction

The system accepts three photographs per subject: front view, side view, and head close-up. After client-side image quality assessment (resolution ≥ 640×640, sharpness and brightness checks), the images are sent to a Gemma3-27B-IT model [5] running locally via Ollama on a single NVIDIA GeForce RTX 4090 GPU. A structured prompt instructs the VLM to analyze specific body parts from appropriate viewpoints (e.g., shoulder height from the side view, nose color from the head close-up) and output eight normalized relative scores in the range [0, 1] as JSON: shoulder height, body weight, muzzle-to-head ratio, nose color, eye color, chest width-to-depth ratio, body squareness, and head breadth index.

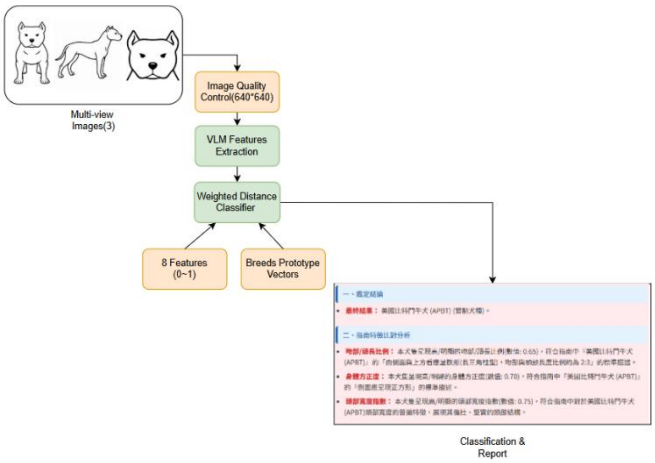


Fig.2. System architecture overview

B. *Weighted Prototype Distance Classifier*

For each of the four target breeds, we define a prototype feature vector p_k and a weight vector w_k ($k \in$ {APBT, AmStaff, SBT, Bully}). These vectors were constructed by referencing the official morphological standards from the Taiwan MOA guideline [4] and international kennel club specifications (UKC, AKC, FCI). Numerical values were calibrated by analyzing a small set of breed-verified reference images (confirmed via Embark DNA testing) through the same VLM pipeline.

Given an input feature vector x, the classification computes the weighted Euclidean distance to each prototype is shown in (1):

$$d_k = \sqrt{\sum_{i=1}^{8} w_{k,i} \cdot \left(x_i - p_{k,i}\right)^2} \qquad (1)$$

The breed with the minimum distance is selected. A user adjustable confidence mechanism allows reducing the weight of occluded features (e.g., if the dog wears clothing that hides the chest). A rejection threshold filters out non-target breeds when the minimum distance exceeds a predefined value.

C. *Explainable Report Generation*

After classification, the system generates a structured Chinese-language report by prompting the VLM with the extracted features, the classification result, and relevant excerpts from the MOA guideline. The report explicitly maps each observed feature to the corresponding breed standard criterion, providing auditable justification (e.g., "the muzzle-to-head ratio of 0.68 is consistent with the APBT standard of approximately 2:3"). This contrasts with CNN-based classifiers that provide no interpretable rationale.

III. Experimental Results

We evaluated the system on 784 images provided by Taiwan's Council of Agriculture, covering the four target breeds with confirmed identities. Due to dataset limitations (single-viewpoint photographs), the same image was used for all three input views during testing; a full evaluation under the intended multi-view configuration remains future work. As a baseline, we directly prompted Gemma3-27B-IT to classify each image into one of the four breeds (Method A). We then compared this against our proposed pipeline of VLM feature extraction followed by weighted distance classification (Method B). Table I summarizes the results.

The proposed method achieved a 4-class accuracy of 46.94%, representing a 45% relative improvement over the VLM direct baseline (32.27%). Confusion matrices (Tables II–III) show that the baseline exhibits severe label bias (only 3.9% APBT recall) while the proposed method produces more balanced predictions.

For binary regulatory classification, the proposed method achieved a precision of 93.08%, outperforming the baseline (90.35%). The proposed method also achieved a specificity of 75.17% versus 68.46%, demonstrating its reliability in preventing unnecessary regulatory interventions for non-regulated breed owners.

TABLE I. Classification Performance Summary

Metric	VLM Direct	Proposed
Dataset size	784 images	784 images
4-class Accuracy	32.27%	46.94%
Binary Precision	90.35%	93.08%
Binary Specificity	68.46%	75.17%
Binary Recall	69.29%	78.43%
Binary F1-Score	78.43%	85.13%

TABLE II. Confusion Matrices (VLM Direct Baseline)

	APBT	AmStaff	SBT	Bully
APBT	16	246	3	145
AmStaff	5	173	4	43
SBT	2	45	1	38
Bully	0	0	0	63

TABLE III. Confusion Matrices (Proposed Method)

	APBT	AmStaff	SBT	Bully
APBT	194	112	94	10
AmStaff	115	77	29	4
SBT	19	18	47	2
Bully	0	0	13	50

IV. Conclusion

By grounding VLM features in official standards, our system provides legally justifiable evidence for frontline regulatory enforcement. Future work includes multi-view evaluation, prototype sensitivity analysis, and automatic feature reliability calibration.

References

[1] Council of Agriculture, Executive Yuan, Taiwan, "Designated animals prohibited from breeding, importing, or exporting," Taiwan, Announcement No. 1110043386, 2022. [In Chinese].

[2] K. R. Olson, J. K. Levy, B. Norby, M. M. Crandall, J. E. Broadhurst, S. Jacks, R. C. Barton, and M. S. Zimmerman, "Inconsistent identification of pit bull-type dogs by shelter staff," The Veterinary Journal, vol. 206, no. 2, pp. 197–202, 2015.

[3] P. Borwarnginn, K. Thongkanchorn, S. Kanchanapreechakorn, and W. Kusakunniran, "Breakthrough conventional based approach for dog breed classification using CNN with transfer learning," in *Proc. 11th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Pattaya, Thailand, 2019, pp. 1–5.

[4] National Taiwan University, "Pit Bull Breed Appearance Identification Guide," Ministry of Agriculture, Executive Yuan, Taiwan. [Online]. Available: https://animal.moa.gov.tw/Frontend/Know/Detail/LT00000658?parentID=Tab0000082.

[5] Gemma Team, Google DeepMind, "Gemma 3 Technical Report," arXiv preprint arXiv:2503.19786, 2025.