SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 April 2026

Image-Based Cat *(Felis Catus)* Facial Expression Recognition Using YOLOv8n

Atanasius Manurip, Marwondo, Venia Restreva Danestiara
Informatika, Universitas Informatika dan Bisnis Indonesia
Email: atanasius.asm21@student.unibi.ac.id; marwondo@unibi.ac.id; veniarestreva@unibi.ac.id.

Accepted:	Accepted After Revision:	Published:
28 February 2026	10 March 2026	16 April 2026

Abstract

Cats (*Felis catus*) express emotional states through subtle facial cues that are often difficult for owners to interpret accurately. Automated recognition systems can provide objective analysis of these expressions using computer vision techniques. This study proposes an image-based cat facial expression recognition system using the YOLOv8n architecture. A dataset of 794 images was collected and expanded to 3,279 images through data augmentation. Four expression classes were defined based on CatFACS and related frameworks: *normal-netral*, *senang-afiliatif*, *stres-takut*, and *marah-agonistik*. The model was trained using a 70:20:10 split for training, validation, and testing. Experimental results show an overall mAP50 of 0.728, with the highest performance achieved in the *senang-afiliatif* class (0.773). However, the *marah-agonistik* class could not be reliably detected due to severe class imbalance in the dataset, indicating that the current model remains insufficient for recognizing anger expressions in cats. Precision and recall reached 0.939 and 0.94 respectively, indicating reliable detection under confident predictions. The trained model was successfully integrated into a Gradio-based dashboard for real-time expression recognition. These results demonstrate the feasibility of lightweight YOLOv8n for feline facial expression recognition while highlighting that accurate detection of *marah-agonistik* expressions requires a more diverse and balanced dataset in future research.

Keywords: Cat Expression, Computer Vision, Image-based Recognition, YOLO

1 INTRODUCTION

Recognizing facial expressions is a fundamental aspect of human communication, allowing individuals to naturally convey and interpret non-verbal emotional cues [14]. While interpreting human expressions is highly intuitive, recognizing the emotions of other species, such as cats, presents a significant challenge. Both humans and animals experience emotions, which are complex responses to internal and external stimuli often observable through facial expressions [8][14]. For cats (Felis catus), scientific tools like the Cat Facial Action Coding System (CatFACS) [15] and the Feline Grimace Scale (FGS) have been developed to link facial expressions to internal states, such as pain. However, human interpretation remains subjective and prone to error, often resulting in delayed or inaccurate detection of a cat's emotional or physical distress [1]. This highlights the urgent need for an objective, automated system to advance feline expression classification.

This research focuses on image-based cat facial expression recognition. While standard image classification approaches using Convolutional Neural Networks (CNNs) can assign a label to an entire image, they are inherently limited in their ability to localize and isolate the specific facial region responsible for an expression. This limitation is particularly critical for feline expressions, which are subtle and spatially concentrated in small facial regions such as the ears, eyes, and muzzle. Without explicit spatial localization, a classification model risks being influenced by irrelevant background features rather than the discriminative facial cues defined by frameworks such as CatFACS. Furthermore, Standard CNN-based image classification models predict a label for the entire image without explicitly identifying the spatial location of the relevant facial features [13]. Object detection architectures, by contrast, generate bounding boxes that constrain the model's attention to the relevant facial area, enabling more precise feature extraction from the subtle action units that differentiate, for example, a *stres-takut* expression from an *marah-agonistik* one. Therefore, this study utilizes the You Only Look Once (YOLO) architecture, a CNN derivative that combines localization and classification in a single forward pass, enabling superior performance and the ability to recognize multiple objects simultaneously [10][13][17].

Specifically, the YOLOv8n architecture was selected due to its optimal balance of high detection accuracy and low computational resource requirements, having demonstrated strong efficacy across various object detection tasks [22][23][24][16]. This research, titled "IMAGE-BASED CAT (FELIS CATUS) EXPRESSION RECOGNITION USING YOLOv8n," aims to train a model to accurately classify feline expressions. By utilizing this trained model, cat owners will be able to distinguish their pets' expressions from facial images, thereby improving the monitoring of their emotional and health conditions.

Based on the background outlined above, this research addresses several key problems regarding the automated detection of feline emotions. Primarily, it investigates how to construct a YOLOv8n model for the image recognition of cat facial expressions and how to effectively evaluate its performance. Furthermore, it explores how to build a user dashboard that allows individuals to easily utilize the trained model for recognizing their cats' expressions via uploaded images.

In alignment with the formulated problems, the main objective of this research is to develop a YOLOv8n-based model capable of accurately recognizing the facial expressions of cats (Felis catus). Additionally, the study aims to evaluate and comprehensively map the metrics detailing the strengths and weaknesses of the model's performance, and to create an integrated dashboard that enables users to identify their cats' facial expressions through uploaded images.

2 LITERATURE REVIEW

2.1 Face Recognition System

Face recognition systems are a critical component of digital image processing and pattern recognition, designed to identify individuals or expressions from visual features. The process encompasses image acquisition, pre-processing, feature extraction, and classification. Common applications include FaceID, FaceUnlock, and Face Authenticator. Generally, face recognition systems identify or verify individuals by matching facial features extracted from input images against stored database entries. Modern systems employ deep learning algorithms such as Faster-RCNN, YOLO, and CNN, all of which require a training phase where features are extracted, layered through neural networks, labeled, and stored in a database for subsequent matching. Beyond human identification, this technology has been applied to animal facial expression analysis, including the classification of cat and dog expressions, enabling detection of subtle facial changes linked to emotional states.

2.2 Artificial Intelligence

Artificial Intelligence (AI) is a branch of computer science focused on creating systems capable of mimicking human thought and decision-making, enabling large-scale automated data processing and pattern learning. In the context of cat facial expression recognition, AI is essential for accurately assessing an animal's emotional state from image analysis [13]. AI can be categorized into Narrow AI designed for specific tasks such as facial recognition and General AI, a hypothetical system with human-level capabilities across domains [18]. This research operates within the scope of Narrow AI. The field encompasses Neural Networks, Machine Learning, Computer Vision, and Natural Language Processing in a hierarchical relationship. Modern AI has shifted from rule-based approaches to data-driven paradigms enabled by big data availability and increased computational power through GPUs.

2.3 Computer Vision

Computer Vision (CV) is a discipline concerned with enabling machines to interpret and understand visual data from images or video. It forms the foundation of facial detection and recognition, applied here for the automated detection of visual patterns on cat faces [19]. As a specific branch of AI, CV replicates the human visual system to automate tasks ranging from simple object recognition to complex scene understanding. Its general workflow involves image acquisition, processing/feature extraction, and decision-making. Core CV tasks include: (a) Image Classification assigning a label to an entire image; (b) Object Detection classifying objects and locating them with bounding boxes; (c) Image Segmentation classifying each pixel into object classes; and (d) Object Tracking identifying and following objects across video frames. This research specifically falls within the object detection domain. Deep learning-based CV models require thousands to millions of labeled examples, highlighting the justification for developing a custom dataset as a key contribution of this study.

2.4 Machine Learning

Machine Learning (ML) is a sub-field of AI that develops algorithms enabling systems to improve performance through experience without explicit programming [18]. ML paradigms are classified into three categories: (a) Supervised Learning training on labeled datasets to learn input-to-output mappings; (b) Unsupervised Learning discovering hidden patterns or structures in unlabeled data; and (c) Reinforcement Learning an agent learns decision-making by interacting with an environment through reward/penalty feedback.

This research employs supervised learning, appropriate for classifying facial expressions into scientifically defined categories using annotated datasets.

2.5 Deep Learning

Deep Learning (DL) is a branch of ML that mimics the human brain's data processing using artificial neural networks particularly Deep Neural Networks (DNN) with multiple hidden layers. Each layer progressively processes and abstracts information, automating the feature engineering process that traditional ML requires manually. This makes DL highly effective for unstructured complex data like images [18][13]. DL is conceptually a subset within ML, which is itself a subset within AI. Key architectures include Convolutional Neural Networks (CNN) for image processing and NLP, Recurrent Neural Networks (RNN) for time series and emotion tasks, Reinforcement Learning for robotics, and Generative Adversarial Networks (GAN) for image and video generation. DL's popularity stems from its superior performance in classification tasks when large training datasets are available, supported by frameworks like TensorFlow and PyTorch.

2.6 Convolutional Neural Networks (CNN)

CNNs are a specialized deep learning architecture designed for grid-structured data such as images, comprising convolutional layers, pooling layers, and fully connected layers. They excel at extracting spatial features and local patterns [11][13]. Introduced broadly by Yann LeCun through LeNet-5 in 1998, CNNs evolved into architectures like AlexNet, VGGNet, and ResNet. Core layers include:

(a) Convolutional Layer applies filters to produce feature maps via
$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n)$ (2.1)
(b) Activation Layer applies non-linear functions like ReLU;
(c) Pooling Layer reduces spatial dimensions via Max Pooling; and
(d) Fully Connected Layer flattens high-level feature maps for final classification.

2.7 You Only Look Once (YOLO)

YOLO is a family of object detection algorithms that reframes detection as a single regression problem, performing classification and localization simultaneously in one network pass. YOLOv8, introduces three main components: (a) Backbone extracts visual features using a CSP-inspired design with the C2f module for richer gradient flow; (b) Neck aggregates multi-scale feature maps using PANet and SPPF for efficient receptive field expansion; and (c) Head generates final predictions using an Anchor-Free approach and Decoupled Head separating classification and bounding box regression. The total loss function is

$$L_{(YOLOv8)} = \lambda_{(cls)} L_{(cls)} + \lambda_{(box)} L_{(box)} + \lambda_{(dfl)} L_{(dfl)} \quad (2.2)$$

combining Binary Cross-Entropy for classification, CIoU for bounding box accuracy, and Distribution Focal Loss for flexible coordinate modeling. YOLOv8 offers size variants (n, s, m, l, x); the Nano (n) variant is optimized for real-time and resource-constrained deployment.

2.8 Cat Facial Expressions

Cat facial expressions are important indicators of emotional and psychological states. Features such as ear position, eye shape, and mouth expression can reflect fear, comfort, or discomfort. Cats (Felis catus) communicate through vocalizations, body language, and specific gestures, though their subtle communication style is often difficult for owners to interpret [15][4]. Facial expressions rank third in frequency among cat communication methods, appearing primarily when vocalizations and body language are insufficient. Notably, cat emotions show heterospecific similarities with human emotions purring parallels human happiness, while hissing parallels anger [19][14].

2.9 Cat Facial Action Coding System (CatFACS)

CatFACS is an objective coding system adapted from the human Facial Action Coding System (FACS) that decomposes cat facial expressions into individual muscle movements called Action Units (AUs) and Action Descriptors (ADs), without subjective emotional interpretation [19][5]. Applications include pain assessment forming the basis for the Feline Grimace Scale (FGS) social interaction analysis (distinguishing affiliative from non-affiliative interactions), and facial mimicry studies [6]. Manual CatFACS coding is time-consuming and prone to bias, leading to computational approaches: (a) Landmark-based methods using anatomical keypoints (e.g., 48-landmark schemes) analyzed with ML models; and (b) Deep learning approaches using CNNs for direct classification without manual AU coding, though these are less interpretable as black-box models.

2.10 Cat Facial Landmarks in the Wild (CatFLW)

Facial landmarks are key points on the face whose positions provide critical information for automated analysis tasks [5]. In computational contexts, these landmarks form the basis of geometric morphometric approaches, where specific points placed

on images represent facial shape and enable quantification of facial changes associated with internal conditions such as pain. Automatic facial landmark detection supports applications including pain classification [9] and emotion recognition [4]. This landmark-based approach also enhances model explainability, as geometric changes can be directly linked to CatFACS Action Units unlike black-box deep learning models that are difficult to interpret.

3 RESEARCH METHODS

The research methodology used consists of five main stages, as shown in the figure 1.

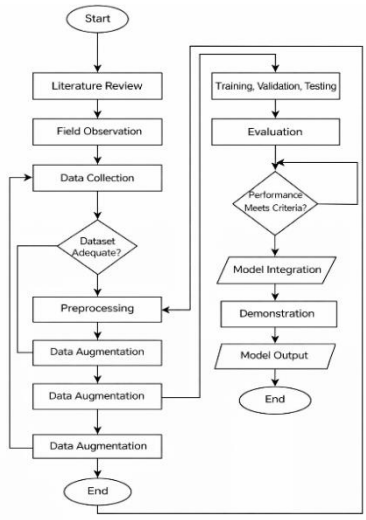

Figure 1. Research Flow

As this study follows eight sequential stages. A Literature Review establishes the theoretical foundations, organized into a Research Comparison Table to define the research focus. Data Gathering collects images from online sources and field acquisition, followed by Pre-Processing where images are cropped and labeled for standardization. The dataset is then split into Training (70%), Validation (20%), and Testing (10%) subsets, with Data Augmentation horizontal flipping, rotation, brightness adjustment, and blurring applied exclusively to training data. Model Training runs iteratively to maximize accuracy and minimize validation loss, after which the model is Evaluated via Confusion Matrix to derive Precision, Recall, F1-Score, and Accuracy. If metrics meet the required threshold, the model is saved and integrated into a user-facing Dashboard.

4 RESULTS AND DISCUSSION

4.1 Data Collection

Data collection was carried out by obtaining datasets from the Roboflow website and field observations. Images from Roboflow were gathered and merged into a new dataset comprising 794 cat images used for model training. Field-collected data was reserved for model demonstration.

4.2 Data Annotation

Data labeling was performed using the annotation feature available on the Roboflow platform. A portion of the dataset was obtained from Roboflow's public dataset repository, where the images already contained validated labels. These labels were retained and converted from image classification format (CNN) to object detection format compatible with YOLO.

For the images collected independently through field observation, annotation was conducted manually by the researchers. Bounding boxes were drawn around each cat face and each image was assigned to one of four predefined expression classes: *normal-netral*, *senang-afiliatif*, *stres-takut*, and *marah-agonistik*.

The labeling process followed the facial action indicators described in the Cat Facial Action Coding System (CatFACS) and related frameworks such as the Feline Grimace Scale (FGS) and CatFLW. These frameworks define feline facial expressions through combinations of observable facial features, including ear position, eye aperture, muzzle tension, and whisker orientation.

During annotation, the researchers examined these facial cues and assigned labels according to the dominant facial configuration observed in each image. This guideline-based annotation process was used to maintain consistency with established scientific interpretations of feline facial expressions.

4.3 Dataset Splitting

The labeled dataset was split with a 70:20:10 ratio for training, validation, and test sets, respectively.

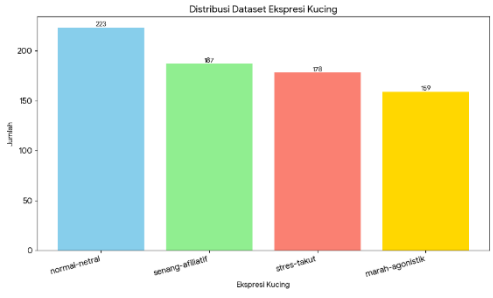

Figure 2. Dataset Split Distribution

4.4 Preprocessing and Data Augmentation

Auto-orientation was applied to prevent the model from failing on images with improper orientation. Data augmentation was then applied exclusively to the training data to increase variation, including: random rotation, horizontal flip, brightness adjustment, blur effect, cutout, and saturation adjustment. This was carried out using Google Collaboratory.

Table 1. Dataset Size Before and After Augmentation

Type	Before	After
Training	556	3,041
Validation	158	158
Test	80	80
Total	794	3,279

The training set grew from 556 to 3,041 images an increase of 312.97% bringing the total dataset to 3,279 images.

4.5 Model Training

The augmented training data was used to train a YOLOv8n model for 300 epochs with an early-stopping patience of 25. Early stopping halted training at epoch 54 due to signs of overfitting, indicated by slight spikes in the validation loss curves.

Figure 3. Model Training Graph

4.6 Testing and Evaluation

After training, the model was evaluated on the validation and test sets. The overall mAP50 was 0.728 (72.8%), with the best per-class result for '*senang-afiliatif*' at 0.773 (77.3%). Table 3 presents the full results.

Table 2. Model Test Results

Class	Imgs	Inst	Box	R	mAP50	mAP 50-95	F1-Score
all	158	148	0.657	0.68	0.728	0.486	0.78
kucing	32	32	0.719	0.75	0.785	0.542	0.68
normal netral	48	49	0.593	0.62	0.655	0.453	0.70
senang afiliatif	38	38	0.699	0.67	0.773	0.505	0.68
stres takut	29	29	0.617	0.70	0.699	0.443	0.67
marah	11	11	0.00	0.00	0.00	0.00	0.00

agnostik

The *marah-agonistik* class obtained a precision, recall, and F1-score of **0.00**, indicating that the model failed to detect any instance of this class in the evaluation dataset. This complete failure suggests that the model did not learn discriminative features for the *marah-agonistik* expression during training.

Figure 4. Testing Process

The failure to detect the *marah-agonistik* class is primarily attributed to severe class imbalance in the dataset. The class contains only **159 images**, significantly fewer than the other classes (*normal-netral* = 223, *senang-afiliatif* = 187, *stres-takut* = 178). In deep learning models, such imbalance causes the model to bias its learning toward the majority classes, reducing its ability to generalize features for underrepresented classes. As a result, the trained model did not produce any confident predictions for the *marah-agonistik* expression, leading to zero precision, recall, and F1-score.

The confusion matrix (Figure 5) confirms that the model performed well on '*normal-netral*' (29 correct), '*senang-afiliatif*' (27 correct), and '*stres-takut*' (20 correct), but failed to recognize '*marah-agonistik*'.

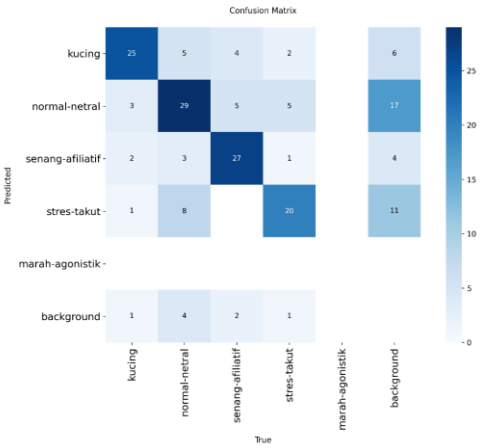

Figure 5. Confusion Matrix

The precision-confidence curve reached 0.939 (93.9%) for all classes combined, indicating the model reliably detects expressions when confident. The recall-confidence curve shows 94% recall, meaning the model can remember 94% of the learned patterns.

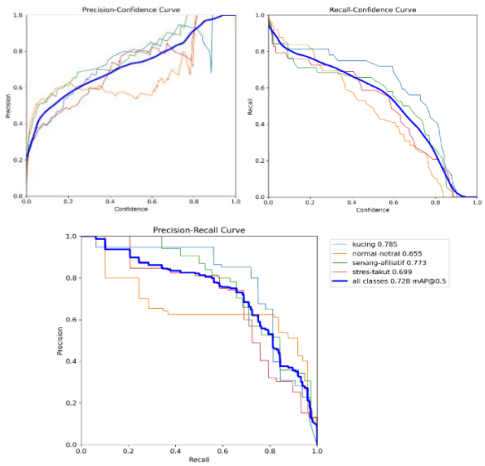

Figure 6. Precision-Confidence , Recall-Confidence, and Precision-Recall Curve

The precision-recall curve shows the combined detection capability at 72.8% (0.728), confirming that the model is suitable for integration into a dashboard interface.

4.7 Dashboard Integration

A Gradio-based dashboard was developed using Google Colab to allow users to interact directly with the trained model. Users can upload a cat image and submit it for expression prediction.

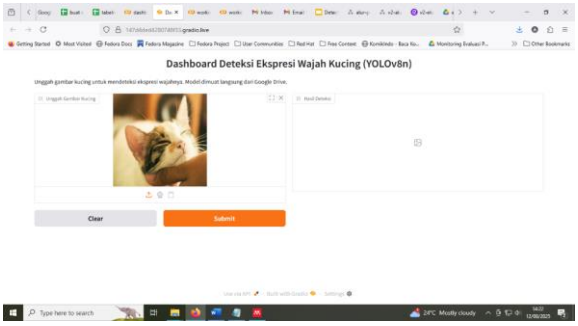

Figure 7. Dashboard Interface #1

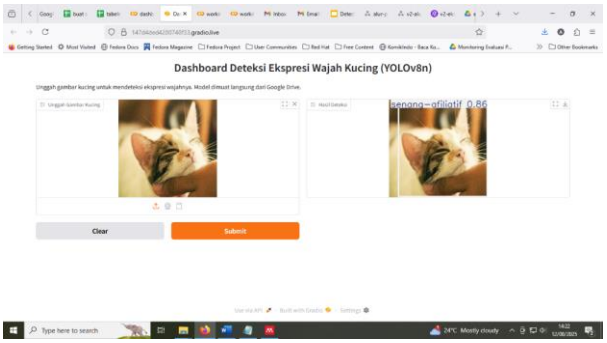

Figure 8. Dashboard Interface #2

4.8 GradCAM Analysis

GradCAM (Gradient-weighted Class Activation Mapping) was applied to the trained model to visualize which regions of the input image the model focuses on. The model consists of 22 layers; layer 0 (input) and layer 21 (pre-output) were selected for comparison.

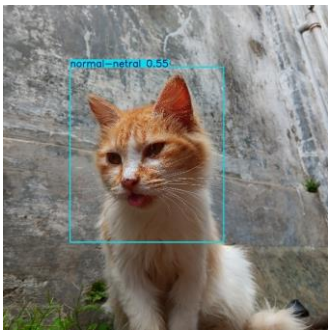

Figure 9. Model Detection Output

Figure 10. GradCAM Layer 0

4.9 Research Analysis

The model trained on 3,279 images across four expression classes successfully recognized three classes. The '*marah-agonistik*' class failed entirely due to class imbalance it contained only 159 images, compared to 223 (*normal-netral*), 187 (*senang-afiliatif*), and 178 (*stres-takut*), as shown in Figure 10. Augmentation widened this imbalance further.

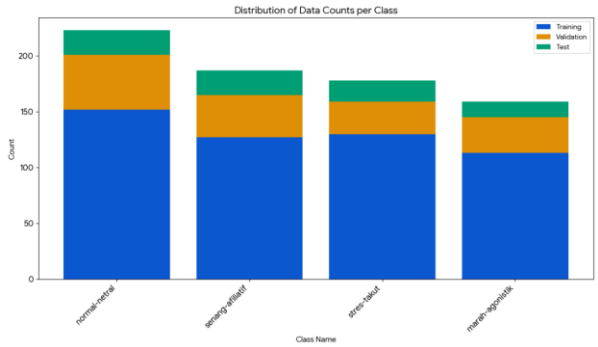

Figure 11. Class Distribution Comparison

Despite this limitation, the remaining three classes achieved mAP50 values above 50%: *senang-afiliatif* at 77.3%, *stres-takut* at 69.9%, and *normal-netral* at 65.5%. These results demonstrate that the model is viable for integration in real-world applications, provided additional data for the '*marah-agonistik*' class is collected to address the imbalance.

5 CONCLUSION

This study presents the development of an image-based cat facial expression recognition system using the YOLOv8n object detection architecture. The experimental results demonstrate that the proposed model is capable of recognizing several feline facial expression categories with satisfactory performance. Specifically, the model achieved reliable detection results for the *normal-netral*, *senang-afiliatif*, and *stres-takut* classes, with mAP50 values above 0.65 and an overall mAP50 of 0.728. These results indicate that the YOLOv8n architecture can effectively extract relevant facial features associated with subtle feline expressions and perform real-time detection when integrated into an application interface.

This research makes several contributions to the field of animal facial expression recognition. First, it demonstrates the feasibility of applying a lightweight object detection architecture (YOLOv8n) for recognizing subtle feline facial expressions from images. Second, the study constructs and organizes a dataset of cat facial images categorized into four expression classes (*normal-netral*, *senang-afiliatif*, *stres-takut*, and *marah-agonistik*), providing a structured dataset suitable for supervised learning in feline expression analysis. Third, the research implements a practical deployment scenario by integrating the trained model into a Gradio-based dashboard, enabling users to interact with the model and perform real-time expression recognition.

However, the evaluation results reveal that the model completely failed to detect the *marah-agonistik* class (F1-score = 0.00). This outcome indicates that the model was unable to learn discriminative features for this expression during training. Further analysis shows that this limitation is primarily caused by severe class imbalance in the dataset, where the number of *marah-agonistik* samples is significantly lower than the other expression classes. As a consequence, the model becomes biased toward the majority classes and fails to generalize to underrepresented categories.

Despite this limitation, the trained model was successfully integrated into a Gradio-based dashboard that allows users to upload cat images and obtain expression predictions in real time. This demonstrates the practical feasibility of implementing lightweight object detection models for assisting cat owners in monitoring their pets' emotional conditions through facial expression analysis.

Future research should focus on improving dataset balance and diversity, particularly by collecting additional samples representing the "*marah-agonistik*" expression. Expanding the dataset, improving annotation consistency, and exploring advanced techniques such as transfer learning, class re-weighting, or alternative deep learning architectures may further improve the robustness and generalization capability of the model for feline facial expression recognition.

REFFERENCE

[1]. J. Cheng, A. Malo, M. Garbin, B. P. Monteiro, and P. V. Steagall, "Construct validity, responsiveness and reliability of the Feline Grimace Scale in kittens," *Journal of Feline Medicine and Surgery*, vol. 25, no. 12, 2023. https://doi.org/10.1177/1098612X231211765

[2]. Vijayakumar and S. Vairavasundaram, "YOLO-based Object Detection Models: A Review and its Applications," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83535–83574, 2024. https://doi.org/10.1007/s11042-024-18872-y

[3]. Peng, M. Sun, K. Zou, B. Zhang, G. Dai, and A. C. Tsoi, "Facial Expression Recognition-You Only Look Once-Neighborhood Coordinate Attention Mamba: Facial Expression Detection and Classification Based on Neighbor and Coordinates Attention Mechanism," *Sensors*, vol. 24, no. 21, 2024. https://doi.org/10.3390/s24216912

[4]. G. Martvel, N. Farhat, I. Shimshoni, and A. Zamansky, "CatFLW: Cat Facial Landmarks in the Wild Dataset," arXiv:2305.04232, 2023. http://arxiv.org/abs/2305.04232

[5]. G. Martvel, L. Scott, B. Florkiewicz, A. Zamansky, I. Shimshoni, and T. Lazebnik, "Computational investigation of the social function of domestic cat facial signals," *Scientific Reports*, vol. 14, no. 1, 2024. https://doi.org/10.1038/s41598-024-79216-2

[6]. G. Martvel, I. Shimshoni, and A. Zamansky, "Automated Detection of Cat Facial Landmarks," *International Journal of Computer Vision*, vol.

132, no. 8, pp. 3103–3118, 2024. https://doi.org/10.1007/s11263-024-02006-w
[7]. L. Scott and B. N. Florkiewicz, "Feline faces: Unraveling the social function of domestic cat facial signals," *Behavioural Processes*, vol. 213, p. 104959, 2023. https://doi.org/10.1016/J.BEPROC.2023.104959
[8]. M. E. Kret, J. J. M. Massen, and F. B. M. de Waal, "My Fear Is Not, and Never Will Be, Your Fear: On Emotions and Feelings in Animals," *Affective Science*, vol. 3, no. 1, pp. 182–189, 2022. https://doi.org/10.1007/s42761-021-00099-x
[9]. M. Feighelstein, I. Shimshoni, L. R. Finka, S. P. L. Luna, D. S. Mills, and A. Zamansky, "Automated recognition of pain in cats," *Scientific Reports*, vol. 12, no. 1, 2022. https://doi.org/10.1038/s41598-022-13348-1
[10]. M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, 2023. https://doi.org/10.3390/machines11070677
[11]. M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," *Computers*, vol. 13, no. 12, 2024. https://doi.org/10.3390/computers13120336
[12]. M. M. Ambali Parambil, L. Ali, M. Swavaf, S. Bouktif, M. Gochoo, H. Aljassmi, and F. Alnajjar, "Navigating the YOLO Landscape: A Comparative Study of Object Detection Models for Emotion Recognition," *IEEE Access*, vol. 12, pp. 109427–109442, 2024. https://doi.org/10.1109/ACCESS.2024.3439346
[13]. M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation*, vol. 11, no. 3, 2023. https://doi.org/10.3390/computation11030052
[14]. M. Mendl, V. Neville, and E. S. Paul, "Bridging the Gap: Human Emotions and Animal Emotions," *Affective Science*, vol. 3, no. 4, pp. 703–712, 2022. https://doi.org/10.1007/s42761-022-00125-6
[15]. P. Kovačič, "Cat Facial Action Coding System (CatFACS) and Scientific Illustration," pp. 159–165, 2024. https://doi.org/10.55295/psl.2024.ii17
[16]. Q. Zhang, K. Ahmed, M. I. Khan, H. Wang, and Y. Qu, "YOLO-FCE: A feature and clustering enhanced object detection model for species classification," *Pattern Recognition*, vol. 171, 2025. https://doi.org/10.1016/j.patcog.2025.112218
[17]. R. Zhang and R. Ma, "Facial expression recognition method based on PSA—YOLO network," *Frontiers in Neurorobotics*, vol. 16, 2023. https://doi.org/10.3389/fnbot.2022.1057983
[18]. R. I. Mukhamediev, Y. Popova, Y. Kuchin, E. Zaitseva, A. Kalimoldayev, A. Symagulov, V. Levashenko, F. Abdoldina, V. Gopejenko, K. Yakunin, E. Muhamedijeva, and M. Yelis, "Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges," *Mathematics*, vol. 10, no. 15, 2022. https://doi.org/10.3390/math10152552
[19]. S. Broomé, M. Feighelstein, A. Zamansky, G. Carreira Lencioni, P. Haubro Andersen, F. Pessanha, M. Mahmoud, H. Kjellström, and A. A. Salah, "Going Deeper than Tracking: A Survey of Computer-Vision Based Recognition of Animal Pain and Emotions," *International Journal of Computer Vision*, vol. 131, no. 2, pp. 572–590, 2023. https://doi.org/10.1007/s11263-022-01716-3
[20]. S. Ren, M. Sun, B. Wang, M. Liu, and S. Men, "High Precision Infant Facial Expression Recognition by Improved YOLOv8," *IEEE Access*, vol. 13, pp. 39621–39630, 2025. https://doi.org/10.1109/ACCESS.2025.3543950
[21]. T. Rakhimzhanova, A. Kuzdeuov, and H. A. Varol, "AnyFace++: Deep Multi-Task, Multi-Domain Learning for Efficient Face AI," *Sensors*, vol. 24, no. 18, 2024. https://doi.org/10.3390/s24185993
[22]. X. Cao, Y. Su, X. Geng, and Y. Wang, "YOLO-SF: YOLO for Fire Segmentation Detection," *IEEE Access*, vol. 11, pp. 111079–111092, 2023. https://doi.org/10.1109/ACCESS.2023.3322143
[23]. X. Lang, Z. Ren, D. Wan, Y. Zhang, and S. Shu, "MR-YOLO: An Improved YOLOv5 Network for Detecting Magnetic Ring Surface Defects," *Sensors*, vol. 22, no. 24, 2022. https://doi.org/10.3390/s22249897
[24]. X. Ren, Y. Bai, G. Liu, and P. Zhang, "YOLO-Lite: An Efficient Lightweight Network for SAR Ship Detection," *Remote Sensing*, vol. 15, no. 15, 2023. https://doi.org/10.3390/rs15153771
[25]. Y. Bu, H. Ye, Z. Tie, Y. Chen, and D. Zhang, "OD-YOLO: Robust Small Object Detection Model in Remote Sensing Image with a Novel Multi-Scale Feature Fusion," *Sensors*, vol. 24, no. 11, 2024. https://doi.org/10.3390/s24113596
[26]. Y. Guo, S. Chen, R. Zhan, W. Wang, and J. Zhang, "LMSD-YOLO: A Lightweight YOLO Algorithm for Multi-Scale SAR Ship Detection," *Remote Sensing*, vol. 14, no. 19, 2022. https://doi.org/10.3390/rs14194801
[27]. Y. Qiu, F. Sha, and L. Niu, "DKA-YOLO: Enhanced Small Object Detection via Dilation Kernel Aggregation Convolution Modules," *IEEE Access*, vol. 12, pp. 187353–187366, 2024. https://doi.org/10.1109/ACCESS.2024.3515201
[28]. Y. S. Dosso, D. Kyrollos, K. J. Greenwood, J. Harrold, and J. R. Green, "NICUface: Robust

Neonatal Face Detection in Complex NICU Scenes," *IEEE Access*, vol. 10, pp. 62893–62909, 2022. https://doi.org/10.1109/ACCESS.2022.3181167

[29]. Y. Shao, Z. Xu, and Q. Zhu, "SH-YOLO: Enhanced Real-Time Detection of Laparoscopic Surgical Instruments in Computer-Aided Surgery Based on Star Operation and Hybrid Attention Mechanisms," *IEEE Access*, vol. 13, pp. 135179–135195, 2025. https://doi.org/10.1109/ACCESS.2025.3593635

[30]. Y. Song, Z. Xie, X. Wang, and Y. Zou, "MS-YOLO: Object Detection Based on YOLOv5 Optimized Fusion Millimeter-Wave Radar and Machine Vision," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15435–15447, 2022. https://doi.org/10.1109/JSEN.2022.3167251

[31]. Y. Su, Q. Liu, W. Xie, and P. Hu, "YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms," *Computer Methods and Programs in Biomedicine*, vol. 221, 2022. https://doi.org/10.1016/j.cmpb.2022.106903

[32]. Y. Wang, Y. Li, D. M. S. Kayes, H. S. Abdullahi, S. Gao, H. Zhang, Z. Song, and P. Lv, "Research on a Lightweight PCB Detection Algorithm Based on AE-YOLO," *IEEE Access*, vol. 12, pp. 109367–109379, 2024. https://doi.org/10.1109/ACCESS.2024.3439523

[33]. Y. Yang, Z. Liu, J. Chen, H. Gao, and T. Wang, "Railway Foreign Object Intrusion Detection Using UAV Images and YOLO-UAT," *IEEE Access*, vol. 13, pp. 18498–18509, 2025. https://doi.org/10.1109/ACCESS.2025.3533304