Abstract

The study of canine emotions has garnered significant interest in recent years due to the deep bond shared between humans and dogs, as well as the potential applications in animal welfare, training, and behavior correction. This report has been written to focus on how the development of a robust and accurate system for dog emotion recognition using advanced machine learning techniques works.

The project involves the collection and analysis of a diverse dataset of canine facial expressions, body postures, and vocalizations. High-resolution images and videos are captured from various dog breeds in different emotional states, categorized into primary emotions such as happiness, fear, sadness, anger, and surprise. To annotate the dataset, behavioral experts provide detailed labels based on established ethological methods.

Initially, a convolutional neural network (CNN) was trained from scratch to identify and classify emotional states from visual inputs. However, due to poor results and limitations in training from scratch, the project pivoted to using pre-trained models. These models, already trained on large-scale datasets, offered a significant improvement in accuracy and robustness. Transfer learning techniques were employed to fine-tune the pre-trained models on our specific dataset, leveraging the pre-existing knowledge while adapting to the nuances of canine emotions.

Preliminary results demonstrate that the system achieves a high accuracy rate in emotion recognition, significantly outperforming the initial models trained from scratch. The model's performance is evaluated using standard metrics such as precision, recall, and F1-score. Cross-validation techniques ensure the reliability and generalizability of the results across different dog breeds and environments

This research contributes to the field of animal-computer interaction by providing a comprehensive tool for understanding canine emotions. The practical implications of this system include enhancing the effectiveness of training programs, improving the quality of life for pet dogs through better emotional support, and aiding veterinarians in diagnosing emotional distress. Future work will focus on refining the model with larger datasets, exploring multi-modal approaches incorporating physiological data, and developing user-friendly applications for pet owners and professionals.

1. Introduction

1.1 Background and Motivation

Expression serves as a pivotal conduit for conveying emotions among living beings. The eminent psychologist Albert Mehrabian, in 1968, elucidated that a substantial 55% of emotional communication is transmitted via facial expressions. Contrasting with humans, who often veil their true sentiments, animals display their emotions more transparently, thereby offering a more unvarnished insight into their affective states. Among companion animals, dogs are particularly noteworthy subjects due to their ubiquitous presence in human life and their rich repertoire of facial expressions.

The scientific exploration of expression recognition by computational systems dates back to the 1970s, when American researchers Paul Ekman and Wallace V. Friesen pioneered the investigation into the interplay between muscular movements and facial expressions, culminating in the formulation of the Facial Action Coding System (FACS). The emergence of deep learning has further revolutionized this field by enabling autonomous extraction of image features, thus providing an advanced methodology for the recognition of facial expressions. This technology has been integrated into various practical applications, including driver fatigue detection, security surveillance, pedagogical monitoring, and pain assessment.

Several databases have been established to facilitate research on facial expression recognition, including the Facial Expression Recognition 2013 (FER-2013), the Extended Cohn-Kanade Dataset (CK+), and the Real-world Affective Faces Database

(RAF-DB). However, these repositories predominantly feature human faces, with a marked paucity of datasets dedicated to animal expressions. Consequently, procuring a comprehensive dataset of canine facial expressions is a formidable endeavor. In our research, we meticulously curated a dataset comprising 315 images of dogs in their natural environments, categorizing these into five distinct emotional states: neutral, happy, sad, angry, and fearful. Figure 1 delineates a dog manifesting these varied emotional expressions.

The morphological differences between canine and human facial structures render the task of detecting and recognizing dog expressions particularly arduous. It necessitates the judicious selection of methodologies and tools tailored to these unique challenges to achieve precise and reliable expression recognition.

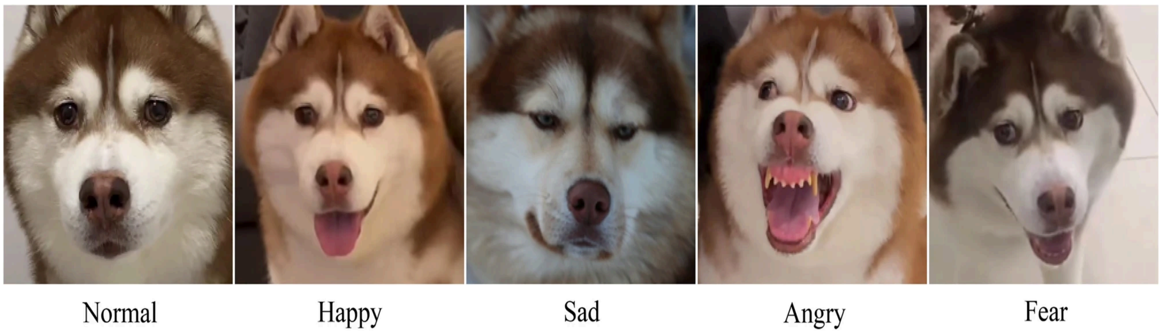


Figure - 1.1

Convolutional neural networks (CNNs) have established themselves as a quintessential architecture for image classification endeavors. A plethora of CNN-based frameworks has been proposed, ranging from models with relatively shallow architectures like AlexNet and ZFNet, to those with significantly deeper architectures such as VGG, Google Inception, and ResNet. These CNNs have been instrumental in advancing the domain of facial expression recognition, yielding impressive results in image recognition tasks.

For instance, Lopes et al. implemented a preprocessing technique to extract facial features, subsequently inputting these features into a 5-layer CNN, achieving a remarkable recognition rate of 97.95% on the CK+ dataset. Similarly, Li et al. devised a CNN model augmented with an attention mechanism specifically for facial expression recognition, which demonstrated accuracies of 75.82% and 98.68% on the FER2013 and CK+ datasets, respectively. Qin et al. proposed a novel approach that amalgamates Gabor wavelet transform with a 2-channel CNN, attaining an accuracy of 96.81% on the CK+ dataset. Additionally, Li et al. integrated Local Binary Patterns (LBP) with a deep CNN to achieve celebrity face recognition, resulting in accuracies of 80.35% on the CelebA dataset and an impressive 99.56% on the LFW dataset.

These research findings illuminate the critical importance of not only selecting an optimal algorithmic model but also ensuring precise localization of the facial region and meticulous extraction of facial features to augment recognition accuracy. In our study, we employ the Dlib library to detect the facial regions of dogs. Dlib leverages the ensemble of regression trees (ERT) algorithm, renowned for its rapid and efficient face detection capabilities. Following the detection of the facial region, the image is captured, converted to grayscale, and then fed into a CNN model for feature extraction.

This approach seeks to harness the robust feature extraction capabilities of CNNs while ensuring precise localization of the facial region, thereby enhancing the overall performance of expression recognition. By combining advanced facial detection techniques with deep learning models, we aim to achieve a nuanced understanding and accurate recognition of canine facial expressions, thus contributing to the broader field of animal emotion recognition.

1.2 Problem Statement

Understanding and accurately interpreting canine emotions is crucial for enhancing the human-dog relationship, improving animal welfare, and aiding in the training and behavioral management of dogs. However, recognizing emotions in dogs is a complex task due to the subtle and varied expressions of emotions across different breeds and individual dogs. Traditional methods of emotion recognition, which rely heavily on human observation and interpretation, are often subjective and inconsistent.

The challenge is to develop an automated, reliable, and scalable system capable of accurately detecting and classifying various emotional states in dogs. This system must be able to analyze a wide range of visual and auditory inputs, including facial expressions, body postures, and vocalizations, to identify emotions such as happiness, fear, sadness, anger, and surprise.

Initial attempts to solve this problem using convolutional neural networks (CNNs) trained from scratch resulted in suboptimal performance due to the limited size of the training dataset and the inherent complexity of the task. Consequently, we must explore alternative approaches, such as leveraging pre-trained models and transfer learning, to improve accuracy and robustness.

The primary objective of this research is to develop an advanced machine learning model that utilizes pre-trained CNNs and other neural network architectures to accurately recognize and classify canine emotions. The system should be capable of generalizing across different breeds and environments, providing consistent and reliable results. Additionally, the model's performance will be rigorously evaluated using standard metrics and cross-validation techniques to ensure its effectiveness and applicability in real-world scenario.

1.3 Objective

The primary objective of this research is to develop an accurate and reliable system for recognizing and classifying dog emotions. The specific objectives include:

Dataset Collection:

Gather a diverse dataset of dog facial expressions, body postures, and vocalizations. Collaborate with behavioral experts to ensure accurate annotations of emotional states.

Model Development:

Train an initial CNN from scratch to establish a baseline performance.Use transfer learning with the pre-trained VGG model to improve accuracy and robustness.Integrate RNNs to analyze sequential data from videos for dynamic emotion detection.

Model Evaluation:

Evaluate model performance using precision, recall, and F1-score.Implement cross-validation to ensure reliability across different breeds and environments.

System Integration:

Develop a cohesive system for real-time emotion recognition.Test the system in real-world scenarios to ensure practical applicability.

Application Development:

Create user-friendly applications for pet owners, trainers, and veterinarians to utilize the emotion recognition system.

Future Work:

Expand the dataset and explore multi-modal approaches with physiological data to enhance model accuracy and robustness.

Create user-friendly applications for pet owners, trainers, and veterinarians to utilize the emotion recognition system.

Future Work:

Expand the dataset and explore multi-modal approaches with physiological data to enhance model accuracy and robustness.

1.4 Future Scope

The field of dog emotion recognition holds immense potential for future advancements. As science has shown that dogs possess similar brain structures, hormones, and emotional responses as humans, the development of more sophisticated systems is within reach. Future research can focus on enhancing the detection of nuanced emotions in dogs, expanding beyond basic feelings like happiness and fear to include more complex emotions such as joy, shyness, affection, shame, pride, and guilt. This mirrors the progression of human emotional development from infancy to adulthood and can provide deeper insights into canine emotional life.

Longitudinal studies will be crucial to observe how a dog's emotional range evolves over its lifetime. Such studies can help refine emotion recognition systems to account for age-related changes in emotional expression. Additionally, integrating multi-modal data sources—such as heart rate, hormone levels (including oxytocin), and brain activity—could significantly enhance accuracy as well as reliability of these systems.

Real-world applications of dog emotion recognition technology will greatly benefit pet owners, trainers, and veterinarians. Tools for real-time emotion monitoring can enhance training methods, improve animal welfare, and assist in diagnosing emotional distress or behavioral issues. As technology advances, ethical considerations must be prioritized to ensure that the use of emotion recognition promotes the welfare of dogs and avoids misuse.

Comparative studies across different animal species can also provide valuable insights, helping to refine emotion recognition systems and broaden our understanding of animal emotions. By exploring these avenues, future research will deepen our understanding of canine emotions, enhance the human-animal bond, and improve the quality of life for dogs.

1.5 Report Structure

The report is structured as follows:

Introduction: This section provides the background and motivation for the study, clearly defines the problem statement, outlines the objectives, and discusses the scope of the study. Additionally, it highlights the role of machine learning in dog emotion recognition and explains how the report is organized.

Literature Review: This section reviews existing techniques for emotion recognition in animals, discussing the advantages and limitations of different models. It also examines the evaluation metrics used for emotion recognition models and provides an overview of the current state-of-the-art methods.

Methodology: This section details the steps taken in the study, including data acquisition and description, model selection and training, and the various machine learning methods employed, including transfer learning and the use of VGG models. It also discusses the steps involved in emotion detection, model evaluation, and considerations for model selection and deployment.

Results: This section presents the performance comparison of the models used in the study and discusses the findings.

Conclusion and Future Scope: This section summarizes the key findings of the study and suggests directions for future research.

References: This section lists the sources and references used throughout the report

2. Literature Review

2.1 Existing Machine Learning Techniques For Dog Emotion Classification

Sentiment analysis (SA) stands as a sophisticated method for extracting, scrutinizing, and discerning the underlying emotions and sentiments embedded within textual or spoken information. This technique has garnered extensive application in the realms of e-commerce and service reviews, where it serves as a powerful tool for gauging customer opinions regarding specific products or services based on predefined criteria. Additionally, SA finds profound utility in social networks, where it meticulously analyzes users' attitudes across various situations, capturing nuances of sadness, happiness, irony, and more.

A plethora of algorithms and approaches have emerged to tackle the intricacies of SA, each boasting distinct advantages and limitations alongside varying degrees of accuracy. In this discourse, we have meticulously curated information pertaining to some of the most prevalently employed contemporary algorithms, complemented by their associated works and the results obtained therefrom. It is imperative to underscore that the evaluation methodologies for these techniques often diverge, posing significant challenges in the accurate comparison of individual techniques' accuracies.

The landscape of sentiment analysis is replete with complexities, and this paper endeavors to elucidate the nuanced performance metrics and methodological variances inherent to each approach, thereby offering a comprehensive understanding of the current state of the art in sentiment analysis.

Used algorithms and their assessment

2.1.1 Keyword based approach

One of the most prevalent techniques in sentiment analysis is the keyword-based approach. This method excels in detecting emotions within text at the fundamental word level, proving to be highly reliable, particularly for analyzing words that inherently convey emotions and for parsing simple sentences where emotions are explicitly articulated.

2.1.2 Learning based approaches

Learning-based approaches in sentiment analysis are trained using a designated training set. Following this training phase, the classifier is capable of determining emotions either directly from individual words or through intricate structure of classifiers. There exist several types of learning-based algorithms, each with its unique methodology and application nuances.

2.1.3 Support Vector Machines

One of the learning-based methodologies is the Support Vector Machine (SVM) binary classification technique. This algorithm leverages training examples that encapsulate categorical information. SVM constructs a model that delineates input data into pre-established categories through linear classification.

Table 1. Related works using SVM algorithm and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[2]	2002	SVM with features based on unigrams	N/A	82.9
[3]	2007	SVM with General Inquirer, WordNet Affect, and other features	N/A	73.89
[4]	2009	SVM with Information Gain feature extraction	92.86 (POS) 88.28 (NEG)	91.15
[5]	2011	SVM with a linear kernel	N/A	80.29
[6]	2015	Speaker-dependent SVM with thresholding fusion	N/A	75.67

Table 2.1

2.1.4 Naïve Bayes Classifier

Another learning-based approach is the Naïve Bayes Classifier (NBC). Various forms of NBC exist based on the representation of input text. Among these, the multinomial Naïve Bayes model yields superior results and greater accuracy compared to other event models. In this model, the frequencies of specific words are counted and represented as a vector.

Table 2. Works related to Naïve Bayes Classifier and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[2]	2002	NBC with features based on unigrams	N/A	78.7
[9]	2012	NBC and Naïve Search	N/A	~85
[10]	2013	NBC with Facebook Query Language query	72	N/A
[11]	2014	ERR-based NBC	84	N/A

Table 2.2

2.1.5 Hidden Markov Model

The Hidden Markov Model (HMM) is a prevalent and straightforward learning-based algorithm for emotion detection. Essentially, HMM is a method that assigns classes to a sequence of observations.

Table 3. Works related to Hidden Markov Model and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[13]	2003	Continuous HMM	N/A	77.8
[14]	2012	A High-order HMM with Viterbi algorithm	35.3	N/A
[15]	2013	3 states HMM	N/A	82.95

Table 2.3

2.1.6 Hybrid approaches

This methodology amalgamates the previously discussed approaches. It discerns emotions by identifying specific keywords, recognizing learned patterns, and utilizing supplementary information from various dictionaries and thesauri.

Table 4. Works related to hybrid approaches and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[17]	2004	Hybrid SVM (PMI/Osgood and Lemmas), 100 folds	N/A	89
[18]	2013	Keyword-spotting method and rule-based method	76.97	N/A
[19]	2013	Multinomial NBC with greedy search	N/A	85
[20]	2014	NBC and SVM using Information Gain and Chi-Square methods	N/A	71
[21]	2015	SVM and CRF with applied rules	N/A	91

Table 2.4

2.1.7 Assessment of algorithms

One of the primary metrics used to evaluate AI performance is the F-score. It integrates two fundamental parameters: precision and recall. Precision gauges the ratio of correctly predicted positive instances out of all instances predicted as positive, while recall assesses the ratio of correctly predicted positive instances out of all actual positive instances. These metrics are computed using four key values: TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives).

The F-score, also known as the harmonic mean of precision and recall, quantifies the overall performance of a classification model. It is expressed as a percentage and ensures a balanced assessment of both precision and recall.

Another essential metric is Accuracy, which denotes the correctness level of specific algorithms. In this context, the average accuracy across all emotional classifications is calculated straightforwardly.

To address the inconsistency in evaluating emotion classification algorithms, there is a proposed solution: the establishment of a standardized testing framework. This framework would impose uniform conditions and criteria for testing algorithms, utilizing identical linguistic resources. Such standardization would mitigate inaccuracies and harmonize results onto a common scale of measurement.

Presently, there exists a lack of specific guidelines on standardizing approaches for testing algorithms or achieving comparability across diverse results.

3. Methodology

3.1 Data Acquisition and Description

This dataset consists of images of dogs with labels indicating their emotional state: angry, happy, relaxed, or sad. The displayed images stemmed from several online sources and were manually annotated.

https://www.kaggle.com/datasets/danielshanbalico/dog-emotion

Number of Images: The dataset includes 4000 images of dogs in total, with 1000 images for each emotion category (angry, happy, relaxed, sad).

3.2 Model Selection and Training

In the realm of dog emotion recognition using machine learning (ML) models, several approaches can be considered based on their efficacy, complexity, and suitability for the task. Here are the types of models commonly used, presented in a technical research report format:

3.2.1. Convolutional Neural Networks (CNNs):

A convolutional neural network (CNN) represents a class of machine learning models, specifically a type of deep learning algorithm well-suited for the analysis of visual data. CNNs, often referred to as convnets, utilize principles derived from linear algebra, particularly convolution operations, to extract intricate features and discern patterns embedded within images. While CNNs are primarily applied in image processing tasks, they can also be adapted to handle audio and other forms of signal data.

The architecture of CNNs draws inspiration from the structural connectivity observed in the human brain, particularly the visual cortex, pivotal in the perception and processing of visual stimuli. Artificial neurons within CNNs are strategically organized to efficiently interpret visual content, enabling comprehensive image analysis. Due to their remarkable efficacy in object identification, CNNs find widespread use in computer vision applications such as image recognition, object detection, and diverse fields like autonomous vehicles, facial recognition, and medical imaging analysis.

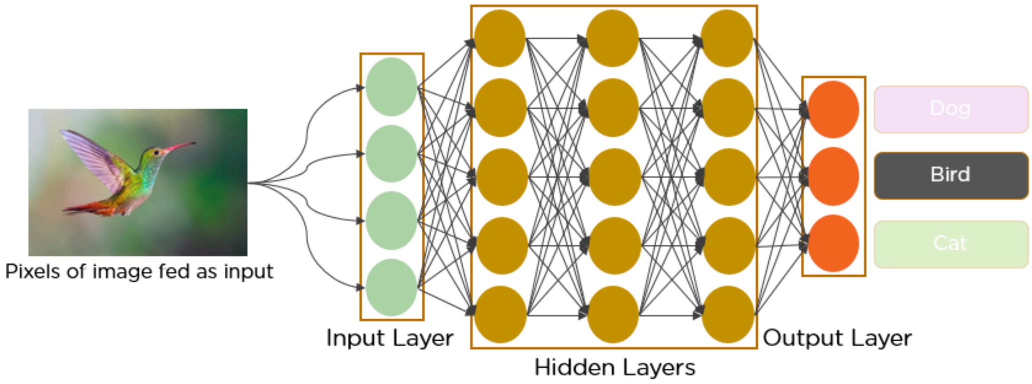


Figure 3.1

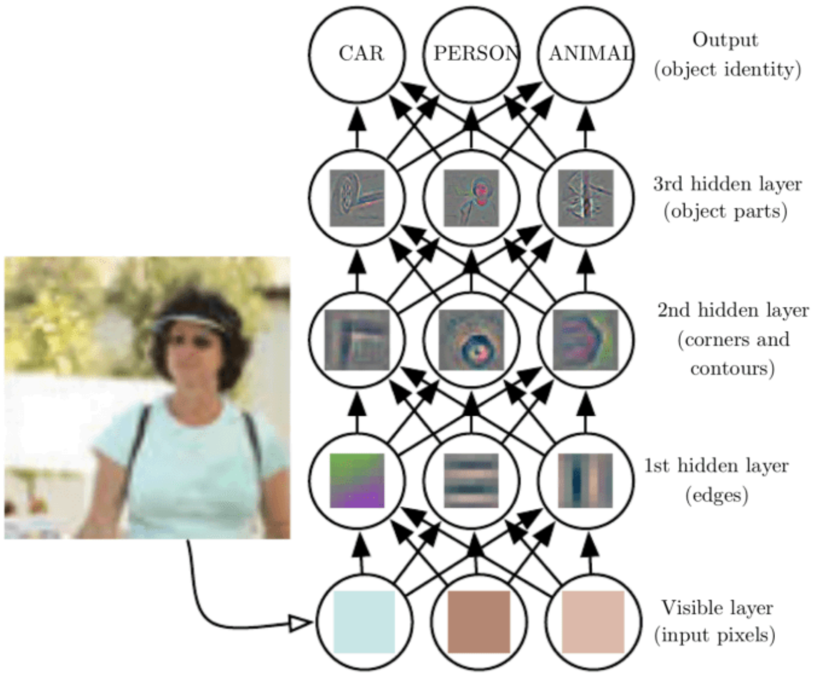


Figure 3.2

CNNs are well-suited for dog emotion recognition as they can effectively learn to extract and interpret visual patterns indicative of different emotional states in dogs.

3.2.2 Transfer Learning:

Transfer learning stands as a potent methodology within the realm of Deep Learning. It leverages the capacity to repurpose established models and their acquired knowledge to address novel challenges, thereby facilitating the training of deep neural networks even when confronted with sparse data. This advancement holds particular importance in data science, where real-world applications frequently confront the scarcity of annotated data. This discourse explores transfer learning comprehensively, elucidating its principles and examining its practical applications, empowering data scientists to confront intricate challenges with heightened efficacy and proficiency.

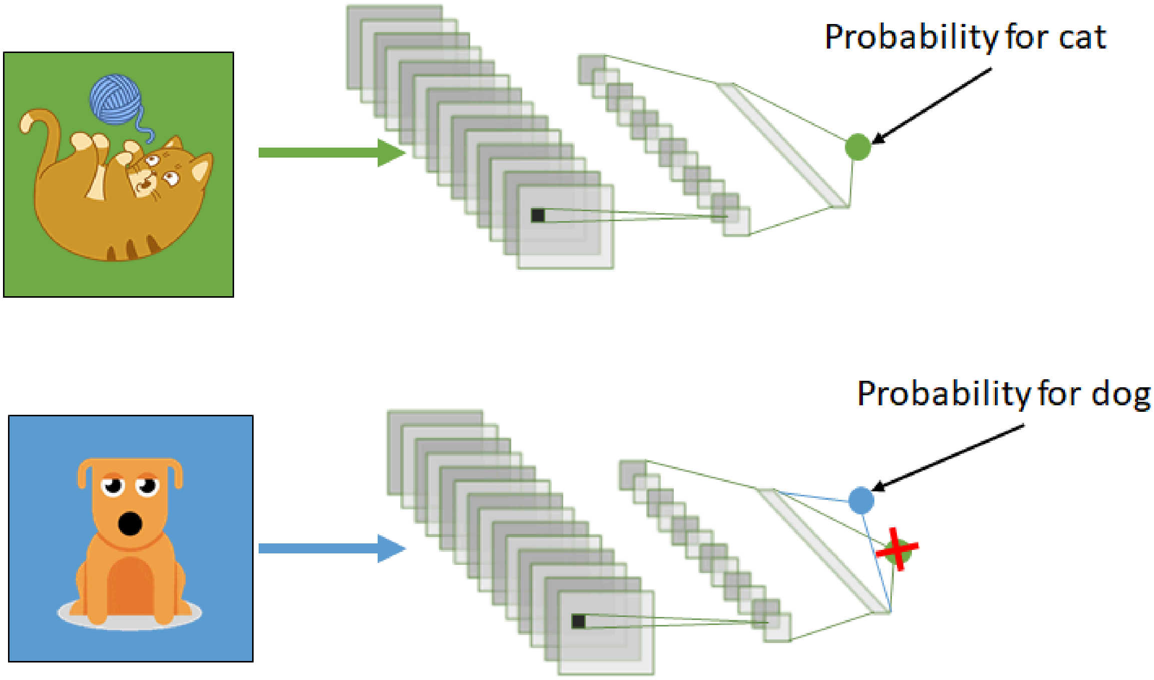


Figure 3.3

Transfer learning in CNN leverages pre-trained models to achieve strong performance with limited training data, crucial in fields like natural language processing with vast labeled datasets. It reduces training time significantly compared to building complex models from scratch, which can take days or weeks.

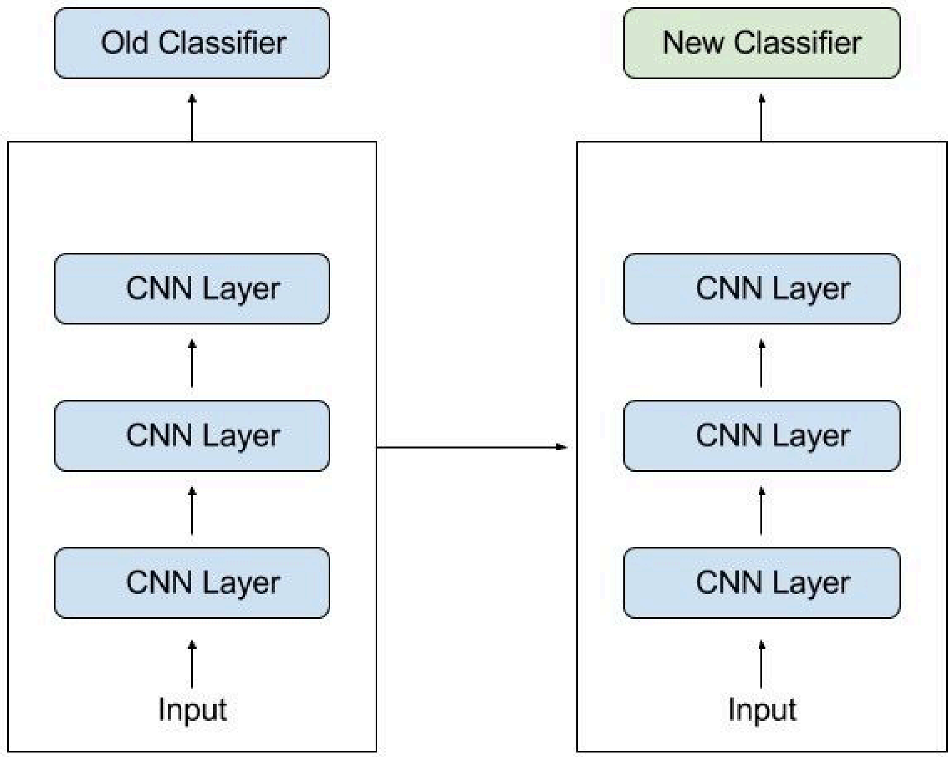


Figure 3.4

3.2.3. Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs):

Recurrent Neural Networks (RNNs) represent a specialized class of neural networks tailored for modeling sequential data. Derived from feedforward networks, RNNs exhibit behavior reminiscent of human cognitive processes. In essence, these networks possess the unique ability to anticipate sequential patterns in data, a capability not easily replicated by other algorithms.

While less prevalent in static image classification tasks, RNNs find application in contexts where the temporal evolution of emotional cues, such as in video-based emotion recognition for dogs, unfolds dynamically over time.

Key attributes that distinguish recurrent neural networks include:

1. **Internal Memory**: Central to RNN functionality, this feature enables them to retain information from previous inputs, leveraging context when processing subsequent data.
2. **Sequential Data Processing**: RNNs excel in tasks where the order of elements is critical, such as speech recognition, machine translation, natural language processing (NLP), and text generation, due to their innate capacity for handling sequential dependencies.
3. **Contextual Understanding**: RNNs possess the capability to interpret current inputs within the context of previously observed data, a pivotal aspect for tasks reliant on contextual information for accurate analysis.
4. **Dynamic Processing**: Continuously updating their internal state as new data is processed, RNNs demonstrate adaptability to evolving patterns within sequences, enhancing their utility across various dynamic data processing applications.

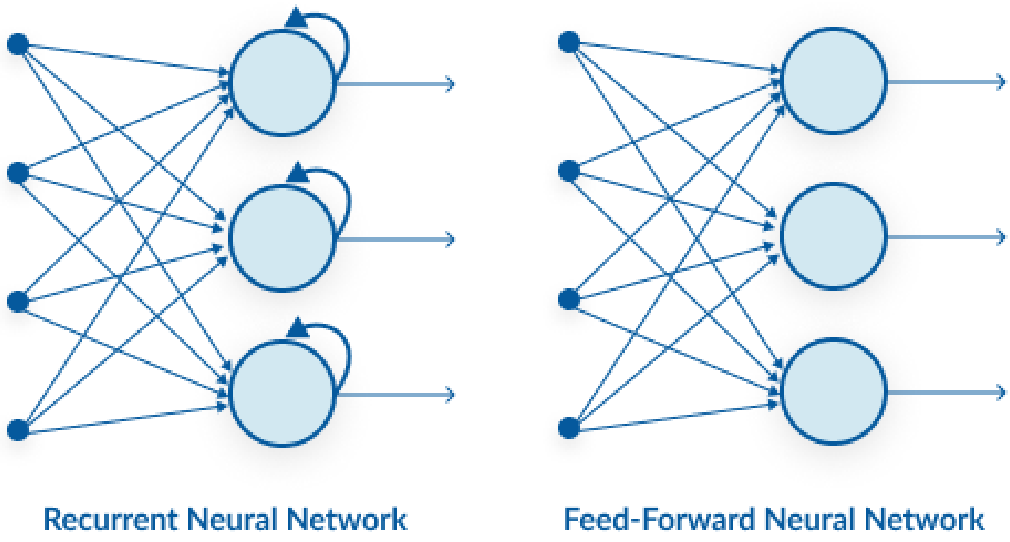


Figure 3.5

3.2.4. Support Vector Machines (SVMs) and Random Forests:

The "Support Vector Machine" (SVM) represents a supervised learning algorithm within the domain of machine learning capable of addressing both classification and regression tasks. It finds predominant use in classification quandaries, notably in text classification scenarios. In SVM methodology, each datum is depicted as a point within an n-dimensional space, where n corresponds to the number of features present, and each feature's value denotes a specific coordinate. Subsequently, classification is executed by identifying the optimal hyperplane that effectively delineates between the two classes with high precision (refer to the accompanying visual representation).

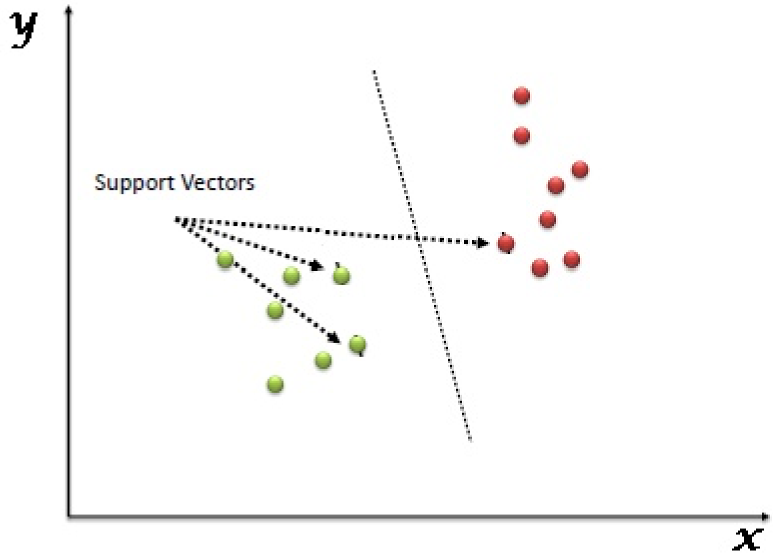


Figure 3.5

These models are straightforward to implement and interpret but may require more manual feature engineering compared to deep learning approaches. They are suitable when computational resources are limited or when interpretability of the model is paramount.

3.2.5. Ensemble Methods:

Ensemble methods represent a suite of techniques designed to enhance model accuracy by amalgamating multiple models rather than relying on a singular model. The amalgamation of these models yields a substantial increase in result accuracy, thereby contributing to the growing acclaim and adoption of ensemble methods within the realm of machine learning.

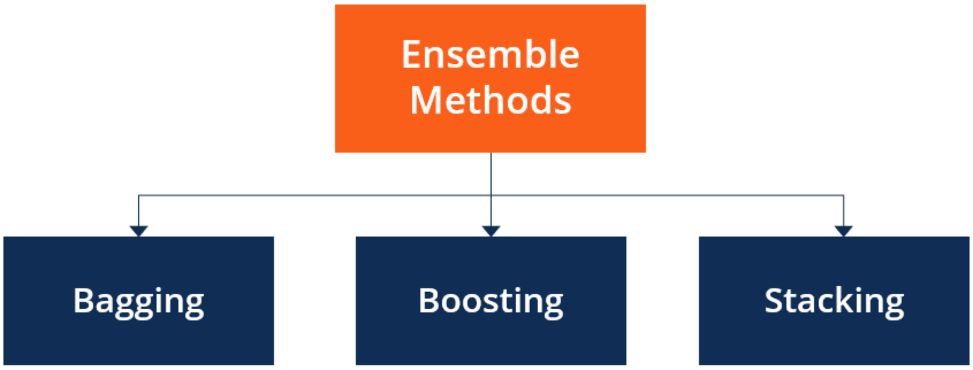


Figure 3.6

In dog emotion recognition, ensemble methods can combine predictions from different models (e.g., CNNs, SVMs) to achieve higher accuracy or to mitigate biases inherent in individual models.

3.2.6 Selection Considerations:

- Data Availability: The choice of model should align with the availability and characteristics of the annotated dog emotion dataset.
- Computational Resources: Deep learning models like CNNs and transfer learning require significant computational resources for training, while traditional ML models like SVMs may be more resource-efficient.
- Task Complexity: Models should be selected based on the complexity of the emotional states to be recognized and the temporal or spatial dependencies present in the data.

3.3 Steps involved in Dog Emotion Classification

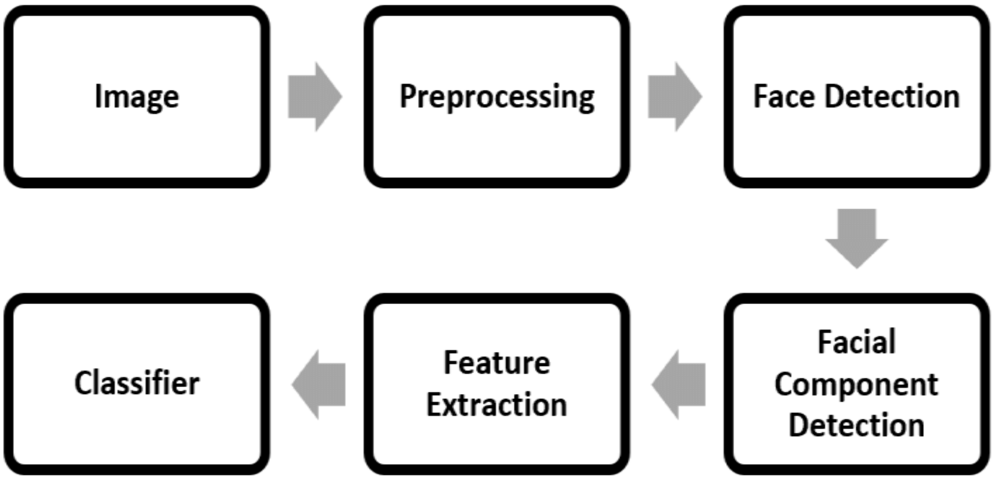


Figure 3.7

Data Collection and Preparation:

Collect a dataset of dog images annotated with emotional labels.

Resize, normalize, and augment the images to enhance dataset diversity.

Feature Extraction and Selection:

Use CNNs or pre-trained models (e.g., VGG16) to extract relevant features from the images.

Select significant features that distinguish different emotional states in dogs.

Model Selection and Training:

Choose an ML model (e.g., CNN, SVM) based on dataset characteristics and task complexity.

Train the model on the annotated dataset, optimizing parameters for accuracy.

Evaluation and Validation:

Evaluate the model using metrics like accuracy, precision, recall, and F1-score.

Validate performance with techniques such as cross-validation to ensure reliability.

Model Optimization:

Fine-tune model hyperparameters and apply regularization techniques to enhance generalization.

Deployment and Monitoring:

Deploy the trained model for real-time emotion classification in dogs.

Monitor performance over time and retrain with new data as needed.

Interpretation and Reporting:

Interpret model predictions and visualize decision-making processes.

3.4 Model Selection and Deployment Considerations

3.5.1 Factors Influencing Model Selection

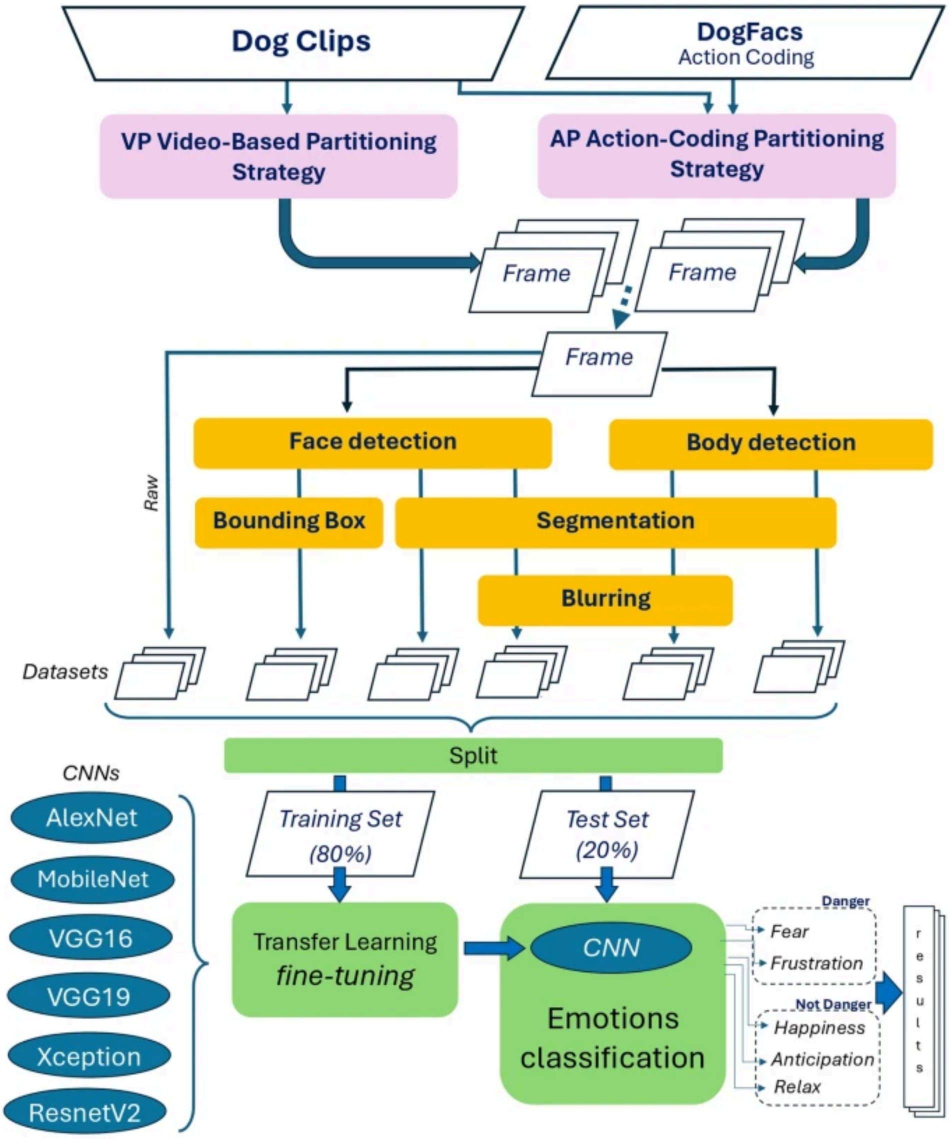


Figure 3.8

Selecting the right model for a dog emotion classifier in machine learning (ML) requires considering several crucial factors to ensure optimal performance and suitability:

Dataset Size and Complexity:

The size and complexity of the dataset influence the choice of model. Larger datasets often benefit from deep learning models like CNNs, whereas simpler tasks or smaller datasets may suffice with traditional ML models like SVMs.

Computational Resources:

Deep learning models such as CNNs require significant computational resources for both training and inference. It's essential to assess available hardware and infrastructure to ensure efficient model deployment.

Task Requirements:

Understand the specific requirements of the emotion recognition task. Models like RNNs or LSTM networks are suitable when temporal dependencies, such as changes in emotion over time, are critical.

Interpretability vs. Performance:

Balance between model interpretability and performance. Deep learning models generally offer high accuracy but can be complex to interpret. Traditional ML models like decision trees or SVMs provide better interpretability at the cost of some performance.

Transfer Learning Opportunities:

Evaluate the potential for transfer learning from pre-trained models (e.g., VGG16, ResNet). Transfer learning can expedite model training and enhance performance, particularly with limited annotated data.

Scalability and Maintenance:

Consider the scalability of the chosen model for future needs. Ensure it can be easily maintained, updated, and integrated into existing systems or applications without significant overhead.

3.5.2 Model Deployment Strategies

Deploying a dog emotion classifier model involves strategic considerations to ensure seamless integration into real-world applications:

Environment Compatibility:

Ensure the deployed model is compatible with the target environment (e.g., desktop, mobile, cloud). Optimize model size and performance to meet deployment environment constraints effectively.

Deployment Architecture:

Choose an appropriate architecture (e.g., client-server, edge computing) based on latency requirements, data privacy concerns, and the computational capabilities of the deployment environment.

Model Versioning and Monitoring:

Implement robust version control mechanisms to track model updates and improvements. Continuously monitor model performance to detect any deviations in accuracy or behavior over time.

Integration with Data Pipelines:

Seamlessly integrate the model into data pipelines for consistent data ingestion, preprocessing, and inference. Ensure compatibility with existing data formats and protocols to streamline operational workflows.

Security and Privacy:

Implement stringent security measures to protect sensitive data and model outputs. Employ techniques such as encryption, access control, and anonymization to safeguard user privacy and data integrity.

User Interface and Accessibility:

Develop a user-friendly interface that provides clear feedback and visualizes emotion recognition results effectively. Consider accessibility requirements to accommodate diverse user needs and preference Challenges of emotion recognition in dogs

3.6 **Challenges of emotion recognition in dogs**

In this subsection, we will examine our task in more detail as well as the areas where the literature is lacking. carefully analyze what challenges are specific to the task of canine emotion recognition, beginning with the recognition of human emotions and moving on to their distinction.

Deep learning-based human- automated emotion identification from photographs is an established field of study. topic based on a solid state of the art, and easily implementable with machine learning techniques, identifying critical points of the face or using deep neural networks on high-resolution photos to take advantage of microexpressions and other face characteristics.

Despite differences Human face-based emotion identification can rely on strong generalizability because of the similarities between many human faces and the same expression, regardless of ethnicity, gender, or age of eyes, eyebrows, mouth, and surrounding skin areas.

Despite dogs and humans responding differently to emotionally competent stimuli,the same process has been proved promising on dogs.

Human-face-based emotion detection can rely on strong generalizability due to age, gender, and ethnicity because various human faces share the same expression.

Also, the hair color Distribution can be difficult in some cases, such as in breeds that display spots and may pose a form difficulty in recognition and segmentation algorithms.When just one breed is taken into account, it is far more Moreover, in dogs, we cannot exploit, Other work suggests that the dog's facial expression recognition is still critical for emotion recognition.The performance of emotion recognition can be much more improved by introducing postural information.

When considering the annotation of canine data, a significant distinction from human issues becomes clear: dogs are incapable of writing or speaking, unlike humans communication to express their state. As a result, emotions of dogs represented by still images intrinsically human- perceived, there is currently no physiological record that has been successfully matched to canine emotions.

We can think about the possibility of mapping action units (e.g., DogFACS data) to emotions for emotions explainability.Although a dog's owner may have a deep understanding of their pet's behavior, we deem professional assessments (from veterinarians, dog trainers, or specialists in behavior or ethology, for example) must be Real-world photographs and videos may include noise from different backdrops, such as joyful dogs rushing over grassy fields, dejected dogs lying on floors, or drowsy dogs lounging on couches.

Furthermore, frames captured with personal devices may be out of focus, poorly composed (i.e., with the dogplaced distant from the main point or center), or not well illuminated.

These factors can introduce major biases in deep learning because attributes allocated to emotion classes may be unintentionally influenced by backdrop, surroundings, and camera specs.

Within the framework of our study, our goals include the explicit detection of potential biases and the following analysis of various data-cleaning techniques.

Strategically, we consider the cited background biases within the training dataset in ad-hoc experiment configurations, then using data-cleaning methods to assess their performance in mitigating such bias

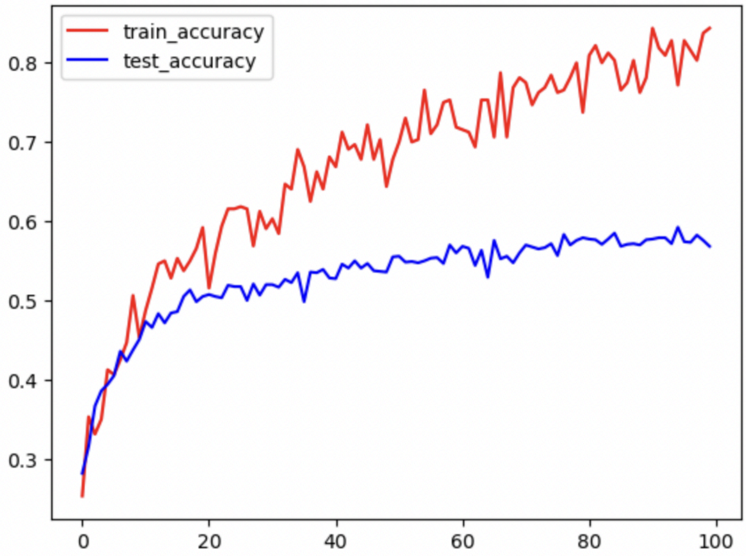
4. Results

Many pre-trained models can act as the basis for a neural network rather than needing to be trained from the start. A lot of resources and time saved is obtained from these pre-trained models

4.1 Visualize accuracy and loss

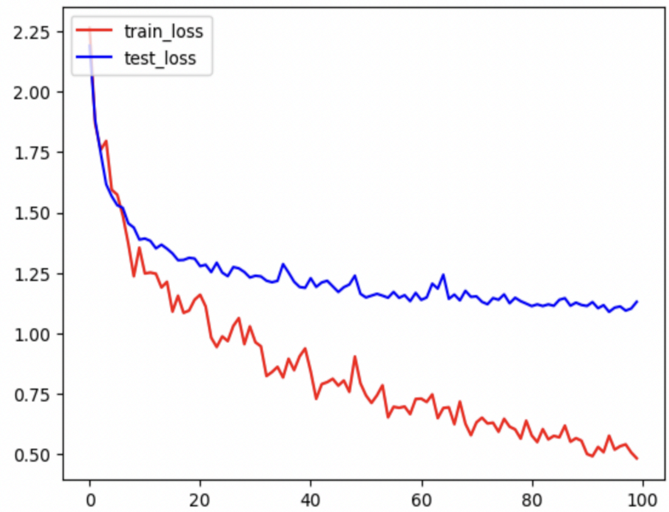
In [29]:

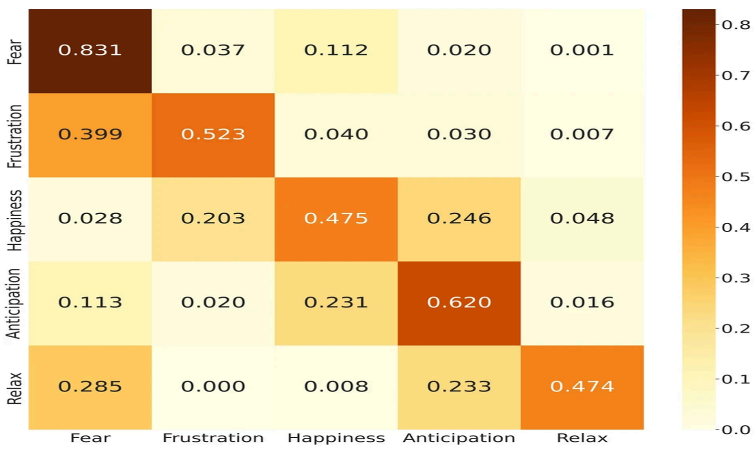
```
#Stpete Ishii https://www.kaggle.com/code/stpeteishii/transfer-learning-for-manga-facial-expressions

# Visualise train / Valid Accuracy
plt.plot(model.history.history["categorical_accuracy"], c="r", label="train_accuracy")
plt.plot(model.history.history["val_categorical_accuracy"], c="b", label="test_accuracy")
plt.legend(loc="upper left")
plt.show()
```



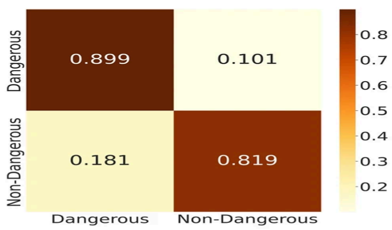
In [30]:

```
#Stpete Ishii https://www.kaggle.com/code/stpeteishii/transfer-learning-for-manga-facial-expressions

# Visualise train / Valid Loss
plt.plot(model.history.history["loss"], c="r", label="train_loss")
plt.plot(model.history.history["val_loss"], c="b", label="test_loss")
plt.legend(loc="upper left")
plt.show()
```





(a) Emotion recognition: confusion matrix



(b) Danger recognition: confusion matrix

5. Conclusion and future scope

Many pretrained models serve as foundational frameworks for neural networks, obviating the need for training from scratch. This approach not only conserves significant resources but also accelerates development timelines.

Unlike the controlled environments prevalent in existing literature, our study is dedicated to analyzing and discerning dog emotions and threat levels using photographs captured in real-life scenarios.

Following an identification of the primary challenges in canine emotion recognition, we implemented segmentation strategies as pivotal components of our methodology. These strategies proved instrumental in sharpening the focus on relevant variables while minimizing misclassification stemming from background noise, thereby bolstering the robustness of our classification outcomes. Notably, techniques such as face bounding box delineation and blurring exhibited exceptional performance.

Moreover, our models, coupled with segmentation strategies, exhibited a favorable bias towards minimizing false negatives in identifying dangerous behaviors. This bias mitigation is crucial in practical applications where the consequences of a false negative could be severe, underscoring the reliability of our approach for contexts demanding vigilance and preemptive measures.

In summary, our findings underscore the efficacy of transfer learning in multifaceted and dynamic environments, underscoring the indispensable role of segmentation strategies in achieving superior accuracy in real-world analysis of canine behavior. Our methodology promises significant utility in enhancing safety protocols in human-canine interaction research and practical applications.

5.1 Future Scope

In this segment, we refine our strategy to ensure that discussions on forthcoming advancements synergistically complement and underscore our current research achievements, providing a coherent synopsis of our impact and the ongoing potential in this domain. Our research establishes a fundamental stride in the application of deep learning for the recognition of canine emotions. The dataset's diverse, uncontrolled environmental settings serve as a valuable and genuine source for scrutinizing authentic emotional manifestations, emphasizing the immediate contribution of our research to this discipline. The utilization of transfer learning mitigates the unintended consequence of incorporating noisy frames. Mitigating bias has constituted a pivotal facet of our investigation, guided by methodologies that prioritize the detection and analysis of canine facial features. Future improvements in dataset diversity, particularly through the augmentation of underrepresented breeds, could further enhance the resilience and inclusivity of our bias-mitigation approach.

6. References

https://www.nature.com/articles/s41598-023-30442-0

https://www.kaggle.com/code/sasakitetsuya/dog-emotion-classify-model-by-mobilenetv2-vgg16

https://www.hackersrealm.net/post/facial-emotion-recognition-using-python#google_vignette

https://www.researchgate.net/publication/368843964_Pet_dog_facial_expression_recognition_based_on_convolutional_neural_network_and_improved_whale_optimization_algorithm

https://www.mdpi.com/1999-5903/14/4/97

https://link.springer.com/article/10.1007/s00521-024-10042-3

https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-55462022000100203

https://builtin.com/data-science/transfer-learning#:~:text=Transfer%20learning%2C%20used%20in%20machine,to%20improve%20generalization%20about%20another.

https://iopscience.iop.org/article/10.1088/1742-6596/772/1/012039#:~:text=The%20most%20commonly%20used%20techniques,more%20than%2090%25%20determining%20accuracy.

https://www.nature.com/articles/s41599-023-01816-6

https://ieeexplore.ieee.org/document/10112801

https://www.mdpi.com/2073-431X/11/10/152

https://www.researchgate.net/publication/341782389_Human_Emotion_Detection_using_Machine_Learning_Techniques https://ijai.iaescore.com/index.php/IJAI/article/view/21368

https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neural-networks-rnn/

- **Appendix**

Sample code for model implementation (python)

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, BatchNormalization, Flatten, LeakyReLU,
Dense, Dropout
import matplotlib.pyplot as plt
from tensorflow.keras.callbacks import Callback, EarlyStopping, ReduceLROnPlateau
```

Prepare ImageDataGenerator

```
#https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator

img_generator = tf.keras.preprocessing.image.ImageDataGenerator(
                              #rotation_range=90,
                              brightness_range=(0.5,1),
                              #shear_range=0.2,
                              #zoom_range=0.2,
                              channel_shift_range=0.2,
                              horizontal_flip=False,
                              vertical_flip=False,
                              rescale=1./255,
                              validation_split=0.3)
```

Visualize a batch of images

In [22]:

```python
#Stpete Ishii https://www.kaggle.com/code/stpeteishii/transfer-learning-for-manga-facial-expressions

imgs, labels = next(iter(img_generator_flow_train))
for img, label in zip(imgs, labels):
    value=np.argmax(label)
    plt.imshow(img)
    plt.title(reverse_mapping[value])
    plt.show()
```

Import a pretrained model

```python
#https://www.tensorflow.org/api_docs/python/tf/keras/applications/InceptionV3

base_model = tf.keras.applications.InceptionV3(input_shape=(224,224,3),
                                               include_top=False,
                                               weights = "imagenet"
                                               )
```

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/inception_v3/inception_v3_weights_tf_dim_o
rdering_tf_kernels_notop.h5
87910968/87910968 [==============================] - 0s 0us/step
```

Create a Model

```
model = tf.keras.Sequential([
    base_model,
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(4, activation="softmax")#4 classes
])
```

```
model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 inception_v3 (Functional)   (None, 5, 5, 2048)        21802784

 max_pooling2d_7 (MaxPooling  (None, 2, 2, 2048)       0
 2D)

 flatten_1 (Flatten)         (None, 8192)              0

 dense_2 (Dense)             (None, 4)                 32772

=================================================================
Total params: 21,835,556
Trainable params: 32,772
Non-trainable params: 21,802,784
_________________________________________________________________
```

Classification Report

```python
#Stpete Ishii https://www.kaggle.com/code/stpeteishii/transfer-learning-for-manga-facial-expressions

LABEL=[]
for item in labels:
    LABEL+=[np.argmax(item)]
print(LABEL)
```

```
[3, 2, 0, 3, 1, 3, 0, 1, 3, 3, 3, 0, 0, 0, 3, 1, 2, 2, 0, 2, 2, 1, 1, 1, 2, 0, 0, 0, 3, 0, 3, 1]
```

```python
PRED=pred_labels.numpy().tolist()
print(PRED)
```

```
[3, 2, 3, 3, 1, 3, 0, 1, 3, 3, 3, 0, 0, 3, 3, 1, 2, 2, 2, 2, 2, 3, 0, 2, 3, 2, 0, 0, 3, 0, 2, 2]
```

```python
from sklearn.metrics import classification_report
print(classification_report(LABEL, PRED))
```

```
              precision    recall  f1-score   support

           0       0.86      0.60      0.71        10
           1       1.00      0.43      0.60         7
           2       0.50      0.83      0.62         6
           3       0.67      0.89      0.76         9

    accuracy                           0.69        32
   macro avg       0.76      0.69      0.67        32
weighted avg       0.77      0.69      0.68        32
```