Paradigma, Vol. 26, No. 2, September 2024, P-ISSN 1410-5063, E-ISSN: 2579-3500, Page. 102-108
Published: LPPM Universitas Bina Sarana Informatika
PARADIGMA
website: http://jurnal.bsi.ac.id/index.php/paradigma/

Classification of Dog Emotion Using Transfer Learning on Convolutional Neural Network Algorithm

Steven Tribethran[1], Nicolas Jacky Pratama Hasan[2], Abdul Rahman[3]

[1,2]Informatics, Universitas Multi Data Palembang, Palembang, Indonesia
[3]Electrical Engineering, Universitas Multi Data Palembang, Palembang, Indonesia

ARTICLE INFORMATION

Artikel History:
Received: Aug. 8, 2024
Revised: Aug. 30, 2024
Accepted: Sept. 18, 2024
Available Online: Sept. 30, 2024

Keyword:

Dog
Convolutional Neural Network
Emotion classification
Transfer Learning
VGG16

ABSTRACT

Recognizing your pet's emotions are very important to improve health, welfare and to detect certain diseases in the animal. The emotions in question are categorized into four categories, namely anger, happiness, calmness, and sadness. The model is designed by utilizing transfer learning techniques using the VGG16 architecture to perform image feature extraction for dog emotion classification based on the image of the animal's facial expression. The research produced an accuracy value of 96.72% on the training set and 88.05% on the validation set, as well as an average F1-Score value of 84.30% on the test set. This research shows the great potential of utilizing transfer learning in dog emotions classification and contributes to more advanced emotion recognition techniques to improve pet's welfare.

Corresponding Author:

Steven Tribethran,
Informatics,
Multi Data Palembang University,
Jl. Rajawali 14, Palembang, Indonesia, 30113,
Email: steven.tribethran@mhs.mdp.ac.id

INTRODUCTION

Emotion is something that is felt by living beings, where different emotions arise from within each individual, such as love, happiness, fear, anger, and sadness that can change based on the circumstances, and the emotions that are being felt can also affect the way each individual thinks and behaves (Sudianto, 2022). In humans, emotions can be shown through facial expression (Dzedzickis et al., 2020; Ko, 2018), through voice (Dzedzickis et al., 2020), through typed text (Alswaidan & Menai, 2020), there are also human emotion recognition systems based on body gestures or body movements (Ahmed et al., 2020; Piana et al., 2016). When annotating canine data, a crucial difference from human subjects is that dogs cannot express their emotions verbally or in writing. Therefore, interpretations of their emotions from images rely solely on human perception. Moreover, there remains a lack of a comprehensive physiological dataset that accurately links to dog emotions, leaving this area largely unexamined

Emotion recognition in animals has been extensively conducted to improve the welfare of farm animals, monitored through their facial expressions and sounds (Neethirajan et al., 2021). However, this research is more focused on farm animals. At the same time, there is still limited research on emotion recognition in other types of animals, especially other pets such as cats or wild animals that interact with humans. Researchers have studied emotion recognition in pets like dogs to identify dog emotions that humans can easily recognize from their facial expressions, such as anger, fear, sadness, and happiness (Burza et al., 2022). However, these studies are generally conducted in environments that are already familiar to dogs, leaving a gap in understanding how new environments or interactions with strangers may affect emotion recognition accuracy. Emotion recognition in dogs is particularly important in the context of human safety, to guard against potential aggressive behavior or threats posed by certain individuals (Amici et al., 2019). Researchers have also used emotion recognition in animals to build an emergency notification system that prevents stray dog attacks on humans through video monitoring (Chen et al., 2023) to detect dogs' physical and emotional conditions (Broomé et al., 2023). To recognize the intent of dog barking (Hantke et al., 2018), developing more adaptive and automated

DOI: https://doi.org/10.31294/p.v26i2.5295

real-time technology for continuous monitoring remains underexplored. Additionally, the limited integration of various modalities, such as sound, facial expressions, and body movements, to improve the accuracy of emotion recognition indicates an opportunity to enhance the quality of animal emotion recognition outcomes in the future.

In image processing, the process of classifying objects is part of the computer vision problem. The purpose of image classification is to classify images into several categories based on needs (Riyadi et al., 2021). Convolutional Neural Network (CNN) is one of the leading Neural Networks in the field of deep learning. Computer Vision using CNN has shown significant achievements, such as in face recognition tasks, autonomous vehicles and others (Li et al., 2022).

Research using the CNN method in handwritten digit recognition that combines the advantages of CNN and Support Vector Machine (SVM) classifiers in handwritten digit recognition achieved a classification accuracy of 99.28% on the MNIST handwritten digit dataset (Ahlawat & Choudhary, 2020). Prediction of animal vocal emotion using CNN resulted in an accuracy of 88% (Totakura et al., 2020). The use of CNN in inferring emotion tags from object images resulted in an accuracy of 85% (Manzoor et al., 2020). Recognition of animal facial expressions for animal welfare assessment on farms using CNN resulted in an accuracy of 96% (Hansen et al., 2021).

CNN has a disadvantage that it needs a lot of data to get optimal performance in object identification (Oquab et al., 2014). Transfer Learning is a technique used to extract knowledge that has been learned from one or more tasks and apply that knowledge to the desired target task, the advantage is obtained from using pre-trained machine learning models that have been trained on a task to use that knowledge on a new task with the context of having a task related to the previous task (Tammina, 2019a). In addition, by applying transfer learning, many trainable parameters become relatively fewer, so the training time required is shorter than training the CNN model from scratch (Sabri, 2023). VGG16 is a type of convolutional neural network recognized for its strong performance in image classification, thanks to its deep architecture and its capability to extract complex features from images. Its design includes small convolutional filters and multiple layers, enabling it to effectively learn hierarchical features. This ability is particularly important for recognizing dog emotions, as it allows for the detection of subtle variations in facial expressions that can reflect different emotional states (R et al., 2022).

There have been several previous studies using transfer learning with the VGG16 architecture. The use of VGG16 transfer learning in classifying dog and cat images achieved an accuracy of 95.40%% (Tammina, 2019b). Emotion recognition through video using VGG16 transfer learning architecture resulted in an accuracy of 87% (Licia Sbattella Co-Supervisor et al., 2018). VGG16 managed to get 89% accuracy in analyzing the effect of VGG16 transfer learning on dog breed classification (Lee et al., 2022).

Based on the literature review, numerous studies have demonstrated the effectiveness of convolutional neural networks (CNNs) in detecting and classifying objects within images. This method has shown promising results in predicting emotions through various approaches and feature extraction techniques. However, a notable limitation of CNNs is their requirement for a substantial amount of data during the training phase to accurately extract image features. To address this challenge, this research investigates the application of CNNs and Transfer Learning, explicitly utilizing the VGG16 architecture, to classify emotions in images of dog facial expressions. VGG16 was selected for its proven efficacy in various image classification tasks, particularly its ability to extract complex and deep features from images. By leveraging knowledge from larger datasets like ImageNet, the VGG16 architecture can help mitigate the limited available image data in dog emotion recognition tasks.

This study aims to enhance the accuracy of emotion classification in dogs while contributing to developing more sophisticated and reliable emotion recognition technologies. Ultimately, it seeks to improve animal welfare by providing better tools for understanding canine emotions, thus paving the way for pet care and training advancements.

RESEARCH METHOD

The research process for designing a dog emotion classification model using transfer learning involves several key stages to ensure a practical and systematic approach. The process begins with preparing a dataset of dog facial expression images, which involves gathering many labeled images representing various dog emotions. Following this, the preprocessing phase occurs, where all images are resized, normalized, and possibly adjusted for lighting or contrast to ensure consistency across the dataset.

Next, the dataset is divided into training, validation, and test sets. The model uses the training set to learn, the validation set to fine-tune hyperparameters and evaluate performance during training, and the test set for final evaluation. To improve the model’s generalization ability, augmentation techniques such as rotation, flipping, and zooming are applied to the training set, creating more variations of the existing images.

Once the dataset is prepared, the transfer learning image classification model is trained using the VGG16 architecture. This architecture leverages pre-trained weights from a large dataset like ImageNet to speed up learning and improve performance. After training, the model is thoroughly evaluated on the validation and test sets to assess its ability to classify dog emotions accurately.

Figure 1 illustrates the flow chart of this research process, starting with the initial dataset preparation, proceeding through the various stages of preprocessing, augmentation, and model training, and concluding with the model evaluation stage. This comprehensive approach ensures that the model is well-trained and robust in classifying dog emotions.

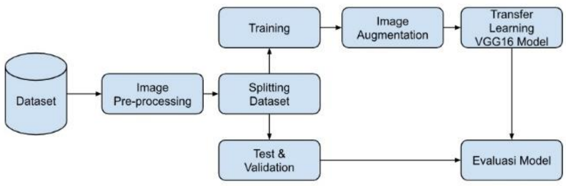

Figure 1. Research Stages Flowchart

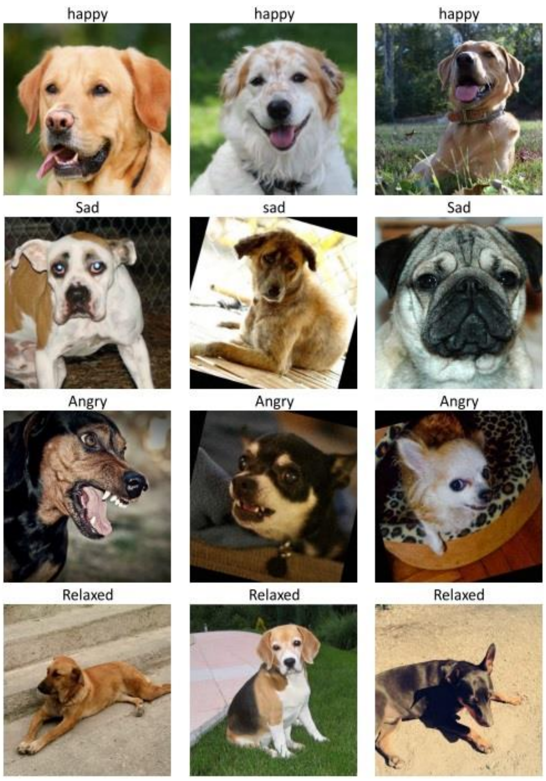

Figure 2. Sample of Dog Facial Expression

The dataset used in this study consists of images of dog facial expressions which are divided into 4 categories of emotions, namely anger, happiness, calmness, and sadness as shown in Figure 2. Each emotion category has 1000 images, totaling to 4000 images which are splitted into 80% of the data used for training, 10% used for validation and 10% used for testing purposes. This research dataset comes from Kaggle with the link https://www.kaggle.com/datasets/danielshanbalico/dog-emotion.

Before being used for model training, the images in the dataset go through a preprocessing stage to ensure the consistency of the data used. The preprocessing steps include resizing, normalization, and augmentation.

Preprocessing is performed on all images by resizing the image to 224x224 pixels while maintaining 3 RGB color channels, based on the default input size of the VGG16 architecture. Then, the pixel values in each image were normalized to a range of 0 to 1 by dividing each pixel value by 255.

To increase the amount of training data and reduce overfitting, data augmentation such as rotation and horizontal flip is performed on the images in the training set. Augmentation is done to help the model learn certain “capture” scenarios for images that are not in the dataset.

This research uses the VGG16 architecture as the base model to classify the emotions contained in dog facial expression images. The default convolution architecture of VGG16 can be seen in Figure 3 (Ferguson et al., 2017). VGG16 is a well-known CNN architecture renowned for its simplicity and efficiency in image recognition tasks. Designed by the Visual Geometry Group at the University of Oxford, it consists of 16 trainable layers, including 13 convolutional layers and 3 fully connected layers.

Each convolutional layer uses small 3×33 kernels with a stride of 1 and the same padding, enabling the network to capture detailed local visual patterns while maintaining spatial information. After each convolutional block, the network applies a 2×22 max pooling layer to reduce the spatial dimensions and compact the feature representations progressively.

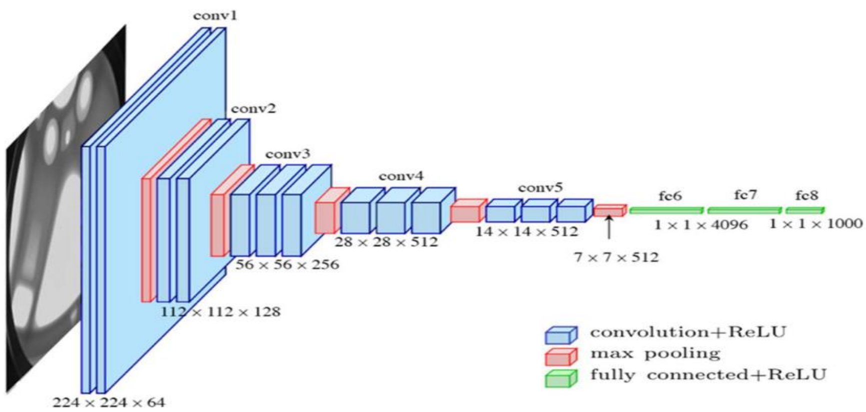

Figure 3. VGG16 Default Architecture

The design organizes the architecture into five main blocks. The first two blocks each have two convolutional layers with increasing filters (64 and 128). The following blocks each consist of three convolutional layers with progressively larger filters (256, 512, and 512). These convolutional layers are followed by three fully connected layers, culminating in a softmax layer for classifying images into 1000 categories.

The main strength of VGG16 lies in its consistent design with small kernel sizes, which simplifies the generalization of visual features. Despite its impressive performance on image recognition tasks, the model has a significant drawback in its high computational and memory demands due to its 138 million parameters. This slows down training and increases resource consumption.

VGG16 has served as a foundational architecture for many research efforts and applications, particularly in transfer learning. Pre-trained weights from VGG16 on the ImageNet dataset can be adapted for other tasks with minimal adjustments, showcasing its versatility and widespread impact on computer vision.

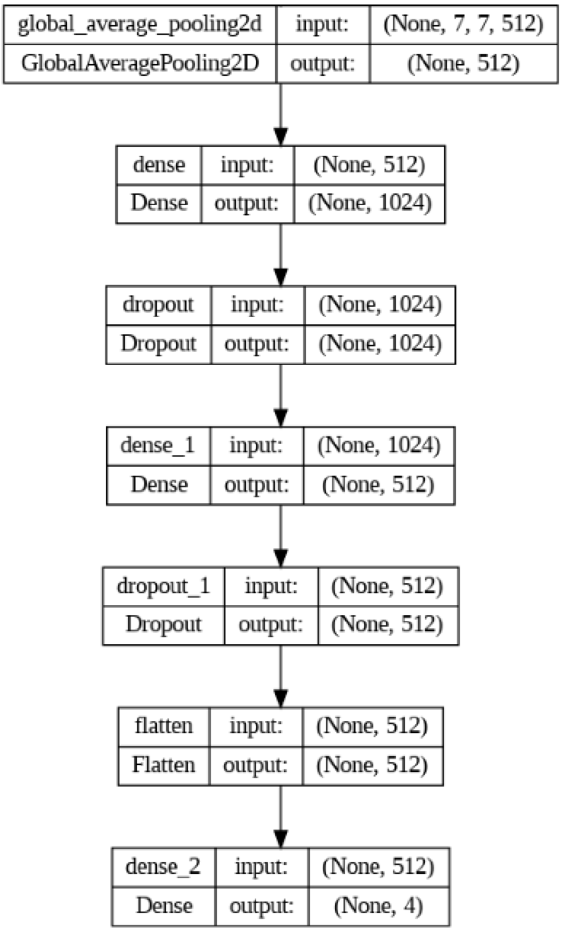

Figure 4. New Layers of Modified Architecture

The steps of applying VGG16 transfer learning start with initializing the VGG16 model with weights that have been trained on the ImageNet dataset. The last layer (fully connected layer) of VGG16 is replaced with 7 new layers, namely GlobalAveragePooling2D, 4 overfitting prevention layers, namely 2 Dense layers with L2 regularization and 2 Dropout, Flatten layers and the final layer is a Dense which has nodes count according to the number of emotion categories used (4 categories).

The initial convolution layer up to the 9th layer before the last layer of the VGG16 architecture is frozen to retain the weights learned from the previous task and the last 8 layers are retrained to match the existing knowledge to the new data. Figure 4 shows the new layers and the number of nodes of each layer added to the default VGG16 architecture

RESULTS AND DISCUSSION

To measure the success rate (performance) of the model in dog emotion classification on the animal's emotion image, the accuracy and loss metrics are used as evaluation metrics when training on the training and validation sets. Accuracy shows how accurate the predictions made by the model against the four emotion categories used. Loss shows how much error the model makes in its predictions.

During the trainng phase, the model achieved its best performance at the 29th epoch. The highest accuracy obtained on the training set was 96.72%, indicating that the model had effectively learned from the training data. Additionally, when evaluated on the validation set, the model achieved an accuracy of 88.05%, reflecting good generalization capability but leaving room for improvement.

Figure 5 illustrates this performance by presenting a graph of the model's accuracy over 30 training epochs. The graph shows a gradual increase in accuracy with minor fluctuations, a common occurrence during training until it reaches the optimal point at the 29th epoch. After this, we observe signs of overfitting or performance stabilization as the accuracy slightly changes or declines at the 30th epoch.

The graph provides a visual insight into the model's accuracy progression during training, aiding in analyzing performance trends and evaluating the stability of the learning process.

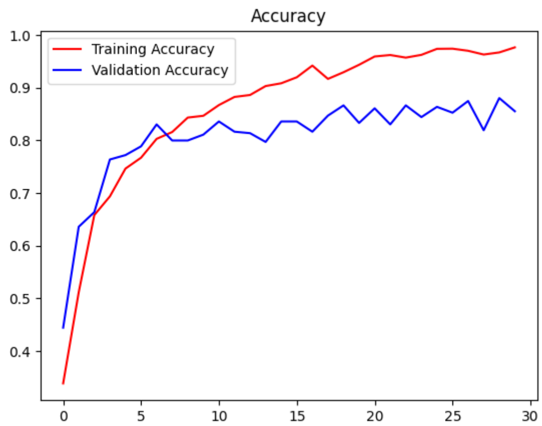

Figure 5. Model Accuracy Graph

In contrast to accuracy metric that shows the higher the value, the better the performance of the model, loss with a high value indicates that the model does not have good performance and more often makes mistakes when making predictions or mispredictions, so the loss value generated by the model is expected to have a positive value that is close to 0.

Figure 6 visually represents the model's loss progression throughout the training phase. The graph tracks the loss values throughout 30 epochs, observing how the model's performance evolved. Initially, the loss decreases steadily as the model learns from the data, but by the 29th epoch, we can see a stabilization in the loss value, particularly in the training set. The loss on the validation set also follows a similar trend, though it tends to fluctuate more, reflecting the challenges the model faced in generalizing to new, unseen examples. This graph is an important diagnostic tool for understanding the model's learning dynamics and identifying potential issues like overfitting.

At the 29th epoch, the model produced a loss value of **0.4466** on the training set, indicating that the model's predictions were relatively close to the actual values in the training data. However, there was still some room for improvement. On the validation set, the loss value was higher at **0.7935**, suggesting that while the model performed well on the training data, it had more difficulty generalizing to unseen data, which is common in machine learning models.

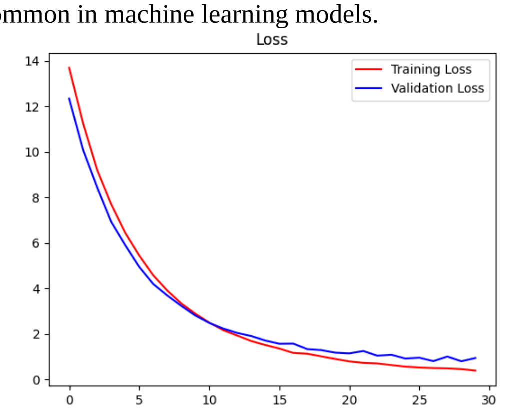

Figure 6. Model Loss Graph

Evaluation of model performance is carried out on the test set to measure the performance of the model when predicting dog emotions in images that have never been encountered or are not in the training set. Confusion matrix shows the results of the accuracy and prediction error made by the model in each emotion category. The results of model evaluation using confusion matrix can be seen in Figure 7.

The resulting model demonstrated a strong ability to predict dog emotions with detailed outcomes: it accurately detected angry emotions in 86 out of every 100 images, achieving high precision and recall for this category. Similarly, it identified happy emotions in 85 images per 100, showing consistent reliability. For calm emotions, the model successfully detected them in 80 images out of 100, indicating a slightly lower but still substantial performance. Lastly, it recognized sad emotions in 86 images of every 100, maintaining a strong detection rate. These results collectively contributed to an overall average F1 score of 84.30%, reflecting the model's balanced performance across all categories.

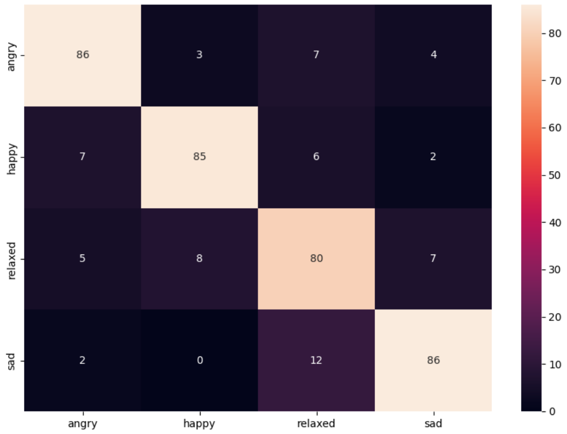

Figure 7. Confusion Matrix Result

Table 1 describes the classification report on the model’s performance on the test set, with the last column (Support) indicating the distribution of each class in the test set. The classification report shows that the model performs well in predicting angry and sad classes with similar precision and recall values (0.86-0.86 and 0.86-0.87) and similar F1-score of 0.86 for both classes, suggesting that it is balanced in identifying its own class. On the other hand, the happy class has a slightly higher recall (0.89) compared to its precision (0.85) with an F1-score of 0.87, showing that the model is better at identifying happy classes but might misidentify other classes as happy. However, the model performs the worst in the relaxed class with the lowest value in both precision (0.80) and recall (0.76) resulting in a lower F1-score of 0.78, indicating the model needs further optimization and more training data to improve the model’s performance in this class.

Table 1. Classification Report

	Precision	Recall	F1-Score	Support
Angry	0.86	0.86	0.86	100
Happy	0.85	0.89	0.87	96
Relaxed	0.80	0.76	0.78	105
Sad	0.86	0.87	0.86	99

CONCLUSION

This research demonstrates that we have successfully applied the transfer learning method with the VGG16 architecture to classify dog emotions based on the animal's facial expressions. The effectiveness of this approach is evident from the model's impressive performance, where it achieved an accuracy of 96.72% on the training set, showcasing its ability to learn and adapt well to the training data. In addition, the model performed reasonably well on the validation set with an accuracy of 88.05%, indicating its capability to generalize to unseen data. Furthermore, when tested on the test set, the model achieved an average F1-Score of

84.30%, highlighting a balanced trade-off between precision and recall across all emotion categories.

However, despite these promising results, the significant gap between the training and validation sets' accuracy signals a potential overfitting issue. This means the model performs well on the data it was trained on but struggles to generalize to new, unseen examples. The overfitting issue is particularly concerning because the images used in training were augmented to introduce more variation, and regularization techniques, such as dropout or L2 regularization, were applied to the additional layers of the model to mitigate this problem. Despite these efforts, the model still appears to have memorized specific features from the training data, which reduced its performance on the validation set.

REFERENCES

Ahlawat, S., & Choudhary, A. (2020). Hybrid CNN-SVM Classifier for Handwritten Digit Recognition. *Procedia Computer Science*, *167*, 2554–2560. https://doi.org/https://doi.org/10.1016/j.procs.2020.03.309

Ahmed, F., Bari, A. S. M. H., & Gavrilova, M. L. (2020). Emotion Recognition from Body Movement. *IEEE Access*, *8*, 11761–11781. https://doi.org/10.1109/ACCESS.2019.2963113

Alswaidan, N., & Menai, M. E. B. (2020). A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, *62*(8), 2937–2987. https://doi.org/10.1007/s10115-020-01449-0

Amici, F., Waterman, J., Kellermann, C. M., Karimullah, K., & Bräuer, J. (2019). The ability to recognize dog emotions depends on the cultural milieu in which we grow up. *Scientific Reports*, *9*(1), 16414. https://doi.org/10.1038/s41598-019-52938-4

Broomé, S., Feighelstein, M., Zamansky, A., Carreira Lencioni, G., Haubro Andersen, P., Pessanha, F., Mahmoud, M., Kjellström, H., & Salah, A. A. (2023). Going Deeper than Tracking: A Survey of Computer-Vision Based Recognition of Animal Pain and Emotions. *International Journal of Computer Vision*, *131*(2), 572–590. https://doi.org/10.1007/s11263-022-01716-3

Burza, L. B., Bloom, T., Trindade, P. H. E., Friedman, H., & Otta, E. (2022). Reading emotions in Dogs' eyes and Dogs' faces. *Behavioural Processes*, *202*, 104752. https://doi.org/https://doi.org/10.1016/j.beproc.2022.104752

Chen, H.-Y., Lin, C.-H., Lai, J.-W., & Chan, Y.-K. (2023). Convolutional Neural Network-Based Automated System for Dog Tracking and Emotion Recognition in Video Surveillance. *Applied Sciences*, *13*(7). https://doi.org/10.3390/app13074596

Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. In *Sensors (Switzerland)* (Vol. 20, Issue 3). MDPI AG. https://doi.org/10.3390/s20030592

Ferguson, M., ak, R., Lee, Y.-T., & Law, K. (2017). *Automatic localization of casting defects with convolutional neural networks*. https://doi.org/10.1109/BigData.2017.8258115

Hansen, M. F., Baxter, E. M., Rutherford, K. M. D., Futro, A., Smith, M. L., & Smith, L. N. (2021). Towards facial expression recognition for on-farm welfare assessment in pigs. *Agriculture (Switzerland)*, *11*(9). https://doi.org/10.3390/agriculture11090847

Hantke, S., Cummins, N., & Schuller, B. (2018). What is my Dog Trying to Tell Me? the Automatic Recognition of the Context and Perceived Emotion of Dog Barks. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5134–5138. https://doi.org/10.1109/ICASSP.2018.8461757

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors (Switzerland)*, *18*(2). https://doi.org/10.3390/s18020401

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Licia Sbattella Co-Supervisor, prof, Roberto Tedesco, I., Cottone, F., & Tsegay Beyene Matr, S. (2018). *Title: Emotion Recognition from Video Using Transfer Learning and Stacking*.

Manzoor, A., Ahmad, W., Ehatisham-ul-Haq, M., Hannan, A., Asif Khan, M., Usman Ashraf, M., Alghamdi, A. M., & Alfakeeh, A. S. (2020). Inferring emotion tags from object images using convolutional neural network. *Applied Sciences (Switzerland)*, *10*(15). https://doi.org/10.3390/APP10155333

Neethirajan, S., Reimert, I., & Kemp, B. (2021). Measuring farm animal emotions—sensor-based approaches. In *Sensors (Switzerland)* (Vol. 21, Issue 2, pp. 1–23). MDPI AG. https://doi.org/10.3390/s21020553

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). *Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks*.

Piana, S., Staglianò, A., Odone, F., & Camurri, A. (2016). Adaptive Body Gesture Representation for Automatic Emotion Recognition. *ACM Trans. Interact. Intell. Syst.*, *6*(1). https://doi.org/10.1145/2818740

Riyadi, A. S., Wardhani, I. P., Widayati, D. S., & Kunci, K. (2021). Klasifikasi Citra Anjing dan Kucing Menggunakan Metode Convolutional Neural Network (CNN). *Universitas Gunadarma Jl. Margonda Raya No*, *5*(1), 12140.

Sabri, A. (2023). Transfer Learning Model CNN Pralatih Untuk Klasifikasi Bunga Iris Berbasis Citra. In *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK)* (Vol. 7, Issue 1).

Sudianto, S. (2022). Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi. *Building of Informatics, Technology and Science (BITS)*, *4*(2). https://doi.org/10.47065/bits.v4i2.2261

Tammina, S. (2019a). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), p9420. https://doi.org/10.29322/ijsrp.9.10.2019.p9420

Tammina, S. (2019b). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), p9420. https://doi.org/10.29322/ijsrp.9.10.2019.p9420

Totakura, V., Hussan, M. I. T., Janmanchi, K., & Rajesh, D. (2020). Prediction of Animal Vocal Emotions Using Convolutional Neural Network. *Article in International Journal of Scientific & Technology Research*, *9*, 2. www.ijstr.org

Lee, D., 동 수이, 구 만박, Park, G., & 약요. (2022). *이동수 외1 인: 반려견 자동 품종 분류를 위한 전이학습 효과 분석 133 (Dongsu Lee et al.: Analysis of Transfer Learning Effect for Automatic Dog Breed Classification) 반려견 자동 품종 분류를 위한 전이학습 효과 분석 Analysis of Transfer Learning Effect for Automatic Dog Breed Classification*. *27*(1). https://doi.org/10.5909/JBE.2022.27.1.133