engineering proceedings

MDPI

Proceeding Paper

Dog Activity Recognition Using Convolutional Neural Network †

Evenizer Nolasco, Jr., Anton Caesar Aldea and Jocelyn Villaverde *

School of Electrical, Electronics, and Computer Engineering, Mapua University, Manila 1002, Philippines; etnolasco@mymail.mapua.edu.ph (E.N.J.); acoaldea@mymail.mapua.edu.ph (A.C.A.)

* Correspondence: jfvillaverde@mapua.edu.ph

† Presented at the 2024 IEEE 6th Eurasia Conference on IoT, Communication and Engineering, Yunlin, Taiwan, 15–17 November 2024.

Abstract: We classified common dog activities, such as sitting, standing, and lying down, which are crucial for monitoring the well-being of pets. To create a new model, we used convolutional neural networks (CNNs) on a Raspberry Pi platform and the InceptionV3 model, optimized on a dataset of Siberian Husky photos. The accuracy was 88% on a test set of 50 samples. In the developed model, TensorFlow Keras was used, while the OpenCV library was also used for system interaction with the Raspberry Pi and its Camera module. The model was effective for the image classification of dog behaviors in various environmental circumstances. The model substantially contributes to the development of pet welfare monitoring systems and improves the care for beloved animal companions.

Keywords: dog activity; image classification; CNN; InceptionV3; Raspberry Pi

1. Introduction

Dogs are incredibly popular due to their loyalty and affectionate nature. They have been companions to humans for a long time and are often considered part of the family [1]. The number of studies on dog behavior has significantly increased over the past ten to twenty years [2]. As we interact with dogs, we naturally interpret their actions as part of our communication with them. Understanding their behaviors forms the basis of how owners treat their pets. Misinterpreting their actions and having incorrect expectations of them can lead to a fear of dogs and cause inappropriate bite injuries, particularly in children [3].

Monitoring pets can be performed in various ways. One common method is using an electronic collar, which tracks the pet's location. However, this method has limitations as it only works within a certain range and does not provide information on the pet's activity or health. The other option is to use a dog monitoring system placed in your home and yard. These devices send notifications when something happens to pets.

Recently, convolutional neural networks (CNNs) have been frequently used to study human behavior. CNNs are also used for pet care. CNNs are a common architecture in deep learning, particularly for image-based detection systems [4]. The main goal of CNNs is to reduce the number of parameters compared to traditional neural networks. This prevents overfitting and improves algorithm speed. In ref. [5], pet activities were detected using accelerometer data, a random forest classifier (RF), a CNN, and a hybrid CNN [5]. The study involved six different dogs performing various activities. Wang et al. utilized CNN models including Alex Krizhevsky's network (AlexNet) and long short-term memory (LSTM) to detect and classify human behaviors [6]. One of the easiest ways to implement CNN models is to use transfer learning. Transfer learning is used to carry the weights

check for updates

Academic Editors: Teen-Hang Meen, Chi-Ting Ho and Cheng-Fu Yang

Published: 30 April 2025

Citation: Nolasco, E., Jr.; Aldea, A.C.; Villaverde, J. Dog Activity Recognition Using Convolutional Neural Network. *Eng. Proc.* **2025**, *92*, 41. https://doi.org/10.3390/engproc2025092041

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

and features of a pre-trained model to a new model to avoid having to create everything from scratch. Then, the time needed to develop and train the model can be shortened [7,8]. These pre-trained models include residual networks (ResNet)-50, visual geometry group (VGG)-16, and InceptionV3. The TensorFlow and Keras libraries are frequently used for training these models [9]. Many researchers implement CNN models with Raspberry Pi, its cameral module, and computer vision. The Raspberry Pi microcontroller is used for image processing due to its flexibility and compatibility [10]. CNN models are widely used in computer vision applications for image processing [11]. The most utilized computer vision library is OpenCV, which is implemented using Python 3 for image processing [12].

Wearable sensors are used in monitoring dog behaviors, while CNN-based image recognition is used for human behavior. In monitoring dog activities, the InceptionV3 algorithm has not been used to recognize the activities of dogs. Therefore, we introduced image processing and a CNN to classify dog activities. To develop a system to detect dog activities, we developed a device to capture an image of a dog and classify its activity at the time of capture. We implemented InceptionV3 as the CNN model for training the dataset and classifying and recognizing images. Finally, a confusion matrix was established to determine the accuracy of the image recognition system.

The developed model in this study benefits animal welfare by understanding dogs' behaviors and providing information. The research of this study serves as a basis for future studies about understanding dog behavior using image processing.

2. Methodology

In this study, three dog activities of sitting, lying, and standing were monitored. The system consisted of Raspberry Pi IV, Raspberry Pi Camera, and LCD. The model's architecture was InceptionV3, which focuses on detection, recognition, and adversarial training. A confusion matrix was used to assess the model performance. The dog used in this study was a Siberian husky. A total of 150 images per activity of the dog were collected. The framework of the study is presented in Figure 1.

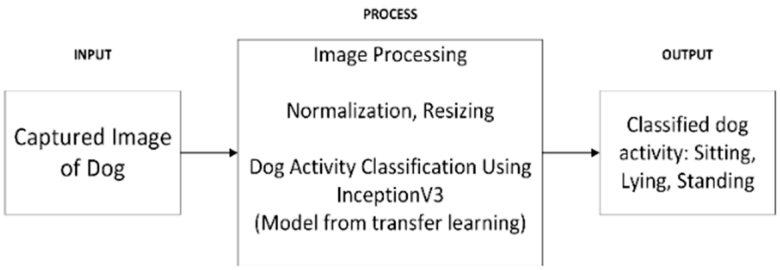


Figure 1. Framework of this research.

Raspberry Pi 4 was used as the microcontroller for processing input images, with the Raspberry Pi Camera integrated to capture images of dog activities. Captured images were preprocessed to fit the model's data requirements. The dog images were resized to 299 × 299 pixels. Then, the pixel values were normalized between 0 and 1 to enhance model performance, stability, and consistency for prediction. The CNN model implemented with transfer learning analyzed the input images and extracted their features for processing. Upon completion, the model output the recognized dog activity based on the features. The model ensured accurate and reliable detection of various dog activities for effective monitoring and analysis.

We designed the device for monitoring using Raspberry Pi parts, a power source to supply 5.1 V and 3 A, and a USB-C [13]. An SD card with a 32 GB capacity was used to run the Raspberry Pi OS and the image classification program. The CNN model

trained on InceptionV3 was programmed into the Raspberry Pi for image classification and processing. The Raspberry Pi camera module was used to capture images of the dog with its 8-megapixel camera [14]. The status of the dog's activity was displayed on the connected 3.5-inch LCD (Figure 2).

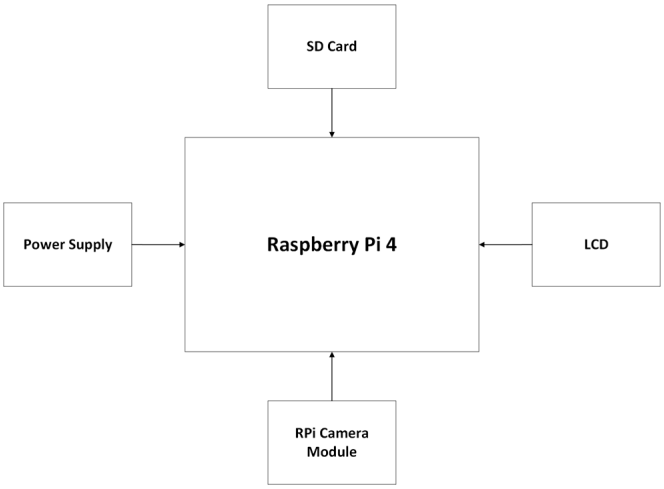


Figure 2. Hardware block diagram.

Figure 3 illustrates the system operation flowchart, beginning with image data input captured with a Raspberry Pi camera. When the system receives an image of the dog, it preprocessed the image, including resizing and normalization. Next, the image was recognized by InceptionV3, followed by classification of the dog's activity. To implement the image classification, the OpenCV library was used. The processed output was displayed on the LCD. After displaying the classified image, another image was captured to recognize and classify images.

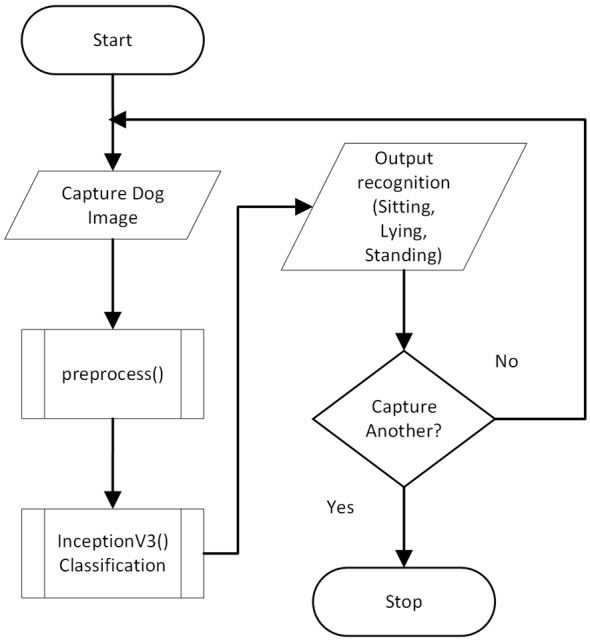


Figure 3. System operation flow.

InceptionV3 is an inception model that combines convolutional filters of various sizes to create a new filter [15]. This design reduces computational complexity by reducing the number of parameters that need to be trained. The InceptionV3 model consists of

convolutional, pooling, and fully connected layers, among others. It uses convolutional filters of different sizes (1 × 1, 3 × 3, and 5 × 5) to gather features at various scales [16]. Larger filters capture more complex features, while 1 × 1 filters reduce the dimensionality of the input data.

In this study, InceptionV3 was implemented through transfer learning by loading the pre-trained model from the TensorFlow Keras library. This model was trained on the ImageNet dataset to learn a variety of visual attributes [17]. In model training, the pre-trained layers were frozen to preserve their learned representations and determine their weights. Transfer learning kept the generalized information from the original dataset while including custom layers for classification. The model was modified to a new task to classify the images of Huskies.

Different dog sizes, shapes, and postures were captured by the model for differentiating between different activities. Minor variations in the dog's body position and orientation that distinguish each activity were examined using the extracted features. A common characteristic of dog standing is that their front and back legs are straight. Its vertical body shape and higher head positions were compared with those of its standing and lying positions to identify the dog's posture.

For the experiment, an LCD monitor and Raspberry Pi camera were connected to the Raspberry Pi 4. The devices were positioned facing the dog to classify its activity. The camera module was set to have a fixed-focus lens with a focal length of 6 mm and an aperture of f/2 and was positioned 3.5 feet from the dog. The recognition results were displayed on the LCD monitor. The model program in Raspberry Pi OS was stored on an SD card integrated into Raspberry Pi 4 (Figure 4).

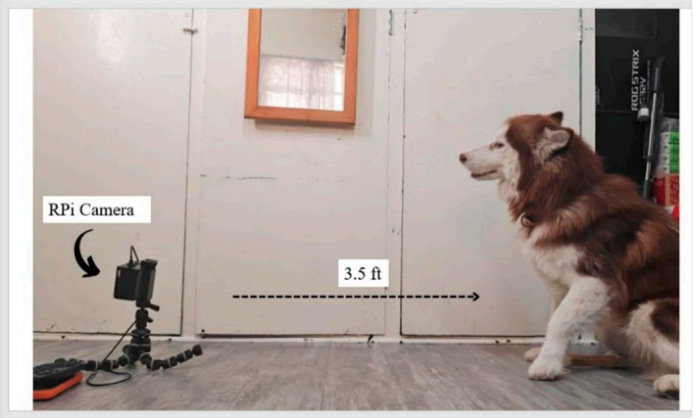


Figure 4. Device setup.

A total of 150 images per activity were collected for model training, with 80% allocated for training and 20% for validation. The dataset consisted of a combination of internet-based pictures showing specific dog activities and real-life images. The model was uploaded to the microprocessor to classify the dog's activity. For testing, 50 images were captured and classified to determine the model's accuracy.

3. Results and Discussion

Figure 5 displays the graphical user interface (GUI) of the system, which was created using the Tkinter Python library. The GUI included a live camera feed and featured two buttons, "Capture Photo" and "Capture Another." When the "Capture Photo" button was pressed, the system captured the current image from the camera. The "Capture Another" button was used to reset the system and capture and classify a new image. The intuitive design of the GUI ensures ease of use and accessibility with limited technical expertise.

By incorporating real-time feedback, users can precisely control the moment of image capture, enhancing the accuracy of the classification process. The simplicity of the two main buttons minimizes confusion and streamlines capturing and classifying images. This setup is beneficial for applications requiring quick and repeated image analysis for efficient and uninterrupted operation.

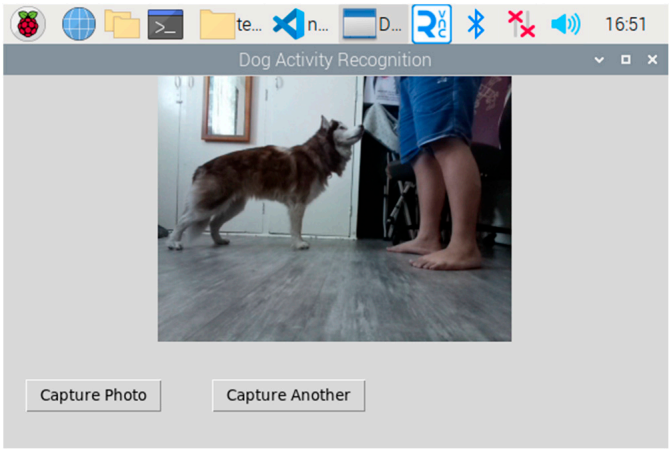


Figure 5. GUI of developed system.

When an image was captured and classified, it was identified to depict its positions. The image recognition system analyzed the visual features of the image to determine posture. The image recognition system successfully processed the image and classified postures and activities. The classification accuracy demonstrated the system's capability to distinguish various postures (Figure 6).

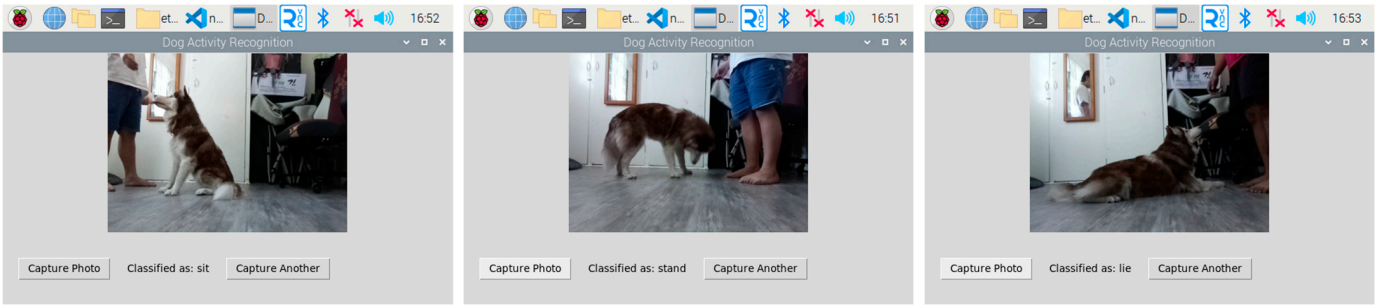


Figure 6. System outputs for sitting, standing, and lying (from up to down).

Table 1 shows the confusion matrix of system performance. From the 50 testing samples, the model predicted 15 out of 16 accurately. The model's ability to identify standing posture was validated by a high true positive rate. For sitting, the accuracy was 100%, while for lying, the accuracy was 71%. The model had difficulty in distinguishing lying postures. The accuracy was calculated using Equation (1).

$$Accuracy = \frac{\sum_{n=1}^{3} A_{nn}}{\sum_{\substack{i=1 \\ j=1}}^{3} Aij} \quad (1)$$

Equation (1) indicates the relationship between expected and actual values using a confusion matrix. The model correctly predicted 44 out of 50 samples. The overall accuracy was 88% across all postures.

Table 1. Confusion matrix.

N = 50		Predicted		
		Standing	**Sitting**	**Lying**
Actual	Standing	15	1	0
	Sitting	0	17	0
	Lying	2	3	12

4. Conclusions and Recommendations

We developed a system to capture and classify the images of dog activities. The system integrated InceptionV3 for classifying dog activities. The accuracy was determined using a confusion matrix. The overall accuracy was 88%. To improve the accuracy of the model, images of larger dogs are needed to better capture the dog's activities, particularly considering that shorter-legged dogs might confuse the system for activity detection. It is necessary to augment the training dataset to include a variety of dog activities. It is also necessary to diversify the dataset with various dog breeds to enhance accuracy. Alternative models need to be constructed for a more accurate classification of dog activities.

Author Contributions: Conceptualization, E.N.J., A.C.A. and J.V.; methodology, E.N.J., A.C.A. and J.V.; software, E.N.J. and A.C.A.; validation, E.N.J., A.C.A. and J.V.; writing—original draft preparation, E.N.J. and A.C.A.; writing—review and editing, E.N.J., A.C.A. and J.V.; visualization, E.N.J. and A.C.A.; supervision, J.V.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Benz-Schwarzburg, J.; Monsó, S.; Huber, L. How Dogs Perceive Humans and How Humans Should Treat Their Pet Dogs: Linking Cognition with Ethics. *Front. Psychol.* **2020**, *11*, 584037. [CrossRef] [PubMed]
2. Chaudhari, A.; Kartal, T.; Brill, G.; Amano, K.J.; Lagayan, M.G.; Jorca, D. Dog ecology and demographics in several areas in the Philippines and its application to anti-rabies vaccination programs. *Animals* **2022**, *12*, 105. [CrossRef] [PubMed]
3. Meints, K.; Racca, A.; Hickey, N. How to prevent dog bite injuries? Children misinterpret dogs' facial expressions. *Inj. Prev.* **2010**, *16* (Suppl. 1), A68. [CrossRef]
4. Hussain, A.; Ali, S.; Abdullah; Kim, H.-C. Activity Detection for the Wellbeing of Dogs Using Wearable Sensors Based on Deep Learning. *IEEE Access* **2022**, *10*, 53153–53163. [CrossRef]
5. Eerdekens, A.; Callaert, A.; Deruyck, M.; Martens, L.; Joseph, W. Dog's Behaviour Classification Based on Wearable Sensor Accelerometer Data. In Proceedings of the 5th Conference on Cloud and Internet of Things, CIoT, Marrakech, Morocco, 28–30 March 2022; pp. 226–231. [CrossRef]
6. Wang, S.; Gao, J.Z.; Lin, H.; Shitole, M.; Reza, L.; Zhou, S. Dynamic human behavior pattern detection and classification. In Proceedings of the 5th IEEE International Conference on Big Data Service and Applications, BigDataService, Newark, CA, USA, 4–9 April 2019; pp. 159–166. [CrossRef]
7. Pangilinan, J.R.; Legaspi, J.; Linsangan, N. InceptionV3, ResNet50, and VGG19 Performance Comparison on Tomato Ripeness Classification. In Proceedings of the 2022 5th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI, Yogyakarta, Indonesia, 8–9 December 2022; pp. 619–624. [CrossRef]
8. Juy, J.N.; Villaverde, J.F. A Durian Variety Identifier Using Canny Edge and CNN. In Proceedings of the 2021 7th International Conference on Control Science and Systems Engineering, ICCSSE, Qingdao, China, 30 July–1 August 2021; pp. 293–297. [CrossRef]

9. Navarro, L.K.B.; Mateo, K.C.H.; Manlises, C.O. CNN Models for Identification of Macro-Nutrient Deficiency in Onion Leaves (*Allium cepa* L.). In Proceedings of the 2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering, ECICE, Yunlin, Taiwan, 27–29 October 2023; pp. 396–400. [CrossRef]
10. Buenconsejo, L.T.; Linsangan, N.B. Detection and Identification of Abaca Diseases using a Convolutional Neural Network CN. In Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Auckland, New Zealand, 7–10 December 2021; pp. 94–98. [CrossRef]
11. Soliman-Cuevas, H.L.; Linsangan, N.B. Day-Old Chick Sexing Using Convolutional Neural Network (CNN) and Computer Vision. In Proceedings of the 5th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET, Kota Kinabalu, Malaysia, 12–14 September 2023; pp. 45–49. [CrossRef]
12. Fitz Bumacod, D.S.; Delfin, J.V.; Linsangan, N.; Angelia, R.E. Image-Processing-based Digital Goniometer using OpenCV. In Proceedings of the 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM, Manila, Philippines, 3–7 December 2020. [CrossRef]
13. Muhali, A.S.; Linsangan, N.B. Classification of Lanzones Tree Leaf Diseases Using Image Processing Technology and a Convolutional Neural Network (CNN). In Proceedings of the 4th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET, Kota Kinabalu, Malaysia, 13–15 September 2022. [CrossRef]
14. Sanchez, R.B.; Esteves, J.A.C.; Linsangan, N.B. Determination of Sugar Apple Ripeness via Image Processing Using Convolutional Neural Network. In Proceedings of the 2023 15th International Conference on Computer and Automation Engineering, ICCAE, Sydney, Australia, 3–5 March 2023; pp. 333–337. [CrossRef]
15. Caya, M.V.C.; Caringal, M.E.C.; Manuel, K.A.C. Tongue Biometrics Extraction Based on YOLO Algorithm and CNN Inception. In Proceedings of the 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM, Manila, Philippines, 28–30 November 2021. [CrossRef]
16. De Goma, J.C.; Divina, F.G.; Isaac, M.K.B.; Pajaro, R.J. Effectiveness of Using Fundus Image Data Containing Other Retinal Diseases in Identifying Age-Related Macular Degeneration using Image Classification. In Proceedings of the 2023 13th International Conference on Software Technology and Engineering, ICSTE, Osaka, Japan, 27–29 October 2023; pp. 113–117. [CrossRef]
17. Legaspi, J.; Pangilinan, J.R.; Linsangan, N. Tomato Ripeness and Size Classification Using Image Processing. In Proceedings of the 2022 5th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI, Yogyakarta, Indonesia, 8–9 December 2022; pp. 613–618. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.