



Article

Predicting Dog Emotions Based on Posture Analysis Using DeepLabCut

Kim Ferres¹, Timo Schloesser¹ and Peter A. Gloor^{2,*}

¹ Department of Information Systems and Information Management, University of Cologne, Pohligstrasse 1, 50969 Cologne, Germany; kferres@smail.uni-koeln.de (K.F.); timo.schloesser@gmail.com (T.S.)

² MIT Center for Collective Intelligence, 245 First Street, Cambridge, MA 02142, USA

* Correspondence: pglloor@mit.edu

Abstract: This paper describes an emotion recognition system for dogs automatically identifying the emotions anger, fear, happiness, and relaxation. It is based on a previously trained machine learning model, which uses automatic pose estimation to differentiate emotional states of canines. Towards that goal, we have compiled a picture library with full body dog pictures featuring 400 images with 100 samples each for the states “Anger”, “Fear”, “Happiness” and “Relaxation”. A new dog keypoint detection model was built using the framework DeepLabCut for animal keypoint detector training. The newly trained detector learned from a total of 13,809 annotated dog images and possesses the capability to estimate the coordinates of 24 different dog body part keypoints. Our application is able to determine a dog’s emotional state visually with an accuracy between 60% and 70%, exceeding human capability to recognize dog emotions.

Keywords: neural networks; machine learning; dog emotion detection; dog emotion



Citation: Ferres, K.; Schloesser, T.; Gloor, P.A. Predicting Dog Emotions Based on Posture Analysis Using DeepLabCut. *Future Internet* **2022**, *14*, 97. <https://doi.org/10.3390/fi14040097>

Academic Editor: Kaushik Roy, Mustafa Atay and Ajita Rattani

Received: 31 January 2022

Accepted: 18 March 2022

Published: 22 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

1. Introduction

Dogs play an important social role in western society and can assist with various tasks. Service dogs assist visually impaired people, patients fighting anxiety issues, and mentally or physically disabled people. Law enforcement uses dogs in several fields, for example, officer protection, track search, or drug inspection. The field of medicine also benefits from dogs since their superior sense of smell can perceive illnesses such as Parkinson’s disease [1].

Along with many other possible uses, all share the need for extensive training adapted to the dog. Even for a pet dog, training and education are usually required to live along with people. In order to train dogs optimally, it is relevant to possess the ability to read and interpret their body language correctly. Hasegawa et al. [2] showed that a dog’s learning performance is best when it displays an attentive body posture and a positive mood. Thus, assessing a dog’s emotions is essential for resource-efficient, successful training.

According to a study by Amici et al. [3], the probability that an adult human of European origin, regardless of his or her experience with dogs, will correctly analyze the emotional state of a dog based on its facial expression is just over 50%. Given this background, a way of reliably and objectively assessing dogs’ emotions seems quite desirable to improve individual training and to enable new ways of utilizing dogs’ abilities. There is also the opportunity of applying the software for educational purposes to enhance peoples’ emotional understanding of dogs. This kind of knowledge has been proven to be of significant impact when it comes to reducing dog abuse [4]. Computer-aided emotion recognition ensures dogs’ well-being across distances or numerous animals simultaneously without high manual effort. Furthermore, dog emotion detection can assist users who have little experience in dealing with dogs to recognize angry or aggressive dogs early, and thus evade a potential attack. Another application area would be veterinary medicine, which could initiate diagnostics earlier if a dog has an increased perception of dissatisfaction.

A variety of scientific work related to dog emotions has already been published. Many analyze the psychological and social aspects of dogs' cohabitation with humans. A group of international researchers developed the Dog Facial Action Coding System (DogFACS), which maps certain canine facial micro-expressions to different emotions [5]. Considering studies that deal with automatic recognition of canine emotions, the University of Exeter implemented an emotion recognition system relying on facial recognition and the previously mentioned DogFACS [6]. Franzoni et al. [7] applied machine learning techniques on emotion image data sets to create a dog emotion classifier relying on facial analysis.

Furthermore, scientific work exists to analyze dog body language by machine learning; Brugarolas et al. [8] developed a behavior recognition application that needs data from a wireless sensor worn by the dog. Aich et al. [9] designed and programmed a canine emotion recognizer based on sensors. In this case, the sensors are worn around the dog's neck and tail, and their data output was fed into an artificial neural network in the training phase. Another relevant scientific contribution was provided by Tsai et al. [10], who developed a dog surveillance tool. The authors combined machine learning techniques for image recognition on continuous image and bark analysis to compute a dog's mood differentiating the emotions happiness, anger and sadness as well as a neutral condition from each other. Maskeliunas et al. [11] analyzed emotions of dogs from their vocalizations, using cochleagrams to categorize dog barking into the categories angry, crying, happy, and lonely. Raman et al. [12] used CNNs and DeepLabCut to automatically recognize dog body parts and dog poses.

2. Chosen Approach

This work aims to create a dog emotion detector by applying machine learning techniques to photographically recorded body postures of dogs. Thereby, landmark points, such as the tail's position, will be extracted first and then used for model training. In contrast to other research in image-based dog emotion recognition, this work should be able to analyze the entire image of the dog and detect subtle differences in the position of the animal. In contrast to work focusing on facial emotions, this software artifact should work more reliably in practice since the dog's head does not have to be constantly aligned with the camera. Especially with moving dogs, this will result in a more practical detector.

At the beginning of this work, possible target values of dog emotions need to be specified. The literature is mainly agreeing on the existence of some basic emotions: Most prominent an emotion appointed with happiness, joy, playfulness, contentment or excitement, further only referred to as happiness, is named. The opposing emotion sadness, distress or frustration is also often listed. Other basic emotions that have been identified to apply to dogs are anger, fear, surprise and disgust [3,13–15]. The findings of the psychologist Stanley Coren [14] should be emphasized here, who, based on the knowledge that the quality of an average dog's intelligence resembles a human child of 2.5 years, concludes that the emotional maturity should also be similar. Therefore, Coren includes the sentiments of affection, love and suspicion to his specification of canine emotions.

For this work, the emotions anger, fear and happiness are taken into account, since the existence of these canine sentiments is accepted in the majority of dog-related research and these sentiments are possible to determine by posture analysis. Although sadness, surprise and disgust are frequently mentioned emotions, they will not be examined here since they are mostly not recognizable through posture observation but through facial analysis [13]. In addition, a state of mind titled "relaxation" is considered for the automatic emotion recognition system to prevent the over-interpretation of only minor physical signs by the algorithm.

The software development is structured as follows (compare Figure 1):

1. Collecting Dog Emotion Images;
2. Image Preprocessing;
 - (a) Detect area(s) containing a dog;
 - (b) Resize image and adjust view direction of the dog;

3. Landmark Detection Model;
4. Dog Emotion Detection Model:
 - (a) Neural Network using all landmark data;
 - (b) Decision Tree with few calculated pose metrics.

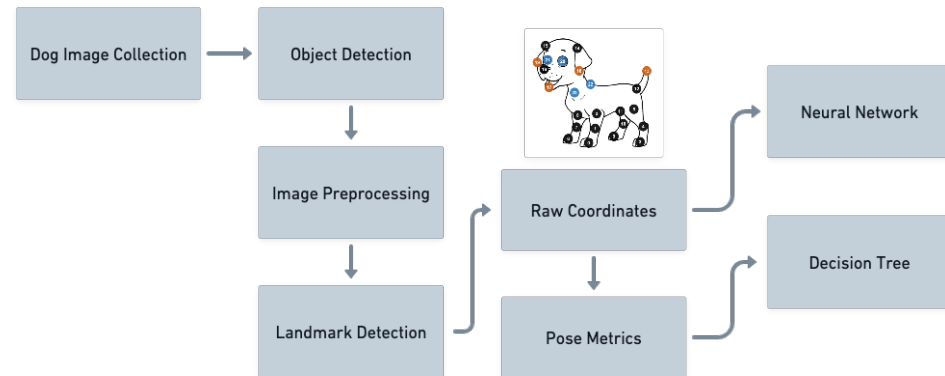


Figure 1. Workflow diagram of the proposed approach.

As a first step, a suitable dataset of dog images reflecting the emotions anger, fear, happiness and relaxation has to be selected. Since there is no appropriate dog emotion data set consisting of labeled full-body images publicly available, photographs showing the entire body displaying one of the selected emotions must be collected. Pictures that depict a puppy are excluded, as they are probably not yet emotionally mature and their physical markers, such as raised ears, might not yet be developed. Furthermore, dogs showing signs of docking, such as cropped ears or tails, are excluded because their emotional communication capabilities are restricted [16]. The actual collection of images is done by a manual search on online video and image platforms. In this case, stimuli describe the circumstances in which a photograph was taken. When it comes to canine emotion studies, the usage of stimuli is a common tactic to assure that a certain emotion is shown in an image. Combining the insights and procedures of some of those kinds of studies, the following mapping of the selected emotions and stimuli is possible:

After establishing the target emotions, several search engines were then queried using various search terms describing situations where the above stimuli would occur. The pictures have been labeled by the first author, verifying and checking the labels initially given by the person uploading a picture or video on the Web. A subset of the images has been verified and relabeled by an additional team of students. The resulting images were subsequently filtered and standardized in terms of size and quality. An object detector was then applied to all images identifying the area where the dog is most likely located. Cropping the images to this area ensures optimal input for landmark detection and prevents potential detection errors.

A landmark detector is then used to capture as many relevant reference points of the dog images as possible. The objective of landmark detection algorithms is to be robust to changes in perspective, changes in size, and motion of the object being observed [17]. From the detected points, the pose or essential aspects of the pose can be computed by relating the points. In particular, coordinates for the animal's joints are necessary to reconstruct its pose as accurately as possible. For this purpose, the landmark detector *DeepLabCut* by Mathis et al. [18] was used, which was previously trained with the *Stanford Extra Data Set* by Biggs et al. [19] and the *Animal Pose Data Set* by Cao et al. [20].

Finally, the extracted landmark points are used to train two models to detect dog emotions. First, a neural network that uses raw coordinates from the landmark detector as training data and second, a decision tree that was trained using calculated pose metrics:

1. **Tail Position:** Angle between tail and spine;
2. **Head Position:** Elevation degree of head;

3. **Ear Position:** Angle of ear in relation to line of sight;
4. **Mouth Condition:** Opening degree of mouth;
5. **Front Leg Condition:** Bending degree of front leg(s);
6. **Back Leg Condition:** Bending degree of back leg(s);
7. **Body Weight Distribution:** Gradient degree of spine.

3. Materials and Methods

3.1. Dog Emotion Data Set

Since there is no publicly available image data set that matches dog images to emotions, various search engines were used to collect images. In order to get a target emotion label for each image, the stimuli described in Table 1 were mapped to search terms as shown in Table 2:

Table 1. Mapping of target emotions to different stimuli [3,13,21–23].

Emotion	Stimuli
Anger	Combination of excitement and an attack Confrontation with an inaccessible resource
Fear	Encounter with a stronger conspecific partner Trimming of toenails Visiting the veterinarian Perception of an active thunderstorm
Happiness	Interaction with a trusted partner, such as the owner Playing (e.g., ball play) Presentation of food
Relaxation	Absence of any events

Table 2. Mapping of target emotions to search terms.

Emotion	Search Terms
Anger	dog attack dog attacking dog dominant dog territory protection dog territorial
Fear	dog thunderstorm dog scared vet dog fear vet dog anxious vet dog scared toenail trim dog fear toenail trim dog anxious toenail trim dog scared submissive dog fear submissive dog anxious submissive
Happiness	dog play dog with toy dog play bow dog plays with owner dog and owner dog with owner dog gets food dog gets treat dog food
Relaxation	dog standing dog waiting

Based on extensive interactive and exploratory Google searches for the most promising language spaces and image websites containing the most dog pictures, these search terms were then translated into Russian, German, Italian, French, Spanish and Chinese, and used for image searches on [bing.com](https://www.bing.com), [depositphotos.com](https://www.depositphotos.com), [duckduckgo.com](https://www.duckduckgo.com), [flickr.com](https://www.flickr.com), [google.de](https://www.google.de), [pexels.com](https://www.pexels.com), [pinterest.de](https://www.pinterest.de), [yahoo.com](https://www.yahoo.com), [youtube.com](https://www.youtube.com) and [yandex.ru](https://www.yandex.ru) (30 January 2022).

3.2. Landmark Detector

Mathis et al. [18] published a keypoint detection framework called *DeepLabCut* (DLC), which is based on a deep learning approach and can determine animal body keypoints after an initial training phase with labeled example images. The authors state that the system allows successful predictions starting from a training dataset of 200 images. Models trained with the DLC framework can provide a probability value for their keypoint predictions. If a keypoint is occluded, i.e., obscured by the dog itself or another object, the detector will attempt to approximate its position, but generally return a lower probability value for the estimated coordinates.

Since DLC works best when the region of interest, i.e., the animal's body parts of interest, is "picked to be as small as possible" [18]. Consequently, object detection is applied to all images to obtain the coordinates of a bounding box containing the dog. If the area containing the dog can be identified, the image is cropped accordingly. For this task the software library *ImageAI* from Olafenwa & Olafenwa [24] is used. Their library can detect 1000 different objects via image recognition, including dogs.

Two datasets were selected as training data for the custom DLC landmark detector:

1. The *Stanford Extra Data Set* builds upon the Stanford Dogs Data Set composed by Khosla et al. [25] that represents a collection of 20,580 dog images of 120 different breeds. Biggs et al. [19] performed annotations on 12,000 images of the set and provided them as a JSON file. Their annotations consist of coordinates for a bounding box enclosing the depicted dog, coordinates for keypoints and segmentation information describing the dog's shape.
2. The *Animal Pose Data Set* was assembled by Cao et al. [20] as part of a paper on animal pose estimation and covers annotations and images for dogs, cats, cows, horses and sheep. In total 1809 examples of dogs are contained in the data set with some images containing more than one animal. Similar to Stanford Extra, the Animal Pose Data Set contains annotations for 20 keypoints in total.

Figure 2 shows the available keypoints of both datasets. This combination results in a total of 23 detectable landmark points, which are later used for model training.

3.3. Dog Emotion Detector

In this work, two different approaches for training a dog emotion recognition model are explored:

1. Coordinate data from landmark detection are used directly as the training input for a neural network. This approach has the advantage of providing all available data to the model.
2. Various pose metrics (e.g., the weight distribution) are calculated by relating multiple landmarks. By using this less complex training data for a basic classification algorithm (e.g., a decision tree), counterfactual explanation techniques can be used to explore further how dogs express emotions.

For the model development, the widely used python library *sklearn* is used.

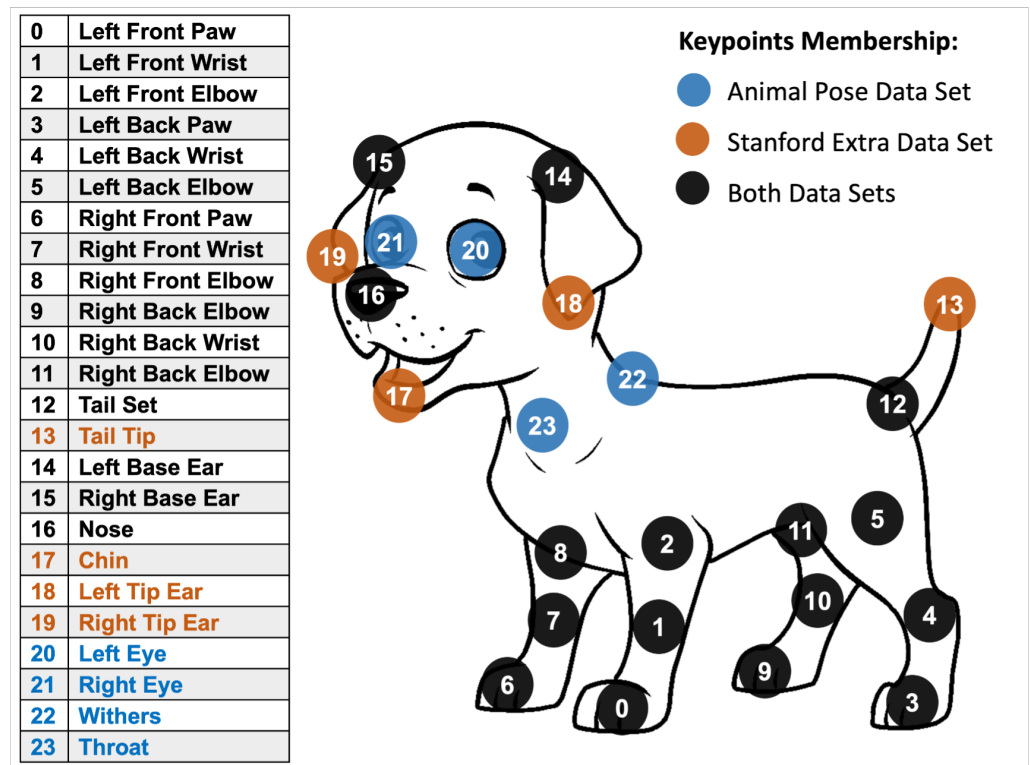


Figure 2. Available keypoints for landmark detection datasets. The colors indicate which data set contains the respective point.

3.3.1. Neural Network

Since the object detector delivers images of different sizes, preprocessing the landmark points is necessary. Comparability between images is achieved by setting all coordinates in relation to the size of the cropped image. In addition, images are mirrored based on the arrangement of the landmark points to ensure that all dogs are facing the same direction.

The emotion classes Anger, Fear, Happiness, and Relaxation were one-hot encoded. Thus, a binary field is added to all data points for each class. In contrast to integer encoding, this prevents the classifier from exploiting the natural order of these numbers in its predictions.

Before the final model training, a grid search with cross-validation as explained in Section 4.3.1 is performed to determine the most appropriate hyperparameters.

3.3.2. Decision Tree

To develop appropriate metrics for predicting a dog’s emotion, various base postures are first broken down into the condition of individual body parts.

For this purpose, seven body postures of a dog [26,27] are considered:

1. A **Neutral** position expresses that a dog is relaxed and approachable since he feels unconcerned about his environment.
2. An **alarmed** dog detected something in his surroundings and is in an aroused or attentive condition. This mood is usually followed by an inspection of the regarding area to identify a potential threat or an object of interest.
3. **Dominant aggression** is shown by a dog to communicate its social dominance and that it will answer a challenge with an aggressive attack.
4. A **defensive aggressive** dog may also attack, but is motivated by fear.
5. An **active submissive** canine is fearful and worried, and shows weak signals of submission.
6. Total surrender and submission is shown by **passive submissive** behavior and indicates extreme fear in the animal.

7. The “play bow” signals that a dog is **playful**, invites others to interact with it and is mostly accompanied by a good mood.
The effects of these postures on individual body parts are listed in Table 3.

Table 3. Dog posture characteristics [26,27].

Posture	Front End	Back End	Head	Ears	Tail	Mouth
Neutral	Normal	Normal	Up	Up	Down	Open, tongue visible
Alarmed	Normal	Normal	Up	Up, forward	Horizontal	Slightly open
Dominant Aggressive	Strongly upright	Strongly upright	Up	Up, forward	Up	Open, teeth visible
Defensive Aggressive	Lowered	Strongly Lowered	Down	Down, back	Down, tucked	Closed
Active Submissive	Lowered	Lowered	Down	Down, flat, back	Down	Closed
Passive Submissive	Underbelly exposed	Down	Down, back	Down	Closed	
Playful	Strongly lowered	Normal	Moving	Up	Up	Open, tongue visible

Based on these postures, pose metrics are defined to represent the four target emotion classes. These are introduced in detail in the next chapter. After calculating the pose metrics for each input image, a grid search with cross-validation is performed in the same way as for a neural network.

3.3.3. Pose Metrics

The general idea underlying the formulas for measuring the dog’s posture is to determine various angles that are as independent as possible of breed, other characteristics of the dog, or other objects in the photograph. Figure 2 can be referred to for interpretation of the landmark numbers. Preliminary testing of the newly trained DLC detector showed that the keypoint detection performs significantly worse when the dog is not standing or sitting. Therefore, no calculations are performed to determine an exposed belly, which is a strong sign of submission in dogs.

The **body weight distribution** of a dog can be calculated using the slope of the hypothetical line between the withers (keypoint 22) and the base of the tail (keypoint 12) as shown in Figure 3. To prevent an uneven surface or the dog’s posture from affecting the slope, the angle α is calculated using the to line between the front (keypoint 0/6) and back (keypoint 3/9) paws of one side. If possible, the paw coordinates of the side facing the camera are used, as these are usually more accurate. If the intersection of the two lines considered is in front of the withers (keypoint 22), i.e., in the direction of the dog’s head, the angle between the lines is recorded as positive. In the opposite case, a negative degree is recorded, and in the case of parallelism, the value zero is recorded.

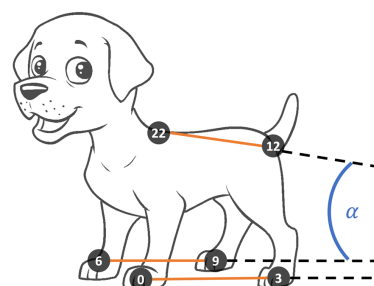


Figure 3. Calculation of the body weight distribution.

If the withers are obscured, it should be possible to measure the body weight distribution using the throat keypoint (23). For comparability of both calculations, the following linear equation was obtained to convert the throat-based angle:

$$\text{body weight distribution} = -14.1 + 0.3 \times (\text{throat-based body weight distribution}) \quad (1)$$

This equation was obtained by applying a linear model to all keypoint pairs (withers/throat).

The **tail position** can be calculated by using the angle α displayed in Figure 4 between the hypothetical line connecting the withers (keypoint 22) and the base of the tail (keypoint 12), hereafter referred to as the dorsal line, and the hypothetical line between the base of the tail (keypoint 12) and the tip of the tail (keypoint 13). A raised tail, characterized by the tip of the tail being above the extended dorsal line, is recorded as a positive value. A lowered tail, recognizable by a tail tip below the extended dorsal line, is given a negative sign. Zero is measured for the tail position if the tail tip is exactly on the extended dorsal line.

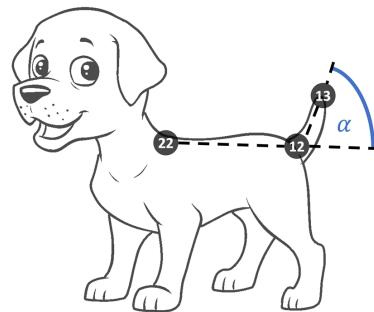


Figure 4. Calculation of the tail position.

Again, it should be possible to perform the calculation using the throat point (23) instead of the withers point (22). The following equation is used for the conversion:

$$\text{tail position} = 18.7 + 0.98 \times (\text{throat-based tail position}) \quad (2)$$

The **head position** measures the degree of elevation of a dog's head by observing the slope of the line connecting the withers (keypoint 22) and the nose (keypoint 16), as shown in Figure 5. Like the body weight distribution calculation, the angle α is calculated using the baseline between a front (keypoint 0/6) and back (keypoint 3/9) paw. A raised head is measured positively, while a lowered head indicates a negative value. In the case of hidden withers, the throat-based calculation can be converted using the following equation:

$$\text{head position} = -2.78 + 0.06 \times (\text{throat-based head position}) \quad (3)$$

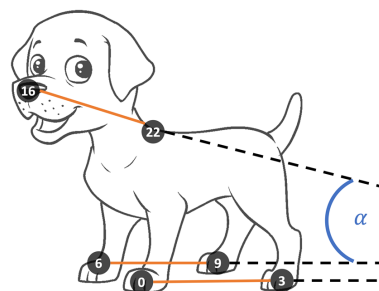


Figure 5. Calculation of the head position.

The **ear position** describes the orientation of the ears in relation to the head as described by the angle α and displayed in Figure 6. Since usually, just the keypoints of one ear are entirely available, only one ear is considered for the calculation and, if possible, the one facing the camera. This procedure is based on the hypothesis that the ears usually point in

a similar direction. If this is not the case, the used calculation method can, of course, be problematic. However, it seems the best option under the given restriction of inaccurate or not recognizable keypoint coordinates. The ear position is determined by regarding a line between the nose keypoint (16) and the ear base keypoint (14/15) in relation to the slope of the line starting at the ear base keypoint (14/15) and ending at the ear tip keypoint (18/19). A raised ear, i.e., the ear tip exceeds the extended head position line, is measured as a positive value, while a lowered ear indicates a negative value.

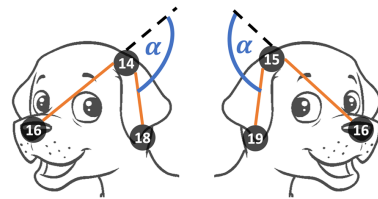


Figure 6. Calculation of the ear position.

The **mouth condition** measures the degree to which the dog’s snout is open (compare Figure 7). For this purpose, the angle α between the nose (keypoint 16), the base of the ear (keypoint 14/15) and the chin (keypoint 16) is calculated. If possible, the ear base keypoint of the dog side facing the camera is used. The calculated mouth opening angles reflect a correct tendency, but a mouth opening of zero is practically impossible since the nose and chin keypoints are used for the calculation. Thus, a slightly higher value than zero represents a closed mouth.

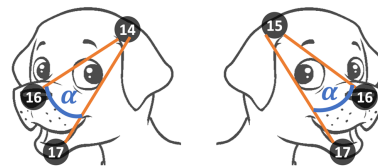


Figure 7. Calculation of the mouth condition.

The **front leg condition** represents the bending angle of one of the front legs by the angle α between a front paw keypoint (0/6), a front elbow keypoint (2/8) and the withers keypoint (22), as shown in Figure 8. This procedure results in high values being assigned to extended legs and low values to flexed legs. Again, replacing the withers keypoint (22) with the throat keypoint (23) should be possible. The conversion formula for the throat-based calculation is as follows:

$$\text{front leg condition} = 83.78 + 0.61 * (\text{throat-based front leg condition}) \quad (4)$$

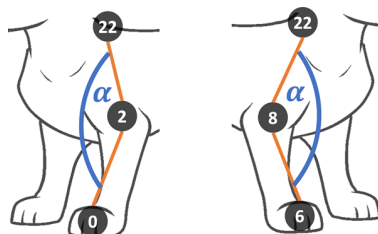


Figure 8. Calculation of the front leg condition.

Similarly, the **back leg condition** displayed in Figure 9 represents the bending angle of one of the back legs. For calculating the bending degree, the angle α between a back paw keypoint (3/9), a back elbow keypoint (5/11) and the tail set keypoint (12) are used.

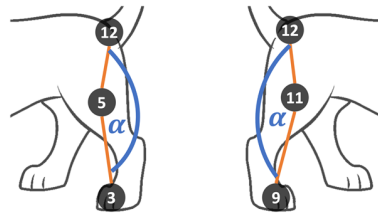


Figure 9. Calculation of the back leg condition.

4. Results

4.1. Dog Emotion Data Set

The objective was to generate a balanced data set in which all classes are equally represented. The class “fear” turned out to be the limiting factor for the total size of the dataset. This may be because such images are less likely to be published, as they may reflect poorly on the holder. Therefore, the search was stopped when no more images were found for the “fear” class. The images are split among the classes as follows:

1. Anger: 111 Images;
2. Fear: 104 Images;
3. Happiness: 109 Images;
4. Relaxation: 106 Images.

For model training, an even, easily dividable number of samples is desirable, since the data set must be split multiple times for cross-validation and final testing. Therefore, all categories were reduced to a number of 100 images each by discarding photos with the lowest probability of correct recognition, such as images with poor exposure, low resolution, blur, unfavorable perspective, or objects too close to the dog. In addition, photos in which dogs are only partially visible or puppies are visible were also excluded.

In total, the dataset contains 36 different breeds and mostly individual dogs, rather than multiple images of one animal.

4.2. Landmark Detector

After the Stanford Extra and Animal Pose datasets were merged, a neural network based on DLC could be trained. As recommended by the authors [6], training was performed for 200,000 iterations and the loss settled between 0.015 and 0.017.

Testing the trained model proved that the detector is capable of achieving very accurate detection results. In some cases, it can even locate the position of occluded body parts accurately and with high prediction probability.

Since the different body parts in the training data do not occur with equal frequency due to two different datasets, it is impossible to use only one uniform probability threshold to accept coordinates. Instead, manual tests on sample images showed that a minimum probability value of 5% is appropriate for all body parts included in the Stanford Extra Data Set (compare Figure 2). For all body parts that are exclusively part of the Animal Pose Data Set (compare Figure 2), a minimum probability value of 0.5% proved suitable. Since significantly less training data was available for these body parts, the model automatically computes a lower value to reflect the ratios of the training data.

In addition to this self-trained detector, the pre-trained detectors **DLC model zoo**, **North DLC model** [6] and **Dog body part detection model** [28] were tested. In order to provide the necessary data for dog posture analysis, the detectors were all used together since each offers a different subset of the needed landmark points. In a manual comparison of both approaches (self-trained vs. combined), the self-trained detector performed better and was selected for further work. In particular, the self-trained detector delivered better results in detecting hidden landmark points and the recognition of the tail. Furthermore, it was not possible to determine the position of the ears with the combined approach.

4.3. Dog Emotion Detector

As the main component of the Dog Emotion Detector, a Neural Network and a Decision Tree were trained and then compared in terms of their performance. Since the neural network can work with missing values, the raw coordinates of the landmark detector are used as training data. For the decision tree, the calculated pose metrics were used as input, for which the respective average filled missing values.

A stratified test set of 40 images (10%) was sampled for the final accuracy evaluation of both classifiers. The remaining 360 images were split into multiple training/validation sets during the grid search using 10-fold cross-validation.

4.3.1. Neural Network

To optimize the neural network, a grid search with typical tuning parameters according to Géron [29] is conducted: Activation function, number of layers and number of nodes. Table 4 summarizes the considered grid search values, the grid search results and all other parameters used for training.

Table 4. Parameters for neural network training.

Name	Grid Search Values	Value
Activation Function (for Hidden Layers)	Sigmoid, ReLU, Tanh	Tanh
Number of Layers	1, 2, 3, 4	1
Number of Nodes in First Hidden Layer	96, 48, 24	15
Number of Nodes in Last Hidden Layer	16, 8, 6	
Batch Size		5
Output Activation Function		Softmax
Optimizer		ADAM
Loss Function		Categorical Cross-Entropy
Epochs		200

Training the model for 200 epochs represents a medium-high number of epochs. It is chosen to get a good impression of the performance of the parameter combination from the grid search while avoiding overfitting. After determining the best parameter combination, an epoch fine-tuning was conducted to balance the network between underfitting and overfitting.

The fine-tuning showed a local minimum of model loss between 100 and 350 epochs on the test set (compare Figure 10). According to Howard and Guggen (2020, pp. 212–213) the local minimum of the loss measured on the test set for each epoch will indicate the best-suited number of epochs. Finally, a model with the best performing hyperparameters was trained on the training and validation set over 240 epochs. Evaluating this model on the test data set achieved an **accuracy of 67.5%**.

4.3.2. Decision Tree

Before training the decision tree classifier, a check was made to verify that the previously calculated pose metrics were suitable for predicting a dog's emotion. The Confusion matrix in Figure 11 shows the correlations between emotion classes and pose metrics. These values suggest that the metrics are suitable for determining emotions. For example, in Table 3, anger is assigned to "defensive aggressive" posture. Its typical signs, raised ears and tail, uplifted head, and a widely opened mouth, show as positive correlation values in the matrix.

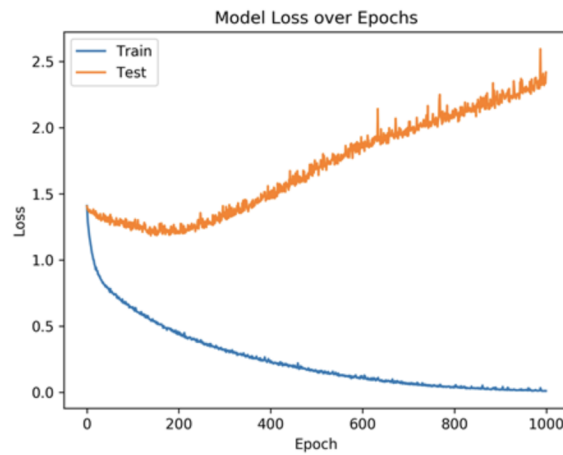


Figure 10. Neural network performance over epochs.

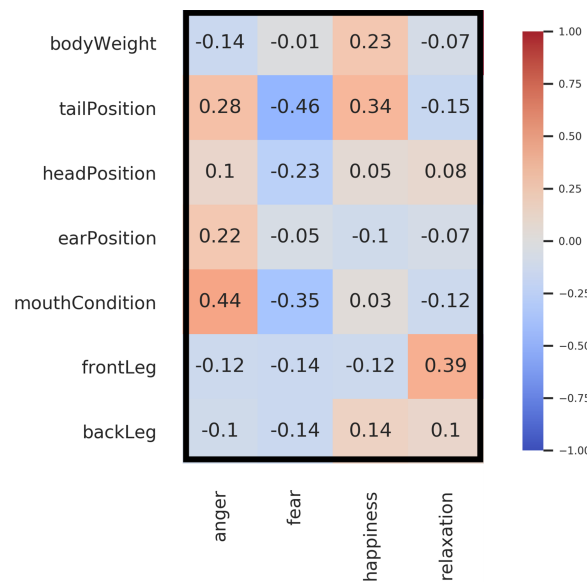


Figure 11. Correlation matrix for pose calculations.

However, there are also deviations from the characteristic features of the assigned attitudes. For example, raised ears are typically an indicator for the attitude “playful”, which is assigned to the emotion class “happiness”. This correlation is not reflected in the data since the posture “playful” might be just one of the possible expressions of happiness in dogs and the strong variation of body signals for “happiness” influences the results.

Since the decision tree classifier cannot be trained with incomplete data sets, the missing values were filled by the average value of each parameter within its emotion class. In this process, values for 32 samples in total (8% of the complete data set) were filled with 1.5 empty parameter cells per sample on average.

The grid search of the decision tree classifier resulted in a set of best hyperparameters listed in Table 5. Training the model with these parameters and then evaluating on the test set resulted in an **accuracy of 62.5%**. To increase the comprehensibility and compactness of the final model, all unnecessary splits, i.e., those where all resulting leaves lead to the prediction of the same class, were removed by pruning. Figure 12 displays an illustrated and annotated graphical representation of the decision tree generated by the learning algorithm.

Table 5. Parameters for Decision Tree Training.

Name	Grid Search Values	Value
Criterion	gini, entropy	entropy
Maximum Depth	2–10	4
Minimum Sample Split	0.025, 0.05, 0.1, 0.15, 0.2	0.05
Minimum Sample Leaf	0.01, 0.025, 0.05, 0.075, 0.1	0.1

Figure 12 highlights the importance of the tail posture for emotion recognition in dogs. Moreover, the decision path relevant to the “happiness” class suggests that only dogs showing the play posture are considered happy. Another interesting finding provided by the graphical tree is the irrelevance of the back leg condition and the ear position for the predictions of this decision tree.

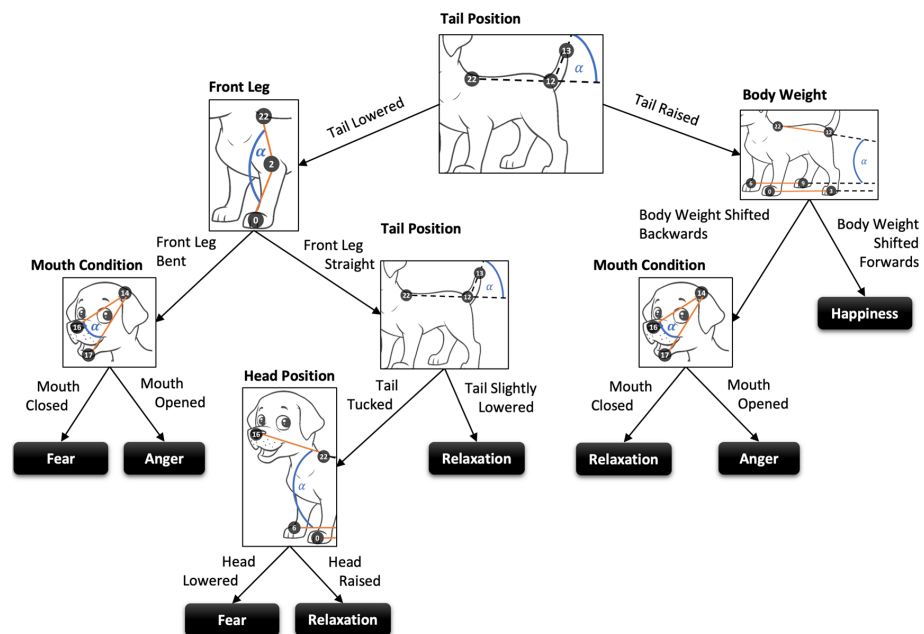


Figure 12. Explanatory illustration of the decision tree.

The neural network confusion matrix shown in Figure 13 shows a balanced model performance for all emotion classes. On the other hand, the decision tree matrix shows strong recognition of the class “fear”, but relatively poor recognition of “happiness”. As mentioned before, predicting the emotion “happiness” might suffer from the large variety of poses in the training set, which are not adequately represented by the pose metrics.

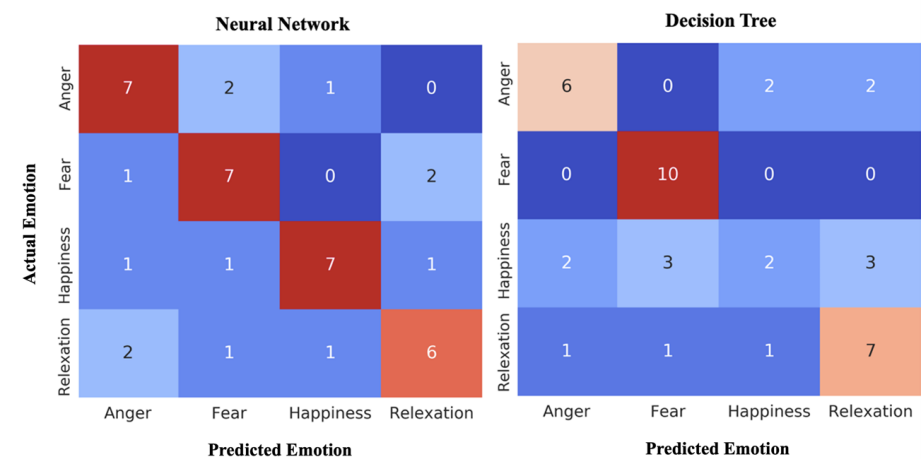


Figure 13. Confusion matrices of the neural network and decision tree.

4.3.3. Additional Classifiers

In addition to the classifiers already described, a logistic regression model and a support vector machine were trained. Table 6 summarizes the accuracies of all examined models.

Table 6. Overview of model performance.

Model	Accuracy	Explanatory Abilities	Ability to Deal with Missing Data
Neural Net	67.5%	-	✓
Decision Tree	62.5%	✓	-
Logistic Regression	62.5%	-	-
Support Vector Machine	67.5%	-	-

5. Discussion

Multiple models were developed to determine a dog's emotional state based on input images with an accuracy between 62.5% and 67.5%. Relying on the scientific findings of Amici et al. [3], this degree of emotion recognition capabilities in dogs exceeds the human one. Image-based emotion recognition promises wide everyday application for automatic emotion analysis of animals. In contrast to sensor-based emotion detection [8,9], the method explored in this work is more cost-efficient and more practicable since no sensor needs to be attached to the animal.

The neural network seems to be better suited for practical applications since it can handle missing landmark points and therefore provides results even for low-quality input images. The Decision Tree can be used not only for prediction, but also for a better understanding of how dogs express emotions.

Nevertheless, several aspects could be improved to increase emotion recognition accuracy further. Since the dataset contains only 100 images per class, an extension could contribute to the better generalizability of the model. Additionally, the chosen stimuli-based labeling approach of training data described in Chapter 2.1 could be questioned and possibly replaced by expert labeling.

Additional emotion classes could be introduced to provide more precise information about the dog's mood. In particular, the emotion "happiness" shows signs of high ambiguity for which a split into several categories might be suitable.

This work focuses mainly on posture and less on facial expression. Therefore, more pose metrics focusing on the facial landmark points could be added.

When comparing the developed emotion predictor to sensor-based canine emotion detection [8,9], it becomes apparent that photography induced pose analysis currently cannot compete with the more accurate information that can be received from sensor data. The sound produced by the animal might also be of interest. For example, refs. [10,30] have already successfully conducted studies that analyzed dog barks in terms of their emotional content. A video recording could capture both the image and the sound information and would allow us to feed more data into the emotion recognition. Nonetheless, the exclusively image-based emotion recognition is still highly relevant since its method of operation promises a wider everyday application for automatic emotion analysis of animals. Additionally, the photographic method can be assumed to be more cost efficient and less disturbing to the animal than the sensor-based analysis.

Author Contributions: Conceptualization, K.F. and P.A.G.; methodology, K.F.; software, K.F.; validation, K.F. and T.S.; data curation, K.F.; writing—original draft preparation, K.F. and T.S.; writing—review and editing, T.S. and P.A.G.; supervision, P.A.G. 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: Data available at request from the authors.

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

References

- Akins, N.J. *Dogs and People in Social, Working, Economic or Symbolic Interaction*; Snyder, L.M., Moore, E.A., Eds.; Oxbow Books: Oxford, UK, 2006; Volume 146.
- Hasegawa, M.; Ohtani, N.; Ohta, M. Dogs' Body Language Relevant to Learning Achievement. *Animals* **2014**, *4*, 45–58. [[CrossRef](#)] [[PubMed](#)]
- Amici, F.; Waterman, J.; Kellermann, C.; Karim, K.; Bräuer, J. The Ability to Recognize Dog Emotions Depends on the Cultural Milieu in Which We Grow Up. *Sci. Rep.* **2019**, *9*, 1–9. [[CrossRef](#)]
- Kujala, M. Canine Emotions: Guidelines for Research. *Anim. Sentience* **2018**, *2*, 18. [[CrossRef](#)]
- Waller, B.; Peirce, K.; Correia-Caeiro, C.; Oña, L.; Burrows, A.; Mccune, S.; Kaminski, J. Paedomorphic Facial Expressions Give Dogs a Selective Advantage. *PLoS ONE* **2013**, *8*, e82686. [[CrossRef](#)] [[PubMed](#)]
- North, S. Digi Tails: Auto-Prediction of Street Dog Emotions. 2019. Available online: <https://samim.io/p/2019-05-05-digi-tails-auto-prediction-of-street-dog-emotions-htt/> (accessed on 30 January 2022)
- Franzoni, V.; Milani, A.; Biondi, G.; Micheli, F. A Preliminary Work on Dog Emotion Recognition. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence—Companion Volume, New York, NY, USA, 14–17 October 2019; pp. 91–96. [[CrossRef](#)]
- Brugarolas, R.; Loftin, R.; Yang, P.; Roberts, D.; Sherman, B.; Bozkurt, A. Behavior recognition based on machine learning algorithms for a wireless canine machine interface. In Proceedings of the 2013 IEEE International Conference on Body Sensor Networks, Cambridge, MA, USA, 6–9 May 2013; pp. 1–5. [[CrossRef](#)]
- Aich, S.; Chakraborty, S.; Sim, J.S.; Jang, D.J.; Kim, H.C. The Design of an Automated System for the Analysis of the Activity and Emotional Patterns of Dogs with Wearable Sensors Using Machine Learning. *Appl. Sci.* **2019**, *9*, 4938. [[CrossRef](#)]
- Tsai, M.F.; Lin, P.C.; Huang, Z.H.; Lin, C.H. Multiple Feature Dependency Detection for Deep Learning Technology—Smart Pet Surveillance System Implementation. *Electronics* **2020**, *9*, 1387. [[CrossRef](#)]
- Maskeliunas, R.; Raudonis, V.; Damasevicius, R. Recognition of Emotional Vocalizations of Canine. *Acta Acust. United Acust.* **2018**, *104*, 304–314. [[CrossRef](#)]
- Raman, S.; Maskeliūnas, R.; Damaševičius, R. Markerless Dog Pose Recognition in the Wild Using ResNet Deep Learning Model. *Computers* **2022**, *11*, 2. [[CrossRef](#)]
- Bloom, T.; Friedman, H. Classifying dogs' (*Canis familiaris*) facial expressions from photographs. *Behav. Process.* **2013**, *96*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Coren, S. Which Emotions Do Dogs Actually Experience? *Psychology Today*, 14 March 2013.
- Meridda, A.; Gazzano, A.; Mariti, C. Assessment of dog facial mimicry: Proposal for an emotional dog facial action coding system (EMDOGFACTS). *J. Vet. Behav.* **2014**, *9*, e3. [[CrossRef](#)]
- Mellor, D. Tail Docking of Canine Puppies: Reassessment of the Tail's Role in Communication, the Acute Pain Caused by Docking and Interpretation of Behavioural Responses. *Animals* **2018**, *8*, 82. [[CrossRef](#)] [[PubMed](#)]
- Rohr, K. Introduction and Overview. In *Landmark-Based Image Analysis: Using Geometric and Intensity Models*; Rohr, K., Ed.; Computational Imaging and Vision; Springer: Dordrecht, The Netherlands, 2001; pp. 1–34. [[CrossRef](#)]
- Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289. [[CrossRef](#)] [[PubMed](#)]
- Biggs, B.; Boyne, O.; Charles, J.; Fitzgibbon, A.; Cipolla, R. Who Left the Dogs Out? 3D Animal Reconstruction with Expectation Maximization in the Loop. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020. [[CrossRef](#)]
- Cao, J.; Tang, H.; Fang, H.S.; Shen, X.; Lu, C.; Tai, Y.W. Cross-Domain Adaptation for Animal Pose Estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019.
- Caeiro, C.; Guo, K.; Mills, D. Dogs and humans respond to emotionally competent stimuli by producing different facial actions. *Sci. Rep.* **2017**, *7*, 15525. [[CrossRef](#)] [[PubMed](#)]
- Racca, A.; Guo, K.; Meints, K.; Mills, D.S. Reading Faces: Differential Lateral Gaze Bias in Processing Canine and Human Facial Expressions in Dogs and 4-Year-Old Children. *PLoS ONE* **2012**, *7*, e36076. [[CrossRef](#)] [[PubMed](#)]
- Riemer, S. Social dog—Emotional dog? *Anim. Sentience* **2017**, *2*, 9. [[CrossRef](#)]
- Olafenwa, M.; Olafenwa, J. ImageAI Documentation. 2022. Available online: <https://buildmedia.readthedocs.org/media/pdf/imageai/latest/imageai.pdf> (accessed on 30 January 2022)
- Khosla, A.; Jayadevaprakash, N.; Yao, B.; Li, F.F. Novel Dataset for Fine-Grained Image Categorization. In Proceedings of the CVPR Workshop on Fine-Grained Visual Categorization (FGVC), Colorado Springs, CO, USA, 25 June 2011.
- Coren, S. How To Read Your Dog's Body Language. Available online: <https://moderndogmagazine.com/articles/how-read-your-dogs-body-language/415> (accessed on 30 January 2022)
- Simpson, B.S. Canine Communication. *Vet. Clin. N. Am. Small Anim. Pract.* **1997**, *27*, 445–464. [[CrossRef](#)]

28. Divyang. Dog Body Part Detection. 2020. Original-Date: 2020-05-31T04:37:34Z. Available online: https://www.researchgate.net/publication/346416235_SUPPORT_VECTOR_MACHINE_SVM_BASED_ABNORMAL_CROWD_ACTIVITY_DETECTION (accessed on 30 January 2022)
29. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed.; O'Reilly Media: Beijing, China, 2019.
30. Hantke, S.; Cummins, N.; Schuller, B. What is my Dog Trying to Tell Me? The Automatic Recognition of the Context and Perceived Emotion of Dog Barks. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5134–5138. [[CrossRef](#)]