EMOTION RECOGNITION OF GOLDEN RETRIEVER USING DEEP LEARNING

Sridhar Chintala
School of CS & AI
SR University
Warangal, India
dr.sridharchintala@gmail.com

Deep Shekhar Acharya
Dept. of EEE
B.I.T. Mesra, Off-Campus Deoghar
Jharkhand, India
dsacharya@bitmesra.ac.in

Veekshitha Adharasani
School of CS & AI
SR University
Warangal, India
veekshithaadharasani@gmail.com

Rida Shireen
School of CS & AI
SR University
Warangal, India
ridashireen08@gmail.com

***Abstract*—Understanding a dog's emotion is crucial in attending to its mental and emotional health. This paper focuses on creating an effective deep learning model capable of accurately classifying the emotional expressions of Golden Retrievers based on still images. A custom data set has been created that contains 900 labeled images to train and evaluate the models, which included three primary emotions: Happy, Sad, and Angry. In developing robust solutions, MobileNetV2 based transfer learning was used as a feature extractor. The architectures included a standalone CNN, a hybrid of CNN and Simple RNN, and our proposed CNN with LSTM. The objective of this work is to demonstrate that sequential learning improves model performance in time-series data. For the performance of the different models, the CNN structure produced results of 82.78%, while CNN plus Simple RNN achieved 81.17%. The ultimate winner was the CNN plus LSTM model, which outperformed the others with an impressive 86.8% accuracy.**

***Index Terms*—CNN, LSTM, Deep Learning, RNN, Golden Retriver**

I. Introduction

Emotions play a crucial role in the way dogs interact with humans, communicate, adapt, and form bonds with their human companions [1]–[3]. Recognizing emotional expressions in dogs helps improve their care, behavior management, and social bonding, respectively. Among all dog breeds, Golden Retrievers are well known for their high sociability and affectionate nature. Understanding how Golden Retrievers express emotions such as happiness, sadness, or anger strengthens the human-animal connection and also supports better care, training, monitoring, and emotional health of pet dogs. The Golden Retriever is a rare kind of dog because of its unique behavioral, emotional, and physical qualities. This makes it a great topic for studies in human-animal interaction, emotion identification, and assistive applications. Golden Retrievers are known for having a stable personality, being very friendly, and having predictable emotional reactions, all of which are good for controlled experimental settings.

The research work in [1] introduces a generative AI ($GPT4$) based deep learning approach for classifying emotions in pet animals, focusing on dogs and cats. Using a modified EfficientNetB5 model enhanced with Dense Residual and Squeeze-and-Excitation blocks, the system categorizes emotions into angry, sad, happy, and neutral with high accuracy (98.2% training, 91.24% validation). The model uses transfer learning and data augmentation to improve performance and robustness. The research highlights the model's potential for real-time emotion detection, aiding pet care and monitoring. Future work aims to expand the dataset to include more species and improve real-time application, contributing to better animal welfare and human–animal relationships. The reconstruction of dogs' basic emotions like joy and anger using machine learning techniques is demonstrated in [4], using a momentum contrast for visual representation. The authors have employed the $SimCLR$, $MoCo$ frameworks and $ResNet50$ architecture on a data set of 2184 images to predict the emotions in dogs using facial expressions and body posture. The supervised $ResNet50$ model, performed noticeably better, achieving 74.32% accuracy, demonstrating the usefulness of supervised learning for similar tasks. The research work presented in [5], demonstrates the development of an emotion recognition system for dogs that identifies anger, fear, happiness, and relaxation. It uses a $DeepLabCut$ machine learning model based on pose estimation. A picture library with 400 images for each emotional state was compiled. 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.

The authors in [6], reconstructed the emotions that affect the behavior of dogs and how they react in various situations to a great extent. The authors demonstrated a new database comprising 15,599 labeled dog images expressing emotions such as aggression and fear, and an automatic classification method. Using advanced $AutoML$-based image classification techniques, they assessed the effectiveness of recognizing dog emotions, even without pre-processing the images. This approach allows for the easy development of systems that understand dog emotions and promotes further research on the topic. The research work in [7], demonstrates the formulation of a neural network based model that identifies canine emotional behavior, based on earlier work. The paper applies photo data of five emotions and targets a single breed. The paper compares the system with alternative systems through the use of transfer learning techniques. Results reveal that VGG16 and VGG19 work optimally, and thus the construction of a deep neural network called mVGG16. The system is tested with internal as well as external data to verify the robustness of the system, with the aim of detecting unsafe behavior and monitoring pets' movements in the absence of their masters. A CNN-based dog facial expression recognition model optimized using the Improved Whale Optimization Algorithm (IWOA), is presented in [8]. To prevent overfitting and reduce complexity, techniques like dropout, $L2$regularization, and Dlib-based face detection were used. The $IWOA$ dynamically adjusted the learning rate to enhance training efficiency. Compared to models like SVM and $LeNet-5$, $IWOA–CNN$ achieved better accuracy, showing the potential of swarm intelligence for tuning deep learning models.

The research work presented in [9], demonstrates the formulation of an automated system that identifies and reacts to certain emotions in dogs, such as aggression, sorrow, pleasure, and relaxation. When people learn to read their dogs' facial expressions, it could help them live more harmoniously with their canine companions. This project describes research on canine face expression recognition utilizing $CNN-VGG16$ deep learning algorithm model. The authors in [10] propose the application of a CNN model to categorize dog emotions into four groups. They identified the need to comprehend these emotions for animal well-being and human-animal relationships. The research applies a Kaggle dataset and an SGD optimizer, explaining pre-processing, model development, and training. The model has a 99.60% validation accuracy, which surpasses SVM and KNN. Some of the prominent research in this field may be found in [11]–[13].

Taking this domain as the inspiration, this work proposes a non-invasive image based deep learning model. The proposed model has been applied, specifically, to Golden Retrievers, however, it is equally applicable to other breeds as well. As per the literature, the existing techniques employ body-worn sensors or videos to assess and determine behavioral and emotional trends. The approach proposed in this article, relies on convolutional neural networks (CNN) and recurrent neural networks (RNN) for classification of dog emotions. In this work, three model, namely, CNN, CNN+SimpleRNN and CNN+LSTM, ahve been applied for emotion detection of Golden Retriever based on images. It has been observed that out of the three, the CNN + LSTM model achieved an of 89.17%, compared to the other models. The proposed approach is simpler and easier to use, practical for real-world applications like pet care, behavior monitoring, or smart home systems.

The rest of the paper is organized as follows: section II describes the dataset and the pre-processing done. Section III explains the methodology and the technique. SectionIV presents the results and discusses their significances, followed by the conclusion presented in sectionV.

II. Dataset Description and Preprocessing

For this research work, a custom dataset, consisting exclusively of Golden Retriever images, was developed to ensure a singular breed profile while minimizing breed diversity. The expressions of the dogs in the photos were classified and organized into three categories: ***Happy***, ***Sad***, and ***Angry***. Each emotion category is balanced since an equal number of images per category is maintained. In order to meet the input needs of the deep learning models, the 900 images were resized to 224 × 224 pixels. To improve model generalization and avoid overfitting, a set of data augmentation methods was employed which included, but are not limited to random rotation, zoom, flipping, width and height shifting, and brightness adjustment. This enhanced the effective size of the dataset while modeling numerous conditions that may be encountered in real-world application.

The dataset was organized into two sets, test and train, for easier access by image generators. The data set was cleaned to remove corrupted and invalid files, after which the data set was split, distributing 80% as training data (480 images) and 20% as testing data (120 images). This provides a clean and organized dataset, structured to enable accurate emotion classification for golden retrievers.

A. Approaches Using Deep Learning For Recognizing Emotions in Dogs

The past few years have seen deep learning take the lead in analyzing and detecting the emotions of animals, especially dogs, using their facial expressions and body movements.One such technique is the use of CNN s and RNN s, which aids in the visual and temporal examination of the data. Generally, these systems employ $MobileNet$ or $ResNet$ which are able to train in depth and retain important details, such as the position of the ears and eyes, with less effort and time. For best results, some researchers strive to make the models accurate and sufficiently general by using data augmentation, transfer learning, and dropout. The recognition of images and their spatial structures is a strength for CNNs, while emotional progressions from frame to frame are best understood by RNNs –and in particular $LSTM$s and $SimpleRNN$s.

For effective monitoring or as an aid during training, these models need to be efficient and operational in real time. Hybrid models have been found to perform better than individual CNNs, with CNNs combined with RNNs exhibiting wider accuracy in emotional states such as happy, sad, or angry. Depending on the model and training configuration, the test accuracies range from 76% to 89%. Although there are still hurdles in the domain of marking emotions across different breeds and dealing with faint emotional expressions.

III. Methodology

This section describes the overall workflow used for dog emotion recognition, including data preprocessing, model design, training strategy, and evaluation process.

A. Image Based Emotion Classification Method

To recognize a dog's emotion from an image, look at its facial expressions. Data collection is resized, normalized, and augmented to preserve uniformity and model learning consistency. This is especially true for dogs, who exhibit the same emotion differently. To fix this, numerical techniques focus on emotion-related patterns such as happiness, sadness, and

Fig. 1. Images of Golden Retriever with different expressions and postures.

rage to include salient elements. The notion is that emotional traits can be identified and learned systematically. Therefore, a well-balanced and labeled data collection is essential. Image-based techniques are a solid start, but further development is needed to create reliable and ubiquitous dog emotion detection systems.

B. Model Architecture

CNN-based deep learning models were tested for this challenge. Various models, including MobileNetV2 and VGG16, were examined for their ability to classify the three dog emotions: Happy, Sad, and Angry. The deep learning network was fed 900 RGB photos scaled to 224 × 224 pixels. The CNN architecture was compatible with this consistent input size, making training efficient. Examples of Golden Retriever photos are given in Figure 1. Each model was initialized with weights using the massive $ImageNet$ dataset, which contains millions of labeled photos in categories, to use existing knowledge and accelerate learning. Transfer learning allowed the model to use previously learned low- and mid-level visual elements, including edges, textures, and patterns, to adapt to Golden Retriever emotion recognition.

Fine-tuning the higher layers allowed the network to identify breed-specific emotional cues for the bespoke Golden Retriever dataset. A global average pooling (GAP) layer was added after the base model to reduce the spatial dimensions of the feature map and summaries the most critical information while minimizing overfitting and trainable parameters. After that, one or more fully connected (dense) layers abstracted the features into a compact categorization representation. A softmax-activated dense layer with three neurons was the final output layer for each of the three emotion classes, happy, sad, and angry. This setup allowed the model to provide

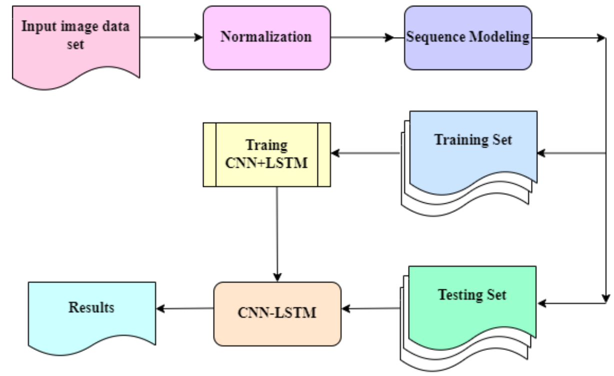

Fig. 2. Block diagram of CNN-LSTM approach.

probability scores for each emotion category, allowing accurate and interpretable emotion categorization from the input image. The CNN-LSTM block diagram is shown in Figure 2.

C. Training Configuration

The data set used to train the emotion recognition model was divided into two subsets: 80% of the images were allocated for training, while the remaining 20% were reserved for testing. This split ensured that the model had sufficient data to learn from, while also allowing for an unbiased evaluation of its performance on unseen data. Training was carried out using a batch size of 32, which balances computational efficiency and gradient stability during optimization.

To optimize model learning, the Adam optimizer was used due to its adaptive learning rate capabilities and effectiveness in handling sparse gradients. The loss function used was the *categorical cross-entropy* (CE), defined in (1), which is well suited for multiclass classification tasks, allowing the model to output probability distributions over the emotion categories.

$$CE = -\sum_{k=1}^{C_T} y_k \log_{10}(\hat{y}_k) \tag{1}$$

where, CE is the loss function, C_T represents the total number of classes, y_i is the true label for the i-th class and $\hat{y}_i$ signifies the predicted probability of the i-th class. To enhance the model's ability to generalize to new and varied data, extensive data augmentation techniques were applied during training. These enhancements simulate different real-world conditions and increase the diversity of the training data without requiring additional manual image collection. The techniques included horizontal flipping, brightness adjustments to simulate lighting variations, zooming to introduce scale variability, rotation to help the model handle orientation changes, and translation to make the model more robust to positioning differences. Together, these strategies significantly improved the model's ability to recognize emotions under varied conditions, leading to a more reliable and adaptable emotion recognition system.

IV. Results

The $CNN + LSTM$ model demonstrated superior performance in classifying emotional states by achieving a test

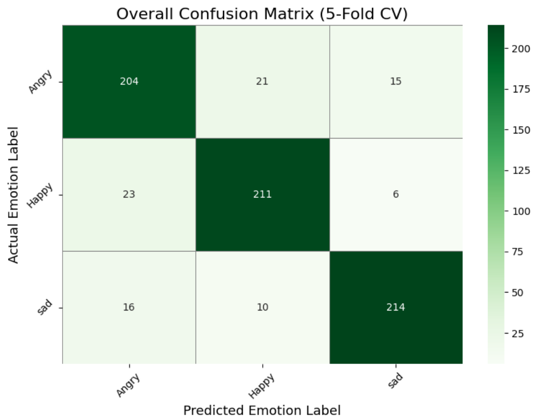

Fig. 3. Confusion matrix of the proposed CNN+LSTM for emotion classification

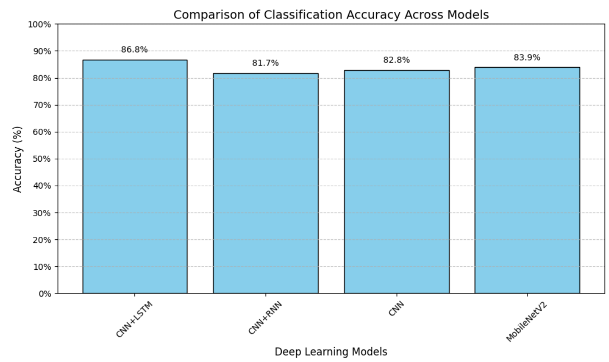

Fig. 4. Comparison of the accuracy of CNN+LSTM with existing methods

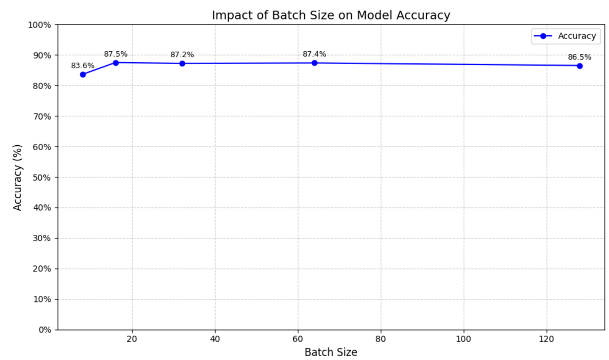

Fig. 5. Accuracy vs. Batch size of proposed CNN+LSTM algorithm

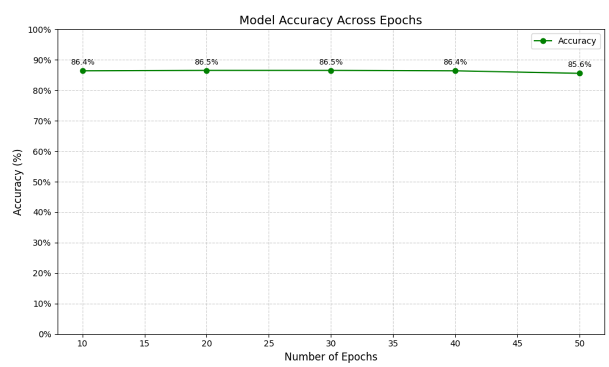

Fig. 6. Accuracy vs Epochs of proposed CNN+LSTM algorithm

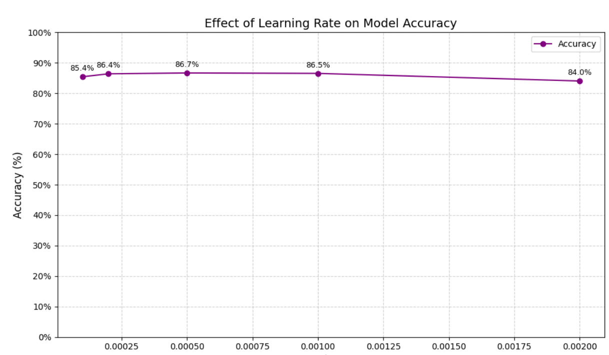

Fig. 7. Accuracy vs Learning rate of proposed CNN+LSTM algorithm

accuracy of 86.8%, indicating its strong ability to generalize to unseen data. This performance significantly surpassed that of the baseline models used for comparison. The CNN-only model, which relied solely on spatial feature extraction without any temporal modeling, achieved a lower test accuracy of 82.78%, suggesting limitations in capturing sequential patterns or contextual cues that may be important for recognizing emotional expressions.

Similarly, the $CNN+SimpleRNN$ model, which incorporated a basic recurrent neural network for temporal processing, performed better than the CNN-only architecture but still fell short of the CNN+LSTM model, achieving an accuracy of 81.7%. The results are summarized in Table I. The improved results of the $CNN+LSTM$ architecture can be attributed to the Long Short-Term Memory ($LSTM$) units' ability to retain and learn from long-term dependencies and temporal patterns within the input data. This made the model more effective in interpreting subtle emotional transitions, especially when working with sequences such as video frames or time-series image data.

The confusion matrix is shown in Figure 3. A comparison of the accuracy of the models is also, illustrated in Figure 4, which corroborates the superior performance of the $CNN+LSTM$ model in emotion classification.

TABLE I
ACCURACY OF CNN, CNN+SIMPLE RNN, MOBILENETV2, AND CNN+LSTM

S.No.	Model	Accuracy (%)
1	CNN	82.78
2	CNN+RNN	81.7
3	MobileNetV2	83.9
4	CNN+LSTM	86.8

Figure 5 represents the effect of varying the batch size on accuracy of the model. It is observed that the accuracy decreases on choosing a batch size less than 20. Also, it is revealed from the plot, that a batch size of 30-60 boosts the classification accuracy. The effect of number of epochs on

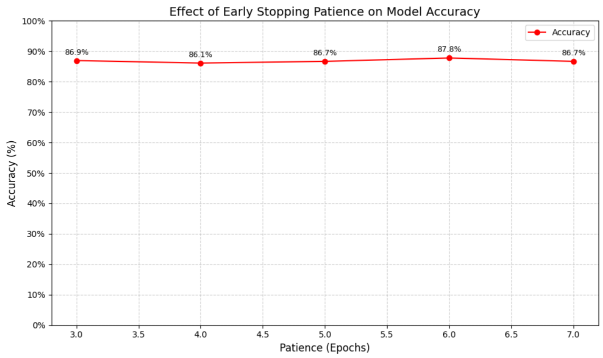

Fig. 8. Accuracy vs Patience of proposed CNN+LSTM algorithm

the accuracy is illustrated in Figure 6. It may be observed that both less and very high epochs degrades the accuracy of classification. In this work, the best accuracy was achieved when no. of epochs was 30.

Figure 7 reveals that a learning rate less than 0.001 reduces the accuracy and that a learning rate $0.00025 \leq \eta \leq 0.001$ results in better accuracy. Patience is also an important parameter. Patience indicates the *number of epochs to wait after which the training will be terminated if no improvement in a specified validation metric is observed*. In this research work, it has been found that a patience of 5.0 produces the best results, as shown in Figure 8. For values < 5.0 and > 5.0, the model exhibited a reduced precision. The aforementioned hyperparameter tunings are selected based on random search, where suitable values have been considered in this work. The values/range of values of the parameters for which the model yielded the best accuracy are listed in Table II.

TABLE II
BEST VALUES OF THE HYPER PARAMETERS

S. No.	Parameters	Best Value
1	Batch size	30-60
2	Epochs	30
3	Learning rate	0.0005 - 0.001
4	Patience	5.0

V. CONCLUSION

This study explores emotion recognition in Golden Retrievers using a custom dataset analyzed by three architectures: CNN, CNN + Simple RNN, and CNN + LSTM. Three emotions—happy, sad, and angry—were classified using visual cues from RGB images. The CNN model, focused on spatial features, reached 82.78% accuracy. Adding temporal dynamics, the CNN+Simple RNN scored 81.7%, underscoring sequential modeling's role. The $CNN + LSTM$ model excelled with 86.8% accuracy, highlighting $CNN+LSTM$'s effectiveness in capturing spatial and temporal patterns critical for detecting dogs' emotions. Future work involves real-time video detection and advanced architectures like Transformers or hybrid attention systems to enhance results. Expanding this model to other breeds and creating a mobile or IoT app could significantly aid pet owners and professionals in animal care.

REFERENCES

[1] B. Cetintav, Y. S. Guven, E. Gulek, and A. A. Akbas, "Generative ai meets animal welfare: Evaluating gpt-4 for pet emotion detection," *Animals*, vol. 15, no. 4, p. 492, 2025.

[2] V. Franzoni, A. Milani, G. Biondi, and F. Micheli, "A preliminary work on dog emotion recognition," in *IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume*, pp. 91–96, 2019.

[3] C. Halkiopoulos, E. Gkintoni, A. Aroutzidis, and H. Antonopoulou, "Advances in neuroimaging and deep learning for emotion detection: A systematic review of cognitive neuroscience and algorithmic innovations," *Diagnostics*, vol. 15, no. 4, p. 456, 2025.

[4] A. Bhave, A. Hafner, A. Bhave, and P. A. Gloor, "Unsupervised canine emotion recognition using momentum contrast," *Sensors*, vol. 24, no. 22, p. 7324, 2024.

[5] K. Ferres, T. Schloesser, and P. A. Gloor, "Predicting dog emotions based on posture analysis using deeplabcut," *Future Internet*, vol. 14, no. 4, p. 97, 2022.

[6] F. Hernández-Luquin, H. J. Escalante, L. Villaseñor-Pineda, V. Reyes-Meza, L. Villaseñor-Pineda, H. Pérez-Espinosa, V. Reyes-Meza, H. J. Escalante, and B. Gutierrez-Serafín, "Dog emotion recognition from images in the wild: Debiw dataset and first results," in *Proceedings of the Ninth International Conference on Animal-Computer Interaction*, pp. 1–13, 2022.

[7] Z. Kowalczuk, M. Czubenko, and W. Żmuda-Trzebiatowska, "Categorization of emotions in dog behavior based on the deep neural network," *Computational Intelligence*, vol. 38, no. 6, pp. 2116–2133, 2022.

[8] Y. Mao and Y. Liu, "Pet dog facial expression recognition based on convolutional neural network and improved whale optimization algorithm," *Scientific Reports*, vol. 13, no. 1, p. 3314, 2023.

[9] D. Shubhangi and S. Fatima, "Emotion classification in dogs by deep learning," in *2024 2nd International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, pp. 838–844, IEEE, 2024.

[10] V. Tanwar, "Understanding dog emotions through deep learning: A cnn-based classification framework," in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 964–969, IEEE, 2024.

[11] C. Yang, S. Hu, L. Tang, R. Deng, G. Zhou, J. Yi, and A. Chen, "A barking emotion recognition method based on mamba and synchrosqueezing short-time fourier transform," *Expert Systems with Applications*, vol. 258, p. 125213, 2024.

[12] K. Guo, C. Correia-Caeiro, and D. S. Mills, "Category-dependent contribution of dog facial and bodily cues in human perception of dog emotions," *Applied Animal Behaviour Science*, vol. 280, p. 106427, 2024.

[13] B. Or, "Transformer-based dog behavior classification with motion sensors," *IEEE Sensors Journal*, 2024.