

A Convolutional Neural Network-Based Approach for Automatic Dog Breed Classification Using Modified-Xception Model



Ayan Mondal, Subhankar Samanta, and Vinod Jha

Abstract The social structure of urban India has been changed and most pet lovers choose the dog over any other kind of pet. The population of adopted dogs is projected at 31.5 million approximately by 2023. With the increase in demand, the fraud cases of selling the right breed are rising day by day. With the demand for different dog breeds, recognizing the correct breed in time by their physical ability, instinct, interaction, and behavior, the body structure is necessary. Recent developments of artificial intelligence have already proven its superiority over the human capability for image classification tasks. The present work has built a Convolutional Neural Network (CNN)-based model to construct a highly accurate dog breed image classifier. In this paper, various state-of-the-art deep CNN models have been applied, and a modified-Xception model has been proposed for improving the overall accuracy. For evaluating the overall classification performance of our proposed methodology, the Kaggle Dog Breed Identification dataset has been used and throughout the experiment, our modified-Xception model has achieved 87.40%, the highest overall accuracy.

Keywords CNN · Dog breed classification · Image classification · Leaky ReLU · Xception

1 Introduction

Nowadays, dogs are the most common pets to be adopted at home. Security personnel also prefer some specific dog breeds for security purposes. As per statistics of the ASPCA [1], nearly 1.6 million dogs are adopted every year. By observation, one can conclude that the intra-class differences in dog breeds are more than the inter-class differences. So, identification of dog breeds becomes very difficult particularly, for the new pet enthusiasts. Dogs are also the most genetically diverse animals on the

A. Mondal (✉) · S. Samanta · V. Jha
School of Electronics Engineering, Kalinga Institute of Industrial Technology (KIIT),
Bhubaneswar, India

V. Jha
e-mail: vjhafet@kiit.ac.in

earth. The current advancements of deep learning, especially Convolutional Neural Network (CNN), have already proved its superiority over human capabilities toward object identification. So, CNN-based dog breed identification is very much essential to decrease the complexity of dog breed classification.

Deep Learning is a widely growing field with continuous up-gradation in various domains like image recognition, speech recognition, object detection, data generation, etc. But still, much work needs to be done in this field to explore complex problems. In recent times with the advent of various CNN architectures like Xception, ResNet, DenseNet, etc. and with the help of Transfer Learning [2], improving classification model performance has become quite easier.

This paper aims to classify different breeds of dogs with improved accuracy compared to the existing dog breed identifiers in the literature. To develop the classifier, we have used the Kaggle Dog Breed Identification dataset [3]. The goal was to make a generalized model which would be able to predict the dog breeds irrespective of any class and with higher accuracy. We have experimented with different state-of-the-art deep CNN models like DenseNet201, Xception, ResNet50, VGG19, etc. We have also made some modifications in the Xception model architecture to improve the overall classification accuracy. In our experiment, our modified-Xception model has obtained the highest, 87.40%, overall accuracy on this applied Kaggle Dog Breed Identification dataset. Based on the previous works on this dataset, our classification model accuracy is a cut above the other proposed models in the literature. Hence, the foremost contributions of this article are as follows: (i) proposed a modified-Xception model to identify the different dog breeds; (ii) compared the adopted method with various pre-trained models; (iii) comparative performance analysis between similar previous works and the proposed approach.

The rest of this article has been structured as follows: a quick overview of the related works of dog breed recognition has been provided in Sect. 2, where our proposed methodology has been explained in Sect. 3 in detail. The experimental results and analysis are presented in Sect. 4 and finally, the article ends with the conclusions and future works.

2 Literature Review

From the past decade, many researchers have earlier tried to construct a dog breed image classifier. In most cases, they have used different CNN architectures. Mulligan et al. [3] have used the Kaggle Dog Breed Identification dataset and experimented with Xception followed by a Multi Linear Perceptron (MLP), but got a very low overall accuracy of 54.80%. There is still a chance of increasing the accuracy by applying more neurons and fine-tuning [4]. In the same context, Shi et al. [5] have also used the same Kaggle Dog Breed Identification dataset to classify the different dog breeds. They have applied various pre-trained CNN models, and among them, DenseNet161 has achieved the best overall accuracy of 85.64%. Kim et al. [6] have also used the same dataset to develop their dog breed classifier model. They have

applied proper data augmentation and got the highest overall classification accuracy of 83.22% using the ResNet152 model.

Sinnot et al. [7] have used the Stanford Dog dataset for identifying the different breeds of dogs. They have also used proper data augmentation and got superior classification model performance using the VGGNet model. Their image classification model has achieved an overall accuracy of 85% for 50 classes but, it drops to 63% for 120 dog breeds. In the same context, Ráduly et al. [8] have also used the Stanford Dog dataset. They have used proper data augmentation and hyperparameter tuning. They have applied the ResNeInception-ResNet-v2 model, which has achieved a decent overall accuracy of 90.69%. But, due to its massive amount of weight, it is computationally quite expensive.

Zou et al. [9] have contributed by developing a new dataset named Tsinghua for dog breed classification. They have removed similar images by computing image structural similarity (SSIM). Further, they have applied three different deep neural networks: PMG, TBMSL-Net, and WS-DAN. Throughout their experiment, WS-DAN has provided superior classification performance over the other models. It has achieved around 86.04% overall accuracy for eighty classes, but this accuracy falls to 58.14% on the Stanford dog dataset.

Liu et al. [10] have used the Columbia dataset for dog breed classification. But they have taken traditional machine learning approaches instead of using CNN. They have used SIFT features descriptor and SVM algorithm for their experiment and got only 67% overall accuracy. Borwarnginn et al. [11] have also experimented on the same Colombia dataset. They have implemented the NASNet model, which has achieved 89.92% overall accuracy. In the same context, LaRow et al. [12] have used the same dataset to classify the different dog breeds. As data pre-processing, they have extracted the facial key point from the dog images and have used a 17-layer CNN architecture for feature extraction. They have used the SVM model for classification. But their CNN-SVM approach has achieved a very low, 52%, overall accuracy. Table 1 represents the classification performance of the various implied methodologies in the literature for dog breed classification.

3 Methodology

This section vividly describes our proposed approach to building a CNN-based dog breed classifier model. The main steps of our methodology are data pre-processing, feature extraction, model training, and prediction on the new images.

3.1 Dataset Description

We have experimented on the Kaggle Dog Breed Identification dataset [3] that is collected from Kaggle. This dataset contains 10,222 training images in 120 classes

3.2 Data Pre-processing

In the Kaggle Dog Breed dataset, the available test set is not labeled, so it is troublesome to evaluate the classification model performance from this test set. To tackle this problem, we have split the training set into a ratio of 80:20 for training and testing purposes, respectively. To seize the overfitting problem, we have also used various data augmentation techniques. We have applied a width-shift-range of 0.25, height-shift-range of 0.25, zoom-range of 0.2, and horizontal-shift and generated many images from each image of the training set. Moreover, we have resized all images in $224 \times 224 \times 3$ for our experiment.

3.3 Convolutional Neural Network (CNN)

CNN is a specialized neural network for image classification. It mimics the visual cortex of the animal brain to recognize and process images. CNNs consist of several building blocks such as a convolutional layer, pooling layer, activation function, and fully connected layer. Figure 2 depicts the basic architecture of a CNN model.

The convolutional layer uses convolution operation to find the features from an image. Equation 1 mathematically expresses the convolutional operation.

$$(X * K)(i, j) = \sum_p \sum_q K(p, q)X(i - p, j - q) \tag{1}$$

where K is the kernel and X denotes the inputs.

Pooling layers generally reduce the dimensions of the feature maps. Thus, the number of trainable parameters decreases and computation time becomes lower. Activation functions introduce non-linearity between the inputs and the outputs. In our methodology, we have used two activation functions: Leaky ReLU and Softmax [13]. We have applied Leaky ReLU and Softmax activation functions in the hidden layers and the output layer, respectively. Leaky ReLU and Softmax functions are mathematically expressed by Eqs. 2 and 3, respectively.

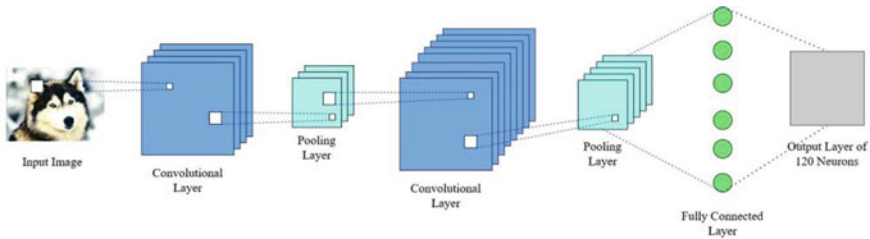


Fig. 2 Basic architecture of a CNN model

$$\text{LeakyReLU}(m) = \begin{pmatrix} am, & \text{if } m \leq 0 \\ m, & \text{if } m > 0 \end{pmatrix} \quad (2)$$

$$\text{softmax}(z)_i = \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)} \text{ for } i = 1, \dots, n \text{ and } z = (z_1, \dots, z_n) \in \mathbb{R}^n \quad (3)$$

where m is the input of Leaky ReLU function, $a = 0.01$ and z_i is the i th element of the input vector z for Softmax function.

3.4 Modified-Xception Model

In this paper, we have applied various deep CNN models like ResNet50, VGG16, DenseNet201, and Xception and have modified the Xception model by replacing its top layers with one Global Average Pooling layer, three consecutive Dense layers, 50% dropout in all of them and finally, one Softmax layer. The dropout layers are added to tackle the overfitting problem during model training. Moreover, we have used the Leaky ReLU activation function rather than ReLU [13] in the hidden layers. ReLU is the most popular activation function, but it offers zero output for the negative inputs. As a result, it causes vanishing gradient problems [14] during model training. On the other hand, Leaky ReLU seizes this vanishing gradient problem by offering a small output for the negative values. As a result, the classification performance of the CNN model increases. Figure 3 depicts the architecture of our modified-Xception model.

3.5 Experimental Setup

To classify the different dog breeds, we have trained our modified-Xception model along with the pre-trained models for 100 epochs. The models are compiled with the Adam optimizer having a 0.0001 learning rate and a loss function, namely categorical cross-entropy. Too many epochs often cause overfitting problems in the learning phase of a classification model. To overcome this problem, we have used the Early Stopping algorithm [15]. It halts the training phase whenever generalization error does not decrease. Moreover, we have reduced the learning rate to avoid any stagnation in the model learning phase. Throughout the experiment, we have used the TensorFlow framework for model training and data pre-processing; Matplotlib and seaborn for data visualization on Google Colaboratory.

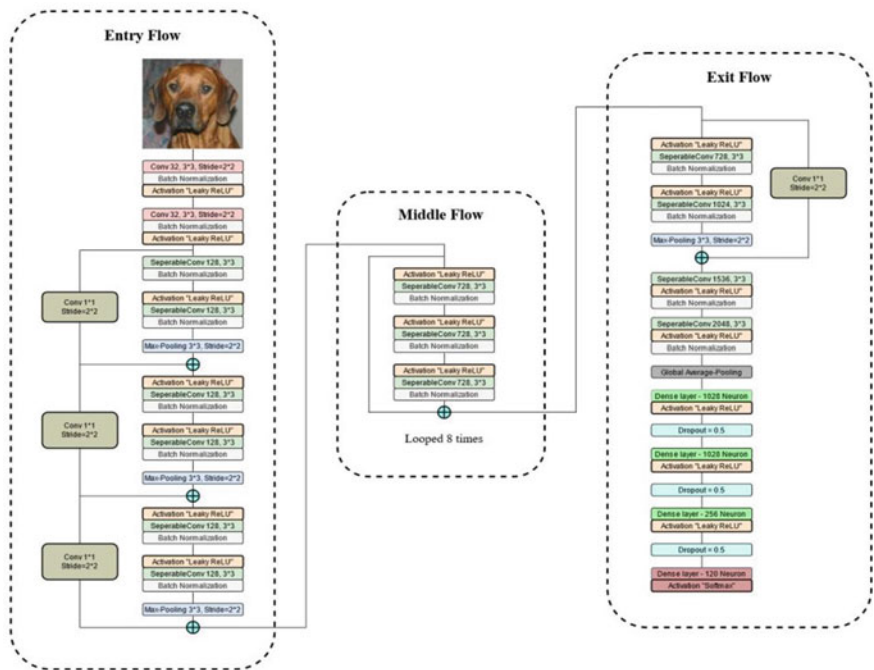


Fig. 3 Model architecture of our modified-Xception model

4 Results and Discussions

Here, we have presented our experimental outcomes and also compared the proposed methodology with the previous works in the Kaggle Dog Breed Identification dataset. We have applied different deep CNN models like ResNet50, VGG16, DenseNet201, and Xception and proposed a modified-Xception model. Figure 4 graphically represents our experimental results.

Throughout the experiment, our proposed modified-Xception model has achieved the highest 87.40% overall classification accuracy, whereas the original Xception model has achieved the second highest 84.15% overall classification accuracy. To find the overall classification accuracy, we have utilized Eq. 4.

$$\text{accuracy} = \frac{A + C}{A + B + C + D} \tag{4}$$

where A is the number of items whose true labels are positive and also classified as positive; B is the number of items whose true labels are negative but predicted as positive; C is the number of items whose true labels are negative and also correctly classified as negative; D indicates the number of items whose true labels are positive but classified as negative.



Fig. 4 Graphical representation of the overall accuracy using deep CNN models

Figure 5 depicts the confusion matrix achieved by the proposed modified-Xception model. To evaluate our model performance, we have presented a comparative performance analysis of the proposed methodology with the existing approaches to classify the dog breed images on the Kaggle Dog Breed Identification dataset in Table 2. It

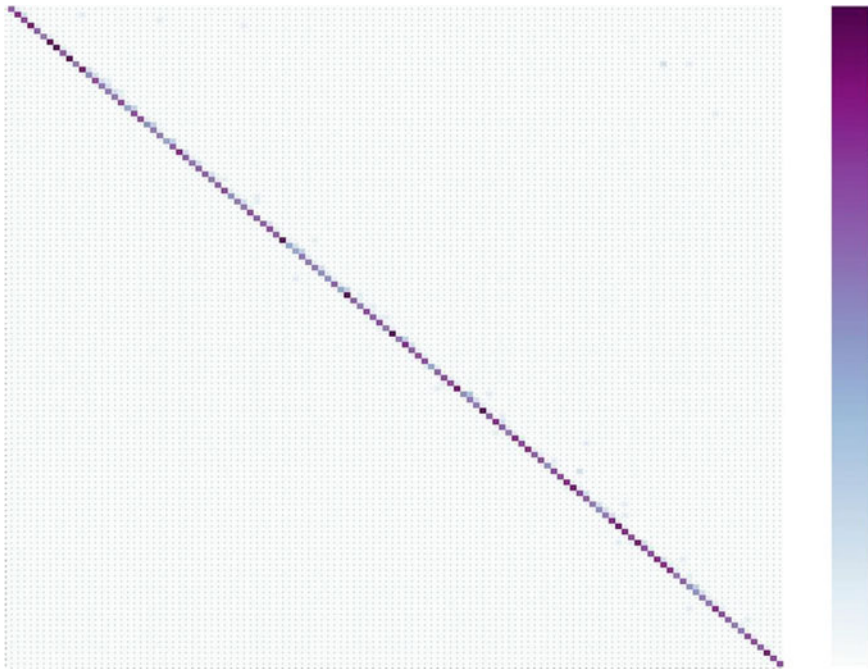


Fig. 5 Confusion matrix achieved by the modified-Xception model

Table 2 Comparative performance analysis between the implied methodology and the previous approaches of literature on Kaggle Dog Breed Identification dataset

Author	Models	Overall accuracy (%)
Mulligan et al. [3]	Xception	54.80
Shi et al. [5]	DenseNet161	85.64
Kim et al. [6]	ResNet152	83.22
Proposed method	Fine-tuned Xception Modified-Xception	84.15 87.40

can be observed that on this dataset, our proposed model has achieved better accuracy as compared to others.

Our methodology also provides a decent recognizing rate to recognize the new images. To recognize a particular image, our proposed modified-Xception model takes around 1.56 ms.

5 Conclusion

This article briefly investigates the ability of the CNN model to classify the different dog breeds. We have also shown the usefulness of using Transfer Learning and fine-tuning techniques for image classification tasks. Here, we have demonstrated the importance of proper data augmentation and pre-processing to improve classification accuracy. Our proposed modified-Xception model has achieved the highest 87.40% overall accuracy on the Kaggle dog breed identification dataset. In the future, we intend to work on a very deep Densely Connected Neural Network and ensemble network models for further improvements. However, our work will inspire the researchers to introduce more developments in dog breed identification.

Declaration The authors have no conflict of interest to declare those are relevant to the contents of this article.

References

1. Weiss E, Gramann S, Spain CV, Slater M (2015) Goodbye to a good friend: an exploration of the re-homing of cats and dogs in the US. *Open Journal of Animal Sciences* 5(04):435
2. Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242–264). IGI global.
3. Mulligan, K., & Rivas, P. (2019). Dog breed identification with a neural network over learned representations from the xception cnn architecture. In *21st International conference on artificial intelligence (ICAI 2019)*.
4. Guo, Y., Shi, H., Kumar, A., Grauman, K., Rosing, T., & Feris, R. (2019). Spottune: transfer learning through adaptive fine-tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4805–4814).

5. Shi, W., Chen, J., Liu, M., & Liu, F. (2018). Dog Breed Identification.
6. Kim, D. Final Project Report-Dog Breed Classification.
7. Sinnott, R. O., Wu, F., & Chen, W. (2018, December). A Mobile Application for Dog Breed Detection and Recognition Based on Deep Learning. In 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT) (pp. 87–96). IEEE.
8. Ráduly, Z., Sulyok, C., Vadász, Z., & Zölde, A. (2018, September). Dog breed identification using deep learning. In 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000271–000276). IEEE.
9. Zou DN, Zhang SH, Mu TJ, Zhang M (2020) A new dataset of dog breed images and a benchmark for finegrained classification. *Computational Visual Media* 6(4):477–487
10. Liu, J., Kanazawa, A., Jacobs, D., & Belhumeur, P. (2012, October). Dog breed classification using part localization. In *European conference on computer vision* (pp. 172–185). Springer, Berlin, Heidelberg.
11. Borwamginn P, Kusakunniran W, Karnjanapreechakorn S, Thongkanchorn K (2021) Knowing Your Dog Breed: Identifying a Dog Breed with Deep Learning. *Int J Autom Comput* 18(1):45–54
12. LaRow, W., Mittl, B., & Singh, V. (2016). Dog breed identification. *Network*.
13. Sharma S, Sharma S (2017) Activation functions in neural networks. *Towards Data Science* 6(12):310–316
14. Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Internat J Uncertain Fuzziness Knowledge-Based Systems* 6(02):107–116
15. Prechelt, L. (1998). Early stopping-but when?. In *Neural Networks: Tricks of the trade* (pp. 55–69). Springer, Berlin, Heidelberg.