Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2021.019020
Article

Tech Science Press

Predicting the Breed of Dogs and Cats with Fine-Tuned Keras Applications

I.-Hung Wang[1], Mahardi[2], Kuang-Chyi Lee[2,*] and Shinn-Liang Chang[1]

[1]Department of Power Mechanical Engineering, National Formosa University
[2]Department of Automation Engineering, National Formosa University, Huwei Township, Taiwan
*Corresponding Author: Kuang-Chyi Lee. Email: kclee@gs.nfu.edu.tw
Received: 29 March 2021; Accepted: 30 April 2021

Abstract: The images classification is one of the most common applications of deep learning. Images of dogs and cats are mostly used as examples for image classification models, as they are relatively easy for the human eyes to recognize. However, classifying the breed of a dog or a cat has its own complexity. In this paper, a fine-tuned pre-trained model of a Keras' application was built with a new dataset of dogs and cats to predict the breed of identified dogs or cats. Keras applications are deep learning models, which have been previously trained with general image datasets from ImageNet. In this paper, the ResNet-152 v2, Inception-ResNet v2, and Xception models, adopted from Keras application, are retrained to predict the breed among the 21 classes of dogs and cats. Our results indicate that the Xception model has produced the highest prediction accuracy. The training accuracy is 99.49%, the validation accuracy is 99.21%, and the testing accuracy is 91.24%. Besides, the training time is about 14 hours and the predicting time is about 18.41 seconds.

Keywords: Image classification; deep learning; Keras; Inception; ResNet; Xception

1 Introduction

Deep learning is commonly used to solve computer vision problems, with researchers building upon each other's work. Dean et al. [1] applied deep learning to speech recognition, and Krizhevsky et al. [2] developed a convolutional neural network (CNN) for image classification. Recognizing that building new, accurate CNNs is difficult due to the data and time required, researchers have fine-tuned existing models to improve results without that expense. Tajbakhsh et al. [3] demonstrated a fine-tuned CNN model that produced better results than an all-new one. Yosinski et al. [4] then demonstrated the feature transferability of CNN models and developed a new image classifier using a pre-built Keras model with a new dataset. Our aim is to tailor a Keras model to develop a classifier for identifying the breeds of dogs and cats (CDC). Our approach can also be applied to develop new image classifiers to enable robots in automated factories to identify objects using appropriate datasets of tools and work pieces.

2 Related Works

Image classification of dogs and cats has been frequently used as a case study for deep learning methods. Parkhi et al. [5] proposed a method for classifying the images of 37 different breeds of dogs and cats with an accuracy of 59%. Panigrahi et al. [6] used a deep learning model to classify images of dogs and cats simply as "dog" or "cat", with a testing accuracy of 88.31%. Jajodia et al. [7] used a sequential CNN to make a similar basic distinction, with 90.10% accuracy. Reddy et al. [8], Lo et al. [9], Deng et al. [10], and Buddhavarapu et al. [11] utilized transfer learning methods with Keras models to create the new models of Resnet, Inception-Resnet, and Xception.

3 CDC Dataset

We adopted Keras models pre-trained with a general image dataset. We used that with our new CDC with images of dogs and cats as listed in Tab. 1. Our data included images of 21 breeds of dogs or cats divided into training, validation, and testing images as shown. All images were taken from Dreamstime stock photos [12]. In total, we used 20,574 training images, 2,572 validation images, and 2,590 testing images.

Table 1: Image details for training and testing the CDC models

Category	Breed	Training	Validation	Testing
Dog	Akita	818	102	103
Dog	Alaskan Malamute	992	124	124
Dog	Basenji	918	114	116
Dog	Basset Hound	967	120	122
Dog	Beagle	956	119	120
Dog	Belgian Malinois	662	82	84
Dog	Bernese Mountain Dog	1233	154	155
Dog	Border Collie	1224	153	153
Dog	Boston Terrier	960	120	120
Cat	Norwegian Forest Cat	976	122	123
Dog	Shiba Inu	1293	163	161
Cat	Abyssinian Cat	1045	132	131
Cat	American Short Hair	1233	154	155
Cat	Birman Cat	896	112	113
Cat	Cornish Rex	757	94	96
Cat	Devon Rex	746	93	94
Cat	Maine Coon	1134	141	143
Cat	Scottish Fold	1128	141	142
Cat	Siamese Cat	718	90	91
Cat	Siberian Cat	1186	149	151
Cat	Somali Cat	732	93	93
Total		**20574**	**2572**	**2590**

4 Models Used with CDC

To test our approach, we adopted the Inception-ResNet v2, ResNet152 v2, and Xception models from the Keras repository and fine-tuned them with the new dataset. We fine-tuned CDC with the training and validation datasets, omitted the top layer of each Keras model, and redefined the number of fully connected layers based on the number of breeds. The training parameters are shown in Tab. 2. For training ResNet-152 v2 and Inception-ResNet v2, we used 50 epochs and a batch size of 16. For training Xception, we used 50 epochs and a batch size of 4. We used the stochastic gradient descent (SGD) as the optimizer in all cases, with a learning rate of 0.0001. Finally, we employed binary cross-entropy as the loss function throughout.

Table 2: Parameters for training the models

Keras Model	Epochs	Batch Size	Learning Rate	Loss Function	Optimizer
Resnet-152 v2	50	16	0.0001	binary cross-entropy	SGD
Inception-ResNet v2	50	16	0.0001	binary cross-entropy	SGD
Xception	50	4	0.0001	binary cross-entropy	SGD

We created the confusion matrix by using the fine-tuned Keras model with the testing dataset as input. We calculated the testing accuracy of each CDC-Keras model combination using the confusion matrix and Eq. (1):

$$\text{Accuracy (\%)} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FN_i} \times 100\%, \tag{1}$$

where TP_i is the true positive value for each breed, FN_i is the false negative value for each breed, and n is the number of breeds.

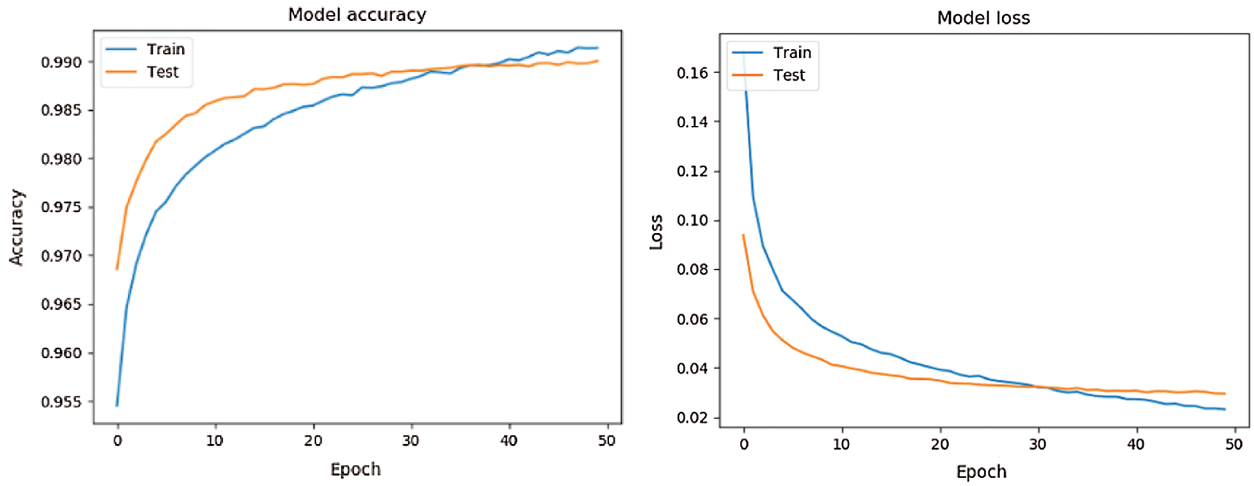
4.1 ResNet-152-Based CDC Model

He et al. [13] proposed ResNet v2 as an improvement of the residual network model. Identity mapping was used to directly propagate the forward and backward signals from one block to another. ResNet v2 offers better performance than the previous version, which has various types with different sizes of hidden layers. ResNet-152 v2 has 152 hidden layers and uses a fixed input image size of 224 × 224 RGB pixels.

The training result for CDC using the fine-tuned ResNet-152 model is shown in Fig. 1, showing accuracy increasing and loss decreasing with the training epoch. The accuracy of the training data with the ResNet-152-based CDC model was about 99.14% with a validation accuracy of about 99.00%. Training the ResNet-152-based CDC required 9 hours and 48 minutes. The confusion matrix of the ResNet-152-based CDC model shown in Tab. 3 indicates the prediction and recall percentage of each breed. The overall accuracy was 89.31%.

4.2 Inception-ResNet-Based CDC Model

Szegedy et al. [14] developed Inception-ResNet v2 model as an improvement to Inception v3 with residual connections to increase training speed and recognition performance. The input images to Inception-ResNet v2 are fixed size 299 × 299 RGB pixels. Training results of the fine-tuned Inception-ResNet-based CDC model are shown in Fig. 2. The final training accuracy and validation were about 98.97% and 98.94%, respectively. Training this combination required 12 hours and 29 minutes. The confusion matrix of the Inception-ResNet-based CDC model is shown in Tab. 4. The testing accuracy was about 89.50%.

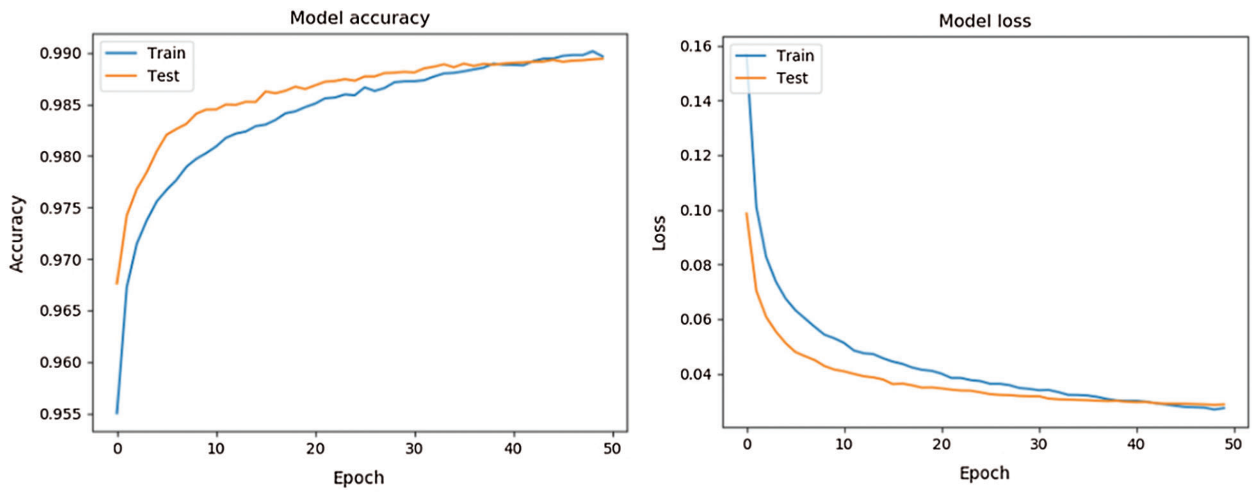


(a) (b)

Figure 1: Accuracy and loss versus epochs when training ResNet-152 v2 (a) Accuracy (b) Loss

Table 3: Confusion matrix of Resnet-152-based CDC model

	Pred. label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Recall
Akita	0	87	2	1	0	0	0	0	3	1	0	9	0	0	0	0	0	0	0	0	0	0	84.47%
Alaskan Malamute	1	0	124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
Basenji	2	1	0	111	0	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	95.69%
Basset Hound	3	0	0	0	117	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	95.90%
Beagle	4	1	0	1	2	113	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	94.17%
Belgian Malinois	5	2	0	0	0	0	79	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	94.05%
Bernese Mountain Dog	6	0	0	0	1	0	1	151	2	0	0	0	0	0	0	0	0	0	0	0	0	0	97.42%
Border Collie	7	1	1	0	1	1	0	4	143	1	1	0	0	0	0	0	0	0	0	0	0	0	93.46%
Boston Terrier	8	0	0	0	0	0	0	0	0	118	0	0	0	1	0	0	1	0	0	0	0	0	98.33%
Norwegian Forest Cat	9	0	0	0	0	0	0	0	3	0	93	0	0	0	1	1	0	13	1	0	10	1	75.61%
Shiba Inu	10	11	1	0	0	0	2	0	0	0	0	144	2	0	0	0	1	0	0	0	0	0	89.44%
Abyssinian Cat	11	0	0	0	0	0	0	0	0	0	0	0	129	0	0	0	1	0	0	0	0	1	98.47%
American Short Hair	12	0	0	0	0	0	0	0	0	0	0	0	1	148	0	0	1	0	4	0	1	0	95.48%
Birman Cat	13	1	0	0	0	0	0	0	0	0	4	0	2	1	85	1	0	1	2	10	6	0	75.22%
Cornish Rex	14	0	0	0	0	0	0	0	0	0	0	0	2	0	0	76	16	0	1	1	0	0	79.17%
Devon Rex	15	0	0	0	0	0	0	0	0	0	0	0	2	0	0	6	86	0	0	0	0	0	91.49%
Maine Coon	16	0	0	0	0	0	0	0	0	0	15	0	0	0	2	0	0	118	0	0	6	2	82.52%
Scottish Fold	17	0	0	0	0	0	0	0	0	0	1	0	0	4	1	1	2	0	130	0	2	1	91.55%
Siamese Cat	18	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	2	1	0	82	0	0	90.11%
Siberian Cat	19	0	1	0	0	0	0	0	0	0	18	0	1	0	10	0	0	8	0	1	109	3	72.19%
Somali Cat	20	0	0	0	0	0	0	0	0	0	4	0	9	1	0	0	1	6	0	2	0	70	75.27%



(a) (b)

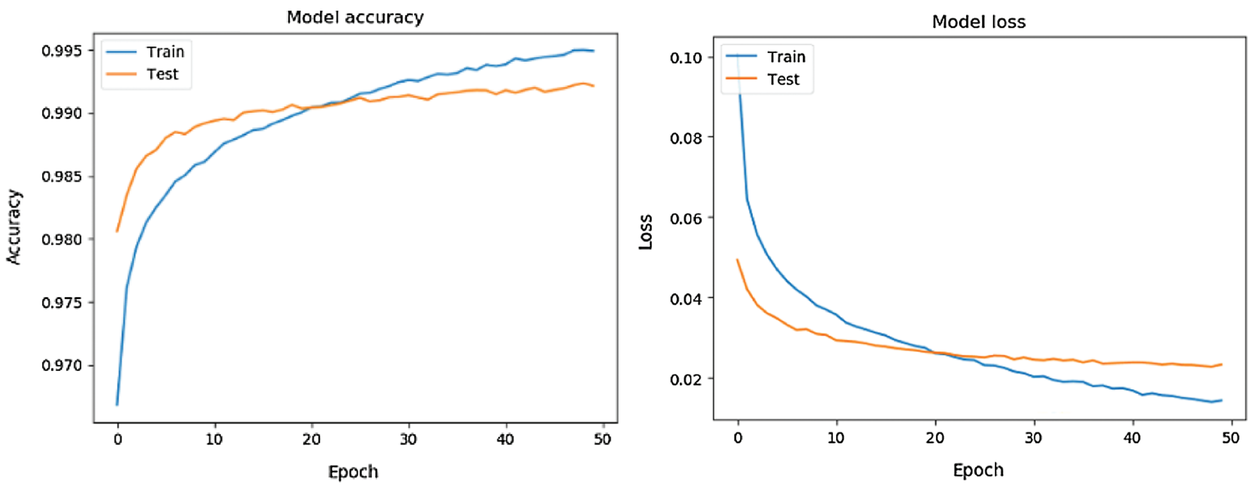
Figure 2: Accuracy and loss versus epochs when training Inception-ResNet v2 (a) Accuracy (b) Loss

Table 4: Confusion matrix of the Inception-ResNet-based CDC model

	Pred. label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Recall
Akita	0	80	2	0	0	1	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	77.67%
Alaskan Malamute	1	3	119	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	95.97%
Basenji	2	0	0	115	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.14%
Basset Hound	3	0	0	0	118	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96.72%
Beagle	4	0	0	2	3	113	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94.17%
Belgian Malinois	5	0	0	0	1	0	82	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	97.62%
Bernese Mountain Dog	6	0	0	0	0	0	0	154	1	0	0	0	0	0	0	0	0	0	0	0	0	0	99.35%
Border Collie	7	0	1	1	0	1	0	1	145	0	2	1	0	1	0	0	0	0	0	0	0	0	94.77%
Boston Terrier	8	0	0	0	0	0	0	0	1	118	0	0	0	0	0	1	0	0	0	0	0	0	98.33%
Norwegian Forest Cat	9	0	0	0	0	0	0	0	0	0	79	1	0	0	3	0	0	19	3	0	18	0	64.23%
Shiba Inu	10	3	1	1	0	0	1	0	0	0	0	154	1	0	0	0	0	0	0	0	0	0	95.65%
Abyssinian Cat	11	0	0	0	0	0	0	0	0	0	0	0	129	0	0	1	1	0	0	0	0	0	98.47%
American Short Hair	12	0	0	0	0	0	0	0	0	0	1	0	2	147	0	0	1	0	3	0	0	1	94.84%
Birman Cat	13	1	0	0	0	0	0	0	0	0	4	0	0	0	81	4	0	2	3	11	6	1	71.68%
Cornish Rex	14	0	0	0	0	0	0	0	0	0	0	0	3	1	0	78	14	0	0	0	0	0	81.25%
Devon Rex	15	0	0	0	0	0	0	0	0	1	0	0	0	1	0	13	73	0	0	5	0	1	77.66%
Maine Coon	16	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	124	1	0	9	0	86.71%
Scottish Fold	17	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	2	0	135	0	1	0	95.07%
Siamese Cat	18	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	89	0	0	97.80%
Siberian Cat	19	1	0	0	0	0	0	0	0	0	8	0	0	1	8	1	0	13	2	1	113	3	74.83%
Somali Cat	20	0	0	0	0	0	0	0	0	0	0	1	3	3	0	0	1	6	2	0	5	72	77.42%

4.3 Xception-Based CDC Model

Chollet [15] developed Xception as an improvement to Inception v3. The Xception and Inception v3 models have the same parameters, but Xception uses separable convolutions. The inputs to Xception are fixed size 299 × 299 RGB images. The training result of the fine-tuned Xception-based CDC model is shown in Fig. 3. The final training and validation accuracies of the Xception-based CDC model were 99.50% and 99.23%, respectively. This combination required 13 hours and 48 minutes for training. The confusion matrix of the Xception-based CDC model is shown in Tab. 5. The testing accuracy was 91.24%.



(a) (b)

Figure 3: Accuracy and loss versus epoch when training Xception (a) Accuracy (b) Loss

5 Model Comparison

Tab. 6 shows the training and validation accuracies for each fine-tuned CDC model. The table also includes values for the fine-tuned VGG19-based CDC model from our previous work [16]. The models we have evaluated for this paper all have higher accuracy than the VGG19-based CDC model. The Xception-based CDC model had the highest accuracy among all the models.

Tab. 7 provides a comparison of the recall values of these same four models. The Xception-based CDC model had the highest class recall among all the models with most of the breeds. However, the highest breed-specific recall values for Alaskan Malamute dog, Norwegian Forest Cat, and Devon Rex cat were achieved by the ResNet-152-based CDC model. The highest breed-specific recall values for Bernese Mountain Dog, Shiba Inu dog, Cornish Rex cat, Scottish Fold cat, and Siamese Cat were achieved by the Inception ResNet-152-based CDC model. The ResNet-152-based CDC model had the highest recall value of 100% for the Alaskan Malamute class and the lowest recall value of 72.19% for the Siberian Cat class. The Inception-ResNet-based CDC model had the highest recall value of 99.35% for the Bernese Mountain Dog class and the lowest recall value of 64.23% for the Norwegian Forest Cat class. Finally, the Xception-based CDC model had the highest recall value of 100% for the Boston Terrier dog class and the lowest recall value of 68.29% for the Norwegian Forest Cat class.

The testing accuracy for the Norwegian Forest Cat class was poor overall. We believe this is due to the similarities between the Norwegian Forest Cat and the Siberian and Maine Coon cats it was commonly misclassified as. Fig. 4 shows images of these three breeds.

Table 5: Confusion matrix of the Xception-based CDC model

	prediction Label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Recall
Akita	0	93	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	90.29%
Alaskan Malamute	1	0	123	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	99.19%
Basenji	2	0	0	115	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99.14%
Basset Hound	3	0	0	0	119	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97.54%
Beagle	4	1	0	2	2	113	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	94.17%
Belgian Malinois	5	0	0	0	1	0	83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98.81%
Bernese Mountain Dog	6	0	0	0	0	0	0	153	2	0	0	0	0	0	0	0	0	0	0	0	0	0	98.71%
Border Collie	7	1	1	0	0	1	1	1	145	0	1	2	0	0	0	0	0	0	0	0	0	0	94.77%
Boston Terrier	8	0	0	0	0	0	0	0	0	120	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
Norwegian Forest Cat	9	0	0	0	0	0	0	0	1	0	84	0	0	2	4	0	0	12	3	0	15	2	68.29%
Shiba Inu	10	6	0	1	0	0	3	0	0	0	0	151	0	0	0	0	0	0	0	0	0	0	93.79%
Abyssinian Cat	11	0	0	0	0	0	0	0	0	0	0	0	128	0	0	0	0	0	1	0	0	2	97.71%
American Short Hair	12	0	0	0	0	0	0	0	0	0	0	0	0	152	0	0	1	0	1	0	0	1	98.06%
Birman Cat	13	2	0	0	0	0	0	0	0	0	2	0	0	1	86	0	1	0	3	9	9	0	76.11%
Cornish Rex	14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	77	15	0	0	0	0	1	80.21%
Devon Rex	15	0	0	0	0	0	0	0	0	1	1	0	0	1	0	8	79	0	0	2	0	2	84.04%
Maine Coon	16	0	0	0	0	0	0	0	0	0	11	0	0	0	1	0	0	124	0	0	6	1	86.71%
Scottish Fold	17	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	1	134	0	2	1	94.37%
Siamese Cat	18	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	88	0	0	96.70%
Siberian Cat	19	0	1	0	0	0	1	0	0	0	10	0	1	1	6	0	0	6	4	0	120	1	79.47%
Somali Cat	20	0	0	0	0	0	0	0	0	0	5	0	5	2	0	0	1	1	1	0	2	76	81.72%

Table 6: Comparison of the accuracy of the CDC models

Model	Loss	Training Accuracy	Validation Loss	Validation Accuracy
VGG19	3.77%	98.59%	4.02%	98.56%
ResNet-152 v2	2.31%	99.14%	2.95%	99.00%
Inception-ResNet v2	2.74%	98.97%	2.88%	98.94%
Xception	1.44%	99.49%	2.33%	99.21%

Tab. 8 shows the training and testing times for all models. VGG19 had the shortest training time, 6 hours and 38 minutes, but the poorest accuracy. The Xception-based CDC model had the longest training time but achieved the highest accuracy and required the least amount of time to identify the breed.

We also used these combined models to identify individual dog and cat images, with results shown in Fig. 5. The prediction times are shown in Tab. 9.

Table 7: Comparison of class recall from testing data

Class	VGG19	ResNet-152 v2	Inception ResNet v2	Xception
Akita	70.87%	84.47%	77.67%	90.29%
Alaskan Malamute	91.13%	100.00%	95.97%	99.19%
Basenji	98.28%	95.69%	99.14%	99.14%
Basset Hound	89.34%	95.90%	96.72%	97.54%
Beagle	92.50%	94.17%	94.17%	94.17%
Belgian Malinois	86.90%	94.05%	97.62%	98.81%
Bernese Mountain Dog	96.13%	97.42%	99.35%	98.71%
Border Collie	92.81%	93.46%	94.77%	94.77%
Boston Terrier	95.83%	98.33%	98.33%	100.00%
Norwegian Forest Cat	68.29%	75.61%	64.23%	68.29%
Shiba Inu	83.23%	89.44%	95.65%	93.79%
Abyssinian Cat	95.49%	98.47%	98.47%	97.71%
American Short Hair	95.48%	95.48%	94.84%	98.06%
Birman Cat	68.14%	75.22%	71.68%	76.11%
Cornish Rex	78.13%	79.17%	81.25%	80.21%
Devon Rex	71.28%	91.49%	77.66%	84.04%
Maine Coon	75.52%	82.52%	86.71%	86.71%
Scottish Fold	88.03%	91.55%	95.07%	94.37%
Siamese Cat	89.01%	90.11%	97.80%	96.70%
Siberian Cat	58.94%	72.19%	74.83%	79.47%
Somali Cat	69.89%	75.27%	77.42%	81.72%
Testing Accuracy	**84.07%**	**89.31%**	**89.50%**	**91.24%**

(a) (b) (c)

Figure 4: Comparison of three cat breeds with similar appearance (a) Norwegian Forest Cat (b) Siberian Cat (c) Maine Coon

Table 8: Comparison of the training and testing times

Model	Training Time (hh:mm:ss.ss)	Testing Time (hh:mm:ss.ss)
VGG19	6:38:47.39	0:16:37.95
ResNet-152 v2	9:48:16.16	0:17:14.17
Inception-ResNet v2	12:29:48.01	0:20:28.19
Xception	13:48:13.13	0:16:19.25

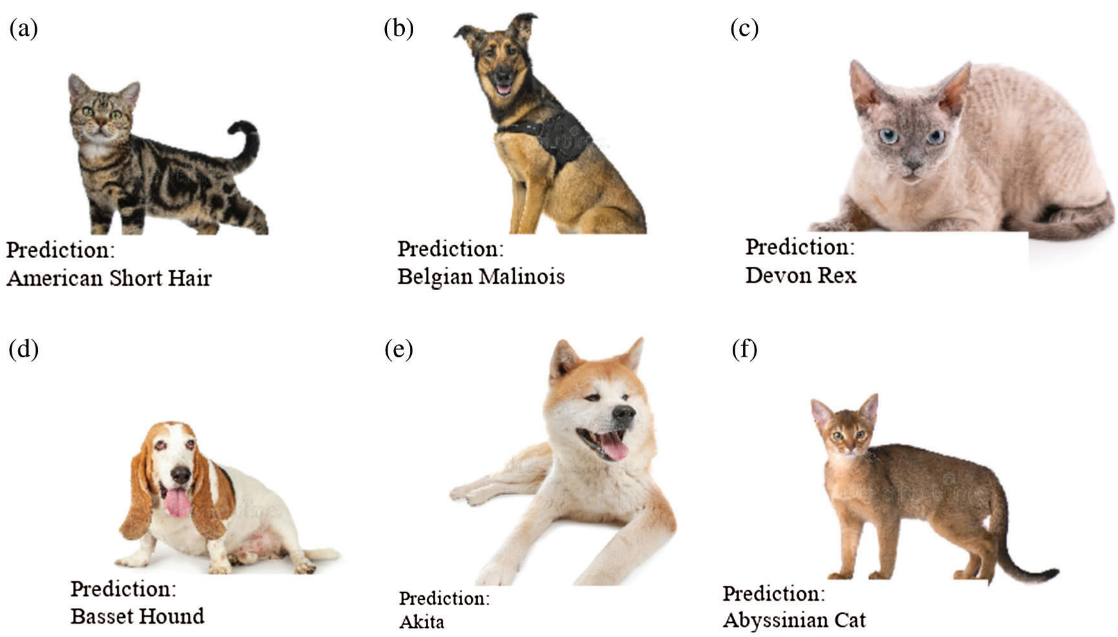


Figure 5: Single image prediction results (a) American Short Hair image (b) Belgian Malinois image (c) Devon Rex image (d) Basset Hound image (e) Akita image (F) Abyssinian Cat image

Tab. 9 shows that the Xception-based CDC was slower than the VGG19-based CDC model but faster than the other two models. Fig. 6 plots the overall performance of all models. Xception offered better accuracy than VGG19, at the cost of slower classification time.

Table 9: Time required to classify a single image

Class	VGG19	ResNet-152 v2	Inception ResNet v2	Xception
Akita	5.42 s	46.50 s	53.41 s	18.50 s
Alaskan Malamute	4.65 s	47.03 s	54.10 s	18.78 s
Basenji	5.37 s	46.03 s	54.31 s	18.68 s
Basset Hound	4.55 s	45.77 s	55.36 s	18.30 s
Beagle	5.30 s	46.23 s	54.15 s	18.15 s
Belgian Malinois	4.53 s	44.63 s	53.67 s	22.16 s
Bernese Mountain Dog	5.37 s	48.71 s	53.72 s	18.22 s

(Continued)

Table 9 (continued).

Class	VGG19	ResNet-152 v2	Inception ResNet v2	Xception
Border Collie	4.44 s	45.16 s	53.64 s	18.31 s
Boston Terrier	5.43 s	47.41 s	53.66 s	18.18 s
Norwegian Forest Cat	4.85 s	45.20 s	53.36 s	18.27 s
Shiba Inu	5.41 s	46.76 s	54.17 s	18.08 s
Abyssinian Cat	4.08 s	47.82 s	55.77 s	17.98 s
American Short Hair	4.46 s	44.83 s	52.51 s	18.02 s
Birman Cat	4.29 s	45.09 s	53.11 s	18.01 s
Cornish Rex	4.41 s	47.77 s	52.64 s	18.05 s
Devon Rex	4.50 s	45.01 s	54.46 s	18.57 s
Maine Coon	4.44 s	45.20 s	53.22 s	18.11 s
Scottish Fold	4.55 s	45.85 s	53.16 s	18.02 s
Siamese Cat	4.46 s	44.78 s	53.27 s	18.08 s
Siberian Cat	4.44 s	45.05 s	53.22 s	18.03 s
Somali Cat	4.44 s	44.97 s	52.92 s	18.16 s
Average	**4.73 s**	**45.99 s**	**53.71 s**	**18.41 s**

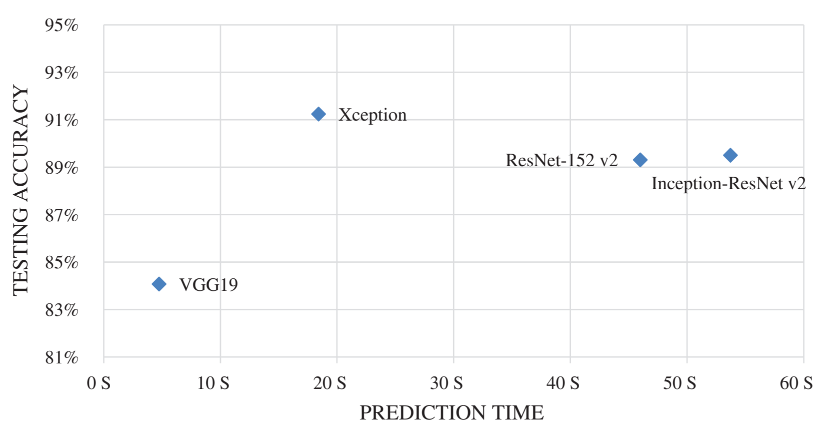


Figure 6: Overall performance plot of all models

6 Conclusion

In this paper, we have presented an image classifier for identifying the breeds of dogs and cats that incorporates fine-tuned deep learning models from Keras. Our results show that the Xception-based CDC model has the highest accuracy among those tested, with training, validation, and testing accuracies of 99.49%, 99.21%, and 99.21%, respectively. Although the training speed of the VGG19-based CDC model from our previous work is the fastest among all the models, it offers the lowest accuracy. The higher accuracy of Xception comes at a cost, with the longest training time of 13 hours and 14 minutes. Nonetheless, Xception's speed at identifying breeds was second-fastest overall, behind only the relatively inaccurate VGG19.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin *et al.,* "Large scale distributed deep networks," in *Advances in Neural Information Processing Systems*, pp. 1223–1231, 2012.

[2] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[3] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall *et al.,* "Convolutional neural networks for medical image analysis: Full training or fine tuning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[4] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, "How transferable are features in deep neural networks," in *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014.

[5] O. M. Parkhi, A. Vedaldi, A. Zisserman and C. V. Jawahar, "Cats and dogs," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Providence, RI, pp. 3498–3505, 2012.

[6] S. Panigrahi, A. Nanda and T. Swarnkar, "Deep learning approach for image classification," in *2018 2nd Int. Conf. on Data Science and Business Analytics (ICDSBA)*, Changsha, pp. 511–516, 2018.

[7] T. Jajodia and P. Garg, "Image classification - cat and dog images," *International Research Journal of Engineering and Technology*, vol. 6, no. 12, pp. 570–572, 2019.

[8] A. S. B. Reddy and D. S. Juliet, "Transfer learning with ResNet-50 for malaria cell-image classification," in *2019 Int. Conf. on Communication and Signal Processing (ICCSP)*, Chennai, India, pp. 0945–0949, 2019.

[9] W. W. Lo, X. Yang and Y. Wang, "An Xception convolutional neural network for malware classification with transfer learning," in *2019 10th IFIP Int. Conf. on New Technologies, Mobility and Security (NTMS)*, Canary Islands, Spain, pp. 1–5, 2019.

[10] G. Deng, Y. Zhao, L. Zhang, Z. Li, Y. Liu *et al.,* "Image classification and detection of cigarette combustion cone based on Inception Resnet v2," in *2020 5th Int. Conf. on Computer and Communication Systems (ICCCS)*, Shanghai, China, pp. 395–399, 2020.

[11] V. G. Buddhavarapu and A. A. Jothi J, "An experimental study on classification of thyroid histopathology images using transfer learning," *Pattern Recognition Letters*, vol. 140, pp. 1–9, 2020.

[12] Dreamtime Stock Photos, n.d. [Online]. Available: https://www.dreamstime.com/photos-images/ dreamtime.html.

[13] K. He, X. Zhang, S. Ren and J. Sun, "Identity mappings in deep residual networks," in *European conf. on computer vision*, Cham: Springer, pp. 630–645, 2016.

[14] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," Thirty-First AAAI Conf. on Artificial Intelligence, 2017. [Online]. Available: https://arxiv.org/abs/1602.07261

[15] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 1800–1807, 2017.

[16] I. H. Mahardi, K. C. Wang Lee and S. L. Chang, "Images classification of dogs and cats using fine-tuned VGG models," in *2nd IEEE Eurasia Conf. on IOT, Communication and Engineering 2020*, Yunlin, Taiwan, pp. 230–233, 2020.