ResNet50 Architecture Based Dog Breed Identification Using Deep Learning

Selvaraj S[1,*] , Thangarajan R[2], Anbukkarasi S[3] and Yuvan Prasad S[4]

[1,3] Dept. of Computer Science and Design, Kongu Engineering College, Perundurai, Tamilnadu-638060, India
[2] Dept. of Information Technology, Kongu Engineering College, Perundurai, Tamilnadu-638060, India
[4] Dept. of Computer Science and Engineering, Kongu Engineering College, Perundurai, Tamilnadu-638060, India

selvaraj.cse@kongu.edu

Abstract. Dog breed identification is one of the challenging task in animal husbandry. Knowing what a dog was bred for can assist you in predicting and comprehending its behavior. Most breeds were created with a specific goal in mind: to have the desire to achieve something specific. Dogs retain part of that drive as pets in our homes, and without a good outlet for that energy, it might manifest in undesirable ways. Developing a deep learning model, that can identify the breed of the dog with high accuracy, using this model one can easily identify the dog breed which makes easier for them to train the dog, so that we can't wait for the experts and no need for spending money for identifying the dog breed, using the mobile phones we can easily identify our dog breed simply using its image. This paper deals with an advanced classification model was designed which identifies the breed of the dog. A training dataset consisting of more than 10000 images is used, CNN is used for the classification and the accuracy is achieved using Resnet50 architecture which is highest when compared to other models.

Keywords: Dog Breed, Convolutional Neural Network, ResNet50.

1. Introduction

In recent times, convolutional neural networks (CNN) are most popular and are being widely used in the field of deep learning: image recognition [1], detection, speech recognition, data generation, etc. an important property of CNN is Scale Invariance with Feature Transform (SIFT), i.e., a CNN is capable of learning feature tensors from the examples. CNNs have gained significant popularity owing to the viability of general recurring structures for a wide range of problems. Using the combined dataset, the current research describes methods and results of hyper parameter tuning CNNs for ResNet topologies. This problem is a classification task as well as a problem of finegrained image recognition task, in which only minor differences separate two groups. CNNs are similar to Artificial Neural Networks [2], in-terms of learnable weights and biases, but contain deep layers of alternating convolutional and pooling layers. Any architecture of CNN has a few dense layers at the end. The filters, which are the learnable feature matrices are effective in image recognition and classification problems. Deep CNNs can handle enormous datasets and are even accurate in video classification on a wide scale. The Resnet50 architec-

tures finetuning procedures and learning results are discussed. This architecture can identify a dog's breed from a photograph provided by the user.

2. Literature Review

At present, approximately the population of dogs worldwide is 1 billion of which 75 % to 85% of them account for stray dogs and wild dogs [3]. Identifying the breed of dogs automatically has been discussed in [4] where two image processing approaches to classify the dog breeds are presented. One of the approach uses the conventional Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG) algorithms. The other approach employs deep learning using convolutional neural networks (CNN) with transfer learning. It is found that CNN model performs better than the conventional method.

In [5], The author has proposed Principal Component Analysis (PCA) to achieve dimensionality reduction and feature selection. After PCA, a classifier is used to classify dog breed based on the dog face images where the template matching technique is applied.

In [6], the author proposed a novel approach to finegrained image classification for dog breed identification and shown that extracting corresponding parts improves classification performance. The author has achieved 67% recognition rate on a large real-world dataset including 133 dog breeds and 8,351 images. The author used the datasets are referenced in [7], [8].

Machine learning based approach [9] has presented by the author to identify images of invasive hydrangea. And also the author has practiced a deep learning technique that extensively applied to image recognition was used.

The work which is carried out by the author in [10], convolutional neural network framework has used to train and categorize dog breeds. And also the author has approached this CNN algorithm based on LeNet [11] and GoogLeNet [12] architectures.
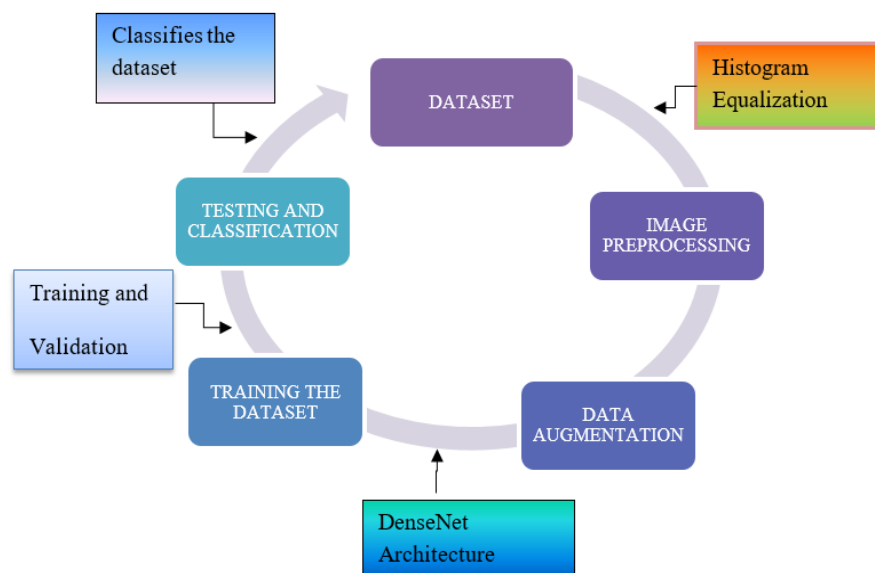
3. System Architecture

Figure 1. Proposed.

This paper is about identifying Dog Breed Classification using ResNet approach, the moles from the images is easily extracted by the convolutional layers. The model can easily identify whether which type of breed is dog. Here we classify 120 classes of Dog breed. After the model trained the comparison of Dog breed image and trained model takes place to predict the Breed, the proposed system is shown in Fig 1.

4. CNN Architecture

CNN is a kind of deep neural network which employes filters to learn image features. CNN when compared to deep neural networks have lesser parameters to learn and mimic the processing of visual cortext in the huma brain. Because the novel architecture, CNNs have achieved near huma level accuracy on image processing task. The basic structure of CNN [13] includes alternating layers of convolutional and pooling as shown in Fig 2. The pooling layers help reduce the number of parameters to be learned from
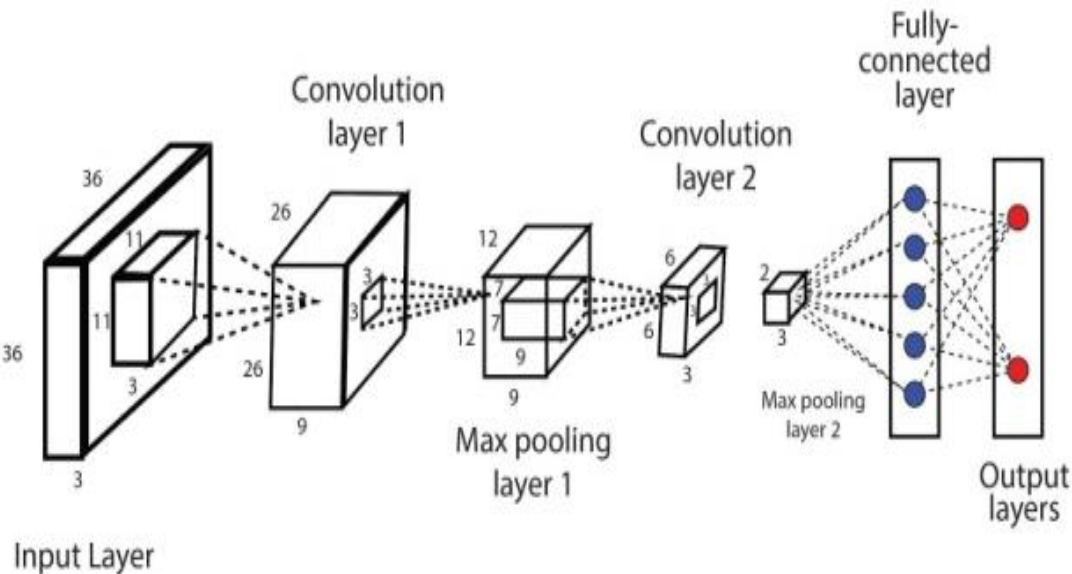

Figure 2. CNN Architecture.

samples. The first layer of any CNN is the convolution layer. The convolution layer extracts features from an input image using its filters [14]. The filters learn the intrinsic features present in the images by computing the convolution of the filter with the subset of the image [15]. Each convolutional layer produces a feature map. A feature map represents the local weighted sums by convoluting a portion of the image with the filter. A different kinds of activation functions are employed in the hidden layers. Among them, Rectified Linear Unit (ReLu) activation function is prepfered as it increase the nonlinearity [16].

At the end stage of CNN, the output of the convolution layers are flattened by means of 1×1 convolution layer and fed into dense layers. Depending on the task differnet output layers are used. For example, in classsification task sigmoid which is also called the logistics function is used [17]. For categorical clasification, softmax layers are used.

5. ResNet50 Architecture

ResNet50's architecture[18] is divided into four stages, as depicted in Fig 3. Generally, ResNet50 design uses a 7×7 and a 3×3 kernel (filter) sizes for initial convolution and max pooling, respectively. The ResNet architecture has 4 stages of concolutional and pooling layers. The identity relationship is addressed by the bended bolts. The ran interfacing bolt shows that the convolution activity in the remaining Square is directed with step 2, bringing about a half size input as far as stature and expansiveness however a multiplying channel width. It is multiplied and the size of the information is diminished to half as we advance starting with one phase then onto the next.

Bottleneck architecture[19] is utilised for deeper networks like ResNet50, ResNet152, and so on. Three layers are placed one on top of the other for each residual function F. Convolutions are used in

the first, third, and eleventh layers. The 11 convolution layers are in charge of shrinking and then expanding the dimensions. With lower input/output dimensions, the 33 layer is left as a bottleneck [20]. Finally, the network has an Average Pooling layer, which is followed by a 1000 neuron fully linked layer, as shown in the Fig 3.
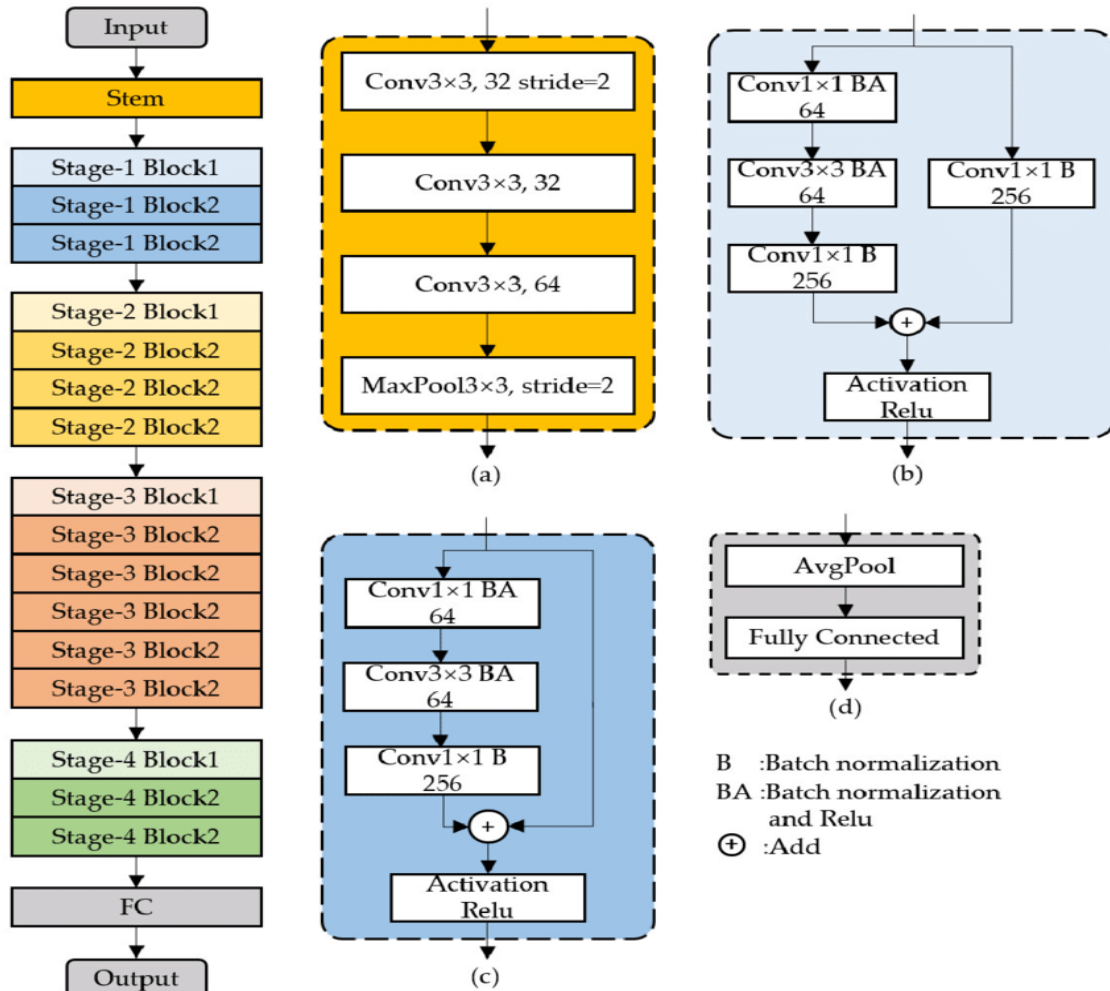

Figure 3. Architecture of Resnet50.

6. Experimental Setup

6.1. Experimental Environment

In this work, Python a high level general purpose programming language that is interpreted, interactive, object oriented, and general purpose is used. Python is open source under the GNU General Public License (GPL) [21]. Python has a dynamic type system and memory management that is automated. Since machine learning and deep learning tasks are resource hungry, it takes huge time to train the models. Therefore, google has come up with the GoogleColab facility on its cloud. Researchers can utilize this facility to train and test their models It has a similar user interface to the anaconda jupyter notebook. To speed up the execution time, the execution might be done in RAM or on the GPU.

6.2. Dataset

Our dataset consists of 10222 images of 120 dog breeds. A sample image is shown in Fig 4. This is the Stanford Dog breed dataset; greater the dataset leads to the better accuracy [22] so we use the dataset of more than 10000 images. It is separated into two which is test and train dataset. Both collections include
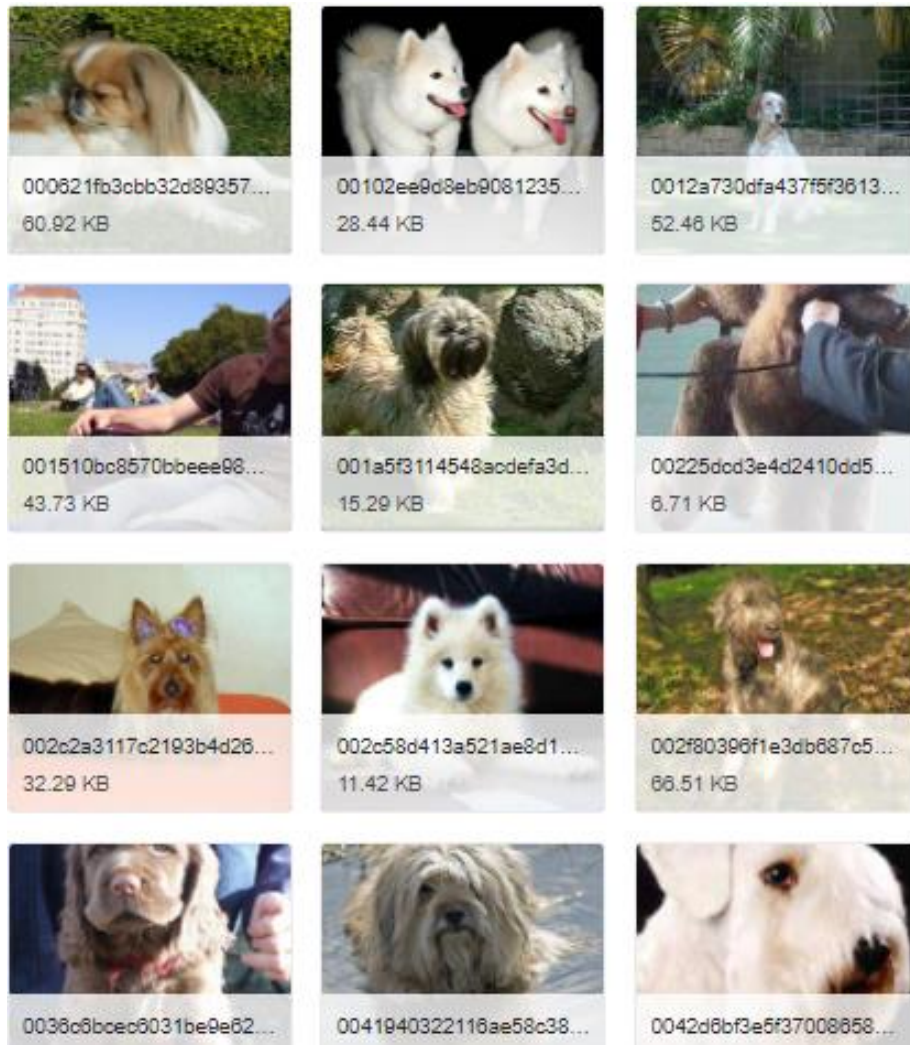

Figure 4. Dataset.

images of various sizes, with each image bearing a label identifying the dog breed.

6.3. Procedure
Our image detection model was done in the platform named google Colab with the help of RESNET model. We have imported all the images to the google Colab using drive [23][24]. We defined ResNet model then starts preprocessing of each images [25]. After preprocessing, the training and validation is carried out as selected in the training regime [26].

7. Result Analysis

7.1 Loss and Accuracy
During the training process several iterations are carried out in which the loss and accuracy is calculated [27]. With each iteration a loss tends to decrease and accuracy increases [28].
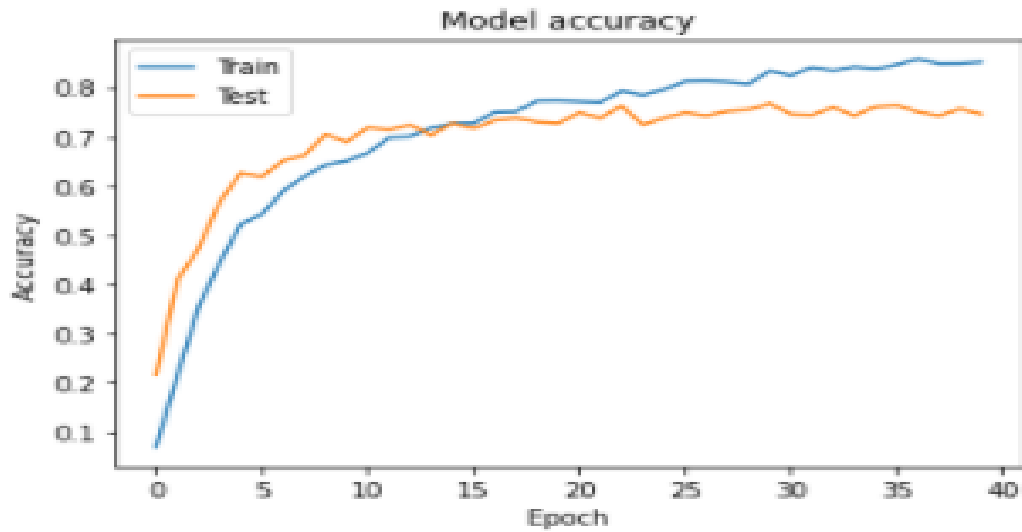
Figure 5. Accuracy Graph.

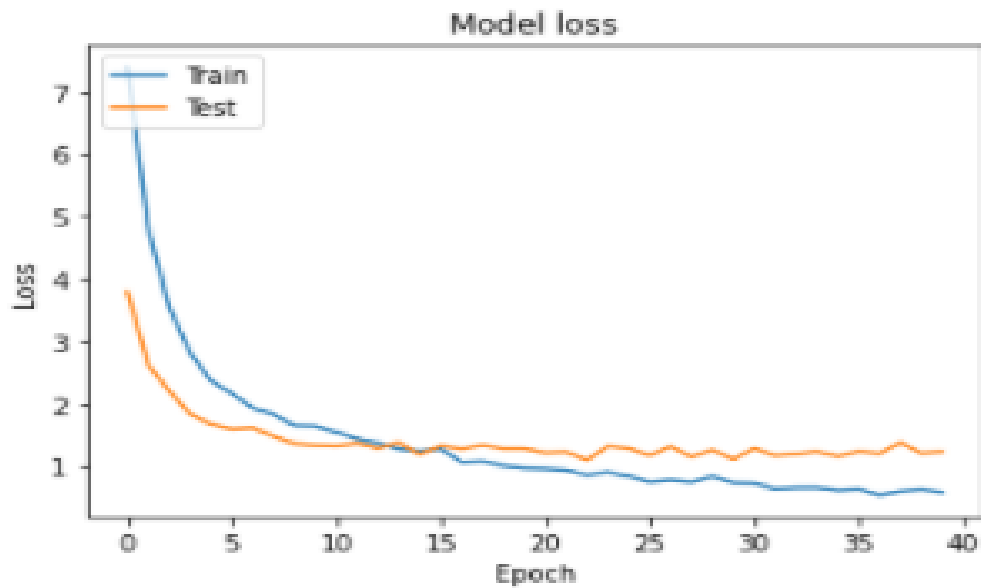Fig 5 represents the accuracy comparison for each epochs, here the accuracy increases for each epoch.

Figure 6. Loss Graph.

Fig 6 represents the loss comparison for each epochs, here the loss decreases for each epoch.

Depending on the training / validation / testing regime, training accuracy and validation accuracy are compared. If training accuracy is high and validation accuracy is low, the system is said to be having high variance. On the other hand, if training accuracy itself is less then, the system is said to be biase..On the training and test datasets, the trained models are tested every 10 minutes. On the training and test datasets, the accuracy is tracked; this statistic shows the mean percentage of correctly categorized classes on a dataset. The Resnet50 models gives the accuracy of 85.15% for the Stanford dataset, which is comparatively higher than the other models. During execution, the accuracy and loss are recorded for

every epoch. It is found that the accuracy starts with 7% and increases to 93% gradually at the 40th epoch.

7.2. Comparison with Other Models

The Resnet50 model gives the better accuracy when compared with other architectures like Alexnet, Googlenet, Nasnet-A and Resnet18. The maximum accuracy achieved by the former is 80% by Nasnet-A, which is also considerably low when compared to Resnet50 which is 93%. The performance of each model is shown in the Table 1. The performance Graph in the Fig 7 shows how clearly Resnet50 out performs other models.

Table 1. Comparison Table.

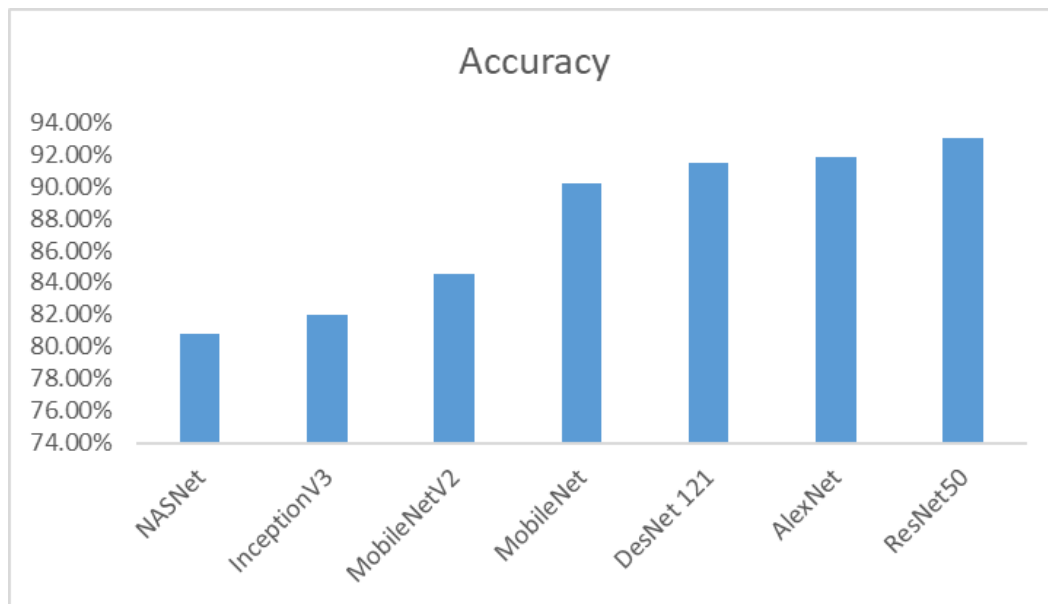| Methods	Recognition Speed
NASNet	80.82%
InceptionV3	82.00%
MobileNetV2	84.58%
MobileNet	90.17%
DesNet 121	91.51%
AlexNet	91.80%
ResNet50	93.00%

Figure 7. Performance Graph.

8. Conclusion

A large number of labelled images is needed to build CNN. To solve this main difficulty, deep learning and image augmentation are applied to a pre-trained ResNet50. It has an accuracy of 85 percent throughout training. Normal deep neural network creates vanishing gradient problem. To overcome this, Resnet50 uses the skip connection. The training was done with the help of RELU and softmax activation. After the model trained the comparison of the new test image and trained model takes place to predict

the Breed. In addition, future effort will include training the model with diverse architecture with better accuracy.

References

[1] Wu, Ren, et al. "Deep image: Scaling up image recognition." arXiv preprint arXiv:1501.02876 7.8 (2015).

[2] Jain, Anil K., Jianchang Mao, and K. Moidin Mohiuddin. "Artificial neural networks: A tutorial." Computer 29.3 (1996): 31-44.

[3] How Many Dogs Are There In The World?, Mar 2019, [online] Available: https://www.worldatlas.com/articles/how-many-dogs-are-there-in-theworld.html.

[4] P. Borwarnginn, K. Thongkanchorn, S. Kanchanapreechakorn and W. Kusakunniran, "Breakthrough Conventional Based Approach for Dog Breed Classification Using CNN with Transfer Learning", 2019 11th Int. Conf. Inf. Technol. Electr. Eng. ICITEE 2019, vol. 7, pp. 1-5, 2019.

[5] M. Chanvichitkul, P. Kumhom and K. Chamnongthai, "Face recognition based dog breed classification using coarse-to-fine concept and PCA", 2007 Asia-Pacific Conf. Commun. APCC, pp. 25-29, 2007.

[6] J. Liu, A. Kanazawa, D. Jacobs and P. Belhumeur, "Dog Breed Classification Using Part Localization", European Conference on Computer Vision, pp. 172-185, 2012.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database", 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.

[8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., "ImageNet Large Scale Visual Recognition Challenge", International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, Dec. 2015.

[9] B. A. M. Ashqar and S. S. Abu-Naser, "Identifying Images of Invasive Hydrangea Using Pre-Trained Deep Convolutional Neural Networks", Int. J. Control Autom, vol. 12, no. 4, pp. 15-28, 2019.

[10] Hsu, David. "Using convolutional neural networks to classify dog breeds." CS231n: Convolutional Neural Networks for Visual Recognition [course webpage] 2 (2015)

[11] Günel, Mehmet. "GoogLeNet." (2016).

[12] Sathishkumar V E, Changsun Shin, Youngyun Cho, "Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city", Building Research & Information, Vol. 49. no. 1, pp. 127-143, 2021.

[13] Sathishkumar V E, Youngyun Cho, "A rule-based model for Seoul Bike sharing demand prediction using Weather data", European Journal of Remote Sensing, Vol. 52, no. 1, pp. 166-183, 2020.

[14] Sathishkumar V E, Jangwoo Park, Youngyun Cho, "Seoul Bike Trip duration prediction using data mining techniques", IET Intelligent Transport Systems, Vol. 14, no. 11, pp. 1465-1474, 2020.

[15] Sathishkumar V E, Jangwoo Park, Youngyun Cho, "Using data mining techniques for bike sharing demand prediction in Metropolitan city", Computer Communications, Vol. 153, pp. 353-366, 2020.

[16] Sathishkumar V E, Yongyun Cho, "Season wise bike sharing demand analysis using random forest algorithm", Computational Intelligence, pp. 1-26, 2020.

[17] Sathishkumar, V. E., Wesam Atef Hatamleh, Abeer Ali Alnuaim, Mohamed Abdelhady, B. Venkatesh, and S. Santhoshkumar. "Secure Dynamic Group Data Sharing in Semi-trusted Third Party Cloud Environment." Arabian Journal for Science and Engineering (2021): 1-9.

[18] Chen, J., Shi, W., Wang, X., Pandian, S., & Sathishkumar, V. E. (2021). Workforce optimisation for improving customer experience in urban transportation using heuristic mathematical model. International Journal of Shipping and Transport Logistics, 13(5), 538-553.

[19] Pavithra, E., Janakiramaiah, B., Narasimha Prasad, L. V., Deepa, D., Jayapandian, N., & Sathishkumar, V. E., Visiting Indian Hospitals Before, During and After Covid. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 30 (1), 111-123, 2022.

[20] Easwaramoorthy, S., Moorthy, U., Kumar, C. A., Bhushan, S. B., & Sadagopan, V. (2017, January). Content based image retrieval with enhanced privacy in cloud using apache spark. In International Conference on Data Science Analytics and Applications (pp. 114-128). Springer, Singapore.

[21] Sathishkumar, V. E., Agrawal, P., Park, J., & Cho, Y. (2020, April). Bike Sharing Demand Prediction Using Multiheaded Convolution Neural Networks. In Basic & Clinical Pharmacology & Toxicology (Vol. 126, pp. 264-265). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[22] Subramanian, M., Shanmuga Vadivel, K., Hatamleh, W. A., Alnuaim, A. A., Abdelhady, M., & VE, S. (2021). The role of contemporary digital tools and technologies in Covid‑19 crisis: An exploratory analysis. Expert systems.

[23] Sathishkumar, V. E., Rahman, A. B. M., Park, J., Shin, C., & Cho, Y. (2020, April). Using machine learning algorithms for fruit disease classification. In Basic & clinical pharmacology & toxicology (Vol. 126, pp. 253-253). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[24] Sathishkumar, V. E., & Cho, Y. (2019, December). Cardiovascular disease analysis and risk assessment using correlation based intelligent system. In Basic & clinical pharmacology & toxicology (Vol. 125, pp. 61-61). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[25] Easwaramoorthy, S., Sophia, F., & Prathik, A. (2016, February). Biometric Authentication using finger nails. In 2016 international conference on emerging trends in engineering, technology and science (ICETETS) (pp. 1-6). IEEE.

[26] Babu, J. C., Kumar, M. S., Jayagopal, P., Sathishkumar, V. E., Rajendran, S., Kumar, S., ... & Mahseena, A. M. (2022). IoT-Based Intelligent System for Internal Crack Detection in Building Blocks. Journal of Nanomaterials, 2022.

[27] Liu, Y., Sathishkumar, V. E., & Manickam, A. (2022). Augmented reality technology based on school physical education training. Computers and Electrical Engineering, 99, 107807.

[28] Rajalaxmi, R. R., Narasimha Prasad, L. V., Janakiramaiah, B., Pavankumar, C. S., Neelima, N., & Sathishkumar, V. E. (2022). Optimizing Hyperparameters and Performance Analysis of LSTM Model in Detecting Fake News on Social media. Transactions on Asian and Low-Resource Language Information Processing.